



SENSORCOMM 2014

The Eighth International Conference on Sensor Technologies and Applications

ISBN: 978-1-61208-374-2

November 16 - 20, 2014

Lisbon, Portugal

SENSORCOMM 2014 Editors

Tapio Saarelainen, National Defence University, Finland

Reinhardt Karnapke, BTU Cottbus, Germany

Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2014

Foreword

The Eighth International Conference on Sensor Technologies and Applications (SENSORCOMM 2014), held between November 16-20, 2014 in Lisbon, Portugal, continued a series of events covering related topics on theory and practice on wired and wireless sensors and sensor networks.

Sensors and sensor networks have become a highly active research area because of their potential of providing diverse services to broad range of applications, not only on science and engineering, but equally importantly on issues related to critical infrastructure protection and security, health care, the environment, energy, food safety, and the potential impact on the quality of all areas of life.

Sensor networks and sensor-based systems support many applications today on the ground. Underwater operations and applications are quite limited by comparison. Most applications refer to remotely controlled submersibles and wide-area data collection systems at a coarse granularity.

Underwater sensor networks have many potential applications such a seismic imaging of undersea oilfields as a representative application. Oceanographic research is also based on the advances in underwater data collection systems.

There are specific technical aspects to realize underwater applications which can not be borrowed from the ground-based sensors net research. Radio is not suitable for underwater systems because of extremely limited propagation. Acoustic telemetry could be used in underwater communication; however off-the-shelf acoustic modems are not recommended for underwater sensor networks with hundreds of nodes because they were designed for long-range and expensive. As the speed of light (radio) is five orders of magnitude higher than the speed of sound, there are fundamental implications of time synchronization and propagation delays for localization. Additionally, existing communication protocols are not designed to deal with long sleep times and they can't shut down and quickly restart.

In wireless sensor and micro-sensor networks, energy consumption is a key factor for the sensor lifetime and accuracy of information. Protocols and mechanisms have been proposed for energy optimization considering various communication factors and types of applications. Conserving energy and optimizing energy consumption are challenges in wireless sensor networks, requiring energy-adaptive protocols, self-organization, and balanced forwarding mechanisms.

We take here the opportunity to warmly thank all the members of the SENSORCOMM 2014 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to SENSORCOMM 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the SENSORCOMM 2014

organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that SENSORCOMM 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of sensor technologies and applications.

We are convinced that the participants found the event useful and communications very open. We hope Lisbon provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

SENSORCOMM 2014 Chairs:

SENSORCOMM Advisory Chairs

Jean Philippe Vasseur, Cisco Systems, Inc., France

Petre Dini, Concordia University, Canada / China Space Agency Center, China

Jaime Lloret Mauri, Polytechnic University of Valencia, Spain

Jens Martin Hovem, Norwegian University of Science and Technology, Norway

Pascal Lorenz, University of Haute Alsace, France

Sergey Yurish, IFSA, Spain

SENSORCOMM 2014 Industry Liaison Chairs

Sarfraz Khokhar, Cisco Systems, Inc., USA

Harkirat Singh, Samsung Electronics Co., Korea

Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain

Michael Niedermayer, Fraunhofer IZM, Germany

SENSORCOMM 2014 Research/Industry Chairs

Hristo Djidjev, Los Alamos National Laboratory, USA Teng Rui, National Institute of Information and Communication Technology, Japan

S. Biju Kumar, Philips Research - Eindhoven, The Netherlands

SENSORCOMM 2014 Special Area Chairs

Embedded systems

Joshua Ellul, Imperial College, London, UK

Security

Yenumula Reddy, Grambling State University, USA

Body networks

Alessandro Pozzebo, Università degli Studi di Siena, Italy

Underwater systems

Mylène Toulgoat, Communications Research Centre - Ottawa, Canada

Applications

Elena Gaura, Coventry University, UK

Performance

Canfeng Chen, Nokia Research Center - Beijing, China

Atmospheric Icing and Sensing

Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2014

Committee

SENSORCOMM Advisory Chairs

Jean Philippe Vasseur, Cisco Systems, Inc., France
Petre Dini, Concordia University, Canada / China Space Agency Center, China
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Jens Martin Hovem, Norwegian University of Science and Technology, Norway
Pascal Lorenz, University of Haute Alsace, France
Sergey Yurish, IFSA, Spain

SENSORCOMM 2014 Industry Liaison Chairs

Sarfraz Khokhar, Cisco Systems, Inc., USA
Harkirat Singh, Samsung Electronics Co., Korea
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Michael Niedermayer, Fraunhofer IZM, Germany

SENSORCOMM 2014 Research/Industry Chairs

Hristo Djidjev, Los Alamos National Laboratory, USA
Teng Rui, National Institute of Information and Communication Technology, Japan
S. Biju Kumar, Philips Research - Eindhoven, The Netherlands

SENSORCOMM 2014 Special Area Chairs

Embedded systems

Joshua Ellul, University of Malta, Malta

Security

Yenumula Reddy, Grambling State University, USA

Body networks

Alessandro Pozzebo, Università degli Studi di Siena, Italy

Underwater systems

Mylène Toulgoat, Communications Research Centre - Ottawa, Canada

Applications

Elena Gaura, Coventry University, UK

Performance

Canfeng Chen, Nokia Research Center - Beijing, China

Atmospheric Icing and Sensing

Muhammad Shakeel Virk, Narvik University College, Norway

SENSORCOMM 2014 Technical Program Committee

Saied Abedi, Fujitsu Laboratories of Europe LTD. - Middlesex, UK
Abdalrahman Al-Qubaa, Newcastle University, UK
Mothanna Alkubaily, Université de Technologie de Compiègne, France
Boushra Alkubily, Université de Technologie de Compiègne, France
Maykel Alonso Arce, CEIT and Tecnun (University of Navarra), Spain
Tariq Alsboui, Manchester Metropolitan University, UK
Adil Al-Yasiri, University of Salford, UK
Isabelle Augé-Blum, INSA Lyon - Laboratoire CITI -Villeurbanne, France
Reza Azarderakhsh, The University of Western Ontario, Canada
Sebastian Bader, Mid Sweden University, Sweden
Valentina Baljak, National Institute of Informatics & University of Tokyo, Japan
Dominique Barthel, Orange Labs Division R&D - Meylan, France
Novella Bartolini, "Sapienza" University of Rome, Italy
Majid Bayanil Abbasy, Universidad Nacional de Costa Rica, Costa Rica
Rezaul K Begg, Victoria University, Australia
Paolo Bellavista, University of Bologna, Italy
Stephan Benecke, Technische Universität Berlin | Fraunhofer Institut Zuverlässigkeit und Mikrointegration, Germany
Ali Berrached, University of Houston-Downtown, USA
Karabi Biswas, Indian Institute of Technology - Kharagpur, India
Alessandro Bogliolo, University of Urbino, Italy
David Boyle, Imperial College London, UK
Lina Brito, University of Madeira, Portugal
Ioannis Broustis, AT&T Labs Research, U.S.A
Tiziana Calamoneri, "La Sapienza" Università di Roma, Italy
Maria-Dolores Cano Baños, Technical University of Cartagena, Spain
Juan Vicente Capella Hernández, Universidad Politécnica de Valencia, Spain
Alexandru Caracas, IBM Research - Zurich, Switzerland
Berta Carballido Villaverde, United Technologies Research Centre, Ireland
Amitava Chatterjee, Jadavpur University, India
Canfeng Chen, Nokia Research Center - Beijing, China
Shu-Ching Chen, Florida International University - Miami, USA
Harsha Chenji, University of Texas at Dallas, USA
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Daniel Curiac, "Politehnica" University of Timisoara, Romania
David Cuartielles, Malmö University, Sweden
Juarez Bento da Silva, Universidade Federal de Santa Catarina, Brazil
Debabrata Das, International Institute of Information Technology - Bangalore, India
Danco Davcev, University for Information Science & Technology "St. Paul the Apostle" - Ohrid, Republic of Macedonia
Javier Del Ser Lorente, TECNALIA-Telecom - Zamudio (Bizkaia), Spain
Jerker Delsing, Lulea University of Technology, Sweden
Behnam Dezfouli, University Technology Malaysia (UTM), Malaysia

Akshaye Dhawan, Ursinus College, USA
Vincenzo Di Lecce, Politecnico di Bari, Italy
Mari Carmen Domingo, Barcelona Tech University, Spain
Wan Du, Nanyang Technological University (NTU), Singapore
Juan Carlos Dueñas López, Universidad Politecnica de Madrid, Spain
Sylvain Durand, LIRMM/Université Montpellier II, France
Imad H. Elhadj, American University of Beirut, Lebanon
Joshua Ellul, University of Malta, Malta
Xiang Fei, Coventry University, UK
Sándor Fekete, Braunschweig Institute of Technology, Germany
Paulo Felisberto, Institut for Systems and Robotics-Lisbon / Universidade do Algarve, Portugal
Gianluigi Ferrari, University of Parma, Italy
Armando Ferro Vázquez, Universidad del País Vasco - Bilbao, Spain
Paul Fortier, University of Massachusetts Dartmouth, USA
Leonardo Franco, Universidad de Malaga, Spain
Pedro Furtado, University of Coimbra, Portugal
Miguel Garcia Pineda, Polytechnic University of Valencia, Spain
Elena Gaura, Coventry University, UK
Hamid Gharavi, National Institute of Standards and Technology (NIST) - Gaithersburg, USA
Chris Gniady, University of Arizona, USA
Stephane Grumbach, INRIA, France
Jayavardhana Gubbi, University of Melbourne, Australia
Jianlin Guo, Mitsubishi Electric Research Laboratories - Cambridge, USA
Malka N. Halgamuge, The University of Melbourne, Australia
Mohammad Hammoudeh, Manchester Metropolitan University, UK
Vincent Huang, Ericsson Research - Stockholm, Sweden
Muhammad Ali Imran, University of Surrey, U.K.
Abhaya Induruwa, Canterbury Christ Church University, UK
Vasanth Iyer, International Institute of Information Technology, India
Shaghayegh Jaber, Islamic Azad University - Tehran, Iran
Imad Jawhar, United Arab Emirates University - Al Ain, UAE
Zhen Jiang, West Chester University, USA
Miao Jin, University of Louisiana at Lafayette, U.S.A.
Adrian Kacso, University of Siegen, Germany
Aravind Kailas, University of North Carolina - Charlotte, USA
Kyoung-Don Kang, Binghamton University, USA
Riad Kanan, The Institute of Engineering Sciences, Switzerland
Reinhardt Karnapke, Brandenburg University of Technology Cottbus – Senftenberg, Germany
Dimitrios A. Karras, Chalkis Institute of Technology, Hellas
Fotis Kerasiotis, University of Patras / Rio-Patras, Greece
Yaser Khamayseh, Jordan University of Science and Technology, Jordan
Abdelmajid Khelil, Huawei European Research Center, Germany
Sarfraz Khokhar, Cisco Systems Inc., USA
Kwangsoo Kim, Electronics and Telecommunications Research Institute (ETRI), Korea
Boris Kovalerchuk, Central Washington University, USA
Thorsten Kramp, IBM Research Zurich, Switzerland
Evangelos Kranakis, Carleton University, Canada
Srđjan Krčo, Ericsson Research, Ireland

Dilip Krishnaswamy, Qualcomm Research - San Diego, U.S.A.
Danny Krizanc, Wesleyan University - Middletown, USA
Sisil Kumarawadu, University of Moratuwa, Sri Lanka
Erlend Larsen, The Norwegian Defence Research Establishment (FFI) - Kjeller, Norway
Seongsoo Lee, Soongsil University - Seoul, Korea
Pierre Leone, University of Geneva, Switzerland
Jacek Lewandowski, Coventry University, UK
Xiuqi Li, University of North Carolina at Pembroke, U.S.A
Chiu-Kuo Liang, Chung Hua University - Hsinchu, Taiwan
Qilian Liang, University of Texas at Arlington, USA
Weifa Liang, Australian National University - Canberra, Australia
Chun-Lung Lin, Industrial Technology Research Institute (ITRI) Hsinchu, Taiwan
Thomas Lindh, STH/KTH - Stockholm, Sweden
André Luiz Lins de Aquino, Federal University of Ouro Preto, Brazil
Aihua Liu, Qingdao Institute of Bioenergy & Bioprocess Technology - Chinese Academy of Sciences, China
Hai Liu, Hong Kong Baptist University, Hong Kong
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain
Abdallah Makhoul, Université de Besancon - Belfort, France
Gianfranco Manes, University of Florence, Italy
Vladimir Marbukh, NIST, USA
Andrew Markham, University of Oxford, UK
José Ramiro Martínez de Dios, University of Seville, Spain
Francisco Martins, University of Lisbon, Portugal
Alireza Masoum, Twente University, The Netherlands
Kovatsch Matthias, ETH Zurich, Switzerland
Natarajan Meghanathan, Jackson State University, USA
Fabien Mieyeville, Ecole Centrale Lyon - INL (Institute of Nanotechnology of Lyon), France
Nader Faisal Jaafar Mohamed, UAEU, UAE
Jose M. Moya, Universidad Politécnica de Madrid, Spain
Mohammad Mozumdar, California State University, Long Beach, USA
Abderrahmen Mtibaa, Texas A&M University, U.S.A.
Umair Najeed Mughal, Narvik University College, Norway
Mohamad Y. Mustafa, Narvik University College, Norway
Deok Hee Nam, Wilberforce University, USA
Enrico Natalizio, INRIA Lille - Nord Europe, France
Mahmuda Naznin, Bangladesh University of Engineering and Technology - Dhaka, Bangladesh
Sarmistha Neogy, Jadavpur University, India
Edith C.-H. Ngai, Uppsala University, Sweden
Michael Niedermayer, Fraunhofer Institute for Reliability and Microintegration, Germany
Gregory O'Hare, University College Dublin (UCD), Ireland
Brendan O'Flynn, Tyndall National Institute/University College Cork, Ireland
Cyril Onwubiko, Research Series Ltd., UK
Knut Øvsthus, Bergen University College, Norway
Carlos Enrique Palau Salvador, Universidad Politecnica de Valencia, Spain
Sung-Joon Park, Gangneung-Wonju National University, South Korea
Lorena Parra Boronat, Universitat Politècnica de Valencia (UPV), Spain
Eros Pasero, Politecnico di Torino, Italy

Leonidas Perlepes, University of Thessaly, Greece
Dirk Pesch, Cork institute of Technology, Ireland
Patrick Pons, CNRS-LAAS, France
Miodrag Potkonjak, University of California - Los Angeles, USA
Shrisha Rao, International Institute of Information Technology - Bangalore, India
Shahid Raza, Swedish Institute of Computer Science (SICS) – Stockholm, Sweden
Yenumula Reddy, Grambling State University, USA
Mark Reed, Yale School of Engineering & Applied Science, USA
Juan Reig Pascual, Polytechnic University of Valencia, Spain
Càndid Reig, University of Valencia, Spain
Tor Arne Reinen, SINTEF ICT, Norway
Biljana Risteska Stojkoska, University "Ss. Cyril and Methodius", Macedonia
Joel Rodrigues, University of Beira Interior, Portugal
Laura Rodríguez, Universidad Popular Autónoma de Puebla (UPAEP), Mexico
Lorenzo Rubio-Arjona, Universidad Politécnica de Valencia, Spain
Ulrich Rückert, Bielefeld University, Germany
Teng Rui, National Institute of Information and Communication Technology, Japan
Jorge Sá Silva, University of Coimbra, Portugal
Sicari Sabrina, University of Insubria, Italy
Husnain Saeed, National University of Sciences & Technology (NUST), Pakistan
Addisson Salazar, Polytechnic University of Valencia, Spain
Ioakeim Samaras, Aristotle University of Thessaloniki, Greece
Francisco Javier Sánchez Bolumar, ADIF, Spain
Olga Saukh, ETH Zurich, Switzerland
Kamran Sayrafian, National Institute of Standards & Technology, USA
Elad Michael Schiller, Chalmers University of Technology, Sweden
Leo Selavo, University of Latvia, Latvia
Sandra Sendra Compte, Polytechnic University of Valencia, Spain
Kuei-Ping Shih, Tamkang University - Taipei, Taiwan
Simone Silvestri, Sapienza University of Rome, Italy
Francesco Simeone, University "Sapienza" of Roma / INFN-Roma, Italy
Jasvinder Singh, Cork Institute of Technology, Ireland
Andrzej Skowron, University of Warsaw, Poland
K M Sunjiv Soyjaudah, University of Mauritius, Mauritius
Arvind K. Srivastava, NanoSonix Inc. - Skokie, USA
Grigore Stamatescu, University Politehnica of Bucharest, Romania
Yannis Stamatiou, University of Patras, Greece
Razvan Stanica, National Polytechnic Institute of Toulouse, France
Kris Steenhaut, Vrije Universiteit Brussel, Belgium
Julinda Stefa, Sapienza University of Rome, Italy
David Stork, Rambus Labs, USA
David Sundaram, University of Auckland, New Zealand
Ewa Szyrkiewicz, Warsaw University of Technology, Poland
Zahra Taghikhaki, University of Twente, the Netherlands
Muhammad Tariq, Waseda University - Tokyo, Japan
Kerry Taylor, CSIRO Canberra, Australia
Lothar Thiele, ETH Zurich, Switzerland
Rolf Thomasius, Technische Universität Berlin, Germany

Bin Tong, Microsoft Corp. - Redmond, USA
Vassilis Triantafillou, Technological Educational Institution of Messolonghi, Greece
Neeta Trivedi, Aeronautical Development Establishment- Bangalore, India
Wilfried Uhring, University of Strasbourg, France
Fabrice Valois, INRIA SWING / CITI, INSA-Lyon, France
Jean Philippe Vasseur, Cisco Systems, Inc., France
Roberto Verdone, Università di Bologna, Italy
Manuela Vieira, UNINOVA/ISEL, Portugal
Muhammad Shakeel Virk, Narvik University College, Norway
Michael Walsh, Tyndall National Institute, Ireland
Wenwu Wang, University of Surrey, UK
You-Chiun Wang, National Chiao-Tung University, Taiwan
Zhibo Wang, University of Tennessee, USA
Chih-Yu Wen, National Chung Hsing University - Taichung, Taiwan
Andreas Wombacher, University of Twente, Netherlands
Fang-Jing Wu, National Chiao Tung University, Taiwan
Hui Wu, University of New South Wales, Australia
Dongfang Yang, National Research Council Canada - London, Canada
Rehana Yasmin, University of Birmingham, UK
Chih-Wei Yi, National Chiao Tung University, Taiwan
Norihiko Yoshida, Saitama University, Japan
Nicolas H. Younan, Mississippi State University, USA
Sergey Y. Yurish, IFSA - Barcelona, Spain
Yifeng Zhou, Communications Research Centre, Canada
Tanveer Zia, Charles Sturt University, Australia

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Electrocardiogram Collection, Pattern Recognition, and Classification Sensor System Supporting a Mobile Cardiovascular Disease Detection Aid <i>Paul Fortier, Patrick DaSilva, and Kristen Sethares</i>	1
A Novel ID Anonymity Preserving Scheme (ID-APS) for Hierarchical Wireless Sensor Networks <i>Ahmed Al-Riyami, Ning Zhang, and John Keane</i>	7
Enhanced Sensitivity in the VIS-NIR Range Under UV Light in a-SiC Pinpin Device <i>Vitor Silva, Paula Louro, Manuel Augusto Vieira, Isabel Rodrigues, and Manuela Vieira</i>	17
WiFi Monitoring Embedded System for Electrical Microgeneration using Renewable Energy Sources <i>Sandro C. S. Juca, Paulo C. M. Carvalho, and Renata I. S. Pereira</i>	21
Feasibility of Charge Transfer Based Atmospheric Ice Sensing <i>Taimur Rashid, Umair Najeem Mughal, and Muhammad Shakeel Virk</i>	27
Evaluation of Torque and Axial Loading Physics for Atmospheric Icing Sensors <i>Umair Najeem Mughal and Muhammad Shakeel Virk</i>	32
Using CFD-Based Virtual Sensor Data to Study the Structure of Air Flow behind A Porous Fence <i>Yizhong Xu and Mohamad Mustafa</i>	39
Improving Distance Estimation in Object Localisation with Bluetooth Low Energy <i>Georgia Ionescu, Carlos Martinez de la Osa, and Michel Deriaz</i>	45
Assessment of Sensor Technologies for Gate-Based Object Counting <i>Colin Lelsie, Alex Vakaloudis, Kostas Anagnostopoulos, Nikolaos Chalikias, and Jian Liang</i>	50
Wide Area Surveillance Using Limited-Flying-Time Helicopters <i>Kenichi Mase</i>	54
BuckshotDV - A Robust Routing Protocol for Wireless Sensor Networks with Unstable Network Topologies and Unidirectional Links <i>Reinhardt Karnapke and Jorg Nolte</i>	60
Ultra Wideband Positioning: An Analytical Study of Emerging Technologies <i>Suheer Alhadhrami, AbdulMalik Al-Salman, Hend Al-Khalifa, Abdulrahman Alarifi, Ahmad Alnafessah, Mansour Alsaleh, and Mai Al-Ammar</i>	66
Integrating Smart Items and Cloud Computing in Healthcare Scenarios <i>Sarfaraz Ghulam, Johannes Schubert, Gerrit Tamm, and Vladimir Stantchev</i>	75

The Impact of Link Lengths on Energy Consumption in Wireless Sensor Networks <i>Knut Ovsthus, Espen Nilsen, Anne-Lena Kampen, and Oivind Kure</i>	82
Towards a Generic Cloud-based Sensor Data Management Platform: A Survey and Conceptual Architecture <i>Vincent C. Emeakaroha, Kaniz Fatema, Philip Healy, and John P. Morrison</i>	88
Towards Tactical Military Software Defined Radio <i>Tapio Saarelainen</i>	96
Small and Low Side Lobe Beam-forming Antenna Composed of Narrow Spaced Patch Antennas for Wireless Sensor Networks <i>Yosuke Sato and Shuzo Kato</i>	107
Accuracy Enhancements in Indoor Localization with the Weighted Average Technique <i>Grigorios G. Anagnostopoulos and Michel Deriaz</i>	112
60GHz Radio Hose for Wireless Harness Communication Systems <i>Yosuke Sato, Vannsith Ith, and Shuzo Kato</i>	117
An Analysis of the Need for Dedicated Recovery Methods and Their Applicability in Wireless Sensor Networks Running the Routing Protocol for Low-Power and Lossy Networks <i>Anne-Lena Kampen, Knut Ovsthus, and Oivind Kure</i>	121
Domain Specific Modeling Language for Object Localization in Marine Observatories <i>Charbel Geryes Aoun, Iyas Alloush, Yvon Kermaec, Joel Champeau, and Oussama Zein</i>	130
A Formal Graph-Based Model Applied to Cluster Communication in Wireless Sensor Networks <i>Laura Margarita Rodriguez Peralta, Lina Maria Pestana Leao de Brito, and Eduardo Ismael Hernandez</i>	137
High Deployability of IEEE 802.15.4k DSSS Systems in Interference Dominated Bands <i>Yasutaka Tada, Yosuke Sato, and Shuzo Kato</i>	147
A Real-Time Bridge Pier Scouring Monitoring System Based on Hall-Effect Sensors <i>Chen-Chia Chen, Ssu-Ying Chen, Wen-Ching Chen, Gang-Neng Sung, Jin-Ju Chue, Chih-Ting Kuo, Yi-Jie Hsieh, Chih-Chyau Yang, Chien-Ming Wu, and Chun-Ming Huang</i>	152
Building Automation: Experience with Dynamic Reconfiguration of a Room <i>Maxime Louvel, Francois Pacull, Safietou Raby Thior, Maria Isabel Vergara Gallego, and Oussama Yaakoubi</i>	157
An Integrated Ambient Intelligence System in the Monitoring and Rehabilitation of the Disorder of Consciousness <i>Francesco Riganello, Luigi Piscitelli, Luigi Flotta, Calogero Pace, Giuliano Dolce, and Walter G. Sannita</i>	163

Electrocardiogram Collection, Pattern Recognition, and Classification Sensor System Supporting a Mobile Cardiovascular Disease Detection Aid

Patrick R. DaSilva, Paul J. Fortier
 Electrical and Computer Engineering Department
 University of Massachusetts Dartmouth
 North Dartmouth, Massachusetts, USA
 pdasilva@umassd.edu, pfortier@umassd.edu

Kristen Sethares
 Adult and Child Nursing Department
 University of Massachusetts Dartmouth
 North Dartmouth, Massachusetts, USA
 ksethares@umassd.edu

Abstract— Current mobile monitoring solutions do not offer the ability to recognize cardiac problems without human interpretation. A combination of electrocardiogram (ECG) detection and classification software running on a mobile cardiovascular disease detection sensor is proposed to replace the need for human interpretation. The ECG is filtered using the Wavelet Transform; the ECG wave points detected using a modified version of the Pan Tompkins rule set and the cardiac rhythm is classified using an N-ary tree. The wireless mobile application is designed on a custom printed circuit board (PCB). Testing results show autonomous classifications are possible using a three lead ECG system while the patient is at rest. The proposed solution serves as a stepping stone towards a fully reliable patient disease management teaching tool with the potential to serve as an aid to the cardiovascular healthcare industry.

Keywords— embedded ECG sensor; real-time algorithm; ECG classification.

I. INTRODUCTION

According to the Centers for Disease Control and Prevention (CDC) Division of Vital Statistics, 24.2% of total deaths in 2010 were directly related to heart diseases [1]. Preliminary 2011 data shows this lethal disease continues to be the number one leading cause of death for 596,339 Americans [2].

Studies have shown as heart failure (HF) symptom severity increased, patient symptom uncertainty levels decreased [3] correlating with previous studies suggesting HF patients delay seeking timely treatment for symptoms [4]. Dedicated nurse staffed care facilities have helped decrease levels of patient symptom uncertainty [5], but a real-time mobile monitoring and motivational solution is desired [6].

Advancements in science and technology have made it feasible to continuously shrink signal processing systems aiding in the development of wearable biometric monitoring equipment and replacing systems that rendered the user with limited mobility. Mobile monitoring systems [7]-[11] are not new, but systems that monitor and interact with users in ways that improve health management are evolving [6][12].

The Electrocardiogram (ECG) subsystem is described as an important part of the overall cardiac wellness system's

ability to aid patients in learning to recognize disease specific symptoms and understand the effect on their health [12]. The purpose of this paper is to describe the development of a mobile cardiovascular disease detection sensor that combines wavelet transformation filtering processes with a modified version of the Pan Tompkins detection rule set and using an N-ary tree for classification of ECG arrhythmias. Research using these three methods has been performed before [13]-[21], but nothing combining all three methods applied to ECG arrhythmia detection and classification in support of a mobile cardiovascular disease detection aid has been.

In Section 2, this paper will discuss ECG raw data collection, waveform extraction, waveform classification and describe the testing process the system underwent. Section 3 will discuss the testing results. A conclusion and recommendation for future work will be presented in Section 4.

II. MATERIALS AND METHODS

The ECG subsystem can be broken down into three serial processes (see Figure 1); collection, extraction, and classification. Each process feeds into the next resulting in a heart health classification. Each process is briefly described below.



Figure 1. ECG subsystem process representation.

A. ECG Raw Data Collection

An ECG is the measured electrical activity representing the heart's conduction system typically recorded on a 1 by 1 millimetre (mm) gridded paper representing 40 milliseconds (ms) by 0.1 millivolts (mV). This paper's ECG sensor was designed using a three lead chest only concept since abnormalities of interest are detected using 3 leads. Abnormalities of interest include normal sinus rhythm, atrial arrhythmias (bradycardia, tachycardia, flutter), conduction abnormalities (1st degree AV block, 2nd degree AV block, 3rd degree AV block), and ventricular abnormalities (premature complex, tachycardia, fibrillation).

The differential voltage between two silver chloride (AgCl) electrodes placed on the right and left side of the chest is measured with reference to body ground, amplified by a gain of 1000, and fed to an analogue to digital converter (ADC) on board an Atmel 32-bit UC3 microcontroller (MCU). The Atmel UC3 MCU uses a 12-bit analogue comparator sampling the ECG at 250Hz with a reference at 60% of the supply voltage. Sampling at 250 Hz, creates 1 sample every 0.004 seconds. Using a 12-bit ADC with an analogue circuit gain of 1000 creates 100 ADC units for every 1 mm or 0.1 mV of ECG signal strength prior to amplification.

B. ECG Waveform Extraction

An ECG waveform is described by its principally important points (PIPs) (see Figure 2). In basic terms, the PIPs are the onset, offset, and peak height of the P wave, T wave, U wave and QRS complex. In total there are twelve PIPs. From these PIPs an ECG’s P wave, T wave, U wave, QRS complex, PR Interval, Atrial Rate, Ventricular Rate, and Rhythm can be calculated. With the PIPs known an ECG waveform can begin to be classified.

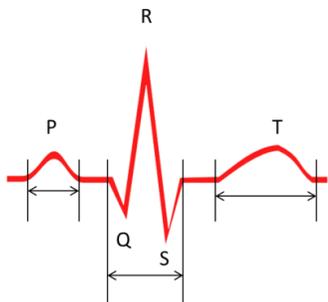


Figure 2. ECG waveform showing principally important point with widths calculated in seconds and heights calculated in millimetres.

To extract the PIPs, the ECG signal is passed through a bank of filters equivalent to the Dyadic Quadratic Spline Wavelet Transform focusing on the time frequency analysis of the signal. Using the dyadic wavelet allows for faster computations on an ECG signal to extract characteristic points by distinguishing between the sharp variations translated into local maxima and minima on different filter scales.

The wavelet transform (WT) equivalent filter described by Li [16] was derived using the work of Mallat [22]. Li derived the WT as a series of high-pass and low-pass filters, used to deduce the equivalent filter as an antisymmetric FIR digital filter with generalized linear phase.

To sync the output of the filters and avoid busy wait loops, an individual delay was added to each filter through the use of the translation property of the Fourier Transform. This caused an overall unified filter bank delay of 62 milliseconds and produced the signal output shown in Figure 3.

The Wavelet Transform equivalent filter has the ability to separate the different characteristic points of an ECG onto

various scales, allowing use of individual filter outputs to find each waveform’s peak, onset and offset.

The amplitude-frequency response of the WT filters (see Figure 4) shows the first five filters in the bank, used to cover the frequency spectrum of an ECG signal sampled at 250 Hz.

To find the QRS complex, originally all five filters were used as discussed by Bahoura [17]. Preliminary testing indicated that filters 2¹ through 2³ are the minimum needed. A QRS complex peak is found by locating the zero crossing of a set of modulus maximum peaks with different polarities on the output of the first three filters simultaneously [18][19][20].

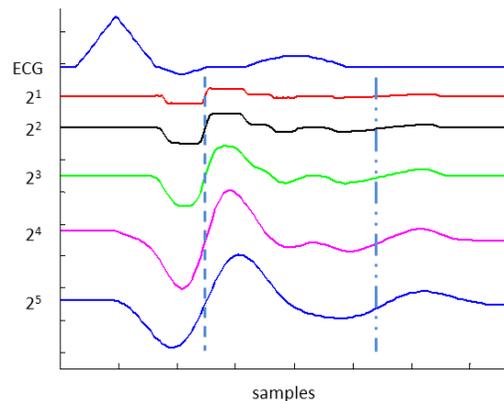


Figure 3. Equivalent WT filter bank output. ECG signal is top line and filter bank out are bottom 5 signals. Dashed line corresponds to QRS complex peak. Dash double dot corresponds to P, T, or U wave peak.

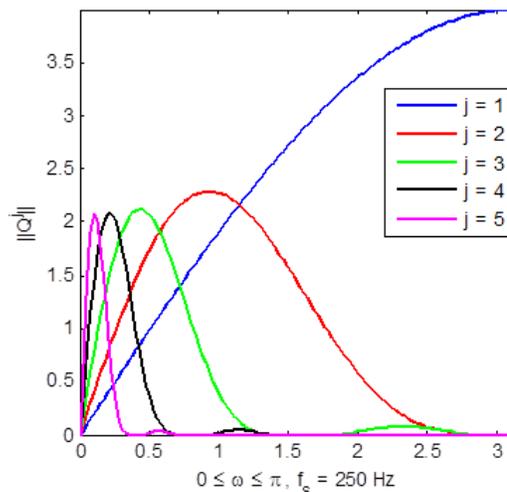


Figure 4. WT filter bank amplitude frequency response.

Once a possible QRS peak is found, all other QRS detections are ignored for a period of 200 milliseconds. An autonomous thresholding technique [23] is employed to find the local minima and maxima peaks on each filter output, allowing the detection algorithm to adjust to a patient’s ECG signal strength. The idea behind the technique was to capture as much of the possible QRS complexes without

low noise with the first threshold and then apply a second threshold to capture missing lower amplitude QRS complexes due to the initialization process. If no QRS complex is found within a 4 second interval, the thresholds are reset [23].

After the QRS peak is found, filter 2^2 is used to find the onset and offset pair using the location of the local minima onset and maxima offset respectively. From preliminary testing using the MIT-BIH databases [24], it was noticed the Q and S waves periodically appear on the filter 2^2 output, so the onset was shifted to the beginning of a second modulus peak found immediately before the first with opposite polarity and the same with the offset, looking at the end of a second modulus peak found immediately after the original offset. Once the QRS complex onset, offset, peak and peak polarity data points are found (see Figure 5), these PIPs are passed to a post-detection scheme to prepare for the classification stage.

Current methods to find P, T, and U waves require first locating the QRS complex and then traversing the ECG signal forwards and backwards in time. The problem with this solution is it's not real-time and does not take into consideration arrhythmias that do not always include a QRS complex for each P and T wave in the ECG, such as Atrial Fibrillation.

Instead, 'blips' which are possible P, T, U waves or just noise, are detected and then categorized in the post-detection scheme. This allows QRS complexes and P, T, U waves to be detected in parallel similar to Bahoura [17]. Preliminary testing proved 2^3 through 2^5 are the only filters required to find blip waves. A blip peak is found by locating the zero crossing of a set of modulus maximum peaks with different polarities on the output of either 2^3 and 2^4 or 2^4 and 2^5 filters, but not necessarily all three filters simultaneously.

Once a possible blip peak is found, the location is saved for 100 ms before reporting to ensure the blip wave is valid and not a QRS complex or noise. If the blip wave is found to be a QRS complex, the information for the wave is transferred to the QRS complex detection to be categorized as such. Again, a thresholding technique is used to find the local minima and maxima peaks on each filter output, except lower than the thresholds used by the QRS. The thresholds adjust to P/T wave amplitudes based upon a pre-calculated ratio between P/T waves and QRS complexes and are re-adjusted every time a new QRS complex is detected.

After the blip peak is found, filter 2^4 is used to find the onset and offset simply because a blip wave will always show on filter 2^4 . The onset/offset pair is found using the location of the local minima onset and maxima offset respectively on filter 2^4 similar to QRS onset/offset detection.

In post-detection, a detected blip is run through a set of test and checks to verify its validity based upon the last detected blip and QRS complex before it is categorized as a P, T, or U wave.

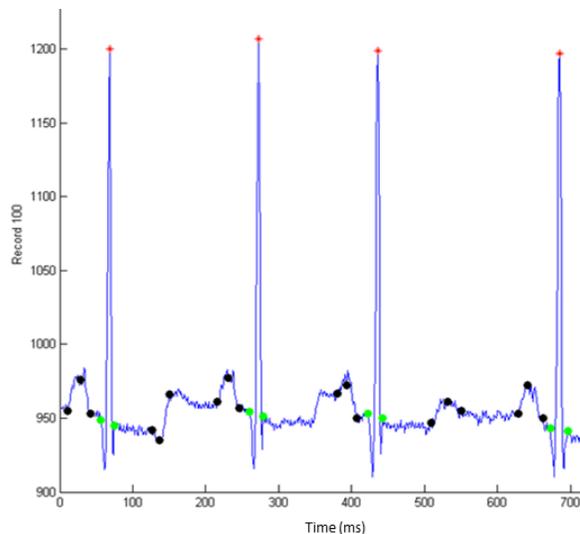


Figure 5. Detected QRS complex (green dots, red stars) and blip waves (black dots) on record 100 from MIT-BIH Arrhythmia database.

This nine point rule set for the test and check is used to construct a finite state machine running on the MCU.

1. Upon start-up, if a blip is detected first, then it is immediately categorized as a P wave and saved in the 'ekgWaveHistory' buffer.
2. Upon start-up, if a QRS complex is detected first, then it is immediately saved in the 'ekgWaveHistory' buffer.
3. A detected blip is invalid and discarded if its onset and/or offset overlap that of the previously detected blip or QRS complex.
4. A detected valid blip is a T wave if its peak is located within 50-75% of the current ventricular heart rate from the offset of the last detected QRS complex and if no other T wave has been detected as of that moment.
5. A detected valid blip is a U wave if its peak is located within 50-75% of the current ventricular heart rate from the offset of the last detected QRS complex, if a T wave has already been detected, and no other U wave has been detected as of that moment.
6. A detected valid blip is a P wave if its peak is located outside 50-75% of the current ventricular heart rate from the offset of the last detected QRS complex or if within that time frame, then it will be a P wave if a T wave and U wave has already been found.
7. A detected QRS complex is always considered valid.
8. A detected QRS complex can invalidate the last detected blip if that blip overlaps the QRS complex in any way.
9. A detected QRS complex can re-categorize a detected P, T, U wave as a T wave that came after the newly detected QRS complex if the last detected blip does not overlap the QRS complex AND comes after the newly detected QRS complex.

Upon type validation, each wave is saved in an ECG buffer by its type, onset, offset, peak and polarity. When the buffer contains one QRS or two P waves, the saved

waveform is sent to classification the next time a third valid P wave is detected or after a second QRS complex has arrived.

The maximum QRS complex detection delay is 462 ms before a QRS complex is detected to when it happened and similarly is 362 ms for a P/T wave detection. This corresponds to 129 beats per minute (bpm) and 165 bpm for a QRS complex detection and P/T wave detection respectively before a lag is seen after each wave occurs. These delays can be attributed to a standard 62 ms filter delay, 200 ms QRS blanking window, 100 ms blip blanking window, and a 200 ms future value collection window.

C. ECG Waveform Classification

Classification is done similar to how a physician classifies an ECG. First the rates are examined, followed by the rhythm, intervals and wave morphology. When a new waveform is detected it's appended to the end of a three waveform historic buffer, shown in Figure 6.

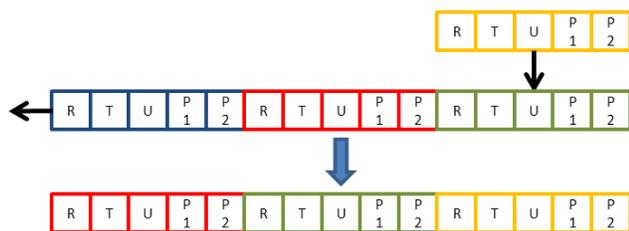


Figure 6. Historic ECG buffer.

The ECG rates are calculated by measuring the lapsed time between each QRS (ventricular) or P (atrial) wave in milliseconds and then dividing that number into sixty thousand milliseconds or equivalently 1 minute to obtain a value in beats per minute (bpm).

The waveform buffer passes through each branch of an N-ary tree (see Figure 7), first eliminating all rhythms that do not correlate. The tree uses cardiac rhythms and classifications along with normal ECG characteristics from best evidence practice literature to determine a waveform's classification.

If there is at least one P wave in the buffer and all available P waves are upright (see Figure 7, branch 00000), then the waveform falls into a Normal Sinus Rhythm (NSR), Atrial Arrhythmia, Conduction Abnormality, Premature Ventricular Contraction (PVC), or Asystole. In this group for a waveform to be considered NSR it must have is a 1:1 P wave QRS complex ratio with normal morphology, PR interval, and ventricular rate.

If there is at least one P wave in the buffer, all available P waves have a negative polarity, are followed by a normal QRS complex with a short PR interval, and all available T waves have normal morphology (see Figure 7, branch 00001), then the waveform could either be Supraventricular Tachycardia (SVT) or a Junctional Rhythm. In this case, the

ventricular rate would be used to differentiate between the two.

If there are no P waves available in the buffer (see Figure 7, branch 00010), then the waveform could be classified as a Ventricular Arrhythmia, Atrial Fibrillation, or SVT. In this group, the morphology of the QRS complex and ventricular rate are used to determine which arrhythmia is present.

By default, if the waveform is unclassifiable, then it is most likely abnormal or if an underlying sinus rhythm is present, but the waveform cannot be classified in the given tree, it is classified as an Abnormal Sinus Rhythm.

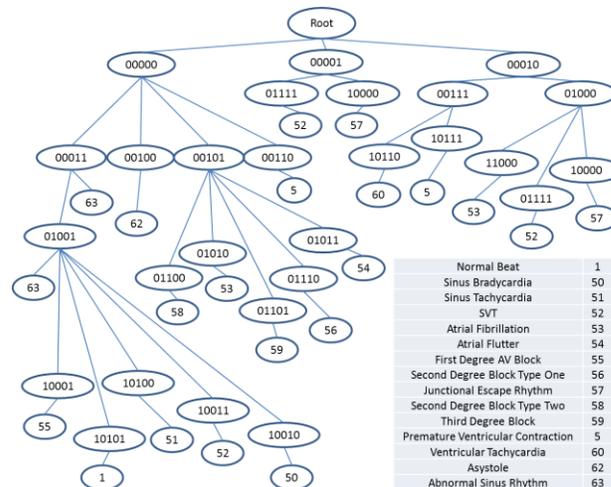


Figure 7. Classification N-ary tree.

The delay between when the heart beats and when that beat waveform is classified by the ECG subsystem is equivalent to one heartbeat. The way the N-ary tree is built takes into account the length of the history buffer in which the algorithm analyses an entire ECG waveform three beats in length at a time. Therefore an NSR classification will only occur when there are three normal heart beats in a row.

D. ECG Sensor System Testing

Testing took place in two stages, the first known canned events followed by live system testing. For the known canned events, ECG signals from the MIT-BIH Arrhythmia and Normal Sinus Rhythm databases [24] were resampled at a frequency of 250 hertz and fed through the algorithm. The results were recorded, reconstructed and analysed using MATLAB.

Only ten minutes of each ECG was used starting at 20 seconds into the signal with results categorized into five areas, positive abnormal classification (PC), positive unknown classification (PU), positive normal classification (PN), negative or missed abnormal classifications (NC), and negative or missed normal classifications (NU). The classifications were cross checked with the annotations included with each signal. If the abnormal annotation matched the abnormal classification, then the result was categorized as a positive abnormal classification. If they

didn't match, but the abnormality was not looked for and the classification wasn't normal, then the result was a positive unknown classification. If neither case, then the result was a negative abnormal classification. In the normal case if the annotation and classification agreed on normal, then the result was a positive normal classification. If the annotation said normal and the classification said anything other than normal, then the result was a negative normal classification.

The ECG records picked from the MIT-BIH Arrhythmia database reflected the classifications the software was attempting to identify. With the exception of records 100, 101, and 222, each record contained PVCs mixed in with various other arrhythmias and normal sinus rhythms. Arrhythmias included Atrial Premature beats, Bundle Branch Blocks, Junctional Premature Beats, Ventricular Tachycardia, Ventricular Flutter, Atrial Fibrillation, Atrial Flutter, and Second Degree Blocks. How well the algorithms could classify normal sinus rhythms mixed with abnormalities and Paced Beats was of interest.

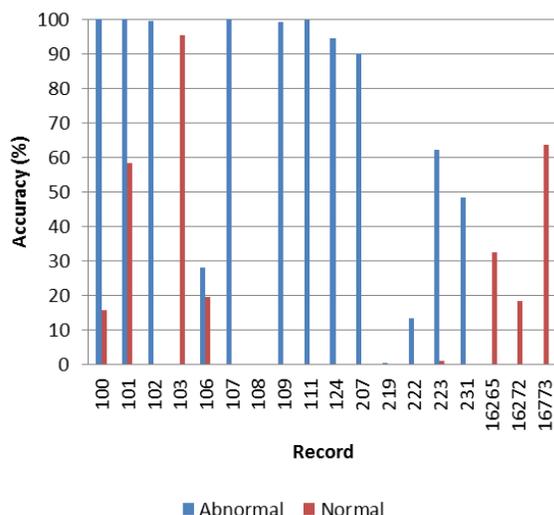


Figure 8. Known canned event testing results. Records 1xx and 2xx are from the MIT-BIH Arrhythmia database. Records 16xxx are from the MIT-BIH Normal Sinus Rhythm database.

In general, the algorithm was able to distinguish between a normal ECG and an arrhythmia (see Figure 8). An analysis of the results showed not one arrhythmia classified as normal, though some were classified as unknown abnormal. Normal classification results shows room for improvement. Majority if not all of the reasons why a given normal rhythm was not classified as such are due to invalid P and T wave detections and increased heart rate from movement.

The results from known canned events testing were collected and analysed and are depicted in Figure 8. To verify the accuracy of the normal sinus rhythm, records from the MIT-BIH Normal Sinus Rhythm database were also run (see Figure 8). Each of those signals mixed with additional noise originating from movement such as running or jogging.

For live system testing and implementation, the ECG algorithms were compiled using the Atmel Studio IDE and uploaded to a custom designed printed circuit board (see Figure 9), running an Atmel 32-bit UC3 microcontroller.



Figure 9. ECG subsystem printed circuit board.

The crucial point of live testing was to ensure the ECG signal collected by the algorithm was the same signal measured and not skewed by the algorithm run time. Using a 2 channel oscilloscope, measurements were initiated during system initialization, when a classification is not found or found using a 16MHz clock source for the MCU. The relative accuracy of the classification process was also studied using a healthy normal ECG by illuminating a series of LEDs corresponding to various classifications.

III. TESTING RESULTS AND DISCUSSION

In some instances, a normal ECG was classified as PVC or could not be identified. This was mainly due to high frequency noise such as in records 108 and 222 or falsely identified and/or unidentified P waves due to wave morphology and P wave proximity to QRS complexes such as in record 222. In other instances a normal ECG would be classified as an Abnormal Sinus Rhythm or Sinus Tachycardia. An Abnormal Sinus Rhythm came about because wave morphology did not fit textbook normal such as in records 100, 101, 103, 106, 223, 16265, 16272, and 16773. Altering the default normal setting in the algorithm would fit this very well. Sinus Tachycardia came about because of an altered heart rate with normal morphology. In a doctor's office, if the patient was running, this would be considered normal, which was the case with records 16265 and 16773.

PVCs in records 106 and 109 were classified as Unknown Arrhythmias because of incorrect QRS polarity detections. In the case of records 109 and 219, PVCs were classified as Unknown Arrhythmias because P waves and underlying sinus rhythm was not detected, both required for a PVC classification. In record 124, the QRS complexes were too wide for this algorithm to be able to detect them and classify the waveform as PVC, but instead classified it as an Unknown Arrhythmia.

The fibrillatory waves' amplitude was too low to detect an atrial rate in records 219 and 222 in order to classify them as Atrial Fibrillation. In record 222, the P waves were

back to back causing incorrect polarity detections to classify the record as Atrial Flutter.

While testing a live healthy and normal ECG signal, the classification was normal for the majority of the test while the user was at rest. The algorithm proved to be resilient to small amounts of movement and noise, but failed as expected when the user began to jog calling for a need to suppress invalid classifications with the addition of an accelerometer and front end smoothing filter.

IV. CONCLUSION AND FUTURE WORK

Overall a new method of autonomously measuring, detecting, and classifying ECG arrhythmias for use in a mobile cardiovascular disease detection sensor system was introduced through combining the Wavelet Transform filtering method with a modified Pan Tompkins detection method and classifying with an N-ary Tree. The main algorithm modifications needed to continue this work would be in the P and T wave detection method and adding an algorithm training method to learn a user's normal sinus rhythm wave morphology.

The training system would include the same detection scheme used throughout the system, but instead of classifying detected sequences, would examine the frequency of the morphology of each P wave, PR interval and QRS complex to determine a proper normal setting within the classification system. Training would be done in the presence of a professional to ensure that a normal ECG is actually occurring rather than an abnormal ECG.

The ECG subsystem would also benefit from a user movement indicator to inform the algorithm that the user is engaged in activity that raises the ventricular heart rate. This battles false classifications of Sinus Tachycardia during a Normal Sinus Rhythm. Movement artifact removal would also be taken care of outside the subsystem by the main monitoring system using motion sensors for scaling.

REFERENCES

- [1] S. L. Murphy, J. Xu, and K. D. Kochanek, "Deaths: Final Data for 2010," National Center for Health Statistics, Hyattsville, MD, National Vital Statistics Reports, vol. 61, no. 4, May 2013.
- [2] D. L. Hoyert and J. Xu, "Deaths: Preliminary Data for 2011," National Center for Health Statistics, Hyattsville, MD, National Vital Statistics Reports, vol. 61, no. 6, Oct. 2012.
- [3] K. A. Sethares, "Predictors of Uncertainty in Adults with an Acute Exacerbation of Heart Failure," *Heart & Lung: The Journal of Acute and Critical Care*, vol. 39, no. 4, Jul. 2010, p. 357.
- [4] C. Y. Jurgens, "Somatic Awareness, Uncertainty, and Delay in Care-Seeking in Acute Heart Failure," *Research in Nursing & Health*, vol. 29, no. 2, Apr. 2006, pp. 74-86.
- [5] A. Stromberg et al., "Nurse-led Heart Failure Clinics Improve Survival and Self-care Behaviour in Patients with Heart Failure: Results from a Prospective, Randomised Trial," *European Heart Journal*, vol. 24, no. 11, Jun. 2003, pp. 1014-1023.
- [6] P. J. Fortier, B. Puntin, and O. Aljaroudi, "Improved Patient Outcomes through Collaborative Monitoring and Management of Subtle Behavioral and Physiological Health Changes," in 2011 44th Hawaii International Conference on System Sciences (HICSS), Kauai, Jan. 2011, pp. 1-10.
- [7] Alive Technologies Pty Ltd, "Alive Heart and Activity Monitor," [Online]. Available: <http://www.alivetec.com/alive-bluetooth-heart-activity-monitor>. [retrieved: September 2014].
- [8] CardioNet, "CardioNet," [Online]. Available: <http://www.cardionet.com/index.htm>. [retrieved: September 2014].
- [9] eCardio, "Arrhythmia Monitoring and Solutions," [Online]. Available: <http://www.ecardio.com/PS/>. [retrieved: September 2014].
- [10] SHL Telemedicine, "SHL Telemedicine," [Online]. Available: <http://www.shl-telemedicine.com/>. [retrieved: September 2014].
- [11] M. K. Suh et al., "An Automated Vital Sign Monitoring System for Congestive Heart Failure Patients," in IHI '10 Proceedings of the 1st ACM International Health Informatics Symposium, New York, Nov. 2010, pp. 108-117.
- [12] P. J. Fortier and B. Viall, "Development of a Mobile Cardiac Wellness Application and Integrated Wearable Sensor Suite," in SENSORCOMM 2011 : The Fifth International Conference on Sensor Technologies and Applications, France, Dec. 2011, pp. 301-306.
- [13] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, Mar. 1985 pp. 230-236.
- [14] V. X. Afonso, "ECG QRS Detection," in *Biomedical Digital Signal Processing*, Englewood Cliffs, NJ, Prentice-Hall, 1993, pp. 236-264.
- [15] P. S. Hamilton and W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database," *IEEE Transactions on Biomedical Engineering*, vol. 33, no. 12, Dec. 1986, pp. 1157-1165.
- [16] C. Li, C. Zheng, and C. Tai, "Detection of ECG Characteristic Points Using Wavelet Transforms," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, Jan. 1995, pp. 21-28.
- [17] M. Bahoura, M. Hassani, and M. Hubin, "DSP Implementation of Wavelet Transform for Real Time ECG Wave Forms Detection and Heart Rate Analysis," *Computer Methods and Programs in Biomedicine*, vol. 52, no. 1, Jan. 1997, pp. 35-44.
- [18] M. Niknazar, B. V. Vahdat, and S. R. Mousavi, "Detection of Characteristic Points of ECG using Quadratic Spline Wavelet Transform," in 2009 International Conference on Signals, Circuits and Systems, Djerba, Nov. 2009, pp. 1-6.
- [19] P. Sasikala and R. WahidaBanu, "Extraction of P wave and T wave in Electrocardiogram using Wavelet Transform," *International Journal of Computer Science and Information Technologies*, vol. 2, no. 1, Feb. 2011, pp. 489-493.
- [20] P. Ranjith, P. Baby, and P. Joseph, "ECG Analysis Using Wavelet Transform: Application to Myocardial Ischemia Detection," *ITBM-RBM*, vol. 24, no. 1, Feb. 2003, pp. 44-47.
- [21] G. K. Prasad and J. S. Sahambi, "Classification of ECG Arrhythmias using Multi-Resolution Analysis and Neural Networks," in TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region, Oct. 2003, pp. 227-231.
- [22] S. Mallat, "Zero-Crossings of a Wavelet Transform," *IEEE Transactions on Information Theory*, vol. 37, no. 4, Jul. 1991, pp. 1019-1033.
- [23] P. S. Hamilton, "Open Source Arrhythmia Detection Software," [Online]. Available: <http://www.eplimited.com/software.htm>. [retrieved: September 2014].
- [24] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiological Signals," *Circulation*, vol. 101, no. 23, Jun. 2000, pp. 215-220.

A Novel ID Anonymity Preserving Scheme (ID-APS) for Hierarchical Wireless Sensor Networks

Ahmed Al-Riyami, Ning Zhang and John Keane

School of Computer Science

The University of Manchester

Manchester, UK

Email: {ahmed.al-riyami, ning.zhang, john.keane}@manchester.ac.uk

Abstract—Node ID anonymity is a critical aspect of privacy in Wireless Sensor Networks (WSNs). Exposing node IDs can help adversaries to learn the underlying network infrastructure and to mount more serious attacks, such as attacks on important nodes (e.g., cluster heads). Therefore, providing ID anonymity should be an integral part of the solution to secure WSNs, and should not hinder other security properties such as accountability and intrusion detection. However, the latter requires that communication nodes be identifiable globally in a dynamic WSN where there is no fixed infrastructure support. Achieving these contradicting security properties effectively and efficiently is challenging. This paper proposes a method, termed ID Anonymity Preserving Scheme (ID-APS), to preserve node ID anonymity with a global node identification capability in hierarchical WSNs. Evaluation results show that ID-APS achieves these properties at a lower level of costs (computational, communication and memory overheads) than comparable methods.

Keywords—Wireless Sensor Networks; privacy; ID anonymity; pseudonym.

I. INTRODUCTION

A critical aspect of privacy in Wireless Sensor Networks (WSNs) is node ID anonymity. By knowing the IDs of the nodes in a WSN, an adversary may obtain information about the communication relationships among the nodes, and infer the network topology which can help to launch successful attacks on important nodes such as cluster heads. Inference of network topology through analysis of node IDs is known as *ID analysis attack* [1].

Passive adversaries launch ID analysis attacks by eavesdropping on network communications; they overhear messages exchanged over wireless media and view source and destination IDs in the message headers. *Active* adversaries launch ID analysis attacks through node compromise; they physically capture a node and access data stored in its memory (e.g., clustering information), identifying the network topology.

Eavesdropping and node compromise attacks are hard to thwart due to the nature of WSNs, e.g., open wireless media and unattended nodes. Existing solutions address these attacks by using dynamic pseudonyms [2]–[5]. These solutions assign a unique pseudonym ID to each node in each transmission, disguising the node's real ID. However, there are a number of issues with these solutions. The first issue is that some solutions, e.g., [3], [4], only consider anonymising node IDs in unicast communications; communication among sensor nodes involves broadcasts as well. If ID anonymity is only provided in unicast communications, then an adversary may still be able to identify nodes through broadcasts. The second issue is how to reduce costs incurred in preserving ID anonymity. Take

pseudonym updating and synchronization as an example, some solutions, e.g., [4], [5], require each pair of communication nodes to use acknowledgments to convey and synchronise the values of a set of pseudonym parameters after each packet transmission between the pair. This raises two further issues. One is that acknowledgments consume network bandwidth and sensor node energy, and increase network traffic levels causing congestions and packet collisions. The second is that acknowledgments are impractical for broadcast traffic, as the resulting acknowledgments will flood the entire network. However, if no mechanism exists to allow communicating nodes to synchronise broadcast pseudonyms, a corrupt or lost message could lead to unsynchronized parameter values hindering further anonymous communication.

Further, preserving node ID anonymity should be accompanied by a secure node global identification facility. This is because the system will also provide other important security properties, e.g., intrusion detection and node compromise detection where a monitoring node needs to report suspicious nodes to the base station. Providing node ID anonymity should enable, rather than hinder, these properties.

This paper proposes a secure solution, the ID Anonymity Preserving Scheme (ID-APS) for hierarchical WSNs. ID-APS offers anonymity and efficiency improvements over the Cryptographic Anonymity Scheme (CAS) [2], the most relevant existing solution. It provides a trade-off between ID anonymity level and computational costs, while minimising communication overheads and memory consumptions. Further, ID-APS also provides an additional built-in functionality to support controlled global node identifiability, i.e., each node in ID-APS can be globally identified by authorised entities. Such flexibility makes ID-APS highly suitable for a wider range of applications as compared with previous solutions.

The rest of the paper is structured as follows: related work is discussed in Section II; the CAS scheme is described in Section III; Section IV discusses how ID-APS improves CAS; Section V presents design preliminaries; the ID-APS scheme is described in Section VI; Section VII analyses the scheme and finally, Section VIII concludes the paper.

II. RELATED WORK

Misra and Xue [2] address anonymity through dynamic pseudonyms in clustered WSNs. They proposed two anonymity schemes, Simple (SAS) and Cryptographic (CAS). In SAS, each node is assigned a randomly distributed set of pseudonym sub-ranges selected from a large space. The node chooses a different pseudonym to identify itself in each transmission. SAS is efficient as it does not involve complicated compu-

tations; however, it requires space to store the pseudonym sub-ranges. CAS reduces the high memory usage of SAS by generating pseudonyms on a per-message basis at run-time by key hashing a sequence number and a pre-defined seed shared between communicating nodes. CAS is more efficient than SAS in terms of memory consumption but at the cost of additional run-time computation. Nevertheless, if an adversary compromises a node in CAS, the adversary may compute the previous IDs of the node using the compromised secret key and the current sequence number.

To protect past used pseudonyms against node compromising attacks, Ouyang et al. [3] proposed the Hashing based ID Randomization (HIR) scheme. Similar to CAS, HIR uses a keyed hash function to generate dynamic pseudonyms. Each outgoing message is identified by a new pseudonym that is generated by key hashing the pseudonym used in the previous message. After each message transmission, the sender creates a new pseudonym and deletes the old one. So, if an adversary obtains the current pseudonym and the secret key, it is still difficult to derive the previous used pseudonyms because of the one way property of the hash function. However, the adversary can still compute the next pseudonym ID and hence impersonate the compromised node to send future messages once the key is captured. To resist such attacks on future messages, the same authors proposed the Reverse Hashing ID Randomization (RHIR) scheme. In RHIR, the one-way hash chain is used in the reverse order, i.e., pseudonyms are assigned backwards from the end of a pre-generated hash chain. This method is more secure against impersonation attack, however, RHIR consumes more memory to store the hash chain and provides each node with a limited number of pseudonyms. In addition, both HIR and RHIR only address anonymous unicast communication not broadcast.

Another scheme to support node ID anonymity is the Anonymous Path Routing (APR) scheme [4]. It hides node IDs using dynamic pseudonyms and uses per-message keys to encrypt each transmitted message. APR only provides node ID anonymity for unicast. For broadcast, (e.g., when a node broadcasts an anonymous path routing request), the broadcasters' IDs are exposed. Hence, an adversary can learn a source node ID from broadcast messages.

Chen and Fang [5] proposed the Efficient Anonymous Communication (EAC) scheme. Unlike other solutions, EAC supports both anonymous unicast and broadcast. Once a message is sent, the sender and receiver generate the next message pseudonym independently using a hash function. The hash function input includes the current message pseudonym and a parameter value shared between the two nodes. To synchronize the unicast pseudonyms used by the two communicating nodes, EAC uses anonymous acknowledgments. After receiving an acknowledgment sent by the receiver, the sender updates the next message pseudonym. However, this work does not mention how local broadcast pseudonyms, which are used between a sender and all its direct neighbours, are synchronised. Without a broadcast pseudonym synchronization solution, the use of anonymous broadcasting may not be practical in areas where message loss and transmission errors are probable.

III. THE CAS SCHEME

Pseudonyms in CAS are generated on a per-message basis using a keyed hash function and a number of parameter

values. The parameters include keys, random seeds and message sequence numbers. Each node maintains a pseudonym table containing the parameter values for each other node in its neighbourhood. Two phases are involved: during the setup phase, nodes exchange and update the parameter values; during the operational phase, they use the values to compute and verify pseudonyms for anonymous unicast and broadcast communications.

A unicast message from node u to the base station via node v is composed as: $M_{uv} = SID \parallel RID \parallel EncryptedPayload \parallel seq_{uv}$ where SID is the end-to-end mutual pseudonym, $SID = Index_v \parallel H_{K_{Bu}}(a_{Bu} \oplus seq_{uv})$ and RID is the next-hop mutual pseudonym, $RID = Index_v \parallel H_{K_{uv}}(a_{uv} \oplus seq_{uv})$, $Index_v$ is the index used to index node u 's parameter values in node v 's pseudonym table, K_{Bu} and a_{Bu} are the hash pairwise key and the random seed shared between node u and the base station, respectively, K_{uv} and a_{uv} are the hash pairwise key and the random seed shared between node u and node v , respectively and seq_{uv} is the current message sequence number for mutual communication between nodes u and v . When v receives message M_{uv} , it retrieves $Index_v$ and searches its pseudonym table for a match. If a match is found, node v uses the corresponding values of K_{uv} and a_{uv} to compute $H_{K_{uv}}(a_{uv} \oplus seq_{uv})$. If the computed value equals the received value, node v accepts the message, otherwise it drops the message. When node v forwards the message to the base station, it includes the received sequence number seq_{uv} . When the message reaches the base station, the base station computes $H_{K_{Bi}}(a_{Bi} \oplus seq_{uv})$ where i refers to every node in the network. Then, it compares the computed hash values with the hash value received in the message to identify its original source.

A local broadcast from cluster head u to its neighbours is composed as: $M_{u*} = SID \parallel RID \parallel EncryptedPayload \parallel seq_{cu}$ where SID and RID are both pseudonyms used to identify the sender, u , $SID = Sentinel \parallel H_{K_{cu}}(a_{cu} \oplus seq_{cu})$ and $RID = Sentinel \parallel H_{K_{cu}}(b_{cu} \oplus seq_{cu})$, where $Sentinel$ is a special character indicating that the broadcast is sent by the cluster head, K_{cu} is the cluster key, a_{cu} and b_{cu} are random seeds shared between the cluster head and all other nodes in the cluster and seq_{cu} is the current message sequence number. Upon receiving the broadcast, a cluster member node identifies the $Sentinel$ to confirm that the message is a broadcast from the cluster head. The node then retrieves the parameter values related to the cluster head from its own pseudonym table and uses the values to compute new hash values $H_{K_{cu}}(a_{cu} \oplus seq_{cu})$ and $H_{K_{cu}}(b_{cu} \oplus seq_{cu})$. If the computed hash values match with the received ones, then the node confirms that the message is from the respective cluster head.

IV. PROPERTIES OF ID-APS

CAS can be improved on three fronts: (i) integrate global node identifiability with ID anonymity, (ii) strengthen the ID anonymity protection level, and (iii) reduce overheads in achieving these properties. These improvements are accomplished using the following ideas:

- For (i), this is achieved by a hybrid combination of fixed (static) and dynamic pseudonyms. Each node can be globally identified using a fixed pseudonym, while their communications are carried out by using per-message dynamic pseudonyms.

- For (ii), two measures are used. The first is to break any association between a currently used pseudonym and the pseudonyms used in the past. So, the use of a random number attribute is introduced in dynamic pseudonym generations. In this way, if the current pseudonym is compromised, the previously used pseudonyms remain safe. The second measure is to hide any patterns or hints that may help adversaries to identify a particular node or a node performing a particular function. For example, CAS uses special characters (*Sentinel*) to identify cluster heads in broadcasts. These sentinels may be used to identify cluster heads as only cluster heads use them. To hide any patterns, we use the same mechanism (indexes) for unicasts and broadcasts and these indexes are randomly picked by the node concerned during the setup phase of the network. We also make the structure of the broadcast pseudonyms identical to that of the unicast pseudonyms, making it harder for an adversary to learn any identifiable information in the network.
 - For (iii), inspired by [6], the sequence number (i.e., message counter) is divided into two parts: one part is explicitly sent as part of the pseudonym in a message and the other part is hidden from transmission (i.e., stored in the node's memory). In this way, the sender and receiver's counters do not need to be tightly synchronized. As long as the number of consecutively lost messages does not exceed a certain threshold, the receiver can still verify the dynamic pseudonym, learn the message sender and synchronise its counter value with the sequence value received. This, along with the reduction in the number of overhead bits each message has to carry, can lead to both bandwidth and node battery savings.
 - *Pairwise Key* (k_{ij}): Each node i shares a unique pairwise key with each of its one-hop neighbours, j . This key is used by node i to secure unicast messages to node j and vice versa.
 - *Broadcast Key* (k_i^*): Each node i shares a unique broadcast key with each of its one-hop neighbours. This key is used to secure local broadcasts by node i to its neighbours.
 - *Cluster Broadcast Key* (k_i^\odot): This key is shared by a CH i and other member nodes in the cluster. It is used by the CH to secure all local broadcasts to cluster members.
- k_N and k_{Bi} are generated by the BS and pre-loaded into each node prior to its deployment. k_{ij} and k_i^* are established using the Energy-efficient Distributed Deterministic Key management scheme (EDDK) [7]. In addition to key establishment, EDDK is also used for node discovery and periodic secure updates of k_{ij} and k_i^* . k_i^\odot is established by the CH. The CH generates the key using a pseudorandom function and securely unicasts it to every cluster member using the pairwise keys.
- Adversaries try to identify the nodes and the communication relationships among them. To do this, they try to access all available information by any means (passive or active attacks). The notations used in this paper are summarised in Table I.

B. Assumptions

- Each node $i \in \mathbb{N}$ has a unique identity, ID_i .
- The BS is always available, trustworthy and protected against physical attacks.
- The clocks of the BS and all the sensor nodes are synchronized.
- Broadcasts from the BS are done by using the μ Tesla authenticated broadcast method [8].
- Sensor nodes are able to obfuscate address fields in their Medium Access Control (MAC) layer header. This assumption is necessary to scope our work without losing generality. A solution to prevent sensor nodes from leaking their MAC level IDs is to use the dynamic pseudonyms at the MAC level as well.
- Each node i maintains a pseudonym table T_i . T_i stores the attribute values associated with the BS and each of the node's neighbours as shown in Table II.

C. Design Requirement Specifications

(a) Node ID anonymity:

- NIP-1: A passive adversary should be unable to learn the identity of the source or destination of a unicast message.
- NIP-2: A passive adversary should be unable to learn the identity of the source of a broadcast.
- NIP-3: An active adversary should be unable to learn the identities of uncompromised nodes through eavesdropping on their unicast communication.
- NIP-4: An active adversary should be unable to learn the past identities of a compromised node.

(b) Node Global Identification:

- GLB-1: A sensor node should be able to globally identify other network nodes without exposing the node IDs to an eavesdropping adversary.

(c) Message Security:

- MSE-1: All messages transmitted should be confidential.

V. PRELIMINARIES

This section presents the system model, assumptions and requirement specifications used in the design of ID-APS.

A. System Model

The WSN consists of a number of homogeneous resource-constrained static sensors and a single resource-rich base station (BS). All network links are assumed to be bidirectional. The network is partitioned into a set of clusters, each with an elected cluster head (CH). Sensor nodes periodically collect measured data and locally aggregate the collected data within each cluster before sending the aggregated data to the BS. Communication modes in this setting include:

- Broadcasts from the BS to all network nodes.
- Unicasts between the BS and any other network node.
- Broadcasts from a node to all its one-hop neighbours.
- Broadcasts from the CH to all cluster member nodes.
- Unicasts among nodes in the same cluster (or tier).

To maximise security, the principle of separation of duties is adopted to establish cryptographic (i.e., symmetrical) keys. In other words, different keys are used for different purposes as follows:

- *Network key* (k_N): This is the network-wide key shared between the BS and all nodes in the network. It is used for securing messages broadcast by the BS.
- *Individual Key* (k_{Bi}): Each node i shares a unique individual key with the BS. The key is used to secure pairwise unicasts between the node and the BS.

TABLE I. ID-APS NOTATION

Notation	Definition
\mathbb{N}	Set of all nodes in WSN where total number of nodes $ \mathbb{N} = N$.
T_i	The pseudonym table of node i .
N_i	Set of all nodes that have their details stored in T_i .
\mathbb{A}_i	Subset of N_i which includes all nodes that node i is allowed to communicate with based on node i 's role.
α	Secret only known by the BS.
β	Shared secret known by the BS and all network nodes.
$H_k^{ID}(\cdot)$	Keyed one-way hash function with output of l_{ID} bits long.
$H_k(\cdot)$	Keyed one-way hash function with output of l_H bits long.
\oplus	Bitwise exclusive-or (XOR) operation.
ID_B	Real identity of BS. All IDs are l_{ID} bits long.
ID_i	Real identity of node i exclusively known to node i and BS.
ID_i^I	Fixed initial pseudonym assigned to node i .
ID_i^O	Fixed operational pseudonym created for node i .
$ID_{i \rightarrow B}$	Dynamic pairwise pseudonym computed by node i ; it is used in unicast messages sent to BS.
$ID_{i \rightarrow j}$	Dynamic pairwise pseudonym computed by node i ; it is used in unicast messages sent to node j .
$ID_{i \rightarrow *}$	Dynamic broadcast pseudonym computed by node i ; it is used in broadcast messages sent to neighbours.
k_N	Network key shared by BS and all network nodes. All keys used are l_k bits long.
k_{Bi}	Individual key shared between node i and BS.
k_{ij}	Pairwise key shared between nodes i and j .
k_i^*	Broadcast key shared between node i and all its one-hop neighbours.
k_i^\odot	Cluster broadcast key shared between a CH node i and all member nodes in the cluster.
$ind_{i \rightarrow j}$	Unicast index shared between nodes i and j . All indexes are l_i bits long.
$ind_{i \rightarrow *}$	Broadcast index of node i shared between node i and its neighbours.
$H_{i \rightarrow j}$	Pairwise hash value used for the construction of dynamic pseudonyms in messages sent from node i to node j .
$H_{i \rightarrow *}$	Broadcast hash value used for the construction of dynamic pseudonyms in messages broadcast by node i to its neighbours.
$h_{i \rightarrow j}$	Implicit counter of messages sent by node i to node j . $h_{i \rightarrow j}$ is l_h bits long.
$s_{i \rightarrow j}$	Explicit counter of messages sent by node i to node j . $s_{i \rightarrow j}$ is l_s bits long.
$C_{i \rightarrow j}$	Unicast counter. This is the counter of messages sent by node i to node j . It consists of two parts: the implicit counter, $h_{i \rightarrow j}$, and the explicit counter, $s_{i \rightarrow *}$. (i.e., $C_{i \rightarrow j} = h_{i \rightarrow j} \parallel s_{i \rightarrow j}$).
$C_{i \rightarrow *}$	Broadcast counter of node i . It is the counter of messages broadcast by node i to its neighbours.
$C_{i \rightarrow \odot}$	Cluster broadcast counter of node i . It is the counter of messages broadcast by node i to other member nodes in the cluster.
$r_{i \rightarrow j}$	Random number generated by node i for the construction of dynamic pseudonyms in messages sent from node i to node j . The random number is l_r bits long.
$r_{i \rightarrow *}$	Random number generated by node i for the construction of dynamic pseudonyms in messages broadcast from node i to its neighbours.
MAC	Message authentication code.
S_T	Threshold value representing the number of lost messages that can be tolerated by the WSN.

 TABLE II. THE PSEUDONYM TABLE (T_i) OF NODE i

	BS	CH	Other neighbours		
Fixed Pseudonym	ID_B^I	ID_j^I	ID_u^I	ID_v^I	...
Pairwise key (k)	k_{Bi}	k_{ij}	k_{iu}	k_{iv}	...
Broadcast key (k)	k_N	k_j^*	k_u^*	k_v^*	...
Cluster broadcast key (k)		k_j^\odot			
Outbound unicast index (ind)	$ind_{i \rightarrow B}$	$ind_{i \rightarrow j}$	$ind_{i \rightarrow u}$	$ind_{i \rightarrow v}$...
Outbound unicast counter (C)	$C_{i \rightarrow B}$	$C_{i \rightarrow j}$	$C_{i \rightarrow u}$	$C_{i \rightarrow v}$...
Inbound unicast index (ind)	$ind_{B \rightarrow i}$	$ind_{j \rightarrow i}$	$ind_{u \rightarrow i}$	$ind_{v \rightarrow i}$...
Inbound unicast counter (C)	$C_{B \rightarrow i}$	$C_{j \rightarrow i}$	$C_{u \rightarrow i}$	$C_{v \rightarrow i}$...
Broadcast index (ind)	$ind_{B \rightarrow *}$	$ind_{j \rightarrow *}$	$ind_{u \rightarrow *}$	$ind_{v \rightarrow *}$...
Broadcast counter (C)	$C_{B \rightarrow *}$	$C_{j \rightarrow *}$	$C_{u \rightarrow *}$	$C_{v \rightarrow *}$...
Cluster broadcast counter (C)		$C_{j \rightarrow \odot}$			

- MSE-2: The authenticity and integrity of a transmitted message should be ensured.
 - MSE-3: A replay attack on a message should be detectable.
- (d) Minimizing Overhead Costs:
- MOC-1: Minimize computational costs.
 - MOC-2: Minimize communication overheads.
 - MOC-3: Minimize memory requirements.

VI. THE ID-APS SCHEME

This section provides an overview of the design of ID-APS including different pseudonyms and protocols used.

A. Pseudonyms

As shown in Figure 1, ID-APS uses a hybrid combination of fixed (initial and operational) and dynamic (pairwise and broadcast) pseudonyms. These pseudonyms are discussed in more details below.

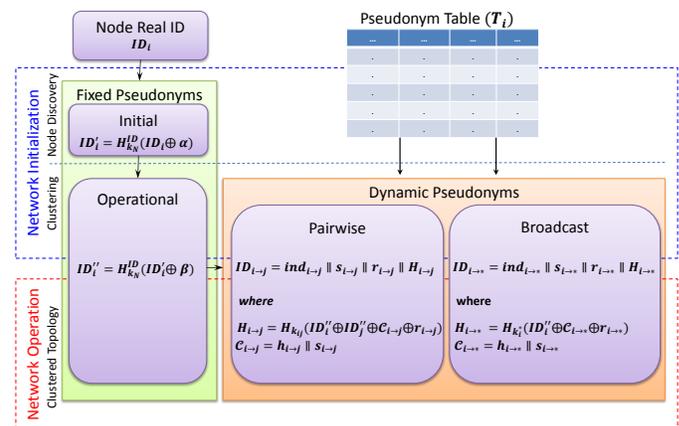


Figure 1. Pseudonym structure used in ID-APS.

(a) Fixed Pseudonyms

- **Initial pseudonym ID_i^I** : A unique identifier of the node used during the initial node discovery and key establishment phase. ID_i^I is computed by the BS as $ID_i^I = H_{k_N}^{ID}(ID_i \oplus \alpha)$ and loaded into each node before its deployment.
- **Operational pseudonym ID_i^O** : A unique global identifier of the node used after the initial phase. It is used when the node should be identified by authorized entities. It is also used to generate dynamic pseudonyms for the node. The confidentiality of this pseudonym is always protected in transit. It is computed as $ID_i^O = H_{k_N}^{ID}(ID_i^I \oplus \beta)$.

(b) Dynamic Pseudonyms

- **Pairwise pseudonym $ID_{i \rightarrow j}$** : This is a dynamically generated pseudonym used as a per-message pseudonym identifying both node i (as a sender) and j (as a receiver) in each message flowing from node i to j . Similarly, node j will use $ID_{j \rightarrow i}$ for its unicast to node i . Only the intended recipient of the unicast can recognise this pseudonym. The pairwise pseudonyms are used in all unicast communications taking place after the initial phase. $ID_{i \rightarrow j}$ is constructed as $ID_{i \rightarrow j} = ind_{i \rightarrow j} \parallel s_{i \rightarrow j} \parallel r_{i \rightarrow j} \parallel H_{i \rightarrow j}$ where $H_{i \rightarrow j} = H_{k_{ij}}(ID_i^O \oplus ID_j^O \oplus C_{i \rightarrow j} \oplus r_{i \rightarrow j})$, $C_{i \rightarrow j} = h_{i \rightarrow j} \parallel s_{i \rightarrow j}$ and,

- $ind_{i \rightarrow j}$ is an outbound unicast index value shared between nodes i and j . It is also called the inbound unicast index of node j for messages from node i . This value is fixed for the life-time of node i . It is used by node j (the recipient) to verify the potential senders of a received message. This verification is designed to filter out any unauthentic messages as soon as they are received, i.e., before more accurate and expensive verifications need to be performed. A recipient only accepts a message if the index value contained in the message matches with one of those recorded in its pseudonym table. Node j will use another index $ind_{j \rightarrow i}$ to send unicasts to node i . These indexes are established as follows: assume that the length of $ind_{i \rightarrow j}$ is l_i bits, node i picks $ind_{i \rightarrow j}$ at random from the range of 2^{l_i} values and sends it to node j encrypted with the pairwise key. Node j does the same to establish $ind_{j \rightarrow i}$ with node i .
- $C_{i \rightarrow j}$ is used to generate sequence numbers to be carried in messages from node i to node j . It is called an outbound unicast counter by node i and inbound unicast counter by node j . It consists of two parts: an implicit counter $h_{i \rightarrow j}$ and an explicit counter $s_{i \rightarrow j}$, (i.e., $C_{i \rightarrow j} = h_{i \rightarrow j} \parallel s_{i \rightarrow j}$). The implicit counter $h_{i \rightarrow j}$ is not included in the pseudonym; rather it is one of the input items used in generating the keyed hash value $H_{i \rightarrow j}$. The explicit counter $s_{i \rightarrow j}$ is explicitly transmitted as part of the pseudonym. $s_{i \rightarrow j}$ increments with each message transmission while $h_{i \rightarrow j}$ increments each time $s_{i \rightarrow j}$ reaches its maximum value, at which point $s_{i \rightarrow j}$ is set to 0. $C_{i \rightarrow j}$ is maintained by both nodes in their pseudonym tables and its value is incremented by the sender after each message transmission and by the receiver after each message reception. Typically, the counters start from 0. However, to increase the anonymity level, each node establishes a random start to its counter value with the receiver. The purpose for using this counter is two-fold: (i) to ensure uniqueness of the pseudonyms carried in each message; (ii) to assure the freshness of a received messages (i.e., for detecting replayed messages).
- $r_{i \rightarrow j}$ is a random number generated by node i to break the link between the pseudonyms used in different messages sent from node i to j , i.e., to make it harder for an adversary to compute node pseudonyms used previously once the current pseudonym is compromised. $r_{i \rightarrow j}$ is regenerated before each message transmission and sent as part of the pseudonym but not stored in the pseudonym tables of the communicating nodes.
- $H_{i \rightarrow j}$ is an outbound hash value carried as part of the pseudonym in a unicast from node i to node j . It is used by the sending node to (i) further obfuscate the identities of the sender and receiver of the unicast; (ii) to ensure the authenticity of the pseudonym. Upon reception, this hash value will also be used by the recipient to verify the pseudonym.

To generate a pairwise pseudonym, the source node first retrieves k_{ij} , $ind_{i \rightarrow j}$ and $C_{i \rightarrow j}$ from its pseudonym table, picks a random number $r_{i \rightarrow j}$, computes $H_{i \rightarrow j}$ and constructs $ID_{i \rightarrow j}$. The node then increments the counter value to make it ready for the next message.

- **Broadcast pseudonym $ID_{i \rightarrow *}$:** This pseudonym is dynamically generated by node i for each message it broadcasts to its neighbours. $ID_{i \rightarrow *}$ can only be identified

by nodes that have node i in their pseudonym tables and are located within the transmission range of node i . The generation of a broadcast pseudonym is similar to that of a pairwise pseudonym, except that a different set of attribute values are used based on whether the node is broadcasting within its own cluster or outside the cluster. If the node is broadcasting within the cluster, then the node uses its broadcast index $ind_{i \rightarrow *}$, cluster broadcast key k_i^\odot and cluster broadcast counter $C_{i \rightarrow \odot}$ to construct $ID_{i \rightarrow *}$. The parameters, $ind_{i \rightarrow *}$, k_i^\odot and $C_{i \rightarrow \odot}$ are shared among all the nodes in the cluster and maintained in each node's pseudonym table. $ID_{i \rightarrow *}$ is constructed as $ID_{i \rightarrow *} = ind_{i \rightarrow *} \parallel s_{i \rightarrow \odot} \parallel r_{i \rightarrow *}$ where $H_{i \rightarrow *}$ where $H_{i \rightarrow *}$ = $H_{k_i^\odot}(ID_i'' \oplus C_{i \rightarrow \odot} \oplus r_{i \rightarrow *})$. However, if the node is broadcasting outside its cluster, then it uses its broadcast index $ind_{i \rightarrow *}$, broadcast key k_i^* and broadcast counter $C_{i \rightarrow *}$ to construct $ID_{i \rightarrow *}$ where k_i^* and $C_{i \rightarrow *}$ are shared among all the neighbouring nodes and maintained in each node's pseudonym table. $ID_{i \rightarrow *}$ is constructed as $ID_{i \rightarrow *} = ind_{i \rightarrow *}$ where $H_{i \rightarrow *}$ = $H_{k_i^*}(ID_i'' \oplus C_{i \rightarrow *} \oplus r_{i \rightarrow *})$.

(c) Dynamic Pseudonym Verification (DP-Ver)

When a message is received by a node, say j , the node executes algorithm DP-Ver (Figure 2) to confirm the validity and authenticity of the pseudonym contained in the message. DP-Ver consists of three verifications, DP-Ver.1, DP-Ver.2 and DP-Ver.3. If the outcome of DP-Ver.1 is negative, node j will discard the message, otherwise it performs DP-Ver.2. If the outcome of DP-Ver.2 is positive, then the node confirms the validity of the received pseudonym, otherwise it performs DP-Ver.3. Similarly, if the outcome of DP-Ver.3 is positive then the node confirms the validity of the pseudonym, otherwise, it discards the message. The three verifications are detailed below.

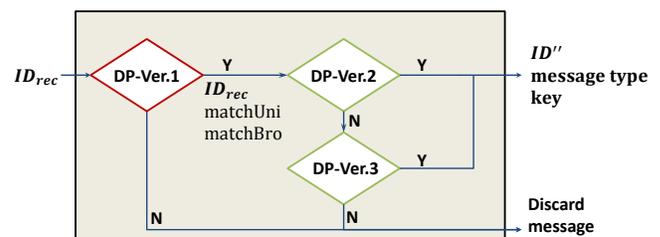


Figure 2. DP-Ver algorithm.

- **DP-Ver.1 - Index Value Verification:** Used to verify the received index value by searching for a match in the receiver's pseudonym table. DP-Ver.1 is a computationally cheap verification that is used to filter out unauthentic messages as soon as possible. Node j maintains a set of nodes with which node j is allowed to communicate, \mathbb{A}_j where $\mathbb{A}_j \subset N_j$. Each node belongs to one cluster and is only allowed to communicate with the nodes (CH and other members) in the same cluster. Only the CH is allowed to communicate outside the cluster with other CHs. Therefore, when node j receives a message, it retrieves index ind_{rec} from the received message and checks it against the index values of the nodes belonging to \mathbb{A}_j . As node j does not know whether the received message is a unicast or broadcast, it has to check against the unicast and broadcast indexes of all nodes in \mathbb{A}_j . If matches are found

within \mathbb{A}_j , then DP-Ver.1 returns two sets of potential senders, one for the unicasts (*matchUni*) and the other for broadcasts (*matchBro*). Otherwise, it returns empty sets, which means that node j is not the intended recipient of the message or the message is not authentic, in which case the node should discard the message.

- *DP-Ver.2 - Keyed hash value verification (no message loss)*: Used to verify the keyed hash value where there are no lost messages (i.e., the received explicit counter = the explicit counter maintained in the receiver's pseudonym table). If either of the two sets, *matchUni* or *matchBro* is not empty, then DP-Ver.2 is used to verify the received keyed hash value H_{rec} . To do this, the algorithm verifies the nodes in *matchUni* first and if the result is negative it continues verifying the nodes in *matchBro*. For each $ID_i'' \in matchUni$, node j retrieves k_{ij} and $C_{i \rightarrow j}$ from T_j and checks if $s_{i \rightarrow j} = s_{rec}$. If they are equal, then node j computes $H' = H_{k_{ij}}(ID_i'' \oplus ID_j'' \oplus C_{i \rightarrow j} \oplus r_{rec})$ and checks if $H' = H_{rec}$. If they are equal, this means that node j is the intended recipient of the unicast message, the message is fresh (not replayed), and it is sent by node i . Node j then increments $C_{i \rightarrow j}$, updates T_j with this incremented value and terminates the algorithm returning the operational pseudonym of the sender ID_i'' , the type of the message (i.e., unicast) and the pairwise key. If the two keyed hash values are not equal, the execution of the algorithm continues by verifying the ID'' s in *matchBro*. The verification here is similar to that for *matchUni*. If this verification is negative, then node j proceeds to execute DP-Ver.3.

- *DP-Ver.3 - Keyed hash value verification (with message loss)*: Used to verify the keyed hash value where there are lost messages (i.e., the received explicit counter \neq the explicit counter maintained in the receiver's pseudonym table). The maximum number of lost messages that can be tolerated equals to a threshold value, S_T . The algorithm starts with verifying the nodes belonging to *matchUni*. First, it retrieves k_{ij} and $C_{i \rightarrow j}$ from T_j for each $ID_i'' \in matchUni$, computes $S_D = |s_{i \rightarrow j} - s_{rec}|$ and checks if $S_D < S_T$. If true, then node j increases $C_{i \rightarrow j}$ based on S_D value, i.e., temporarily creates a new counter $C^{Temp} = C_{i \rightarrow j} + S_D$ and uses C^{Temp} to compute $H' = H_{k_{ij}}(ID_i'' \oplus ID_j'' \oplus C^{Temp} \oplus r_{rec})$. Then, it checks if $H' = H_{rec}$. If they are equal, node j confirms that it is the intended recipient of the unicast message. It increases $C_{i \rightarrow j}$ based on the computed number of lost messages and updates T_j with this value, thus, synchronising the counter value with the sender. Then, node j terminates the algorithm returning the operational pseudonym of the sender ID'' , the type of the message (i.e., unicast) and the pairwise key. If $H' \neq H_{rec}$, then, node j continues verifying if the sender is one of the ID'' s in *matchBro*. If the outcome of this verification is negative, then node j discards the message.

B. Anonymous Communication Protocols

There are three communication protocols in ID-APS: anonymous one-hop unicast (AOHU), anonymous multi-hop unicast (AMHU) and anonymous local broadcast (ALBR).

(a) Anonymous one-hop unicast (AOHU)

This protocol consists of AOHU-Send and AOHU-Receive procedures. A node i sends data to its one-hop neighbour, node j , by executing AOHU-Send, for which node i generates

a pairwise pseudonym $ID_{i \rightarrow j}$, encrypts the data using k_{ij} , generates a MAC of the pseudonym and the encrypted data, and constructs a message $m_{i \rightarrow j} = ID_{i \rightarrow j} \parallel Payload_{i \rightarrow j} \parallel MAC_{i \rightarrow j}$ where $Payload_{i \rightarrow j}$ is the encrypted data. Then, node i sends $m_{i \rightarrow j}$ to node j . Node j executes AOHU-Receive to receive $m_{i \rightarrow j}$, for which it verifies the pseudonym using DP-Ver to find the source of the message, verifies the MAC and decrypts the payload using k_{ij} .

(b) Anonymous multi-hop unicast (AMHU)

This protocol is built on AOHU. It consists of AMHU-Send, AMHU-Forward and AMHU-Receive procedures. When node i is to send data using a unicast to the BS multi-hops away (e.g., through an intermediate node j), node i executes AMHU-Send for which it generates an end-to-end pairwise pseudonym $ID_{i \rightarrow B}$, encrypts the data using k_{Bi} and generates a MAC of the pseudonym and the encrypted data. Next, node i constructs a message $m_{i \rightarrow B} = ID_{i \rightarrow B} \parallel Payload_{i \rightarrow B} \parallel MAC_{i \rightarrow B}$. As the BS is not a direct neighbour of node i , node i checks its routing table for the next hop (i.e., node j) in the route to the BS. Next, node i uses the AOHU-Send procedure to send a unicast, which wraps $m_{i \rightarrow B}$ to the one-hop neighbour j . Upon receiving the one-hop unicast message from node i , node j executes AMHU-Forward to forward node i 's message to the BS. First, it retrieves $m_{i \rightarrow B}$ using the AOHU-Receive procedure. Then, it checks its routing table for the next hop (i.e., the BS). After that, it uses the AOHU-Send procedure to forward node i 's message to the BS. When the BS receives the one-hop unicast from node j , the BS executes AMHU-Receive procedure to retrieve the data sent by node i . This is done by executing AOHI-Receive first to retrieve $m_{i \rightarrow B}$. After that, the BS verifies $ID_{i \rightarrow B}$ using DP-Ver to find the original source of the message. Then, it verifies the received MAC and decrypts the payload using k_{Bi} to retrieve the data sent by node i .

(c) Anonymous local broadcast (ALBR)

This protocol consists of ALBR-Send and ALBR-Receive procedures. When node i is to send a broadcast to its one-hop neighbours, it executes the ALBR-Send procedure. In this procedure, node i first reads the respective broadcast key from its memory, depending on whether it is broadcasting within the cluster or outside it. Then, node i generates the broadcast pseudonym $ID_{i \rightarrow *}$, encrypts the data and generates a MAC of the pseudonym and the encrypted data. Next, node i broadcasts the message $m_{i \rightarrow *} = ID_{i \rightarrow *} \parallel Payload_{i \rightarrow *} \parallel MAC_{i \rightarrow *}$ to the neighbours. Upon receiving the broadcast by a neighbouring node j , node j executes ALBR-Receive for which it verifies the pseudonym using DP-Ver to find the source of the broadcast, verifies the MAC and decrypts the payload to retrieve the data sent by node i .

C. ID-APS in Action

There are three phases in ID-APS: sensor node initialization, network initialization and network operation.

(a) Sensor Node Initialization

Prior to deployment of the network, each sensor node is initialized with specific parameter values and functions. For each node $i \in \mathbb{N}$, the BS generates the following parameter values: ID_i , ID_i' , β , k_N , k_{Bi} , $ind_{B \rightarrow i}$, $ind_{i \rightarrow B}$, $ind_{B \rightarrow **}$, $C_{B \rightarrow i}$, $C_{i \rightarrow B}$ and $C_{B \rightarrow **}$. The indexes and starting counter values are picked at random for each node. The installed

functions are two keyed hash functions, $H_K^{ID}(\cdot)$, $H_k(\cdot)$ and a pseudorandom function.

(b) Network Initialization

In this phase, nodes discover their neighbours, establish pairwise and broadcast keys and perform clustering, preparing for normal network operation. In the node discovery and key establishment phase, the nodes use EDDK methods to discover their neighbours, and establish broadcast and pairwise keys. In this process, the nodes identify their neighbours using the initial fixed pseudonyms and populates a pseudonym table that includes the initial fixed pseudonym and the established keys for each neighbour. The nodes also pick the indexes ($ind_{i \rightarrow *}$, $ind_{i \rightarrow j}$) and starting counter values ($C_{i \rightarrow *}$, $C_{i \rightarrow j}$) at random and exchange these values with their neighbours. Then, each node generates an operational pseudonym for itself and for each neighbour in the node's pseudonym table and deletes the initial pseudonyms. By the end of this process, each node should have updated its pseudonym table and be ready for the clustering process to be performed. During clustering, nodes organise themselves into clusters and elect CHs. To ensure that the clustering process is done in a secure and anonymous way, we utilise the Private Cluster Head Election (PCHE) [9] protocol. This protocol is designed to hide the identities of the CHs from an adversary who can monitor the clustering process. To further increase the protection level of the ID anonymity, we use our ALBR protocol for all broadcast messages involved in the PCHE. Upon completion of clustering, each node should know if it is a CH or a cluster member. After that, the elected CH in each cluster generates a cluster broadcast key and picks a cluster broadcast counter at random and unicasts both values to each member in its cluster. It is worth noting that the elected CHs will also need to establish their routes to the BS. ID-APS is independent of the routing mechanism; however, ID-APS requires that each CH uses the dynamic pseudonyms when sending the route request and route update messages. Once clustering is complete and the routes to the BS established, the nodes are ready to perform normal network operation.

(c) Network Operation

In this phase, the nodes start to carry out their normal functional operations such as data collection and aggregation. Cluster member nodes can only communicate with the CH and other cluster members (intra-cluster communication). Communication outside the cluster (inter-cluster communication) is carried out via the CH.

In intra-cluster communication, each node senses the physical phenomenon at a pre-defined sampling period and sends its reading to the CH. The CH waits for the readings to be received from different cluster members before applying an aggregation function to compute an aggregate value for the readings. Intra-cluster communication can either be unicast or broadcast among cluster members as follows:

- Intra-cluster unicast is carried out using the AOHU protocol and the pairwise keys to secure the communication.
- Intra-cluster broadcast is carried out using the ALBR protocol and the cluster broadcast key to secure the communication.

In inter-cluster communication, each CH forwards the locally aggregated data within its cluster to the BS through other CHs. CHs which are closer to the BS may collect the aggregated data from other CHs that are further away and

aggregate the received data before sending them to the BS. The inter-cluster communication encompasses the following:

- Inter-cluster unicast between a CH and the BS is carried out using the AOHU protocol if the CH is a direct neighbour of the BS or the AMHU protocol if the CH is multi-hops away from the BS. In both protocols the pairwise keys are used to secure the communication.
- Inter-cluster unicast among neighboring CHs is carried out using the AOHU protocol and the pairwise keys to secure the communication.
- Inter-cluster broadcast from the BS is done through the authenticated broadcast protocol μ Tesla [8].

VII. ANALYSES AND EVALUATION

The effectiveness and efficiency of ID-APS are analysed based on the parameter values presented in Table III.

TABLE III. ID-APS EVALUATION PARAMETER VALUES

Sensor mote	CrossBow TelosB [10]
Block cipher	RC5 [11] (CBC mode)
Message authentication	CBC-MAC-RC5
Key length l_k	8 bytes
ID length l_{ID}	8 bytes
Index length l_i	1 byte
Counter length ($l_h + l_s$)	(3+1) bytes
Pseudonym random number length l_r	2 bytes
Pseudonym Hash value length l_H	4 bytes

A. ID Anonymity

NIP-1: As described earlier, a single pairwise pseudonym $ID_{i \rightarrow j} = ind_{i \rightarrow j} \parallel s_{i \rightarrow j} \parallel r_{i \rightarrow j} \parallel H_{i \rightarrow j}$ is used to identify both the sender (node i) and the receiver (node j) of a unicast message, where $H_{i \rightarrow j} = H_{k_{ij}}(ID_i'' \oplus ID_j'' \oplus C_{i \rightarrow j} \oplus r_{i \rightarrow j})$. It is hard for an adversary to gain any identity information regarding the source or destination of the message from the hash value. This is because (i) it is computationally hard to reverse the hash value due to the one-way property of the hash function and (ii) to compute a matching hash value, the adversary needs to know the pairwise key that was used in the computation, which is only known by the message sender and receiver, and in addition, the adversary needs to guess the correct values of ID_i'' , ID_j'' and $h_{i \rightarrow j}$ to succeed. The probability of guessing the correct values of all these parameters is: $(1/2^{l_K + 2l_{ID} + l_h} = 2^{-216})$, which is sufficiently low if the values of l_K , l_{ID} and l_h are sufficiently large.

NIP-2: The analysis of NIP-2 follows that of NIP-1. The hash value in a broadcast is computed as $H_{i \rightarrow *} = H_{k_i^*}(ID_i'' \oplus C_{i \rightarrow *} \oplus r_{i \rightarrow *})$. The probability of guessing the correct values of k_i^* , ID_i'' and $h_{i \rightarrow *}$ is $(1/2^{l_K + l_{ID} + l_h} = 2^{-158})$.

NIP-3: For a neighbourhood of n nodes where an active adversary is present, we analyse the impact on ID anonymity when the adversary compromises m nodes where $m \leq n - 3$. Let us consider the worst case scenario when the adversary compromises all the nodes except three nodes a , b and c (i.e., there are $n - 3$ colluding nodes). The adversary may learn ID_a'' , ID_b'' and ID_c'' assuming that each of the non-compromised nodes lies in the transmission range of at least one compromised node. When node b for example sends a

unicast to node c , node b constructs the dynamic pseudonym as $ID_{b \rightarrow c} = ind_{b \rightarrow c} \parallel s_{b \rightarrow c} \parallel r_{b \rightarrow c} \parallel H_{b \rightarrow c}$ where, $H_{b \rightarrow c} = H_{k_{bc}}(ID_b'' \oplus ID_c'' \oplus C_{b \rightarrow c} \oplus r_{b \rightarrow c})$. The additional information available to the adversary through overhearing the message is $s_{b \rightarrow c}$ and $r_{b \rightarrow c}$. However, the adversary does not know k_{bc} and $h_{b \rightarrow c}$ which are only known to nodes b and c . Therefore, from the adversary's point of view, each of the three non-compromised nodes has equal probability to be the sender or receiver of the message. To learn the exact sender or receiver of the message, the adversary needs to guess the values of k_{bc} and $h_{b \rightarrow c}$ and apply the keyed hash function to get the correct $H_{b \rightarrow c}$. The probability to get the correct hash value is $1/2^{l_k+l_h} = 2^{-88}$.

NIP-4: ID-APS uses a random number, r , to break the link between different pseudonyms used by the same node for the same destination. If a node i is compromised, the adversary may obtain all the necessary items (i.e., crypto keys, indexes, operational fixed pseudonyms and counters) to compute node i 's pseudonyms. Let us denote the current pseudonym of node i in its unicast communication to node j as $ID_{i \rightarrow j}^m$. If $ID_{i \rightarrow j}^m$ is compromised, then to learn the pseudonym used in the previous unicast message to node j , i.e., $ID_{i \rightarrow j}^{m-1}$, the adversary has to decrement the current counter value and guess the random number used in the previous message. As the length of $r_{i \rightarrow j}$ is l_r bits, the probability of guessing the correct random number used by the node in the previous message is $1/2^{l_r} = 2^{-16}$. If higher anonymity is required, the length of $r_{i \rightarrow j}$ can be set to a higher value.

B. Degree of anonymity vs. computational costs analysis

The anonymity level of ID-APS is analysed and quantified based on the concept of degree of anonymity [12]. The entropy of a system is calculated after the attack and compared with the maximum entropy of the system. This provides a measure of information gained by the adversary after the attack. Assuming a WSN with a total number of N nodes distributed uniformly in a field, if an eavesdropping adversary is present while a message is being sent, then the only fixed part of the pseudonym that the adversary can identify a sender by is the index value. However, this index value may also be used by several other nodes while sending messages to their neighbours. As the index values are picked at random, the probability of a node i picking a certain index (ind_a) to identify itself as a sender to one of its neighbours is $1/2^{l_i}$. Let the average size of $|N_i| = n$, then the node will pick n indexes at random to identify itself in unicast communications and another index for broadcast communications. As broadcasts and unicasts messages have similar structure, the total probability, P_{ind_a} , of a node using index ind_a to identify itself in any message sent by the node is given by (1).

$$P_{ind_a} = 1 - \left(\frac{2^{l_i} - 1}{2^{l_i}} \right)^{n+1} \quad (1)$$

This also implies that approximately NP_{ind_a} nodes in the network will be using ind_a to identify themselves when communicating with other nodes (assuming large network). Through continuous monitoring of the index values appended to messages exchanged in the network, the adversary may assign certain probability to each node as being the sender of an eavesdropped message. Let X be the discrete random

variable with probability mass function $P_i = P(X = i)$, where i represents a single sender in a set of potential senders, N . Let us assume that, in the worst case scenario, the adversary is able to identify all the nodes that are using ind_a as identifier. So, the adversary assigns a probability of $(1/NP_{ind_a})$ to NP_{ind_a} nodes as being the senders of a message that has an index value = ind_a and a probability of 0 to the rest of the nodes in the network. The probability P_i is given by (2).

$$P_i = \begin{cases} \frac{1}{NP_{ind_a}} & \text{for } NP_{ind_a} \text{ nodes} \\ 0 & \text{for } N(1 - P_{ind_a}) \text{ nodes} \end{cases} \quad (2)$$

From [12], the entropy of the system after the attack is computed as $E(X) = -\sum_{i=1}^N P_i \log_2(P_i)$, hence:

$$E(X) = \log_2 \left(\left(1 - \left(\frac{2^{l_i} - 1}{2^{l_i}} \right)^{n+1} \right) N \right) \quad (3)$$

The maximum entropy of the system E_M is achieved when the adversary assigns a probability $P_i = 1/N$ to each node in the network as being the sender of a message, i.e., $E_M = \log_2(N)$. Hence, the degree of anonymity d is estimated as:

$$d = \frac{E(X)}{E_M} = \frac{\log_2 \left(\left(1 - \left(\frac{2^{l_i} - 1}{2^{l_i}} \right)^{n+1} \right) N \right)}{\log_2(N)} \quad (4)$$

We see that the degree of anonymity depends on three parameters N, n and l_i . As N and n are fixed values, then the only parameter that can affect the anonymity of the nodes is l_i . Figure 3 shows the impact of l_i for a WSN with $N = 1000$ and $n = 100$. It is clear from the figure that as l_i increases, the degree of anonymity decreases. According to [12], to achieve an acceptable level of anonymity, the degree of anonymity should be greater than 0.8. Therefore, l_i should be selected as (≤ 8 bits) in the above example.

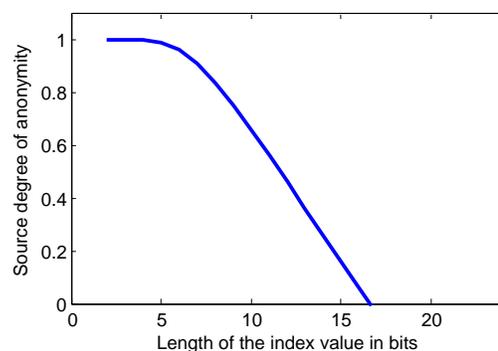


Figure 3. The impact of the length of the index value on the source degree of anonymity for $N=1000$ and $n=100$.

Let us now analyse the computational costs of receiving a message. Depending on the number of matches of the received pseudonym with those contained in a receiver's pseudonym table, the receiver may need to execute more than one keyed hash operation before successfully confirming the validity of the pseudonym. Suppose that each node i has an average size of $|A_i| = m$. The value of m is directly proportional to the size of the cluster. The pseudonym table of node i , T_i , includes an inbound unicast index and a broadcast index for each node in T_i . Therefore, when node i receives a message, it has to

find a match within $(2m)$ indexes in T_i . As described earlier, the inbound unicast indexes and broadcast indexes are picked at random during the node discovery and key establishment phase. Based on this, there is a probability that more than one index in T_i could match the index of the received dynamic pseudonym. The probability of getting exactly r indexes in T_i that matches the index of a received pseudonym (ind_{rec}) is given by (5).

$$P_{match}(r) = \binom{2m}{r} (P_{ind})^r (1 - P_{ind})^{2m-r} \quad (5)$$

Figure 4 shows the probability $P_{match}(r)$ against different r values for $m = 100$ nodes and $l_i = 8$ bits. It can be seen from the figure that the probability beyond 3 matches is very low. In other words, in 99% of the cases, the pseudonym matches occurs in less than 3 matches. Therefore, in 99% of the cases, the verification of a pseudonym involves less than 3 keyed hash operations.

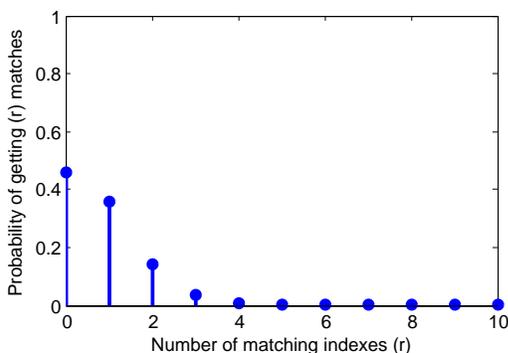


Figure 4. The probability of getting exactly r indexes in T_i that matches ind_{rec} for $m = 100$ and $l_i = 8$.

This computational cost may be further reduced by tuning the parameter values in (5) as follows:

- Reduce the size of the cluster and therefore the value of m drops: This will result in a reduction in the list of potential indexes. Figure 5 depicts the three cases where $m = 100, 50$ and 10 nodes, respectively. As can be seen, reducing the cluster size from 100 to 10 nodes reduces the likelihood of matching indexes from 3 to 1 in 99% of the cases.
- Increase the size of the index: Figure 6 shows the probability of getting r matches when l_i is set to 8, 10 and 12, respectively, where $m = 100$. As shown in the figure, the probability $P_{match}(r)$ drops when l_i increases from 8 to 12 bits. For $l_i = 12$, the receiver need to compute less than 2 hash values in 99% of the time.

From the above we can see that by reducing the cluster size and/or increasing the bit length of the index, we may reduce the computational costs incurred as a result of anonymity protection. However, increasing the bit length of the index reduces the degree of anonymity as illustrated by (4). Therefore, for an acceptable level of anonymity, we may cut computational costs by reducing the size of a cluster.

C. Performance Evaluation

In this section, the performance of ID-APS is evaluated in terms of computation, communication and memory consumption and compared to the performance of CAS and EAC.

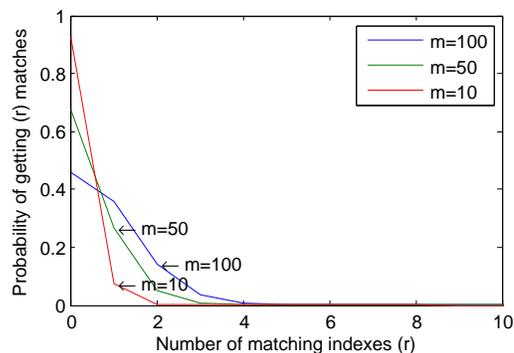


Figure 5. The probability of getting exactly r indexes in T_i that matches ind_{rec} for $l_i = 8$ and different m values.

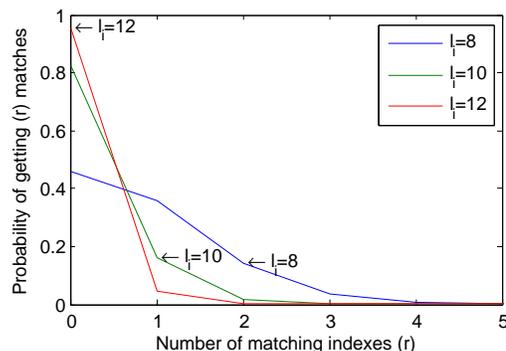


Figure 6. The probability of getting exactly r indexes in T_i that matches ind_{rec} for different l_i values and $m = 100$.

MOC-1: In ID-APS, each message construction prior to transmission involves the following computational costs: one random number generation, one dynamic pseudonym generation using keyed hash function, one payload encryption and one MAC generation protecting the pseudonym and the encrypted payload using a keyed hash function. Based on the experiments with RC5 execution time for TelosB motes [13], constructing a message in ID-APS with a maximum payload size of 29 bytes, as suggested by TinyOS [14], takes less than 20 ms. According to [8], sensor nodes support a maximum throughput of twenty 30-bytes messages/s with the microcontroller being idle for about 50% of the time. Therefore, our scheme is able to anonymize the node IDs, encrypt and generate a MAC for every message sent by the sensor node. The computational costs of receiving the message at the receiver side are comparable to that of the sender. Table IV shows a comparison of computational costs with CAS and EAC.

TABLE IV. COMPUTATIONAL COSTS COMPARISON

Scheme	Sender	Receiver	Other neighbours
CAS	One keyed hashing operation	One keyed hashing operation	One keyed hashing operation
EAC	Two hashing operations	One hash operation	-
ID-APS	One random number generation and one keyed hashing operation	One keyed hashing operation	-

As can be seen from Table IV, the computational costs of ID-APS are comparable to that of EAC and better than CAS.

This is because with the optimization of l_i and cluster size in ID-APS, there are negligible computational costs involved by other nodes in the neighbourhood.

MOC-2: A pseudonym in ID-APS is 8 bytes in size and it is used to identify both the sender and the receiver of a message. Table V shows the comparison between the three schemes in terms of message overheads. It can be seen that ID-APS exhibits the least communication costs in comparison with the CAS and EAC schemes. EAC has the highest message overheads as the receiver in EAC is required to send an acknowledgment to the sender to synchronise the pseudonym parameters.

TABLE V. MESSAGE OVERHEADS COMPARISON

Scheme	Sender	Receiver	Total
CAS	18-byte (ID+Index+Seq)	-	18-byte
EAC	32-byte (2 IDs)	20-byte (Ack)	52-byte
ID-APS	8-byte (ID)	-	8-byte

MOC-3: The memory requirements for the parameters used in ID-APS are shown in Table VI in comparison with CAS and EAC. In this table, n denotes the size of the node neighbourhood. Figure 7 shows a comparison of memory consumption between the three schemes for different n sizes.

TABLE VI. MEMORY CONSUMPTION COMPARISON

Scheme	Parameters	Memory consumption (byte)
CAS	Key size $K = 8$ bytes Other parameters size $P = 8$ bytes Index size $I = 2$ bytes Sequence number size $S = 8$ bytes	$2K + 3P + S + n(2K + 2I + 3P + 2S)$
EAC	Pseudonym size $ID = 16$ bytes Key size $K = 16$ bytes Other parameters size $P = 16$ bytes Link direction size $L = 0.25$ byte	$3ID + 2K + 2P + n(3ID + 2K + 2P + L)$
ID-APS	ID/pseudonym size $ID = 8$ bytes Key size $K = 8$ bytes Index size $I = 1$ byte Counter size $C = 4$ bytes	$3ID + 4K + 4I + 5C + n(ID + 3K + 3I + 4C)$

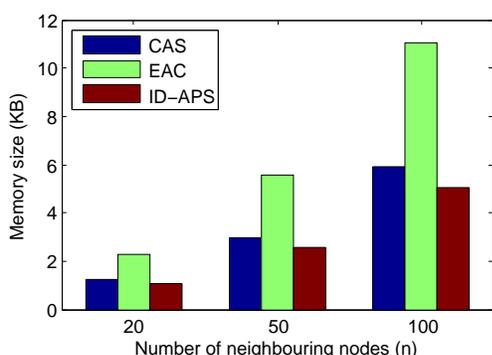


Figure 7. Comparison of memory consumption for different n values.

There are two observations from Figure 7. The first is that the memory requirements of ID-APS can be easily accommodated in the sensor memory. This is because TelosB mote has an architecture where the entire memory (48KB of flash memory, 10KB of RAM and 1MB of external EEPROM) is accessible for code and data. The second observation is that ID-APS consumes least memory space compared to the other two schemes.

VIII. CONCLUSION AND FUTURE WORK

This paper has presented the design and evaluation of a novel ID anonymity scheme, ID-APS. The scheme is designed to achieve a high level of ID anonymity with as less overhead costs as possible in terms of computation, communication and memory space. ID-APS provides ID anonymity in unicast and broadcast communication against passive adversaries and against attacks on past used pseudonyms by active adversaries. Theoretical evaluation of the anonymity level provided by the scheme has been carried out using the concept of the degree of anonymity, and the scheme has also been compared with two most relevant schemes in literature. The evaluation and comparison results have shown that ID-APS achieves a good level of ID anonymity protection with least level of overhead costs in comparison with related solutions. However, with ID-APS, if a node is compromised, the adversary may be able to impersonate the node and compromise the privacy of other nodes. Our future work will examine this issue in depth.

REFERENCES

- [1] X. Luo, X. Ji, and M.-S. Park, "Location privacy against traffic analysis attacks in wireless sensor networks," in Information Science and Applications (ICISA), 2010 International Conference on. IEEE, 2010, pp. 1–6.
- [2] S. Misra and G. Xue, "Efficient anonymity schemes for clustered wireless sensor networks," International Journal of Sensor Networks, vol. 1, no. 1, 2006, pp. 50–63.
- [3] Y. Ouyang et al., "Providing anonymity in wireless sensor networks," in Pervasive Services, IEEE International Conference on. IEEE, 2007, pp. 145–148.
- [4] J.-P. Sheu, J.-R. Jiang, and C. Tu, "Anonymous path routing in wireless sensor networks," in Communications, 2008. ICC'08. IEEE International Conference on. IEEE, 2008, pp. 2728–2734.
- [5] J. Chen, X. Du, and B. Fang, "An efficient anonymous communication protocol for wireless sensor networks," Wireless Communications and Mobile Computing, vol. 12, no. 14, 2012, pp. 1302–1312.
- [6] M. Gouda, Y.-r. Choi, and A. Arora, "Antireplay protocols for sensor networks," Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks, (ed. Jie Wu), CRC, 2005.
- [7] X. Zhang, J. He, and Q. Wei, "Eddk: energy-efficient distributed deterministic key management for wireless sensor networks," EURASIP Journal on Wireless Communications and Networking, vol. 2011, 2011, p. 12.
- [8] A. Perrig, R. Szewczyk, J. Tygar, V. Wen, and D. E. Culler, "Spins: Security protocols for sensor networks," Wireless networks, vol. 8, no. 5, 2002, pp. 521–534.
- [9] L. Buttyán and T. Holczer, "Private cluster head election in wireless sensor networks," in Mobile Adhoc and Sensor Systems, 2009. MASS'09. IEEE 6th International Conference on. IEEE, 2009, pp. 1048–1053.
- [10] Crossbow, "Telosb datasheet," April 2014. [Online]. Available: http://www.willow.co.uk/TelosB_Datasheet.pdf
- [11] R. L. Rivest, "The rc5 encryption algorithm," in Fast Software Encryption. Springer, 1995, pp. 86–96.
- [12] C. Diaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in Privacy Enhancing Technologies. Springer, 2003, pp. 54–68.
- [13] D. He et al., "An enhanced public key infrastructure to secure smart grid wireless communication networks," IEEE NETWORK, vol. 28, no. 1, 2014, pp. 10–16.
- [14] P. Levis et al., "Tinyos: An operating system for sensor networks," in Ambient intelligence. Springer, 2005, pp. 115–148.

Enhanced Sensitivity in the VIS-NIR Range Under UV Light in a-SiC Pinpin Device

V. Silva, P. Louro, M.A. Vieira, I. Rodrigues,
M. Vieira

Electronics Telecommunication and Computer Dept.
ISEL – Instituto Superior de Engenharia de Lisboa
Lisboa, Portugal

vsilva@deetc.isel.ipl.pt, plouro@deetc.isel.ipl.pt,
mvieira@deetc.isel.ipl.pt irodrigues@deetc.isel.ipl.pt,
mv@isel.pt

V. Silva, P. Louro, M. A. Vieira, M. Vieira

CTS-UNINOVA,
UNL - Universidade Nova de Lisboa,
Caparica, Portugal.

vvd.silva@campus.fct.unl.pt, plouro@deetc.isel.ipl.pt,
mvieira@deetc.isel.pt, mv@isel.pt

Abstract—The use of ultra-violet (UV) steady state illumination increases the spectral sensitivity of a pi'npin photodiode beyond the visible spectrum. Increased sensitivity in the range of 400-850 nm is experimentally demonstrated. The pi'npin photodiode can be illuminated back and front sides. Under front irradiation and low power intensity the gain is high and presents a well defined peak at 750 nm and strongly quenches in the visible range. As the power irradiation increases the peak shifts to the visible range and can be deconvoluted into two peaks, one in the red range and another in the green range. Under back irradiation the gain is high in the violet/blue ranges and strongly quenches for wavelengths higher than 550 nm, whatever the intensity of the background. Results show that, front background UV illumination enhances the light-to-dark sensitivity of the medium, opening a window in the visible and near infrared range (VIS-NIR). Background illumination with UV at the back enhances only the magnitude in short wavelength range opening a window in the UV and blue ranges.

Keywords-UV; VIS-NIR; Visible Light Communication; SiC technology.

I. INTRODUCTION

Light-emitting Diode (LED) is a very effective lighting technology due to its high brightness, long life, energy efficiency, durability, affordable cost, optical spectrum and its colour range for creative use. Their use as communication devices with a photo-diode as receptor, has been used for many years in hand held devices, to control televisions and other media equipment, and with higher rates, between computational devices [1]. This communication path has been used in the near infra-red (NIR) range, but due to the increasing LED lighting in homes and offices, the idea to use them for visible light communications (VLC) is present in many working groups. The Institute of Electrical and Electronics Engineers (IEEE) task group 7 has come up with the IEEE 802.15.7 VLC PHY/MAC standard proposal for the physical (PHY) and medium access control (MAC) for VLC communications [2]. The Internet use and its most popular protocols also have been studied for their performance over VLC [3]. Economic issues that will eventually guide the VLC outcome are also on the run [4].

The sensor presented in this paper, is based on amorphous silicon carbon technology (a-SiC:H) [5], it consists of a pi'npin structure. The front pi'n is thinner and has carbon a-SiC:H in the intrinsic layer while the back pin is based on a-Si:H. Two electrical optically transparent contacts interface the sensor at the front and back. Due to the asymmetric lengths of each pin structure and to their difference in materials, the sensor has interesting properties, namely a light selective filter [6]. The sensor is capable of discriminating different wavelength signals and with convenient signal modulation has the capability of logic transformations (AND, OR, NOT and XOR) over the input signals [7], which may play an important function in specific VLC applications. A multiplexing / demultiplexing function has been identified using the sensor with 4 different channels allowing WDM applications [8]. The study of UV and IR response of the sensor is important for the possible use of these wavelengths as communication channels for out-of-visible band information interchange. In this paper, we demonstrate the use of near-UV steady state illumination to increase the spectral sensitivity of the a-SiC/Si pi'npin photodiode beyond the visible spectrum. Section II describes the pi'npin structure. Section III presents the theoretical optoelectronic model and linear state equations. Section IV shows in graphical plots the experimental spectrum of the pi'npin photocurrent, and the photocurrent output for input signals in the visible and NIR spectrum ranges. Section V presents the results and their discussion. The conclusions and acknowledgement close the article.

II. MATERIAL AND METHODS

The sensor is composed by two stacked p-i-n structures (p(a-SiC:H)- i'(a-SiC:H)-n(a-SiC:H)-p(a-SiC:H)-i(a-Si:H)-n(a-Si:H)) sandwiched between two transparent contacts at each end. The thicknesses and optical gap of the i'- (200 nm; 2.1 eV) and i- (1000 nm; 1.8 eV) layers are optimized for light absorption in the blue and red ranges [9]. The working range of the sensor is in the visible spectrum, although light sources within the ultra-violet (350-400nm) and near infra-red (700-880nm) ranges are also used.

This 1cm^2 structure can be seen in Figure 1, where the wavelength arrows indicate the absorption depths during operation [10] and the digital light signals $\lambda_V, \lambda_B, \lambda_G, \lambda_R, \lambda_{IR}$, where the V index is in the UV range, R,G,B in the Visible and IR in the NIR range.

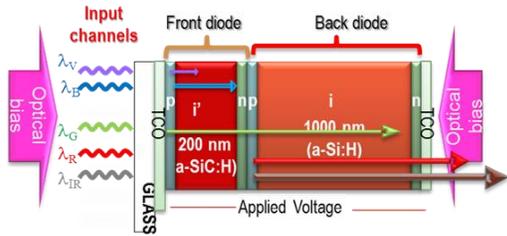


Figure 1. Device structure and operation.

General purposes LEDs are used as light sources. These light sources are used as digital signals and background lighting. The signals are impinged on the front side of the sensor. The background lighting is either at the back or at the front side. The intensity of the signal sources is very low when compared to the background intensity. The LEDs that shine on the front surface, signal and background, are set at a distance of 3cm, while the back led is almost touching the surface. Different wavelength signal sources are used: violet (400 nm), blue (470 nm), green (524 nm), red (626 nm) and infra-red (850 nm). For background lighting a 390 nm violet wavelength is applied. The signals are thus subject to constant background lighting and sampled at the midpoint of each bit. To change the background intensity different currents were used to drive the LED ($0 < I_{LED} < 30 \text{ mA}$). The sensor working point is set at -8 V.

III. THEORY

Based on the device configuration and experimental results an optoelectronic model was developed [11] and upgraded to include several inputs. The model is shown in Figure 2 as an ac equivalent circuit.

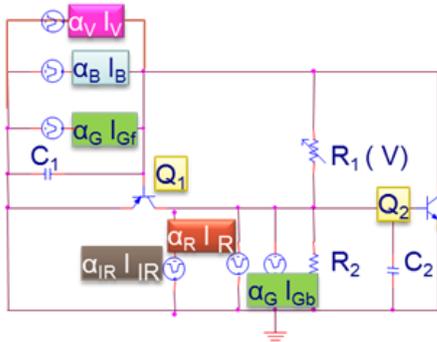


Figure 2. ac equivalent circuit of the pi-npin photodiode.

The corresponding block diagram of the ac equivalent circuit of the pi-npin photodiode, and the linear state equations, are displayed in Figure 3. Input signals, $\lambda_{IR,R,G,B,V}$ model the input channels, and $i(t)$ the output signal. The

amplifying elements, α_1 and α_2 are linear combinations of the optical gains of each impinging channel, respectively into the front and back phototransistors, and provide gain ($\alpha > 1$) or attenuation ($\alpha < 1$) depending on the background effect.

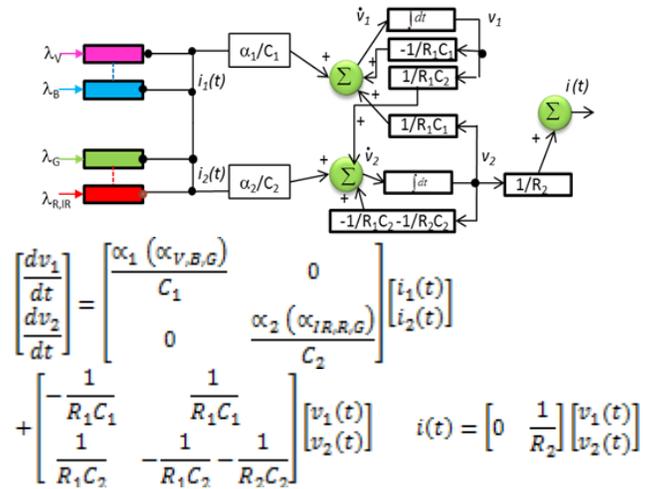


Figure 3. Block diagram and linear state equations.

The control matrix takes into account the enhancement or reduction of the channels, due to the steady state irradiation: front irradiation $\alpha_2 \gg \alpha_1$ and back irradiation $\alpha_1 \gg \alpha_2$. This affects the reverse photo capacitances, $(\alpha_{1,2} / C_{1,2})$ that determines the influence of the system input on the state change (control matrix).

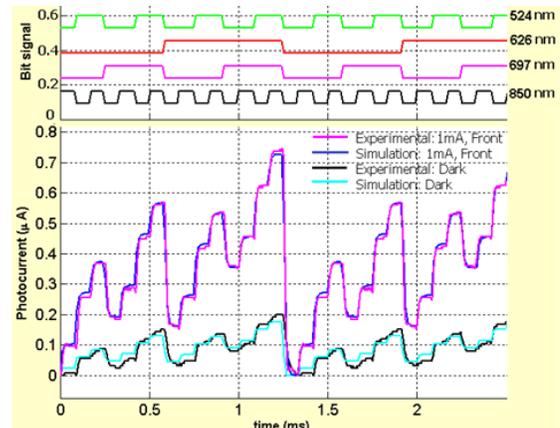


Figure 4. Simulation example of the MUX signal with and without front $\lambda=390 \text{ nm}$ irradiation.

A graphics user interface (GUI) computer program was designed and programmed with MATLAB® to ease the task of numerical simulation. This interface allows the selection of model parameters, and plots the photocurrent results of the simulated and experimental data. The simulation uses as a solver, one of two alternative algorithms: a one-step solver based on an explicit Runge-Kutta formula and an implementation of the trapezoidal rule using a "free" interpolant. An example is presented in Figure 4.

IV. EXPERIMENTAL

A. Spectral measurement

Using a monochromator with 1mm slits and a chopper frequency of 3500 Hz, the spectrum measurements were made from 400 to 800 nm in 10 nm steps in three conditions: dark, front, and back. In the dark condition, the sensor is only subjected to the monochromator’s light source.

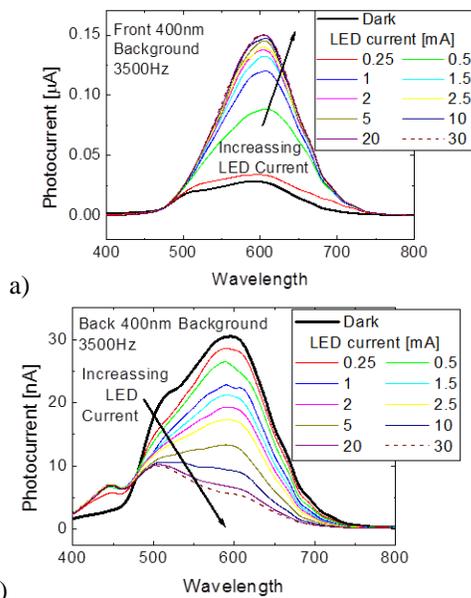


Figure 5. Photocurrent spectrum with different background lighting intensities at the a) front and b) back side of the device.

These measurements (dark) are considered as a reference. The back or front conditions refer to a steady illumination of the back or front side of the sensor.

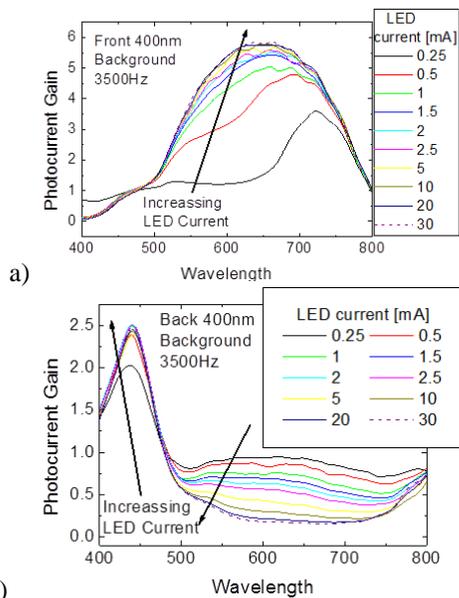


Figure 6. Photocurrent gain with a) front and b) back illumination

Several measurements were made with different background illumination intensities by applying currents in the 0.25 mA to 30 mA range, to the background LEDs with a wavelength of 400 nm. The results are presented in Figure 5.

Experimental output plotted in Figure 5a), shows that with increasing front illumination (see arrow) by the violet LED there is an increase in the gain, from 470 to 750 nm. This increase is very effective, as a current of 0.5 mA is enough to produce an output 5 times higher than the dark level, which is represented in both Figure 5a) and b) by a thick black curve. The influence of the back lighting in Figure 5b) with the same violet wavelength reduces the 470-750 nm bandwidth with increasing LED current. Photocurrent gain is defined as the ratio of the output in relation to the dark curve. The spectral gain for the different LED currents is shown in Figure 6.

B. Visible / Infrared tuning

Several monochromatic (850 nm, 697 nm, 626 nm, 524 nm, 470 nm, 400 nm) individually pulsed lights (input channels) at 12000 bps, and then combined (MUX signal), illuminated the device while steady state 390 nm bias at different intensities ($0 < I_{LED} < 30$ mA) were superimposed separately, from either side of the sensor, and the photocurrent was measured.

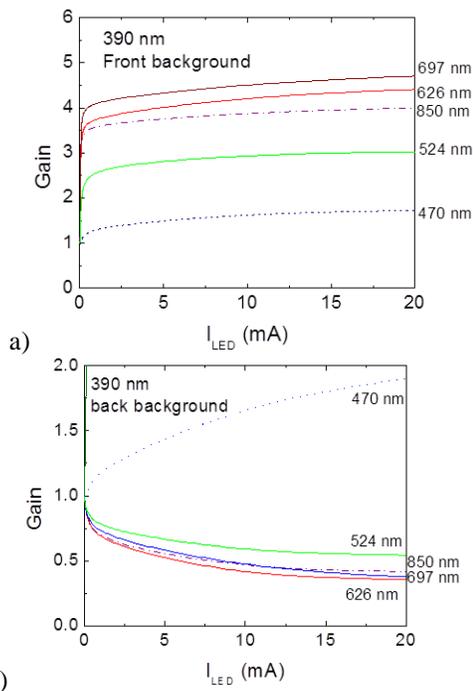


Figure 7. a) Front and b) back optical gain at $\lambda=390$ nm irradiation and different input wavelengths.

For each individual channel the photocurrent was normalized to its value without irradiation (dark), and the photocurrent gain determined. Figure 7 displays the different gain as a function of the drive currents of the lighting LED under front and back irradiation.

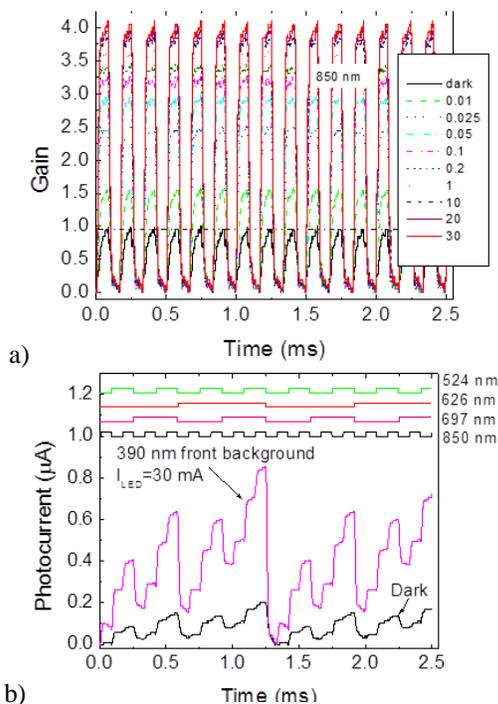


Figure 8. a) Optical gain at 390 nm front irradiation with different intensities. b) Combined polychromatic signal with and without 390 nm front irradiation and different bit sequences.

To exemplify, in Figure 8a), the gain of the 850 nm input channel, under front irradiation at different intensities, is displayed. In Figure 8b), the MUX signal due to the combination of the 850 nm, 697 nm, 626 nm and 524 nm input channels, is presented. At the top of the figure, there are the signals used to drive the input channels, and are shown to guide the eyes into the on/off states.

V. RESULTS AND DISCUSSION

Results show, that background illumination of the front side, Figure 3a), reduces the 400-470 nm spectral range while it augments the 470-800 nm range. It acts as a selective filter for the long pass wavelengths. Background illumination of the back, Figure 3b), enhances the 400-470 nm range and reduces the 470-800 nm wavelengths. It also acts as a selective filter but for the short pass wavelengths. This fundamental behaviour is the basis of the other possible functions that can be derived from it, besides a low pass or a long pass filter. Observing the gain values of the 850 nm IR wavelength presented in Figure 7a) and b), there is a significant difference that distinguishes the background lighting at the back or front side of the device by a wavelength of 390 nm. This allows for the recovery of a NIR channel by a UV background illumination which permits a communication channel outside the visible range. The individual magnitude of each input channel without background lighting ($I_{LED}=0$ in Figure 7), was used to simulate the input channels, and for each bit sequence shown in Figure 8b), the corresponding gain at the

simulated background intensity is presented in Figure 7. The results of the numerical simulation and the experimental results for the same sequence are presented in Figure 4, and show that a good fitting between experimental and simulated results was achieved, showing the robustness of the proposed optoelectronic model.

CONCLUSIONS

An increased sensitivity in a SiC pi-npin device in the VIS-NIR range under UV light was experimentally and theoretically demonstrated. Results show that under front 390 nm irradiation the sensor sensitivity was enhanced in the red/infrared ranges, leading to linearly profiled collection areas that allow the incoming wavelength recognition. An optoelectronic model was presented to explain the observed data, and allow decoding of multiplexed data in the visible/infrared range. The results of the numerical simulation are coherent with the experimental results for the same input sequence.

ACKNOWLEDGMENT

This work was supported by FCT (CTS multi annual funding) through PIDDAC Program funds and PTDC/EEA-ELC/111854/2009 and PTDC/EEA-ELC/120539/2010.

REFERENCES

- [1] T. Komiyama, K. Kobayashi, K. Watanabe, T. Ohkubo, and Y. Kurihara, "Study of visible light communication system using RGB LED lights," in Proceedings of SICE Annual Conference, IEEE, 2011, pp. 1926–1928.
- [2] S. Rajagopal et al, "IEEE 802.15 Wireless Personal Area Networks," 2009, pp. 1–121.
- [3] V. Mai, N. Tran, T. Thang, and A. Pham, "Performance analysis of TCP over visible light communication networks with ARQ-SR protocol," 2014, pp. 1–9.
- [4] H. Chowdhury, H. Bagheri, I. Ashraf, S. Tamoor-ul-hassan, and M. Katz, "Techno-Economic analysis of Visible Light Communications," Proc. Tenth Int. Symp. Wirel. Commun. Syst. (ISWCS 2013), 2013, vol. 9, pp. 96–100.
- [5] G. De Cesare, F. Irrera, F. Lemmi, F. Palma, and M. Tucci, "a-Si:H/a-SiC:H Heterostructure for Bias-Controlled Photodetectors," MRS Proc., Feb. 2011, vol. 336, p. 885.
- [6] M. A. Vieira, "Three Transducers for One Photodetector : essays for optical communications," 2012, p. 222.
- [7] V. Silva et al, "Logic functions based on optical bias controlled SiC tandem devices," Phys. Status Solidi, vol. 11, no. 2, Feb. 2014, pp. 211–216.
- [8] M. A. Vieira, M. Vieira, P. Louro, A. Fantoni, and A. Garção, "Demux operation in tandem amorphous Si-C devices," i-ETCISEL Acad. J. Elect. Comput., vol. 2-1, 2013, p. ID-13.
- [9] M. A. Vieira et al, "Double Pin Photodiodes with Two Optical Gate Connections for Light Triggering," Sensors & Transducers, vol. 10, no. Special issue, 2011, pp. 96–120.
- [10] A. Fantoni, M. Vieira, and R. Martins, "Simulation of hydrogenated amorphous and microcrystalline silicon optoelectronic devices," Math. Comput. Simul., vol. 49, no. 4–5, Sep. 1999, pp. 381–401.
- [11] M. Vieira, P. Louro, M. Fernandes, and M. A. Vieira, "Three Transducers Embedded into One Single SiC Photodetector," in Advances in Photodiodes, vol. March Ch19, G. F. D. Betta, Ed. Intech, 2011, pp. 403–426.

WiFi Monitoring Embedded System for Electrical Microgeneration using Renewable Energy Sources

Sandro C. S. Jucá

Telematics Area

Federal Institute of Ceará – Maracanaú Campus

Maracanaú, Brazil

E-mail: sandrojuca@ifce.edu.br

Paulo C. M. Carvalho, Renata I. S. Pereira

Electrical Engineering Department (DEE)

Federal University of Ceará (UFC)

Fortaleza, Brazil

E-mail: {carvalho, renata}@dee.ufc.br

Abstract— The present paper describes how to design and assemble a low cost online monitoring and WiFi data acquisition system using free software applied to microgeneration based on renewable energy sources worldwide. The development of online monitoring systems for microgeneration plants based on renewable energy sources is becoming more important, considering that monitoring and data acquisition systems are applicable in stages of the microgeneration process. The monitoring and data acquisition WiFi system was developed using an embedded WiFi modem (Wifly) coupled to a microcontrolled board based on the free tool SanUSB. This monitoring system was applied to a photovoltaic (PV) water pumping plant without batteries, so as the control system and the wireless communication with the online server, which is also autonomous and powered by PV panel. The free software for online monitoring and WiFi data acquisition allows the analysis of stored data and charts through mobile devices as notebooks, tablets, and smartphones.

Keywords- WiFi monitoring; Data acquisition systems; Microgeneration; Renewable energy sources.

I. INTRODUCTION

Based on the advent of Brazilian Electricity Regulatory Agency (ANEEL) resolution 482 [1], establishing general conditions for the connection of microgeneration plants (power up to 100 kW) in the Brazilian distribution grid, makes more relevant the development of online monitoring systems [2] of microgeneration plants based on renewable energy sources, like solar, wind, and biomass. Therefore, monitoring and data acquisition systems are applicable in various stages of the microgeneration process, for example, in the energy resource evaluation, generation failures prognosis, practical verification of project data and efficiency of generation.

As a first positive reaction, in August 2013 started working in Rio de Janeiro a PV microgeneration plant connected to the grid, able to produce 50% more than the total house electricity consumption [3].

Data acquisition and online monitoring systems are found in Brazil mainly at large power plants, with complex monitoring and relatively high costs, making impracticable the incorporation into domestic clients and in other clients

that are inside the range of microgeneration power. In this way, the present article intends to develop efficient techniques for online monitoring with open source software, sensing and WiFi data transmission to contribute with the diffusion and installation of microgeneration systems based on renewable energy sources. Until the 90s, Internet was used as a file, news and electronic messages exchange tool, used by students in academy and universities [4]. After the 90s, the World Wide Web started to draw the user attention, changing the way people interact with other people.

The Web offers a colorful and easy way to navigate in graphic interface, with a content not limited to text and hypertext, also offering icons, lines, drawing, maps and pictures. Additionally, there are also hypermedia contents, hypertext pages with sound and video. The Web content can be visualized via a browser, which verifies the files and displays the contents stored on the server, as the open source software for online monitoring developed in the present research.

This paper is organized as follows. Section II describes the hardware and software of the WiFi microcontrolled system, followed by the description of the WiFi communication protocol in Section III. Section IV provides information about the open source software and the principles of the WiFi acquisition system implemented to receive data from the photovoltaic (PV) microgeneration plant described in Section V, including sensing and conditioning. This section contains also charts obtained from the values of the microgeneration plant. Finally, conclusions are presented in Section VI.

II. WiFi ACQUISITION SYSTEM

This section describes the developed data acquisition board with WiFi connection, shown in Figure 1. The WiFi embedded system programming is based on open source and free software, which is an advantage of the proposed system. The board sends monitored data to an online server also programmed in open source software.

This modulated hardware consists of a microcontrolled board, called as SanUSB, connected to an WiFi modem via an adapter board that was developed for this application in order to adjust the pin connections, as well as to convert the voltage from 5 V to 3.3 V.

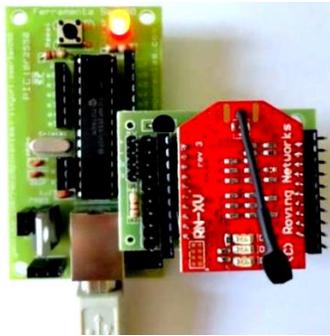


Figure 1. Acquisition data board and WiFi modem.

After coupling the embedded WiFi modem to the microcontroller pins (VCC, GND, Rx, and Tx), it is necessary to program the microcontroller with the firmware developed by Wifly configuration commands. In this way, support libraries were programmed to develop the firmware available as an Android application [5]. The users can also improve the open source firmware in order to adapt for a specific monitoring system.

A wireless system using RF signal to power a temperature sensor for on-body temperature monitoring is shown in [6]. There are also other studies on remote wireless sensing, such as those developed by [7] and [8].

Wireless sensor nodes inside buildings are used to read out sensor data and to control actuators by a reference value in [9]. The nodes need to operate for a long time with a single battery. When using a standard WiFi connection, the node battery would be depleted after a few hours due to idle currents in receive state. The use of sensor nodes with included wake-up receivers can prolong the lifetime of the sensor network to several years, because the sensor measure data and send information only when a command is received from the Internet.

The developed data acquisition board, based in a microcontroller USB programmable tool and using open source software, is patented by National Institute of Industrial Property (INPI) with register number 088503, executable in multiplatform like Linux, Mac OSX and Windows® and available in the files from the SanUSB group [10]. This tool is composed by a preprogrammed bootloader in the embedded system and a graphic interface used to program the microcontroller via USB.

SanUSB tool allows a friendly interface and also debugs directly through the virtual serial emulation via USB. This step can be implemented in a fast and efficient way when the microcontroller is connected to a computer via USB. A developed equipment for specific applications tend to be cheaper, have better cost benefit relation and make possible an easy understanding and operation. With this same data acquisition board, a monitoring system was developed for implementation of a device for high temperature control applied to thermoelectric microgenerators [11]. Figure 2 shows the program process using the designed embedded system.



Figure 2. Program process using the designed embedded system.

Also using SanUSB tool, students were champions of:

- IFCE Robotics Competition (2007, 2008, and 2009) in the Location category, Brazil;
- Brazilian Science and Engineering Fair (FEBRACE-2009) at Federal University of São Paulo (USP), Brazil, in Engineering category;
- Innovation Award for Technology Application at the *Feria Explora Medellin* in Colombia in 2009;
- International Forum of Science and Engineering 2010 in Chile in the *Supranivel* Category;
- FEMECI IFCE 2010 in labyrinth robotics category, Ceará, Brazil.

The mentioned open source software is based on contributions of Internet developers throughout the world. This software offers a better performance, encourages creativity, allows dedicated applications and stimulates finding and correcting code errors faster than private software [12]. It is important to emphasize that USB ports are used as power source by the tool microcontroller only in the code development phase. The data acquisition system is autonomous; that means, a battery or a PV panel is used as an external power source.

III. WiFi COMMUNICATION PROTOCOL

WiFi nets use radio wave technology with Institute of Electrical and Electronics Engineers (IEEE) standards, such as IEEE 802.11a, 802.11b, and 802.11g. These standards provide reliable and safe wireless connectivity. WiFi nets can be used for connection between computational devices and also to connect these devices to Internet [13]. WiFi net operates in the not licensed radio waves 2.4 GHz, in the IEEE 802.11b and IEEE 802.11g technologies and in the 5 GHz frequency of the 802.11a technology [14].

The WiFi modem (Figure 3) is based in the RN-171 module to promote connection to the wireless networks. The connection of this modem needs only four pins designated to power and WiFi communication. This device has an independent antenna to increase its reach and offers stronger signal and support for the most common communication protocols like Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and File Transfer Protocol (FTP).

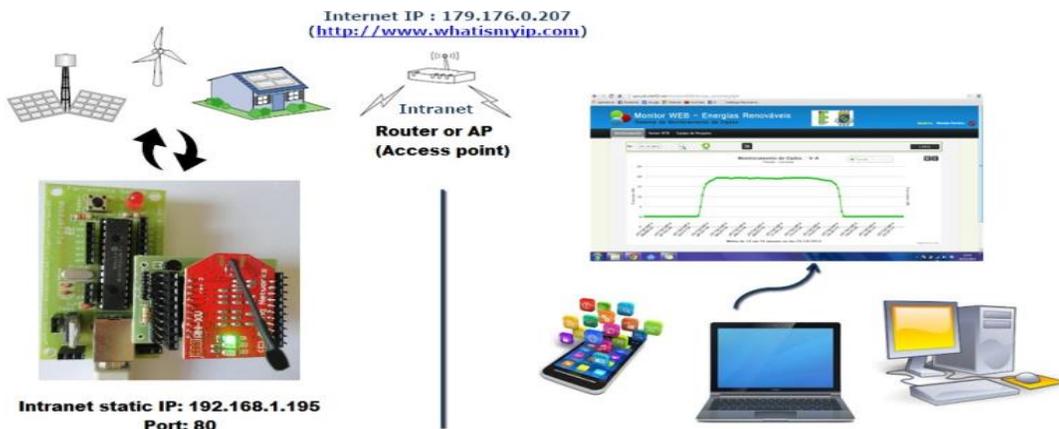


Figure 3. Online monitoring process via WiFi modem.

If more than one WiFi modem or computational device is connected on the Internet in a home network using a broadband router or a gateway, most of the time only this router will contain an Internet IP address and each device in the internal net utilizes an IP local address given by the router.

The intranet IP local address is usually created dynamically via a service called Dynamic Host Configuration Protocol (DHCP) of the gateway, or defined in a static way by the user, according to the network.

In the proposed case, the gateway is 192.168.1.1. Therefore, a fixed static IP address was defined for the WiFi embedded modem, i.e., 192.168.1.195. The proposed embedded system can be, with this firmware, server (e.g., switching a load through the instruction "192.168.1.195/YT" in address bar) or also client (e.g., a sensor value is posted to an online databank server).

IV. PROPOSED CONFIGURATION OF ONLINE MONITORING

In this item, the open source software and the principles of WiFi acquisition are described. In the present study, a PV powered pumping plant without batteries is used, in which the control and communication system with the online server is also autonomous and powered by a PV panel.

In this case, information that is sent to the online data bank is configured with remote access over WiFi without charging, different from the applications that use General Packet Radio Service (GPRS) protocol.

In this context, the development, programming and application of a WEB monitoring system and wireless data acquisition using open source software are described.

The online data bank can be queried by any computational device connected to the Internet by using a personal password. The query can be made at any time, updated every minute. Figure 4 shows the online monitoring development in two stages:

- Sensing, conditioning and data transmission;
- Data bank uploading from an online server and presentation to the user.

The first stage establishes communication between the sensors connected to the data acquisition board. The second stage, i.e., the user presentation layer, was developed in PHP with MySQL database. In this way, a free option for online monitoring applied to renewable energy sources is introduced.

The first stage of sensing, conditioning, and wireless acquisition reads the data from the sensors every minute and saves the information in an internal Electrically-Erasable Programmable Read-Only Memory (EEPROM).

Every ten minutes, the average value of the sensor measurement is calculated and sent to the stage two. It is important to mention that the actuators may be also connected to the board, allowing WiFi load control through the server. In the server, a communication interface to the acquisition system and another one working in parallel to the user communication were developed.

The first interface is responsible for data receiving, to store them in the persistent database and to send a confirmation to the board. The user interface delivers a friendly front-end that allows data visualization in the form of list or chart.

The embedded firmware was programmed using a digital filter to identify and exclude abnormal values that differ by more than 100% of the maximum and minimum values of the variable reference measured every minute.

V. PHOTOVOLTAIC (PV) MICROGENERATION PLANT DESCRIPTION

The microgeneration plant used in the present study is installed on the laboratory of alternative energy (LEA) at the Federal University of Ceará (UFC). This plant consists in a water pumping system powered by PV panels, with a maximum power (P_{max}) of 87 Wp, open circuit voltage of 21.7 V (Voc) and short-circuit current of 5.34 A (Isc).

The motor-pump has a voltage of 12 Vdc, a maximal pressure of 20 psi and a pump flow of 5.9 liters per minute (2,205 rpm) and a nominal current of 4.2 A.

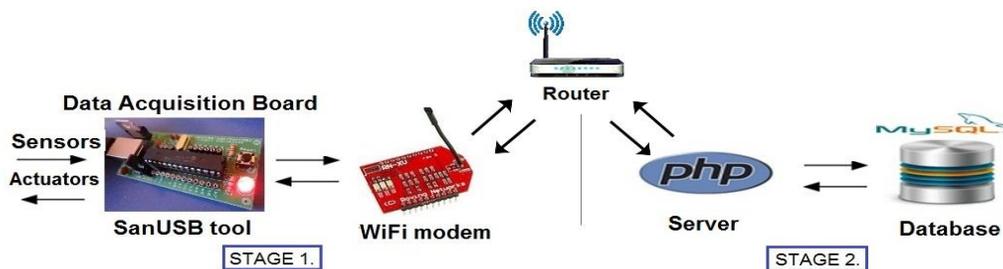


Figure 4. Steps of the open source software monitoring system.

The developed online monitoring and wireless acquisition open source software was installed to record PV panels voltage and current, motor-pump pressure and flow, ambient temperature and solar irradiance. The electronic devices were located in plastic boxes in order to avoid the influence of thermal variation.

A. Sensing, conditioning and data acquisition

To adapt the sensor signal to the analog-to-digital converter (AD) from the data acquisition system is used the principle of conditioning by voltage division, when the measured value is greater than the maximum voltage of the AD converter. There is also the principle of signal amplification with circuits that use instrumentation amplifiers to adjust the voltage or current of the sensor signal to the voltage full scale of the microcontroller AD converter.

Furthermore, the instrumentation amplifiers are also used to read sensors, with minimal or virtually no signal interference due to the high impedance input amplifiers, functioning as a signal insulator, since the amplifier output voltage is supplied by the amplifier power and not by the input voltage. For amplification of the signal from the current sensor of the PV panels, conditioning circuit for difference amplifying is used. This configuration responds to the difference between voltage signals applied to the input and ideally rejects signals that are common to both inputs. Regarding the PV panel voltage, it was necessary to deploy a board with conditioning circuit by voltage division to collect the data, because the signal exceeds the AD converter threshold voltage. In the case of voltage, the PV panel used provides maximum voltage of ca. 19 V.

B. Open source software for online monitoring

The Web Monitor is an application developed on online server with the PHP programming language, in conjunction with a structured database in Database Management System (DBMS), in order to serve as visualization of monitoring data (signals sensors) from renewable energy sources and to send them via a wireless communication (GPRS or WiFi). Figure 5 shows the initial login screen of the data monitoring system.

To access the monitoring system, it is necessary to enter the website link in PHP; after that, the system main authentication page appears, containing an authentication form (username and password). In this way, only pre-registered users can access the system. After entering the data access in the authentication step, the user is redirected to the system home page.



Figure 5. Monitor WEB login screen.

Through the WEB monitor, the user can query the stored data in the online database via smartphones or PC whereas the Ethernet standard is the solution currently used in private systems to promote network interconnectivity; the cost of monitoring systems for microgeneration in wireless network with open source software tends to be dwindling.

At the submenu *Monitoramento* (monitoring), there are two submenus: *Gráfico* (chart) and logs, as shown in Figure 6. By clicking the submenu *Gráfico* a chart of the sensor values stored in the database is generated, for example the chart showed in Figure 7, with voltage, current and electrical power data.



Figure 6. Menu screen expanded.

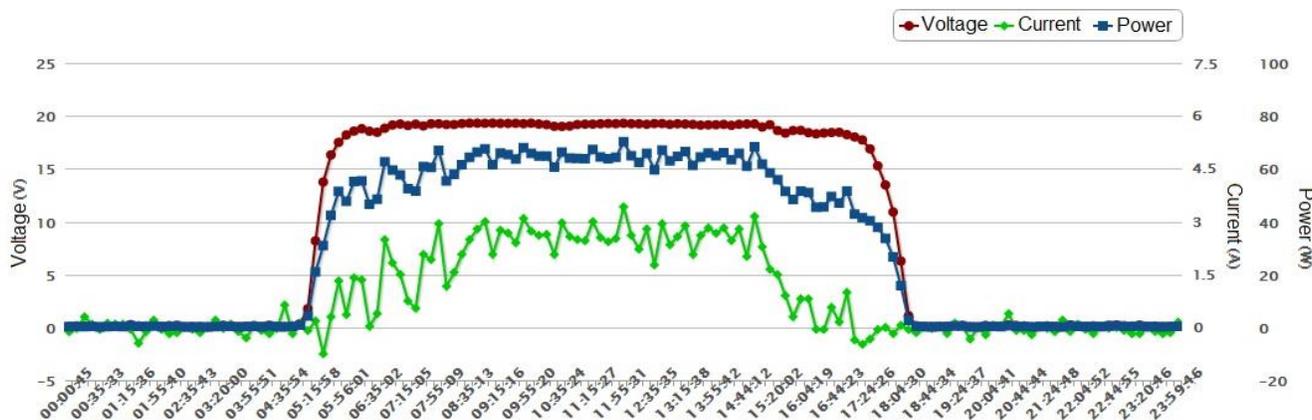


Figure 7. PV voltage, current, and electric power.

To generate the charts, a library called Highcharts, written in Java Script, was used. The chart automatically restarts every 30 seconds. To display the sensors values involved in the chosen application in a specific day, it is necessary to inform or to select a valid date and then click the magnifying glass icon. Shortly after, the chart will be updated showing for each point a representative value averaged over the sensor values every 10 minutes.

In order to view the corresponding logs to the current values displayed in the chart, it is necessary to click the logs button. The system also provides the option to print or export (PNG, JPEG, PDF, and SVG) the entire structure of the chart view.

C. Charts of the open source software for online monitoring

In this topic, charts obtained from the open source software for online monitoring related to voltage, current and power will be presented. The PV panel voltage shape remains stable in sunny days, with a maximum of 19.3 V at noon and remaining around 19 V from 6 am to 5 pm.

The PV current chart showed in Figure 7 tends to follow the solar irradiance, reaching a maximum of 3.7 A at 12 pm, a peak of irradiance. The electrical power is obtained by multiplying the voltage and the current of the PV generator.

The power is maintained around 70 W, for a voltage of 19.29 V and a current of 3.64 A.

The flow sensor has a signal from 0 to 10 V for a range from 0 to 25 liters per minute. In order to perform the sensor reading, it is necessary that the variation is from 0 to 5 V, due to the full scale voltage of the microcontrolled data acquisition system. Thus, the l/min value for a voltage range from 0 to 5 V is 0 to 12.5 l/min (1 liter per minute corresponding to 0.06 m³/h).

In the present study, a pressure sensor PN20 (2005) was deployed. Considering a 5 V maximum voltage of the analog-to-digital converter and a digital maximum value of 255, the sensor shows 70 psi. For a reading in kPa, the conversion factor from psi to kPa is 6.895; then the value of the actual pressure (PR) related to read digital value (Vd) is given by

$$P_R = \frac{70 \cdot V_d}{255} \cdot 6,895 \Rightarrow P_R = 1,893 \cdot V_d \quad (1)$$

Q x H product, measured in m⁴/h is defined as the product of water flow (in cubic meters per hour) and pressure (in meters of water column); this product multiplied by constants indicate the hydraulic power in different units. Pressure, flow, and Q x H product of a monitoring day are shown in Figure 8.

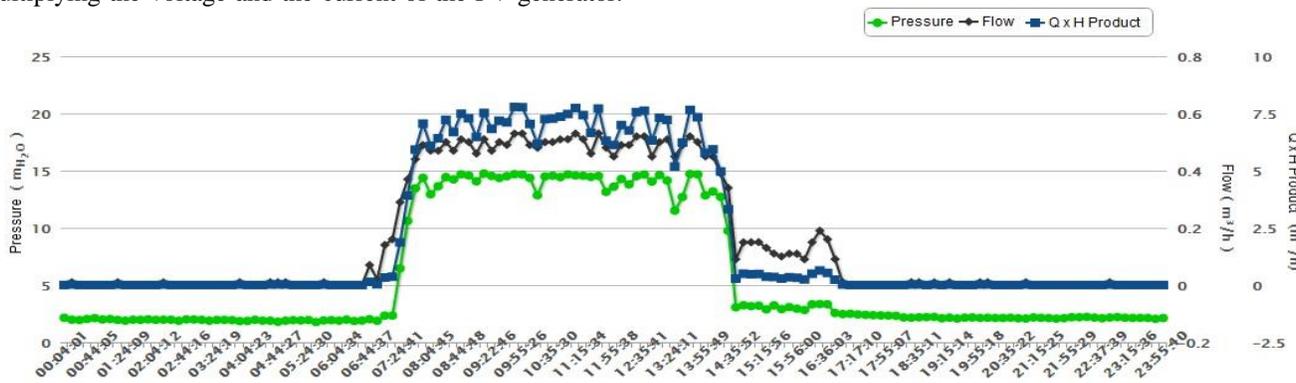


Figure 8. Pressure, flow, and Q x H product.

In the flow chart of Figure 8, the pumping system begins operation with a mean value of 0.5 m³/h at 6 am and reducing the flow at 3 pm. This water flow reduction is due to the decrease in solar irradiance.

VI. CONCLUSIONS

As financial resources in developing countries are generally limited, solutions dedicated to monitoring and identification of local energy resources and for real-time monitoring of decentralized plants using renewable energy sources can contribute to a policy of decentralization of electric power generation in these countries. The present paper proposed a wireless data acquisition system, online monitoring applied to decentralized microgeneration of electrical power from renewable energy sources.

With the advent of ANEEL normative resolution n. 482 introducing general conditions for electrical microgeneration connected to the grid using units up to 100 kW, it becomes more important to develop online systems for monitoring processes of microgeneration based on renewable energy sources, such as solar, wind, and biomass.

The monitoring and data acquisition system proposed was developed in open source software and multiplatform (Linux, Windows®, and Mac OSX) in order to facilitate the dissemination of the computational tool developed for different user profiles. The open source monitoring software proposed enables monitoring the microgeneration plant via smartphones, tablets or other mobile devices with Internet access. As a case study for the monitoring system, a water pumping plant powered by PV panels was used.

The online database can be queried by any computing device connected to the Internet via password. Queries can be made at any time by updating the database every 10 minutes, which is usually the maximal time step for data acquisition systems of renewable energy plants. The WEB monitoring and designed data acquisition system of the microgeneration plant was efficient because of the online query possibility and real-time operation of the electrical microgeneration plant, showing a behavior according to the project. The use of tools based on open source software for online monitoring applied to microgeneration systems allows greater accessibility to general users. The online monitoring and data acquisition model proposed can be expanded to record data from other types of analog or digital sensors, as well as other types of microgeneration plants using renewable energy sources.

ACKNOWLEDGMENT

The authors would like to thank CNPq and Vale for the financial support through the *Forma-Engenharia* program, CAPES and Deutscher Akademischer Austauschdienst (DAAD) for research grants awarded, and UFC and IFCE for the availability of laboratories and equipment.

REFERENCES

- [1] ANEEL, Resolution n° 482. [Online]. Available from: <http://www.aneel.gov.br/cedoc/ren2012482.pdf/> 2014.05.19.
- [2] D. G. Costa, L. A. Guedes, F. Vasques, and P. Portugal, "Partial Energy-Efficient Hop-by-Hop Retransmission in Wireless Sensor Networks," International Conference on Industrial Informatics, Jul. 2013, pp. 146-151, doi: 10.1109/INDIN.2013.6622873.
- [3] Energy Journal, Light interconnects first microgeneration system. [Online]. Available from: http://jornaldaenergia.com.br/ler_noticia.php?id_noticia=14464&id_secao=8/ 2014.04.10
- [4] F. J. Kurose and K. W. Ross, Computer networks and Internet: a new approach. São Paulo, SP: Addison Wesley, 2003.
- [5] S. Jucá, "SanUSB IP access Wifly". [Online]. Available from: https://play.google.com/store/apps/details?id=appinventor.ai_sandro_juca.Wifi_SanUSB/ 2014.07.25.
- [6] G. C. Martins and F. R. de Sousa, "An RF-Powered Temperature Sensor Designed for Biomedical Applications," 26th Symposium on Integrated Circuits and Systems Design (SBCCI), Sept. 2013, pp. 1-6, doi: 10.1109/SBCCI.2013.6644861.
- [7] M. S. Durante and S. Mählknecht, "An Ultra Low Power Wakeup Receiver for Wireless Sensor Nodes", The Third International Conference on Sensor Technologies and Applications (SENSORCOMM), Jun. 2009, pp. 167-170, doi: 10.1109/SENSORCOMM.2009.34.
- [8] H. Raisigel, G. Chabanis, I. Ressejac, and M. Trouillon, "Autonomous Wireless Sensor Node for Building Climate Conditioning Application", Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM), Jul. 2010, pp. 68-73, doi: 10.1109/SENSORCOMM.2010.17.
- [9] G. U. Gamm, S. Sester, and L. M. Reindl, "SmartGate – connecting wireless sensor nodes to the Internet". [Online]. Available from: <http://www.readcube.com/articles/10.5194/jsss-2-45-2013/> 2013.03.27.
- [10] SanUSB Group, *SanUSB tool*. [Online]. Available from: <http://tinyurl.com/SanUSB/> 2014.05.15.
- [11] S. C. S. Jucá, P. C. M. Carvalho, R. I. S. Pereira, D. Petrov, and U. Hilleringmann, "Design and Implementation of a High Temperature Control Monitoring Applied to Micro Thermoelectric Generators". [Online]. Available from: <http://www.icrepq.com/icrepq13/425-juca.pdf/> 2014.06.05
- [12] J.W. Paulson, G. Succi, and A. Eberlein, "An empirical study of open-source and closed-source software products," IEEE Transactions on Software Engineering, Apr. 2004, pp. 246-256, doi: 10.1109/TSE.2004.1274044.
- [13] G. S. V. Radha, K. Rao, and G. Radhamani, WiMAX: A Wireless Technology Revolution. New York, NY: Auerbach Publications, 2007.
- [14] IEEE802, Official IEEE 802.11 working group timelines. [Online]. Available from: <http://www.ieee802.org/11/> 2014.04.17.

Feasibility of Charge Transfer Based Atmospheric Ice Sensing

Taimur Rashid, Umair N. Mughal & Muhammad S. Virk

Atmospheric Icing Research Team
 Narvik University College
 Narvik, Norway
 Email: unmq@hin.no

Abstract—The atmospheric icing parameters are being measured nowadays with the aid of more customized yet limited commercial equipment. The parameters include atmospheric ice detection, icing load and icing rate. The robustness of such equipment is usually under scrutiny when it comes to cold/harsh environment operations. This phenomenon was experienced consistently by the atmospheric Icing research team at Narvik University College during data retrieval exercises from its atmospheric icing stations installed at Fargnesfjellet during 2012-13. In this paper it is aimed to address the potential feasibility to produce a robust hardware addressing the icing measurements signals, which includes instrumentation hardware giving icing indications, icing type and icing rate measurements in a single platform (not commercially available till date).

Keywords—Atmospheric Icing Sensor; Icing Type; Icing Rate; Charge Transfer; Zero Crossover; Cold Regions.

I. INTRODUCTION

Charge transfer method adhere to the capacitive principle and can detect anything that is either conductive or has different dielectric properties than the sensors electrodes surroundings. Any object, conductive or non-conductive, that is brought in the vicinity of the electrode, has its own dielectric properties that will alter the capacitance between the electrode and the sensor housing and, in turn, will produce a measurable response. Also, certain sensor gauge the change by generating an electric field (e-field) and measuring the attenuation suffered by this field. The prime area of focus is to detect the ice at the first instance based on its physical properties. The measured signal is then required to be calibrated in order to achieve reasonably wide range of measurements; differentiated based on the liquid water content in the ice. This would enable to lay the foundation of measuring icing thickness and icing rate.

Atmospheric icing on structures occurs in conditions where cooling of an air mass causes super cooling of water droplets resident in the air mass. Water droplets in the earth atmosphere can remain in the liquid state at air temperature as low as $-40^{\circ}[C]$, before spontaneous freezing occurs [2][3]. Generally an icing event is defined as periods of time where the temperature is below $0^{\circ}[C]$ and the relative humidity is above 95%. According to ISO 12494 standard [4], ice accretion can be defined as any process of ice build-up and snow accretion on the surface of the object exposed to atmosphere [4], [5]. Primarily Atmospheric icing can be classified in two main categories on the basis of the physical properties of atmospheric ice. These will vary depending upon the meteorological conditions and are termed as *precipitation icing* (freezing precipitation and wet snow) and *in-cloud icing* (rime/glaze, including fog). In Northern part of Europe particularly in Norway, it is primarily

freezing fog which causes ice accumulation on structures and this occurs mainly at high altitudes [6]. The ice accretion relies heavily on temperature, liquid water content and droplet size. The ice appearance and its physical properties are influenced by the atmospheric and weather conditions. Table I mentions these affecting parameters, whereas other elements such as compressive strength (yield and crushing), shear strength, etc., might describe the nature of ice accretion on the subject item exposed in the environment. The object exposed will have different ice accretion based on its dimensions and its orientation to the direction of icing wind blowing. Table I shows major outline of the basic meteorological parameters handling the ice accretion process.

TABLE I. TYPICAL PROPERTIES OF ACCRETED ATMOSPHERIC ICE [5]

Type of ice	Density (kg/m ³)	Cohesion	General appearance	
			Colour	Shape
Glaze	900	strong	transparent	evenly distributed/icicles
Wet snow	300-600	weak, strong	white	evenly distributed/eccentric
Hard rime	600-900	strong	opaque	eccentric, pointing windward
Soft rime	200-600	low to medium	white	eccentric pointing windward

This paper is divided in six sections. Section II is an overview of different ice sensing techniques followed by section III, which outlines the critical reasons of system failure in cold region domains. Section IV deals with a detailed description of Charge Transfer Scheme whereas section V deals with the associated tradeoff. This paper is ended with conclusion section VI.

II. ICE SENSING TECHNIQUES

The ISO issued in 2001 a standard for ice accretion on all kinds of structures, except for electric overhead line conductors. In this recommendation, a standard ice-measuring device is described in ISO 12494 [4] as a smooth circular cylinder having diameter of 30 mm placed vertically in the axis and rotating around the axis. The cylindrical length should be of 0.5m, but, in case of heavy ice accretion expectation, the length should be altered to 1m. If the cylinder cannot rotate freely due to wind drag, it may be provided with a motorized mechanism to force the rotation. The speed of rotation is not critical in terms of vertical collection. The vertical cylinder is not fully appropriate for freezing rain in the wet growth stage. To achieve this, it is preferred to use sets of horizontal collectors (rods), which are oriented orthogonally as in case of

Soviet standard ice collector Popov [9] or the Canadian Passive Ice Meter (PIM) as described in IEC 60826 [10]. Ice Sensing techniques come under the umbrella of icing instrumentation. To know about Ice sensing it is required to comprehend the phenomena of icing caused by the Meteorological parameters. Drage [6] describes the complexity of Meteorological Icing based on the factors such as object shape, wind speed, air temperature, liquid water content and droplet size distribution.

There is a lot of room to improve these measurements as new developments being carried out on focusing on these parameters, Makkonen [11]. Different research institutes and industries are involved in ice detection instruments manufacturing. But it is important to notice that many of them are still in the prototyping phase and few amongst those have launched their products in the market. The ISO 12494 standard ice sensor has been manufactured in one version in Swedish Combitech: automatic weighing, free rotation) and two in Finland (Digita: automatic weighing, forced rotation, Lehtonen [12], FMI: manual weighing, forced rotation). The devices/instruments identical with ISO ice collector have been used in the past at some locations as well Rothig [13]. For the purpose of the meteorological icing detection, few market systems (Rosemount Goodrich) and few prototypes are available in the name of Holo Optics, Infralytic, Vibrometer/Boschung. Also some available prototypes available for the ice rate analysis uses active infra-red techniques but the robustness and reliability of the equipment under harsh conditions is still questionable. The electrical impedance and weight measurement based icing equipment are more specialized, focusing on a specific parameter. There is a definite need of an icing system capable of measuring the instantaneous icing rate and thickness along with the ice type detection mechanism. This could provide an advantage to anticipate the ice accretion and load based on the valid detection of icing type.

III. GENERAL REASONS OF SYSTEM FAILURE IN COLD ENVIRONMENTAL CONDITIONS

As mentioned in Rashid et. al. [26] that the dimension of operational problems faced in cold climate is quite different from the operations in normal atmospheric conditions. More often the factors not significant at all in the normal conditions become extremely critical in cold climate regions. Investigations were carried out to track down possible reasons of the HiN icing station's components failure from operational point of view. Analysis showed that in addition to the harsh weather conditions, a combination of various design and operational aspects could also lead to the system's failure in harsh conditions. Following are some noteworthy causes in this regard.

A. Intermittent Power Source

The system installed at the location takes power from the available commercial facility, where high load machines are being operated. Due to the demographic location of the site in terms of accessibility and complicated power infrastructure available in terms of maintenance, several power breakdowns have been frequently reported. The instantaneous power surge could be one reason that has affected the sensors operations.

B. Electrostatic Discharge

The electrostatic discharge phenomena could not be fully neglected in weather station breakdown. For snowstorms,

temperature gradients in the ice particles produce charge separation because the concentration of H^+ and OH^- ions in ice increases rapidly with increasing temperature. H^+ ions are much more mobile within the ice crystal than OH^- ions. As a result, the colder part of an ice particle becomes positively charged, leaving the warmer part charged negatively [21]. The resulting electrostatic phenomena due to blizzard can be hazardous for the control circuitry inside the sensor module, provided the said consideration is not catered for in the design. Over and above this fact, the proper maintenance of earthing at the site becomes all the more critical in this perspective.

C. Data Links / Interfaces Winterization

Interface links between the components are data and power based. Data links might include the Ethernet/serial links with supporting routing cables or interface panels, whereas, power links have distribution panels, supplying power requirements to the computing and sensing equipment. Interface links along with power support systems have direct and/or indirect exposure to cold climatic conditions and they are under sudden transitional states, hence are most vulnerable to degradation and failure.

D. Power Cable Insulation

Electrical insulation of external power cables can be another possible cause of system failure. Many of the insulations normally used on electrical wires and cables are not compatible with colder temperatures. Cracking of the insulation exposes the conductor to the environment creating a serious hazard. This is particularly a problem for the extension cords used outdoors. Several polyvinylchloride (PVC) insulations that are commonly used as electrical insulation do not withstand flexing at low temperatures, in the range of or below $-30^\circ[C]$, PVC insulations crack and peel off leaving exposed conductors, which can cause short circuiting or develop grounding problems making data unreliable [24], [25].

E. Material & Winterization

The sensitivity of problems encountered in cold regions is largely a function of materials used in the sensor construction and degree of stress, under which they are operated. Some materials get stronger at cold temperatures while other materials can be altered to become more cold tolerant [24], [22]. Similarly, sensor winterization can be another possible reason for this failure. Sensors must be properly winterized to make them possible to use during winter and reduce cold related wear and breakage [23].

IV. CHARGE TRANSFER CAPACITANCE BASED ICE DETECTION

It can be said that a diversified sensing technique is required for ice detection, which is robust enough to face the harsh environmental conditions. The reliability and consistency of the measured results add on to the significant requirements during the measurement process. The ice detection and measurement through its capacitive properties could be a viable option in this scenario as suggested by Mughal et. al. [14]. Capacitive sensors are considered amongst the reliable and robust sensing options. A capacitive ice detector for monitoring ice formation in power lines has been reported in Michael J. Moser [17]. Furthermore, charge induction based ice detectors

is reported in [16] (mounted in the surface of the road). This ice detector can detect the presence of icing over the road/runway without measuring the icing rate and type of the ice.

To consider other viable options, proximity and touch sensing techniques can be considered. These are widely used in sensing human capacitance. Most of the daily home appliances, mobiles, industrial applications and gadgets use this detection principle as an integral part of their system design. The technique for proximity or nearness detection of human skin has matured over the years and customized microelectromechanical systems MEMS devices are available for integration as per the requirement of the design. However, use of this approach is not tested for the purpose of ice detection and has not been reported yet. This technique could be applied for ice detection including the water layer along with icing because of its advantages based on implementation flexibility and design robustness.

A. Working Principle

The nearness of the ice on the surface can be measured with the appropriate use of selection of dielectric material and charge transfer process. The field generated due to charge transfer can be thought of as a forced field applied, which ultimately can be measured in return when the presence of ice is detected on the surface. To amicably use this technique sensor’s electrode should be adequately designed for detecting the proximity of the target material with a quantified output. The field is self-generated through any type of conductive material and the response is processed in a customized way to compute the changes in the measurement. The technique can be thought of as an active measurement methodology where self-generated field is repeated to a known threshold level and de-activated to analyze the properties of the material during the silent phase field generated. The charge transfer technology can be implemented / tested for two different schemes based on capacitive sensing

- i Self-capacitance oriented
- ii Mutual capacitance oriented

In *Self-Capacitance* approach, the electrode used for sensing is a single conductive plate; second plate of capacitor is in fact the circuitry or earth ground. The sensing electrode is merely an open circuit plate or alternatively describing we can measure the self-capacitance of this plate. Electrode is underneath the dielectric panel so there is no direct galvanic connection to measuring circuit in the presence of ice or other substance. This technique is aimed to make detection measurements in case of external object presence near electrode. The detection is made because of the effect that its presence has on the capacitance of the electrode. The equivalent circuit of the self-capacitance circuit is shown in Figure 1.

Here, GND is the printed circuit board PCB ground and EARTH is earth ground. The coupling capacitance between ground and earth is few tens of pico-farad. Sampling capacitors store the charge during burst of pulses applied to the electrodes and normally have recommended values in nano-farads, which are further tuned to the design requirement to achieve larger detection threshold. Therefore, key highlights are

- Assuming $C_s \gg C_x$, and $C_f \gg C_x$ and C_t
- C_x and C_t are of parameters of interest

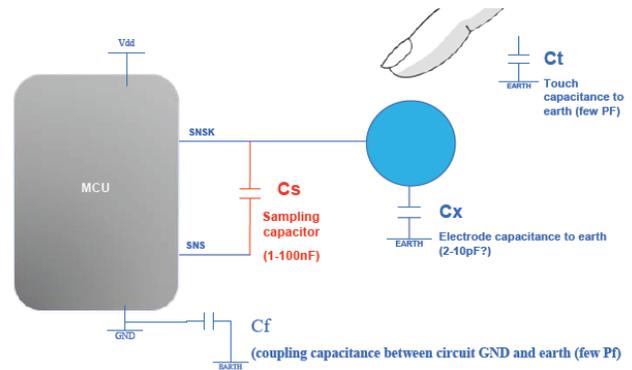


Figure 1. Self Capacitance Equivalent Circuit, [1]

- Increasing C_s = Increased differential sensitivity and makes the burst length longer and improves resolution

In Mutual capacitance approach, each sensing electrode pair contains a field drive electrode and a receive electrode as shown in Figure 2.

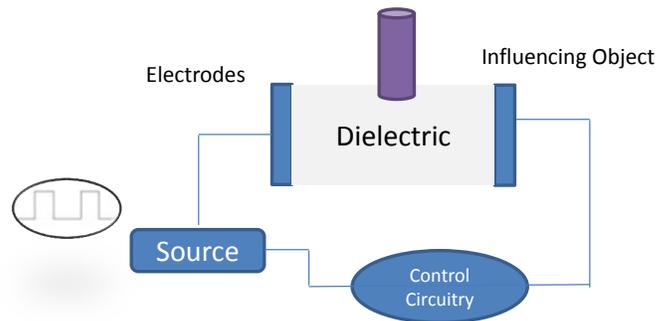


Figure 2. Ice Detection by Mutual Capacitance (Broad Scope)

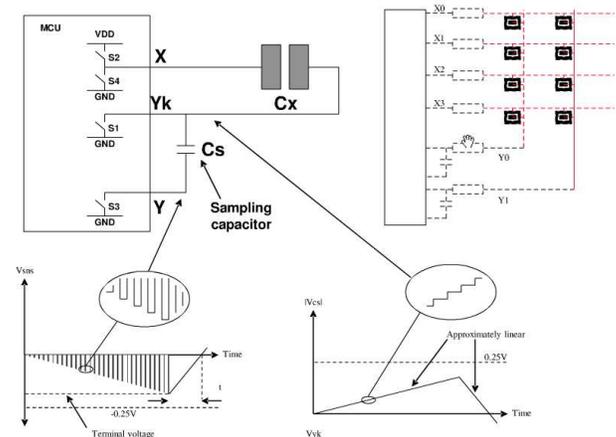


Figure 3. Mutual Capacitance Equivalent Circuit, [18]

Signal that couples through the mutual capacitance of the electrode structure is collected onto a sample capacitor, which is switched by the chip synchronously with the drive pulses.

A burst of pulses is used to improve the signal to noise ratio; the number of pulses in each burst also affects the gain of the circuit, since more pulses will result in more collected charge and hence will provide more signal [18].

By modifying the burst pulse length, the gain of the circuit can be easily changed to suit various electrode sizes, panel materials, and panel thicknesses. After the burst completes, the charge on the sample capacitor is converted using a slope conversion resistor which is driven high, and a zero crossing is detected to result in a timer value, which is proportional to the pair electrode charge coupling, which also reflects charge absorption caused by external intruding material. The presence of intruding object absorbs charge, so the measured signal decreases with its presence. The burst phase causes the charge on the sample capacitor to ramp in a negative direction, and the slope conversion causes a ramp in the positive direction on the capacitor; the net effect is that the conversion process is dual slope, and is largely independent of the value of the sample capacitor and is highly stable over time and temperature [19].

B. Cold Climate Implications & Performance

Both capacitive mechanism are majorly comprised of electrode sets and PCB hardware. The electrode design has the flexibility to adopt any shape as it is flexible enough to be mounted on any type of support. The electrode plate can be covered with different types of dielectric material much more resistant to harsh weather and climatic effects; since there are large number of thin dielectric materials available commercially nowadays. Figure 4 shows the simple operational scheme of the overall charge transfer based icing sensor design.

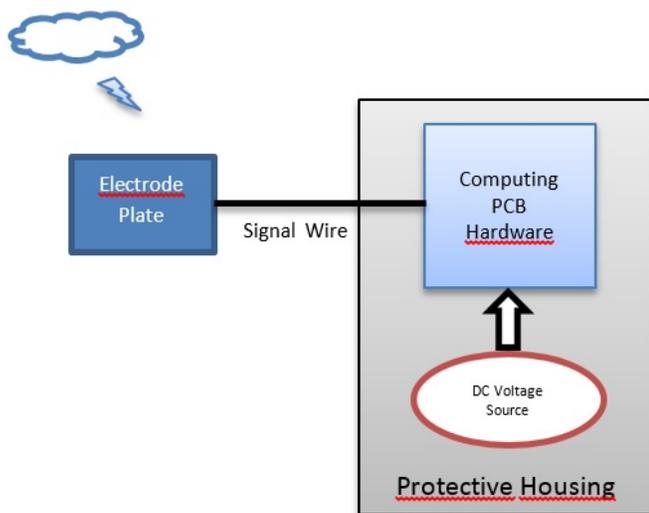


Figure 4. Operational Scheme of Charge transfer Ice sensing

The only exposed part is the electrode, which may be the copper trace on a printed circuit board covered with weather resistant dielectric material of known parameters. The weather resistant dielectric coating/covering can ensure the protection required to avoid corrosion of the plate. The low voltage requirement [milli-volts] of the MEMS devices makes the design feasible for battery operated system in a remote location.

1) *Electrode Panel Selection:* The flexibility of electrode configuration provides different schemes in terms of shapes and forms. The layout of the electrode for charge transfer scheme is focused on the maximum transferring of charge either to the adjacent electrode (receive) or provide effective ground loading. Common electrode materials include copper, carbon and silver ink. The lower the $[\Omega/sq]$ resistivity of the material, the better as it makes control of any RC time constants (which play key role in detection measurement) much easier. The $[\Omega/sq]$ rating choice is coupled with the shape and size of the electrode. Larger in length and thin electrodes or traces build up resistance extremely quickly, even for relatively low resistivity. Common front panel materials include glass, plexiglass and polycarbonate. The panel thickness and its dielectric constant ϵ_r play a large part in determining the strength of electric field at the surface of the control panel.

Glass has higher ϵ_r than most plastics as higher numbers deduce the fields to propagate more effectively. Thus a 5[mm] panel with ϵ_r of 8 will perform similarly in sensitivity to a 2.5[mm] panel with an epsilon of 4, considering all other factors equal. A panel up to 10[mm] thick is quite usable, depending on electrode spacing and size [17]. The circuit sensitivity needs to be adjusted during development to compensate for panel thickness, dielectric constant and electrode size. With increase in thickness of material signal to noise ratio SNR worsens hence always thickness of the front panel material is to be kept small. Materials with high relative dielectric constants are also preferable for front panels as they help to increase SNR.

V. TRADE-OFF

The self and mutual charge transfer techniques are primarily differentiated based on the electrode configuration. The single electrode and its multiple scheme implanted on a single surface adheres to ground loading influence for the measurement, which eventually will attributes to the sensitivity of the ice measurement. The main tradeoffs amongst the self and mutual capacitance techniques are sensitivity, range, noise rejection, ground loading and probability of false detection. The self-capacitance design is more sensitive in nature. The field is spreads outwards the electrode in the dielectric environment and ground loading is provided by the external influencing of the object, in our case will be ice or a water film. But with increase in sensitivity comes the inclusion of the noise in the circuitry, which is un-desirable. The noise in self capacitance might be increased so sensitivity tuning is the vital for design based on this methodology. The sensitivity in case of capacitive based design has several factors ranging from the electrodes design to the charging/discharging mechanism affecting the sensitivity of the sensor. This includes electrode dimensions and shape, ground loading, return path, supply voltage and charging pulses duration.

VI. DISCUSSION AND CONCLUSIONS

The atmospheric icing detection and measurement in harsh cold climate is a demanding challenge. The need is more demanding with the latest developments in the high north regions to explore for energy ventures, which have initiated the infrastructure and channelizing of resources. The ice accretion and winterization factors can be very easily overlooked during the development process, which has the lasting impact in cost

related damages. Therefore, icing parameters like rate, type, thickness could play vital role in areas for instance deicing feed-back mechanism for efficient ice removal, creation of geographical ice maps of the particular region and many more. The icing parameters discussed need reliable sensing methodology for acquisition and measurement in the harsh cold regions.

A preliminary series of experimentation were performed in Cold Room Chamber of Narvik University College. The ice samples used were collected from the freezing process of the commercially available freezers. The charge transfer technique outputs zero crossover due to the dielectric variation between different samples shows clear potential this effective/potential technique for Cold Regions. The zero crossover is a real time technique and hence the delays associated with this technique are minimum. The results can be seen in Fig. 5.

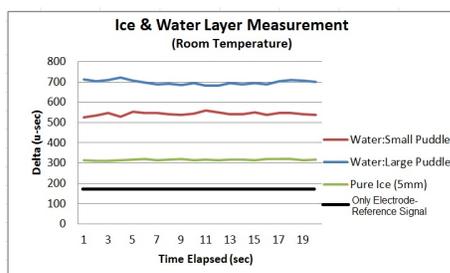


Figure 5. Ice and Water Detection Ranges

The charge transfer based techniques well known for human capacitance application can be utilized for icing rate, type and thickness. The self or mutual capacitance basics can be employed in the design based on charge transfer method. The electrode configuration schemes can be used in pairs or as an individual based on the optimum tradeoff amongst the design options. A small printed circuit board aided with the specific electrode configuration can be manufactured as a basic prototyping icing sensor. The benefit of reshaping the electrodes and design to any form with capability to change different dielectric material suitable to harsh environment can be used as a starting point to develop a robust prototype. This would enable to detect and measure the icing parameters in real time embedded platform with low power consumption; which is ideal for remote installations.

ACKNOWLEDGMENT

We acknowledge the research funding from Research Council of Norway, project no. 195153/160 and partially by the consortium of the ColdTech RT3 project - Sustainable Cold Climate Technology.

REFERENCES

[1] http://www.ineltek.com/ru/files/Capacitive_Touch_Intro_2011_05.pdf. Last accession date 23rd March 2014.
 [2] K. F. Jones and K. Z. Egelhofer, "Computer model of atmospheric ice accretion on transmission lines", CRREL Report, 1991.
 [3] L. J. Battan, "Cloud physics: A Popular Introduction to Applied Meteorology", Dover Publications, ISBN 978-0486428857, 2003.
 [4] H. F. Foder, "ISO 12494 - Atmospheric icing on structures and how to use it", ISO 12494:2001, ISBN 1-880653-51-6, June 2001.

[5] S. Fikke, G. Ronsten, A. Heimo, S. Kunz, M. Ostrozlik, P. E. Persson, J. Sabata, B. Wareing, B. Wichura, J. Chum, T. Laakso, K. Santti, and L. Makkonen, "Cost 727 - Atmospheric icing on structures; measurement and data collection on icing", ISSN 1422-1381, MeteoSwiss, 2007.
 [6] M. Drage, "Atmospheric icing and meteorological variables - Full Scale Experiments and Testing of Models", Phd Thesis, Department of Geophysics, Bergen University, Norway, 2005.
 [7] J. Martinec, "Expected snow loads on structures from incomplete hydrological data", Journal of Glaciology, 1977, pp. 185195.
 [8] W. Gareth, "Remote sensing of snow and ice", Chapman & Hall Publishers, ISBN 978-0415298315, 2006.
 [9] N.I. Popov and V. V. Holodov, "On the applicability of icing measurements (in russian)", Trudy GGO 408, 1978, pp. 4448.
 [10] , "Design criteria of overhead transmission lines", IEC 60826, 2003.
 [11] L. Makkonen, "Analysis of rotating multicylinder data in measuring cloud droplet size and liquid water content", Journal of Atmospheric and Oceanic Technology, 9(3), 1992, pp. 258263.
 [12] P. Lehtonen, "Experience on the iso reference collector", Experience on the iso reference collector. 8th IWAIS, Reykjavik, Iceland, 2000, page 357.
 [13] H. Rothig, "A device for continuous measuring and recording of ice accretion", Abhandlungen-Meteorologische Dienst DDR 107 (in German), 1973, pp. 2630.
 [14] U. N. Mughal, T. Rashid, M. S. Virk, and P. O. Nyman, "E driven technology for Cold Regions", CogInfo Comm, ISBN 978-1-4799-1546-0/13, 2013, pp. 683686.
 [15] J. M. Moser, B. George, and H. Z., "Icing detector for overhead power transmission lines", IEEE I2MTC, 2009, pp. 11051109
 [16] A. Troiano, E. Pasero, and L. Mesin, "New system for detecting road ice formation", IEEE Transactions on Instrumentation and Measurement, 60(3), March 2011, pp. 1091 1101.
 [17] <http://www.atmel.com/Images/doc10752.pdf>. Last accession date 2rd May 2014.
 [18] http://www.atmel.com/images/qmatrix_white_paper_100.pdf. Last accession date 25th Feb 2014.
 [19] http://www.atmel.com/Images/qt60240_r8.06.pdf. Last accession date 25th March 2014.
 [20] K. B. Shraavan, U. N. Mughal, and M. S. Virk, "Experimental study of relative permittivity of atmospheric ice", International Journal of Energy and Environment, 4(3), 2013, pp. 369-376.
 [21] J. Latham, "The electrification of snowstorms and sandstorms", Journal of Royal Meteorological Society, 90(383), January 1964, pp. 91-95.
 [22] P. K. Dutta, "Behavior of materials at Cold Region Temperatures", U.S Army Corps of Engineers, Cold Regions of Engineering Lab (CRREL), Hanover, 1988.
 [23] D. Diemand, "Winterization and winter operation of automotive and construction equipment", U.S Army Corps of Engineers, Cold Region Research and Engineering Lab. Hanover, 1992.
 [24] D. R. Freitag and T. T. McFadden, "Introduction to Cold Regions Engineering", ASCE Press. 75, 1966.
 [25] D. Rosato and R. T. Schwartz, "Environmental effects on polymeric materials", John Wiley and Sons, 1968.
 [26] M. S. Virk, T. Rashid, U. N. Mughal, K. Zaman, and M. Y. Mustafa, "Multi Sensor Atmospheric Icing Station Performance in Cold Climate - A Case Study", The Seventh International Conference on Sensor Technologies and Applications, ISBN 978-161208-296-7, August 25-31, 2013, pp. 220:224.

Evaluation of Torque and Axial Loading Physics for Atmospheric Icing Sensors

Umair N. Mughal and Muhammad S. Virk

Atmospheric Icing Research Team
Narvik University College,
Narvik, Norway
Email: unm@hin.no

Abstract—Measuring icing load and icing rate are important parameters for an atmospheric icing sensor. A new icing sensor has recently been designed and developed at Narvik University College for measuring atmospheric icing rate, icing load and icing type. Unlike the existing atmospheric icing sensors commercially available in market, which uses the axial loading for measuring icing load and icing rate, this new sensory system measures icing load and icing rate using the torque loading physics. The performance of this new icing sensor has been tested and validated at cryospheric environment simulator Japan. This research work focus upon the lab based experimentation and evaluation of axial loading and torque loading sensory physics during an icing event. Results show a significant performance difference between torque and axial loading physics for atmospheric icing sensors.

Keywords—Atmospheric Icing Sensor; Icing Load; Icing Rate; Axial Loading; Torque Loading

I. INTRODUCTION

A. Atmospheric Ice

Atmospheric icing primarily occurs due to the accretion of ice on structures or objects under certain conditions. Generally an icing event is defined as periods of time where the temperature is below $0^{\circ}[C]$ and the relative humidity is above 95%. Ice accretion can be defined as, *any process of ice build up and snow accretion on the surface of objects exposed to the atmosphere* [4]. Liquid below $0^{\circ}[C]$ is called supercooled liquid, which creates atmospheric icing. This accretion can take place either due to freezing precipitation or freezing fog. It is primarily freezing fog that causes this accumulation, which occurs mainly on mountaintops, which is particularly dominant in Norway [1]. It depends mainly on the shape of the object, wind speed, temperature, liquid water content (amount of liquid water in a given volume of air) and droplet size distribution (conventionally known as the median volume diameter). Sometimes ice crystals have a thin coating of water even at temperatures well below freezing, which form clouds. These ice crystals join together to form flakes and reach ground as snow via air passage with temperature less than zero [5]. Snow crystal forms when tiny supercooled cloud droplets (about $10[\mu m]$ in diameter) freeze. However, if the outer coating of water freezes of the combined ice crystals during its path via a air passage with temperature less than zero then they form snow pellets, which are sometimes called as *graupel*. Also, if there is a hot layer of air just below the cold layer of air then they reach ground as *sleet*. Sometimes these small droplets with water coating are not successful in combining with other droplets and they get affected by the surrounding air currents but eventually they fall on the surface, they are called *drizzle*, which is different than *fog* as it doesn't

fall. Drizzle drops are of $0.5[mm]$ but drops larger than this are raindrops. *Hail* is another form of ice. Ice crystals when drop towards the surface they are sometimes passed through very moist air passage due to which they are coated with liquid water, which by strong wind is moved upward where water freezes and then they move down and get covered by liquid water and then moves up to become solid. This process is continued and then it becomes to fall. *Cloud formation* is motivated by a seed or crystalline skeleton on which very tiny, supercooled water droplets can freeze to form snowflakes or soft ice (graupel). Naturally, these seeds are random particles of soil, dust, sand, and salt. Artificially they are of two types *glaciogenic* (ice forming e.g., silver iodide or dry ice crystal) or *hygroscopic* (water attracting e.g., small salt particles as potassium chloride) [8].

B. Applications of Atmospheric Icing Sensors

Atmospheric icing is a natural phenomenon, which cannot be avoided in Cold Regions. However, it definitely has some physical loading characteristics on human activities and their associated inventories. On the basis of its loading aspects we can distinguish the effects of atmospheric ice on three classes [8], which are

- a *Static Loads*: Atmospheric ice, particularly (rime and glaze) is when deposited on some static structure, it increases its mechanical weights. Hence it constrains, the design characteristics (particularly factor of safety) of any civil or mechanical structure to be developed in Cold Regions.
- b *Dynamic Loads*: This atmospheric ice, when deposited on the dynamic structures e.g., free dynamic structure as like power cables and motorized dynamic structures as like wind turbines or automobile or ships/boats create additional dynamic loads on these surfaces, which need to be overcome by either anti/de icing techniques in case of free dynamic structures or through increasing the power delivered to such systems.
- c *Wind Action on iced structure*: It is expected that if the structure is iced, its effective geometry will be altered, which in turn reduces the aerodynamic efficiency of structures. This additional drag cannot be completely controlled, however efficient anti/de icing techniques through a good feedback from atmospheric icing sensor may reduce the losses.

Atmospheric ice can be a big problem for different industries working in Cold Regions. The potential affected stakeholders can be [8],

- i *Wind Turbine Industry*: One can improve the efficiency of Wind Turbines by installing an atmospheric icing sensory

network, which should control turbine pitch by providing feedback to filter out the atmospheric icing loading and rate errors

- ii *Oil and Gas Industry due their Onshore and Offshore Installations:* These sensors can be utilized in the big installations of oil and gas platforms in Arctic Region. The output from these sensors can be a good feedback for active monitoring of atmospheric icing activities and its remedies through a good anti/de icing system
- iii *Automobiles Industry:* In Automotive industry these sensors are required to be installed in order to sense the real time atmospheric icing activity on the road surfaces. These sensors can be interconnected with the Road Services via GPS so that accelerated responses for road maintenance can be conducted.
- iv *Power Industry due to the ice on the long power networks:* On power networks, one cannot completely calculate the icing load (which can be very critical). This load if it remains on the power line can be less dangerous than if it suddenly falls its reaction can damage the system. However, active monitoring of icing rate and icing load may reduce this problem if it is connected with some semi active dampers (e.g., magnetorheological damper) mounted on the power cable connections on the pole.

This paper is divided in five sections. In section II, it is aimed to discuss the atmospheric icing load measurement techniques through International Standards/Recommendations and some details about the commercially available atmospheric icing load sensors along with their necessary physics. Section III deals with the torque loading basics and the need to utilize this technique for measuring atmospheric icing load and rate. Section IV starts with an introduction of Cryospheric Environmental Simulator and leads towards performing series of experiments to validate the usefulness of torque loading over axial loading. The concluding section V is divided into smaller parts to comment on various aspects of experimental results.

II. ATMOSPHERIC ICING LOAD MEASUREMENT

To design a new atmospheric icing sensor, in order to measure icing load and icing rate, it is important to understand the existing atmospheric icing rate and icing load sensors. Power requirements for the removal of snow and ice is different hence to distinguish between snow and ice can be considered to be a limiting factor for de-icing system, because most devices for the removal of snow are normally ineffective for the efficient removal of ice or hard packed snow [3]. It is therefore important to distinguish between different types of atmospheric ice. Presently as found, there are commercially available atmospheric icing load sensors, which are only based upon axial loading physics e.g., Ice Load Monitor and Ice Meter [2]. As recommended in international standard ISO 12494 [7], a standard way of measuring ice accretion is to measure the load of atmospheric ice on a steel rod that is 0.5[m] high (1[m] if heavy icing is expected) and has the diameter of 30[mm]. The rod must be freely rotating or *forced to rotate at a constant speed*. When ice accumulates on the steel rod, aerodynamic drag will cause it to rotate in the case of free rotating icing load monitor, always facing the least amount of the iced part towards the wind. This doubling of length of the steel rod due to heavy icing is primarily aimed

to uniform the drag distribution along its profile. By measuring the weight of the iced steel rod with the help of load cells, the amount of ice that has accreted can be determined. Two ice sensors (ice monitor & ice meter) have been developed on the basis of this technique. The following sections describes these two sensors. However, till date none of the atmospheric icing sensor is utilizing torque loading physics, which is also recommended in ISO 12494 [7].

A. The Ice Monitor

The Ice Monitor measures the mass of accumulated ice gravimetrically. The working element is a freely rotating steel pipe resting on a rod placed on a load cell. As ice accumulates on the freely rotating steel pipe, the ice load is weighed by the load cell, see Figure 1. The Ice Monitor is manufactured by SAAB Technologies and was initially developed for power line surveillance systems. It can measure the rate because the readings from the load cell are recorded with time. The Ice-Monitor is not able to detect ice over a wide area and cannot distinguish between the two types of in-cloud icing.



Figure 1. Ice Load Monitor [13]

B. Ice Meter

The ice meter was developed by the Institute of Atmospheric Physics, Prague, Czech Republic. It measures the mass of icing accumulated on the surface of the collector. It has a horizontal rod, which is coupled with a cylindrical collector to the tensometric. The cylinder is orientated vertically in order to eliminate the detection of wet snow as much as possible but the orientation of this cylinder can be changed to horizontal, if required, see Figure 2. In this sensor, the mass of accumulated ice is measured by means of a tensometric bridge (strain gauge load sensor) the output of which is tied to the precise AD converter. The digital signal is preprocessed by a micro-controller, which assigns the time and stores the data into the device memory. In order to prevent the freezing of the horizontal rod, which couples the cylindrical collector to the tensometric, which is located together with the electronics in the housing. The passage through the housing may be heated depending on the passage temperature. A test electromechanical impulse is applied each hour to verify the free force transition to the tensometric, and thus to check whether the acquired data are reliable or not [6].

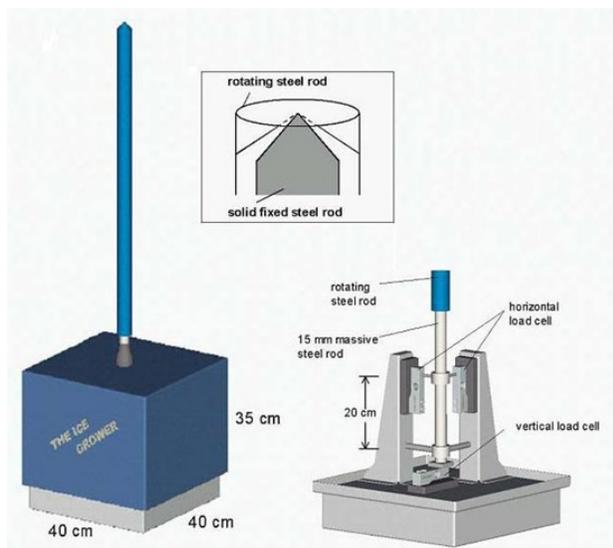


Figure 2. Ice Meter [6]

C. Study of the Physics of Axial Load Based Ice Sensing Methodology

The load is normally described by the way they are applied. If the line of action of the loads is applied parallel to the surface they are called shear loads and when they are applied perpendicular to the surface they are called axial loads, see Figure 3. Although there are many techniques to measure these loads but the most common are associated with piezoelectric sensor and strain gauge sensing elements.

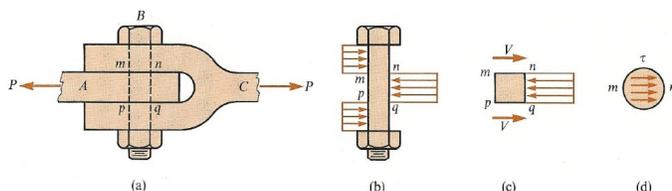


Figure 3. Axial and Shear Load (a). Typical clevis joint, (b). FBD of bolt, (c). FBD of section mnqp, (d). Shear Force on Section mn

D. Piezoelectric Sensing

This sensing technique works by converting the force or strain by converting it into electric charge. Its primary application is in the vibration industry, however it can also be used for load sensing. Piezoelectricity is a property, which is possessed by some materials, which is activated when are acted by some force. Piezoelectric materials have a recoverable strain of 0.1 % under electric field; they can be used as sensors. They may be polymers (e.g., polyvinylidene fluoride *PVDF*) or ceramic materials (e.g., Lead Zirconate Titanate *PZT*). If you apply a static force to a piezoelectric force sensor, then the charge output generated initially will leak away and the output of the sensor will ultimately return to zero. This discharging rate of the charge is exponential and is based on the sensor's discharge time constant, *DTC*. Mathematically the charge characteristics can be described as [9],

$$\begin{aligned} Q_x &= d_{xy} F_y \frac{b}{a} \\ q_x &= Q_x e^{-\frac{t}{RC}} \end{aligned} \quad (1)$$

where Q_x is the amount of charge stored in the piezoelectric material due to the orthogonal force F_y , a and b are the characteristic dimensions of the sensor parallel and orthogonal to the F_y respectively, d_{xy} is the characteristic piezoelectric coefficient (or calibration coefficient of the material), q is the instantaneous charge, R resistance prior to amplified and C is the total capacitance prior to amplifier.

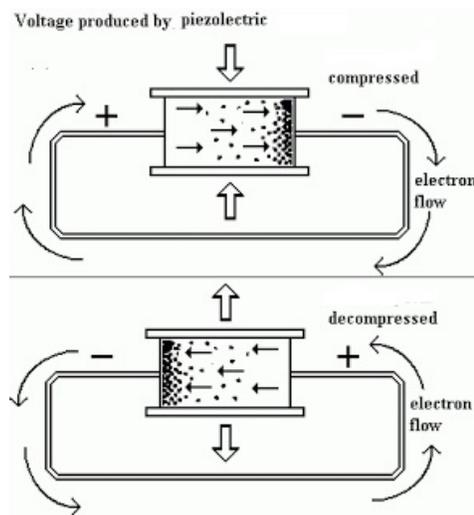


Figure 4. Piezoelectric Sensing [9]

E. Strain Gage Sensors

This technique is more suitable for precise measurement of static loads or a quasi dynamic loads. Within the elastic limits of the material, when a stress is applied to an electric conducting material its electrical conductance R_x changes due to the change in its geometry. This change in conductance is then further measured using a wheatstone bridge (special combination of resistors), which is then further used to measure the load. The resistance change in the wheatstone bridge delivers a voltage, which deliver the strain and hence the force. This special combination of wheatstone bridge is shown in Figure 5b.

$$\begin{aligned} q_x &= Q_x e^{-\frac{t}{RC}} \\ \frac{R_2}{R_1} &= \frac{R_x}{R_3} \\ R_x &= \frac{R_2}{R_1} R_3 \\ V_G &= \left(\frac{R_x}{R_3 + R_x} - \frac{R_2}{R_1 + R_2} \right) V_s \end{aligned} \quad (2)$$

This concept has been utilized in the Ice Load Monitor Sensor by SAAB Technologies, see Sec. II-A.

F. Problems with existing Ice Load and Icing Rate Measurement Systems

1) *Measurements in Czech Republic [7]*: The icemeter has been operated on the Milesovka peak from 2000. Although icemeters installed in the site measured correctly most of the time, but there were also some time periods when the instruments gave obviously wrong negative values. These times

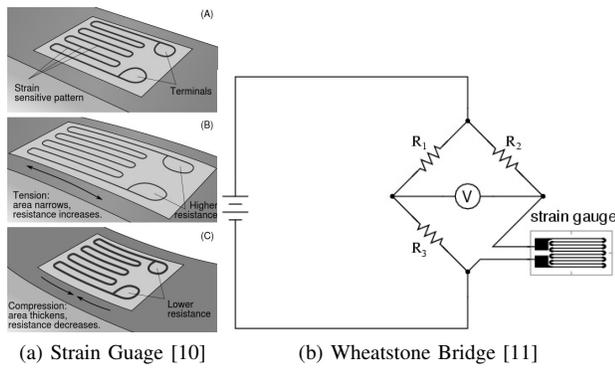


Figure 5. Strain Gauge Sensing

can usually be associated with the periods when the horizontal rod coupling the tensometric with the vertical collector became icebound to the instrument housing. Thus there was no free force transition, which can be identified by not observing a proper electromechanical pulse in the data. The following changes are expected to be made in icemeter for its better performance,

- i Possibility to build an instrument with a rotating collector.
- ii More focus on the sensors that measure accumulated icing.

III. TORQUE MEASUREMENTS

An analytical expression between mass moment of inertia and the supplied current (3) was developed (for more see Mughal et. al. [14]),

$$J = aI_{in} - b \quad (3)$$

where $a = \alpha V_{in} / \beta \gamma \omega_m^3$ and $b = \eta / \beta \gamma \omega_m^3$ will be constant. The (3) relation was further modified to take the form (4) in order to compare the results of this new sensor with a standard ice load sensor.

$$\delta m_{ice} = A \delta I_{in(ice)} + B \quad (4)$$

Where $A = \frac{a}{R_{ice}^2}$ and $B = \frac{aI_{in} - J - b}{R_{ice}^2}$. The (4) relation is named as *MuVi Current-Mass ($\delta I - \delta m$) Relation*.

A. Requirement of Torque Loading

As it is understood from [7] that presently there is no atmospheric icing sensor in the commercial market that can measure the icing load, which is uniformly distributed along the sensory surface. All the commercially available sensors (e.g., Ice Load Monitor), measure the non uniform loading on a freely rotating surface, however it is recommended in [7] that a rotary sensory system will be more suitable. A new sensor have been developed by the authors to measure icing load and icing rate. This sensor is based upon the rotary dynamics where the collector rotates at a constant speed in order to allow ice to uniformly distribute around the rotary collector.

IV. CRYOSPHERIC ENVIRONMENT SIMULATOR CES, SHINJO JAPAN

Cryospheric Environmental Simulator is an experimentation facility of NIED at Shinjo, Japan. They provide the capacity to perform basic and applied studies related with snow and ice disasters using a snow fall machine and icing wind tunnel. For this experimentation to evaluate the performance of this new sensor, Cryospheric Environmental Simulator CES was the most suitable choice, which had the facility to test this sensor both in Icing Wind Tunnel and Snow Simulator. The facilities and specifications of CES are provided in Table I, which are taken from [12],

TABLE I. FACILITIES AND CONTROLLABLE PARAMETERS IN COLD ROOM AT CES, NIED JAPAN

Facilities	Specifications		
	Variables	Limitations	Comments
General Conditions in the Cold Room	Temperature	$-30^\circ \rightarrow +25^\circ$	—
Snowfall Machine A	Snowfall Intensity	0 —	water equivalent
	Crystal Type	Dendrites	$0.5 \rightarrow 5[mm]$
	Area	$3 \times 5[m]$	—
Snowfall Machine B	Snowfall Intensity	0 —	water equivalent
	Crystal Type	Sphere	Diameter $0.025[mm]$
	Area	$3 \times 5[m]$	—
Rainfall Machine	Rainfall Intensity	0 —	—
	Area	$3 \times 5[m]$	—
Solar Simulator	Solar Radiation	0 —	—
	Area	$3 \times 5[m]$	—
Experiment Table	Size	$3 \times 5[m]$	—
	Inclination	$0 \rightarrow 45^\circ$	—
Wind Tunnel	Wind Speed	0 —	—
		$10[m/s]$	—

A. Calibration of Ice Load Monitor and new atmospheric Icing Sensor

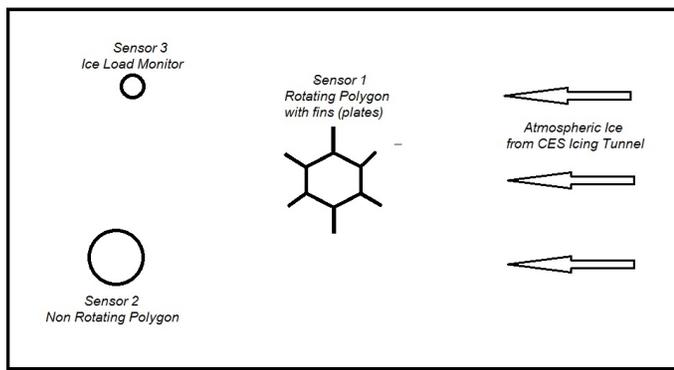
1) *Ice Load Monitor*: It is found that this standard sensor have never been tested in any standard icing tunnel or snow simulator as like Cryospheric Environmental Simulator and if it is tested then it is not published/reported (the details of CES Simulator will be discussed in Sec IV). The Ice Load Monitor deliver a current output as a measure of icing load. This sensor was calibrated at Narvik University College and the calibration equation is given by (5). For more details on this calibration, see Mughal and Virk [15].

$$I = 0.0017 \delta \delta m_{ILM_{cal}} + 4.435 \quad (5)$$

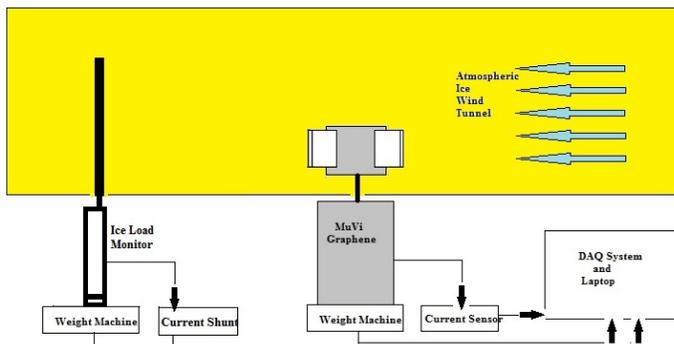
2) *New Atmospheric Icing Sensor*: This was calibrated using standard rotary masses on the sensor shaft and measuring the output current as a measure of mass moment of inertia. The electronic setup for calibration can be seen in Figure 6c. The calibration equations thus obtained is given by (6) (for more details on this, see Mughal et. al. [14]),

$$I = 2.91 \times 10^{-7} J_{MuVi_{cal}} + 328.64 \quad (6)$$

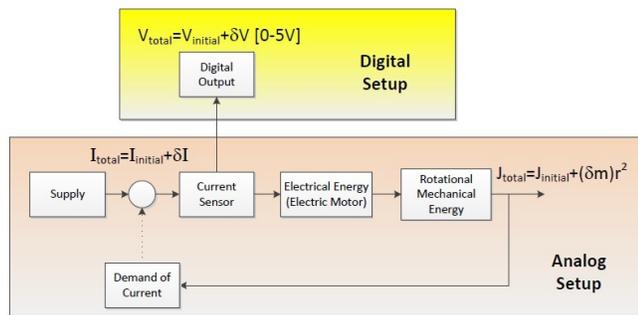
Where J is in $[gm\ mm^2]$ and $I_{measured}$ is in $[mA]$.



(a) Experimental Table - Top View



(b) Experimental Table - Side View



(c) Electronic Setup

Figure 6. Experimental and Electronic Setup, CES Japan [15]

B. Experiment performed at CES, Japan

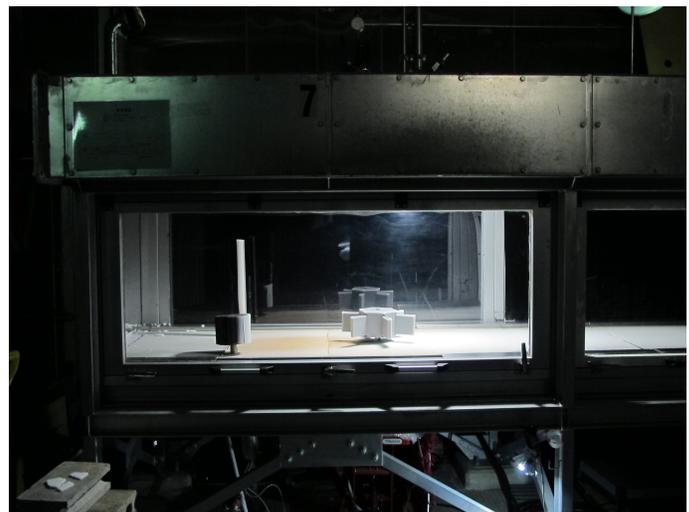
The experiment was performed in Icing Wind Tunnel using three different sensors, arranged on the sensory table as can be seen in Figure 7a where Sensor 1 is the rotary sensor with complete electronics of this new sensor, Sensor 2 is the additional sensor, which is just a geometric shape to understand the deposition of atmospheric ice in order to optimize the non rotary parts parts of new sensor. The experimental conditions during the experiment can be seen in Table II, which is followed by Figure 7b reflecting the measurement setup under the experimental table. The graphical results of this experiment are shown in Figure 8 [8].

TABLE II. EXPERIMENTAL CONDITIONS FOR EXPERIMENT 1

Before Experiment	
Condition/Variables	Specifications
Date of Experiment	17/02/2014
Experiment Number	1
Experimental Facility	CES Icing Wind Tunnel
Sensor 1	Rotating hexagon with plates
Sensor 2	Non rotating cylinder
Sensor 3	Freely rotating Ice Load Monitor
Tunnel Temperature	-15°
Time of Experiment	170[<i>min</i>]
Wind Speed for first 60[<i>min</i>]	6[<i>m/s</i>]
Wind Speed for next 110[<i>min</i>]	10[<i>m/s</i>]
Mass of Sensor 1 (without ice)	8936[<i>gm</i>]
Mass of Sensor 2 (without ice)	591[<i>gm</i>]
Mass of Sensor 3 (without ice)	0[<i>gm</i>]
After Experiment	
Mass of Sensor 1	9735[<i>gm</i>]
Mass of Sensor 2	866[<i>gm</i>]
Mass of Sensor 3	480[<i>gm</i>]



(a) Sensors Arrangement in Icing Tunnel - Front View



(b) Sensors Arrangement in Icing Tunnel - Side View

Figure 7. Experimental Setup in Icing Tunnel

The icing rate on this new sensor can be determined using the relation (7) whereas the icing rate on Ice Load Monitor can be determined using the relation (8). The associated curves for these equation are shown in Figure 8a, which clearly reflect that this new sensor, have more capability of holding uniform distribution of atmospheric ice then a freely rotating Ice Load Monitor. Also, the manufacturers of Ice Load Monitor claim that after 40[gm] their sensor will be able to predict icing load and icing rate and as found by experiments this 40[gm] on ice load monitor is collected in around 30[*min*] whereas the same 40[gm] of mass on new sensor is collected in first 5[*min*] and that too uniformly. This uniform distribution is very important for additional mass moment of inertia on the sensor to deliver additional current as a measurable quantity. The calibrated relation between mass moment of inertia and current loading is given by (6) whereas the experimental mass moment of inertia and current loading equation is given as (9) and graphical results are shown in Figure 8b. The experimental section was bounded by walls (in icing wind tunnel), and it may be one of the possible reason of deviation between calibrated and experimental $I - J$ Relation but nevertheless this can be adjusted through proper calibration. The additional mass and additional current calibrated equation of Ice Load Monitor is given by (5) and the experimental relation is given by (10) and the results are shown in Figure 8c. These results of Ice Load Monitor shows that for a mass range of 0 → 500[gm] the current variation is around 4.5 → 5.3[*mA*] ideally and 5 → 6.8[*mA*] experimentally, which have more potential of noise interference. The mass current relation of this new sensor was also determined experimentally and is given by (11) and the graphical representation can be seen in Figure 8d. This equation of this new sensor was then compared with the (10), which clearly reflect that this new sensor for a mass domain of 40 → 800[gm] was having a current range of 0 → 150[*mA*], which is quite reasonable to filter noise and other calibration errors. The R^2 value reflect that linear relations of experimental of this new sensor are 96% whereas for Ice Load Monitor this value is 88% linear.

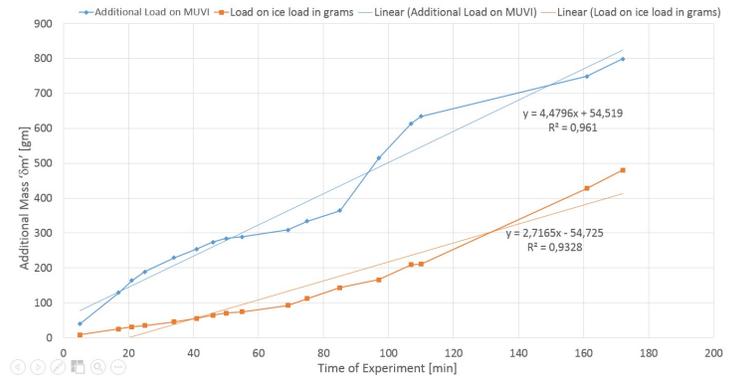
$$\delta m_{MuVi} = 4.480t + 54.519 \tag{7}$$

$$\delta m_{ILM} = 2.716t - 54.725 \tag{8}$$

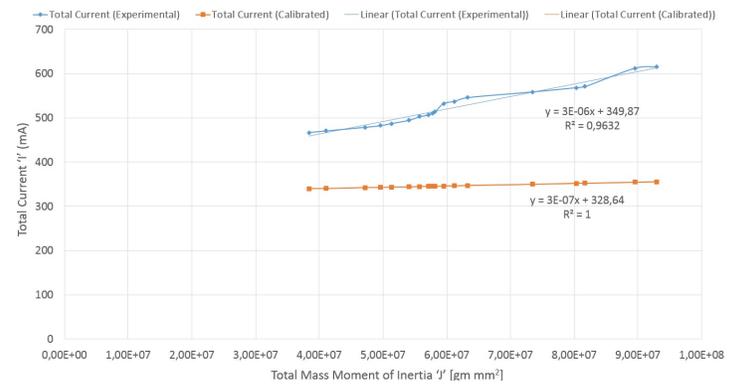
$$I = 2.84 \times 10^{-6} J_{MuVi_{exp}} + 349.87 \tag{9}$$

$$\delta I = 0.0031\delta m_{ILM} + 5.259 \tag{10}$$

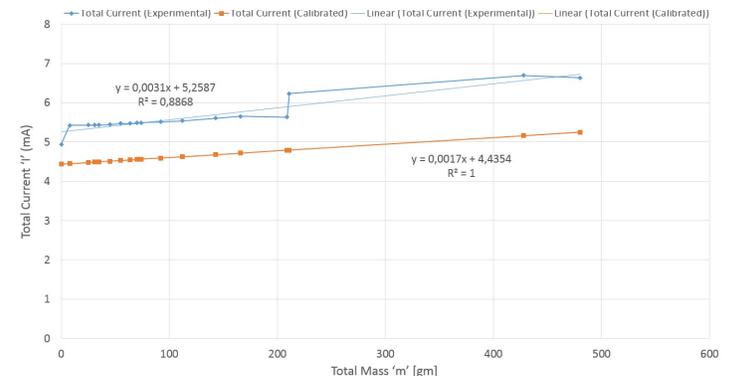
$$\delta I = 0.1936\delta m_{newsensor} - 7.21 \tag{11}$$



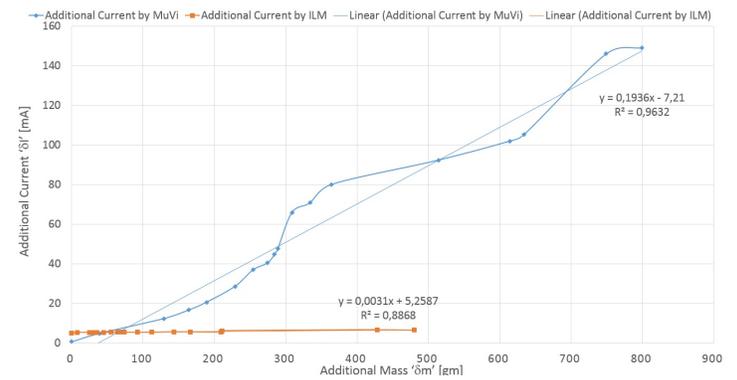
(a) Atmospheric Ice Growth Rate (Mass-Time Curve)



(b) Current Mass Moment of Inertia Curve (I-J Curve) of New Sensor at 6rpm



(c) Current Mass (I-m Curve) of Ice Load Monitor



(d) Additional Current Additional Mass ($\delta I - \delta m$ Curve)

Figure 8. Graphical Results of Experiment 1 in Icing Wind Tunnel [15]

V. CONCLUDING DISCUSSIONS

1) *RPM Fluctuations*: The rotational speed of this sensor was selected to be 6rpm. It is found that if the rpm was 5 there were current fluctuations in the range of mA during rotation without any change in load. Also, during experiment it was found that if the rpm was increased to 8 chances of non uniform deposition of ice on the sensor, which may lead to additional errors.

2) *Sensitivity*: The results of experiment clearly reflect that new sensor provide sufficient deposition of atmospheric ice ice at an optimized rpm of 6. Also, the current range to measure icing load and icing rate is quite reasonable to determine the required parameters. This new sensor is more sensitive than Ice Load Monitor because if it is said that 40[gm] of mass for Ice Load Monitor as the starting mass for new sensor than it is collected within first 5 → 10[*min*] on new sensory unit and it is more than 30[*min*] for Ice Load Monitor. The reason of this mass deposition is due to the rotary physics associated with new sensor and the surface area of Ice Load Monitor is around $50 \times 10^3[mm^2]$ and new sensor is around $112 \times 10^3[mm^2]$.

3) *Saturation*: The saturation limit for the motor was adjusted to be around 1[A]. If the demand in current from the rotary unit of new sensor exceeds 1[A] than the motor would stop. The maximum mass for new sensor was tested for a mass of 800[gm] for a time period of 170[*min*] and it is found that new unit have still more potential to hold the ice mass.

4) *Performance*: The sensor have been tested in the Cryospheric Environmental Icing Simulator at a temperature of $-15^{\circ}[C]$ and have performed satisfactory without any error signal. This sensor have not been tested in the field, however it is aimed to test this sensor in the field in the upcoming Winter 2014 – 15. The calibration error may have been associated with the poor performance of the sensor, however this have been true if the results of experiment were not linear. The experimental results show that the new sensor have a good performance potential than Ice Load Monitor, however Ice Load Monitor is already in use by industry in the field whereas new sensor still has not yet been commercialized.

5) *Error Diagram*: The error diagrams from experiment 1 to experiment 4 reflect that new sensor have very low error in the start but it increases linearly due to the deviation between the experimental slope and calibrated slope of new sensor but nevertheless it can be improved by adopting proper calibration. The error is calculated using the following (12),

$$\%Error = \left| \frac{I_{exp} - I_{cal}}{I_{cal}} \right| \times 100. \tag{12}$$

The error diagram of Experiment between new sensor and Ice Load Monitor as calculated by (12) is shown in Figure 9. The error results shows that the % error range of new sensor is starts from 0% and increases linearly because experimental slope of J-I Relation of new sensor (9) is one order higher than the calibrated slope of (6). The error on Ice Load Monitor is around 10 – 30%. These all results reflect new sensor as a good solution for measuring Icing and Snow Load Rate if properly calibrated as it has more potential to deposit ice and give reasonable reading count, which is required to filter the cold climate current errors.

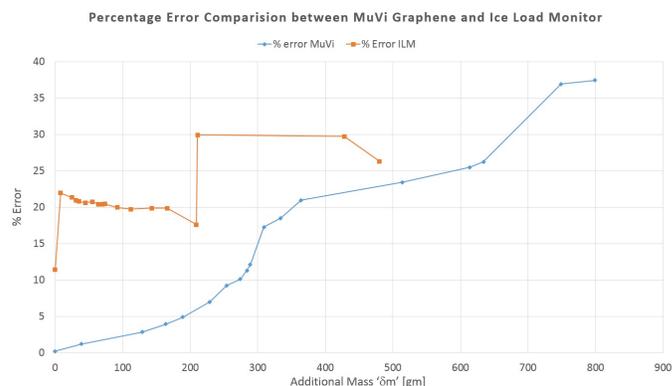


Figure 9. Experiment 1 Percentage Error Diagram

ACKNOWLEDGMENT

We acknowledge the research funding from Research Council of Norway, project no. 195153/160 and partially by the consortium of the ColdTech RT3 project - Sustainable Cold Climate Technology.

REFERENCES

- [1] M. C. Homola, "Atmospheric Icing on Wind Turbines", Doctoral Theses at NTNU 2011:259, 2011.
- [2] U. N. Mughal and M. S. Virk, "Atmospheric icing sensors - An insight", The Seventh International Conference on Sensor Technologies and Applications, ISBN 978-161208-296-7, August 25-31, 2013, pp. 191-199.
- [3] C. D. Kuhumonen, "Ice removing machine", U.S. Patent 4186967, 1980.
- [4] H. F. Foder, "ISO 12494 - Atmospheric icing on structures and how to use it", ISO 12494:2001, ISBN 1-880653-51-6, June 2001.
- [5] S. W. Trimble, "Encyclopedia of Water Science - 2nd Edition", CRC Press, ISBN 978-0-8493-0-9627-4, 2008.
- [6] <http://www.ufa.cas.cz/html/upperatm/chum/namraza/1cingmeasczech2.pdf>, Last Access date: September 10, 2014.
- [7] S. Fikke, "Cost 727 - Atmospheric icing on structures; measurement and data collection on icing", ISSN 1422-1381, MeteoSwiss, 2007.
- [8] U. N. Mughal, "Design and Development of Robust Atmospheric Icing Sensor", PhD Thesis submitted at the Department of Physics, University of Oslo, July 2014.
- [9] <http://migyanesfertechno.blogspot.no/2012/01/piezoelectricidad.html>, Last Access date: September 10, 2014.
- [10] http://en.wikipedia.org/wiki/Strain_gauge, Last Access date: September 10, 2014.
- [11] http://www.allaboutcircuits.com/vol_1/chpt_9/7.html, Last Access date: September 10, 2014.
- [12] http://www.bosai.go.jp/seppy/jikkentou/jikkentou_syoukai_e.html, Last Access date: September 10, 2014.
- [13] http://www.combitech.com/Documents/IceMonitor_Produnktblad_Combitech.pdf, Last Access date: September 10, 2014.
- [14] U. N. Mughal, M. S. Virk, K. Kosugi, and S. Mochizuki, "Experimental Validation of Icing Load Using Rotary Physics", Manuscript submitted in International Journal of Cold Regions Science and Technology, August 2014.
- [15] U. N. Mughal and M. S. Virk, "Performance Evaluation of MuVi-Graphene and Ice Load Monitor in icing conditions", Manuscript submitted in International Journal of Cold Regions Science and Technology, August 2014.

Using CFD-Based Virtual Sensor Data to Study the Structure of Air Flow behind A Porous Fence

Yizhong Xu and Mohamad Y. Mustafa

Atmospheric Icing Research Team

Narvik University College

Narvik, Norway

Yizhong.Xu@hin.no and MohamadYazidF.Mustafa@hin.no

Abstract— Physical experiments have difficulties to thoroughly investigate the full structure of air flow behind a porous fence. Physical measurement sensors have their limitations of data acquisitions in turbulent air flow. Computational Fluid Dynamics (CFD) technique provides an infinite number of virtual sensors that allows producing quantitative CFD-based virtual sensors data for users. In this paper, a 3D CFD model is assessed by the physical sensors data, and the simulation has provided comprehensive information for studying the structure of airflow in a 3D domain.

Keywords—physical sensor; virtual sensor; wind tunnel experiment; 3D CFD model; CFD-based virtual sensor data

I. INTRODUCTIONS

In cold regions like Norway, the outdoor environment can be extremely hostile towards human activities. Porous fence is one of common devices widely applied in these regions. It serves as windbreaks to effectively mitigate the damages caused by strong wind and transported sediments. Therefore it can create an operable and habitable space for human needs.

The structure of air flow behind a porous fence is complex due to the presence of the bleed flow passing through the pores in the fence and the displaced flow passing over the fence. Figure 1. shows a comparison of flow regimes behind porous fences as porosity above and under critical porosity, where β is porosity of fence, and β_{crit} is critical porosity of fence.

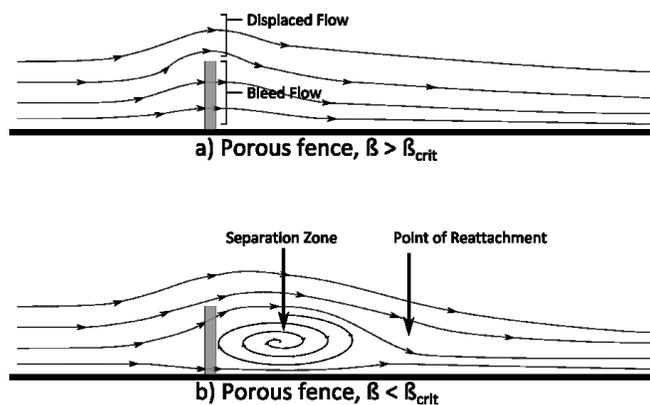


Figure 1. Comparison of flow regimes behind porous fences as porosity approaches critical porosity

Critical porosity β_{crit} is defined as the maximum fence porosity below which flow separation and reversal occurs [1]. Above the critical porosity, the airflow in the leeward is dominant by bleed flow and there is no flow separation (Figure1, a). Below the critical porosity, the leeward airflow directly behind the fence reverses, resulting in a region of recirculating air (Figure1, b). In general, fence porosity in the range of 0.20-0.50 is considered to give noticeable changes of flow structures behind fences [2] [3] [4] [5].

Physical experiments to investigate the structure of air flow behind porous fences are quite challenging, due to the presence of turbulence. Conventional cup-type anemometer is the earliest device to give a rough estimate of turbulence intensity in the field tests [6]. Hot-Wire anemometer (HWA) and Pulsed-Wire anemometer (PWA) must be positioned at specific measurement points to obtain results, that will distort the airflow field. Laser Doppler anemometer (LDA) and Phase Doppler anemometer (PDA) are non-intrusive to the measured airflow field. However, like HWA and PWA, LDA and PDA only provide time-averaged velocity and turbulence intensity values at discrete measuring points, and have difficulties to measure the near-wake regions behind fences. Particle Image Velocimetry (PIV) and Particle Tracking Velocimetry (PTV) as recently developed non-intrusive measurement techniques can obtain instantaneous velocity measurements and are related properties in a target area. The air is seeded with tracer particles, which must be sufficiently small to be assumed to faithfully follow the air dynamics. In practice, PIV and PTV are costly, and usually are not applied in field measurements. Overall, the above physical sensors have their limitations, and have difficulties to obtain a high-resolution data set within a space of airflow influenced by a porous fence.

Over the last three decades, with the rapid development of computer technology and Computational Fluid Dynamics (CFD) techniques, numerical simulation has been increasingly employed in porous fence researches. Wilson[7] introduced a momentum sink involving the fence resistance coefficient to simulate a porous fence solved by Reynolds-averaged Navier-Stokes (RANS) equation. His results demonstrated a promising prediction in terms of the flow structure around the fence. Under different fence porosities and different turbulence models, Packwood [8]

examined numerical results against wind tunnel experimental results through a 2D thin fence model in a thick boundary layer. He found that the k-ε incorporated a Preferential Dissipation Modification (PDM) model worked better. Bourdin and Wilson [10] confirmed the suitability of CFD with regard to windbreak aerodynamics, based on the comparison of the numerical data (2D and 3D model simulation) against the experimental data. Alhajraf [9] introduced a CFD model for 2D and 3D simulations of drifting particles at porous fences. His model showed good agreement with the field observations and the wind tunnel measurements.

Virtual sensor is a smart sensor, it can be used for computing estimating complex variables that otherwise should require very expensive equipment or laboratory tests [11]. Recently CFD-based virtual sensor data as alternatives to physical sensor data are increasingly adapted by researchers. Jang et al. [12] implemented CFD-based virtual sensor data in a micro-scale air quality management system. Sun and Wang [13] used CFD-based virtual sensor data to control indoor environment and space ventilation.

In this paper, a porous fence with porosity of 0.23 has been selected for the case study that ensures recirculating air occurred in the flow regime behind the porous fence. Section 2 is the theoretical framework discussed about the novelty and robustness of the CFD-based virtual sensor data. Section 3 and Section 4 are the case study carried out in physical wind tunnel and virtual wind tunnel (numerical), where the detailed procedures of physical test and numerical simulation are presented. In Section 5, the 3D CFD model has been assessed against wind tunnel experiment, and the simulation results have been demonstrated and discussed. Finally, the capability of CFD-based virtual sensor data to study the structure of airflow behind a porous fence has been concluded in Section 6.

II. THEORETICAL FRAMEWORK

Blocken [14] reviewed a perspective on the past, present, future of Computational Wind Engineering (CWE), and made a statement: “CFD offers some particular advantages compared with on-site measurements and reduced-scale wind tunnel measurements. They can provide detailed information on the relevant flow variables in the whole calculation domain, under well-controlled conditions and without similarity constraints.” CFD technique is an efficient, flexible and relatively cheap alternative to physical experiment that has been widely recognized in the porous fence research industry nowadays. Effective application of CFD is the combination of knowledge in domain physics and numeric. When adequate physical models are selected and supplied with the correct data, essentially, CFD allows for an infinite number of virtual sensors to assess the performance of a unit.

Reynolds-averaged-Navier-Stokes (RANS) equations are the most popular governing equations to describe turbulence

flow behind porous fences so far, which are mathematically expressed as follows:

$$\frac{\partial u_j}{\partial x_j} = 0 \quad (1)$$

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \frac{\mu}{\rho} \frac{\partial^2 u_i}{\partial x_i \partial x_j} - \frac{\partial}{\partial x_j} (\overline{u'_i u'_j}) + g_i \quad (2)$$

where u_j is the j component of velocity, t is the time, x_j is the j coordinate, ρ is the air density, μ is the dynamic viscosity, and g_i is the gravitational body force.

The RANS are time-averaged equations of motion for airflow that need to solve a closure problem because of the non-linear term from the convective acceleration, known as the Reynolds stress. For the porous fence research, k-Epsilon and k-Omega are the two most popular turbulence closure models used in CFD simulations [7] [8] [15] [16]. The main difference between them is that k-Epsilon model solves kinetic energy and turbulence dissipation, while the k-Omega model solves kinetic energy and turbulence frequency. Although it is well acknowledged that the selection of turbulence models is sensitive to the accuracy of numerical results, the suitability of turbulence models varies individually. It is still open to debate the issue of turbulence model selections in the research field.

The main limitation of RANS modeling is incapable to simulate the inherently transient features of the airflow field such as separation and recirculation downstream of windward edges and vortex shedding in the wake. Large-eddy simulation (LES) can explicitly resolve these large-scale features. However, LES increases computational requirements and has the difficulty in specifying appropriate time-dependent inlet and wall boundary conditions. Nevertheless, mathematical model based on the RANS equations has been used successfully for studies of the structure of airflow behind porous fences.

CFD simulation provides virtual sensor data to estimate product properties or process conditions based on mathematical models. These mathematical models use other physical sensor readings to calculate the estimations. Inlet velocity profile and boundary conditions are those of physical sensor data that will be introduced into the mathematical model as the pre-set data, which reflects to the real scenario. Consequently, CFD simulation creates a channel through which a virtual system (CFD-based virtual sensor data) has communicated with a natural system (physical sensor data and empirical knowledge) in a way that improves understandings for researchers.

Care for high quality and reliability of CFD simulations is crucial. Numerical and physical modeling errors must be assessed. Without validation against physical experiments, the robustness of CFD-based virtual sensor data is questionable.

III. CASE STUDY SETUP

The physical experiment for this case was conducted in a closed return wind tunnel at Narvik University College. A porous fence was placed at the center of the cross section of the wind tunnel with a distance of 1000mm from the leading edge of the test section (upstream). The configuration of the fence is 650mm width * 200mm height * 3mm thickness, and it is oval holed with porosity of 0.23. The CFD simulation domain was configured by the exact size of the physical domain, which makes the 3D domain with dimensions: 655mm height, 4000mm length and 1160mm its maximum width. Figure 2 shows the physical wind tunnel experiment setup. Figure 3 displays the 3D virtual wind tunnel domain.

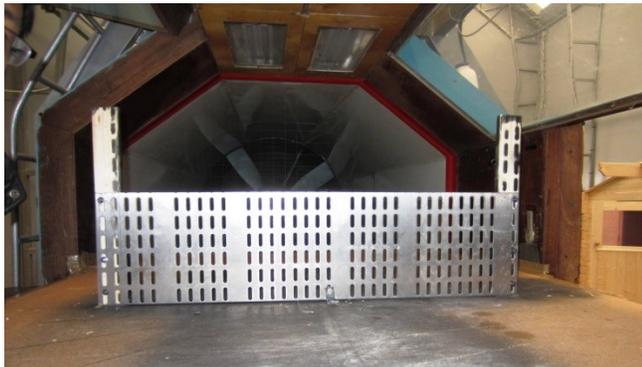


Figure 2. Physical wind tunnel experiment setup

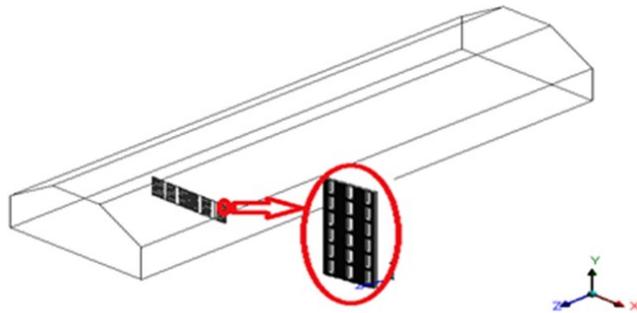


Figure 3. 3D virtual wind tunnel simulation domain

The physical wind tunnel experiment is designed to investigate the structure of airflow behind the fence under free upstream velocities of 15m/s and 20m/s respectively. A Pitot static tube was placed at the entrance of the test section to monitor the upstream velocities, and a traverse attached with a Hot-Wire Anemometer (HWA) was positioned at a longitudinal distance of 925 mm downstream of the fence. Test data were taken by moving the traverse at steps of 0.2 inch in the vertical direction. To improve the accuracy of the data, 50 readings have been taken for each step, and then time averaged data were recorded.

The physical experiment revealed the inlet velocity profiles were fully developed and obeyed the power law profile:

$$u = U_{free}(y/\delta)^\alpha \tag{3}$$

Where the U_{free} is the free stream inlet velocity which is measured 15m/s and 20m/s here respectively. δ is the boundary layer thickness which is equal to 10mm. The exponent α is 0.11.

Equation (3) was written in program C language and was interpreted into the CFD model. As such, the real sensor data have been transferred into the numerical simulations.

IV. NUMERICAL SIMULATION

The CFD simulations were performed under ANSYS 14.0 Fluent workbench package. To optimize resources the meshed domain was reduced down to half since it was symmetrical in the YZ plane and an air box (length * width * height = 3000mm * 400mm * 300mm with the upstream length of 500mm) was created to dense the elements around the fence. Figure 4 demonstrates the creation of the CFD domain and its meshed symmetry wall.

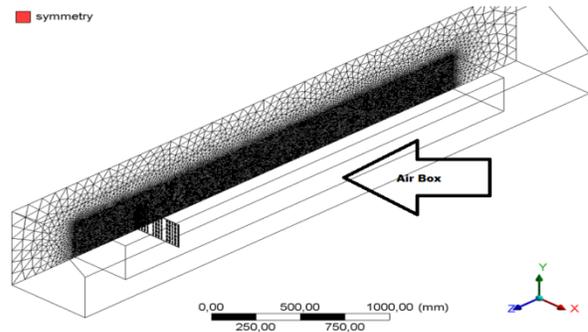


Figure 4. CFD domain & its meshed symmetry wall

In this paper, realizable k-Epsilon turbulence model with Non-Equilibrium Wall Function is employed, as it is in remarkable agreement with the considerable testing results [17].

Mesh sensitivity study was carried out under the same simulation conditions, where tetrahedral and polyhedral elements have been performed through 6 different meshing scales. The parameters selected to check the mesh independent condition of grid were velocity magnitudes and turbulent kinetic energies (TKEs). It was examined that the mesh with 6.3 million tetrahedral elements was desirable.

Turbulent intensity ratio and viscosity ratio were set at 1% and 10% respectively after the inlet velocity profiles were hooked. The gauge pressure at the pressure outlet was set at 0 Pascal with the backflow turbulent intensity ratio and viscosity ratio as 5% and 10%. All of the rest boundary conditions were treated as no-slip stationary wall with 1mm roughness height and 0.5 roughness constant. The solution method was the pressure-velocity coupling the Semi-Implicit Method for Pressure-Linked Equation (SIMPLE) scheme, since the scheme has been extensively used for atmospheric flows [16] [17].

The convergence criteria were set the scaled residuals below $1 \cdot 10^{-4}$, and mass flow rates between velocity inlet and pressure outlet have been checked afterwards.

V. RESULTS AND DISCUSSIONS

A. Assessment of 3D CFD model

The comparisons of velocity magnitudes between the numerical and experimental results is presented in Figure 5, where H/h is the ratio of the measuring height to the fence height. The acquired data were taken along a vertical line $925mm$ downstream of the fence in the symmetry wall, which is correspondent to the exact position of the experimental measurement line.

Comparisons among CFD and experimental results

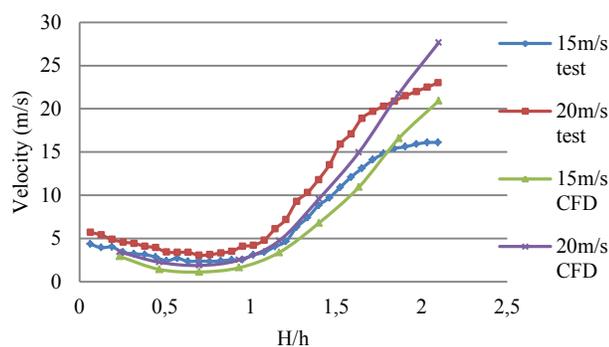


Figure 5. Comparisons between CFD and wind tunnel results

It is observed that the numerical results are in good agreement with the experimental results especially within the range of $H/h = 1.2$. Beyond the range of $H/h = 1.8$, velocity magnitude in the CFD grows faster than that in the experiment. The reason can be attributed to the fact that the blockage ratio of the wind tunnel in the current setting is 9.8%. It is slight high that increases the effects of the top wall boundary layer on the regional velocities in the physical test, while for the case of CFD, the roughness height of the top wall is set to $1mm$.

The CFD simulation over-predicted the reduction of velocity when compared to the physical measured results. In general, it is describable.

B. Domain structure of air flow

The 3D CFD simulation provides an infinite number of virtual sensor data to form a comprehensive structure of air flow in the targeted domain. It allows assessing the performance of any unit. Figure 6 and Figure 7 display the contours of velocity magnitude and kinetic energy respectively, where the free stream velocity is at $20m/s$, plane-1 is parallel to the symmetry wall with $x = 11mm$, plane-2 is parallel to the floor with $y = 0mm$, and plane-3 is parallel to the velocity inlet with $z = 1500mm$. The domain structure of air flow is agreed to the descriptions of other researchers [1] [18].

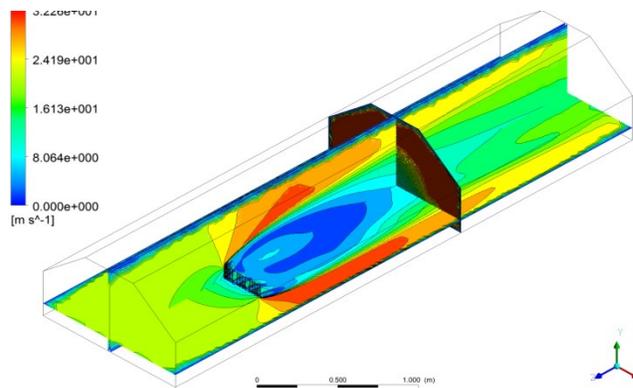


Figure 6. Velocity magnitude contours in the 3 planes

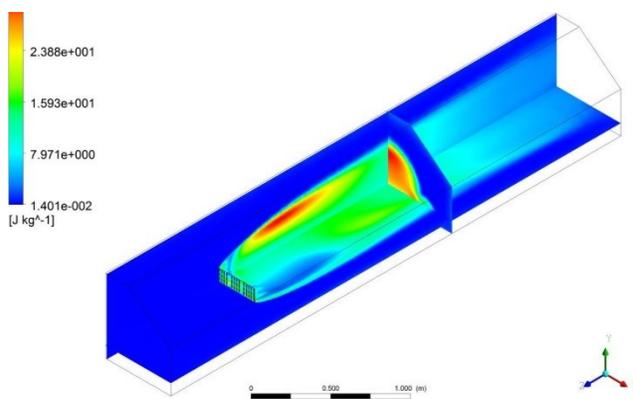


Figure 7. TKE contours in the 3 planes

C. Shear stress distributions in fence porous zone

One of advantages in 3D CFD simulation is that it allows scrutinizing shear stress and pressure distributions in porous fence zone. Unlike 2D model, 3D model can directly reflect this information in detail without modifying momentum and inertial loss within the porous zone. Figure 8 shows the shear distribution in the porous zone. These data are not possible to be obtained by the real sensors equipped in the current wind tunnel experiment.

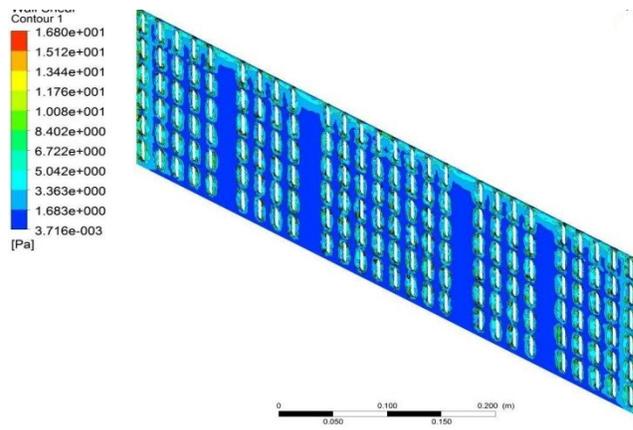


Figure 8. Contour of wall shear stress in porous zone

D. Position of the reattachment point

Reattachment point is determined by examining the horizontal velocity component at ground level to determine the point, where the horizontal velocity changes sign from upwind (negative) to downwind (positive) [1]. It is an important parameter to assess the shelter distance of a porous fence. It is unlikely to accurately capture a reattachment point under the current setup of wind tunnel experiment, since the sensor produced time averaged data that it is no possible to generate negative data.

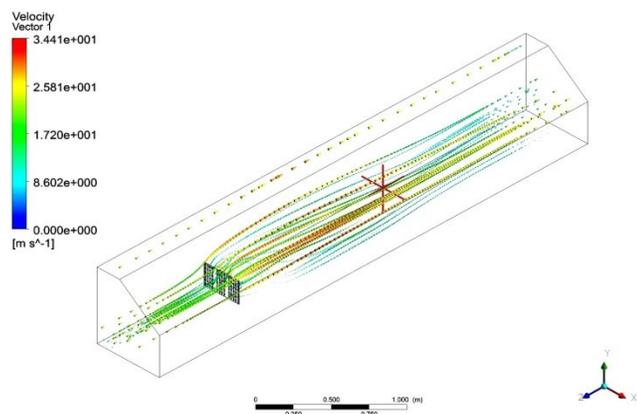


Figure 9. Reattachment point in the domain

In CFD simulation, as the time averaged bed shear stress reflects the velocity in the cell next to the boundary, then the reattachment is defined as the point where the near-wall velocity is zero. Therefore it can conveniently allocate the position of reattachment point in the domain. Figure 9 displays the red-cross is the position of the reattachment point, which is at $x, y, z = 3.1e-5, 0, -1668mm$.

E. Algorithmic outputs of CFD-based virtual sensor data

A power feature of CFD simulation is to generate algorithmic outputs of CFD-based virtual sensor data for analysis. It takes CFD generated data of dynamic head as an example, since dynamic head is a variable commonly used in the fluid dynamics research. Calculating dynamic head is based on the following formula:

$$q = \rho * (|u|)^2 / 2 \tag{4}$$

Where q is the dynamic head, ρ is the air density, and $|u|$ is the velocity magnitude.

In ANSYS Fluent, Using the Define command by opening the Custom Field Functions, the formula is easily to be written into the model. Figure 10. shows the contours of dynamic head in the symmetry wall.

CFD simulation can also generate its virtual sensor data by defining algorithm in its User Defined Functions. Outputs of data can be written to files by applying XY plot.

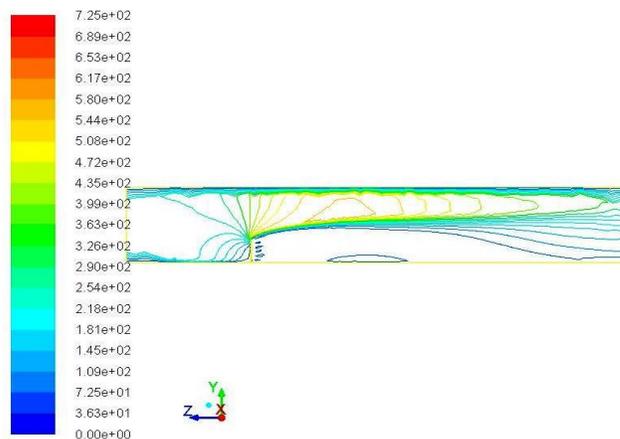


Figure 10. Contours of dynamic head

VI. CONCLUSIONS

In this paper, the CFD model has been assessed and its data of velocity magnitudes are in desirable agreement with the physical sensor data. Through an infinite number of virtual sensors, the model provides quantitative CFD-based virtual sensor data to comprehensively study the structure of air flow behind a porous fence. Comparing with physical experimental test, the CFD model has shown its strength with regard of flexibility, efficiency, relatively low cost and productivity. The model can be used in evaluating and designing porous fences.

It must be pointed out that CFD-based virtual sensor data are valid only after the model has been proved sound, which means that CFD modeling has to be examined and assessed by essential physical sensor data.

Future work on this research will apply this model for two-phase flow simulation (wind driven sediments like sands and snows).

ACKNOWLEDGEMENT

Norwegian Research Council financed this work under project number 195153 (ColdTech). The authors would also like to acknowledge the contribution of the industrial partner in this task: IKM dsc AS, Norway.

REFERENCES

- [1] Z. Dong, W. Luo, G. Qian, P. Lu and H. Wang, "A wind tunnel simulation of the mean velocity field behind upright porous wind fences," *Agricultural and Forest Meteorology*, vol. 146, pp. 82-93, 2007.
- [2] M. Jensen, Shelter effect: investigations into Aerodynamics of shelter and its effects on climate and crops, Copenhagen: Danish Tech. Press, 1954.
- [3] N. Tani, "On the wind tunnel test of the model shelter hedge," in *Bulletin of the Natinal Institute for Agricultural Sciences*, 1958.
- [4] J. Raine and D. Stevenson , "Wind protection by model fences in simulated atmospheric boundary layer," *Industrial Aerodynamics*, vol.

- 2, pp. 159-180, 1977.
- [5] S. Lee and H. Kim, "Laboratory measurements of velocity and turbulence field behind porous fences," *Wind. Eng. Ind. Aerodyn.*, vol. 80, pp. 311-326, 1999.
- [6] L. Hagen and E. Skidmore, "Turbulent velocity fluctuations and vertical flow as affected by windbreak porosity," in *Trans.ASAE*, 1971A.
- [7] J. Wilson, "Numerical studies of flow through a windbreak," *Wind Eng. Ind. Aerodyn.*, vol. 21, pp. 119-154, 1985.
- [8] A. Packwood, "Flow through porous fences in thick boundary layers: comparisons between laboratory and numerical experiments," *Wind Eng. Ind. Aerodyn.*, vol. 88, pp. 75-90, 2000.
- [9] S. Alhajraf, "Computational fluid dynamic modelling of drifting particles at porous fence," *Environmental Modelling and Software*, vol. 19, pp. 163-170, 2004.
- [10] P. Bourdin and J. Wilson, "Windbreak aerodynamics: is computational fluid dynamics reliable?," *Boundary-Layer Meteorology*, vol. 126, pp. 181-208, 2008.
- [11] E. Wilson, "Virtual sensor technology for process optimization," in *Symposium on Computers and Controls in the Metals Industry in Iron and Steel Society*, 1997.
- [12] A. Jang, A. S. Man, J.-J. Kim, H. W. Choi, K. Hong, S. B. Lim, H. Kim and J.-H. Woo, "Implementation of CFD, sensor network, and distributed data management in support of micro-scale air quality management," in *6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, Seogwipo, 2011.
- [13] Z. Sun and S. Wang, "A CFD-based test method for control of indoor environment and space ventilation," *Building and Environment*, vol. 45, pp. 1441-1447, 2010.
- [14] B. Blocken, "50 years of Computational Wind Engineering: past, present and future," *Wind Engineering and Industrial Aerodynamics*, vol. 129, pp. 69-102, 2014.
- [15] P. Richards and R. Hoxey, "Appropriate boundary conditions for computational wind engineering models using the k-e turbulence model," *Wind Eng. Ind. Aerodyn.*, vol. 46 (47), pp. 145-153, 1993.
- [16] J. Wilson and C. Mooney, "A numerical simulation of boundary-layer flows near shelterbelts comments," *Boundary-Layer Meteorology*, vol. 85, pp. 137-149, 1997.
- [17] J. Santiago, F. Martin, N. Bezdeneznykh and A. Sanz-Andres, "Experimental and numerical study of wind flow behind windbreaks," *Atmospheric Environ.*, vol. 41, pp. 6406-6420, 2007.
- [18] N. Zhang, J.-H. Kang and S.-J. Lee, "Wind tunnel observation on the effect of a porous wind fence on shelter of saltating sand particles," *Geomorphology*, vol. 120, pp. 224-232, 2010.

Improving Distance Estimation in Object Localisation with Bluetooth Low Energy

Georgia Ionescu, Carlos Martínez de la Osa and Michel Deriaz

Institute of Services Science

University of Geneva

Switzerland

Email: {*georgia.ionescu, carlos.martinez, michel.deriaz*}@unige.ch

Abstract—The arrival of Bluetooth Low Energy (BLE) creates opportunities for great innovations. One possible application is object localisation. We present our unique software that can track objects and help finding their location within a house perimeter. With the help of Bluetooth beacons that can be attached to different items, we can estimate the distance between the mobile device and the object with an accuracy of less than one meter. In this paper, we describe our system and the techniques we use, the experiments we conducted along with the results. In addition, we briefly present some work in progress using an indoor positioning system that helps locating the objects.

Keywords—object localisation; mobile application; Bluetooth Low Energy

I. INTRODUCTION

Improving the quality of life of elderly people with the help of technology is a key topic in the current research. There are numerous advantages of using technology in order to complete tasks that otherwise would be very difficult for a human. Technology can help elderly overcome different challenges. One particular problem that we try to solve is finding lost personal items. Locating objects has a wide variety of practical applications, not only for elderly people. On a daily basis people deal with losing important items (like wallets, keys, etc.). Having your phone telling you where a desired item is, or giving you information about how far the object is from you, could save a lot of time and effort for a person. Additionally, it would help people for which it is difficult to remember where they keep such items.

Currently, there exists a wide range of systems and technologies that provide real-time locating. Different technologies include WiFi [1], RFID (Radio-Frequency Identification) [2], ZigBee [2], etc. They all differ in terms of cost, infrastructure complexity, availability, maintenance costs, etc. These differences make each of them more appropriate for certain problems with different characteristics, further presented.

We describe distance estimation improvements when locating lost objects using the Bluetooth Low Energy technology. We developed a unique approach to reduce the noise and lower the error up to 1 meter for short distances and below 3 meters for long distances.

In this paper, we present firstly previous work that is relevant to our research along with some of the existing systems used for locating objects. Next, we provide a detailed description of our proposed model, and the techniques we use.

In Section IV, we describe the results of our experiments. Finally, in Section V, we discuss several directions of current and future research along with the conclusions.

II. RELATED WORK

Locating objects is not a new problem. Over the past years, there were many attempts to find a generic solution and currently there are numerous approaches and technologies that address this issue. Each of them is more suitable for certain contexts and constraints that we present below.

One of the most common approaches is tracking objects and their movements in a video sequence, using a camera. This is well suited for traffic, surveillance and robots that need to identify objects based on images [3][4][5]. The vision-based techniques are not ideal in a mobile context, due to several reasons. Firstly, the lost object might be visually inaccessible, despite the fact that it can be close to the person. Secondly, these techniques are very computationally expensive, which makes it not suitable for a mobile device.

A different category of Real-Time Location System is based on WiFi. The main components are the tag (active or passive) that is attached to the object, and the reader, that can establish a wireless communication with the tag. It has been successfully used for autonomous mobile robots [6] and for locating people in the underground (subway) [7]. However, WiFi is better suited to locate smart devices rather than objects.

Although RFID is primarily used for identifying objects, real-time location based on RFID has been studied extensively over the past years, with applications in health-care [8] and warehouse operations [9]. However, RFID is not a technology accessible to most of the people, expensive devices (RFID readers) are required.

Bluetooth Low Energy has a lot of potential for object tracking. It is a mainstream technology available on latest mobile devices. The distance between the tags and the readers can reach up to 50 meters.

Bluetooth has been successfully used for indoor positioning. L. Pei et al. [10][11] present their system that finds the location using fingerprinting. The position is calculated using the RSSI (Received Signal Strength Indication) probability distribution combined with the Weibull distribution. The accuracy obtained has a standard deviation of 10 meters. A different approach is proposed by F. Subhan et al. [12] that use trilateration for computing the position. The distance is estimated based on the

radio propagation model combined with the Gradient filter for reducing the noise. The accuracy obtained using this method is 2.67 meters.

In contrary, the technology proposed in our paper is based on BLE, as most portable devices come already equipped with it, hence we can benefit without additional cost and effort. The system setup cost is low which makes it more affordable in comparison to above mentioned solutions. In addition, the accuracy of the distance we achieve in our system is proven to be higher than in previous approaches.

There are several recent commercial products that offer the hardware and software to help finding objects [13][14], based on BLE. However, they all give only information regarding the intensity of the communication, or whether the object is in range or not.

III. SOLUTION OVERVIEW

Our goal is to provide a mobile application that helps people find their belongings, without the need of expensive infrastructure. Secondly, the set-up must be minimal and easy to use by a person without technical background. Thirdly, the technology must be accessible to everyone who owns a smart device.

Based on all these criteria, Bluetooth Low Energy was chosen. Our system uses the StickNFind [14] beacons which can be attached to the objects. Their battery lasts up to one year based on 30 minutes a day use [14].

The distance between the emitter and the receiver can be estimated using the Log Normal Shadowing model (LNS)[15] detailed in (1).

$$P_d = P_0 - 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \quad (1)$$

- P_d represents the power of the signal strength (RSSI),
- P_0 is the offset (the signal strength at the reference distance d_0),
- n is a coefficient characteristic to the device and the surrounding environment, d represents the distance between the emitter and the receiver,
- X is the noise added to each measurement.

The parameters that can be tuned are P_0 and n . These are specific to the environment and the layout of the room. Using the appropriate values of P_0 and n we can estimate the distance between the receiver and the emitter, using (1). Computing the distance is not just a simple estimation. The process consists of several steps showed in Figure 1.

A. Data Acquisition

The StickNFind beacons broadcast data one time per second when they are not paired with the smart device. Once paired, they broadcast every 100 milliseconds. The RSSI (Received Signal Strength Indication) value of the Bluetooth signal received by the mobile device is used for estimating the distance between the beacon and the mobile device.

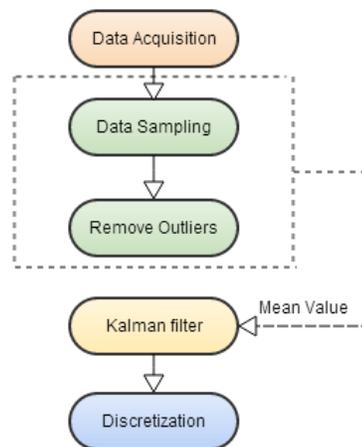


Fig. 1. Diagram

B. Data Sampling

The Bluetooth signal is not sufficiently stable to estimate the distance based solely on one measurement. In Figure 2, we show an example of different measurements of the RSSI for the same distance between the emitter and the receiver. In order to reduce the noise and attenuate the extreme values, we use a sample of ten measurements for computing the distance, instead of a single one. For every value of the distance, we use the last ten measurements to which we apply filtering techniques which we detail later.

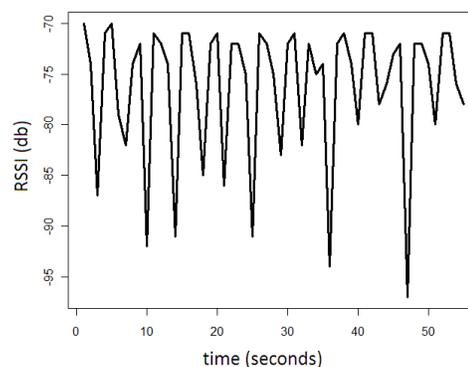


Fig. 2. RSSI variation for a fixed distance

C. Outliers Removal

The experiments (Figure 2) show that the signal is very unstable. Due to noise, spikes appear in the signal which should not be taken into consideration. We remove the outliers with the help of Chebyshev Outlier Detection based on Chebyshev’s theorem detailed in (2). We apply the inequality for $k=2$, which was proven to be a good choice according to other researchers [16]. As a consequence of (2), 75% of the data must be in the range of maximum two times standard deviation distance from the mean value. Based on this probability, we remove all the values that fall outside the domain. The values

inside the domain are used to calculate the mean value.

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (2)$$

D. Filtering

We consider the mean value calculated above as the current value of the signal strength. We use the Kalman filter [17] to better estimate the mean, which is proven to add more stability to signal.

E. Discretisation

The final step is to approximate the value of the distance with the closest value from a predefined set. From a human point of view, the little variation of the value creates more difficulty in interpreting the distance. For instance, we approximate all the values between 0.76 and 1.25 to 1, all between 1.26 and 1.75 to 1.5, etc. This gives better stability to the estimated distance displayed to the user. In the same time, the little error added does not create an impediment in the process of finding the object.

IV. RESULTS

A. Tuning the parameters

The distance can be estimated based on the RSSI value as shown in (1). The different parameters described in (1) can be tuned, in order to acquire a good accuracy of the estimation.

We conducted measurements at distances up to 25 meters. A set of measurements contains 20 values of the RSSI for every value of distance. The average of the set is calculated. The parameters are further calculated using linear regression, that minimize the error of the measurements. The parameters are used in estimating the distance between the beacons and the mobile device. The best values of these parameters were found to be $P_0 = -63.506$ and $n = 1.777$. We used this values in all our experiments.

B. Experimental setup

The experiments were conducted in two different ways. For the first set of experiments the beacon was placed at a distance of 0 meters from the mobile device, and moved with a constant speed from 0 to 18 meters away from the target. The RSSI was measured and the distance computed based on (1), and following the steps presented in Section III. The results obtained are shown in Figure 4, Figure 5 and Table I.

For the second set of values, we performed different measurements of the RSSI at distances from 0 to 25 meters. In Figure 3 we show the distance computed based on the RSSI value, the estimation using our method, along with the actual measured distance.

C. Results

We can clearly see an improvement of the distance estimation as a result of applying Chebyshev's theorem combined with the Kalman filter. Moreover, the discretisation of the distance values adds stability and reduces the noise as shown in Figure 4.

In Figure 3, we plot the estimated distance against the measured distance between the tag and the smart device. The points tend to be closer to the first diagonal (pictured in green) when the distance is smaller which shows that we obtain a smaller error for a shorter distance. This has a direct implication on the process of finding an item. The closer we are to the object, the more accurate the distance becomes, which makes it easier to find the object.

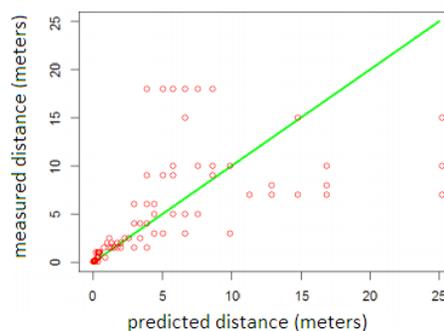


Fig. 3. Estimated distance against measured distance

In order to have a numerical quantification of the error, we defined 3 domains of accuracy for which we compute the average error in meters, and as a percentage. The first chosen range (4 meters) is based on the average room size in the typical elderly residence we collaborate with. This will give us an indication of the efficiency of our system, in the perimeter of a room. The second domain (4-10 meters) corresponds to the average overall length of the whole apartment. The last range (10-18 meters) represents the size of bigger, non typical household.

We show the results in Table I. The first two rows of the table show the error produced when we estimate the distance based on the signal measurement without any filters applied. Although the absolute value seems to increase with the distance, the percentage decreases. On the other hand, the use of our techniques produce different results, as we can see in the third and fourth row. The average error is roughly 3 times better after applying the filters, with an average error below one meter for distances up to 4 meters. Even though an error of 43% can be considered not reliable in certain circumstances, it is a significant improvement in the context of our application. While searching for a lost item, this accuracy gives a good indication on how close the object is. Additionally, the average error obtained with our approach is proven to be smaller than the previous approaches found in the literature [12].

TABLE I. Error results

Distance	0-4 meters	4-10 meters	10-18 meters
Error Measured %	182%	127%	111%
Absolute value Measured	3.61 m	8.95 m	8.95 m
Error Filtered %	43%	34%	36%
Absolute value Filtered	0.865 m	2.45 m	2.90 m

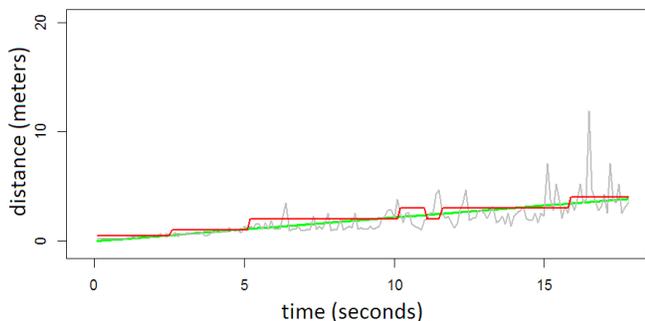


Fig. 4. Distance over time

In Figure 4 and Figure 5, we show the evolution of estimated distance for a moving receiver, as described in Section IV-B. On the x-axis we have the number of discrete values for which we measured the RSSI. The green straight line shows in both pictures the measured distance between the beacon (receiver) and the emitter (the tablet in our case). In Figure 5 the estimated distance based on the RSSI is shown in grey, while the value computed with our method is displayed in dark blue. Our method is clearly improving the distance estimation. We show the results for short distance (lower than 4 meters) in Figure 4. Similar to Figure 5, the green line represents the actual distance, and the grey one shows the estimated value based on one RSSI value. In red we show the values computed using our method. The stability is clearly improved with our method, and the extreme values affect very little the distance estimation.

It is important to mention some limitations that our system still has, which will be addressed in the future. Firstly, the effect of the battery level of the beacons is not studied yet. Additionally, the setup of our experiments didn't include walls or other obstacles that can interfere and increase the noise.

V. CONCLUSION AND FUTURE WORK

We presented our model that can successfully guide people into finding their lost items. As opposed to the currently available products which give only a non quantifiable information about how close the object is, our application can locate objects by estimating the distance between the BLE beacons with a good accuracy. We showed that Chebyshev inequality combined with the Kalman filter can be successfully applied to Bluetooth. In addition, discretising the distance adds a certain stability to the signal, which makes the application easier to use from a human-machine interaction point of view.

We are currently working on a more advanced version of the application that is able to locate the object in the perimeter

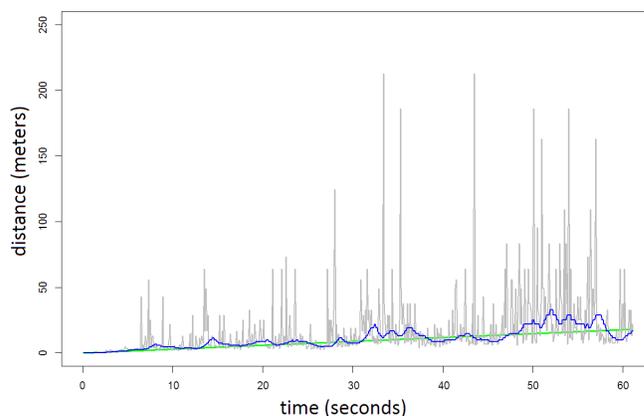


Fig. 5. Distance over time

of the home. We use an indoor positioning system to locate the user, and based on the person's movement (location at different moments) and using the different distance between the person and the object we can determine the location of the object. The location is displayed as an area and not as a point.

Lastly, we will study the impact that the battery level and the interference of obstacles have on the distance estimation.

ACKNOWLEDGMENT

The work of this paper was supported by the AAL EDLAH project (aal-2012-5-062). The authors would like to thank Anastasija Collen who helped to improve the quality of this paper.

REFERENCES

- [1] "WIFI", 2014, URL: <http://standards.ieee.org/about/get/802/802.11.html> [accessed 2014, 09, 19].
- [2] "PANs", 2014, URL: <http://standards.ieee.org/about/get/802/802.15.html> [accessed 2014, 09, 19].
- [3] M. Irani, B. Rousso, and S. Peleg, "Detecting and tracking multiple moving objects using temporal integration", *Computer Vision ECCV 1992, Lecture Notes in Computer Science*, Volume 588, 1992, pp. 282-287.
- [4] A. Gilbert and R. Bowden, "Tracking Objects Across Cameras by Incrementally Learning Inter-camera Colour Calibration and Patterns of Activity", *Computer Vision ECCV 2006, Lecture Notes in Computer Science*, Volume 3952, 2006, pp. 125-136.
- [5] F. Fayad and V. Cherfaoui, "Tracking objects using a laser scanner in driving situation based on modeling target shape", *Intelligent Vehicles Symposium*, 2007.
- [6] B. G. Amidan, T. A. Ferryman, and S. K. Cooley, "Data outlier detection using the Chebyshev theorem", *Proceedings of Aerospace Conference*, 2005.
- [7] N. Kawaguchi, et al., "Underground Positioning: Subway Information System using WiFi Location Technology", *Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, 2009.
- [8] P. Najera, J. Lopez, and R. Roman, "Real-time location and inpatient care systems based on passive RFID", *Journal of Network and Computer Applications*, Volume 34, Issue 3, May 2011, pp. 980989.
- [9] H. K. H. Chowa, K. L. Choya, W.B. Leea, and K.C. Lau, "Design of a RFID case-based resource management system for warehouse operations", *Expert Systems with Applications*, Volume 30, Issue 4, May 2006, pp. 561576.

- [10] L. Pei, et al., "Inquiry-based Bluetooth Indoor Positioning via RSSI Probability Distributions", *Second International Conference on Advances in Satellite and Space Communications*, 2010, pp. 151-156.
- [11] L. Pei, et al., "Using Inquiry-based Bluetooth RSSI Probability Distributions for Indoor Positioning", *Journal of Global Positioning System*, Volume 9, Issue 2, 2010, pp. 122-130.
- [12] F. Subhan, H. Hasbullah, A. Rozyyev, and S. T. Bakhsh, "Indoor Positioning in Bluetooth Networks using Fingerprinting and Lateration approach", *International Conference on Information Science and Applications*, April 2011, pp. 1-9.
- [13] "TrackR", 2014, URL: <http://www.thetrackr.com/> [accessed 2014, 09, 19].
- [14] "StickNFind", 2014, URL: <https://www.sticknfind.com/> [accessed 2014, 09, 19].
- [15] T. S. Rappaport, "Wireless communications principles and practices", *Proceedings of Aerospace Conference*, 2002.
- [16] B. G. Amidan, T. A. Ferryman, and S. K. Cooley, "Data outlier detection using the Chebyshev theorem", *Proceedings of Aerospace Conference*, 2005.
- [17] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transaction of the ASME Journal of Basic Engineering*, 1960, pp. 35-45.

Assessment of Sensor Technologies for Gate-Based Object Counting

Colin Leslie, Alex Vakaloudis, Kostas Anagnostopoulos, Nikolaos Chalikias, Jian Lang

Nimbus Centre Cork Institute of Technology

Cork Ireland

{colin.leslie, alex.vakaloudis, kostas.anagnostopoulos, nikolaos.chalikias, jian.liang }@cit.ie

Abstract— On the area of automating object counting, we concentrate on uniform, disposable products stored on a pile, queue or a stack (e.g., a shelf) and examine a number of different technologies for sensing input and output through a gate to the storage area. We define a set of comparison criteria with practical flavor in order to examine and evaluate twelve different types of sensors. The intention for our study is to form a baseline for anyone needing to implement gate-based input/output control.

Keywords: *sensor evaluation; gate control; evaluation of sensor technologies*

I. INTRODUCTION

Product counting is an important aspect of stock control and may take many forms depending on the material being stocked and counted [4]. Automating this process of object counting can bring many benefits by simplifying the renewal process, eliminating errors etc. [2].

In this paper, we concentrate on the context of a gate-based, disposable material placed in a casing, stored in a pile or queue in and examine the case-study of detecting the addition or removal of a product and consequently calculating reliably the number of items in stock. The question this paper deals with is assessing the different types of sensors available for automating this process. Towards this objective a set of criteria with a practical orientation is defined [5].

Under this perspective, the paper is organized as follows. Section II describes current techniques in Supply-Chain Management (SCM) and Section III records the factors for sensor assessment. Section IV discusses the different types of sensors giving a brief definition and listing their advantages and disadvantages with reference to the assessment factors. This critique is subsequently summarized and we finish with outlining a case study and listing conclusions and pointers to future work.

II. RELATED WORK

Most of the today's stock control systems use extensive use of Radio-frequency identification (RFID) technology [1] [2] due to the effective way of recording movement of objects within a networked system [3]. In these systems the stock control is achieved automatically by the information system that supports the Supply-Chain Management (SCM) and unauthorized removal of an object is detected by

electronic gates that monitor continuously the physical entry points of the area that the system is installed. To add system intelligence, traditional RFID based SCM are integrated with Wireless Sensor Networks (WSN) [4] [5]. These systems incorporate advanced sensing technologies while for the monitoring of the position of an object, we could use localization techniques.

The above practices though, are applied in large scale deployments and are not cost effective for ad hoc or small scale deployments. They require changes in the existing physical infrastructure and the deployment of sophisticated Information Systems for which the support cost is not insignificant.

SCM systems are expensive because of the nature of the sensors (RFID) that are deployed and the extensive features that are available nowadays. The cost of the SCM is high and the problem we set in this paper is the monitoring of stock in a pile or queue. The basic requirements are

1. low cost solution,
2. easily deployable, and
3. no modifications in the existing physical infrastructure.

III. SENSOR EVALUATION FRAMEWORK

As mentioned above our assessment has a practical orientation and envisages the scenario when products such as cans or bottles are stores on shelves, stacks or piles. The main, driving requirement of our assessment is to be able to count adding or removing products in a reliable way defined as

1. Reject false positives, for example the hand of the carrier is not measured as a product.
2. Spot attempts of cheating, for instance trying to extract a product from an input-only entrance.
3. Work under environments that can be cold/warm and/or wet (e.g., freezer)

Under this context, we define the assessment factors of Table 1 under which the different types of sensors are evaluated and consider two main areas namely, usability and technology. The rationale behind this partition is that in an industrial setting the decision makers are usually business people who are interested in the costs of the sensors and the installation process and technical personnel who focus on suitability of the sensor from a technical perspective. We

decided not to assign weights to these criteria in order to leave the decision making open to each case-study.

TABLE I. CONSIDERATIONS DRIVING EVALUATION

Usability Considerations	
Cost	
Installation procedure	Purchase pricing per unit of coverage, e.g., per shelf
Proven in Industry	Technical expertise and time required contributes to ease of use
Technical Considerations	
Power consumption	Influences simplicity of installation and ease of use
Form Factor	Could turn particular choices inappropriate for certain applications
Robustness	Working in diverse (e.g., wet, warm) environments
Reliability	Detecting false positive/negatives, e.g., human hand
Integration	

IV. MATERIALS AND METHODS

This section lists twelve different types of sensors we identified as suitable for gate-based object counting. Due to space constraints, it outlines the way each sensor operates, where our survey has revealed that it is mainly used and what are the main advantages and disadvantages.

A. Capacitive Sensors

Capacity sensors detect changes in their electric field to estimate proximity. Industrial Style Sensors detect objects passing under/past them. They are widely used in large Form Factors and appear in a number of different packages, suitable mostly for manufacturing lines. They are relatively cheap and can operate under harsh environments. They rely on power supply and while they can be in due course accurate, their calibration can be very tricky and lengthy process requiring software intelligence.

B. Capacitive array sensors

Capacitive Arrays are placed on the “floor” side and sense an object placed on it. A modern analogy on how they operate is how a Smart Phone screen detects touch. They are a fairly novel technique and being innovative they are not yet industry proven and therefore risky. Occupying a whole side, they could be an accurate solution and filter false positives.

However, they may not work well in potentially wet environment and would require employing a clever grounding mechanism.

C. Inductive Sensors

In a method similar to capacitive sensors, inductive sensors employ changes in their magnetic field to pinpoint proximity to detect proximity and hence measure distance of metal objects. Industrial variants have the same type of form factors as capacitive sensors and for all intents and purposes,

operate the same as Industrial Capacitive Sensors but only detect metal objects.

They are also proven in industry, are accurate, come in robust packaging and large form factors, suitable mostly for manufacturing lines. Most important could only work with products of a metal element like metal caps.

D. Ultrasonic Sensors

Ultrasonic sensors transmit a sound wave (“Ping”) of ultrasonic frequency and measure the time it takes for the wave to return in order to measure distance. They are widely used in both industrial and hobby circles and come in different form factors. A simple everyday day example of the technology is reversing sensors on modern cars.

They are accurate and cheap but potentially inaccurate since they would for example count employee hands.

E. Camera-based Sensors

A camera can be paired with a smart controller to detect patterns, shapes or colours to detect objects entering a storage area. The camera could be trained to watch for a distinctive feature of the object, for instance a branding logo, a characteristic shape or a specific colour.

With the appropriate software processing, cameras could be extremely accurate. Nevertheless they are a complex solution. Cameras are an expensive and sensitive component and they would need a more powerful MCU to process images while also being susceptible to environmental changes, such as lighting.

F. Switch Technologies

The usage of simple mechanical switches positioned in a way that they switch and count objects as they enter or exit the storage area. They are scalable; one can use just 1 on each point of entry/exit or an array of switches to reduce the chance of false switching and infer direction.

An interesting aspect is a rotary encoder to detect direction as well as actuation. Switches are reliable and simple, proven in industry and a very cheap solution. Since they require power only when activated, they consume minimum power. Although they cannot differentiate between object and detect false positives if placed in clever way they can minimize such occurrences. They are easy to install and replace.

G. Magnetic switches

These can be described as small, mechanical switches encapsulated in a small glass or plastic enclosure, activated by changes in the magnetic field. They are more reliable and space efficient to the simple mechanical switches and are used in industrial applications. They are cheap, however could be fragile especially the ones with the glass enclosure.

H. Hall-effect switches

They come in the form of integrated circuits, also activated by changes in a magnetic field. They are smaller and more sensitive than the magnetic switches and therefore can achieve better performance. They are available in analog and digital form providing higher resolution and can detect

not only presence but also distance and speed (through change). Also proven in industry, they add a degree of complexity compared to other solution since they are an electronic component with power and enclosure requirements.

I. Optical Sensors

There are various types of optical sensors, two of which are applicable in our domain of interest. The first are Interrupt Beam type sensors, where the transmitting beam and receiving sensor are placed opposite each other. When the object passes the beam the interrupted beam would cause a switch.

The other choice are reflected Beam type Sensors, where the transmitter and receiver are next to each other and the changed angle of the reflected beam is detected when an object passes through it.

Optical sensors have no mechanical parts, so virtually there is no wear. They have many industrial variants that are well proven and can be easily simplified to a light source and photodiode. They may have very small form factors. Their performance though, may be erratic not be able to detect different opacities or false positives.

J. Strain gauge

Load cells are a practical application of Strain Gauges. Ultimately, they measure weight; they consist of an array of piezoelectric materials that generate a voltage proportional to the stress applied. There exist three types worth considering:

- a. Bending Beam
- b. Shear Beam
- c. Pancake Cell

This solution would be the most accurate representation of how many objects are in a storage area if these objects have uniform weight. Since it counts objects individually, it would have the least number of false positives. They are in cheap price if bending or shear beam but they need to be customized for different products.

K. Barcode

A barcode is an optical machine-readable representation of data relating to the object to which it is attached. Originally barcodes systematically represented data by varying the widths and spacing of parallel lines, and may be referred to as linear or one-dimensional (1D) [6]. In order to use barcode sensors extensive rework must happen to the mechanical part of the system in order the items to leave from a particular exit and with a certain orientation so that the sensor would be able to identify them correctly. The implementation of the sensors is cost effective since low-end microcontrollers have the necessary processing power to cope with the algorithm of the barcode decoding [7].

L. QR code

A QR code [8] consists of black modules (square dots) arranged in a square grid on a white background, which can be read by an imaging device (such as a camera) and processed using Reed–Solomon error correction until the image can be appropriately interpreted; data is then extracted

from patterns present in both horizontal and vertical components of the image [7]. The need for a camera in order to capture the image makes the technique expensive, while the algorithms in order to decode the QR code are quite sophisticated [9].

V. CASE-STUDY AND DISCUSSION

Our survey outlines twelve different types of sensors. Table 2 presents in an organised format our findings which generated the following outcomes:

1. Although some are relative to each other (e.g., QR and Barcode), the number of different available approaches makes evident that even in quite specialised situations, an optimum solution could be achieved exactly because of the sheer quantity of solutions.
2. Cost does not appear to be a forbidding factor for any of the technologies, switch-based approaches seem to be the most suitable when cost and simplicity are the main requirements.
3. Sensors are available in three forms, namely, mechanical, magnetic mechanical and half-effect providing different levels of sensitivity.
4. As Internet of Things (IoT) grows in importance, choosing the right technology of sensor to link to the communication module of an IoT device will also be pivotal.
5. Requirements, such as power consumption and durability will play a major role as well as compatibility among modules and the existence of a communication protocol or a feasibility to create one.

This research is currently investigating the applicability of these solutions using as a case-study shelves in which objects embedded in a cylinder shaped boxes are stored. The setup is located in the lab setting and is used to organize and store any material used by the Nimbus centre researchers, ranging from chips to small mechanical and wooden tools. Placing them inside a cylinder box ensures uniformity and eases stock control. Moreover it is a shape that is compatible with switches, capacitive arrays, QR code and optical sensors.

Different shelves with the twelve sensors are used to measure their reliability, endurance, easiness of installation and sensitivity. The results up to now indicate that the mechanical switches suit better. They perform better in the sense of minimum cases of false positives and false negatives especially when they are used in numbers to verify each other's triggering. The capacitive arrays also provide accurate measurements however their installation process is longer and need to be refitted for every new shelf. The optical sensors require the usage or development of extra software to filter out hand movements of the hand placing objects in our out of the shelf. Finally, the QR and barcode choices need special care from the user to scan the code and from the system to give feedback (e.g., sound).

This case-study, while installed in a lab environment can be classified as pre-production because of the volume of use and the number of users. However, we acknowledge that additional levels of testing are needed to extract more results and also test different setups.

VI. CONCLUSIONS AND FUTURE WORK

This paper examined the applicability of twelve different types of sensors in facilitating gate input/output control of simple, every day products. As requirements and their gravity vary on each application, there is no single preferred solution. It is remarkable however how technologies of diverse orientation could be applied to solve a single problem.

Current and future research work involves not only the further specification of the assessment factors but the application of techniques, such as fuzzy logic with sets associating factors and sensors and different fuzzy rules to apply in order to get a qualitative or quantitative marking. This will enable to objectify the suitability of every sensor technology for specific application needs. We also intend to extend this study and perform quantitative analysis with specific application domains.

REFERENCES

[1] T. Engel, S. Lunow, J. Fischer, F. Kobler, S. Goswami, H. Krcmar. "Value Creation in Pharmaceutical Supply Chains using Customer-Centric RFID Applications", Proceedings of 2012 European Conference on Smart Objects, Systems and Technologies, 2012. pp 1-8.

[2] Y. Miaji, M.A.A Mohamed, and N. B. Daud, "RFID Based Improving Supply Chain Traceability", Proceeding of the 2013 IEEE International Conference on RFID Technologies and Application, 2013, pp. 1-6.

[3] Y. Wu, D. Ranasinghe, Q. Sheng, S. Zeadally, J. Yu, "RFID enabled traceability networks: a survey", Distributed and Parallel Databases October 2011 vol. 29 pp. 397-443.

[4] L. Evers, P. Havinga, J. Kuper, and M.E.M. Lijding, "SensorScheme: Supply Chain Management Automation using Wireless Sensor Networks", IEEE Conference on Emerging Technologies and Factory Automation, 2007 pp. 448-455.

[5] S. Vellingiri, A. Ray and M. Kande, "Wireless Infrastructure for Oil and Gas Inventory Management", 39th IEEE Annual Conference of Industrial Electronics IECON 2013 pp. 5461-5466.

[6] <http://en.wikipedia.org/wiki/Barcode>, June 2014

[7] X. Wang, J. Liu, H. Yi, "Design of Multifunctional Barcode System Based on MEGA128L", Automation and Systems Engineering (CASE) 2011 International Conference on Control pp. 1-4.

[8] http://en.wikipedia.org/wiki/QR_code, June 2014

[9] Y. Liu, J. Yang, M. Liu, "Recognition of QR Code with Mobile Phones" Chinese Control and Decision Conference 2008 pp. 203-206.

Table II. SYNOPSIS OF SENSORS AND THEIR FEATURES

Sensor	Costing	Installation	Industry	Power	Form	Robustness	Reliability
Capacitive Sensor	Cheap	medium complexity	Bottling and Manufacturing plants	High	Small to Medium	High	High if no unpredicted obstacles
Capacitive Array	Cheap	complex / difficult	None	High	Small to Large	Medium	Unproven
Inductive Sensor	Cheap	medium complexity	Manufacturing Bottling	High	Small to Medium	High	High
Ultrasonic Sensor	cheap to medium	medium	Diverse	Medium to High	Small to Large	Low to High	High if no unpredicted obstacles
Optical Interrupt Beam	medium to expensive	medium	People Counting Bottling and Manufacturing plants	Medium to High	Medium	High	High if no unpredicted obstacles
Optical Reflected Beam	medium to expensive	medium	Bottling and Manufacturing plants	Medium to High	Medium	High	High if no unpredicted obstacles
QR	Expensive	Medium	Diverse	High	Medium	Low to High	High
Optical Camera	expensive	difficult	None	High	Big	Low to High	High
Mechanical Switch	Cheap	easy	Diverse, Variety of equipment and applications	No	Small	Low to High	High if no unpredicted obstacles
Magnetic Switch	Cheap	easy	Diverse, Variety of equipment and applications	No	Small	Low to Medium	High
Hall-effect Switch	Cheap	easy	Diverse, Variety of equipment and applications	No	Small	Medium to High	High
Weight Load Cell	cheap to expensive	difficult	Weighing scales, Mini Bar fridges,	Low	Small to Large	Low to High	High
Barcode	cheap to expensive	medium	Diverse	High	Small to Medium	Low to High	High

Wide Area Surveillance Using Limited-Flying-Time Helicopters

Kenichi Mase

Academic Assembly

Niigata University

Niigata-Shi, Japan

e-mail: mase@ie.niigata-u.ac.jp

Abstract—Sensor networks can be useful for disaster recovery. It is essential how to realize time-efficient and pervasive surveillance over a wide disaster-affected area. A surveillance architecture based on collaboration among multiple electric helicopters is presented. Area division principles are used considering cost minimization as well as considering the requirements of allowable surveillance time, and transmission range and effective bandwidth of the wireless link. A simple model and methodology for area division considering the battery-capacity-limited flying range are presented. Numerical examples show the feasibility of the area division approach.

Keywords—disaster; surveillance; ad-hoc network; wireless; electric vehicle.

I. INTRODUCTION

Information and Communication Technology (ICT) can contribute to the creation of disaster-resilient societies in two ways. One way is to increase the resilience and tolerance of communication infrastructure toward disasters. Second way is to develop new ICT technologies for promoting disaster recovery activities. In this study, we focus on use of ICT for surveillance of wide areas such as disaster-affected areas.

A disaster, such as big earthquake, hurricane, etc. can affect quite a large area, e.g., as large as 100 km square. It would take enormous time to perform close, detail, and pervasive surveillance to cover the entire area and deliver the obtained surveillance results in real-time to the remote disaster recovery headquarter. It is thus essential how to realize time-efficient and pervasive surveillance over a wide disaster-affected area.

Surveillance can be performed on the ground or from the air. The latter provides a wider field of vision than the former, and is, therefore, especially efficient when the disaster-affected area is wide. For realizing aerial surveillance, a single aerial vehicle such as air plane or helicopter can be used to cover the entire disaster-affected area. However, this solution may not be acceptable since it would take much time to complete one-round surveillance for a large area due to speed limit of the aerial vehicle. Furthermore, it is also a problem how to deliver the obtained surveillance results to the disaster recovery headquarter in real-time and continuously. An alternative approach using a number of very small Unmanned Aerial Vehicles (UAVs) with fixed or rotary wings is attractive from the viewpoints of the investment required and operational ease. Indeed, this approach can be applied in various applications including surveillance and forms an active research area [1]-[16].

Multiple UAVs working in parallel may be required for monitoring a wide area. An approach called “three-Dimensional Mobile Surveillance (3DMS),” was proposed in [15], where multiple Electric Vehicles (EVs) and very small, lightweight unmanned UAVs with rotary wings (helicopters), termed “Electric Helicopters (EHs),” were cooperatively engaged in surveillance over a wide area.

The contributions of this paper are as follows:

- 1) A time-efficient and pervasive surveillance architecture based on the collaboration of multiple EHs, each with its partner EV, is presented.
- 2) Area division principles are given based on cost minimization, as well as considering requirements of allowable surveillance time and transmission performance.
- 3) A simple model and methodology for area division devised by considering the battery-capacity-limited flying range are presented.

The remainder of this paper is organized as follows:

Section II summarizes related works. Section III discusses surveillance strategies for designing wide-area surveillance architectures. Section IV outlines the basic approach and surveillance architecture of the proposed system. Section V presents a surveillance area division scheme, and Section VI presents numerical examples. Finally, the conclusions of this paper are presented in Section VII.

II. RELATED WORK

There exist many studies on Vehicular Ad hoc NETWORKS (VANETs) [17]. In these studies, vehicles with engines consuming fuels such as oil and gas were assumed implicitly, and applications useful during driving were extensively explored and developed. The benefits of using EVs in a disaster-affected area for providing emergency communication networks have been discussed [18].

UAVs can be used for improving the connectivity and performance of ad hoc networks on the ground [8]-[10]. It is easier to control UAV position in air given the lack of obstacles or physical boundaries. Therefore, UAV relay positioning methods have been studied extensively [11]-[14].

When multiple UAVs work in tandem for surveillance in a disaster-affected area, inter-UAV coordination for task assignment and responsibility is essential to improve surveillance efficiency. Coordination among helicopter UAVs for forest-fire surveillance was explored; here, the focus was on observing the fire’s expanding perimeter and

devising a parallel, non-overlapping, uniformly time-consuming patrolling task assignment method [6]. In disaster surveillance, all areas should be monitored continuously, and repeatedly. This problem has not been fully discussed in the previous works.

III. SURVEILLANCE STRATEGY

A. Ground-based vs. Air-based

Ground-based surveillance is indispensable for close, detailed, and pervasive surveillance, as well as the subsequent rescue activities. Rescue team members should be equipped with means of transportation such as vehicles that enable them to quickly move around the wide area within limited time. In addition to surveillance using ground vehicles, surveillance can be performed from the air. It is easier to view the entire disaster-affected area from the air. Aerial vehicles can monitor locations that ground vehicles cannot access owing to a lack of roads or the presence of obstacles.

B. Manned vs. Unmanned

Aerial vehicles can be either manned or unmanned. The former type is expensive, and the availability of human pilots is limited. Focusing on UAVs, the need for a human pilot on the ground should be avoided. In principle, the UAV can fly and work through autonomous piloting or remote automatic control by computer without human manual operation.

C. Fuel vs. Electricity

Both ground vehicles and UAVs can be powered by fuel oil or electricity. The use of EVs and electric UAVs is attractive in the disaster recovery phase because their batteries can be recharged using local power generation facilities such as solar panels, even under long blackouts, whereas fuel-based vehicles and UAVs cannot be refueled under a shortage of oil stock in a disaster-affected area. Focusing on UAVs, the spare battery can be immediately substituted for the spent battery, while it takes some time to refuel the oil tank.

D. Supply of EV

EVs can reduce the air pollution due to automobile exhaust gases. The EV market has recently experienced significant growth.

Moreover, a very small EV with one or two seats, called a mini-EV, can potentially grow EV market in the near future, achieving significantly greater penetration in the community, especially in aging societies [18]. In such an environment, in the event of a large-scale disaster, it would be easy to divert EVs available in the community toward disaster recovery activities, thus reducing the reserves required for disaster recovery. Therefore, the supply of EVs for disaster recovery is expected economically feasible.

E. UAV options

UAVs include airships, helicopters, and airplanes.

Airships allow for relatively long flying and hovering times over the disaster-affected area, which are desirable for continuously monitoring the same area in disaster surveillance. However, support services such as the supplementing of helium gas are required. Helicopters can hover as well, but require a greater amount of energy and have limited continuous operation times. Airplanes find it difficult to both hover as well as to adapt to the route according to the situation. In addition, similar to helicopters, their continuous operation times are limited. An airship requires a relatively large space for periodic topping-up of its helium gas, takeoff, and landing, and an airplane needs a long runway or special care for takeoff and landing. These support services are unnecessary for a helicopter.

IV. SURVEILLANCE ARCHITECTURE

A. Overview

In the proposed architecture, an EV–EH pair is the key component for the surveillance of disaster-affected areas. Each EV functions as the carrier of its partner EH during the round trip to the designated destination and surveillance, providing its roof area for accommodation, takeoff, and landing of the EH. It also may support automatic piloting for its partner EH. Each EV and EH are equipped with sensing and wireless communication devices. For conducting time-efficient and pervasive surveillance activities over a wide disaster-affected area, a number of EV–EH pairs may be used for simultaneous and parallel surveillance over said area. To start surveillance, each EV accompanied by its partner EH is driven to and parked at its target point in the disaster-affected area. During this phase, the EH may get and provide information useful for EV driving, e.g., existence of obstacles and traffic congestion on streets, to the EV driver.

A designated EV within the area is assigned the role of the data collection node (simply center EV, hereafter). The sensed data are delivered from each source EV or EH by means of wireless multi-hop communication relayed by a number of EVs or EHs on the way to the center EV, and then to the remote disaster recovery headquarter, using, for instance, a satellite communication link. Disaster-affected area surveillance by EVs on the ground and that by EHs in the air can work in a complementary, cooperative manner, thus significantly improving the efficiency and accuracy of surveillance.

B. Area division requirements

Area division is useful for coordinating and assigning surveillance activities among multiple EV–EH pairs, where the entire surveillance area is divided into multiple non-overlapping sub-areas filling the entire area and an EV–EH pair is assigned to each sub-area. Surveillance activities are independently and simultaneously performed in parallel within each sub-area. In general, the greater the number of EV–EH pairs, the less is the time required for one round of surveillance of the entire area; the required time can be reduced further, although at the cost of EV–EH pair resources.

C. Wireless multi-hop data delivery

Upon reaching the target point in the assigned sub-area, each EV remains at this initial position during the following surveillance activities as the support station, while its partner EH flies over the entire assigned sub-area for surveillance and sends the gathered surveillance data to its partner EV, which works as the temporal data storage point. The EV then forwards the data to the neighbor EH flying over the adjacent sub-area, which is located on the designated path toward the center EV. The neighbor EH receives this data and forwards it to its partner EV, in addition to the data gathered by itself. This process is repeated and the data obtained by the source EH are eventually delivered to the center EV. In addition to EHs, EVs may gather and transmit surveillance data. Area division needs to be performed to ensure that an EH and its partner EV, and an EV and its upstream EH in the adjacent sub-area on the designated path toward the center EV are within the transmission range (Requirement 1). Each EV and EH may have multiple down-stream EHs and EVs, respectively. Simultaneous packet transmission to the same EH and EV may result in interference, and, thus, the use of non-overlapping channels is desirable for avoiding interference.

D. Line-of-sight transmission using flying EH

It is noteworthy that data packets can be transmitted between the EV and the EH in the proposed architecture, thereby avoiding EH-EH or EV-EV transmission. In the EH-EH, packet transmission could be unstable due to variations in the EH position or attitude in air even in the hovering state. In the EV-EV, line-of-sight may not be assured due to the obstacles on the ground, thus resulting in poor packet transmission performance. In contrast, it is easier to assure line-of-sight in the EH-EV packet transmission. Moreover, a directional antenna can be attached to the side of the EV, which can significantly extend the packet transmission range [19].

E. Flight time and range requirement

To minimize EH total weight, the weight of the battery is strictly limited. This means that the EH's battery capacity is also limited, allowing only relatively short continuous flight times and ranges. The EV's battery can be used for recharging the partner EH's spent battery, thus allowing the EH to engage in airborne surveillance repeatedly [15]. This functionality is termed as "on-EV charging." A carrier EV can be equipped with spare EH batteries. By substituting a spare EH battery for the landed EH's spent battery, the waiting time for on-EV charging can be shortened as mentioned in Section III.C. The spent battery from the EH is recharged on the EV for reuse. EH flight time must be set at least longer than the round trip time between the positions of its partner EV as the on-EV charging station and the farthest part within the sub-area (Requirement 2).

V. SURVEILLANCE SCHEME

A surveillance area is a square of size $L \times L$, which is uniformly divided into $n \times n$ square sub-areas, each of size $L/n \times L/n$, as shown in Figure 1. An EV-EH pair is

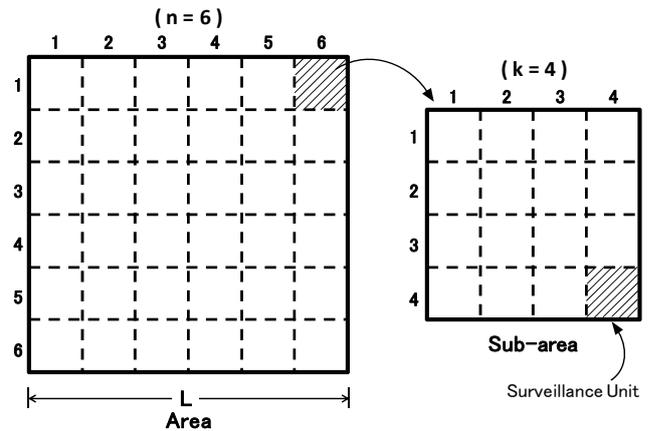


Figure 1. Example of area division in square area modeling.

assigned to each sub-area. Each EV is positioned in the center of the respective assigned sub-area. The center EV is positioned in the center sub-area. A sub-area is further divided into uniform squares, each of which, termed "a surveillance unit," is the field of vision, with the length of each side, d' , closest to and no greater than the given value, d , to make uniform division. Each surveillance unit can be surveyed by an EH flying over the center line crossing the center of the surveillance unit without approaching the edges of the surveillance unit. Let F_{\max} be the maximum flight distance divided by the side length of the surveillance unit (normalized flight distance) of an EH. Requirement 2 in Section IV.E can then be represented by

$$\begin{aligned} F_{\max} &\geq 2k && (k : \text{even}) \\ &\geq 2(k-1), && (k : \text{odd}) \end{aligned} \quad (1)$$

where $k = L/n/d$.

Because the EH needs to return to its partner EV before running out of battery for on-EV charging to continue further surveillance, the normalized flight distance, D , required for completing one-round surveillance of a sub-area is represented as the following function of k and F_{\max} :

$$D = f(k, F_{\max}). \quad (2)$$

The upper bound of D , D_{ub} , is then given by substituting F_{\max} with its lower bound in (1)

$$\begin{aligned} D_{ub} &= f(k, 2k) && (k : \text{even}) \\ &= f(k, 2(k-1)). && (k : \text{odd}) \end{aligned} \quad (3)$$

In contrast, if there is no limit on the battery capacity, the EH can continue surveillance by taking an optimum route over the sub-area without returning to its partner EV. The lower bound of D , D_{lb} , is thus obviously given by

$$\begin{aligned} D_{lb} &= k^2 & (k : \text{even}) \\ &= k^2 + 2. & (k : \text{odd}) \end{aligned} \quad (4)$$

The time required for completing one-round surveillance of a sub-area (one-round surveillance time), T is given as

$$T = D\tau, \quad (5)$$

where τ is time required to cross the side length of the surveillance unit in flight. It is noteworthy that the time of on-EV charging is not included in (5) based on the remark in Section IV.E. The surveillance time, T , should be designed to be no greater than its upper bound T_{\max} , which may be given as the design target parameter in surveillance system development, as shown below.

$$T \leq T_{\max}. \quad (6)$$

Let β (MB/km²) be surveillance data size per unit area. The total surveillance data size of the surveillance area is then βL^2 . As described in IV. C, a designated path toward the center EV should be set for data transfer from each sub-area to the center area. Because surveillance data from each sub-area are merged at each relay EV and EH, the transmission load increases for the upstream EVs and EHs closer to the center EV. Four most-upstream EHs are adjacent to the center EV in the square area modeling considered in this section. Assuming that the set of designated paths for the area is designed to balance the transmission load among the links, and neglecting the transmission overhead, the maximum link load, B (Mbps), can be represented as

$$\begin{aligned} B &= \frac{2\beta L^2}{T} & (n : \text{even}) \\ &= \frac{2\beta L^2(n^2 - 1)}{Tn^2}. & (n : \text{odd}) \end{aligned} \quad (7)$$

An example of the designated path configuration is given in Figure 2.

The transmission range of a wireless link can be defined as the maximum distance over which the packet loss rate is no greater than the given threshold. In general, the transmission range becomes shorter as the effective bandwidth increases; thus, the transmission range is a function of B , represented by $R(B)$. The transmission distance can be the longest when the EV in the center of each sub-area transmits packets to the upstream EH at the far-end corner of the adjacent sub-area on the designated path to the center EV. For satisfying Requirement 1 in Section IV.C, the following relationship should be satisfied.

$$\frac{\sqrt{10}L}{2n} \leq R(B). \quad (8)$$

In contrast, the number of EV-EH pairs covering the surveillance area should be maintained as low as possible for

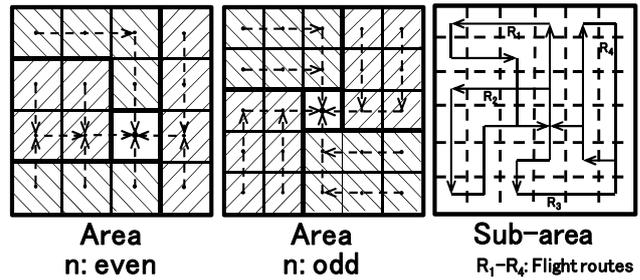


Figure 2. Example of designated path configuration to center EV.

Figure 3. Example of flight path configuration.

cost saving.

An area division solution for the square area and sub-area models should thus obtain the minimum value of n for satisfying conditions (6) and (8), where B is substituted by (7).

VI. NUMERICAL EXAMPLES

Investigation of the characteristics of the function in (2) is an interesting research issue, but is out of the scope of this study. Instead, the upper bounds of the normalized flight distance, D_{ub} , in (3) can be given numerically using the flight path configuration obtained by a heuristic approach (See Figure 3). The results, including the D_{lb} of (4), are shown in Figure 4. The ratio of D_{ub} to D_{lb} is also shown in Figure 4. It is expected that the ratio roughly converges to around 1.2.

The upper bounds of the one-round surveillance time are obtained by substituting D with D_{ub} in (5) and shown in Figure 5, where L and EH velocity are 50 km and 60 km/h, respectively. As expected, the one-round surveillance time decreases with an increase in the number of divisions or the field of vision. Similarly, the transmission range requirement given in (8) decreases with an increase in the number of divisions, as shown in Figure 5.

Assuming use of an adequate transmission system meeting required effective bandwidth and transmission range for each EH-EV or EV-EH link, link quality such as packet delivery ratio can be assured, while end-to-end path quality degrades as the number of divisions of sub-area increases. To improve end-to-end path quality, packet retransmission mechanism either in link or end-to-end may be required.

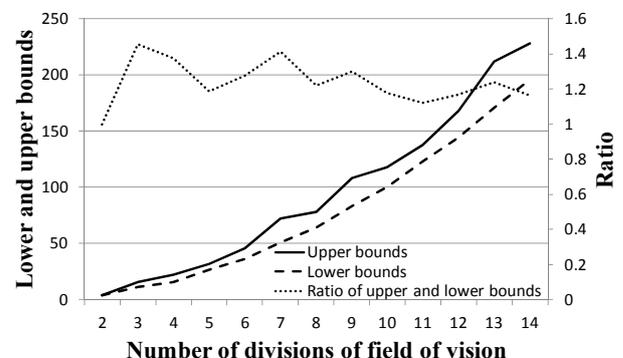


Figure 4. Lower and upper bounds of normalized flight distance, D and ratio of upper bounds to lower bounds.

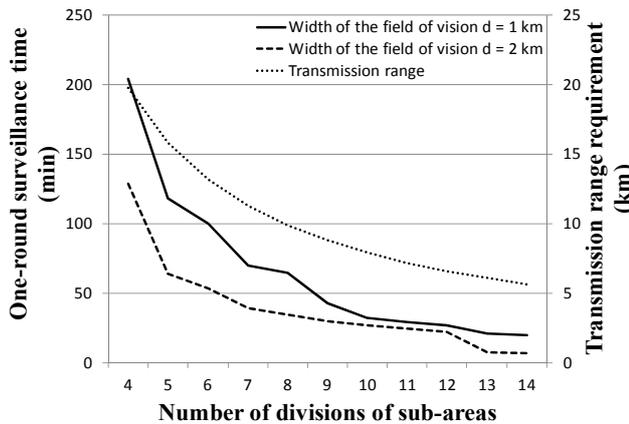


Figure 5. Upper bounds of one-round surveillance time and transmission range requirement with regard to number of divisions, n .

The number of divisions, n , meeting the given one-round surveillance time requirements is obtained from (3), (5) and (6) and shown in Figure 6. The higher the one-round surveillance time requirement, the lower is the number of divisions of a sub-area. The corresponding maximum link load in (7) and transmission range requirements in (8) are shown in Figure 7. As the one-round surveillance time requirements rise, the surveillance data transmission frequency decreases, resulting in a decrease in the maximum link load. By contrast, the transmission range requirements increase owing to an increase in the sub-area size. Therefore, given the one-round surveillance time requirements, an adequate transmission device should be selected based on the requirements of decrease in the maximum link load and increase in the transmission range.

In summary, the numerical examples in Figure 6 show that wide area as large as 100 km square can be monitored pervasively in one hour or so, operating a few hundred EV–EH pairs. Transmission devices meeting the requirements as shown in the numerical examples in Figure 7 can be in the scope of the existing technologies.

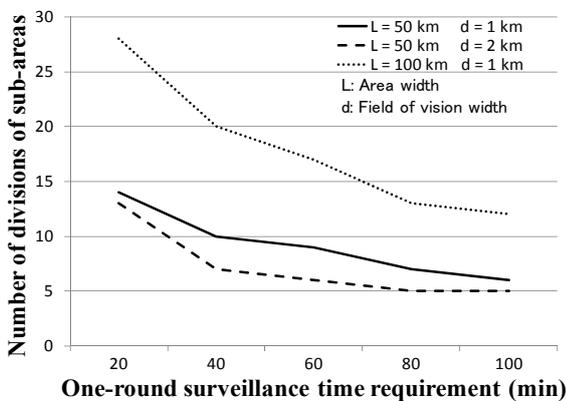


Figure 6. Number of divisions, n , meeting given one-round surveillance time requirement.

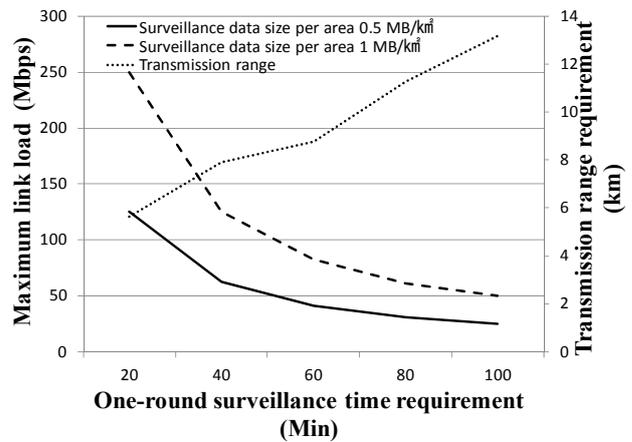


Figure 7. Maximum link load and transmission range requirement meeting given one-round surveillance time requirement.

VII. CONCLUSION AND FUTURE WORK

A time-efficient and pervasive surveillance architecture based on the collaboration of multiple EHs was presented. A wide area is divided into a number of sub-areas. An EH and its partner EV, which supports automatic piloting and battery charging for the EH, are assigned to each divided sub-area. A wireless link is established between an EH and its partner EV, as well as between an EV and its immediate upstream EH for forming a wireless multi-hop path toward the data collection node in the area. Area division principles were used considering cost minimization as well as considering the requirements of allowable surveillance time, and transmission range and effective bandwidth of the wireless link. A simple model and approach for area division based on explicit flight route distances of an EH considering the battery-capacity-limited flying range were presented. Numerical examples showed the feasibility of the area division approach.

Further studies on this subject include the development of a general algorithm for obtaining the explicit flight route of EHs for not only simple square area models but also for geographical areas of any shape under flight distance limitations. Communications aspects, such as performance evaluation and improvement of wireless multi-hop communication path composed of multiple EH–EV and EV–EH links, are also included in the further study issues.

ACKNOWLEDGMENT

This work was supported in part by the JSPS Grants-in-Aid for Scientific Research (KAKENHI) under Grant 24246068.

REFERENCES

[1] D. Kingston, R. Beard, T. McLain, M. Larsen, and W. Ren, "Autonomous Vehicle Technologies for Small Fixed Wing UAVs," In AIAA 2nd Unmanned Unlimited Systems, Technologies, and Operations-Aerospace, Land, and Sea

- Conference and Workshop & Exhibit, San Diego, CA. Paper no. AIAA-2003-6559, pp. 1-10, 2003.
- [2] R. Beard, D. Kingston, M. Quigley, D. Snyder, R. Christiansen, W. Johnson, T. McLain, and M. A. Goodrich, "Autonomous Vehicle Technologies for Small Fixed-Wing UAVs," *AIAA Journal of Aerospace Computing, Information, and Communication*, vol. 2, pp. 92-108, 2005.
- [3] D. W. Casbeer, D. B. Kingston, R. W. Beard, T. W. McLain, S. Li, and R. Mehra, "Cooperative Forest Fire Surveillance Using a Team of Small Unmanned Air Vehicles," *International Journal of Systems Science*, vol. 37, no. 6, pp. 351-360, 2005.
- [4] G. M. Saggiani and B. Teodorani, "Rotary wing UAV potential applications: An Analytical Study through a Matrix Method," *Aircraft Engineering and Aerospace Technology*, vol. 76, no. 1, pp. 6-14, 2004.
- [5] T. J. Koo, D. H. Shim, O. Shakernia, B. Sinopoli, Y. Ma, F. Hoffman, and S. Sastry, "Hierarchical Hybrid System Design on Berkeley UAV," *International Aerial Robotics Competition*, 1998.
- [6] K. Alexis, G. Nikolakopoulos, A. Tzes, and L. Dritsas, "Coordination of Helicopter UAVs for Aerial Forest-Fire Surveillance," *Applications of Intelligent Control to Engineering Systems*, pp. 169-193, 2009.
- [7] F. N. Webber and R. E. Hiromoto, "Assessing the Communication Issues Involved in Implementing High-Level Behaviors in Unmanned Aerial Vehicles," *IEEE Military Communication Conference (MILCOM)*, pp. 1-7, 2006.
- [8] I. Rubin, A. Behzad, H. Ju, R. Zhang, X. Huang, Y. Liu, and R. Khalaf, "Ad hoc wireless networks with mobile backbones," *Proc. 15th IEEE Int'l Conf. on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 566-573, 2004.
- [9] D. L. Gu, H. Ly, X. Hong, M. Gerla, G. Pei, and Y. Lee, "C-ICAMA, a centralized intelligent channel assigned multiple access for multilayer ad-hoc wireless networks with UAVs," *Proc. 2000 IEEE Wireless Communications and Networking Conference*, vol. 2, pp. 879-884, 2000.
- [10] Z. Han, A. L. Swindlehurst, and K. J. R. Liu, "Smart Deployment/Movement of Unmanned Air Vehicle to Improve Connectivity in MANET," In *Proceedings of WCNC 2006*, vol. 1, pp. 252-257, 2006.
- [11] P. Zhan, K. Yu, and A. L. Swindlehurst, "Wireless Relay Communications with Unmanned Aerial Vehicles: Performance and Optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 2068-2085, 2011.
- [12] S. Gil, M. Schwager, B. Julian, and D. Rus, "Optimizing Communication in Air-Ground Robot Networks Using Decentralized Control," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1964-1971, 2010.
- [13] O. Burdakov, P. Doherty, K. Holmberg, J. Kvarnstrom, and P-M. Olsson, "Positioning Unmanned Aerial Vehicles as Communication Relays for Surveillance Tasks," In *Proceedings of the 5th Robotics: Science and Systems Conference (RSS)*, 2009.
- [14] C. Dixon, E. W. Frew, "Optimizing Cascaded Chains of Unmanned Aircraft Acting as Communication Relays," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 883-898, 2012.
- [15] K. Mase and T. Saito, "Electric-vehicle-based Ad Hoc Networking in Cooperation with Electric Helicopters and Surveillance for Disaster Recovery," *The Ninth International Conference on Networking and Services*, pp. 87-94, 2013.
- [16] K. Mase, "Wide-Area Disaster Surveillance Using Electric Vehicles and Helicopters," *The 24th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 3466-3471, 2013.
- [17] H. Hartenstein and K. P. Laberteaux, "A Tutorial Survey on Vehicular Ad Hoc Networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164-171, 2008.
- [18] K. Mase, "Information and Communication Technology and Electric Vehicles—Paving the Way towards a Smart Community," *IEICE Trans. Commun.*, vol. E95-B, no. 6, pp. 1902-1910, 2012.
- [19] H. Okada, H. Oka, and K. Mase, "Network Construction Management for Emergency Communication System SKYMESH in Large Scale Disaster," *IEEE International Workshop on Management of Emerging Networks and Services*, pp. 875-880, 2012.

BuckshotDV - A Robust Routing Protocol for Wireless Sensor Networks with Unstable Network Topologies and Unidirectional Links

Reinhardt Karnapke, Jörg Nolte
 Distributed Systems/Operating Systems Group
 Brandenburg University of Technology Cottbus - Senftenberg
 Cottbus, Germany
 Email: {Karnapke, Jon}@informatik.tu-cottbus.de

Abstract—Experiments have shown that the number of asymmetric and unidirectional links often exceeds the number of bidirectional ones, especially in the transitional area of the communication range of wireless sensor nodes. Still, most of today's routing protocols ignore their existence or try to remove their implications. Also, links are not stable over time, and routes become unusable often, resulting in a need for new routing protocols that can handle highly dynamic links and use unidirectional links to their advantage. In this paper, we present BuckshotDV, a routing protocol which is resilient against link fluctuations and uses the longer reach of unidirectional links to increase its performance. Furthermore, its distance vector nature makes it scalable for large sensor networks.

Keywords—wireless sensor networks; routing; unidirectional links

I. INTRODUCTION

In recent years, asymmetric and unidirectional links have been shown to be common in wireless sensor networks. Depending on the used hardware and the distance between nodes, different regions (transitional region [1], grey area [2]) have been defined, in which unidirectional links are common and can even represent the majority of links. Also, most links are not stable over time [3].

In traditional routing protocols, unidirectional links and unstable links are ignored and not used for forwarding purposes. Bidirectional, stable links make routing decisions much easier. Unfortunately, this approach neglects a lot of potential optimizations, as unidirectional links often have a greater reach than bidirectional ones. Thus, unidirectional links reduce the number of hops needed to deliver a message to its destination. However, using unidirectional links is often considered to induce too much overhead [4]. An example for this overhead is the need to inform upstream nodes of their outgoing links.

In this paper, we present BuckshotDV, a routing protocol specifically designed to use unidirectional links implicitly. The overhead which results from the need to inform upstream neighbors of their outgoing unidirectional links in other protocols is eliminated. BuckshotDV is based on a multi path approach, enabling the usage of unidirectional links and making it resilient against link changes and node failures. Moreover, a node implicitly updates its routing table each time a message is received.

The remainder of this paper is structured as follows: the nature of unidirectional links and their commonness in wireless sensor networks are presented in Section II. Selected state of the art routing protocols that were used in the evaluation are presented in Section III, followed by the description

of our protocol BuckshotDV in Section IV. In Section V the evaluation of BuckshotDV and selected state of the art protocols in simulations and real experiments is shown before concluding remarks are given in Section VI.

II. UNIDIRECTIONAL LINKS IN WIRELESS SENSOR NETWORKS

Different classifications of link quality are used in literature. Examples are included in [1][2][3][5], which all use different classifications (see below).

The most commonly used classifications divide links into bidirectional links, asymmetric links and unidirectional links. A bidirectional link is always defined as a link between two nodes which can be used to transmit a message from either of those two nodes to the other one. In contrast, the terms asymmetric link and unidirectional link are not always defined clearly, and sometimes used synonymously. Common definitions for asymmetric links focus on a variation of either Received Signal Strength Indication (RSSI) values or packet loss (delivery ratio). When the delivery ratio is used, unidirectional links can be seen as a subclass of asymmetric links where the delivery ratio in one direction is 0. However, this definition requires quite a lot of message transmissions in order to evaluate the delivery ratio. For this paper, a unidirectional link is defined as follows: a link from node A to node B is unidirectional, if node B can receive messages from A, but not vice versa.

Woo et. al. focus on link quality estimation in [1]. They measured link quality for a sensor network deployment consisting of 50 Mica Motes from Berkeley. All nodes within a distance of about 10 feet (about 3 meters) or less from the sender received more than 90% of the transmitted packets (called the *effective region*). It is followed by the *transitional region* which reaches roughly from 10 feet to 40 feet (between 3 and 13 meters) distance. Nodes in this region cannot be uniformly characterized as some of them have a high reception rate while others received no packets at all. The last region is the *clear region* and contains only nodes that did not receive any transmissions.

Zhao and Govindan measured the properties of wireless sensor networks on the physical and medium access control layers [2]. These measurements were conducted using up to 60 Mica motes, which were placed in three different environments: an office building, a parking lot and a habitat. The experiments for the physical layer were realized with a single sender and multiple receiver nodes, and have shown the existence of a *grey area* in reception which can consist of up

to one third of the network (similar to the *transitional region* described above). Another result described by the authors is that in the parking lot and indoor environments nearly 10% of measured links were unidirectional (called asymmetric links in the paper).

The Medium Access Control (MAC) layer evaluation used a simple Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) protocol, which is the default implementation for TinyOS. It was augmented with a retransmission scheme, to make use of the link-layer acknowledgments that were being transmitted anyway. The authors have defined the packet loss difference for two nodes as the difference between the packet delivery efficiency of both nodes. Unidirectional links are quite common: more than 10% of the surveyed links have a difference of more than 50%.

Ortiz and Culler studied the feasibility of using multiple channels in wireless sensor networks [5]. They evaluated link quality in three different testbeds: a machine room, a computer room and an office building, using up to 60 sensor nodes. During the experiments, each node transmitted 100 messages and each other node recorded the number of received messages, enabling easy calculation of the packet reception rate.

The authors found that unidirectional links were indeed common in their testbeds. In the machine room 32 - 36% of links were unidirectional, 18 - 34% in the computer room and 10 - 46% in the office building.

In previous work [3], we described connectivity measurements conducted using eZ430-Chronos sensor nodes from Texas Instruments. We evaluated different placements (desk, lawn, stones), different heights (ground or poles) and two radio channels. Connectivity graphs were gathered every minute, for 60 minutes in each experiment. The results show that unidirectional links were extremely common in those experiments, there were always more unidirectional than bidirectional links. Also, the increased communication range that resulted from the higher placement on the poles led to a stronger increase of unidirectional links than of bidirectional ones. On average, we measured about four to five times more unidirectional than bidirectional links. Furthermore, we found that all links were extremely unstable, with lots of link changes between measurements (minutes).

All these experiments show that unidirectional links are normal in wireless sensor networks and should be taken into account when routing decisions are made. Using them can increase connectivity, which may prevent network separation and increase performance.

III. RELATED WORK

AODVBR [6] is an enhancement of Ad-Hoc On Demand Distance Vector Routing (AODV) [7][8], that uses a mesh structure to supply multiple paths. The main achievement of the protocol is to build multiple routes without sending additional control messages. This is possible because of the broadcast character of the medium. Every node that overhears a route reply packet and is not the addressed next hop discards this packet in AODV. In AODVBR, these nodes enter the node from which the route reply was received as next hop to the destination into their routing cache. This way, a structure similar to a fish bone is constructed.

When a link breaks, the node that detected the break (re-)broadcasts the data packet with a flag indicating that this message should be sent using an alternative route. A neighboring node that receives this message and has overheard the route reply that created this route forwards the message to the next hop. This way, a detour of one hop is taken, which may enable the delivery of the data packet. Also, a route error packet is transmitted to the source, so that a new and possibly better route can be established. However, the message still has to traverse all nodes that are on the original route.

Dynamic Source Routing [9][10][11] is one of the first routing protocols that took unidirectional links into account. The authors specify two different modes of operation for DSR: one for the usage of only bidirectional links, and another which should be used when unidirectional links are common (used here). In this version, route request messages (RREQ) are flooded in the usual way. Route reply messages (RREP) however, are not sent back the inverted path of the RREQ message. Instead, the destination (D) inserts the path the RREQ has taken into a RREP message, which is also flooded. Once this message has arrived at the originator of the RREQ message (the source, S), S inserts the path taken by the RREQ into its routing table and transmits an additional routing message to node D, which contains the path taken by the RREP. Once the destination has received this message, the routes from S to D and from D to S, which can differ strongly, have been established.

Virtual coordinates are used by ABVCap_Uni [12] to enable the usage of geographic routing in networks without location information. ABVCap_Uni uses clusters and rings to enable the usage of unidirectional links. The overhead of maintaining clusters and rings is high, though. When links change often, the performance of ABVCap_Uni decreases drastically.

In previous work, we introduced Buckshot Routing [13], a source routing protocol for dense ad-hoc networks. It uses a multi path approach to circumvent broken links, unidirectional links or dead nodes. These multiple paths are implemented by a limited directional flooding: when a node receives a message, the forwarding decision differs from that used in traditional source routing protocols. Normally, a node that receives a message only checks if it is the intended next hop. In Buckshot Routing, only the one after that is important, the next-but-one hop. All nodes that have this next-but-one hop in their neighbor table forward the message.

IV. BUCKSHOTDV

Buckshot Routing and BuckshotDV are both based on a limited directional flooding. When a node S wants to transmit a message to a node D and a path is already known, messages are not only sent along this path, but also within a certain tunnel around the original route.

An example of the forwarding mechanism is depicted in Figure 1. The original path from node S to node D is a straight line in the middle of the figure. Where in traditional routing protocols a node only forwards the message if it is the intended next hop, nodes forward it if they have the hop after the next in their neighbor table in Buckshot Routing and BuckshotDV. This results in a higher message load, but also adds redundancy to the forwarding mechanism.

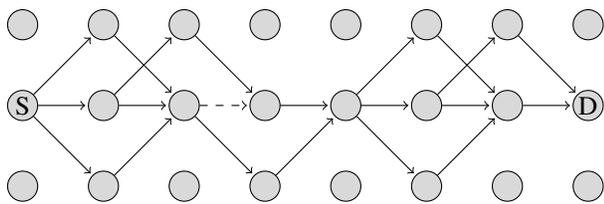


Figure 1. Multiple paths taken by a message in Buckshot Routing and BuckshotDV

The usefulness of the created redundancy can also be seen in Figure 1. The dashed link between the second and the third node on the path is now broken, which would usually result in a delivery failure. In Buckshot and BuckshotDV, this broken link is implicitly circumvented, removing the need for a new route discovery.

Buckshot Routing and BuckshotDV are based on the same forwarding mechanism. However, while Buckshot Routing works quite well in networks with a small diameter, wireless sensor networks are assumed to consist of thousands of nodes in the future. The source routing character of Buckshot Routing means that the size of messages grows with the route length, which can become a problem in state of the art wireless sensor networks where the upper bounds for message size can be quite low (e.g., 64 Byte on the eZ430-Chronos from Texas Instruments [14]).

To make the forwarding principle of our Buckshot Routing usable in large scale networks, we developed its distance vector version called BuckshotDV, which reduces the message size while at the same time increasing the robustness of the routing protocol and increasing the delivery ratio.

In traditional distance vector routing algorithms like DSDV [15] or AODV [7], each node maintains a routing table, with entries consisting at least of the ID of the destination, the distance, and the next hop. Using the same entries in Buckshot Routing with Distance Vectors (BuckshotDV) is simply not possible. As described in [13], Buckshot Routing needs to know the next-but-one hop, which means that this value has to be determined and kept in the routing table, too.

In BuckshotDV, a node enters its own ID along with the ID of the node from which it received a RREQ message into the RREQ before retransmitting it. A node that receives a RREQ message now knows its neighbor's neighbor, and thus the next-but-one hop on the reversed path, which it enters into its routing table in the form (source of RREQ, next-but-one hop, distance).

Figure 2(a) shows an example of a RREQ message in BuckshotDV. The first value in the RREQ is the type of message, followed by the sequence number of the originating node and its identity, which are used for duplicate suppression and to build the reversed route. The destination ID is of course necessary to terminate the route discovery once the destination has been reached. All of these values are fixed throughout the lifetime of a RREQ message.

The first value being subject to change is the hop count which is incremented by one on each hop. Please note that of course any other weight function, e.g., energy, would also be possible. The hop count is followed by the identities of the previous and the current hop.

When a node receives a RREQ and determines that it is the destination of this packet, it creates a routing entry for the source of the RREQ message and transmits a route reply (RREP). RREP messages also contain the ID of the node from which the RREP was received and the identity of the next-but-one hop in BuckshotDV (Figure 2(b)). The next-but-one hop is needed to find the route to the source of the RREQ message, the identity of the previous node is needed to build the backward route. Thus, contrary to Buckshot Routing in its source routing variant, RREP messages are also used to build new routes. Nodes that receive a RREP message check their neighbor table for the next-but-one hop listed there, which is the next hop from their perspective. If and only if there is an entry, they look up the next-but-one hop from their perspective in their routing table, adjust the values in the RREP message and retransmit.

Please note that the forwarding mechanism of BuckshotDV results in a spreading of messages across neighboring nodes (nodes adjacent to the initial path). This redundancy is what makes BuckshotDV resilient against link breaks and node failures. It is also the reason, why unidirectional links are implicitly circumvented.

Once the RREP message has arrived at the source and the routing table entry has been created, the DATA packet can be transmitted. As no new route needs to be learned from a data packet, the identities of the previous and current hop are omitted in DATA packets, resulting in a smaller header.

The data packet format used in BuckshotDV is shown in Figure 2(c). Just like when forwarding a RREP message, each node that receives a DATA message checks its neighbor table for the next-but-one hop listed in the message and replaces it with its own next-but-one hop for the listed destination if and only if it has found the neighbor in its neighbor table.

When compared to pure Buckshot Routing, BuckshotDV is complicated and requires more computation and copying on each node. Still, when comparing it to protocols like AODV, it remains simple. Its main advantage compared to Buckshot Routing is its scalability. In Buckshot Routing, as in all source routing protocols, the message headers grow with increasing network diameter. In BuckshotDV the header size is constant for each type of packet, making it usable in large networks.

V. EVALUATION

The evaluation includes simulations using the OMNeT++ framework [16], as well as outdoor experiments on 36 eZ430-Chronos sensor nodes from Texas Instruments [14]. These feature an MSP430 micro controller with an integrated CC1101 sub-gigahertz (868MHz) communication module [17].

Figure 3 shows the used eZ430-Chronos sensor nodes in three different placements which were used in the experiments. An external battery pack has been soldered to the nodes, which replaces the internal coin cells. This enables the usage of freshly charged batteries for each protocol.

Apart from the modification for the batteries, the sensor nodes were used as they were delivered, no calibration was made. The transmission power was also left at the preset level of 0 dBm, which lead to a small transmission range. This small transmission range is also due to the absence of a real antenna on the eZ430-Chronos: the metal surrounding the display acts as antenna.

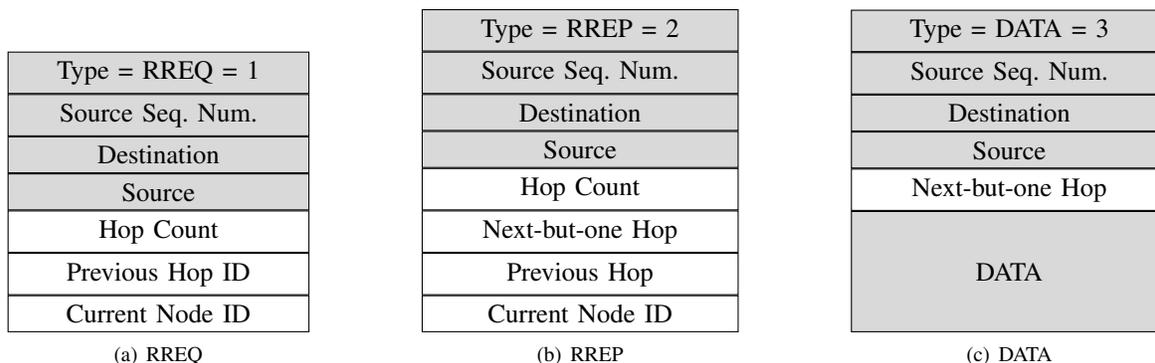


Figure 2. Packet formats in BuckshotDV

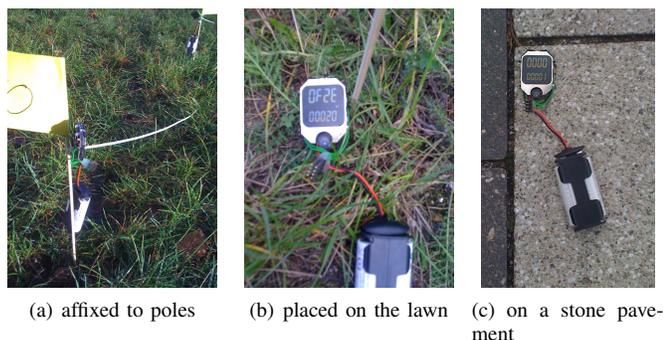


Figure 3. A modified eZ430-Chronos sensor node from Texas Instruments

Five different routing protocols were chosen as competitors for BuckshotDV in the evaluation: Flooding, Tree Routing, AODVBR, DSR and Buckshot Routing.

A. Simulation

The simulated networks consisted of four different sizes of grids: 100 nodes (10x10), 400 nodes (20x20), 900 nodes (30x30) and 1600 nodes (40x40). A grid alignment was chosen to represent applications that need area coverage, where each node is equipped with sensors that have a range of one distance unit. To simulate a certain connectivity between nodes, we used the matrix-based simulation approach presented in [18]. As the largest networks, consisting of 1600 nodes, needed to be simulated for the longest time, they also needed the highest number of connectivity matrices: for a single simulation 17761 connectivity matrices were needed. In each of these matrices, a (directed) link from node A to node B exists with a probability of α/d^6 where d is the distance between node A and node B. The inverse link, from node B to node A, exists with the same probability. Therefore, the link is bidirectional with a probability of $(\alpha/d^6) \times (\alpha/d^6)$, unidirectional (in any one direction) with $\alpha/d^6 \times (1 - (\alpha/d^6))$ and non existing with $(1 - (\alpha/d^6))^2$. The quotient (d^6) reflects the dampening induced by the distance between nodes while α represents the probability that a link between geographically adjacent nodes exists. Nodes that are directly above, below, right or left of a node are called direct neighbors and their distance was defined as 1. α was varied between 0.9, 0.95 and 1, and for each value of α ten sets of matrices with different seeds for the

random number generator were generated, leading to 30 sets of matrices per network size, and a total of 996120 connectivity matrices containing between 10.000 and 2.560.000 entries. Please note that due to the fact that connectivity matrices were generated randomly, there is no guarantee that there always was a path from sender to destination. Therefore, no upper limit can be calculated, but Flooding is used as reference protocol: the number of application messages delivered by Flooding is taken as 100% and the delivery ratio of all other protocols calculated accordingly.

The delivery ratio of Buckshot Routing, BuckshotDV, DSR, AODVBR and Tree Routing is shown in Figure 4(a). For a small network containing only 100 nodes, the delivery ratio of Buckshot Routing in its source routing version and that of BuckshotDV are still close to each other. However, when the number of nodes and thus the network diameter and route length increase, the performance of Buckshot Routing declines while that of BuckshotDV improves.

Indeed, the performance of all protocols declines, except for BuckshotDV. This is due to the forwarding mechanism of BuckshotDV, which always uses the next-but-one hop from the perspective of the node which is currently handling a message. As this next-but-one hop might be different for different nodes, the limited directional flooding gets broader with increasing network diameter, increasing redundancy and robustness against unidirectional links and link breaks.

However, this increase in robustness comes at a price: the increased redundancy means that a higher number of messages is transmitted. Figure 4(b) shows the number of messages transmitted by each protocol. While Flooding naturally transmitted the most messages, BuckshotDV nonetheless transmitted about twice as many messages as Buckshot Routing. The least number of transmitted messages can be seen on Tree Routing as expected: when a link breaks, two retransmissions are tried before the message is discarded, keeping the cost of delivery failure low. In the case of DSR, a failure to deliver a message to the next hop results in a route error message being transmitted to the originator of the message, and a subsequent new route discovery, which includes two floodings of the whole network. AODVBR should in theory be robust against message losses due to the fish bone structure it uses to reclaim lost data messages. However, this reclaiming mechanism is only used for data messages, meaning that AODVBR needs a completely bidirectional path during route discovery.

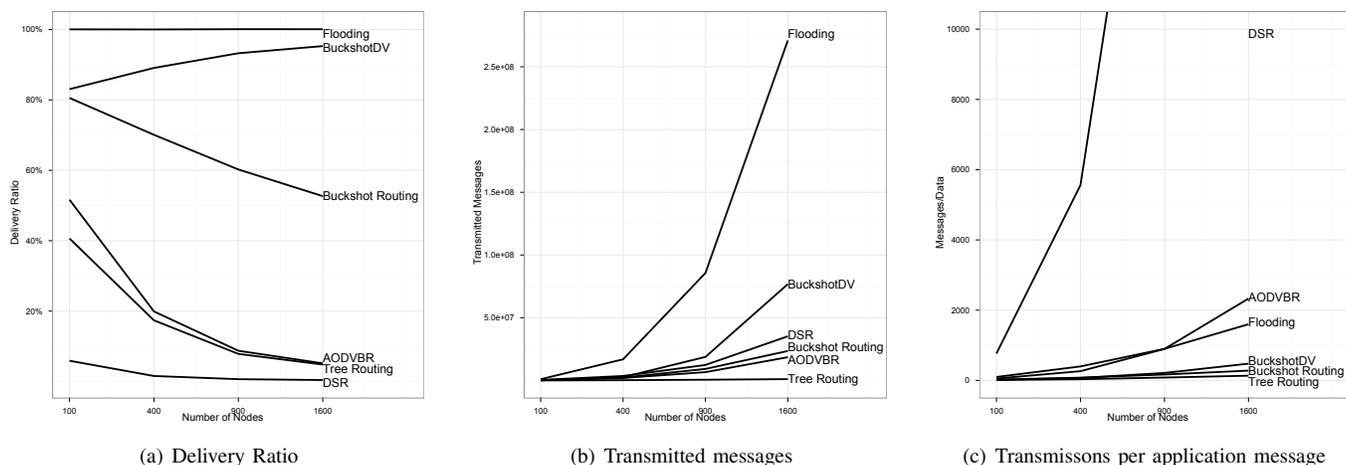


Figure 4. Performance of all evaluated protocols achieved in the simulations

The cost of delivering a single application message to the destination measured in transmitted messages is shown in Figure 4(c). With a delivery ratio of 40% and a low number of overall transmissions, Tree Routing can be a good choice for small networks if network load is more important than delivery ratio. DSR represents the other end of the spectrum - the low number of delivered application messages compared to the fairly high number of transmitted messages results in a very bad cost ratio. Therefore, DSR should not be used in such dynamic environments. The ratio of AODVBR is also worse than that of Flooding, therefore it should not be used for the resource constrained sensor networks. The ratios of Buckshot and BuckshotDV are close to each other, with BuckshotDV a little worse. The decision which of these two should be used in a certain scenario depends on the importance of data and network load: if the delivery ratio is more important, BuckshotDV should be chosen, and the increased network load tolerated. If network load needs to be reduced, Buckshot Routing should be used.

B. Experiments

In the experiments, four different placement were used: a desk, a lawn, poles, and stones. The desk placement is a one hop environment with all 36 nodes lying directly next to each other. In the other experiments, nodes were placed one meter from each other, on the grass of a lawn, on a stone pavement or affixed to poles at a height of 20 cm above ground. Each placement has different radio characteristics. In the experiments the delivery ratio was defined as the number of received application messages divided by the number of application messages handed to the routing protocol.

The delivery ratio of each protocol, divided by different placements, is shown in Figure 5. The figure shows that all protocols work well on the desk, and fairly well in the pole experiments. But when the sensor nodes are placed on the ground, AODVBR and DSR show a steep decline in delivery ratio. Tree Routing works much better, but still not as good as Buckshot Routing or BuckshotDV. Even Flooding shows a strong decline, which is due to problems with the MAC layer. However, BuckshotDV outperforms all protocols chosen for comparison.

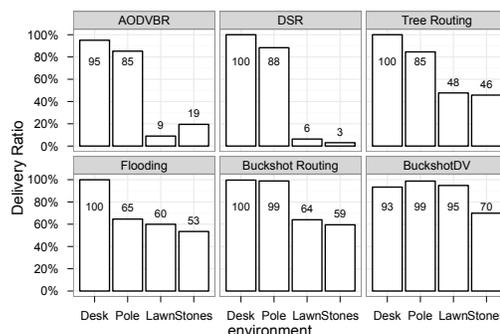


Figure 5. Delivery ratio of each protocol achieved in the experiments

The total number of messages transmitted by each protocol is shown in Figure 6. It can be seen that Flooding transmits the most messages in all placements, with Buckshot Routing and DSR following for the placements on the ground (Lawn, Stones). Tree Routing transmitted the lowest number of messages in the placements on the ground. However, the number of transmitted messages needs to be correlated to the delivery ratio.

The cost of delivering a single application message to its destination measured in transmitted messages is shown in Figure 7. DSR performs worst due to the high number of transmissions and low number of delivered application messages. However, BuckshotDV performs at least as well as all other protocols, often outperforming its competitors. Even TreeRouting which has a better cost function in the stones placement has a lower delivery ratio for that same scenario. In the desk and pole placements Buckshot Routing has the same ratio of 1 as BuckshotDV and in the lawn placement Tree Routing and BuckshotDV share a value of 7. When the delivery ratio is taken into consideration, this means that BuckshotDV always outperforms its competitors, even for our relatively small testbed consisting only of 36 nodes.

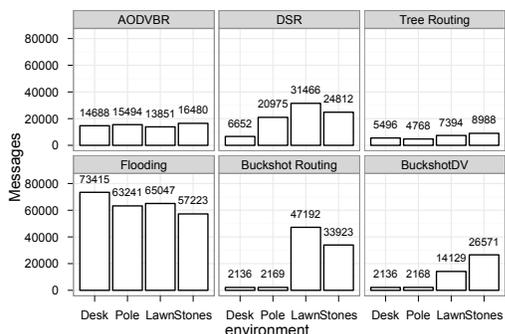


Figure 6. Number of messages transmitted by each protocol in the experiments

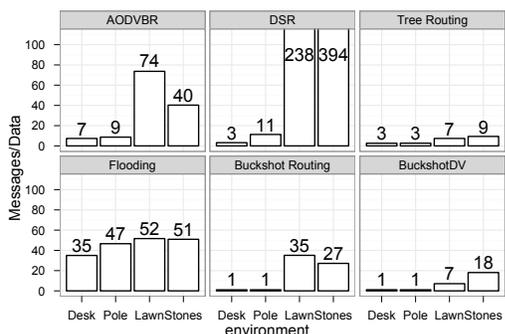


Figure 7. Number of messages transmitted to deliver a single application message in the experiments

VI. CONCLUSION

In this paper, we presented BuckshotDV, a distance vector routing protocol for wireless sensor networks which uses unidirectional links implicitly. We evaluated its performance and compared the results to those achieved by AODVBR, DSR, Tree Routing, Flooding, and our original source routing version of Buckshot Routing. The experiments that we conducted with 36 sensor nodes from Texas Instruments show the feasibility of our approach, while the simulations of up to 1600 nodes were used to evaluate the scalability. Simulation results indicate that while Buckshot Routing can only be used for sensor networks with a moderate diameter, BuckshotDV can indeed be used in large scale networks. However, we did not possess enough hardware to prove this indication in large scale experiments.

The evaluation shows that BuckshotDV can operate in sensor networks with unidirectional links, and use them to increase its delivery ratio without introducing additional overhead. BuckshotDV does not need to inform upstream nodes of unidirectional links. Rather, those links are used implicitly. The implicit usage of multiple links makes BuckshotDV resilient against link changes and node failures, and removes the need for explicit route maintenance. Routing tables are implicitly updated with each received message, introducing no communication overhead and only negligible computation overhead on the nodes. The fact that the route maintenance overhead is marginal in BuckshotDV shows its usability for networks with frequent topology changes.

REFERENCES

- [1] A. Woo, T. Tong, and D. Culler, "Taming the underlying challenges of reliable multihop routing in sensor networks," in *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2003, pp. 14–27.
- [2] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2003, pp. 1–13.
- [3] S. Lohs, R. Karnapke, and J. Nolte, "Link stability in a wireless sensor network - an experimental study," in *3rd International Conference on Sensor Systems and Software*, 2012, pp. 146 – 161.
- [4] M. K. Marina and S. R. Das, "Routing performance in the presence of unidirectional links in multihop wireless networks," in *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing, ser. MobiHoc '02*. New York, NY, USA: ACM, 2002, pp. 12–23.
- [5] J. Ortiz and D. Culler, "Multichannel reliability assessment in real world wsns," in *IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*. New York, NY, USA: ACM, 2010, pp. 162–173.
- [6] S.-J. Lee and M. Gerla, "AODV-BR: Backup routing in ad hoc networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2000)*, Chicago, IL, September 2000, pp. 1311–1316.
- [7] C. E. Perkins and E. M. Royer, "ad hoc on-demand distance vector routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, New Orleans, LA, FEB 1999, pp. 90–100.
- [8] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (aodv) routing," <http://www.ietf.org/rfc/rfc3561.txt>, 2003.
- [9] D. Johnson, D. Maltz, and J. Broch, *DSR The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks*. Addison-Wesley, 2001, ch. 5, pp. 139–172.
- [10] D. Johnson, H. Yu, and D. Maltz, "The dynamic source routing protocol (dsr) for mobile ad hoc networks for ipv4," <https://tools.ietf.org/html/rfc4728>. [Online]. Available: <https://tools.ietf.org/html/rfc4728> last accessed September 2014
- [11] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*. Kluwer Academic Publishers, 1996, pp. 153–181.
- [12] C.-H. Lin, B.-H. Liu, H.-Y. Yang, C.-Y. Kao, and M.-J. Tsai, "Virtual-coordinate-based delivery-guaranteed routing protocol in wireless sensor networks with unidirectional links," in *INFOCOM 2008. 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies*, 13–18 April 2008, Phoenix, AZ, USA. IEEE, 2008, pp. 351–355.
- [13] D. Peters, R. Karnapke, and J. Nolte, "Buckshot routing - a robust source routing protocol for dense ad-hoc networks," in *Ad Hoc Networks Conference 2009, Niagara Falls, Canada*, 2009, pp. 268–283.
- [14] "Texas instruments ez430-chronos." [Online]. Available: <http://focus.ti.com/docs/toolsw/folders/print/ez430-chronos.html?DCMP=Chronos&HQS=Other+OT+chronos> last accessed September 2014
- [15] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (dsv) for mobile computers," in *SIGCOMM '94: Proceedings of the conference on Communications architectures, protocols and applications*. New York, NY, USA: ACM Press, 1994, pp. 234–244.
- [16] A. Varga, "The omnet++ discrete event simulation system," in *Proceedings of the European Simulation Multiconference (ESM'2001)*, Prague, Czech Republic, Jun. 2001, p. 185.
- [17] "Texas instruments cc430f6137," <http://focus.ti.com/docs/prod/folders/print/cc430f6137.html>. [Online]. Available: <http://focus.ti.com/docs/prod/folders/print/cc430f6137.html> last accessed September 2014
- [18] R. Karnapke, S. Lohs, A. Lagemann, and J. Nolte, "Simulation of unidirectional links in wireless sensor networks," in *7th International ICST Conference on Simulation Tools and Techniques*, Lisbon, Portugal, 2014, pp. 118–125.

Ultra Wideband Positioning: An Analytical Study of Emerging Technologies

Suheer Alhadhrami,
AbdulMalik Al-Salman
and Hend S. Al-Khalifa

Abdulrahman Alarifi,
Ahmad Alnafessah
and Mansour Alsaleh

Mai A. Al-Ammar

Computer Science Department,
King Saud University,
Riyadh, KSA

Computer Research Institute,
King Abdulaziz City for
Science and Technology,
Riyadh, KSA

Department of Computer Science,
Al-Imam Mohammad bin Saud
Islamic University,
Riyadh, KSA

Email: 430204094@student.ksu.edu.sa Email: aarifi@kacst.edu.sa Email: mai.alammam@ccis.imamu.edu.sa
 salman@ksu.edu.sa aalnafessah@kacst.edu.sa
 hendk@ksu.edu.sa maalsaleh@kacst.edu.sa

Abstract—In recent years, indoor positioning has emerged as a critical function in many end-user applications; including military, civilian, disaster relief and peacekeeping missions. To cope with this surge of interest, much research effort has focused on meeting the needs of these applications and overcoming their shortcomings. Ultra WideBand (UWB) is an important technology in the field of indoor positioning and has shown great performance compared to others. In this work, we identify and analyze existing ultra wideband positioning technologies and present a detailed comparative survey. We also provide a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis, a method generally used in management science to evaluate the strengths, weaknesses, opportunities, and threats involved in a product or technology, to analyze the present state of UWB positioning technologies.

Index Terms—Ultra wideband; localization; positioning; indoor positioning; wireless sensor networks; networking; SWOT

I. INTRODUCTION

Positioning is the process of determining a position of people, equipment, and other objects. Recently, it has been an active research area where much of the research focuses on utilizing existing technologies to address the problem of positions' determination. Positioning can be classified, according to the environment where the positioning is conducted, into two types: outdoor positioning and indoor positioning. Outdoor positioning is performed outside buildings and indoor positioning is performed inside buildings such as houses, hospitals, malls and others. Different applications may require different types of positioning technologies that fit their needs and constraints. For example, Global Positioning System (GPS) is a technology that is suitable and efficient for the outdoor rather than the indoors since satellite radio signals cannot penetrate solid walls and obstacles [1][2][3][4].

Indoor Positioning Systems (IPSs) determine the position of something in a physical space continuously and in real-

time [5]. IPSs use numerous positioning approaches, which vary greatly in terms of accuracy, cost, precision, technology, scalability, robustness and security [2]. There are five main quality metrics of the indoor positioning systems, which should be considered: (1) accuracy and precision of the system; (2) coverage and resolution of the coverage; (3) latency in producing location updates; (4) impact of a building's infrastructure; and (5) effect of random errors on the system such as errors caused by signal interference and reflection [6]. Indoor positioning has many applications such as indoor navigation systems for blind and visually impaired people, locating devices through buildings, aiding tourists in museums, finding an emergency exit in a smoky environment, tracking kids in crowded places, and tracking expensive equipment. Indoor positioning applications may require different quality attributes and thus IPSs should be carefully selected to meet the requirements of the application.

Indoor location-based services are an important application of indoor ubiquitous computing. Accurate position measurement is a critical requirement for indoor positioning techniques. Given that UWB is one of the key techniques that is proven effective in indoor positioning, a comparative analysis of the state-of-the-art UWB indoor positioning systems is indeed required. Furthermore, due to the U.S. Federal Communications Commission (FCC) recent allowance for the use of unlicensed UWB communications, UWB civilian applications have been studied and explored intensively worldwide. Also, the development of international wireless communication standards that adopts UWB technology has encouraged research and development efforts on UWB. Consequently, developing new algorithms to improve UWB positioning performance is becoming an active research area [7].

This work is motivated by the fact that UWB is the most promising technology for indoor positioning and tracking.

Further, to the best of our knowledge, this paper is the first analytical study of the state-of-the-art UWB indoor positioning systems. Our study analyzes a wide range of positioning algorithms that have empowered UWB positioning systems and tackle different aspects of applications needs and requirements. The nature of the application in question plays a major role in determining the appropriate solution for achieving certain quality attributes. Hybrid positioning approaches have future potentials as they combine features of different approaches to improve the performance.

Our contributions in this paper are twofold.

- We provide an updated literature review for existing and recent research in UWB positioning systems (see Sections II and III).
- We conduct a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis for UWB technology, which provide new directions and deeper insights into the state of this technology beyond its well-known pros and cons (see Section IV).

II. UWB POSITIONING

UWB is one of the most recent, accurate, and promising technologies [8]. The precursor technology of UWB is referred to as a base-band, impulse, and carrier-free technology. The US Department of Defense is the first to use the term ultra wideband. At the late of 1990, the UWB was commercially available [8]. UWB radio is a method of spectrum access that can provide high speed data rate communication over the personal area network space. UWB is based on transmitting extremely short pulses and it uses techniques that cause a spreading of the radio energy over a wide frequency band with a very low power spectral density [8]. This high bandwidth offers high data throughput for communication. The low frequency of UWB pulses enables the signal to effectively pass through obstacles such as walls and objects.

There are three main application areas for using UWB, namely (1) communication and sensors; (2) positioning and tracking; and (3) radar [8][9]. UWB positioning techniques can provide real time indoor precision tracking for many applications such as locator beacons for emergency services and mobile inventory, indoor navigation for blind and visually impaired people, people or instruments tracking, and military reconnaissance. The characteristics of UWB signals provide the potential of highly accurate position and location estimation for indoor environments [10][8].

A. Why UWB has recently gained attention?

In general, UWB has different features that are explored in the literature [8][3][11]. The high data rate of UWB can reach 100 Megabits per second (Mbps), which makes it a good solution for near-field data transmission. Also, the high bandwidth helps in reducing the effect of multipath interference, which makes UWB a more desirable solution for indoor positioning than other technologies [12]. In fact, UWB

provides a high accuracy rate in which error can be minimized to sub-centimeters. Therefore, UWB is considered to be one of the most suitable choices for critical positioning applications that require highly accurate result.

UWB technology, unlike other positioning technologies such as infra-red and ultrasound sensor, does not require line-of-site and does not get affected by the existence of other communication devices or external noise due to its high bandwidth and signal modulation [13][14]. Furthermore, the UWB equipment cost is low and it consumes less power compared with other competitive solutions.

Many IPSs were implemented commercially using UWB. One of the most known positioning systems that use UWB is UbiSense system. In UbiSense system, a user carries tags that transmit UWB signals to fixed sensors that use the signals to determine the user's positions using Time of Arrival (TOA) method [15].

B. Signal Modulation

Signal modulation is the process of carrying information on the impulse signal (the carrier signal) by modifying one or more of the signal properties. In general, signal modulation can be categorized based on the signal state into three categories; binary modulation, ternary modulation, and M-ary modulation. Also, signal modulation can be categorized based on signal properties that need to be modified into four categories; amplitude modulation, frequency modulation, phase modulation, and hybrid modulation.

Signal modulation is a crucial phase in signal transmission, which can greatly improve the quality of transmitting signals to achieve certain quality criteria. For example, UWB signals are usually transmitted in the existence of other signals in the air as well as reflected signals which may cause multipath interference. Thus, UWB must have high modulation efficiency as signals must be recognized correctly in the presence of noise and interference [9].

There are various signal modulations that are used for UWB, such as Pulse Position Modulation (PPM), On-Off Keying (OOK), Pulse Amplitude Modulation (PAM), and Pulse Width Modulation (PWM) [9][16]. Signal modulation is utilized to enhance the accuracy of UWB localization [9]. In UWB, Time-Hopping Spread Spectrum (TH-SS) impulse radio could be used to solve multipath problems and generate UWB signals with relatively computational cost. There are other modulations that are used by UWB such as Pseudo Random (PR) time-modulation, Binary Phase Shift Keying (BPSK), Time-Hopping Binary Phase Shift Keying (TH-BPSK), Time-Hopping Pulse Position Modulation (TH-PPM), and Minimum-Shift Keying (MSK) [6][17]. Further details about using these modulation technologies in positioning are presented in the next sections.

III. UWB POSITIONING ALGORITHMS

UWB technology is well-suited for indoor positioning applications. In order to employ this technology, different

TABLE I. COMPARISON OF UWB SYSTEMS

No.	Authors	Year	Accombined Technology	Algorithm	Environment
1	Ch'oliz et al.[26]	2011		TOA	LOS/NLOS
2	Guangliang Cheng[10]	2012		TOA	LOS/NLOS
3	Gunter et al.[27]	2010		TOA	LOS/NLOS
4	Sivanand et al.[14]	2007		TDOA	NLOS
5	Rowe et al.[29]	2013		TOA	LOS
6	Jiang et al.[34]	2010	GPS	AOA, TDOA	NLOS
7	Jiang et al.[34]	2010		TDOA, TDMA	LOS
8	Pittet et al.[31]	2008	MEMS	AOA, TDOA	
9	Shahi et al.[11]	2012		AOA, TDOA	LOS/NLOS
10	Segura et al.[12]	2012		TOA, TDOA and TDMA	LOS/NLOS
11	Cao and Li[47]	2012		ROA, TDOA and DOA	
12	Mucchi et al.[37]	2010		AOA, TOA	LOS
13	Liu et al.[48]	2012	GPS	TDOA, RSS	NLOS
14	Kuhn et al.[17]	2011	GPS, RFID and WLAN		
15	Zhang et al.[49]	2010		TDOA	
16	Deissler et al.[38]	2012		TOA	
17	Tuchler et al.[50]	2005		TOA	LOS
18	Digel et al.[39]	2013			LOS/NLOS
19	Jiang et al.[40]	2012		TOA,RSS	LOS/NLOS
20	Tome et al.[28]	2010		TOA	LOS
21	Arias-de-Reyna and Mengali[13]	2013		TOA	LOS
22	Kilic et al.[51]	2013		TOF	LOS
23	Mahfouz et al.[52]	2011		TDOA	LOS
24	McCracken et al.[53]	2013	RSS Sensor	RSST	
25	Jiang et al.[54]	2013		TOA, TDOA	
26	Yang et al.[55]	2013		STBCS-TDOA	LOS/NLOS
27	Mirza et al.[56]	2012	Ultrasound sensor and compass	AOA,TDOA	
28	Ubisense[57]	2010		AOA,TDOA	LOS/NLOS
29	Ubisense[58]	2011		TOA,TDOA	LOS/NLOS
30	Ubisense[59]	2010		TOA,TDOA	LOS/NLOS
31	Ubisense[60]	2010		TOA/TDOA	LOS/NLOS

positioning algorithms have been developed in which position information is extracted from radio signals traveling between the target node and the reference nodes as well as position information of the reference nodes. There are many positioning algorithms that can be classified into five main categories based on estimating measurements: (1) Angle of Arrival (AOA); (2) Time of Arrival (TOA); (3) Time Difference of Arrival (TDOA); (4) Received Signal Strength (RSS); and (5) hybrid algorithm. In this section, we give a detailed review of these algorithms for UWB indoor positioning. Also, we compare the algorithms in different aspects such as accuracy,

environment, estimation technique, range, purpose of use and others. A summary and comparison of UWB positioning algorithms is presented in Table I.

A. AOA-based Algorithms

In AOA technique, the estimation of the signal reception angles from at least two sources are compared either with signal amplitude or carrier-phase across multiple antennas. From the intersection of the angle line for each signal source, the location can be found. AOA estimation algorithms are very sensitive to many factors, which may cause errors in

their estimation of target position. Furthermore, AOA estimation algorithms have a higher complexity compared to other methods. For instance, the antenna array geometry has an important role in the estimation algorithm [18]. Increasing the distance between the sender and receiver may decrease the accuracy [19]. AOA technique can be used with other techniques to increase the accuracy [20].

AOA based algorithms have been used in many literatures. Xu et al. presented a new cooperative positioning method based on the AOAs, which utilizes pairwise AOAs information among all the sensor nodes rather than relying only on anchor nodes [21]. Lee proposed the using of signal model and weighted-average to estimate AOA parameters for Low data Rate UWB (LR-UWB) [22]. A Kalman filter based AOA estimation algorithm was introduced by Subramanian, which rely on a new linear quadratic frequency domain frequency invariant beamforming strategy [23].

Furthermore, many studies have been conducted to evaluate the performance of AOA for different applications, environments, hardware, and configurations. Mok et al. studied the feasibility and performance of AOA for UWB in Ubisense Real-Time Location System (RTLS) when integrated with GPS to facilitate resource management in underground railway construction sites [24]. The influence of UWB directional antennas on the performance of AOA estimation was presented in detail by Gerok et al. [25]. Gerok et al. presented a corrected AOA estimation algorithm, which mitigates the error resulting from the UWB directional antenna.

B. TOA-based Algorithms

TOA is based on the intersection of circles for multiple transmitters. The radius of those circles is the distance between the transmitter and receiver. This distance is calculated by finding one-way propagation time between them [19]. The time synchronization of all transmitters is required while the receiver synchronization is unnecessary so that any possibility of significant delays must be accounted for during calculation of the correct distances.

Choliz et al. identified a realistic indoor scenario, defined by a layout of walls and corridors, and a specific indoor UWB ranging model to evaluate different kinds of TOA based algorithms for UWB such as Trilateration, Weighted Least Square with Multidimensional Scaling (WLS-MDS), Least Square with Distance Contraction (LS-DC), Extended Kalman Filter (EKF) and Particle Filter (PF) [26].

TOA based algorithms have been used to locate targeted objects for various applications and environments. Cheng et al. designed TOA-based personnel localization system for coal mine using UWB technology, which can be very helpful to locate workers effectively in case of accidents [10]. For mobile robot tracking, Segura et al. proposed a novel UWB navigation system for indoor environment, which employ a TOA based estimation algorithm to accurately locate mobile robot [12]. Fischer et al. designed a monolithic integrated

transceiver chipset for UWB to use them in indoor localization systems where TOA techniques have been used for position estimation [27]. The system was implemented for Line-of-Sight environment and its accuracy was estimated to be 8.3 cm. On the other hand, Tom'e et al. designed and built a large-scale deployable UWB-based Local Positioning System (LPS) in which TOA is used for position estimation [28].

C. TDOA-based Algorithms

TDOA is based on measuring the time difference of arrival of a signal sent by an object and receive by three or more receivers. In this manner, the location of the object (transmitter) will be determined. Also, the scenario can be flipped where a single receiver can determine the target location by measuring the difference in arrival times of two transmitted signals [19].

Typically, only one transmitter is available that requires the multiple receivers to share the data and cooperate to determine the location of the transmitter. This cooperation requires significant bandwidth in comparison with other algorithms.

Krishnan et al. have used TDOA for UWB indoor positioning system where the site has been divided into cells and each cell has four UWB readers mounted on the top corners to have line-of-sight with user tag. In this manner, the readers will be able to receive the signals from the user tag then send the time of arrival to a central processing unit to determine TDOA and find user location [14]. Rowe et al. designed one dimensional system with two sensors and one tag using TDOA based algorithm to determine the tag location [29]. On-Off Keying (OOK) modulation was used to overcome the collision deduced by synchronous tag transmission, increase the performance, and decrease the cost and power at the same time.

D. RSS-based Algorithms

In RSS-based algorithms, the tracked target measures the signal strength for received signals from multiple transmitters in order to use signal strength as an estimator of the distance between the transmitters and receiver. This way, the receiver will be able to estimate its position relative to the transmitter nodes. Although RSS is sensitive to multipath interference and small scale channel effect, which cause a random deviation from mean received signal strength, it is used frequently with unrealistic assumptions such as transmitted power and path loss exponent are known, and transmitter antennas are isotropic [19][30]. According to Pittet et al., the accuracy of RSS for Non-Line-Of-Sight (NLOS) and multipath environment is low, which shows clearly that RSS is not the right estimation method for indoor positioning systems [31]. Gigl et al. explored the performance of RSS algorithms for positioning using UWB technology [32]. They also studied the effect of small scale fading on the system accuracy; however, a simulator based on the UWB channel model 802.15.4a was used to evaluate the algorithms rather than relying on real scenarios for indoor environments.

RSS based algorithms can be classified into two main types: trilateration and fingerprinting [33]. Trilateration algorithms use RSS measurements to estimate the distances to three different reference node and hence estimate the current location. On the other hand, fingerprinting requires collecting a dataset of RSS fingerprints of a scene, which is later used to match on-line measurements with the closest fingerprint in the dataset in order to estimate the location.

E. Hybrid-based Algorithms

When multiple positioning techniques are used, they can complement each other or target different parts of the site that fit with their strengths. Overall accuracy will increase and the complexity and cost will increase, too. Jiang et al. presented a tracking system for staff, patients, and instruments in a hospital environment [34]. They used GPS for outdoor tracking and UWB for indoor tracking. Furthermore, the site was divided into cells where each cell has at least 4 UWB readers and GBS repeater. They used PDA, which has built-in GPS receiver and it was connected to UWB tag in order to work with both GPS and UWB at the same time. The UWB subsystem uses both AOA and TDOA received by UWB readers to estimate the user position. Similarly, Kuhn et al. designed a multi-tag access scheme for UWB localization system, in which Minimum-Shift Keying (MSK) modulation was used with 2.40-2.48, 5.40-10.6 Gigahertz (GHz) frequency and refresh rate of 1-20 Hz in the range of 1m-100m [17]. Also, they have used Time Division Multiple Access (TDMA) for channel access control. TDOA was used to discover new tags and identify its position in 3D. Experimentally, it uses two tags and switches between them 20 times per second.

A new pedestrian navigation solution has been introduced by Pittet et al., which combines UWB localization system and Micro Electro Mechanical Sensors (MEMS) to improve the performance of pedestrian positioning [31]. AOA and TDOA were used to know the presence of multipath and position estimation. Furthermore, they used an Extended Kalman Filter (EKF) based algorithms to couple the measurement of these two subsystems in order to combine the complementary advantages of UWB and MEMS. Another system has been introduced by Shahi et al., which consists of a network of tags and receivers communicating over 68 GHz signals [11]. The path from transmitter to receiver is measured to locate the tag. The true location is determined by the direct path signals; however, the error was produced by reflections of the signals. The direct path signal can be distinguished from reflection using UWB, so the accuracy increases. The computation is calculated in one master server, which uses AOA and TDOA for estimation. Also, FuCheng and MingJing designed UWB localization and tracking system based on Kalman, linear H and extended H filters to accurately estimate the target position using DOA and TOA [35]. Their system was implemented in 30x30 meter cell with one access point, which is equipped with 4 elements array and noise statistics.

Several other systems have been developed for critical missions to help in tracking people and object. An UWB indoor/outdoor NLOS localization system has been implemented for disaster aid, in which GPS is used for outdoor localization while UWB is used for indoor localization [36]. TDOA and RSS are used to improve localization performance. Another UWB tracking system for athlete has been presented by Mucchi et al. in order to determine the athlete's speed and acceleration and analyze his/her performance after medical surgery [37]. They have implemented their system for outdoor environments with different cell sizes and for indoor environments using 4 sensors. The system was implemented for Line-Of-Sight (LOS) environment setup and uses TOA and AOA for positions' estimation with good accuracy. Another system was designed by DeiBler et al., which tackles the problem of simultaneous localization and mapping in an emergency like an earthquake, fire, or terrorist attacks [38]. The system was designed to perform UWB indoor mapping using a mobile antenna array with two receiver antennas and one transmitter between them. DeiBler et al. used Kalman filter for position estimation and Rao-Blackwellized particle filter for data association and initialization of new objects.

Furthermore, a new UWB indoor navigation system was proposed by Segura et al., which includes two sub-systems: the location system and Mobile Robot (MR) control system [12]. They detect the first arrival of signal by designing a novel dynamic threshold crossing algorithm and using TOA/TDOA for estimation. Time Division Multiple Access (TDMA) is used to avoid multi-users interference.

Several other efforts have been done to improve positioning in UWB using hybrid based algorithms. Digel et al. designed and improved a digitizer of no-coherent Impulse Radio Ultra WideBand (IR-UWB) [39]. Jiang et al. designed a technique to mitigate NLOS error by using Biased Kalman Filtering (BKF) and Maximum Likelihood Estimation (MLE) where both AOA and RSS were used [40]. Srimathi and Kannan made a comparison between Time-Hopping Spread-Spectrum (TH-SS), Time-Hopping Binary Phase-Shift Keying (TH-BPSK), and TH-SS coded and un-coded scheme UWB systems [41]. Zebra is a commercial UWB positioning system, that offers a UWB Real-Time Location System (RTLS) integrated with other RTLS, which can use technologies, such as GPS, Radio Frequency Identification (RFID) and Wireless Local Area Network (WLAN) [17].

IV. SWOT ANALYSIS

SWOT analysis is a useful analysis tool to understand and evaluate a technology, solution or business. SWOT analysis aims to identify the key internal (strengths and weaknesses) and external (opportunities and threats) factors that may affect the success of an analyzed target. SWOT analysis has been applied in many areas such as; industry, management and engineering. Here, we apply the SWOT analysis to evaluate UWB in terms of strength, weakness, opportunities and threats

TABLE II. SUMMARY OF SWOT ANALYSIS FOR UWB TECHNOLOGY

Internal	
Strengths	Weaknesses
<ul style="list-style-type: none"> • License free • Low power consumption • Does not interfere with most of the existing radio systems • High level of multipath resolution • Large bandwidth • High data rate communication • High processing gain in communication system • Involve very short pulses • Carrierless transmission property offers the advantage of hardware simplicity. • Work well with low SNR • Low Probability of Intercept and Detection • Resistance to jamming • Penetration through different kinds of material 	<ul style="list-style-type: none"> • Potential interference to the existing systems (ex Wimax in USA) • Affect GPS and aircraft navigation radio equipment. • Potential interference to the existing systems • Very short pulses in UWB may take long time for synchronization
External	
Opportunities	Threats
<ul style="list-style-type: none"> • Robot guidance • Tracking systems • Medical and surgeries that require sub-millimeters of accuracy • Indoor localization systems • UWB short pulses. This short pulses signals can be utilized in non-communication purposes • Sensor, positioning, and identification network (SPIN) • Industrial warehouses applications • Shipboard environment application • Military applications. • Application for noisy environments • Video streaming 	<ul style="list-style-type: none"> • Commercially expensive compared to other technologies • They are in some cases not totally immune to multipath effects • Design and implementation of UWB antennas can be more challenging

to have a deep understating of UWB. A summary of SWOT analysis is shown in Table II.

A. *Strength*

One advantage of using UWB is being licensed free because of its low power. UWB is not classified as a radio equipment as its low power signal does not interfere with most of the existing radio systems [42]. In addition, UWB has a very high level of multipath resolution because of its large bandwidth. Large bandwidth provides frequency diversity that makes Time Modulated Ultra WideBand (TM-UWB) signal resistant to the multipath problems and interference [42]. Time Modulated UWB has a low probability of interception and detection, which it is used in some applications, particularly in the military.

The large bandwidth is the main feature of the UWB wireless systems. This feature offers an improved channel capacity and high data rate communication in digital communication systems [43]. The channel capacity is defined by Shannons

law, where is the channel capacity is proportional to the bandwidth (B) and the log of Signal to noise ratio $\frac{S}{N}$ plus one.

$$C = B \log_2 \left(1 + \frac{S}{N} \right) \tag{1}$$

In addition to the advantage of large bandwidth, it is potential for high processing gain in communication systems. Processing gain for real Direct-Sequence of UWB (DS-UWB) modulation systems is defined as two times ratio of noise bandwidth at the front end of the receiver to the noise bandwidth of symbol rate. Here is the formula of DS-UWB processing gain [43]:

$$PG = 2 \times \frac{\text{Noise Bandwidth}}{\text{Symbol Rate}} \tag{2}$$

The large processing gain offers a greater immunity distortion and noise. It allows negative Signal to Interference and Noise Ratio (SINR) to be recovered [43].

UWB signals have greater penetration of obstacles (such as

walls) than conventional signals, and they achieve same data rate [44]. Furthermore, UWB transmissions involve very short pulses, which have recently received significant interest. Very short pulses offer an advantage in terms of resolvability of multipath components [44]. Many received signals in an environment that are characterized by multipath is a superposition of the delayed replicas of the signal. This has been avoided in UWB because the reflection from objects and surfaces near the path between transmitter and receiver tend to not overlap in time because of the very short pulses of UWB. This means UWB has a desirable direct resolvability of direct multipath components.

UWB technology's carrierless transmission property offers the advantage of hardware simplicity and small hardware. UWB transceivers can be built with much simpler radio frequency architecture than narrowband systems with fewer components. Also, there is no need for power amplifier because of its low power consumption [44]. In general, UWB hardware is considered to be simple and the hardware simplicity depends on the application and operational scenario. For example, the transmitter does not need Analog to Digital (A/D) converter, digital pulse shaping filter, or equalizer to correct carrier phase distortion [44].

B. Weakness

Although UWB has many strengths for different applications, it has some weakness. One of these weaknesses is the possibility of interference with nearby systems [43]. In the United States, the UWB frequency range for communication applications is 3.1 to 10.6 GHz, which is operating in the same frequencies as popular communication products such as Worldwide Interoperability for Microwave Access (WiMAX) and digital TV. In some countries, it may also interfere with some systems such as third-generation 3G wireless systems [43]. There are some concerns that several UWB devices may cause harmful interference to GPS and aircraft navigation radio equipment [44]. To overcome those concerns, different techniques have been developed to eliminate harmful interference with other sensitive services, such as Detection and Avoidance (DAA) [43].

Also interference may happen from the existing system to the UWB system. The UWB systems signals may spread over other bandwidths that contain existing frequency of narrowband systems [44]. This interference can be elevated by using Minimum Mean-Square Error (MMSE) multiuser detection schemes to reject strong narrowband interference.

Although using very short pulses in UWB has many advantages, the UWB receiver requires signal acquisition, synchronization and tracking to be done with very high precision in time relative to the pulse rate. These steps of processes are time-consuming and take a long time to be performed [44]. There are some techniques for reducing this time such as using preamble sequence for rapid acquisition.

C. Opportunities

UWB becomes a choice for many systems that require high accuracy such in building robot guidance and tracking systems to utilize its advantages. Furthermore, UWB is used for medical applications that require sub-millimeters of accuracy [15]. In addition, UWB is used in radars in order to improve their high performance [45].

For indoor localization systems, there are multipath reflections from objects inside rooms, which negatively impact radio signals. However UWB signals have time resolution, which offers a high resolution positioning applications to solve the multipath problems [43].

As mentioned before, UWB communication signals have short pulses. Those short pulse signals can be utilized in non-communication purposes [44]. For instance, the low power UWB RFID tag transmitters have been used to locate objects with an accuracy proportional to the inverse of the signal bandwidth.

UWB could be beneficial for industry and service providers in many applications such as Sensor, Positioning, and Identification Network (SPIN) systems [46]. These systems need a large number of devices (sensors and tags) in industrial warehouses to transmit low-rate data combined with position information. This allows the devices to operate over long distance (around 100m) between mobile tags and sensors of UWB.

There are some challenges for using Radio Frequency (RF) operation for the shipboard environment. Using UWB and network analyzer measurements offers good opportunities for NLOS communication for indoor and on ships [44]. It allows signals to propagate well aboard ships and around objects, which provide reasonable accuracy to determine positions. UWB is used in radar in order to improve its high performance [45].

D. Threats

UWB usually does not have a negative impact on the neighbor's devices because there are some techniques that are used to avoid the interference with other devices [15]. However, UWB is still commercially expensive compared to other technologies (see [9] for further limitations).

While UWB systems are known to be robust against multipath reflection issues, they are not totally immune to multipath effects [46]. One of these cases is when there is an extreme ratio of link distance to antenna height. This may result in signal losses and propagation delay that lasts to tens or even hundreds of nanoseconds.

The design and implementation of antennas for UWB systems can be more challenging than the bandwidth and variable conditions of operation [46]. This may add some limitations to UWB systems in comparison with conventional RF.

V. CONCLUDING REMARKS

Positioning is one of the most important and challenging phases in navigation systems where different technologies have been developed to improve performance. In this paper, we presented an analytic study of UWB positioning, in which a detailed and updated overview of UWB indoor positioning techniques has been presented. Furthermore, we presented a SWOT analysis of UWB technology, focusing mainly on positioning applications, in which internal factors (strengths and weaknesses) as well as external factors (opportunities and threats) have been identified and analyzed. UWB systems have emerged as one of the leading technologies for indoor positioning and have been used in many more applications than before.

ACKNOWLEDGMENT

This project was supported by NSTIP strategic technologies program number (11- INF1855-02) in the Kingdom of Saudi Arabia.

REFERENCES

[1] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE computer*, vol. 34, no. 8, pp. 57–66, 2001, ISSN: 0018-9162.

[2] H. Huang and G. Gartner, *A survey of mobile indoor navigation systems*. Springer, October 2010, chapter 20, pp. 305–319, ISBN: 978-3-642-03293-6.

[3] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1067–1080, 2007, ISSN: 1094-6977.

[4] S. Ram and J. Sharf, "The people sensor: a mobility aid for the visually impaired," in *Second International Symposium on Wearable Computers*. IEEE Society, October 1998, pp. 166–167.

[5] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 13–32, 2009, ISSN: 1553-877X.

[6] H. Wu, A. Marshall, and W. Yu, "Path planning and following algorithms in an indoor navigation model for visually impaired," in *Internet Monitoring and Protection, 2007. ICIMP 2007. Second International Conference on*. IEEE Society, July 2007, pp. 38–38.

[7] M. Al-Ammar *et al.*, "Comparative Survey of Indoor Positioning Technologies, Techniques, and Algorithms," in *Cyberworlds (CW), 2014 International Conference on*. IEEE Computer Society, October 2014, pp. 1–8.

[8] M. Ghavami, L. B. Michael, and R. Kohno, Eds., *Front Matter Ultra Wideband Signals and Systems in Communication Engineering*. John Wiley & Sons, Ltd, February 2006, ISBN: 9780470867532.

[9] K. Siwiak and D. McKeown, Eds., *Ultra-Wideband Radio Technology*. John Wiley & Sons, Ltd, 2005, ISBN: 978-04-70-85-93-39.

[10] G. Cheng, "Accurate toa-based uwb localization system in coal mine based on wsn," *Physics Procedia*, vol. 24, pp. 534–540, 2012, ISSN: 1875-3892.

[11] A. Shahi, A. Aryan, J. West, C. Haas, and R. Haas, "Deterioration of UWB positioning during construction," *Automation in Construction*, vol. 24, pp. 72–80, 2012, ISSN: 0926-5805.

[12] M. Segura, V. Mut, and C. Sisterna, "Ultra wideband indoor navigation system," *IET Radar, Sonar & Navigation*, vol. 6, no. 5, pp. 402–411, 2012, ISSN: 1751-8792.

[13] E. Arias-de Reyna and U. Mengali, "A maximum likelihood UWB localization algorithm exploiting knowledge of the service area layout," *Wireless personal communications*, vol. 69, no. 4, pp. 1413–1426, 2013, ISSN: 0929-6212.

[14] S. Krishnan, P. Sharma, Z. Guoping, and O. Woon, "A uwb based localization system for indoor robot navigation," in *Ultra-Wideband, 2007. ICUWB 2007. IEEE International Conference on*. IEEE Society, September 2007, pp. 77–82.

[15] U. Company, "Ubisense website," 2009, URL: <http://www.ubisense.net/en/> [accessed: 2014-04-01].

[16] S. Cui, "Modulation and multiple access techniques for ultra-wideband communication systems," Ph.D. dissertation, Cleveland State University, 2011.

[17] M. Kuhn, M. Mahfouz, J. Turnmire, Y. Wang, and A. Fathy, "A multi-tag access scheme for indoor UWB localization systems used in medical environments," in *Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireless), 2011 IEEE Topical Conference on*. IEEE Society, January 2011, pp. 75–78.

[18] S. Al-Jazzar, A. Muchkaev, A. Al-Nimrat, and M. Smadi, "Low complexity and high accuracy angle of arrival estimation using eigenvalue decomposition with extension to 2D AOA and power estimation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 1–13, 2011, ISSN: 1687-1499.

[19] N. Reddy and B. Sujatha, "TDOA Computation Using Multicarrier Modulation for Sensor Networks," *International Journal of Computer Science & Communication Networks*, vol. 1, no. 1, pp. 85–90, 2011, ISSN: 2249-5789.

[20] S. Gezici *et al.*, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 70–84, 2005, ISSN: 1053-5888.

[21] J. Xu, M. Ma, and C. Law, "AOA Cooperative Position Localization," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*. IEEE Society, November 2008, pp. 1–5.

[22] Y. Lee, "Weighted-Average Based AOA Parameter Estimations for LR-UWB Wireless Positioning System," *IEICE transactions on communications*, vol. 94, no. 12, pp. 3599–3602, 2011, ISSN: 1745-1345.

[23] A. Subramanian, "UWB Linear Quadratic Frequency Domain Frequency Invariant Beamforming and Angle of Arrival Estimation," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*. IEEE Society, April 2007, pp. 614–618.

[24] E. Mok, L. Xia, G. Retscher, and H. Tian, "A case study on the feasibility and performance of an UWB-AoA real time location system for resources management of civil construction projects," *Journal of Applied Geodesy*, vol. 4, no. 1, pp. 23–32, 2010, ISSN: 1862-9024.

[25] W. Gerok, M. El-Hadidy, S. El Din, and T. Kaiser, "Influence of the real UWB antennas on the AoA estimation based on the TDoA localization technique," in *Antennas and Propagation (MECAP), 2010 IEEE Middle East Conference on*. IEEE Society, October 2010, pp. 1–6.

[26] J. Chóliz, M. Eguizabal, A. Hernandez-Solana, and A. Valdovinos, "Comparison of Algorithms for UWB Indoor Location and Tracking Systems," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*. IEEE Society, May 2011, pp. 1–5.

[27] G. Fischer, O. Klymenko, D. Martynenko, and H. Luediger, "An impulse radio UWB transceiver with high-precision TOA measurement unit," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE Society, September 2010, pp. 1–8.

[28] P. Tome *et al.*, "UWB-based Local Positioning System: From a small-scale experimental platform to a large-scale deployable system," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE Society, September 2010, pp. 1–10.

[29] N. Rowe, A. Fathy, M. Kuhn, and M. Mahfouz, "A UWB transmit-only based scheme for multi-tag support in a millimeter accuracy localization system," in *Wireless Sensors and Sensor Networks (WiSNet), 2013 IEEE Topical Conference on*. IEEE Society, January 2013, pp. 7–9.

[30] S. Wang *et al.*, "System implementation study on RSSI based positioning in UWB networks," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*. IEEE Society, September 2010, pp. 36–40.

[31] S. Pittet, V. Renaudin, B. Merminod, and M. Kasser, "UWB and MEMS based indoor navigation," *Journal of Navigation*, vol. 61, no. 3, pp. 369–384, 2008, ISSN: 0373-4633.

[32] T. Gigl, G. Janssen, V. Dizdarevic, K. Witrissal, and Z. Irahhtauten, "Analysis of a UWB Indoor Positioning System Based on Received Signal Strength," in *Positioning, Navigation and Communication, 2007. WPNC '07. 4th Workshop on*. IEEE Society, March 2007, pp. 97–101.

[33] N. Kodippili and D. Dias, "Integration of fingerprinting and trilateration techniques for improved indoor localization," in *Wireless And Optical Communications Networks (WOCN), 2010 Seventh International Conference On*. IEEE Society, September 2010, pp. 1–6.

[34] L. Jiang, L. Hoe, and L. Loon, "Integrated UWB and GPS location sensing system in hospital environment," in *Industrial Electronics and*

- Applications (ICIEA), 2010 the 5th IEEE Conference on.* IEEE Society, June 2010, pp. 286–289.
- [35] F. Cao and M. Li, “An Algorithm for UWB Signals Tracking Based on Extended H Filter,” *Physics Procedia*, vol. 33, pp. 905–911, 2012, ISSN: 1875-3892.
- [36] J. Liu, Q. Wang, J. Xiong, W. Huang, and H. Peng, “Indoor and Outdoor Cooperative Real-Time Positioning System,” *Journal of Theoretical & Applied Information Technology*, vol. 48, no. 2, pp. 1066–1073, 2013, ISSN: 1992-8645.
- [37] L. Mucchi, F. Trippi, and A. Carpini, “Ultra Wide Band real-time location system for cinematic survey in sports,” in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 3rd International Symposium on.* IEEE Society, November 2010, pp. 1–6.
- [38] T. Deissler, M. Janson, R. Zetik, and J. Thielecke, “Infrastructureless indoor mapping using a mobile antenna array,” in *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on.* IEEE Society, April 2012, pp. 36–39.
- [39] J. Digel *et al.*, “Integrator and Digitizer for a non-coherent IR-UWB Receiver,” in *Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 2013 IEEE 13th Topical Meeting on.* IEEE Society, January 2013, pp. 93–95.
- [40] X. Jiang, H. Zhang, and W. Wang, “NLOS error mitigation with information fusion algorithm for UWB ranging systems,” *The Journal of China Universities of Posts and Telecommunications*, vol. 19, no. 2, pp. 22–29, 2012, ISSN: 1005-8885.
- [41] S. Srimathi and P. Kannan, “Literature Survey for Performance evaluation of various time hopping ultra-wideband communication system,” *International Journal of Scientific & Engineering Research*, vol. 4, no. 2, pp. 1–3, 2013, ISSN: 2229-5518.
- [42] M. Hämäläinen, V. Hovinen, and M. Latva-aho, “Survey to Ultra Wideband Systems,” *European Cooperation in the Field of Scientific and Technical Research*, vol. 262, pp. 1–7, 1999.
- [43] R. Aiello and A. Batra, Eds., *Ultra wideband systems: technologies and applications.* Newnes-Elsevier, June 2006, ISBN: 978-0750678933.
- [44] L. Miller, “Why uwb? a review of ultrawideband technology,” Tech. Rep., 2003, technical Report to NETEX Project Office, DARPA by Wireless Communication Technologies Group, National Institute of Standards and Technology, Gaithersburg, Maryland, URL: <http://www.ntis.gov/search/product.aspx?ABBR=PB2012113101> [accessed: 2014-08-01].
- [45] S. Ye, J. Chen, L. Liu, C. Zhang, and G. Fang, “A novel compact UWB ground penetrating radar system,” in *Ground Penetrating Radar (GPR), 2012 14th International Conference on.* IEEE Society, June 2012, pp. 71–75.
- [46] D. Porcino and W. Hirt, “Ultra-wideband radio technology: potential and challenges ahead,” *Communications Magazine, IEEE*, vol. 41, no. 7, pp. 66–74, 2003.
- [47] F. Cao and M. Li, “An Algorithm for UWB Signals Tracking Based on Extended H Filter,” *Physics Procedia*, vol. 33, pp. 905–911, 2012, ISSN: 1875-3892.
- [48] J. Liu, Q. Wang, J. Xiong, W. Huang, and H. Peng, “Indoor and Outdoor Cooperative Real-Time Positioning System,” *Journal of Theoretical and Applied Information Technology (JAITT)*, vol. 48, no. 2, pp. 1066–1073, 2012, ISSN: 1875-3892.
- [49] “Realtime non-coherent UWB positioning radar with millimeter range accuracy: theory and experiment,” *Microwave Theory and Techniques, IEEE Transactions*, vol. 58, no. 1, pp. 9–20, January 2010, ISSN: 0018-9480.
- [50] M. Tuchler, V. Schwarz, and A. Huber, “Location accuracy of an UWB localization system in a multi-path environment,” in *Ultra-Wideband, 2005. ICU 2005. 2005 IEEE International Conference on.* IEEE Society, September 2005, pp. 414–419.
- [51] Y. Kilic, H. Wymeersch, A. Meijerink, M. Bantum, and W. Scanlon, “UWB device-free person detection and localization,” *CoRR*, vol. abs/1303.4092, 2013.
- [52] M. Mahfouz, M. Kuhn, Y. Wang, J. Turnmire, and A. Fathy, “Towards sub-millimeter accuracy in UWB positioning for indoor medical environments,” in *Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireless), 2011 IEEE Topical Conference on.* IEEE Society, January 2011, pp. 83–86.
- [53] M. McCracken, M. Bocca, and N. Patwari, “Joint ultra-wideband and signal strength-based through-building tracking for tactical operations,” in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Conference on.* IEEE Society, June 2013, pp. 309–317.
- [54] H. Jiang, Y. Zhang, H. Cui, and C. Liu, “Fast three-dimensional node localization in UWB wireless sensor network using propagator method digest of technical papers,” in *Consumer Electronics (ICCE), 2013 IEEE International Conference on.* IEEE Society, January 2013, pp. 627–628.
- [55] D. Yang, H. Li, Z. Zhang, and G. Peterson, “Compressive sensing based sub-mm accuracy UWB positioning systems: A space-time approach,” *Digital Signal Processing*, vol. 23, no. 1, pp. 340–354, 2012, ISSN: 1051-2004.
- [56] R. Mirza, A. Tehseen, and A. Kumar, “An indoor navigation approach to aid the physically disabled people,” in *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on.* IEEE Society, March 2012, pp. 979–983.
- [57] C. Brown, “Real-time location of Jena’s buses and trams with Ubisense RTLS,” April 2010, ubisense Report, URL: <http://www.ubisense.net/en/news-and-events/press-releases/real-time-location-of-jenas-buses-and-trams-with-ubisense-rtls.html> [accessed: 2014-09-01].
- [58] M. Baum, “RTL in Longueuil selects bus yard management solution provided by Solotech, ISR Transit and Ubisense,” Oct 2011, ubisense Report, URL: <http://www.ubisense.net/en/news-and-events/press-releases/rtl-in-longueuil-selects-bus-yard.html> [accessed: 2014-09-01].
- [59] C. Brown, “Ubisense launches Intrinsically Safe location tracking tags for personnel safety in the Oil and Gas industry,” May 2010, ubisense Report, URL: <http://www.ubisense.net/en/news-and-events/press-releases/ubisense-launches-intrinsically-safe-location-tracking-tags-in-the-oil-and-gas-industry.html> [accessed: 2014-09-01].
- [60] U. Cambridge, “Ubisense tag combines uwb and gps for ‘best-of-both worlds’ tracking named as csr location summit fast-pitch finalist,” Oct. 2010, ubisense Report, URL: <http://www.ubisense.net/en/news-and-events/press-releases/ubisense-tag-combines-uwb-and-gps-.html> [accessed: 2014-09-01].

Integrating Smart Items and Cloud Computing in Healthcare Scenarios

Sarfaraz Ghulam, Johannes Schubert, Gerrit Tamm, Vladimir Stantchev

Institute of Information Systems (IIS)

SRH Hochschule Berlin

Berlin, Germany

e-mail: opsit@srh-hochschule-berlin.de

Abstract-Use of cloud computing in healthcare is a promising trend, particularly in utilization of smart items. Access, affordability and quality to proper healthcare is a great challenge in world society. With the utilization of smart sensors, there are possibilities to improve the quality of healthcare services whenever needed. Hence, such smart items services lead to economic advantages for the whole healthcare system. Keeping in mind these opportunities as well as challenges such as security or legal issues, it is important to bring healthcare and IT together. As a result, healthcare business processes need to be modeled in order to provide IT solutions tailored for practice-oriented applications. In this paper, we present a three-level architecture for a smart healthcare infrastructure. Our approach is based on recent literature work and results from conducted expert interviews with healthcare specialists and IT professionals. To demonstrate the applicability of this architecture model, we provide an example use-case as a template for any kind of smart sensor-based healthcare infrastructure.

Keywords-sensors; cloud computing; smart healthcare; business processes; reference model.

I. INTRODUCTION

During the time of its emergence, it was believed that cloud computing might not be successful for long. But against all the predictions made by IT pundits, cloud computing turned into a big success and kept benefitting different sectors which is continue till today [1]. Cloud computing is the focus of interest for institutions with different backgrounds, including pharmaceuticals, healthcare outfits and medical researches where institutions are highly engage in finding cure and solutions for chronic diseases. Integration of cloud computing with smart items which allows healthcare professionals to monitor the patients' treatment is a significant achievement, which was difficult a decade ago. One reason for its success is the economical features. Sophisticated Information and Communication Technology (ICT) is often expensive and not affordable for medium size businesses including hospitals, where smaller budget hinder the adaptation of latest ICT [2]. Contrary to other ICT, cloud computing and smart items are relatively cost effective and easily adaptable [3]. Adopting cloud services can relatively reduce the

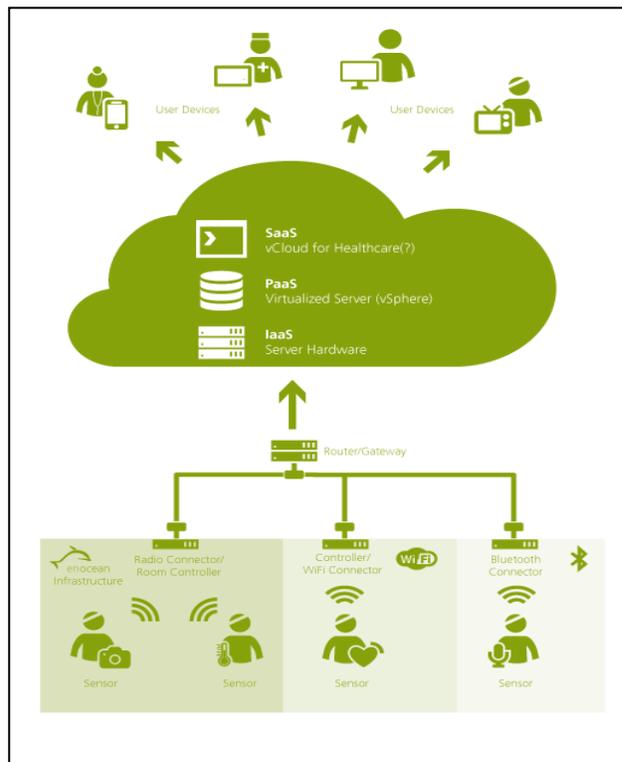


Figure 1. OpSIT architecture prototype

burden of infrastructure and save the maintenance cost and provide relieve while maintaining the infrastructure. The usage of smart items like collaborative radio frequency identification (RFID) solutions in the textile and medical industry have already shown a positive impact, such as continuous quality improvement [4]. Furthermore, patients can privately maintain it in a private environment which is supportive to medical adherence. A report issued by world health organization (WHO) states that there is a majority of patients who do not adhere with their medical prescriptions as they are supposed to which results in delay or failure in the treatment of the diseases [5]. In such scenarios smart healthcare items can support in monitoring patient adherence and support the better treatments by providing solutions to such challenges like smart pill boxes, which come up

with solutions like opening at accurate time and recording of pill time. Furthermore smart healthcare items like physiological parameter monitors can automatically record the health parameter and update it to the cloud, which become available for patient and the healthcare staff who can make comparison with record history and can take corrective decisions on behalf it.

While the economic opportunities of cloud computing in healthcare are quite promising and often subjects of recent research [6], optimization of smart items infrastructure for healthcare processes is underrepresented in research [7]. There exist approaches that define composability for such embedded architecture [8] with a specific focus on dependability in distributed environments [9]. In this regard the approach of architectural translucency [10] [11] aims to offer standardized ways of controlling such architectural properties via an architecture-wide view while using standardized measures such as replication [12] [13]. Applications range from position sensing [14] to project portfolio management [15] and consider applications in the healthcare domain specifically [16] [17].

The project “Optimaler Einsatz von Smart-Items-Technologien in der Stationären Pflege”, Germany (OpSIT) builds on these existing works and is conducting literature research, workshops, and expert interviews with healthcare specialists as well as IT professionals to model reference processes for practice-oriented cloud applications in the healthcare domain. Figure 1 shows an overview of the working mechanism of the project. In smart environment, smart sensors are responsible for monitoring different vital signs of human health, and upload and update it to the respective cloud server where the data become available and accessible for individuals who are involved in the treatment process or directly related to the patient.

The aim of this paper is to introduce an architectural approach for smart items in healthcare environments that is being developed as part of the research objectives of the OpSIT project. In Section II, we will introduce the benefits of sensors in healthcare. Subsequently, we will explain how sensors are used in cloud computing environments (Section III). As a consequence, in Section IV, we will define eHealth in the context of cloud computing. Section V presents our

architecture model, including an application example and how it can be modified for individual purposes. Managerial implications will be presented in Section VI for smart homecare and in Section VII for smart stationary care. We will finish the paper with a compact conclusion in Section VIII.

II. BACKGROUND

The use of smart items has successfully been practiced in the hospitals and other healthcare institutes. Nowadays, sensors are embedded in a variety of instruments for use at home, elderly houses, clinics, and hospitals and providing a critical evaluation of physiological, physical and mental state of the patients. Most of the diagnosis will not be possible nor affordable without using even simple sensors, such as thermometer, blood glucose monitors, electrocardiography, electro-encephalography etc.

These sensors are composed of transducers and capable of detecting electrical, thermal, optical, chemical, genetic and other signals with physiological origin. Signal processing algorithms can help to calculate, forecast and measure different features of human health based on input from these sensors. Measuring the vital state of a patient is also important for devices like pacemakers and insulin pumps [18].

The dimensions which are helping in advancing computer technologies in healthcare can be structured as follows:

- **Sensing:** Involvement of new, effective and cheap technologies with the ability to diagnose and provide immediate results and solutions to the healthcare and public sectors on promising basis. For example, a non contact electrocardiogram can be helpful in detecting the symptoms like heart diseases, [19].
- **Cost:** During past few years, mostly available sensors were quite expensive and hardly affordable for small clinics. But now, the revolution in this industry and availability of cheap sensors, such as RFID, made it possible for clinics to afford it even making personalized treatment possible.
- **Size:** The sensors are quite handy and easily adjustable in pockets which facilitate to carry them anywhere. These devices are available in different shapes and types like blood pressure

monitors, pulse reading wrist watches, blood glucose monitors, etc.

- **Data:** Data collected by the sensors are relatively in mass, and require proper evaluation by experts from time to time. The only solution in this regard is to transfer it to the cloud, which is the best solution to handle and maintain the big data.
- **Cloud services:** The three fundamental cloud services models which are available to any type of businesses are also available to healthcare to support the industry in their information systems management [19]:
 - a) **IaaS:** Enables the healthcare service provider to rent the fundamental cloud resources like storing, processing, and networking.
 - b) **PaaS:** Helps the healthcare service providers to deploy their own or rented applications, software, libraries and tools on cloud infrastructure.
 - c) **SaaS:** Configured software running on cloud infrastructure is available where healthcare institutes can subscribe under the management of the cloud contractor.

III. CLOUD COMPUTING AND SMART ITEMS

A large variety of definitions for cloud computing exists, but there is no single universal upon agreed definition. Authors and experts define cloud according to their own understandings. The most commonly used definition comes from NIST (National institutes of standard and technology) “*Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service providers interaction*“ e.g., networks, servers, storage, applications, and services that can be rapidly provisioned and released with minimal management efforts or service providers interaction [20]. The utilization of the word "cloud" gives reference to the two essential concepts [21]:

- **Abstraction:** The notion of the technology is not disclosed to the users and developers. Unspecified physical systems are used to run the applications, location of the data is undisclosed, further it allows to delegate the administration of the infrastructure while maintaining ubiquitous access of the users.
- **Virtualization:** It is a pool of resource sharing. Centralized storage capability offers storage

provisioning and can be done when required. The cost model is similar to utility model. Moreover it offers multi tenancy, while providing dynamic resource allocation.

When we speak about cloud computing in healthcare, it becomes very obvious that like for other industries, healthcare can also take full advantage out of it. Internally it can ease the burden of the infrastructure and number of people associated with it [22] and allows the institutes to focus on their core competencies. Synergistically cloud computing and smart items are helping patients, clinics and insurance providers to access the health records of the patient whenever needed. As a result, the development in smart healthcare technologies such as mobile healthcare, wireless sensors and cloud computing lessen the requirement for visiting medical facilities and consultation, which can be remotely fulfilled and significantly reduce the manpower requirements, while providing quality treatment to the patient by making remote treatment and consultancy possible and achievable [19]. However, the implementation of such technologies implicates different challenges such as security and privacy issues [23], technological restrictions, or management and governance issues [6].

On one side, it's highly important to understand the evolving business processes occurring in healthcare environments before developing a cloud and smart items infrastructure. On the other side the technical possibilities and requirements needs to be understood to create an architectural approach for smart healthcare.

Wireless sensors can easily be deployed in any environment and with the help of cloud computing, the information can be gathered and saved from these sensors at any time. Smart devices are evolving at a rapid pace in health monitoring, while meeting the needs and demands of assisted living and healthcare providers. This system focuses on the different features for a mobile healthcare system.

The active smart systems with the help of mobile devices collect physiological signals such as body temperature, pulse rate, etc. Once the data collection is completed, it will be transmitted through Wi-Fi or another compatible system network, which will be stored, synchronized, and shared instantaneously on a server. Health symptoms can be diagnosed continuously and immediately. The collected health

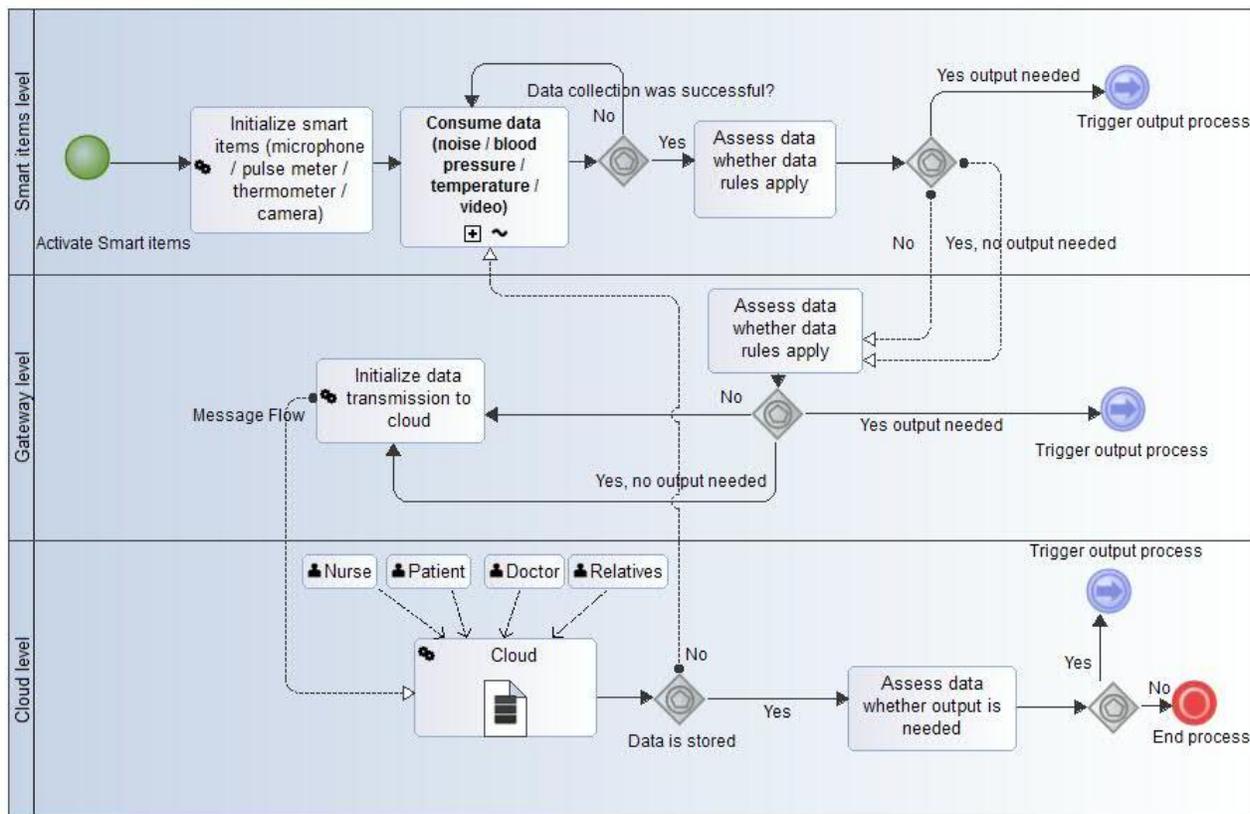


Figure 2. BPMN Model of process steps of the application scenario

data can be consolidated and accessed with a cloud service in order that health professionals can analyze the information and continue patient treatment on the conclusion of the collected data.

IV. EHEALTH AND CLOUD GOVERNANCE

WHO defined health as a "state of complete, physical, mental and social well-being and not merely the absence of infirmity" [24]. As a consequence, healthcare can be defined as the maintenance of physical and mental well-being, supported by available medical services. Finally, eHealth is the maintenance of physical and mental well-being while employing computer technology resulting in improved and better quality of services.

Cloud Governance denotes the idea of applying corporate governance concepts to the world of cloud computing [25]. This can be of particular relevance for healthcare, due to the complex requirements regarding data privacy and information security [26]. Furthermore, aspects of performance evaluation [27] or ensuring the reliability of systems, e.g., through replication [12] [13].

V. COMMUNICATION IN A SMART ITEM ENVIRONMENT

Communication among the devices is empowered by proper communication channels between them. In

the past, when sensors' connectivity was isolated to an integrated user interface, it was difficult to connect with. Contrary to that, modern day sensors are capable of connecting with a large number of interfaces, such as wired, RS 232, USB and Ethernet as well as wireless ones, including Bluetooth, Zigbee, UWB for short distance and Wi-Fi, radio frequency, or ZigBee for longer range [18] [28]. Some devices come with the capability of connecting directly to Smartphones and computers allowing sensors to save data in the non-volatile memory and make them available for later usage.

It is a recursive process, where the respective sensors remain active and observe even minor changes in the pattern of human body which human cannot easily observe. As shown in Figure 2, this process is divided in three different levels. In each level, devices are responsible to perform the assigned tasks and forwarding the data to the further level, or triggering directly an output process. The entire process accomplish in a short period of time (in nanoseconds) making it very fast while helping human to have good control over their health and take corrective measures and steps as soon as possible.

In general, Figure 2 shows the process flow modeled with Business Process Modeling Notation (BPMN) based on the working mechanism of smart

sensors architecture from the OpSIT project. As BPMN is an established way to analyze and design operational process flows regarding multiple kinds of resources [29] it is utilized in the presented architecture which aims to enhance current healthcare processes. Moreover, the architecture model can be seen as a process flow template for any specific use-case, importing it and revising its tasks with modeling software such as Modelio.

As a very basic example use-case, we consider a blood pressure measurement process in home care. To analyze the traditional process, it is important to keep in mind three different levels matching with the levels presented in our approach. Hence, the “smart items level” can be entitled in the traditional process as “measuring device level” where a patient uses the blood pressure gauge that directly outputs the data on a display. The “gateway level” can be compared with a “short-term storage level” where the patient will be manually note down the value shown on the display. The “cloud level” illustrates permanent access and evaluation of the data, while its equivalent in the traditional process may be called “evaluation level”, when the doctor receives the patients’ data in order to evaluate whether medical intervention is necessary or not. This simple comparison points out the disruption within the traditional process flow. Obviously, every level in the traditional process ends with an output task and is not directly connected to the next level. Analyzing this problem is made possible by considering the proposed BPMN model even at the stage of modeling the status quo process.

VI. SMART HOMECARE

Mobile networks give rapid access to the patient data provided by the servers of the healthcare cloud computing environment.

Smart healthcare items benefits are broad range and dual natured (benefiting healthcare providers as well as the patients). These are the physically interconnected hardware performing their tasks seamlessly through a network. For healthcare providers, it is valuable for several reasons: At first, it reduces the number of labor due to the replacement of human based monitoring. Moreover, it reduces the chance of human error. Secondly, wearable sensors are capable of sensing even very small changes in vital signs and recording it, which a human cannot easily

observe, such as pulse and blood oxygen level. Thirdly, on-time measurement can help in saving lives while collected health data of the patient stored in the cloud can support in decision making process whenever needed as enabling the doctor in making well informed decisions. And last but not least human physicians will be able to get valuable assistance from the “electronic physician” [30].

Patients will be benefitted in several ways. They can get quality healthcare treatment at home within their own private environment while living close to their family members while feeling more comfortable and relaxed resulting in less mental stress. Other significant factors involved are dignity and convenience which are highly supported while the patient gets treatment at home.

VII. SMART STATIONARY CARE

Most of the smart items available in the healthcare sector can be utilized in stationary care scenarios as well. Due to cost pressure, hospitals as well as intensive care units or other patient care facilities, are facing challenges like less financial resources. As a result, reduction of labor cost becomes the critical criterion for the implementation of smart items infrastructure in a stationary setting.

According to a study conducted by the European Commission, the large majority (81%) of the hospitals within the EU are connected to the internet and two third (66%) of all surveyed hospitals have an in-house wireless infrastructure, an increase of twelve percents points within two years. Moreover, most of the used IT services are managed in-house which means that qualified IT staff is generally available [31]. Consequently, the basic infrastructure for smart items implementation is already provided in the majority of hospitals. The ongoing challenge is the integration of IT professionals and medical experts to realize smart items solutions supporting healthcare processes in a proper way.

VIII. CONCLUSION

Cloud computing based smart healthcare solutions and stationary items can play a vital role in improving the quality of health services in near future by remarkably supporting care staff in fulfilling their tasks in a well managed manner.

Wireless communication channels between sensors, gateways and other devices enables continuous data transfer to the cloud and making it available for doctors, care staff and patient to access the health status and maintain the records time to time. Hence, synergistic combinations among smart healthcare solutions and cloud help the doctors and other care staff in making highly informed decision and produce better output with quality of care service provided.

Understanding healthcare processes and modeling them for IT application is an inevitable task to create practice-oriented cloud solutions with the aim of providing high quality smart healthcare services. The approach presented in this paper can be used as a basic design for further developments.

ACKNOWLEDGMENT

This publication is based on the research project “OpSIT” which received financial funding from the ministry of education, science, research and technology (BMBF) under the funding sign 16SV6048 within the program „IKT 2020 – Forschung für Innovationen“ and by the German Federal Ministry of Economics and Technology through project PrevenTAB (KF3144902DB3). The authors are responsible for the content.

REFERENCES

- [1] N. Sultan, "Making use of cloud computing for healthcare provision: Opportunities and challenges," *International Journal of Information Management*, vol. 34, no. 2, pp. 177-184, 2014.
- [2] European Commission, "Benchmarking deployment of eHealth services," Luxembourg, 2013.
- [3] Y. Han, "On the clouds: a new way of computing," *Information Technology and Libraries*, vol. 29, no. 2, pp. 87-92, 2013.
- [4] C.M. Birkmeyer et al., "Will Electronic Order Entry Reduce Health Care Costs?," *Eff Clin Pract.*, vol. 5, no. 2, pp. 67-74, 2002.
- [5] World Health Organization, "Adherence to long-term therapies: evidence for action," Geneva, Study ISBN 9241545992, 2003.
- [6] A. M.-H. Kuo, "Opportunities and challenges of cloud computing to improve health care services," *Journal of medical Internet research*, vol. 3, no. 13, 2011.
- [7] Division of Engineering and Applied Sciences, "Technical Report TR-08-05," Harvard University, 2005.
- [8] M. Werner et al., "Composability concept for dependable embedded systems," in *Proceedings of the International Workshop on Dependable Embedded Systems in conjunction with SRDS*, 2003.
- [9] P. Ibach et al., "CERO: CE RObots Community," *IEE Proceedings Software*, vol. 152, no. 5, pp. 210-214, 2013.
- [10] V. Stantchev, *Architectural Translucency*, 8th ed. Berlin: GITO mbH Verlag, 2008.
- [11] V. Stantchev and M. Malek, "Architectural translucency in service-oriented architectures," *IEE Proceedings-Software*, vol. 153, no. 1, pp. 31-37, 2006.
- [12] V. Stantchev, "Effects of replication on web service performance in WebSphere," International Computer Science Institute, Berkeley, 2008-03, 2008.
- [13] V. Stantchev and M. Malek, "Addressing web service performance by replication at the operating system level," in *Internet and Web Applications and Services*, 2008, pp. 696-701.
- [14] P. Ibach, V. Stantchev, and C. Keller, "DAEDALUS—A Peer-to-Peer Shared Memory System for Ubiquitous Computing," in *Euro-Par 2006 Parallel Processing*, 2006, pp. 961-970.
- [15] V. Stantchev, M.R. Franke, and A. Discher, "Project portfolio management systems: Business services and web services," in *Fourth International Conference on Internet and Web Applications and Services*, 2009, pp. 171-176.
- [16] V. Stantchev, "Intelligent systems for optimized operating rooms," in *New Directions in Intelligent Interactive Multimedia Systems and Services - 2*. Berlin Heidelberg: Springer, 2009, pp. 443-453.
- [17] V. Stantchev, "Enhancing health care services with mixed reality systems," in *The Engineering of Mixed Reality Systems*. Berlin Heidelberg: Springer, 2009, ch. 17, pp. 337-356.
- [18] J.G. Ko et al., "Wireless Sensor Networks for Healthcare," in *Proceedings of the IEEE 98.11*, 2010, pp. 1947-1960.
- [19] E.-M. Fong and W.-Y. Chung, "Mobile Cloud-Computing-Based Healthcare Service by Noncontact ECG Monitoring," *Sensors*, vol. 13, no. 12, pp. 16451-16473, 2013.
- [20] NIST. (2011) The NIST Definition of Cloud Computing. Special Publication 800-145.
- [21] B. Sosinsky, *Cloud computing Bible*. Indiana: Willey, 2011.
- [22] Hitachi Data Systems, "How to Improve Healthcare with Cloud Computing," California, 2012.
- [23] J. Schubert et al., "Datenschutzrechtliche Rahmenbedingungen für die Datenerhebung durch

- AAL-Infrastruktur bei der Pflege," in *Wohnen–Pflege–Teilhabe–"Besser leben durch Technik"*, Berlin, 2014.
- [24] World Health Organization. (1948, Apr.) WHO website. [Online].
<http://www.who.int/about/definition/en/print.html>
(Accessed: 20 September 2014)
- [25] K. Petruch, V. Stantchev, and G. Tamm, "A survey on IT-governance aspects of cloud computing," *International Journal of Web and Grid Services*, vol. 7, no. 3, pp. 268-303, 2011.
- [26] S. Dzombeta et al., "Governance of Cloud Computing Services for the Life Sciences," *IT Professional*, vol. 16, no. 4, pp. 30-37, 2014.
- [27] V. Stantchev, "Performance evaluation of cloud computing offerings," in *Advanced Engineering Computing and Applications in Sciences*, 2009, pp. 187-192.
- [28] J.-S. Lee, Y.-W. Su, and C.-C. Shen, "A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi," in *The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON)*, Taipei, 2007, pp. 46-51.
- [29] W.M.P. van der Aalst et al., "Business process management: A survey," in *BPM 2003 Proceedings*, Eindhoven, 2003, pp. 1-12.
- [30] J.A. Stankovic et al., "Wireless Sensor Networks for In-Home Healthcare: Potential and Challenges," in *High confidence medical device software and systems (HCMDSS) workshop*, Virginia, 2005, pp. 2-3.
- [31] European Commission, "European Hospital Survey: Benchmarking Deployment of eHealth Services," Brussels, 2014.
- [32] M.H. Baeg et al., "Building a Smart Home Environment for Service Robots Based on RFID and Sensor Networks," in *International Conference on Control, Automation and Systems*, Seoul, 2007, pp. 1078-1082.

The Impact of Link Lengths on Energy Consumption in Wireless Sensor Networks

Knut Øvsthus¹, Espen Nilsen¹

1. Dep. Electrical Engineering
Bergen University College
Bergen, Norway

kovs@hib.no, espen.nilsen.no@gmail.com

Anne-Lena Kampen^{1,2}, Øivind Kure²

2. ITEM, NTNU
Trondheim, Norway

alk@hib.no, okure@item.ntnu.no

Abstract— Conservation of energy is one of the main challenges in designing a wireless sensor network (WSN). The reason is that these large scaled networks cannot be arranged, configured, or maintained manually. Thus, automated deployment and configuration are required. One important factor determining the total energy consumption is the network topology. This article evaluates the relation between the maximum distance (link lengths) between the nodes in a WSN and the total energy consumed. The optimal topology for the two most commonly used medium access control (MAC) protocols were found. A WSN based on a Time Division Multiple Access (TDMA) protocol is limited by the maximum available or allowed emitted radio power. Thus, the criterion for optimal link lengths is related to the expected number of transmissions over the links. By including the retransmissions over the links we found an optimal internode distance. A Carrier Sense Multiple Access (CSMA) based WSN, on the other hand, is limited by the consumed energy of the overhearing nodes. In an analysis including only the overhearing nodes, the link lengths should be as short as possible and the connectivity of the network limits the link length used. However, we found that in a sparsely populated WSN, the total energy consumption increased for shorter link lengths as they were decreased from the optimal link length.

Keywords; WSN; Energy Efficiency; Multi-Hop Routing; Hop Length; Network's Life Time.

I. INTRODUCTION

A typical wireless sensor network (WSN) consists of several battery powered autonomous devices. The devices are equipped with a unit for sensing targeted environmental attributes and a communication unit that enables communication with a designated node, that provide data collection (sink). The communication capability of a sensor node mainly serves two tasks: transmitting the sensor data generated by the node and relaying packets on behalf of other nodes. This article focuses on the transmission and transport of data packets through the WSN.

The characteristics of radio units in a WSN differ from traditional radios. Many find it strange that the receiver consumes approximately the same amount of energy receiving (RX) a packet as the sender consumes in transmitting (TX) the packet. The reason is the low power emitted from the sender, as explained in [1][2]. The datasheet of the RF Transceiver CC2420 [3] verifies the statement. Due to the equality in energy consumption, it is very important to reduce the number of overhearing nodes. Overhearing nodes are in the reception range of the sender,

but they are not the destination or next hop node. However, they learn this after receiving the packet and analyzing the packet's Medium Access Protocol (MAC) address. The conclusion in [4] was that for a WSN using a Carrier Sense Multiple Access (CSMA) protocol the energy consumption is dominated by the overhearing nodes. In a WSN using the time division multiple access (TDMA) protocol it was found [4] that the distance between two communicating nodes (the terms internode distance and link lengths are also used) should be as long as feasible. Alternative solution would include for example a relay node. It would consume much energy during reception and retransmission.

Based on published data, this article first proposes a function that relates the packet delivery ratio (PDR) and the distance between two communicating nodes. Next, an expression is derived for the expected number of transmissions required along a path. Simulations are used for validating and comparing the statistical result. Finally, we analyzed a 2-D WSN. Following the randomly node deployment, a routing protocol defines the topology. The routing protocol had the core functionality of RPL [5] with an object function [6] that limited the candidate link according to a link quality requirement. We used the expected number of transmissions as requirement. Finally, the energy consumption in a WSN with different link lengths is found for WSN using CSMA and TDMA. We found that both protocols have an optimum internode distance.

The novelty of the article is a more realistic function relating the packet loss probability and the node distance. Currently, most publications are based on a model where the packet loss probability is zero if the node distance is less than the transmission range, and if the node distance is greater the losses are 100%. Although this is not a good model, it is used mainly due to a lack of alternatives. The second contribution is that the model is applied for determine the optimum link distance in WSNs using TDMA and CSMA.

The paper first presents related works, and then the model for packet losses is introduced and applied for assessing packet losses over a single path. Next, the model is applied for a real WSN, where energy optimum link lengths are found for a WSN using TDMA and a WSN using CSMA. Finally, the paper presents the conclusion.

II. RELATED WORK

Energy saving in WSN is vital for the operation of the network and many publications have assessed this issue. This article revisits the challenges addressed in [4]. In [4], no lower optimum transmission range was found for a WSN using a CSMA MAC protocol. Another result was that in a WSN using a TDMA MAC protocol, the transmission should be as long as possible and no upper limit was found. In addition to the related publications presented in [4], this section supplements related works.

Several publications have addressed the topic of transmission power in WSN. The conclusion found in the survey [7] was that the transmission power should be as low as possible to reduce the energy consumption in overhearing nodes. This conclusion has been the starting point of several publications on transmission power control. Several earlier publications have stated the opposite conclusion, as in [8]. It lists twelve reasons for having long internode distances (using high transmission power). One reason is that longer hops are more efficient as they are closer to the Euclidean distance between source node and sink. The same results are found in [9][10], where sparsely populated random networks were shown to cause long paths. Our simulation confirms these results. A general observation regarding transmission power is that the conclusion depends on whether the assessment has included all effects that determine the total energy consumption. The effects are the physical radio, the link quality, the MAC layer, and the routing layer. Furthermore, the conclusion depends on the technology used, for example which MAC technology that is used. The conclusion of [8] is valid for a TDMA, but as the MAC layer is not included in the assessment the conclusion is not valid for a CSMA. However, the conclusion found in [7] is valid for CSMA as the overhearing nodes dominate the total energy consumption.

Statistical analyses of WSN paths are presented in several publications [11]-[14]. Several of these articles consider path reliability under high traffic loads where the queuing of packets is included in the intermediate nodes. Our analysis is of WSNs with low traffic load, which is a valid assumption for many WSN that reports data at a low duty cycle.

Another contribution of this article is the proposal of applying a *Fermi-Dirac* function for expressing the relation between PDR and distance between the two communicating

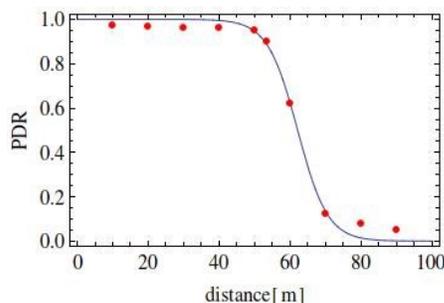


Figure 1. Packet delivery ratio (PDR) as a function of distance between sender and receiver. The red dots are measured data adopted from [17].

nodes. Based on this model we derive an expression for the losses along a path in a WSN. The model for the link properties was also used in assessing an ordinary randomly deployed WSN. We found that the *Fermi-Dirac* function was proposed as a function relating the packet losses and Receiver Signal Strength Indicator (RSSI) in [15]. Our proposal is to use this function for relating distance and PDR.

III. PACKET LOSS RATE FOR A LINK

The performance of the individual radio links determines much of the overall network performance in a WSN. This section presents the fundamental performance issues related to a single radio link. The radio used follows the characteristics of the RF Transceiver CC2420 [3].

Some publications present measured performance of sensor nodes [15]-[18], like the PDR as a function of distance. The observed relation between packet loss and RSSI are presented in [15]. In [15], the *Fermi-Dirac* function is proposed as a function relating the packet losses and RSSI. Here, we propose to use the same function for relating the PDR and the distance between two communicating nodes:

$$f(x) = \frac{1}{1 + e^{\frac{x-x_0}{x_1}}} \tag{1}$$

Figure 1 show data extracted form [17] and the approximated function. The internode distance is x , and x_0 and x_1 are fitting parameters. The PDR_1 , for a single trial of transmitting a packet between two nodes is: $PDR_1(x) = f(x)$. The packet loss rate (PLR) is $PLR_1(x) = 1 - f(x)$.

Radio links are made more robust by retransmitting packets that are not acknowledged. However, the number of retransmissions has to be limited. Without an upper limit, a high number of retransmissions depletes the node's energy and causes large, unpredictable delays over the links. Therefore, a maximum number, m , of transmission trials are permitted each packet before the sender discards it. This means that the sender tries to retransmit the packet a maximum of $(m-1)$ times. Based on the PDR_1 for a single packet, it is possible to estimate the $PDR(m)$ of a link:

$$PDR(m) = p_1 * \text{Sum}[(1-p_1)^i, \{i, 0, m-1\}] = 1 - (1 - p_1)^m \tag{2}$$

where $p_1 = PDR_1$ and the summation of the series is written using the notation: $\text{Sum}[a_i, \{i, 1, m-1\}] = a_1 + a_2 + \dots + a_{m-1}$. Solving for the packet loss rate, $PLR(m)$, over the link: $PLR(m) = (1 - p_1)^m$

IV. PACKET ERROR RATE FOR A PATH

A WSN is a two dimensional (2-D) network as shown in the example in Figure 2. Each node, depicted as blue dots, may produce data destined for the sink located in the lower left corner. As can be seen in this example WSN, not all nodes were directly connected to the sink and they required

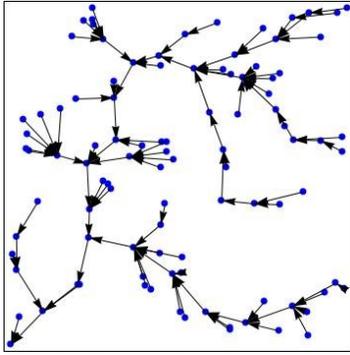


Figure 2. A WSN with 100 nodes (blue dots), including the sink in the lower left corner. The paths are show as arrows.

help of neighboring nodes to relay their packets. The pattern of arrows was found by a routing protocol. Each link was chosen among a set of candidate links. Routing is explained in the next section.

In this section, we present the results of data packet transmission along a single path. This one-dimensional network (1-D network) of N nodes is illustrated in Figure 3. In our assessment, the data-producing node, Node1, was to the left and the sink (the destination) was to the right. Between these two nodes were a number of relaying nodes. Their only task was to relay the data packet produced by Node1.

Our goal was to forward a data packet using as little energy as possible. The comparison was between different node arrangements using different link lengths. We sought the node arrangement that consumed least energy. The only energy consumption that differed between the node arrangements was due to the difference in the number of packet transmitted and received (TX/RX). The network was designed according to common WSN principles where only the communicating nodes (that is the current sender and receiver) were consuming TX or RX energy. Nodes not participating in the communication were in sleep mode. The number of packet TX/RX for a given node arrangement was proportional to the energy difference between the node arrangements, this comparison is commonly used [11]. The MAC protocol used in the 1-D WSN was TDMA. In TDMA, each pair of communicating nodes was assigned a time slot. The optimal node arrangement was found as the configuration that required least packet treatments. The assessment of the 1-D network was based on a statistical description of the link performance and simulated performance.

The statistical analysis was based on the performance of the individual links. The distance from the source node (Node1) to the sink was L_{tot} . L_{tot} was a fixed parameter in our assessment. The distance between the nodes was L_{link} . L_{link} was assumed to apply for all inter node distances, except the last link to the sink, which may be shorter, as explained below. Given the distance between nodes (L_{link}), the number of nodes required to connect the source and sink

was found as $\text{Ceiling}[L_{tot}/L_{link}]$. The $\text{Ceiling}[x]$ operator returns the smallest integer greater than or equal to x .

The expected performance of the 1-D network depending on L_{link} is: Short internode distances will have few retransmissions at each link, but the path will consists of a higher number of links compared to a path of longer internode distances. However, as the link distances approaches the zone of less quality, retransmission over the individual links starts to limit the gained benefit of reduced hop counts. In the end no packets get through, when the link distance causes disconnected links. Then the sender at Node1 only transmits the packet m times over the first link, before discarding the packet. The description given above is supported in the following statistical derivation of the expected performance.

1) Number of transmissions along a path

The expected number of transmission (ETX) over a single link, allowing maximum m transmissions is:

$ETX(m) = (1-q_1) \text{Sum}[\{n \ q_1^{n-1}\}, \{n, 1, m\}] + m \ q_1^m$, using the notation $q_1 = PLR_1$. The summation gives the following result:

$$ETX(m) = (1 - q_1^m)/(1 - q_1) \tag{3}$$

Along the path shown in Figure 3, the probability of reaching node number z (Node $_z$) is:

$$P[\text{Packet reaching Node}_z] = (1 - q_1^m)^{(z-1)} \tag{4}$$

We assume statistically independent links. Thus, the expected number of transmissions along the path is:

$$ETX_{path}(N) = ETX(m) * \text{Sum}[(1 - q_1^m)^{(i-1)}, \{i, 1, N-1\}]$$

$$ETX_{path}(N) = \{1 - q_1^m - (1 - q_1^m)^N\} / \{q_1^m (1 - q_1)\} \tag{5}$$

The performance of a 1-D network can be derived from (5). If only high quality links are used, PDR_1 approaches one (q_1 approaches zero). Then the number of transmissions required is $N-1$, which equals the hop count. However, if disconnected links are used, PDR_1 approaches zero (q_1 approaches one). Then it is found from (5) that the number of transmissions equals the maximum transmissions of the first link, which is m .

Figure 4 illustrates the average number of transmissions required to reach the sink from a node located 500 meters away. The parameter that was changed was the distance between the nodes. Both theoretical and simulated results are presented. The effect of discontinuous changes in the

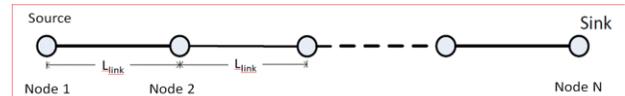


Figure 3. A path from source node to the sink represented as a 1-D network.

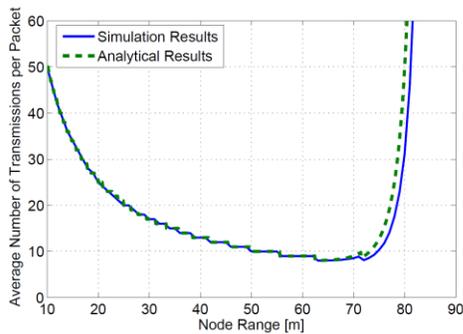


Figure 4. Number of transmission as a function of node separation.

number of links appears as discontinuous changes in the number of transmissions. Fewer transmissions are needed when longer link lengths are permitted since fewer hops are required. However, poor link quality caused retransmission. This can be seen by the increase in number of transmission at link lengths above 70 meters. The theoretical results are given by (5) while the simulation results were produced using the OMNeT++ [20] simulation framework and simulation models available in the MiXiM simulator [21]. To mimic a real world sensor node the theoretical and simulation were according to the CC2420 datasheet [3] and the IEEE802.15.4 standard [19].

V. OPTIMAL TOPOLOGY IN A WSN

The previous analyses established some fundamental understanding of the performance of packet transfer along a path subjected to packet errors. In this section, the performance of the individual links are used to find the optimal network configuration in terms of link length that minimize the total energy consumption in a two-dimensional (2-D) WSN. The comparison follows the method presented earlier where the performance is according to the energy consumption for different network configurations. The difference is presented in number of times packets transmitted and received (TX/RX). Here, all nodes, except the sink, generate data. This implies that the nodes are both data generators and some are also relay nodes.

The simulation was done by randomly deploying the nodes in a 400x400m² area, establish routing based on link criteria, and then count the number of times each packet had to be transmitted and received in order to get to the sink. The sink was placed, as shown in Figure 2, in the lower left corner. The motivation was that we wanted to investigate the consequences of long paths. For a given maximum internode distance, each node arrangement was repeated at least 200 times in order to gain statistical confidence. The nodes were interconnected as the routing protocol determined which links each node should use to forward their data. However, not all node configurations produced a connected graph. These arrangements were not included in our results. Especially for the short maximum internode distances (producing sparsely WSN), it was difficult to produce a fully connected WSN.

The number of transmission permitted at each link was introduced above as the number m . In our analyses of a single path, it was also pointed out that when link lengths become too large the end-to-end performance is compromised, as the packets do not get through to the sink. A low value for m would produce an optimum network with disconnected links. Therefore we used $m=100$ in our simulations.

Before the number of TX and RX was found the nodes had to be arranged and interconnected. The nodes were arranged by randomly positioning them in the area. The next step was to interconnect the nodes using a routing protocol. Here, we used the selection criteria of the RPL routing protocol [5], with an object function that selected only links fulfilling the defined EXT requirement [6]. The requirement is given in the following text. The ETX was found from the proposed function relating PDR and internode distances and the ETX of the individual links derived above. Thus, this selection criteria and the random node positioning produced links with different lengths, but with an upper limit with respect to internode distance. This means there was a maximum link length (or internode distance).

Different MAC protocols produce different results. Therefore, the two most popular MAC protocols were evaluated according to their characteristic performance as presented below. In the 1-D network, it was assumed that only the communicating nodes were active. In our analyses of the 2-D WSN we used the same assumption. The first MAC protocol analyzed was a TDMA protocol, where the two communicating nodes were assigned a time slot. Time slots were assumed allocated during network establishment (as in, e.g., WirelessHART™ [22]). The overhead due to network configuration is not included in our assessment. The motivation for omitting this is that the total energy consumption is determined by the long-term operation. Furthermore, several assumptions would have had to be made regarding signaling and initial connections. The second MAC protocol analyzed was a CSMA MAC protocol. In CSMA, all nodes in the reception area of the sender decoded the packet in order to determine if the packet was destined to them. Since all nodes in the sensing range overheard the messages, it was concluded in [4] that the transmission range (link length) should be as short as possible.

Before the energy consumption results are presented, some observations are given regarding paths length in WSN.

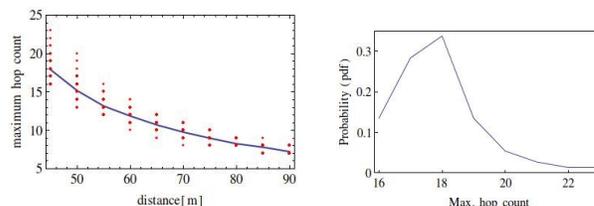


Figure 5. Maximal hop count (left) as a function of maximal permitted node distance. The figure to the right shows the estimated probability distribution function (pdf) of maximal hop for a node distance of 45m

The reason for this discussion is that the characteristics of the paths in WSN determine the performance of the network. It was pointed out in [9][10] that short link length in randomly deployed networks results in long paths. This was clearly observed in our simulations. In Figure 5, the maximum number of hops is presented for a network of 200 nodes. The figure shows that the variation in maximum hop count is larger for networks using short link lengths. In addition, as illustrated in the graph on the right side in Figure 5, the distribution of maximum length has a tail towards long paths. These long paths demand many transmissions in order to forward packets to the sink. Clearly this is negative as they consume much energy, and in a CSMA based WSN the long paths will have many overhearing nodes along the paths.

A. Optimal topology for TDMA

The TDMA simulations are based on a radio with a fixed transmission power. The motivation was that in TDMA, the link length should be as large as possible, following the conclusion in [4]. However, the emitted power cannot be increased beyond a limit. The limit is determined by the design of the radio and/or the regulations.

In the TDMA simulations, we sought optimal performance with respect to optimum internode distance. The limitation on the distance was the ETX of the individual links. ETX is proposed as an alternative metric to filter out only links fulfilling a defined quality determined by the object function [6]. In our model there is a direct relation between ETX and the distance between the nodes, as presented above. The challenge of estimating the ETX was thus avoided, and the routing protocol did not have to establish this information based on over the air communications. The assumption was according to the assumption stated earlier that the management traffic was not included.

In Figure 6, a typical result is presented. Starting with the shortest allowed link length, it can be seen that by allowing longer links the total energy is reduced as the paths get shorter. However, the reduced energy consumption reaches a minimum, in the figure at approximately 60m. Further increase in allowed hop length causes increased total energy consumption. The reason is that longer

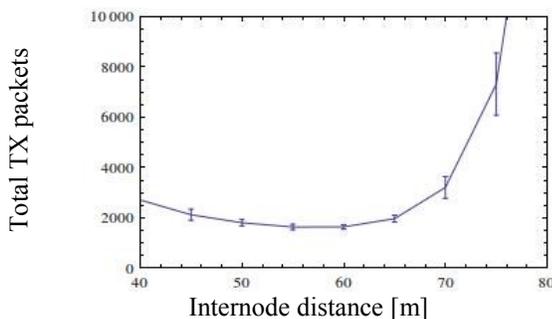


Figure 6. Number of packets transmitted as a function of maximum allowed link length through a TDMA WSN.

internode distance causes higher ETX. The probability of retransmission over the links increases and this retransmission consumes energy.

B. Optimal topology for CSMA

The CSMA simulations were performed differently from the TDMA simulations. Here, the emitted power of the sender is changed. The reason is that for a CSMA WSN, we seek the optimal link lengths for the radio as in [4]. Change in the TX power causes changes in the parameters x_0 and x_1 , in (1). In our simulation, x_0 was increased and the original ratio between x_1 and x_0 determined the new x_1 . Thus, the grey zone increased with increasing TX power according to observations [17].

The ETX was used as a criterion for link selection in the CSMA simulations. Only links having a PDR_1 equal to or better than 90% were considered candidates for routing. The derived relations above defined the relation between ETX and PDR_1 . Clearly, nodes outside the link length overheard the transmission. These nodes were in the sensing range (gray zone), where they received packets, but most often with bit error. However, the nodes consumed energy in receiving the packets and this energy was included in our analyses. The node’s sensing range was set equal to x_0 . It might be argued that this is too short, but here we used this as a first approximation.

Figure 7 presents the results for a WSN based on CSMA. In the figure, the left graph shows all data points in addition to the average result of number of TX/RX. The curve to the right shows only the average number of TX/RX required (it is similar curve as the line in the left graph). The interesting observation is that there is a point where the number of required TX/RX starts to increase if the transmission range is decreased beyond an optimal hop length. The reason for this increase is the long paths that are likely to occur as the network is operated close to its connections limits. A second, interesting observation is the variation in number of TX/RX that is shown for the short internode distances. From the figure, it can be seen that for a maximum link length of 45m, there are some very low values and some very high values for the maximum hop count. From these results it can be concluded that it is advantageous to use long-hop also for CSMA.

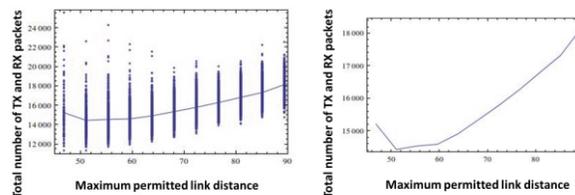


Figure 7. Total number of TX and RX packets handling as function of maximum link distance. The figure to the left shows both average and all simulated results while the figure on the right, average numbers are given.

VI. CONCLUSION

A *Fermi-Dirac* function was suggested for relating the packet delivery ration (PDR) and the distance between the nodes. The function enables analytical evaluation of WSN. The parameters of the function were found by curve-fitting to published results. Next, the *Fermi-Dirac* function was applied in estimating the total number of transmissions along a path. The estimated number of transmissions showed good correlation to simulated results.

In a WSN with randomly deployed nodes, the link quality of candidate links available for routing was evaluated for a TDMA based WSN. The consequences of having too strict requirement on the links caused higher energy consumption due to the long paths. However, beyond an optimal point the total energy consumption increased rapidly as each link had to retransmit the packet due to the reduced link quality. Thus, it was found that TDMA based WSN has an optimal internode distance.

An optimal node distance was also found for a CSMA based WSN. This is contrary to an intuitive conclusion, where the maximum link length should be as short as possible in order to reduce the unnecessary energy consumption in the overhearing nodes. The increasing total energy consumption for short link lengths was found to be due to the long, none optimal paths that occurred when short link lengths were used.

Our findings can be used as a tool for designing a WSN as a more realistic description of the relation between the link quality and node distance is found. It was found that an optimum node distance exist for WSN based on both CSMA and TDMA.

REFERENCES

[1] R. Min and A. Chandrakasan, "Top Five Myths about the Energy Consumption of Wireless Communication", *ACM Mobile Computing and Communication Reviews*, vol. 7, No 1, Jan. 2003, poster

[2] H. Karl and A. Willig, "Protocols and Architectures for Wireless Sensor Networks", ISBN: 978-0-470-51923-3, John Wiley & Sons, 2007

[3] CC2420, Chipcon Products from Texas Instruments, <http://www.ti.com/lit/ds/symlink/cc2420.pdf>, visited July 2014

[4] A-L. Kampen, K. Ovsthus, L. Landmark, and O. Kure, "Energy Reduction in Wireless Sensor Networks by Switching Nodes to Sleep During Packet Forwarding", *The Sixth International Conference on Sensor Technologies and Applications, SENSORCOMM 2012, Italy, 2012*, ISBN 978-1-61208-207-3. s. 189-195

[5] T. Winter, et al., "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks", RFC 6550, Internet Engineering Task Force (IETF), 2012

[6] JP. Vasseur, M. Kim, K. Pister, N. Dejean, and D. Barthel, "Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks", RFC6551, Internet Engineering Task Force (IETF), 2012

[7] I. Khemapech, A. Miller, and I. Duncan, "A Survey of Transmission Power Control in Wireless Sensor Networks", *The 8th Annual Postgraduate Symposium The Convergence*

of Telecommunications, Liverpool John Moores University, 28th-29th June 2007

[8] M. Haenggi, "Twelve reasons not to route over many short hops", *VTC2004-Fall. 2004 IEEE 60th Vehicular Technology Conference*, Vol. 5, 2004, pp. 3130 – 3134.

[9] P. Szczytowski, A. Khelil, A. Ali, and N. Suri, "TOM: Topology oriented maintenance in sparse Wireless Sensor Networks", *8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2011, pp. 548 – 556.

[10] S. Vural and E. Ekici, "On Multihop Distances in Wireless Sensor Networks with Random Node Locations", *IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 9, NO. 4*, 2010, pp. 540 – 552.

[11] Z. Rosberg, R.P. Liu, T. L. Dinh, Y. F. Dong, and S. Jha, "Statistical reliability for energy efficient data transport in wireless sensor networks", *Wireless Netw*, 16, 2010, pp. 1913-1927.

[12] R P. Liu, Z. Rosberg, I. B. Collings., C. Wilson, A. Y Dong, and S. Jha, "Energy Efficient Reliable Data Collection in Wireless Sensor Networks with Asymmetric Links", *Int J Wireless Inf Networks*, 16, 2009, pp- 131-141-

[13] T. Zheng, E. Kamel, and S. Wang, "Node performance model of wireless sensor networks", *2013 9th Asian Control Conference (ASCC)*, 2013, pp. 1 – 6.

[14] N. Farzaneh and M. H. Yaghmaee, "Probability based hop selection approach for resource control in Wireless Sensor Network", *2012 Sixth International Symposium on Telecommunications (IST)*, 2012, pp. 703 – 708.

[15] A. Bildea, O. Alphand, F. Rousseau, and A. Duda, "Link quality metrics in large scale indoor wireless sensor networks", *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2013, pp. 1888 – 1892.

[16] J. Zhao and R. Govindan, "Understanding Packet Delivery Performance In Dense Wireless Sensor Networks", *Proceedings of the 1st international conference on Embedded networked sensor systems, SenSys '03*, Pages 1 - 13, ACM New York, USA 2003, pp. 1 – 13.

[17] Y-D. Lee, D-U. Jeong, and H-J. Lee, "Performance analysis of wireless link quality in wireless sensor networks", *5th Int. Conf. on Computer Sciences and Convergence Information Technology (ICCIT)*, 2010, pp. 1006 – 1010.

[18] L. Siyu and G. Hongju, "Propagation characteristics of 2.4GHz wireless channel in cornfields", *IEEE 13th International Conference on Communication Technology (ICCT)*, 2011, pp. 136 – 140.

[19] Available at <http://www.ieee802.org/15/pub/TG4.html>

[20] OMNeT++, <http://www.omnetpp.org/>, visited July, 2014

[21] MiXiM, <http://mixim.sourceforge.net/>, visited July 2014

[22] WirelessHART™, IEC 62591 ed1.0, http://webstore.iec.ch/webstore/webstore.nsf/ArtNum_PK/43964, visited July 2014

Towards a Generic Cloud-based Sensor Data Management Platform: A Survey and Conceptual Architecture

Vincent C. Emeakaroha, Kaniz Fatema, Philip Healy, John P. Morrison
 Irish Centre for Cloud Computing and Commerce (IC4)
 University College Cork, Ireland
 Email: {vc.emeakaroha, k.fatema, p.healy, j.morrison}@cs.ucc.ie

Abstract—With the increasing numbers of sensors and smart devices being deployed worldwide, the volume of data they generate is becoming difficult to store and process on local platforms. Cloud computing provides scalable resources that could address these issues. However, platform-independent methods of gathering and transmitting sensor data to Clouds are not widely available. This paper presents a survey of Cloud-based sensor monitoring and data gathering platforms. It reviews the state-of-the-art and discusses their strengths and weaknesses. Informed by the survey, the paper further proposes a generic conceptual architecture for achieving a platform-neutral Cloud-based sensor monitoring and data gathering platform. We also discuss the objectives, design decisions and the implementation considerations for the conceptual architecture.

Keywords—Sensor Monitoring Platforms; Sensor Data Gathering; Cloud Computing; Generic Sensor Cloud Platform; Interoperable Communication

I. INTRODUCTION

Data gathering IT techniques such as those underlying entity tracking and control systems, for example, have shown the commercial value of real-time control of real-world devices. Based on this, more generalised applications for sensor devices are becoming significant in real world usages. Sensors enable access to remote objects and environmental information providing the raw materials from simple monitoring through to next-generation applications such as smart cities. The gathering, storage and processing of sensor data locally is becoming very costly due to the increase in the number of sensors and the volume of data they generate.

In a similar manner, the Internet of Things (IoT) [1], which promise to connect objects, devices and humans, generate large volumes of data. Harnessing these data by organisations can be a complex process due to heterogeneous operating systems, varying connectivity protocols and legacy application compatibility. Furthermore, the ability to draw meaningful insights from the volume of data unleashed by these technologies is another challenge.

Cloud platform offers scalable compute and storage resources to support the management of these data [2]. However, solutions [3]–[6] available today are mostly customized for particular usages. To unlock the business potential in this area, generic solutions are required to address the core challenges such as communication bottle-neck, data interchange formats, security and interoperability.

This paper presents a survey reviewing the state-of-the-art of Cloud sensor monitoring platforms and data gathering techniques. Based on the survey findings, it further

proposes a platform-neutral architecture providing a generic Cloud interface that addresses the identified challenges. The main contributions are (i) a survey of industrial, commercial and open source sensor monitoring and data management platforms; (ii) analysis of the platforms to identify trends, issues and challenges; and (iii) the proposal of a generic architecture utilising standardised data interchange formats and interoperable communications.

The rest of the paper is organised as follows: Section II presents the related work. The state-of-the-art analysis is performed in Section III. Section IV discusses the identified challenges and objectives for a platform-neutral architecture while Section V presents the proposed architecture and its implementation considerations. In Section VI, we conclude the paper and discuss the follow-up work.

II. RELATED WORK

In this area, previous efforts have been focused mostly on particular issues like the virtualisation of physical sensors on Clouds, data privacy and security or the provision of efficient platforms for sensor data storage and processing. Catrein *et al.* [7] propose a Cloud design for user-controlled storage and processing of sensor data to ensure privacy. They identified the importance of using the Cloud for sensor data processing. However, their approach focuses on security-related issues such as data privacy and access control. Aoki *et al.* [8] present a Cloud architecture to enable fast response to real world applications in spite of the flood of sensor data. The authors used the strategy of reducing network latency to achieve this goal but they did not consider creating a generic interface for the diverse sensor data gathering.

Piyare *et al.* [9] propose an architecture for integrating Wireless Sensor Networks (WSN) into Cloud services for real time data collection. In their approach, WSN is considered an important paradigm for Internet of Things since they consist of smart sensing nodes with embedded Central Processing Units (CPU) and sensors for monitoring different environments. This work concentrated on connecting WSNs to Clouds and does not consider other sensor types. The authors in [3] [4] discuss particular approaches for gathering medical sensor data. Alamri *et al.* [10] present a survey on sensor-Cloud, architecture, applications and approaches. The survey analyses the current efforts and challenges in this area. It shows that many of the existing efforts are geared towards creating virtual sensors from physical ones on Clouds.

Salehi *et al.* [11] present Global Sensor Network (GSN), which is a middleware to interconnect diverse sensor network

technologies. This work focus on providing a mechanism for easy integration of existing sensor networks. With this approach, the management of sensor networks are simplified. For example, changing or updating components within a sensor network does not interfere or hinder communications with other sensor networks. However, it does not consider the gathering of single sensor data. Other efforts like the Sensor Observation Service (SOS) [12] provide standards in accordance with the Open Geospatial Consortium (OGC) for discovery and retrieval of real-time data from diverse sensors in the context of geospatial data processing and sharing.

Thus, to the best of our knowledge, none of the existing work provide a generic Cloud-based monitoring platform for gathering sensor data. In the next section, we present our analysis of the state of the art.

III. STATE-OF-THE-ART ANALYSIS

This section details the survey of the existing tools and platforms for monitoring, gathering and processing sensor data. It highlights their characteristic features and practical use cases.

A. IBM Mote Runner

Mote Runner is IBM's infrastructure for WSNs [13]. It is based on a virtual machine targeted to resource-constrained hardware environments and consists of two parts: a run time for mote-class hardware, such as Libelium Waspote or MEMSIC Iris motes, and a development environment for WSN applications.

At its core, Mote Runner is designed to run on very small, standard, embedded controllers, including low-power 8-bit processors, thereby reducing both initial investments as well as post-deployment and maintenance costs. It has been shown to be light in energy consumption [14]. It provides a high-level, language-friendly, resource-efficient and high-performance Virtual Machine (VM) that shields portable applications from hardware specifics. In addition, Mote Runner allows programmers to use object-oriented programming languages such as Java and a development environment based on Eclipse to develop portable WSN applications that may be dynamically distributed. Its features include [15]:

- Low-power,
- Support for harvesting solar power as a source of energy,
- Wirelessly connected embedded systems,
- Provides Software Development Kits (SDK) for developers,
- Supports multiple high-level languages such as Java and C#,
- Runs on 8 bit micro-controllers with as little as 4 KB of Random Access Memory (RAM) and 32 KB of flash memory,
- Leverages integrated development environments such as Eclipse, Visual Studio and MonoDevelop.

This tool focuses on WSN related applications and mote actuators. However, little or no information was available about its support for wired or other sensor types. The details about the used data interchange formats are not publicly available.

B. SensorCloud Platform

MicroStrains SensorCloud offers a sensor data storage, visualisation and remote management platform that leverages Cloud computing technologies to provide data scalability and rapid visualisation [16]. It was initially designed to support long-term deployments of MicroStrain wireless sensors. However, it now supports any web-connected third party devices, sensors, or sensor networks through a simple OpenData Application Programming Interface (API). It aims to provide virtually unlimited storage for the sensor data [17]. Its features include:

- Presumably unlimited data storage with triple-redundant reliability,
- A time series visualisation and graphing tool,
- A MathEngine analytic tool facilitating quick user application development using their data on the Cloud,
- Provides Short Message Service (SMS) and Electronic mail (E-mail) alerting capabilities,
- Provides OpenData API and Representational State Transfer (REST) API for data transport,
- Supports only eXternal Data Representation (XDR) and Comma Separated Value (CSV) data interchange formats.

Some of its usage scenarios include structural health monitoring and condition based monitoring of high value assets where commonly available data tools are often not capable in terms of accessibility, data scalability, programmability, or performance. All communication with the SensorCloud is performed over Secure HyperText Transfer Protocol (HTTPS), which is secure. It restricts however, other forms of interactions such as low-level communication thereby forcing the use of HyperText Transfer Protocol (HTTP)-capable middleware to connect the devices and the platform. Furthermore, XDR and CSV are the only data interchange format types currently supported.

C. Ostia Portus Platform

Ostia Portus is designed to mediate between multiple vendor technologies where each vendor has particular platforms, databases and programming languages. It achieves this by taking the isolated data sets from individual systems and packaging them into a standard service [18]. This platform connects a variety of devices including sensors, networks and platforms. Its features include:

- Support for relational databases,
- Uses Secure Socket Layer (SSL) over HTTP to achieve security,
- Supports Simple Object Access Protocol (SOAP), REST, Publication/Subscription (PUB/SUB) protocols,
- Supports Java Message Service (JMS), RabbitMQ, Transmission Control Protocol (TCP)/Internet Protocol (IP), Message Queue (MQ) transport protocols,
- Easy installation and use.

Portus is built using open technologies and exploits open standards in accessing organisations data and presenting them using business defined views. Its core components are: (i)

server written in C and C++ and hosted by Apache; (ii) control centre for administration that is written in Java and built to run with the Eclipse Framework and (iii) front-end providing web services.

D. TempoDB Platform

TempoDB is a database as a service aiming to store and analyse time series data from sensors, smart meters, servers and automotive telematics. It is a commercial tool consisting of the following features [19]:

- Simple REST API for data storage and retrieval,
- Allows data storage at full resolution (no downsampling),
- Guarantees data availability with its three times data replication,
- Offers SSL encryption for all data transfer,
- API clients available in multiple programming languages like Java, .Net, Python, Ruby,
- Supports Internet of things.

The primary aim of TempoDB is the management of time series data sets with timestamps in ISO8601 format [20]. In querying data from TempoDB storage using a client API, the returned data is formatted only as JavaScript Object Notation (JSON). Other data interchange formats are currently not available.

This platform is currently expanded and renamed into TempoIQ [21]. It now offers flexible sensor data monitoring and alerting mechanism. The alerts are based on thresholds and informs the user about the status of its applications that are using the sensor data. It offers also analytic tools to support user applications.

E. FreshTemp Temperature Monitor

FreshTemp is a Cloud-based monitoring system for perishable goods [5]. It automates temperature collection during production, transportation and storage of any perishable product by providing the capability of integrating with the different temperature sensors monitoring such products. It offers real-time data logs, configurable alerts and online Dashboard. It is a commercial tool aiming to provide solutions for food services, transportation, health care and industrial usages. Its core features include:

- Wireless temperature monitoring,
- Bluetooth food probes,
- Bluetooth data loggers,
- Real time data logs,
- Alerting mechanism with SMS E-mail and phone,
- Online dashboard.

This tool focuses solely on temperature sensors, which limits its usability for managing and controlling other sensor device types. Furthermore, it does not offer any programming API for application development or any means of accessing the data stored on the platform by developers.

F. SensaTrack Monitor

SensaTrack is a multi-platform independent monitoring service that is ideal for Machine-to-Machine (M2M) sensor monitoring. It is developed by Cannon Water Technology Inc. and released in November 2012 [6]. The SensaTrack Cloud-based monitoring software solution allows users to visualize sensor data from any web enabled device. It is designed for enterprises with distributed assets like chemical storage facilities, bulk storage tanks, bins and silos. Its features include:

- Ability to quickly scan multiple locations,
- Secure data servers,
- Wireless data gateways,
- Track trends and find problems early,
- Easy installation by non-technical personnel.

SensaTrack uses wireless communications equipment based on Zigbee protocols developed by the Digi Corporation. It also supports hybrid networks of wired and wireless sensors. This tool has no support for application developments and does not provide an API for external access.

G. Bluwired S-Cloud Platform

Bluwired S-Cloud provides a platform for sensor data exploration, interaction and analysis [22]. It facilitates the management of sensor and device data from any web enabled location in the world, and to deploy data processing and analysis applications that rely on gathered data on the Cloud. Its features include:

- A platform for sensor data exploration, interaction and analysis,
- Management of wireless sensor data from any web enabled location in the world,
- Support for user development and deployment of data processing and analytic applications using their data on the cloud,
- Facilitates real-time data storage and retrieval,
- Provides alerting mechanism for abnormal events,
- Provides visual management interface.

Bluwired S-Cloud promises to offer reliable storage, tracking and analysis of sensor data that comes from monitoring and control solutions for most applications, including: Factory Automation, Process Control, Agriculture and Irrigation monitoring, Patient Monitoring Systems, Oil and Gas. It is a commercial tool and does not offer an open source API for accessing the platform.

H. Xively Platform

Xively is a Cloud Service platform that harnesses the power of IoT to quickly and easily transform connected product vision into market reality. It was formerly known as "pachube" and later as "COSM" [23]. Xively is currently a division of LogMeln Inc. and strives to provide business solutions through the IoT. It offers a platform that connects devices and products with applications to provide real-time control, management and storage. Its features include:

- Secure real-time messaging,
- Time series data storage,

- Selective data sharing,
- Provides easy connection to external Cloud services like Twitter and Facebook,
- Encryption with Transport Layer Security (TLS) and SSL,
- Real-time message bus based on Message Queue Telemetry Transport (MQTT).

An example use case for Xively is the “Park-A-Lot” project [24], which is designed to support an automated parking management system. Although it is a commercial product, but, it provides open APIs and libraries for easy usage by developers to create smart applications and interacts with the platform. Xively provides effective means to reason about devices and actuators at high level but fails to provide detailed context information within which all these devices are being placed, especially when it comes to small scale setups such as individual houses.

I. Nimbits Platform

Nimbits is a platform as a service (PaaS) for developing software and hardware solutions that seamlessly connect to the Cloud and each other. It has the ability to record and share sensor data on the Cloud [25]. Within Nimbits, sensor data are stored as data points using textual, JSON or eXtensible Markup Language (XML) formats. It provides REST web services for logging and retrieving time and geo stamped data (such as a reading from a temperature sensor). Nimbits servers can run on both powerful Cloud platforms like Google App Engine and on the smallest Raspberry Pi device. Its graphic user interface is tree structured having parent and child structures, which allow user content to be organised according to a parent-child structure and could be dragged and dropped as desired [26]. Its features include:

- Ability for recording and sharing data,
- Data storage as data points,
- Easy connection of data to analytic tools,
- Graphic user interface for visualisation,
- Ability to generate alerts based on defined thresholds or events.

Nimbits is an open source platform for the Internet of things. It is freely available and provides libraries, APIs and documentations for different programming languages.

J. ThingSpeak Platform

ThingSpeak is an open source Internet of things platform that provides API to store and retrieve data from things using HTTP over the Internet or via a Local Area Network [25]. It has the ability to facilitate the creation of sensor logging applications, location tracking applications, or a social network of things providing status updates.

In excess to its ability to store and retrieve numeric data and alphanumeric data, its API allows for numeric data processing such as time scaling, averaging, median, summing and rounding. ThingSpeak is organised in Channels, which is where a user application can store and retrieve data. Each Channel supports data entries of up to 8 data fields. The channel feeds support JSON, XML, and CSV data formats for integration into applications. Its features include [27]:

- Open API for developers,
- Real-time data collection,
- Geolocation data gathering,
- Data processing and analytic tools,
- Data visualisations on web and mobile devices,
- Device status messages and event alerting,
- Supports diverse programming languages like Java, JavaScript, .Net, Ruby,
- Allows easy plugins integrations.

ThingSpeak offers also a hosted service that is different to the open source version. Open.Sen.se [9] is another Internet of Things tool that is very similar in characteristics and features to ThingSpeak.

K. Microsoft Azure Intelligent System Service

The Microsoft Azure Intelligent Systems Service aims to securely connect, manage and capture machine-generated data from industry devices, sensors and other line-of-business (LoB) assets across a range of operating system platforms. The intelligent service represents the efforts of Microsoft to address the challenges of IoT and to help businesses utilise its potentials.

This tool promises to offer enterprises the ability to extend their Microsoft Azure Cloud across connected devices and sensors in order to capture vital data, analyze them with familiar Microsoft tools like HD Insight and Power BI for Office 365 in order to facilitate taking quick and appropriate actions that drive impact. Its features include [28]:

- Secure connection and management of devices and data,
- Support for real-time control of heterogeneous environments and accelerated implementations,
- Supports efficient capture, store, join, analyse, visualise and share data,
- Provide a trusted platform that can be extended easily for industrial specific requirements,
- Improve operations and unlock new business opportunities by harnessing machine-generated data from connected sensors and actuators.

The intelligent service is a fully commercial software as the other Microsoft services however, it is not yet production ready. A limited preview version was released in April 2014.

L. Paho Platform

The Paho project aims to provide scalable open-source implementations of open and standard messaging protocols to facilitate new, existing, and emerging applications to enable machine-to-machine and Internet of things usage scenarios. It is a part of the Eclipse foundation and its features include:

- Enables levels decoupling between devices and applications,
- Encourages the growth of scalable web and enterprise middleware and applications,
- Supports resource constrained embedded platforms,
- Based on MQTT,

- Provide MQTT client implementations in Java, Python, C, C++,
- Provide open libraries, API and client implementations,
- Enables integration of wide range of middleware, programming languages and messaging models.

Paho strives to be a software for constrained networks, devices with limited processing resources and embedded platforms [29].

M. Platform Summary and Comparison

This section presents a summary of the identified platform features in a table, which enables a quick comparison. The analysis offers an opportunity to identify trends and the gaps in technological advancements in this area. A goal of this analysis is to organise our thoughts and to see what is available in order to design a conceptual architecture.

The described and analysed platforms represent the current efforts towards addressing the challenges of Cloud-based sensor monitoring and data gathering, and their real-time analysis in order to facilitate informed decision making and smart applications. This information is summarised in Table I.

As shown in Table I, the analysed platforms exhibit some similar characteristics in terms of data storage, support for web technologies and availability of a REST API. The commercial platforms are closed source, thus, details regarding programming languages, API, data interchange formats were not accessible. It can be observed that the platforms have poor support for resource constrained deployments, energy efficiency, communication protocol implementation, software development environments and data interchange formats. These issues represent the challenges facing the existing platforms.

The commercial tools promise many features and functionalities, but details of their implementations are completely hidden from the public, which slows down technical know-how establishment and thereby makes it difficult for developers to leverage such functionalities when creating applications. This problem obstructs market growth and poses challenges to technology adoption. Furthermore, it hinders the realisation of portable solutions facilitating interoperability among different platforms. This shows the need for generic solutions with open implementations. Openness is a means to mitigate this problem. Providing open source APIs and interfaces enables easy creation of interoperable software and assures adherence to defined standards.

Another interesting point is that most of the open source tools are developed in the context of Internet of Things. This means that they could be easily used in designing generic solutions. We therefore view the open source tools as providing a good basis for the development of open solutions supporting standardised interfaces and data interchange formats.

According to Table I, most of the open source analysed tools are using textual data interchange formats, which have the advantage of being human readable and easy to understand. However, the serialised data in those formats are not very compact in size for efficient transportation without consuming large amounts of bandwidth. This identifies the need for integrating standardised binary data interchange formats that have the ability to achieve compact serialisation of data.

In general, the commercial solutions tend to surface many features. But, they are mainly closed source and implement proprietary technologies instead of the standardised ones. Most of them also do not provide open-source API for the general public, which could enable quick adoption and development of generic solutions. On the side of the open source solutions, they provide basic features and are not quite advanced. But they offer basis for further developments and possible realisation of generic solutions.

IV. ISSUES AND OBJECTIVES

In this section, we discuss the challenges informed from the state-of-the-art analysis and present our objectives for the proposed architecture.

A. Challenges and API Requirements

As can be observed in Table I, none of the analysed platform provides a complete set of features representing a generic and open solution. The implementation of such generic platform is complex and difficult. Many challenges exist in this spectrum especially regarding the heterogeneity of the targeted hardware and software components. At the lower levels, there is a plethora of sensor devices out there, which gather data in different formats. To design a generic interface for these data types requires a comprehensive study of the existing data types and how they could be aggregated or adapted/converted to a standard format.

At the higher levels, another challenging factor is the security and data access control on the Cloud platform. The authentication of a gateway server before allowing the forwarding of the sensor data to the Cloud platform is not enough. Many customers worry more about the privacy of their data and what the Cloud providers might do with them on the Cloud. To address these issues, different levels of security assurances are required for securing the data on the Cloud and controlling their usage.

Furthermore, the communication mechanism for transferring the sensor data and the actuator control information is challenging as shown in Table I. Such a mechanism must be simple, pluggable and reliable in order to ensure, robust and reliable communications.

B. Objectives

Currently, the ability to gather data from different sensor devices and feed them into Cloud platforms in a unified manner is lacking. This hinders the availability of raw data on computational capable platforms that can quickly and efficiently analyse the data to derive knowledge, which could support numerous usage scenarios such as real-time critical applications in health and medicine. Therefore, the goals of the generic Cloud-based sensor monitoring and data processing platform are to provide standard and open mechanism for collecting diverse remote sensor data and processing them irrespective of the sensor devices or usage platform. Furthermore, we aim to achieve:

- 1) Easy setup and portability,
- 2) Platform-neutral data formatting and serialisation,
- 3) Interoperability among heterogeneous Cloud platforms and
- 4) Simple configurations and ease of usage.

In the next section, we describe our proposed architecture.

TABLE I. SURVEYED PLATFORM FEATURES

Capability/ Features	IBM Mote Runner	SensorCloud	Portus	TempoDB	FreshTemp	SensaTrack	Bluwired S-Cloud	Xively	Nimbits	ThingSpeak	MS Intelligent Service	Paho
API/Library	SDK for implementations	OpenData API, REST API	No	REST API, client API	No	No	No	RESTful API, client libraries	REST API, libraries	Open API and libraries	No	Open client libraries
Data Formats	N/A	XDR, CSV	JSON, XML, IDoc	JSON	N/A	N/A	N/A	JSON, XML, CSV	Textual, JSON, XML	JSON, XML, CSV	N/A	N/A
Programming Language	Java, C#	Python, Java, C#, C++	C, C++, Java, PHP	Java, .Net, Ruby, Python	N/A	N/A	N/A	Objective C, C, Java, JavaScript, Ruby	Java, JavaScript	Java, JavaScript, Python, Ruby, .Net, node.js	.Net	Python, Java, JavaScript, C, C++
Open Source / Commercial	Commercial with open SDK	Commercial solution	Commercial Solution	Commercial with open source client	Commercial Solution	Commercial Solution	Commercial Solution	Commercial with Open API libraries	Open source solution	Open source with hosted version	Commercial solution	Open source solution
Visualisation	No	Yes and graphing tool	No, but with external web clients	Yes	Online dashboard	Web enabled	Visual management interface	Management console	Graphic user interface	Web capable devices	Yes	N/A
Analytic Tool	No	MathEngine	No	Yes, for time series data	No	No	Blu Automation Studio	Yes	No	Yes	HD Insight, Power BI	No
Messaging Protocol	N/A	No	RabbitMQ, JMS, TCP/IP, IBM MQ	No	No	No	No	MQTT	N/A	N/A	N/A	MQTT
Notification / Alert	No	SMS and email	No	No	SMS, phone email	Text message, email	Custom messages	Yes	Yes	Yes	N/A	N/A
Energy Efficiency	Low power, harvest solar power	No	No	No	No	No	No	No	No	No	No	N/A
Connection Type	wireless	Http	Http	Http	Wireless	Wireless gateway, Zigbee, Wired	Wireless	N/A	Http	Http, wireless, Zigbee	N/A	Http
Platform Resource Requirement	Resource constrained devices, embedded controllers	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Resource constrained devices, embedded platforms
Software Development Tool	Eclipse, Visual Studio, MonoDevelop	N/A	Eclipse	N/A	No	No	No	Developer workbench	Arduino, Java compatible tools	Arduino, Java compatible tools	N/A	Eclipse
Security	N/A	Https, SSL	SSL over http	SSL encryption	N/A	Secure servers	N/A	Encryption with TLS and SSL	Keys, OAuth	Write API Keys	Enterprise-grade security	Authentication, SSL
Database Type	N/A	N/A	Relational databases	Relational databases	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Storage Space	Yes, limited	Seemingly unlimited with triple redundant	Yes	Yes, with 3 times replications	Yes	Yes	Yes	Yes, time series archiving	Yes	Yes	Yes	Yes, limited

V. GENERIC CLOUD-BASED SENSOR MONITORING PLATFORM

The details of our proposed generic Cloud-based sensor monitoring platform is presented in this section. We discuss the conceptual architecture and its implementation considerations.

A. Conceptual Architecture

The conceptual architecture is designed to address the identified challenges posed by the existing platforms. Ease of use, portability, openness and interoperability are the key factors driving this design. This architecture is further motivated by what we think could be of value to customers with a system of distributed sensors and actuators.

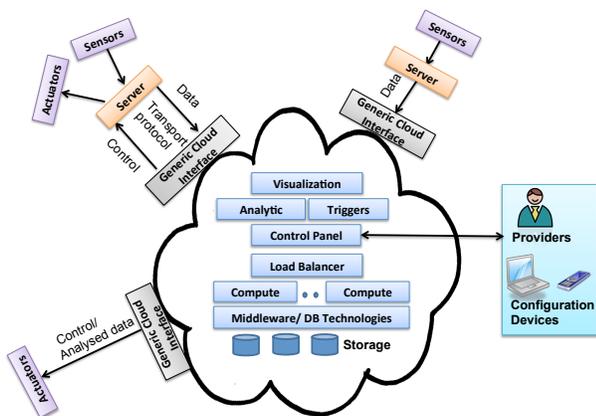


Figure 1. Conceptual Architecture

Figure 1 presents the proposed architecture. As shown on that figure, it is a distributed environment designed to

accommodate efficient sensor monitoring, data gathering and sending of control information based on analysed data to actuators. Different usage scenarios are considered in this conceptual architecture such as:

Dual interactions: In this case, a local or public server is attached to both sensors and actuators thereby making it capable of forwarding the captured sensor data and as well receiving the control information to be passed on to the actuators.

Data gathering: This scenario includes situations where the aim is to store and analyse the sensor captured data in the Cloud. This could also be to make the data public for applications to use or to use the analysed data results to derive control decisions for actuators.

Controller actuating: This represents the cases where the stored sensor data are directly being used by applications or to control an actuator.

As shown in Figure 1, the conceptual architecture consist of different components. However, we focus on the key ones due to space constraints.

Database Technologies

The storage of the sensor data on a Cloud platform requires efficient management for inputting and querying the data. This process is supported by appropriate database technologies. In our design, we aim to make this component pluggable so as to support diverse DB types depending on the usage platform. This enables the use of both SQL and NoSQL databases. The relational SQL database types would provide easy compatibility since most of the existing applications, according to our review and experiences, are based on this technology. The NoSQL databases are aimed for quick and

less complex scaling of the data storage system.

Generic Cloud Interface

The generic Cloud interface shown in Figure 2 is responsible for mediating between the sensor servers and the Cloud platform. It provides a simple interface and supports numerous formats for time series and alphanumeric data coming from the sensor servers. On the side of the Cloud platforms, it uses platform-neutral data interchange formats to achieve interoperability among heterogeneous Cloud environments.

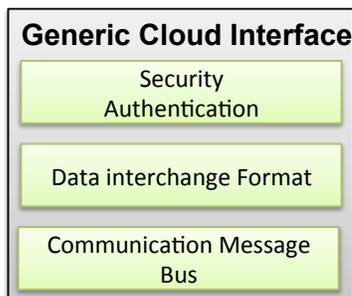


Figure 2. Generic Cloud Interface

As shown in Figure 2 the generic Cloud interface includes three essential components - i) security, ii) data interchange format and iii) communication message bus. These components provide the core functionalities of this interface and therefore deserve more explanations.

Security

This component ensures the privacy of the data interactions and guarantees their authenticity. It provides authentication mechanism to validate the access credentials of the sensor servers to access and transfer data to the Cloud platform. It also enforces data location constraints specified as a policy by the sensor data owners. This policy controls the location of the sensor data storage in the Cloud. It also informs the data owner whenever their data are being copied to some other location.

Data Interchange Format

The format of the sensor data impacts how they can be transported and analysed. This component plays the role of formatting the sensor data into a platform-neutral data interchange format enabling portability, interoperability and efficient transportation. The data interchange formats can be categorised into two groups: (i) self-describing data interchange formats such as XML and JSON and (ii) binary (schema-based) data interchange formats like MessagePack & Protocol Buffers. Based on our previous findings [30], the two format groups have their advantages and drawbacks. The self-describing data interchange format group has the strength of being human readable and easy to understand. But, from the transmission perspective, they contain redundant components, which affects the size of data transfers. The binary data interchange format group is not human readable. But, they are more efficient for transmission according to the performance results in [30]. For this generic interface, we propose a hybrid data interchange format combining the

strength of self-describing and binary data interchange formats.

Communication Message Bus

This component provides the communication mechanism for transferring the sensor data to the Cloud platform and also for sending control information to the actuators. It uses the data interchange formats for formatting and serialising the data. The message bus consist of three internal components: (1) Producer, (2) Messaging infrastructure and (3) Consumer

The producer feeds in the data to be transmitted by connecting to the external data producing devices. It integrates the data interchange formats to appropriately prepare the data for transmission. The messaging infrastructure provides the functions of a message broker by asynchronously delivering messages from the producer to the consumers (synchronisation decoupling). The producer does not need to know the nature or location of a consumer. It simply delivers its messages to the broker, which in turn routes them to the appropriate consumer (space decoupling). The broker therefore enables space, time and synchronisation decoupling [31]. This feature facilitates the necessarily loose relationship between a producer and a consumer, which is essential in distributed systems like Clouds. To realise the messaging infrastructure, we base our implementation on the well established Advanced Message Queuing Protocol (AMQP) [32]. The consumer connects the receiving end of the communication. It ensures that the data is appropriately deserialised for the end target platform.

B. Implementation Considerations

For a prototype implementation of this proposed architecture, we suppose the following considerations to be important.

Software artifacts: We envision the use of well established open source software components and standardised technologies as the basis for the implementation. As shown by our survey, there are some promising open source projects that could be considered. This would support the vision for an open and generic solution. Furthermore, it would avoid unnecessary re-implementations.

Component interactions: The generic interface is designed to interact with sensors, actuators and servers mediating between remote sensors and actuators in different fashions. In implementing these interactions, care has to be taken in choosing the data interchange format and communication mechanism since these two factors are very relevant in achieving a wide portability and interoperability of a solution.

Openness: Efforts should be made to make any implemented solution accessible to the general public. This would support quick adoptions and provide demos to the industry to encourage uptake among startup enterprises.

VI. CONCLUSION AND FUTURE WORK

This paper presented a survey of the current Cloud-based sensor monitoring and data gathering platforms. It analysed the state-of-the-art and identified open challenges. Both open source and commercial solutions were reviewed to provide a comprehensive survey. We used a table to compile all the features of the analysed platforms, which created a basis for quick comparisons and identification of trends.

According to our survey, the commercial solutions are generally closed source and implement proprietary technologies,

which hinders portability and interoperability of their solutions. Furthermore, there is poor support for messaging protocols, energy efficiency and resource constrained environments by the analysed platforms. Also, the platforms use mostly textual data interchange formats, which are not very compact in size for efficient transportation. To address these identified challenges, we proposed a generic Cloud-based sensor monitoring and data gathering platform. The goal is to provide an open solution implementing standardised technologies to promote fast adoptions.

In our next steps, we intend to implement a prototype of our proposed conceptual architecture as a proof-of-concept. We aim to provide it to the general public as an open source software to enable fast testing, deployment and adoption by the market. This contributes to our vision of facilitating interoperable data management in Clouds.

ACKNOWLEDGMENT

This work is supported by the Irish Centre for Cloud Computing and Commerce (IC4), an Irish National Technology Centre funded by Enterprise Ireland and the Irish Industrial Development Authority.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, 2010, pp. 2787 – 2805.

[2] B. Rao, P. Saluia, N. Sharma, A. Mittal, and S. Sharma, "Cloud computing for internet of things and sensing based applications," in *Sensing Technology (ICST), 2012 Sixth International Conference on*, Dec 2012, pp. 374–380.

[3] J.-C. Liu, K.-Y. Chuang, and Y.-L. Wen, "An efficient data gathering system for home medical treatment," in *Genetic and Evolutionary Computing (ICGEC), 2012 Sixth International Conference on*, Aug 2012, pp. 460–463.

[4] C. Rolim, F. Koch, C. Westphall, J. Werner, A. Fracalossi, and G. Salvador, "A cloud computing solution for patient's data collection in health care institutions," in *eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED '10. Second International Conference on*, Feb 2010, pp. 95–99.

[5] FreshTemp, "Cloud-based temperature monitoring tool," <https://freshtemp.com/> [retrieved: 23-09-2014].

[6] SensaTrack, "Online sensor monitoring platform," <http://www.sensatrack.com/index.html> [retrieved: 23-09-2014].

[7] D. Catrein, M. Henze, K. Wehrle, and R. Hummen, "A cloud design for user-controlled storage and processing of sensor data," in *Proceedings of the 2012 IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), ser. CLOUDCOM '12, 2012*, pp. 232–240.

[8] Aoki, H. et al., "Cloud architecture for tight interaction with the real world and deep sensor-data aggregation mechanism," in *Software, Telecommunications and Computer Networks (SoftCOM), 2010 International Conference on*, September 2010, pp. 280–284.

[9] Piyare, R. et al., "Integrating wireless sensor network into cloud services for real-time data collection," in *ICT Convergence (ICTC), 2013 International Conference on*, Oct 2013, pp. 752–756.

[10] A. Alamri, W. S. Ansari, M. M. Hassan, M. S. Hossain, A. Alelaiwi, and M. A. Hossain, "A survey on sensor-cloud: Architecture, applications, and approaches," *International Journal of Distributed Sensor Networks*, 2013, <http://dx.doi.org/10.1155/2013/917923>, In-press.

[11] A. Salehi and K. Aberer, "GSN, quick and simple sensor network deployment," *European conference on Wireless Sensor Networks (EWSN), Netherlands, January 2007*.

[12] Open Geospatial Consortium, "Sensor observation service," http://www.ogcnetwork.net/SOS_Intro [retrieved: 23-09-2014].

[13] IBM, "Ibm mote runner for wireless sensor platform," <http://www.zurich.ibm.com/moterunner/> [retrieved: 23-09-2014].

[14] Caracas, A. et al., "Energy-efficiency through micro-managing communication and optimizing sleep," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, June 2011, pp. 55–63.

[15] A. Caracas, T. Kramp, M. Baentsch, M. Oestreicher, T. Eirich, and I. Romanov, "Mote runner: A multi-language virtual machine for small embedded devices," in *Sensor Technologies and Applications, 2009. SENSORCOMM '09. Third International Conference on*, June 2009, pp. 117–125.

[16] MicroStrain, "Microstrain sensorcloud platform," <http://www.sensorcloud.com/> [retrieved: 23-09-2014].

[17] V. K. Sehgal, A. Patrick, and L. Rajpoot, "A comparative study of cyber physical cloud, cloud of sensors and internet of things: Their ideology, similarities and differences," in *Advance Computing Conference (IACC), 2014 IEEE International*. IEEE, 2014, pp. 708–716.

[18] Ostia, "Portus Platform," <http://www.ostiasolutions.com/index.php/product/platform-overview> [retrieved: 23-09-2014].

[19] TempoDB, "The Time Series Database Service," <https://tempo-db.com/> [retrieved: 15-08-2014].

[20] Zhang, Jia et al., "Sensor data as a service—a federated platform for mobile data-centric service development and sharing," in *Services Computing (SCC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 446–453.

[21] TempoIQ, "Sensor Analytics for the Measured World," <https://www.tempoi.com/?r=1> [retrieved: 23-09-2014].

[22] Bluwird, "BLUWIRED S-CLOUD Platform," <http://bluwired.com/Pages/Discover/BluwiredCloud.aspx> [retrieved: 23-09-2014].

[23] J. Hong and M. Baker, "Interaction platforms, energy conservation, behavior change research, and more," *IEEE Pervasive Computing*, vol. 12, no. 3, 2013, pp. 10–13.

[24] Yang, Kuo-pao et al., "Park-a-lot: An automated parking management system," *Computer Science and Information Technology*, vol. 1, no. 4, 2013, pp. 276–279.

[25] C. Doukas and I. Maglogiannis, "Bringing IoT and cloud computing towards pervasive healthcare," in *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*. IEEE, 2012, pp. 922–926.

[26] A. A. Chandra, Y. Lee, B. M. Kim, S. Y. Maeng, S. H. Park, and S. R. Lee, "Review on sensor cloud and its integration with arduino based sensor network," in *IT Convergence and Security (ICITCS), 2013 International Conference on*. IEEE, 2013, pp. 1–4.

[27] ThingSpeak, "ThingSpeak Platform," <https://thingspeak.com/> [retrieved: 23-09-2014].

[28] Microsoft, "Microsoft azure intelligent system service," <http://www.microsoft.com/windowseembedded/en-us/intelligent-systems-service.aspx> [retrieved: 23-09-2014].

[29] M. Prihodko, "Energy consumption in location sharing protocols for android applications," in *The Institute of Technology, Linköping University, 2012, Master Thesis*.

[30] V. C. Emeakaroha, P. Healy, K. Fatema, and J. P. Morrison, "Analysis of data interchange formats for interoperable and efficient data communication in clouds," in *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, ser. UCC '13, 2013*, pp. 393–398.

[31] N.-L. Tran, S. Skhiri, and E. Zimanyi, "EQS: An elastic and scalable message queue for the cloud," in *2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), 2011*, pp. 391–398.

[32] S. Vinoski, "Advanced message queuing protocol," *IEEE Internet Computing*, vol. 10, no. 6, 2006, pp. 87–89.

Towards Tactical Military Software Defined Radio

Major Tapio Saarelainen, PhD, IARIA Fellow

Research and Development Division

Army Academy

Lappeenranta, Finland

tapio.saarelainen@mil.fi

Abstract—This paper presents an idea-phase introduction of a tactical level communication system (battalion and below), which enables utilizing different waveforms and frequencies when communicating by using Software Defined Radio. The discussed idea-stage solution relies on using Unmanned Aerial Vehicles as hub-stations in order to ensure secure communication and reliable data transmission as regards wide bandwidth transmission to base-stations. The need for timely and accurate analyzing of the increasing amount of Situational Awareness and Common Operational Picture –related data collected keeps looming large in the battlespace. Similarly, the type and amount of different waveforms and frequencies also keep increasing. Software Defined Radio with its Graphic User Interface application, as discussed in the Results section, may offer one way of freeing the hands of a warrior to handle his or her firearm instead of a myriad of communication devices while in combat. This paper briefly looks at communication enabled by using swarms of Unmanned Aerial Vehicle and Self Organizing Networks in a military context and provisionally examines what testing and creating such a system would require and does so only at an early idea phase of a concept development process.

Keywords- *Unmanned Aerial Vehicle, Self-Organizing Networks, Software Defined Radio/Graphic User Interface.*

I. INTRODUCTION

This paper presents an idea-phase introduction of a tactical level (battalion and below) communication system to be used by a tactical end-user performing in a battlespace. As this paper's contents represent an early idea stage of concept development, the system drafted and its features described have neither been operationalized nor field-tested. When creating any new functional system, the first step of the development process concerns outlining an idea of what a functioning system necessarily needs to comprise. This idea phase of a concept development process turns into a fully-fledged concept with operationalized features to be tested once the end-user devices discussed in this paper have first been brought into being. Testing the devices described in an environment similar to what the paper outlines requires resources, i.e., political decisions allowing funding, personnel, and time. Any testing in lab conditions becomes impossible as no combat settings can be neither operationalized nor modelled in laboratories.

The key issue in modern warfare continues to be communication. Without communication there are neither coordinated operations nor success. High Data Rates (HDR) are needed for the type of data necessary in Battle Management Systems (BMS) and in Command Posts (CP) where operations are commanded and controlled and orders issued onwards for tasks to be executed by lower echelons.

Military environment is challenging also from the perspective of communications. Hostile military environment possesses challenges of several types for the need to communicate. First of all, the communication environment, a battle zone, is hostile. An adversary party tries to deny the free use of the frequency spectrum. Similarly, attempts of jamming the adversary's communication devices are typical of military actions executed in different frequencies and waveforms in the battlespace. Secondly, the soldier operating in a hostile territory using Cognitive Radios (CR) needs to establish mutual contact by using CR equipment to forming an ad hoc network. The challenge to create a functioning network in this case is exacerbated by the likely lack of accurate knowledge of the usage patterns of a radio spectrum in the hostile territory. If we compare this situation with a civilian case, where the frequencies and platforms are known in advance, the challenge in the military case has to be solved somehow in order to create a functioning communication network.

This paper introduces the challenges of the Future Force Warrior from the perspective of a consumer of communication services. This issue is essential for both the research community and the relevant industry. Once problems are pointed out, the process of finding solutions is easier. This paper presents one solution for how to facilitate fighters' need for constant capability to communicate in the battlespace.

This paper examines tactical level military operating referring to commanded tasks being executed at the level of company and below. Both soldiers and commanders of any kind are dependent on radios to execute missions. A single soldier relies on radio communication in order to be commanded. This asks for a reliable communication tool and a robust ubiquitous network system that allows for precision and minimized collateral damage. A battlespace can be understood as an environment, where operations are being executed including land, sea, air, underwater and cyber

operational environments. Warriors, sensors and Unmanned Vehicles (UVs) of several types operate and communicate in the versatile, constantly changing battlespace. As indicated in [1], militaries are using sensors as part of their battlefield strategy. As mentioned in [1], the integration of data collecting capabilities is in an essential role in expanding the communication platform capabilities. The key issues involve collecting the data, analyzing these data, and forwarding the analyzed data reliably and in an intact form to the end-user, a Future Force Warrior (FFW). Networks have to be organized to cover the needs of the end-user at all levels, as explained in [2]. Issues such as Quality of Service (QoS) and Speed of Service (SoS) are seminal in tactical communications [2].

Militaries concentrate on sustaining and developing their capability to communicate in a battlespace. To optimize performance, a Future Force Warrior needs only one communication device, which can be Software Defined Radio (SDR) for the reasons listed. First, one communication device covers all the communication needs of an FFW instead of him or her needing to use several communication devices. The focus of the FFW has to be in fighting. This means that an FFW keeps his or hands around the weapon, monitors the threats in the battlespace and fights. Secondly, mobility- and action-critical matters from the perspective of an FFW, such as size, weight, and power and cost (SWAP-C), are relevant. Thirdly, one power source should cover the need of communication devices instead of several sources (i.e., power source for radio and Personal Digital Assistant). Lastly, one communication device will ease the communication burden of an FFW and he or she can focus on the main functions: to monitor the environment and fight to survive. Present military communications are based on combat net radios (CNR), dominated by the enhanced position location reporting system and the single channel ground and airborne radio systems. This means that the varying levels of tactical communication can comprise several actuators. The distances between the communicating elements can vary from only a few meters to tens of kilometers and more. Tactical communication utilizes unmanned vehicles, drones and satellites acting as hubs or relay stations. The term tactical refers to the operative capabilities of a given military force. For example, a maneuver, which is tactical for the U.S. Army with its special forces, can be an operative manoeuvre for an army smaller in size and its operative capability.

The basic problem in communication is that the High Data Rate gives shorter range in communication (e.g., $4 \times \text{rate} = \frac{1}{2} \text{ range}$). Therefore we have to solve this problem with different means than just increasing the data range with increased transmission power. Despite the system characteristics, the communication system for military use in lower echelons (i.e., companies and below) has to fulfil the requirements of operational security, coverage, connectivity and Low Probability of Detection (LPD) and Low Probability of Identification (LPI). Military operations are dependent on covert high-speed networks, which also represent functional requirement of modern infantry and special operations warfare [3].

This paper introduces a solution utilizing a swarm of Unmanned Aerial Vehicles (UAVs) and SDRs. The swarm of UAVs is seen as a platform for a communications system, in which the distances between the UAVs must be short to ensure the message throughput in a hostile communication environment. This aims at ensuring a reliable data exchange process and fulfilling the requirements of LPD and LPI. SDRs are used by FFWs performing at the tactical level and also embedded into each UAV to ensure a reliable data exchange process. The swarms of UAVs in this system are in a central role to ensure the message throughput in a case when one or several UAVs are destroyed. Once a UAV becomes incompetent to act as a relay-station, a neighboring UAV takes over its functions and, with the assistance of Self-Organizing Networks (SON), the routing of communication can be reorganized and ensured.

This paper is organized as follows: Section II discusses Military Communication Environment, Section III focuses on the Challenges of FFW, Section IV introduces SDR, Section V introduces Universal Software Radio Peripheral (USRP). Section VI concentrates on explaining the idea of Cognitive Radio (CR). Section VII discusses the significance of a Graphic User Interface, and Section VIII introduces the new system. Section IX examines the strengths and weaknesses of the introduced system and Section X concludes the paper with Section XI describing the demands for future work.

II. MILITARY COMMUNICATION ENVIRONMENT

Military communication environment differs from its civilian counterpart. Any civilian communication environment tends to be non-hostile and its features thoroughly known in that the transmission distances, frequencies and waveforms used are common knowledge. In contrast, the military communication environments are part of a battlespace and abound in uncertainties in connectivity and latency may vary uncontrollably due to incessant hostile electronic warfare attempts. Communication break-down in a battlespace typically results in compromising someone's life. A given battlespace comprehends also the communication environment, in which war is waged. Military frequencies tend to be mandatory and always commanded from higher echelons in order to control the electromagnetic spectrum most effectively from the execution perspective of own military operations.

Military troops equipped with varying end-user devices transform the battlespace of 21st century into network-centric warfare with Network Centric Operations in a central role. SDRs will provide a flexible tool suited for the changing military environments in that they allow versatile communication in the battlespace [4]. In a constantly changing battlespace the commanded troops can be mobile or static. Often the communication tools, end-user devices, are handheld and lightweight. A ground level military performer, soldier, has to be able to communicate also via satellites in order to contact higher echelons, for example, when executing special operations in rocky terrain. The size of the used system, together with its weight, power and cost

(SWAP-C) become vital from the perspective of the system user and provider.

For the end-user, to be able to cover multiple battlespace scenarios, the simultaneous requirement of communication requirements such as voice, video and data together with the capabilities and megabit bandwidths set design challenges. Moreover, to sustain secure communication by means of the end-user devices, new military waveforms have been designed to fulfil the requirements of the end-user, a soldier. An example of such waveforms suitable for an SDR-based system implementation is WiMAX 802.16e, which has been modified to operate in the military frequency range of the NATO UHF band of 225 – 400 MHz [5].

A future military SDR-platform should support multiple radio frequency frontends. Depending on the available and sufficient frequency bands, different frontends could be installed. The flexible use of different frontends and waveforms enable finding a suitable system configuration for all the planned operational scenarios. The next generation SDR-based platform should enable at least the following benefits for tactical networking: First, mobility support for mobile ad hoc network (MANET). Second, sufficient communication capacity must be guaranteed at the tactical level, this is a minimum throughput of 1 Mbps to support mobile user. Third, from a perspective of life cycle management, a SDR platform must be independent from the waveforms and frequencies used. Fourth, communication flexibility has to be ensured with radio frequencies in SDR frontends and with used waveforms. Fifth, interoperability with national and coalition waveforms has to be granted [5].

In a civilian communication environment an end-user can benefit from reliable and fast communication, high throughput of messages, issues of low latency, the constant capability to communicate, adequate bandwidth, good Quality of Service (QoS), and Speed of Service (SoS). Civilian communication systems offer the possibility to benefit from constant power supply or the capability to recharge the battery of the used communication device when necessary. The communication process usually suffers from only slight if any hostile interference or jamming.

Communication systems utilized in a military environment can confront all types of interferences. These include jamming and all means electronic warfare with a constant threat of becoming annihilated by the adversary if a communication tool has been detected, pinpointed, and placed in the targeting process to be destroyed. Table I lists the differences between civilian and military communication environments.

TABLE I. DIFFERENCES BETWEEN CIVILIAN AND MILITARY COMMUNICATION ENVIRONMENT

Characteristics of Communication	Military communication environment	Civilian communication environment
Free use of spectrum	Restricted	More possibilities
Latency	Varying, sometimes high	Typically low
Energy	Limited, hard to recharge	Possibility to recharge fast
Hostility	High	Low
Jamming	Possible	Low
Adequate bandwidth	Limited, altering	Typically high
Limitations in use	Often restricted	No limitations

As Table I indicates, in a military communication environment the characteristics of communication involve restrictions and constant uncertainty due to hostility in the battlespace.

III. CHALLENGES OF FFW

The following overview lists six challenges which have been identified within the military community. The writer has encountered these challenges while conducting research on related issues, such as a nationwide Company Attack Study performed during 2004 – 2007 in Finland. The first challenge is related to the main task of a fighter: the main task of an FFW is to fight in performing the given mission. This means he or she has to monitor the environment to stay alive and to be able to execute the commanded tasks. He or she engages the enemy with all the weaponry available. This also means that the FFW relies on connectivity and capability to communicate at all times. The constant connectivity poses the second challenge. Mission success requires the capability to transmit and receive data and commands. The constant communicating ability requires that a single warrior be capable of acting as an executor of an operation, or a military commander at some level. Connectivity remains the key. Some sort of a network must be available at all times. Location data and commands can be forwarded only by means of a functioning network and a reliable hand-held or soldier-mounted communication device.

The third challenge is linked to the usage of warrior platforms. A warrior platform consists of several subsystems and their control units. For example, the systems can be controlled via a wrist-worn user-interface presented here. All the communication controls can be easily and rapidly found from the wrist-held device which is embedded onto the arm and acts as a supporting hand when using a personal firearm. Contrary to a visor-embedded system, this wrist-held device

does not hamper viewing the environment with a constant data flow. Figure 1 features a wrist-held device.



Figure 1. A wrist-worn control system into which SDR can be embedded.

The fourth challenge involves the number of networks and data sources on which an FFW relies. If the communication network or operating unit malfunctions, FFWs get in trouble because they lack the necessary resources either because of the different frequency of waveform used or the network becoming out of coverage. This slows down a single FFW and usually harms the whole military operation.

The fifth challenge equals the access to different types of Battle Management Systems (BMSs). BMSs support the efficient utilization of military units at all levels. The access process into a BMS requires more bandwidth than using voice and text messaging when issuing commands. Figure 2 features BMS.

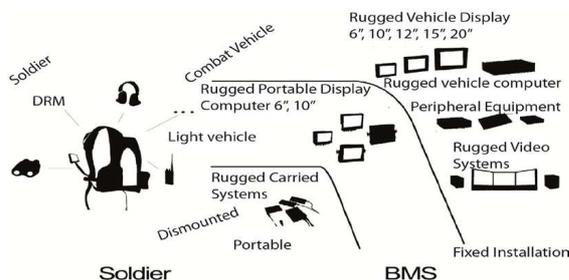


Figure 2. Soldier Systems linkage into Battle Management Systems.

This sets more bandwidth demands for SDRs used. The data may surface in varying waveforms and frequencies. In order to benefit from BMSs, the location and identification of friend or foe are relevant. These data are needed from the battlespace to ensure the effective use of different weapon systems.

Systems related to location data and identification systems keep improving. The command and control capabilities necessary for units serving at battalion level and below are provided by means of BMS. Present combat net radios can only be seen as beneficial tools in supporting geographical based situational awareness. When several operational BMSs are integrated, it is possible to maximize the amount quantitative and qualitative data for analysis purposes. This involves applying the concept of System of Systems (SOS) and utilizing robotic platforms. The sixth challenge is the type of a Graphic User Interface (GUI) of a

communication device, which can be used for various purposes. GUIs can be utilized in creating the Situational Awareness and Common Operational Picture as well as for location services and for weapon selection process. Although one of the challenges is the variety of used waveforms and the bandwidths, they fall outside of the scope of this study. This paper discusses a possible solution for the listed challenges.

IV. SOFTWARE DEFINED RADIO

SDR is a radio communication system in which components that have typically been implemented in hardware, for example, mixers, filters, amplifiers, modulators/demodulators, and detectors, are instead implemented by means of software on a personal computer or embedded system. Usually SDR can be programmed to support frequencies from 100 MHz to 6 GHz with using 130 nm Complementary Metal Oxide Semiconductor (CMOS) technology [6]. Depending on the system configuration, typically supported signal bandwidths can vary between 700 kHz and 40 MHz or 200 kHz and 40 MHz, depending on the CMOS used [6]. In other words, SDR is a wireless communications system where the traditional hardware is replaced by software modules [7]. While the concept of SDR is not new, the rapidly evolving capabilities of digital electronics enable executing many such processes which earlier used to be only theoretically possible.

Currently, most SDR related products and studies focus on analog communication and voice transmission. The SDR platform consists of Field Programmable Gate Array (FPGA)-based radio hardware and open source SDR software module [7].

The main features of SDR include: 1) radio spectrum sensing; 2) reconfigurable radio modules and 3) link for digital data communication. These features form an important basis to accomplish Cognitive Radio technologies.

The mobile devices can afford the high speed and complex computation owing to the advance in computing ability of the processor, such as Personal Digital Assistant (PDA), Smart Phone, or Ultra-Mobile PC (UMPC).

Most of these mobile devices equipped with Wi-Fi, WiMAX or other wireless modules enable end-users to access services anywhere. The traditional hardware radio system comprises a variety of analogy elements such as filters, converters, modulators and demodulators. The hardware is expensive and has low compatibility with other components.

The reason why SDR becomes increasingly popular is that it allows using SDR technology for realizing many applications relatively effortlessly in the integration of different components. The most used software architecture for SDR is the Software Communications Architecture (SCA), which is considered as the standard for military domain [4]. The novelty of SCA lies in the availability of SCA-based tools to allow designers to create component-based SDR-applications as assemblies of components and logical devices. When these types of systems are being created, the communication between the components and devices must be carefully orchestrated. In this process it is

possible to benefit from the use of Common Object Request Broker Architecture (CORBA) [4].

When discussing SDRs, security issues must be considered. When new Software (SW) is being loaded, the consequent threat of having unauthorized and potentially malicious SW installed on the platform becomes possible, if security precautions have not been taken, by, for example, adding a digital and verified signature in the code before the new software is being transmitted.

In order to successfully benefit from the products and performance of SDR, we have to focus only the performance produced via SDR. The added value from SDR can be seen via tactical communication requirements for the FFW operating in Battlespace which are: Situational Awareness (SA), Common Operational Picture (COP), Command and Control systems, identification friend or foe (cf. Figure 3. below), (IFF)/Blue Force Tracking, capability to co-operate with UAVs and Unmanned Ground Vehicles (UGVs) and robots, data from sensor to shooter, Voice, Navigation, messaging, Imaging, Video, Security.

One interesting possibility is to embed Radio Frequency Identification (RFID) system into SDR by using Quadrature Amplitude Modulation (QAM). The system has been explained in [8]. The identification friend or foe (IFF) process can be embedded as part of SDR functions. Figure features 3 an IFF process.

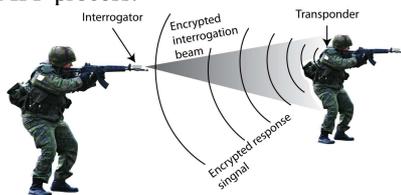


Figure 3. Identification process Friend or Foe (IFF) in progress.

By simply downloading a new program, a SDR is able to interoperate with different wireless protocols, incorporate new services, and upgrade to new standards. One solution is depicted below above while the process is introduced in Figure 4.

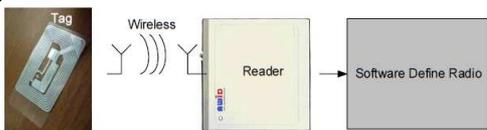


Figure 4. RFID system with SDR [8].

As Figure 4 indicate, by combining the Radio Frequency Identification tags and IFF –process it is possible to decrease the total mass of the gear a FFW carries. Similarly, this enables simultaneously decreasing the amount of transmission energy necessary for identification purposes. RFIDs can be sensitive to electromagnetic interrogation signal by nature and need little energy when responding once only a slight transmission signal focuses on them. Thereby the amount of response energy transmitted towards of the interrogator can be significantly small. This means Low Probability of Detection as regards the surveillance tools spread in the battlespace. Relevant for operational security

purposes, the interrogation process remains undetected and discreet.

V. UNIVERSAL SOFTWARE PERIPHERAL RADIO

In Universal Software Peripheral Radio, we examine only the functionality of USRP. In doing so, we notice it offers more performance than its predecessor SDR. The technology used in USRP is located in the hardware implemented in frontend for sending and receiving waveforms. USRP offers different frequencies, bandwidths and frequencies for specific purposes. The USRP can be fixed to respond to the end-users’ requirements by selecting appropriate motherboards for controlling the frequencies and waveforms [7].

USRP can be divided into two parts based on the transmission path. These are the transmitting signal path and receiving signal path. For example, on transmit signal path, users can define the setting parameters by software on personal computer such as radio protocols, modulation types, frequency of spectrum modulation.

Then the USRP receives the parameters, and FPGA executes Intermediate Frequency (IF) processing on Digital Up Converter (DUC) and Digital Down Converter (DDC). After the Intermediate Frequency process, users adjust the baseband to the frequency band selected before.

The last step on USRP motherboard is that digital to analog (DAC) converts the digital signal into analog signal. Finally, the analog signal is transmitted to the antenna through the interface side on the daughterboard, as illustrated in USRP block diagram in Figure 5.

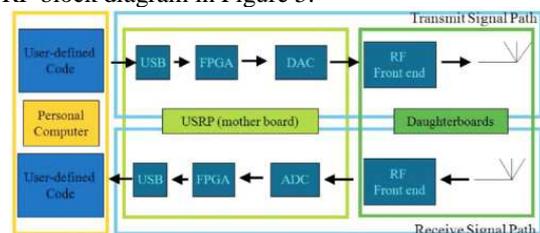


Figure 5. Composition of USRP [7].

As Figure 5 demonstrates, the composition of the introduced USRP system offers flexibility in using different waveforms and frequencies. The flexibility can be offered by different daughterboards which can be tailored to meet the requirements of different frequencies and waveforms. By changing and tuning the performances of daughterboards, the FFWs have improved communication devices as communication tools in a constantly altering battlespace and varying missions.

VI. COGNITIVE RADIO

When moving on towards the communication device suitable for an FFW, we have to take a quick glance at Cognitive Radio (CR). As widely known, Software defined Radio is a platform for Cognitive Radio [9]. Without going into the details of the technical structure or composition of CR, we focus on the listed and wanted end-products and functionalities of the CR from an FFW perspective.

Cognitive Radio capabilities and functionalities include the following features. First, Spectrum Awareness (SAw), which means being able to detect quickly and robustly the presence of incumbent (preemptive) users to avoid causing interference. Second, Dynamic Spectrum Access (DSA), which means CRs will access the spectrum on an opportunistic basis. Third, Dynamic Spectrum Sharing (DSS), which means CRs must be aware of other CRs' coexistence. Lastly, CRs are Spectrum Agile (SAg), which means that CRs should provide seamless operation over multiple channels. Also challenges related to adaptive coding, modulation and multi-access have to be solved as indicated in [9].

There is a long way towards CR, which meets the listed requirements. From the perspective of an end-user, a consumer, an FFW, it is essential to meet the requirements listed. This also applies as regards the scientific community and the industrial community. The necessary requirements have to be identified prior to being able to produce a wanted end-product.

One solution for designing the suitable architecture and configurations for the future force radio communication device could be Software Communications architecture. SCA has been created to assist in the development process of SDR communication systems. SCA allows for waveform application software to be more easily ported across radio platforms. At the moment publicly available specifications can be found for SCA 2.2.2 and 4.0, as well as for SCA Appendices and SCA APIs. As a matter of fact, it is possible in next generation products to take full advantage of the following: First, SCA 4.0, which is to empower more freedom to do the SDR implementation. Second, Programmable SDR chip sets. Third, offering more efficient SDR development tools and use of more efficient higher level modelling methods. Fourth, adding a new approach to waveform portability and full utilization of SDR work done in commercial domain. Lastly, designers are focusing on developing of sophisticated RF front-end technologies.

SOA -technology involves assisting processes performed in military operations. As indicated in [10], the SOA has been used to design and construct the CR systems. When an FFW can benefit from the possibilities offered by a successful adoption SOA, also in communication services, the result can be improved overall performance in military operations.

Figure 6 features how the data are gathered, processed, analyzed and then transmitted as commands to an FFW. If the data are correctly collected, analyzed and successfully transmitted to the performer, an FFW, the process of waging war can be improved and collateral damage minimized. Various battlespace sensors transmit data to a context-aware reasoning layer. In this layer, data are converted to context and an inference engine transmits the data to a ubiquitous main layer for analyzing purposes. The data are verified, analyzed and transmitted as information for the execution of the operation [11]. This process is depicted in Figure 6.

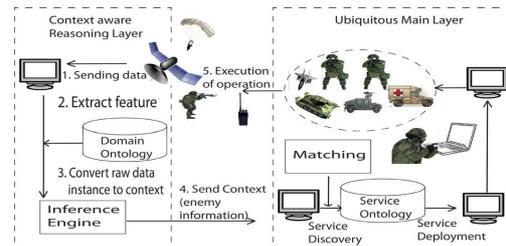


Figure 6. Increased FFW performance can be gained via successful data utilization and analyzing process [11].

Once the collected data have been analyzed, they can be forwarded to the military performers who need these data most. The transmission process has to be automated to ensure sustaining overall performance.

VII. GRAPHIC USER INTERFACE

A new type of communication device for an FFW has to fulfil the specific communication needs of an FFW. The communication system and the GUIs have to be defined to fulfil the needs. Figure 7 features the actuators affecting the system definition process.

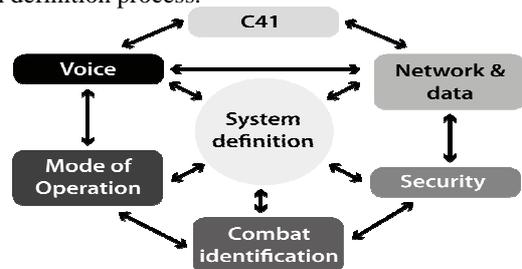


Figure 7. The idea of configuration of the future communication device from a perspective of a consumer, an FFW.

According to results presented in [12], the visualization of events can improve the human capability to accelerate the Military Decision Making Process (MDMP) by offering necessary information in required time and understandable form. From the perspective of a consumer, an important role is set for the type of a GUI. A functional GUI is a means to present collected data, a control panel to access networks and guide UVs, a tool for a weapon selection process, and, of course, a communication tool for the entities of higher and lower echelons. Figure 8 features one possible figure caption of a functional GUI.

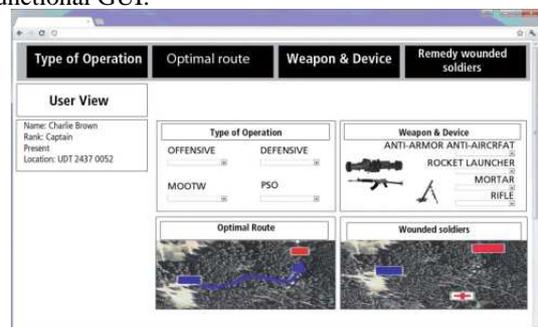


Figure 8. A view of a Graphic User Interface.

Apart from the mentioned facts, an FFW has to be able to access BMSs of various types and different databases with a new type of communication tool. Figure 9 features one type of BMS where an FFW can be constantly connected to optimize the performance while executing tasks.

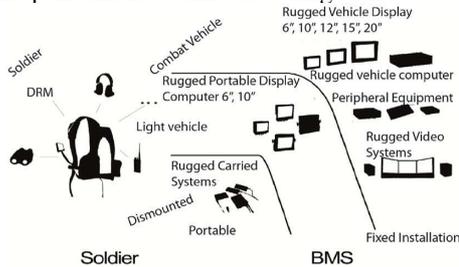


Figure 9. A view of a BMS.

Lastly, an FFW has to be able to control the systems embedded on his Battle Dress Uniform (BDU). Figure 10 below features the FFW's electronic skeleton and its functions [13].

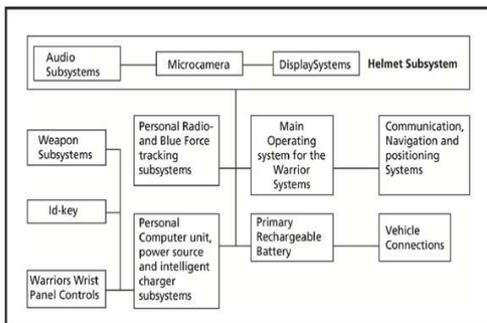


Figure 10. A composition of a FFW's electronic skeleton [13].

An FFW uses the control unit embedded into his BDU as an essential tool to control and monitor the functions of own gear. In Figure 11 below, the controlling system is wrist-worn.

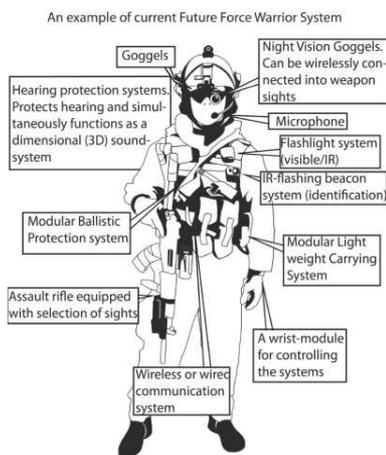


Figure 11. An example of an FFW as a platform with selected gear [14].

By embedding all the control units of the FFW's electronic gear into one wrist-worn controller, the number of controlling units can be decreased. This may increase the overall performance of the FFW as the FFW can find all the control units in one location instead of needing to separately control each different embedded system listed in Figure 11.

VIII. A NEW COMMUNICATION SYSTEM FOR THE FFW

As noted in [3], mobile ad-hoc networking of dismounted combatants is necessary as regards the future involving net-centric operations. The amount and variety of data transmitted in the battlespace keep increasing. Issues such as bandwidth, type of waveform, frequency and security are only a few of the issues that have to be accounted for. Low Probability of Detection and Low probability of Identification remain critical in covert operations, as mentioned in [3]. Single UAVs are utilized as tools tailored for Special Forces and a system relying on Advanced Encryption System (AES) encrypted network with a range of 3 kilometers [12].

A new communication system is possible to create if we utilize the capabilities of SDR, swarms of UAVs, SDR and Self-Organizing Networks (SON) implemented in 4G networks. It has to be highlighted that in this system, SDRs are implemented in FFW gear and inside an UAV. As noted, SON aims to configure and optimize the network automatically, in a manner that the interaction of human can be reduced and the capacity of the network can be increased.

The main functionality of SON includes the following: self-configuration, self-optimization and self-healing. SON is described as a part of 3GPP LTE and it is a key feature for effective and automatic operation and maintenance (O&M) of 4G networks. Besides that, SON maximizes overall performance of network and reduces the cost of installation and need of management by simplifying operation and maintenance through self-configuration, self-optimization and self-healing. SON also reduces the power consumption and results in reduced operational expenses and produces an environmentally friendly approach. Figure 12 features the SON as seen in [16].

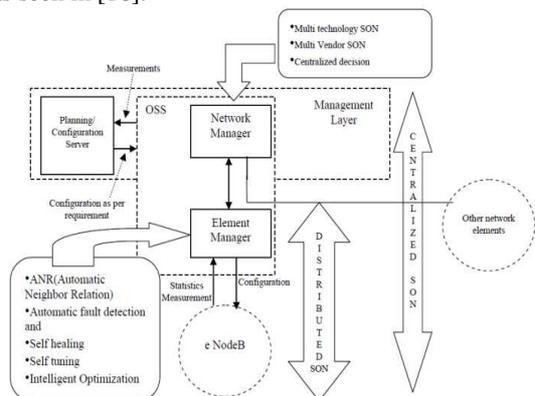


Figure 12. A composition of a SON implementation in 4G as seen in [16].

When a swarm of UAVs is utilized, the distances needed to communicate with an FFW-worn SDR must be minimal in order to ensure the message throughput in this system.

Different types of data can be transmitted from a soldier to a higher echelon via an UAV. Security issues remain essential when dealing with UAVs utilized in Network Centric Warfare at a tactical level. This means opting for low transmission power and thus minimizing the chances of the UAV becoming detected, targeted, and destroyed.

When transmission distances remain short between the ground (FFW) and aerial stations (UAV), the accrued and transmitted data can be better secured and requirements of LPD and LPI can be achieved. When SON utilizes all the SDRs, those embedded into the UAVs and those embedded into soldier worn systems, this data exchange can be executed successfully and thus ensure that the data remain intact and coherent. Figure 13 features the data-exchange process via a command post and UAVs with the assistance of embedded SDRs into the mentioned entities.

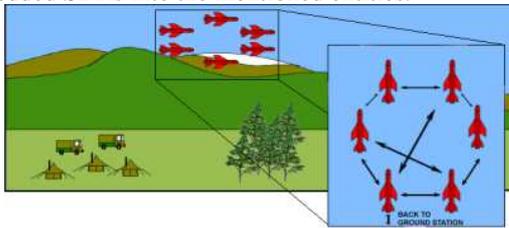


Figure 13. A data exchange process with the swarm of UAVs.

The swarms of UAVs will forward the data automatically via the network system created by UAVs. Demands of LPD and LPI can be fulfilled, because the transmitting energy used via the transmission protocols by means of UAVs remains low.

IX. DISCUSSION

A military environment, battlespace, differs from a civilian environment. In battlespace both constant stress and uncertainty continue to dominate. The fear of losing one's life prevails. An FFW has to monitor his or her environment when fulfilling the commanded mission and stay alive. The mental capacity of an FFW must be focused on the matters at hand. The lower the number of gadgets an FFW has to monitor, the longer his or her life with an increased possibility to continue performing.

Militaries aim at developing SDR into a communication tool for all the troops at the tactical level. The process of embedding a functional SDR as part of military troops' communication devices is still globally ongoing in militaries, with no existing, operationally fully functional end-user devices anywhere in combat use able to transmit large amounts of data in various waveforms and frequencies. Meeting the requirements of mobile users in a battlespace remains challenging. Issues related to SWAP-C have to be solved. One critical challenge related to military SDR use involves achieving sufficient computational capacity. This is a problem when processing wide-band high-bit rate waveforms consisting large amounts of data. In terms of SWAP-C, an FFW needs the selected communication end-user device to be reasonably tailored with optimally minimal total mass of a device, its batteries and recharging units. This means that Data Processing Units (DPUs) and Event Driven

Administrative and Control Components (EDACCs) have to be carefully selected and orchestrated to meet the operational requirements of the end-user [7].

The issues related to energy cannot be over emphasized. The energy requirement of a typical handheld device can be between a few hundred milliwatts [15] to few Watts [8]. The system specifications of the SDR are significant in defining the energy needed as well as the amount of data transmitted.

Challenges related to operational security are essential in reconfiguring the SDR-systems, especially as regards software. While loading a new program, new waveform or new hopping sequence, issues of transmission security during the different uploading processes have to be guaranteed. If this part becomes neglected, the SDR will not act as a useful tool in own hands in net-centric operations but rather becomes a novel tool to be exploited by the adversary.

A communication system that serves fighters' needs is creatable if we utilize the capabilities of SDR, swarms of UAVs, SDR and SON implemented in 4G networks and combine them as depicted in Figure 13. This means that an FFW's end-user device, SDR, is connected with a swarm of UAVs via SON. The swarms of UAVs form an own data communication system in which the data transmission distances between UAVs are short and operationally secure. This in turn will fulfill the requirements of LPD and LPI. The described delicate system introduced is a new one and based on ideas that can be executed by utilizing existing Commercially Off-the-Shelf (COTS) technology. The system is not yet bullet-proof and can malfunction for a number of reasons. Challenges related to creating the described system have to be solved to enable the function of different processes. The orchestration of the system can also fail because of intentional enemy action (jamming, a virus, a worm). The system needs to be equipped with an analyzing program, which indicates when the system functions properly before using the system. This asks for an easily replaceable and fault-tolerant system with inbuilt check-in routines. Otherwise, traditional methods in orchestrating services need to be adopted.

The introduced system offers an access to communication processes, which are created to support command and control systems. An FFW relies on communication services. To enhance SA and COP, it is essential to have user-friendly GUIs of some kind for presenting data. As noted, time remains a critical factor in tactical-level operations and the main function of an FFW is to fight, not to spend time browsing different databases in search for vital data.

Creating a new communication tool requires resources, such as personnel, time, money, troops and space to execute the use of the tailored device in pre-defined drills. The vendors and the end-users have to co-operate to create a functioning communication and control tool for the use of an FFW. In the development process, the use-cases and usage methods of SDRs have to be defined. This includes defining use-cases of operations, training-scenarios, types and timing of operations, training practices, data gathering during the exercises, and After Action Reviews (AARs) together with debriefing-sessions for system designers and troops after

implementing the training-scenarios. The system-creation process requires strict timing in a well-orchestrated series of field testing in which the system users and system developers have to attend the tests at the same time. The time required for field testing equals approximately a decade, the number of training drills necessary a hundred, and the number of military personnel committed to executing the drills a hundred. The complexity of the SDR system requires that a handsome number of designers and engineers from the vendor's attend the drills, ideally one on one. The estimated funding requirements equal at least a 100 M€. However, only average ballpark figures can be estimated as the actual realized costs and their approximations would by default value be labelled classified. The number of personnel and funding required described in this paper rely on the experience accumulated over twenty years on active duty as a field-test participant. The military personnel, designers and engineers have to be fully committed to this work in order to achieve results. Table II below lists the identified resources needed for the described study. In order to create a feasible testing system, an amount of work equal to producing a dissertation is required.

TABLE II. ESTIMATED RESOURCES TEQUIRED FOR FIELD-TESTING

Types of resources required	Vendor side	Military side
Money	30 M€	70 M€
Personnel	100	100
Time for planning	2 years	2 year
Time for testing the system	5 years	5 years
Time for evaluating the results and the system	2 years	2 years
Reserve time	1 year	1 year
Total resources	10 years and 30 M€	10 years and 70 M€

Table II features only a rough estimate. More precise data requires pre-planning for a period of twelve months. Different sources of funding, such as industrial and / or academic contributions of personnel and / or equipment, need to be estimated before any final estimation is doable.

X. RESULTS

The main result is an idea-phase introduction of a tactical level (battalion and below) communication system to be used by a tactical end-user performing in a battlespace. As this paper's contents represent an early idea stage of concept development, the system drafted and its features described can neither be operationalized nor field-tested.

An FFW performs in a battlespace filled with ubiquitous networks and communication systems. He or she has to cope

with actions involving humans and machines, such as databases and UVs. Equipped with a reliable communication tool, an FFW can perform tasks with improved speed and efficiency. Bespoke SDR can enhance the performance of an FFW by answering the defined challenges listed in Section III.

To sustain optimal performance, an FFW has to be able to use only one single device for command and control. This device can be an SDR with a GUI. This way there is only one single device, Software Defined Radio/Graphic User Interface (SDR/GUI), for an FFW to communicate and use controls with instead of being exposed to several communication tools. Compared to a stable civilian environment, a military environment equals a constantly altering battlespace. An FFW has to be able to monitor the events in the prevailing environment instead of needing to update the status and monitor his or her command and communication tool. An FFW has to have hands on a weapon and be ready to act when necessary. If unable to do so, the FFW will become incapacitated by a splinter or a bullet. It is essential to enable an FFW to receive and transmit data with the assistance of SDR/GUI, into which SDR can be embedded.

First, an improved SDR/GUI can act as a control-station for all of the digital systems. This allows an FFW to focus on his or her main task, to fight. He or she can monitor the prevailing environment and use his or her weaponry in a time-critical environment. Second, an FFW can concentrate on one device, SDR/GUI, instead of monitoring several screens and displays. He or she saves time and can focus on the task commanded. Third, the control units of his or her own warrior skeleton and communication controls can be found from one communication device, SDR/GUI. The fourth challenge can also be solved by adopting SDR/GUI which will take care of the various networks and waveforms and switch automatically to the free and appropriate channel to transmit or receive data.

The fifth challenge was a problem concerning the access to BMS via different communication tools. The problem is linked to the issues of bandwidth, frequency and waveform. The accessing process into BMS involves utilizing the performance provided by the SDR and SON. This means offering the frequencies, bandwidths and waveforms required from the fighters' perspective. The swarm of UAVs serves as a secure and replaceable communication gateway for the data exchange process. This allows for keeping the transmission power low and transmission ranges relatively short. The Sixth challenge was linked to the GUIs. To comprehend the prevailing operational situation and a holistic list of events at a tactical level, an informative presentation of SA and COP is in an essential role. The presented system together with improved SDR/GUI can be seen as a feasible solution to solve all the listed challenges. Figure 14 features a possible view of a functional SDR/GUI for the FFW operating at low level (company and below) tactical operations.



Figure 14. An example of tactical SDR/GUI.

Tactical SDR/GUIs can serve as command and control tools offered for the FFWs. SDR/GUIs can be utilized as tools for possibly enhancing the overall performance of the soldier. An FFW needs only a short period of time to take a look at the SDR/GUI and notice if something significant has changed in the overview. An FFW can concentrate on his or her main mission, which is to fight instead of constantly monitoring all the controlling units of his or her gear.

XI. FURTHER WORK

When moving towards tactical military SDR/GUI, the system presented requires funding and field testing to be able to create a functional end-product. The rough estimate of the resources required is evaluated above in Section IX and in Table II based on the experience gained in twenty years spent as a participant in different military tests. Automated systems and allocation of diminishing resources force militaries to consider the facilitated performance offered by means of exploiting SDR/GUI. The result could be an agile and modular military performer with ever-improved capabilities and SA completed with the capability to utilize the diminishing resources more optimally with decreased instances of collateral damage.

Further work related to creating a functional system based on the idea-phase description outlined in this paper needs to pay attention to operational security issues of using software and hardware in a digitized battlespace. Issues such as adequate level of constant energy flow and protection against violations caused by electronic warfare must be studied, tested and solved before the adoption of the system in any type of operational use.

The introduced idea-phase description of a communication system aims at being battle-proof from the perspective of using a swarms of UAVs in particular in that it guarantees the usability of the system functionality as the swarms of UAVs can automatically recreate a functional communication network and maintain an adequate distance between each UAV in all the circumstances and situations: once a UAV is destroyed or shot down, it has been programmed to destroy itself mechanically and electronically. The remaining fleets of UAVs corrects the formation of flying UAVs automated, to maintain a functioning and reliable communication system to ensure the communication system remains intact. SON supports the communication system formed by the swarms of UAVs together with SDRs. Unfortunately, when resources have been invested in further researching, developing, and, finally, implementing the system introduced here, the follow-up

papers cannot any longer be accessed in any public domain data sources.

REFERENCES

- [1] K. A. Hempenius, A. R. Wilson, M. J. Kumar, N. Hosseini, M. E. Cordovez, and M. S. Sherriff, "A More Cost-Effective Unattended Ground Sensor Using Commercial Off-the-Shelf Products", Proc. of 2012 IEEE Systems and Information Engineering Design Symposium, pp. 62 – 67, ISBN: 978-1-4673-1286-8/12.
- [2] R. A. Syed and R.S. Wexler, "Army Warfighter Information Network-Tactical (WIN-T) Theory of Operation", Military Communications Conference, MILCOM2013, 18–20 Nov. 2013, San Diego CA, pp. 1453 – 1461, doi: 10.1109/MILCOM.2013.246.
- [3] S. L. Cotton and W.G. Scanlon, "Millimeter-Wave Soldier-to-Soldier Communications for Covert Battlefield Operations", IEEE Communications Magazine, pp.72 – 81, doi:10.1109/MCOM.2009.5273811, ISSN 0163-6804/09.
- [4] T. Ulversøy, Software Defined Radio: "Challenges and Opportunities", IEEE Communications Surveys & Tutorials, Vol. 12, Issue: 4, Fourth Quarter 2010, pp. 531 – 550, doi:10.1109/SURV.2010.032910.00019.
- [5] J. Mölsä, J. Karsikas, A. Kärkkäinen, R. Kettunen, and P. Huttunen, "Field test results and use scenarios for a WiMAX based Finnish broadband tactical backbone network" The 2010 Military Communications Conference, San Jose, CA, Oct. 31–Nov. 3, 2010, pp. 2014 – 2019. doi:10.1109/MILCOM.2010.5680446, ISSN 2155-7578.
- [6] Scaldio, Reconfigurable Radio Transceiver IP, www2.imec.be/content/user/File/Brochures/GR2010_Leaflet_Scaldio.pdf, accessed on 10th September 2014.
- [7] C-Y. Chen, F-H. Tseng, K-D. Chang, H-C. Chao, and J-L. Chen, "Reconfigurable Software Defined Radio and Its Applications", Tamkang Journal of Science and Engineering, Vol. 13, No. 1, 2010, pp. 29 – 38.
- [8] M. Islam, M.A. Hannan, S.A. Samad, and A. Hussain, "Software Defined Radio for RFID Application", Proc. of the World Congress on Engineering and Computer Science 2009 Vol I WCECS 2009, October 20–22, 2009, San Francisco, USA, pp. 415 – 418, ISBN:978-988-17012-6-8.
- [9] G. Thomas, "Situation Awareness Issues in Tactical Cognitive Radio", Proc. of 2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situational Awareness and Decision Support, 6–8 March 2012, pp. 287 – 293, ISBN 978-1-4673-0345-3/12.
- [10] X. Dong, S. Wei, Y. Li, L. Wang, and L. Bai, "Service-Oriented Radio Architecture: A Novel M2M Network Architecture for Cognitive Radio Systems", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2012, Article ID 762953, 8 pages, <http://dx.doi.org/10.1155/2012/762953>.
- [11] T. Saarelainen, "Modeling War as a Business Process with the Assistance of Service Oriented Architecture", ICDT 2014 The Ninth International Conference on Digital Communications, Feb. 23–27, 2014 Nice, France, pp. 1 – 7, ISSN: 2308-3964, ISBN: 978-1-61208-316-2.
- [12] R. Adams, "Special Operations Forces Skills", Military Technology, Vol. XXXVIII, Issue 5/2014, p. 38 – 42, ISSN 0722-3226.
- [13] T. Saarelainen and J. Jormakka, "Computer-Aided Warriors for Future Battlefields", The 9th European Conference on Information Warfare and Security ECIW2009, Lisbon, 6–7 July, 2009, Portugal, pp. 224 – 233, ISBN: 9781906638337.

- [14] T. Saarelainen, "Improving the Performance of a Dismounted Future Force Warrior by Means of C⁴I²SR", Doctoral Dissertation, 212 pages, ISSN: 1796-4059, ISBN 978-951-25-2457-0, 2013.
- [15] Y. Lin, H. Lee, M. Who, Y. Harel, S. Mahlke, and T. Mudge, "SODA: A Low-Power Architecture for Software Radio", Proc. of 33rd annual international Symposium on Computer Architecture, pp. 89 – 101, doi: 10.1109/ISCA.2006.37, 0-7695-2608-X, ISBN: 0-7695-2608-X.
- [16] P. Pancholy, J. Kushwaha, and R. Chitranshi, "Smart 4G network:-Implementation of SDR & SON with Automatic EM Radiation control", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 1, Issue 2, November 2012, pp.14 – 18, ISSN:2319-5967.

Small and Low Side Lobe Beam-forming Antenna Composed of Narrow Spaced Patch Antennas for Wireless Sensor Networks

Yosuke Sato, Shuzo Kato

Research Institute of Electrical Communications, Tohoku University
Sendai, Japan
y-sato@riec.tohoku.ac.jp

Abstract—In order to reduce sensor node power consumption and interference, high gain antennas are required for the base station (BS) but this means multiple BS antennas are required for 360° area coverage. A high gain beam-forming antenna is the key to base stations covering “spotty service areas” more effectively and reducing interference, especially in the ISM bands. To solve the size and side lobe issues of conventional beam-forming antennas, this paper proposes the “narrowly spaced” patch antenna array. The measured performance of a prototype confirms that while being 2/3^{rds} smaller than conventional antennas, it has lower side lobe levels (-20dBc) and high antenna gain (10dBi).

Keywords— Antennas, Linear antenna arrays, Planar arrays, Patch antennas, Wireless sensor networks

I. INTRODUCTION

In order to realize wide area sensor networks, low energy critical infrastructure monitoring (LECIM) has been attracting attention resulting in the 802.15.4k standard of the IEEE802 Standards committee. Since sensor nodes in wide area sensor networks are expected to offer long lifetimes with battery power, high gain base station antennas are required for reducing sensor node power consumption. High gain base station antennas are also a good solution to mitigating interference, a major problem in Industry-Science-Medical (ISM) band wireless communications systems [1-3]. Base station antennas that offer beam steering capability are also suitable for some wide area sensor networks, such as those for agriculture applications, where sensor nodes are closely located in multiple spots. Furthermore, 2-dimensional (2-D) beam-forming is more suitable for multiple spot communications. The beam forming antenna composed of phased array antennas is a good candidate due to its high gain and electrical 2-dimensional beam steering. However, the traditional phased array antenna with more than $\lambda/2$ element separation is not suitable for wide area sensor network base station antennas since it is physically large and the antenna pattern side lobes are significantly high. The sheer size increases installation difficulties and decreases wind tolerances. Since conventional uniform distribution phased array antennas [4-6] have larger side lobe level than fixed directive antennas, such as sector antennas [7-9], the traditional phased array antenna increases interference to/from

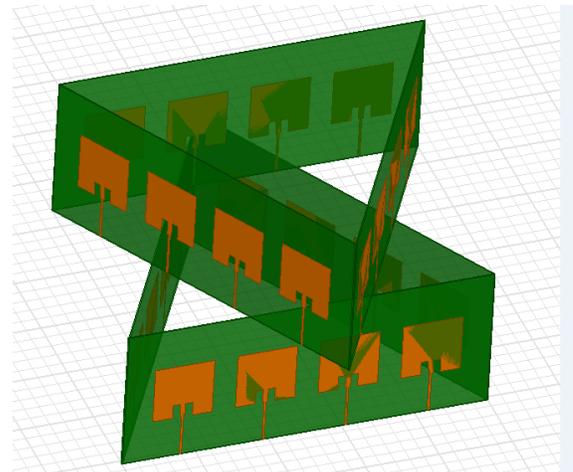


Fig.1. Antenna combination for 360 degree coverage

other systems. Recently, narrow beam phased array antenna were proposed [10-11] for radar applications. Since the narrow beam antenna required more than 10-elements with more than $\lambda/2$ element separation, these antennas are not suitable for base station antennas due to antenna size issue. In order to realize small, low side lobe, and high gain beam-forming antennas for wireless sensor networks base stations, we have proposed a beam-forming antenna composed of narrowly-spaced patch antennas. Due to electromagnetic simulator advancement, the phased array antenna with less than $\lambda/2$ element separation can be designed and simulated well. By using an unconventionally small element separation and balancing antenna gain against side lobe levels, we can realize small, low side lobe and high gain antennas. Since the proposed antenna offers a half power beam width (HPBW) of 60 – 80 degrees, 6 elements of the proposed antenna will be combined for 360 degree coverage as shown in Figure 1.

This paper is organized as follows. Section II proposes a 1-D low side lobe beam forming antenna to separate the service areas by horizontal degrees, such as 60 degrees per beam and Section III describes the corresponding 2-D beam forming antenna design which will be used to radiate the sensors by spots with horizontal and vertical beam-steering. Section IV summarizes our work on the proposed small, low side lobe and high gain beam-forming antennas

II. 1-DIMENSIONAL BEAM-FORMING ANTENNA

A high gain base station antenna is preferable to separate the service areas by horizontal degrees such as 60 degrees per beam. This creates following advantage against Omni base station antenna in star topology networks: (i) each sensor node needs to transmit less power by assuming up-link and down link budgets are balanced under maximum transmit EIRP is limited by radio regulation, (ii) the system will receive much less interference such as 1/6 when 6 element antenna is used, (iii) the system can achieve much higher spectrum efficiency such as 3 times when 6 element antenna is used.

For practical applications, base station antenna beam forming capability is preferable to adjust the direction of each beam to radiate the power efficiently for where sensor nodes are located. For this beam-forming antenna, we propose a methodology to give mutual couplings between antenna elements intentionally contrast to the traditional antenna design. A small, high gain and planer antenna element is desired for the narrow-element-separation phased array. Patch type antennas are easy to connect to micro strip lines (MSL) and antenna size is reduced by using high relative permittivity substrates. The designed single patch element is shown in Figure 2. The patch antenna is fed by a MSL and radiates on the Z-axis. The antenna has 10mm wide notches for impedance matching. The substrate material is FR-4 whose permittivity is 4.12 and $\tan \delta$ is 0.019 at around 900MHz. Patch antenna element width and height are set at 81.2mm to suit the working frequency of 920-928MHz that is allocated

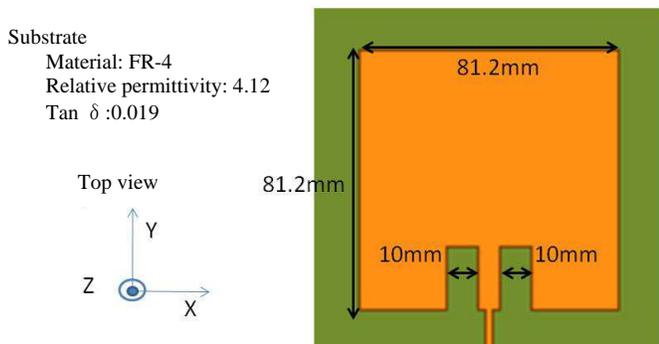


Fig.2. Single patch antenna

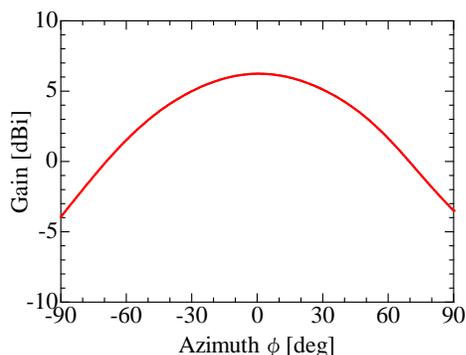


Fig.3. Single patch antenna gain

for sensor networks in Japan; MSL width is 6.95mm with input impedance of 50 ohm. The element size, 81.2mm (0.25λ) is suitable for small element separation arrangements. Gain and input reflection of the patch element were simulated by an electromagnetic simulator (HFSS v12.1). As shown in Figure 3, its peak gain is 6.28dBi and HPBW is 98degrees. Reflection is less than -10dB from 920MHz to 928MHz as shown in Figure 4. These characteristics confirm that the proposed single patch antenna is suitable for a beam-forming antenna element.

A four-element beam-forming antenna composed of narrowly-spaced linear array antennas was simulated and measured. The results confirm that the proposal offers small, low side lobe and high gain beam-forming antennas. The key design procedure is to determine the best element separation for more than 10 dBi gain, and -20dBc side lobe. This trade off was balanced by using an electromagnetic simulator. As shown in Figure 5, in order to achieve better than -20dBc side lobes, the element separation should be 113mm or smaller. For more than 10dBi gain, element separation should be 113mm or larger as shown in Figure 6. From these results, the separation of 113mm can realize small, high gain and low side lobe beam-forming antennas.

The beam-forming antenna composed of the resulting antenna is shown in Figure 7. When compared to traditional $\lambda/2$ separation designs, this structure reduces antenna size by 35 percent. The simulated reflections of each element

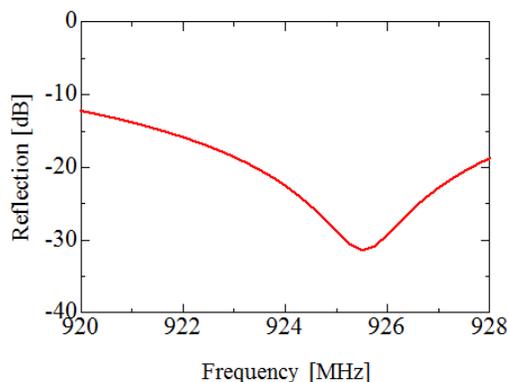


Fig.4. Single patch antenna reflection

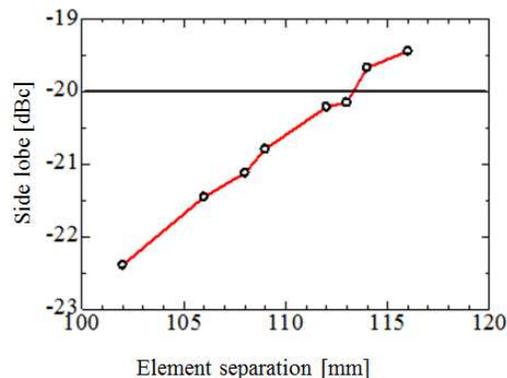


Fig.5. Element separation optimization for side lobe (4-element 1-D beam-forming antenna)

is shown in Figure 8. Element reflection is less than -10dB from 920MHz to 928MHz. The proposed antenna has more than 10dBi gain and -20dBc side lobe as shown in Figure 9. Based on electromagnetic simulation results, a prototype was designed, fabricated and measured. Measured reflection results are shown in Figure 10. The measurement results show the resonance point shift by 1MHz from the simulation and this is caused by fabrication error. The designed antenna gain / directivity have been measured in an anechoic chamber by comparing with an Omni antenna as shown in Figure 11. The prototype antenna is connected with the phase shifter (PS) board and signals are fed from a signal generator (SG) (The measured frequency is 923MHz). Receiver antenna, which is a sleeve antenna with 2dBi gain, is set at 4.8m from the

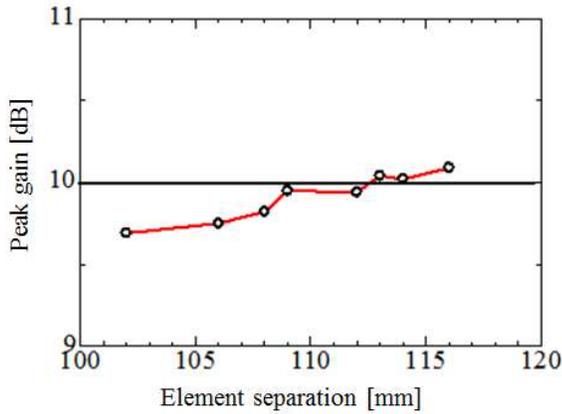


Fig.6. Element separation optimization for side lobe (4-element 1-D beam-forming antenna)

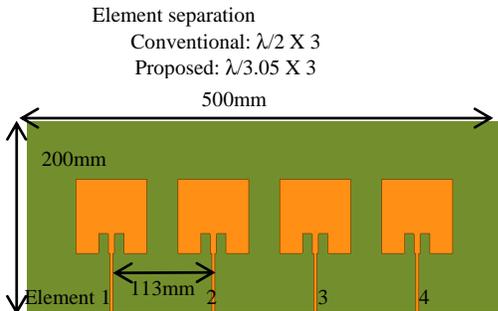


Fig.7. Beam-forming antenna composed of 4-element patch phased array antenna

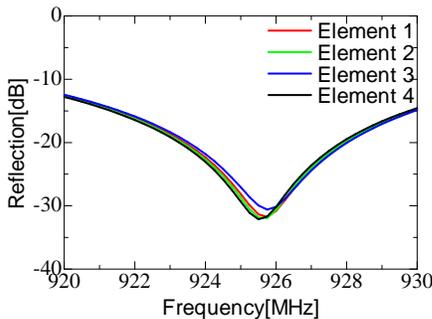


Fig.8. Simulated reflection (4-element 1-D beam-forming antenna)

transmitter antenna and received signals are fed to a spectrum analyzer (SA). The difference from the Omni transmission antenna shows the performance of the prototype antenna. A comparison of simulated and measured directivity is shown in Figure 12. Measured reflection is less than -10dB. The measured peak gain is 8.1 dBi, smaller than simulated value, and the measured side lobe level is -15.3dBc. Since the difference between measured and simulation results is small,

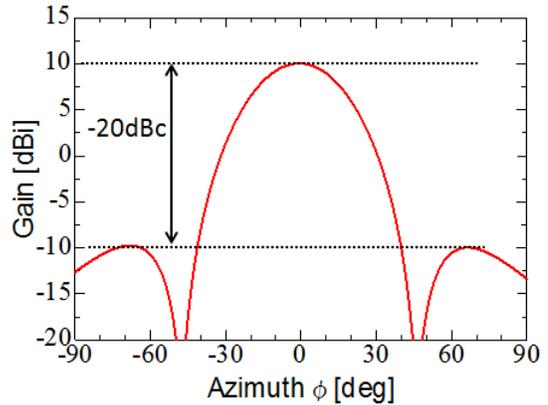


Fig.9. Simulated directivity (4-element 1-D beam-forming antenna)

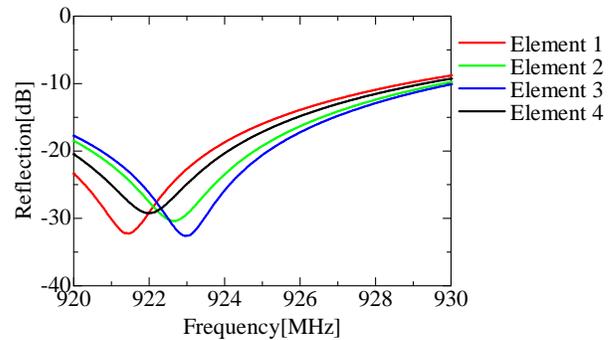


Fig.10. Measured reflection (4-element 1-D beam-forming antenna)

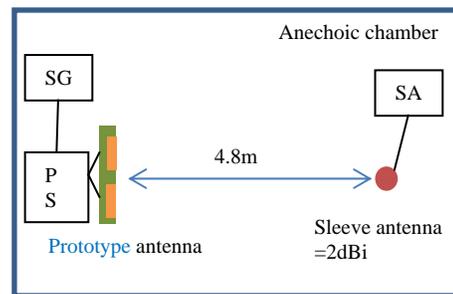


Fig.11 Measurement setting

the results show that proposed concept is suitable for creating 1-dimensional beam-forming antennas

III. 2-DIMENSIONAL BEAM-FORMING ANTENNA

For wide area sensor networks, 2-D beam-forming antennas are more useful and size reduction is more important than is true for 1-D ones. In order to design 2-D beam forming antennas composed of 16-element narrowly-spaced patch antenna arrays, the element separation was optimized by using an electromagnetic simulator. As shown in Figure 13, in order to achieve better than -20dBc side lobes, the element separation is 98mm or smaller. Since the 2-D phased array antenna can emit narrow beams in E-plane and H-plane, the gain at the element separation of around 98mm exceeds 10dBi as shown in Figure 14. From these optimization results, the element separation is set at 98mm. The resulting beam-forming antenna is shown in Figure 15. The substrate permittivity is 4.12 which is same in Figure 2 and 7. The antenna elements are separated by 98mm ($\lambda/3.38$). When compared to the traditional $\lambda/2$ separation, this structure reduces the antenna size by 41

percent. The simulated reflections of elements 1, 2, 5, and 6 are shown in Figure 16. Since the antenna is symmetrical, these four plots are enough to show overall reflection performance. All reflection values are less than -10dB from 920MHz to 928MHz. The antenna gain exceeds 13dBi and side lobes are at most -20dBc as shown in Figure 17. A matching prototype was designed, fabricated and measured. Measured reflection results are shown in Figure 18, and a comparison of simulated

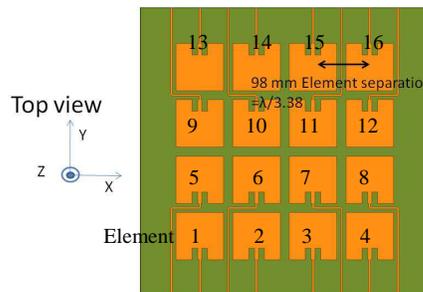


Fig.15 2-D beam-forming antenna

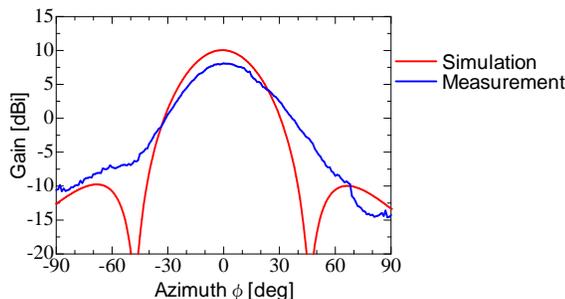


Fig.12. Measured directivity (4-element 1-D beam-forming antenna)

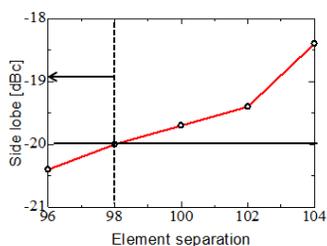


Fig.13. Element separation optimization for side lobe (16-element 2-D beam-forming antenna)

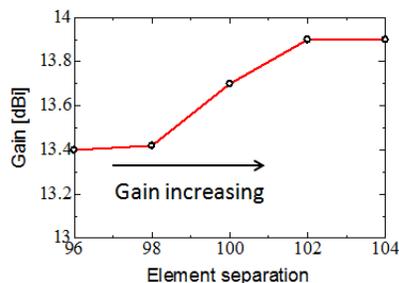


Fig.14. Element separation optimization for peak gain (16-element 2-D beam-forming antenna)

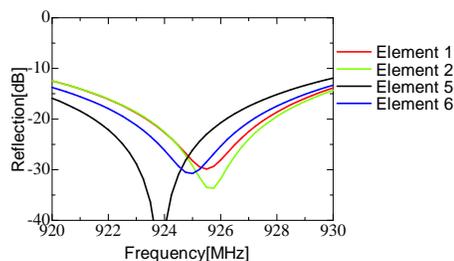


Fig.16. Simulated reflection (16-element 2-D beam-forming antenna)

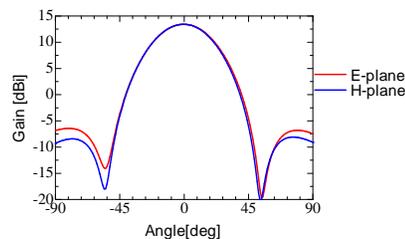


Fig.17. Simulated directivity (16-element 2-D beam-forming antenna)

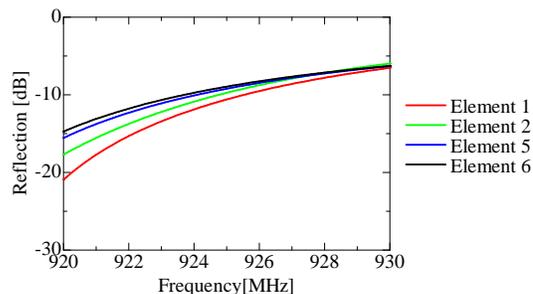


Fig.18. Measured reflections (16-element 2-D beam-forming antenna)

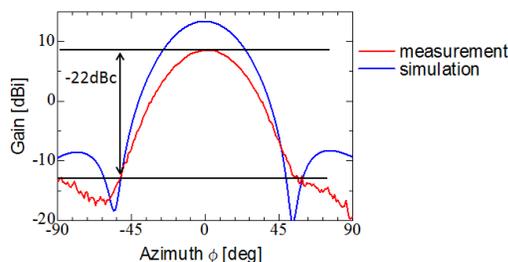


Fig.19. Measured directivity (16-element 2-D beam-forming antenna)

and measured directivity is shown in Figure 19. Measured reflection values are better than -10dB from 920MHz to 928MHz. In the measurement reflection, the resonance point is out of measurement range. Since this difference is comes from substrate permittivity error, the difference can be canceled. The measured peak gain is 8.59 dBi, smaller than the simulation value, and the measured side lobe level is -22.3dBc

IV. CONCLUSION

This paper has proposed the low side lobe beam-forming antenna composed of narrowly-spaced patch antenna arrays. To validate the proposed antenna, 1-D and 2-D beam-forming antennas composed of patch antenna arrays were simulated and designed, and prototypes were measured. The simulation results show that both antennas have, small (65 percent size reduction) and low side lobe level (better than -20dBc), and high gain (more than 10dBi). Measured results show close agreement with the simulation results and the 2-D beam-forming antenna achieved side lobe levels better than -20dBc. Thus, the proposed beam-forming antenna is a good candidate for base station antennas of wide area sensor networks.

ACKNOWLEDGMENT

This project is partially supported by Ministry of Internal Affairs and Communications, SCOPE and also partially supported by Japan Science and Technology Agency, A-STEP

REFERENCES

[1] J. Park, S. Park, D. Kim, P. Cho, and K. Cho, "Experiments on Radio Interference Between Wireless LAN and Other Radio Devices on a 2.4 GHz ISM Band", Vehicular Technology Conference 2003, vol. 3, April 2003, pp.1798-1801.

[2] S. Unawong, S. Miyamoto, and N. Morinaga, "Techniques to improve the performance of wireless LAN under ISM interference environments", APCC/OECC '99, vol.1, Oct. 1999, pp. 802 – 805.

[3] M. R. Souryal, D. R. Novotny, D. G. Kuester, J. R. Guerrieri, and K. A. Remley, "Impact of RF interference between a passive RFID system and a frequency hopping communications system in the 900 MHz ISM band" Electromagnetic Compatibility (EMC), July 2010, pp 495 – 500.

[4] C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-Lobe Level", Proceedings of the IRE, Volume:34, Issue: 6, June 1946, pp.335 – 348.

[5] T.T. Taylor, "Design of line-source antennas for narrow beamwidth and low side lobes", Antennas and Propagation, Transactions of the IRE Professional Group on, Volume:3, Issue: 1, Jan. 1955, pp.16-28.

[6] R. C. Hansen, "Fundamental Limitations in Antennas", Proceedings of the IEEE, Volume:69, Issue:2, Feb. 1981, pp170 – 182.

[7] T. Seki, H. Yamamoto, and T. Hori, "Low-loss and compact sector antenna that adopts omni-directional characteristics", Antennas and Propagation Society International Symposium, Vol 3, July 2000, pp. 1306-1309.

[8] A. Edalati, and T. A. Denidni, "Beam-switching Antenna Based on Active Frequency Selective Surfaces", APSURSI, July 2011, pp. 2254-2257.

[9] T. D. Dimousios, S. A. Mitilneos, S. C. Panagiotou, and C. N. Capsalis, "Design of a Corner-Reflector Reactively Controlled Antenna for Maximum Directivity and Multiple Beam Forming at 2.4 GHz", IEEE Transactions on Antennas and Propagation, Apr. 2011, pp. 1132 – 1139.

[10] D. G. Kurup, M. Himdi and A. Rydberg, "Synthesis of Uniform Amplitude Unequally Spaced Antenna Arrays Using the Differential Evolution Algorithm", IEEE Trans. Antennas Propagation, Vol. 51, No. 9, Sep. 2003, pp. 2210-2217.

[11] B. Pompeo, L. Pralon, M. Pralon, and R. Mendes, "Phase-only pattern synthesis using a modified least squares method for phased arrays", Radar Conference (EuRAD), Oct. 2013, pp. 443-446.

Accuracy Enhancements in Indoor Localization with the Weighted Average Technique

Grigorios G. Anagnostopoulos, Michel Deriaz

Institute of Services Science
University of Geneva
Geneva, Switzerland

Email: {grigorios.anagnostopoulos, michel.deriaz}@unige.ch

Abstract—Indoor localization is a topic that has drawn great attention over the last decade. One of the main goals of the research in the field is to improve the achieved accuracy. Along with the accuracy, factors like the easiness of deployment and reconfiguration, the cost, the computational complexity, and the ability to tune the desired accuracy in specific areas are also important. In this study, we used Bluetooth Low Energy (BLE) technology, that offers a low cost and is easily deployed and reconfigured. The weighed average method, combined with the selection of the closest beacons and the averaging of the received signal strength indication (RSSI) at the distance domain proposed in this paper, offers an accuracy down to 0.97 meters, depending on the deployment configuration. This method was tested in our lab and was following installed at the hospital in Perugia, Italy, in the context of the Ambient Assisted Living (AAL) Virgilius project, where users can navigate with a smartphone.

Keywords—Indoor Localization; Received Signal Strength; Positioning; Bluetooth

I. INTRODUCTION

During the last years, the field of indoor positioning has drawn an increasing attention of researchers. Outdoor positioning has been ahead, having reached many users through commercial applications. Nowadays, almost all new mobile devices are equipped with global positioning system (GPS) technology, which has familiarized most users with the concept of positioning. On the other hand, no indoor positioning method has been broadly recognized as a standard one, and the research in this domain has led to having multiple alternatives.

A technology commonly used for positioning in indoor environments is the Wi-Fi signal [1][2]. One advantage of using Wi-Fi is that most buildings have several Wi-Fi access points, in order to provide internet access, so the hardware required is already installed. On the other hand, usually the access point network is not dense enough to facilitate a satisfactory precision of localization. Moreover, the transmission of the Wi-Fi access points is unstable, as well as the reception in big distances due to multipath effects, and therefore using the RSSI of them can be problematic.

In this study, we work with BLE technology. BLE is a wireless technology used for transmitting data over short distances. It has a low energy consumption and cost, while maintaining a communication range similar to that of its predecessor, Classic Bluetooth. As transmitters, we used iBeacons, an Apple technology standard, and more precisely the Tod iBeacons [3].

The iBeacon technology allows mobile applications (running on both iOS and Android devices) to listen periodically for signals from beacons and react accordingly. Each beacon broadcasts a self-contained packet of data periodically. The packets contain the mac address of each beacon, so that the receiver can distinguish among them. The RSSI can be used to estimate the distance between the mobile device and the transmitting beacon [1][4][5][6]. Due to their low cost and low consumption, a dense network can be deployed. Having a dense deployment can lead to a reliable distance estimation, at least from the closest beacons.

This distance estimation is used to derive the actual position, usually by using lateration methods [7][8]. These methods can have some drawbacks, as for example, when the estimated distances are wrong, or when the beacons used are aligned, an estimated position that is far from the real one may be returned. Furthermore, using different mobile devices with different receiving capabilities can add a systematic error to each distance estimation that will dramatically affect the lateration outcome.

In our method, we proceeded with another approach. We propose a placement of beacons such that the beacons surround all the area that we want to cover. The position prediction is limited to the area that is defined by the polygon that the beacons' positions define. We get a distance estimation from each beacon, by averaging the estimated distances that correspond to the latest RSSI measurements from this beacon. In this way, we cope with the instability of the RSSI.

Having this filtered distance estimation, we focus only on the B closest beacons. In this work, we propose $B = 4$, as will be discussed later. We use the inverse value of the distance estimation as weight, in order to perform a weighted average of the positions of the 4 closest beacons. This weighted method, is also met in [9]. That work, to our knowledge, does not have RSSI filtering, nor does it focus on the closest beacons. The weighted method in [9] is used with radio frequency identification (RFID) technology, as an area/room selection first step, before performing a server side supervised machine learning positioning method.

The lateration approach uses the assumption that distance estimations are close to accurate, which is unlikely. Our proposed method anticipates the uncertainty of this estimation as an absolute value. It firstly utilizes that the expected error is

smaller in small distances, and secondly, the main conceptual idea of RSSI methods, that a stronger RSSI reception from beacon *a* as compared to beacon *b* is interpreted as being closer to beacon *a*, especially after averaging distance estimations.

The advantages of the proposed technique are numerous. The Bluetooth beacons are easily deployed and rearranged in order to cover new areas of a building or to improve accuracy with a more dense placement. For example, the deployment could be more dense in a corridor with many doors, where accuracy is critical, compared to a long corridor with few doors, that may simply link two buildings. Another advantage is that this method does not require the creation of a radio map [10], where measurements of RSSI from all access points should be stored for many points of the area where it will be used. One can reconfigure the deployment, by adding for example one beacon, with no need to retake any measurements for a radio map, but simply by storing the position of the new beacon. Our method also offers low computational and implementation complexity. Finally, as all RSSI techniques, it has the advantage that most modern mobile phones can offer the RSSI of a Bluetooth reception, and thus no extra hardware or modification of the devices is required.

The rest of this paper is organized as follows. In Section 2, we present the propagation model and how it is used to derive distance estimations from the RSSI values. In Section 3, we present the idea of performing a weighted average of the known beacon positions. Measurements and both theoretical and experimental results are reported and discussed in Section 4. Finally, future work directions along with conclusions drawn are presented in Section 5.

II. PROPAGATION MODEL AND RSSI METHODS

In RSSI methods used in localization, the received signal strength is used as a measure from which the distance between the transmitter and the receiver can be inferred. Nevertheless, the RSSI received at a given time and space depends on many other factors other than the relative distance of the two devices. Even the slightest changes in position and orientation can provoke dramatic variations to the RSSI values [11]. Moreover, the movement of people and objects in the environment often has great effect on the signal. In general, RSSI is vulnerable to strong multipath effects, especially indoors [11]. Furthermore, factors like temperature and humid conditions can affect the propagation of the signal [11]. Using a set of RSSI measurements instead of a single instantaneous measurement can improve the accuracy of the distance estimation [5].

The RSSI, apart from the propagation channel, depends on the transmitter and the receiver. For a given installed system the transmitters are the same, but it would be desirable that each user could use the system by using her personal mobile device (smart phone, tablet, etc.). Taking under consideration all these factors, we propose the propagation model and its parameters that will be used in our system.

The propagation model commonly used for the RSSI to distance correspondence, where the expected RSSI r_i in distance d_i is calculated, is the following:

$$r_i = r_0 - 10 n \log_{10}(d_i/d_0) \tag{1}$$

In this formula, r_0 is the received RSSI at a reference distance d_0 , and n is the path loss exponent which depends on the transmission channel, the transmitter and the receiver. Using 1 meter as reference distance, and solving for d_i , the formula is simplified to:

$$d_i = 10^{\frac{r_0 - r_i}{10 n}} \tag{2}$$

A Tod Bluetooth iBeacon was placed at the center of the corridor of Figure 3, and RSSI measurements were performed at several points at a known distance from the iBeacon. We performed these measurements using a Samsung Galaxy S4, and then ran a regression to find the parameters of the best fitting curve described by (2). In Figure 1, we see the measured RSSI values (black dots) in several distances (at 0.25,0.5,1,2,3,4,5 and 6 meters), and with blue color the resulted propagation model as the best fitting curve. The estimated values of these parameters are $r_0 = -62.72$ and $n = 2.2853$. Substituting in (2), we get:

$$d_i = 10^{\frac{-62.72 - r_i}{2.2853}} \tag{3}$$

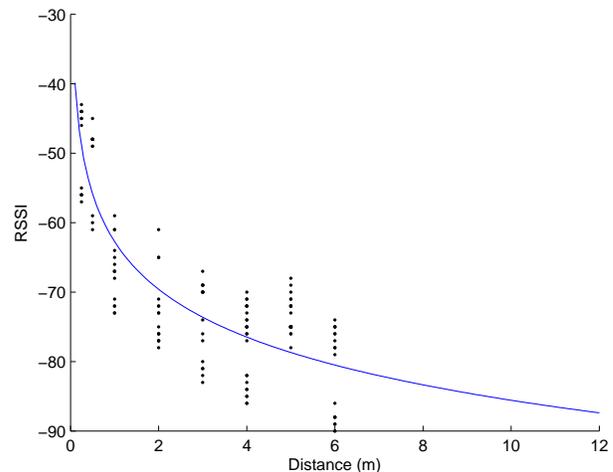


Fig. 1. RSSI measurements at several distances (in black) and the resulting propagation model (in blue) as the best fitting curve described by (2).

In order to have a more reliable distance estimation we do not use just the latest RSSI reception but a set of the latest ones. Due to the non linear relation of distance with RSSI, it is important to decide if we will average the RSSI values and then get the distance estimation as their average, or if we should calculate the corresponding distance of each RSSI reception and then use the average of these distances. In Figure 1, we see that for distances from 0 to 5 meters,

the RSSI differs significantly. On the other hand, for distances greater than 15 meters, RSSI differences are minor. Since we target to utilise the reliability and distinguishability of the small distance measurements, we direct our method to this part of the propagation model. Given that the derivative of the RSSI curve changes with the distance, averaging the RSSI values inserts an intrinsic error to the estimation.

To state this argument we explain a simple example. Assume that users are moving and the RSSI measurements they receive from each beacon correspond exactly to the real distance from it at each moment. Averaging in the distance domain, provides the users' average distance. On the other hand, due to the non linear relation of distance and RSSI, averaging RSSI values first, will give a distance estimation different than the average distance.

Furthermore, in case where the measurements of the RSSI values are biased, in short distances, small RSSI errors have small consequences to the distance estimations. On the other hand, for big distances, a small fluctuation of the RSSI values can have a dramatic consequence to the estimated distance. In the proposed method, only the RSSI values of the closest beacons are used, so averaging in the RSSI domain would introduce a bias. For these reason, averaging in the distance domain was selected.

III. WEIGHTED AVERAGE OF BEACON POSITION

Initially, it is worth noting that the method proposed can give estimations of positions only inside the polygon area that is defined by the positions where the beacons are placed. For an indoor localization system and its applications, it may be desirable to constrain the prediction inside a specific area, i.e., inside the building. In case where map matching is used to provide navigation, a jump of the estimation outside a building could lead to problematic navigation. Thus, in practice, to provide coverage in a rectangular room or a corridor with the proposed method, $B = 4$ is the minimum number of beacons to cover this area. Later in this work, we discuss proposed configurations.

Having obtained an estimation about the distance of the mobile device from each beacon, we proceed to the position estimation. Due to the phenomenon of multipath effects, it is unrealistic to claim that at every moment the distance estimation will be precise. Especially in big distances, just a small difference in the RSSI values is translated to big distance differences. On the other hand in small distances the RSSI values are quite distinguishable. We utilize this fact, in the following way. From the list of beacons that are detected, we keep the 4 closest beacons. Assuming that the mobile device is inside the coverage area (beacon defined polygon), the estimated position will also be inside the quadrilateral defined by these four beacons. Let $[e_1, e_2, e_3, e_4]$ be the estimated distances from the 4 closest beacons, while $[lat_1, lat_2, lat_3, lat_4]$ and $[lon_1, lon_2, lon_3, lon_4]$, the corresponding latitude and longitude of their positions. We calculate the latitude Lat_{est} and longitude Log_{est} of the estimated position as follows:

$$Lat_{est} = \frac{\sum_{i=1}^4 \frac{lat_i}{e_i}}{\sum_{i=1}^4 \frac{1}{e_i}}, \quad Lon_{est} = \frac{\sum_{i=1}^4 \frac{lon_i}{e_i}}{\sum_{i=1}^4 \frac{1}{e_i}} \quad (4)$$

In (4), we calculate the weighted average of the four closest beacons' positions, using as weight $1/e_i$, that is the inverse of the estimated distance from beacon i . By using this weighted average the prediction is limited inside the quadrilateral of the closest beacons, and with the specific weight that is proposed, the prediction is pulled towards the closest beacon, although allowing the rest of the beacons to contribute inversely proportionately to their distance.

The minimum number of closest beacons that could be used is 3, since 3 points define a plane. In the case where the 3 closest beacons were used, the defined area would be a triangle. In the middle area of this triangle, the estimation is slightly better comparing to the case where 4 beacons are used. The drawback with the usage of only 3 closest beacons is that when the user is moving and passes from one triangle to the other, the accuracy of estimation near the common edge of the two triangles is significantly degraded. Using the 4 closest beacons offers a smooth transition from one triangle to another. In the following section, along with the system's accuracy measurements, an error analysis of these two cases is presented.

IV. MEASUREMENTS, RESULTS AND DISCUSSION

The weighted average method proposed may have an error in the location estimation even when the distance estimations are precise. We model this error in Figure 2, for the cases where 4 and 3 closest beacons are used. We simulate the deployment at the corridor where our system was tested. The beacons are placed at a height of 2.40 meters, following a zigzag pattern (alternatively at the left and right wall of the corridor), every four meters along the direction of the corridor. The orientation of the beacons is towards their opposite wall. We observe that, for the first scenario (4 closest beacons), the error is lower at the center of the corridor, and changes smoothly as we move along the length of the corridor. On the other hand, using just 3 beacons degrades significantly the accuracy estimation at a broader area. The error increases rapidly when approaching at the edges of the triangles that the 3 closest beacons define.

Our method was tested in the corridors of the Centre Universitaire d'Informatique (CUI), of the University of Geneva. Measurements were taken with two ways. First, by letting the mobile device at a specific place and receiving the estimated positions it calculated. A second approach was to test the localization while having the mobile device moving, which better corresponds to real life scenarios.

In order to get a broad estimation of the positioning accuracy, we took 1000 measurements at three points in the corridor. Two mobile devices were used for these measurements,

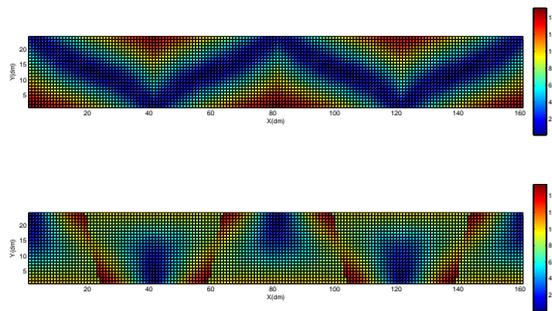


Fig. 2. Error in decimeters (color scale) of position estimation in a corridor at the position x,y, when using 4 (upper plot) and 3 (lower plot) closest beacons method.

a Samsung Galaxy S4, which was used for the creation of the propagation model (3), and a Samsung Galaxy Note 3. The goal was to test the adaptability of the system to different devices, with different reception characteristics.

The results are represented in the following table. In Figure 3, the beacon positions are highlighted with black color, and the places that the measurements were taken with red. We placed point A to be at the center of the corridor, that better represents the usual usage area. In order to test the accuracy of the system at its limits, we place point B exactly at the wall on the side of the corridor, and point C at the end of the corridor. Both points B and C lay exactly at the limit of the beacon’s polygon.

TABLE I. ACCURACY

		Mean error (m)	σ of error (m)
Point A	S4	1.22	0.82
	Note 3	0.97	0.48
Point B	S4	3.08	0.76
	Note 3	3.18	1.09
Point C	S4	3.82	2.20
	Note 3	3.50	1.78

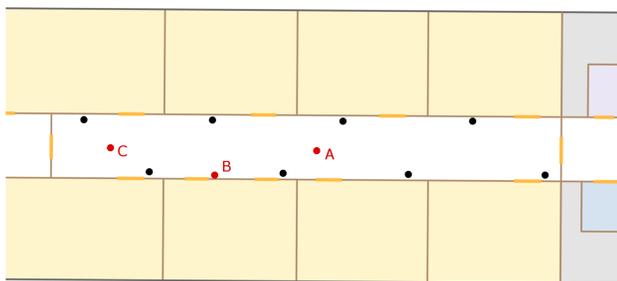


Fig. 3. Test environment. Beacon positions are highlighted with black color, and the places that the measurements were made with red color.

The position estimation is very reliable throughout the corridor, with an average accuracy of 1.22 and 0.97 meters for the two devices. The accuracy drops at the limits of the polygon, but remains reliable, with an average error of 3.08

(and 3.18) meters next to the wall. It is worth mentioning that the accuracy with the Note 3 is really similar to the one with the S4, that is the device used for the propagation model calculation.

Apart from the static measurements, we performed also a dynamic test in the same environment. The corridor is 2.5 meters wide and the trajectory of the mobile phone’s movement was 25 meters, at a straight line, and in a constant pace of 1 m/s. In Figure 4, the real trajectory is represented by the grey straight line segments, and the corresponding trajectory of the position estimations by the red crooked line segments. The mean value and standard deviation of the distances between the estimated and the true positions are measured to be $\mu = 2$ and $s = 1.28$ respectively.

The error in the dynamic version is higher than the static one, as expected. Nevertheless, a precision of 2 meters for a moving device that drops to 0.97 when the device is static, can be satisfactory for most indoor position applications. The distance measure used for calculating the error was the two dimensional euclidean distance. Of course, in cases where map matching is used, all position estimations would be projected on the axis of the corridor, reducing the error only at its length-wise component, removing the width-wise component.



Fig. 4. Real trajectory (grey straight line segments) and path from position estimations (red crooked line segments).

V. CONCLUSION AND FUTURE WORK

An innovative indoor positioning system is presented in this paper. The wireless technology used is BLE, which has low cost, and offers ease of deployment. It does not require the creation of a radio map, neither a calibration stage, but simply the awareness of the positions where the beacons were deployed. The scenarios that the technology was tested with were directed towards localization in the corridor area of buildings. Localization in corridors can assist a navigation system to guide a user to a room in a building. The achieved accuracy of localization is 0.97 meters (depending on the device).

Our goal is to further improve this method by testing other beacon configurations [12] (apart from the zigzag pattern) that might optimize the accuracy of estimation by also keeping a low density of beacon deployment. In the context of the the AAL Virgilius Project, beacons were deployed following the zigzag pattern in the corridors of the hospital of Perugia, Italy. The goal was to navigate a user from the entrance of the hospital to the door of the room that the user would choose. Regarding a corridor deployment, the zig-zag pattern is very efficient, but it remains to design a general pattern that works for all room shapes and sizes.

Moreover, the weight used (inverse distance) has been selected after comparison with other possible weights, but remains to be further studied. Lastly, after obtaining a series of position estimations, we intend to apply an extra filtering step, in order to smooth the transition among position estimations.

ACKNOWLEDGEMENTS

This work is supported by the AAL Virgilius project (aal-2011-4-046).

REFERENCES

- [1] S. Mazuelas *et al.*, "Robust indoor positioning provided by real-time rssi values in unmodified wlan networks." *J. Sel. Topics Signal Processing*, vol. 3, pp. 821–831, 2009.
- [2] M. Lee and D. Han, "Voronoi tessellation based interpolation method for wi-fi radio map construction." *IEEE Communications Letters*, vol. 16, no. 3, pp. 404–407, 2012.
- [3] "Tod iBeacons," 2014, URL: <https://www.todhq.com/> [accessed: 2014-09-09].
- [4] F. Vanheel, J. Verhaevert, E. Laermans, I. Moerman, and P. Demeester, "Automated linear regression tools improve rssi wsn localization in multipath indoor environment." *EURASIP J. Wireless Comm. and Networking*, vol. 2011, p. 38.
- [5] C. Papamantou, F. P. Preparata, and R. Tamassia, "Algorithms for location estimation based on rssi sampling," in *ALGOSENSORS*, 2008, pp. 72–86.
- [6] M. Saxena, P. Gupta, and B. N. Jain, "Experimental analysis of rssi-based location estimation in wireless sensor networks," in *In Proc. Int. Conf. Communication System Software and Middleware*, 2008, pp. 503–510.
- [7] Z. Yang, Y. Liu, and X.-Y. Li, "Beyond trilateration: On the localizability of wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, December 2010, pp. 1806–1814.
- [8] K. Lu, X. Xiang, D. Zhang, R. Mao, and Y. Feng, "Localization algorithm based on maximum a posteriori in wireless sensor networks," *IJDSN*, vol. 2012, pp. 1–7.
- [9] H. Zou, L. Xie, Q.-S. Jia, and H. Wang, "An integrative weighted path loss and extreme learning machine approach to rfid based indoor positioning," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, Oct 2013, pp. 1–5.
- [10] S. Sorour, Y. Lohan, S. Valace, and K. Majeed, "Joint indoor localization and radio map construction with limited deployment load," *IEEE Transactions on Mobile Computing*, vol. 99, 2014, no. PrePrints, p. 1.
- [11] G. Zanca, F. Zorzi, A. Zanella, and M. Zorzi, "Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks," in *Proceedings of the Workshop on Real-world Wireless Sensor Networks*, ser. REALWSN '08. New York, NY, USA: ACM, 2008, pp. 1–5.
- [12] N. Bulusu, D. Estrin, and J. S. Heidemann, "Adaptive beacon placement." in *ICDCS*. IEEE Computer Society, 2001, pp. 489–498.

60GHz Radio Hose for Wireless Harness Communication Systems

Yosuke Sato, Vannsith Ith, Shuzo Kato

Research Institute of Electrical Communications, Tohoku University
Sendai, Japan
y-sato@riec.tohoku.ac.jp

Abstract— In order to lower car fuel consumption and increase communication capacity, a wireless harness communications system that uses metal coated plastic/rubber hose (radio hose) as the communications channel is proposed. Measurements results show that the proposed system reduces conventional harness weight to 1/10, and increase reliability even if the hose is cut and/or pinched. This indicates a possibility for wireless communications systems to be adopted not as a soft safety system but a hard safety system for the first time.

Keywordst— *Sesor; millimeter wave communications; wireless communications; automobiles, harness*

I. INTRODUCTION

Since the weight of wire harnesses in cars has been increasing due to increase of number of Electronic Control Unit (ECUs), there is strong need for “harness weight saving”. A typical wire harness weighs around 30 to 50 Kg and various weight saving approaches have been proposed [1]. Conventional wire harness communication systems use copper wires, to interconnect the many ECUs. Since the number of such units continues to increase with the rapid adoption of hybrid and electrical systems, conventional wire harness systems are becoming excessively complex and heavy. Another trend is the increase use of intra-car multimedia systems that demand high data rates. In “intra-car communication systems”, there is not only front display for car navigation systems but also are many multimedia monitors for TV and movies. In order to realize these applications, a high data rate communication channel is also required in vehicles [2-3]. Currently, USA automobile regulation requires tire pressure monitor. This has been done by using Bluetooth technology and it is so called “soft safety” application where some erroneous transmissions are allowed. Meanwhile, engine and brake control systems cannot allow erroneous transmission at all and use wired connections. They are called hard safety and not even short term disruption is permitted. This is naturally the only choice given by current technology. To meet these requirements, the “wireless harness communication system with radio hose” has been proposed. The proposed radio hose is a hybrid of wire and wireless communications. This system uses a “radio hose” which is lighter than conventional wire harnesses while offering the wide bandwidth needed for multimedia services, the robustness demanded by the control systems, and multiplexing transmissions by a single radio hose. The radio hose is made by thin metal coated low dielectric material, such as polyolefin whose weight is much lighter

than wire harness. Since the proposed radio hose aims to replace the current wired harness, we name the radio hose system as “wireless harness”. The proposed system has high reliability since communication remains possible even if the hose is cut or pinched. Metal coated hose offers high immunity from electromagnetic interference as well. The proposed system offers light weight, wide bandwidth, and highly reliable wireless harness systems. Since conventional wired connects fail once a wire or connector fails, the proposed system offers higher communication reliability and well suits “hard safety” applications. The proposed system is assumed to use 5 and 60GHz unlicensed bands. Although all wireless transmissions are contained within the hose, there is no need to use unlicensed bands. We have selected these bands since relatively cheap transceivers are available for 5 and 60GHz band applications. When transceiver physical size is considered, the 60GHz band is promising, while the 5GHz band offers long range communications. In our previous work [1], the large radius hoses (25-40mm) were found to be suitable with low loss for wireless harness communication systems. In order to replace the current wired harness (with about 6.5 mm in diameter), however, this paper focuses on small radius hoses, such as 6.5 mm in diameter.

This paper shows that the transmission loss of a 4m length hose is estimated 39.4dB, which is 40.6 dB lower than the loss of over-the-air transmission (80dB) [4] and a delay spread is as small as 0.2ns, which is small enough for effective communications without equalizers and no equalizers will contribute to realize low latency systems. Although this paper describes 3 Gbps transmission systems mostly, a variety of transmission rates, such as 1 Gbps or lower are possible supported by IEEE802.15.3c standard. We used the measured results to design the channel model and BER simulations of both control and entertainment signal transmission channels. The simulation results show BER penalty is less than 1dB with Forward Error Correction (FEC). Since these results show that there is no need to use Orthogonal frequency-division multiplexing (OFDM), essential if the delay spread is significant, the proposal offers high speed communication systems with single carrier links and simple equalizers.

This paper is organized as follows. Section II describes the structure and basic performance of the proposed wireless harness communications system. Section III describes its reliability. Section IV describes multi channel communication performance. Section V summarizes our work on wireless harness communication systems.

II. WIRELESS HARNESS COMMUNICATION SYSTEMS

As shown in Figure 1, the model considered here consists of control systems to send control signals from sensors, and entertainment signal transmission systems. The central hub is assumed to be the car navigation system. For general vehicles, the maximum communication length is estimated to be 4m. Because 5GHz communication systems are suitable for big vehicles like trucks, this paper targets the 4m wireless harness performance since this will be the most common application.

The wireless harness is a plastic hose with conductive coating. Our previous research examined a wireless harness with large radius ($\phi=40\text{mm}$) and proved a possibility of low loss and highly reliable transmission over wireless harness. Given the severe space constraints of modern vehicles, 40mm radius hose is not so practical and this paper examines radio hoses with smaller radius.

Figure 2 shows trial small radius wireless hose appearance and its cross section view. Its outside diameter is 6.5mm and inside diameter is 6mm. The hose made of polyolefin has a relative permittivity of 2.3-2.4. As shown in Figure 3, its transmission loss is 16.5dB smaller than non-coated hose. This shows that wireless harness is suitable for establishing practical communication channels. Table I compares the performance of this small radius hose with the performance of the large radius hose measured in previous work [1]. In table I, the delay spread is also improved by metal coating. As for the latency, we do expect much shorter than the current wired harness since the transmission rates of the proposed systems are 10^4 times faster than the current one, however, we need to wait for further detailed analysis.

In order to clarify wireless harness communication capacity, we measured transmission losses and delay spreads of 1, 0.5, and 0.25m hoses. The measurement system is shown in Figure 4 and results are shown in Figure 5. The transmission loss is calculated by (1)

$$\begin{aligned} \text{Total transmission loss} = & \\ & \text{Transmission loss constant} \times \text{Hose length} \\ & + \text{Antenna conversion loss} \times 2 \end{aligned} \quad (1)$$

From (1) and Table II, the transmission constant is 8.2dB/m, antenna conversion loss is 3.3dB, and the estimated transmission loss is 39.4dB. By assuming the sensitivity of a 3Gbps receiver is -55dBm at 60GHz (typical sensitivity defined by IEEE802.11ad standard), and maximum transmission power is 10dBm (typical transmission power of Transmitter modules in unlicensed bands), the margin of the small radius wireless harness for 3Gbps communication is around 35dB. In IEEE 11ad standards, multiple communication speed is supported. The assumed systems are based on IEEE 802.11.ad standard which supports High (more than 3Gbps) and low (1Mbps) speed communication. In these standards, the communication is possible if the received signal level is higher than the receiver's sensitivity.

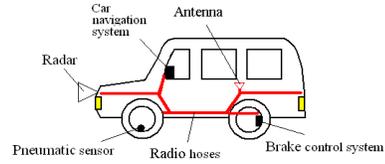


Fig.1 Wireless harness for vehicles

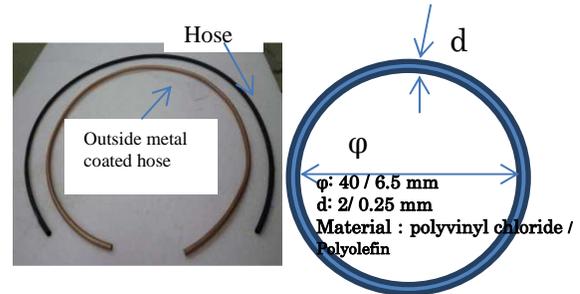


Fig.2. Metal coated hose

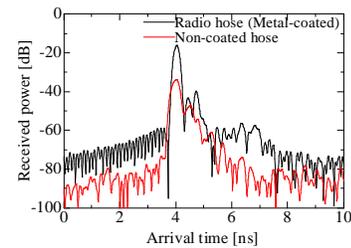


Fig.3 Transmission performance of wireless harness

TABLE I. RADIUS COMPARISON OF WIRELESS HARNESS

	Transmission Loss@1m	Delay spread@1m
Small radius($\phi=6.5\text{mm}$) wireless harness	14.8	0.08
Large radius($\phi=40.0\text{mm}$) wireless harness	31.3	0.16

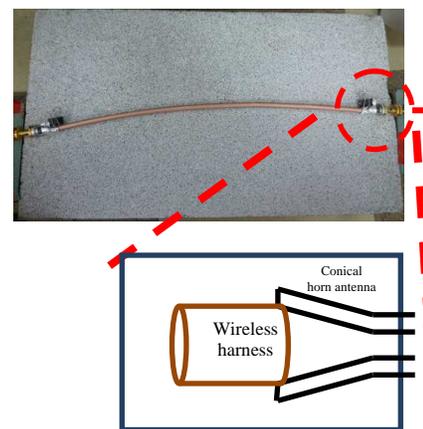


Fig.4 Measurement setting

Since IEEE11.ad standard supports high and low speed communications and our proposed systems work in these standards, the proposed systems cover from low speed to high speed communications.

III. RELIABILITY OF WIRELESS HARNESS

A. Immunity to engine noise

In order to clarify interference mitigation performance of the small diameter hose to engine noise, penetration loss of radio hose has been measured. Hose material is Polyvinyl chloride for this test and 2 horn antennas were set facing each other and the wireless hose was set on between them. For comparison, metal coated and non-coated hose penetration loss was measured. Measured results are shown in Figure 5. Non-coated hose penetration loss is matched with the free space loss, while the coated hose had 30dB higher penetration loss. Since intra-car noise in the microwave band and millimeter wave band is small, the 30dB penetration loss shows the proposed wireless harness has good noise immunity and can realize highly reliable communications

B. Communications continuity against radio hose cut and pinching

Our previous study found that the hose bend effect was less than 4dB loss increase, i.e., negligible for 3Gbps communications. This paper examines hose cut and pinching.

The definition of cut length is shown in Figure 6. The cut loss for lengths of 0, 1, 2, 4, and 5cm were measured as shown in Figure 7. The loss increases with cut length; it is 15dB for 5cm cut length.

Pinch depth is defined in Figure 8. Transmission loss was measured for pinch depths of 0, 2, 4, and 6mm as shown in Figure 9. The loss increases with pinch depth and is less than 20dB for depth of 6 mm – 100 % pinched case. Since wireless harness has 30dB margin for 3Gbps communications as shown in the previous section, the wireless harness remains operational even if radio hose is cut and separated by 5 cm or radio hose is pinched virtually flat.

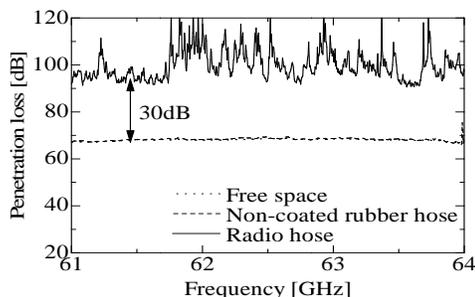


Fig.5 Penetration loss performance of radio hose

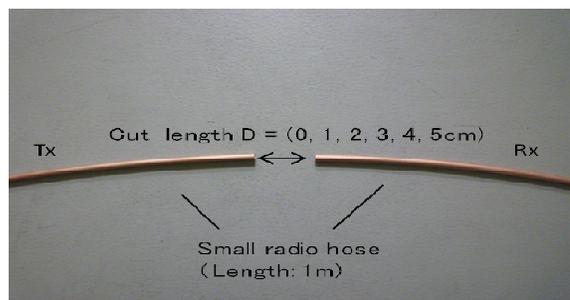


Fig.6 Cut length definition

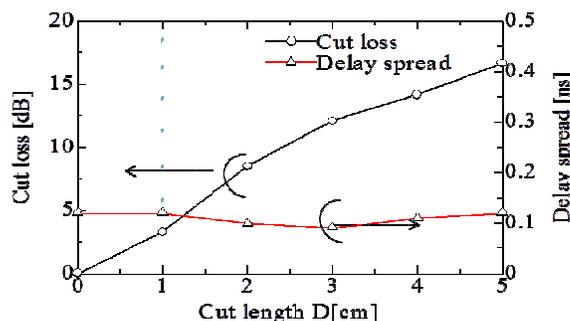


Fig.7 Measured cut loss results

TABLE II. LOSS AND DELAY SPREAD VS. HOSE LENGTH

Hose length [m]	Transmission loss [dB]	Delay spread [ns]
1	14.8	0.08
0.5	10.7	0.1
0.25	9.4	0.2

Radio hose
length: 1m,
Outer diameter: 6.5mm
Inner diameter: 6mm

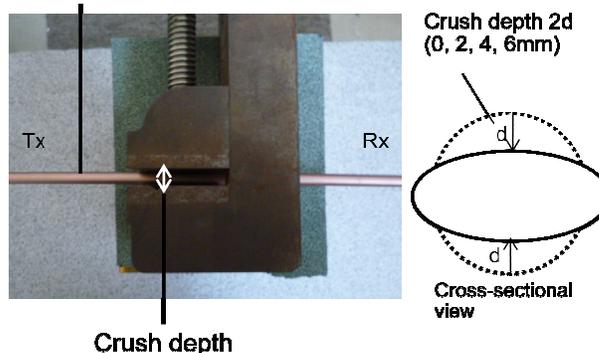


Fig.8 Pinch depth definition

IV. MULTI CHANNEL COMMUNICATIONS

Wireless harness communication systems can realize multi point communications. A typical system configuration is shown in Figure 10. Tx and Rx1 antennas are conical horn antennas with 13dBi gain with 45HPBW while Rx2 antenna is a monopole antenna with a 1.6dBi Omnidirectional gain. The monopole antenna consists of semirigid cable. Tx and Rx1 use circular polarization while Rx2 uses vertical polarization. The possible situations are Sub Length Communication (SLC) which connects the inserted antenna to the inserted antennas, and Whole Length Communication (WLC) where hose ends are directly connected. SLC measurement results are shown in Table III. SLC transmission loss at 4m is 54.5dB by (1). When transmission power is 10dBm, received power at SLC antenna is -44.5dBm which is higher than receiver sensitivity for 3Gbps communication in 60GHz band. WLC measurement results are shown in Table IV. The increase in WLC transmission loss due to by SLC antenna insertion is less than 0.3dBe. Thus, WLC still has 30dB margin for 3Gbps communication in the 60GHz band.

V. CONCLUSION

This paper proposed a wireless harness system consisting of metal-coated plastic hose (6.5 mm OD) to replace wire harness in automobiles. Measurements showed that the proposed system can realize low loss, small delay spread, and high reliability intra-car communication systems. The proposed system has the following features. Transmission loss of a 4m wireless harness is 39dB, 41dB smaller than free space loss. For the typical receiver sensitivity of 3Gbps communication, the system margin is around 25dB. The delay spread under the condition of hose cut or pinch is less than 0.3ns and loss is smaller than the margin. In multi channel transmission, the loss increase is less than 19dB, which again is smaller than wireless harness communication margin.

These measurement results confirm that the proposed wireless harness is a good candidate for future, high efficiency, intra-car communication systems.

REFERENCES

[1] K. Fujita, H. Sawada, and S. Kato, "Intra-car Communications System Using Radio Hose", APMC2010, Dec. 2010, pp. 57-60.
 [2] H. Sawada, T. Tomatsu, G. Ozaki, H. Nakase, S. Kato, K. Sato, and H. Harada, "A Sixty GHz Intra-car Multi-media Communications System," Proceedings of IEEE 69th Vehicular Technology Conference, Apr. 2009, pp. 1-5.
 [3] H.Sawada, H.Nakase, K.Sato, and H.Harada, "A Sixty GHz Vehicle Area Network for Multimedia Communications," IEEE Journal on Selected Areas in Communications, Vol. 27, No. 8, Oct. 2009, pp.1500-1506.

[4] S. Kato et al., "Single Carrier Transmission for Multi Gigabit 60 GHz WPAN Systems," IEEE Journal on Selected Areas in Communications, Vol. 27, No. 8, Oct. 2009, pp 1466-1478.
 [5] M. Heddebaut, V. Deniau, and K. Adouane, "In-vehicle WLAN frequency communication characteristics", IEEE trans. Intell. Transport Syst., Vol.5, no.2, June 2004, pp. 114-121.
 [6] IEEE 802.11 Task Group, Nov. 2014, http://www.ieee802.org/11/Reports/tgp_update.htm

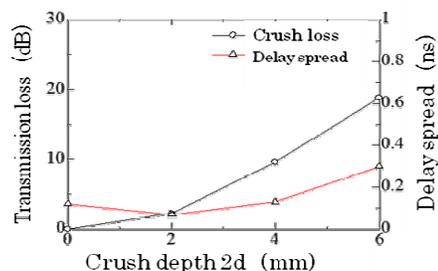


Fig.9 Measured pinch depth results.

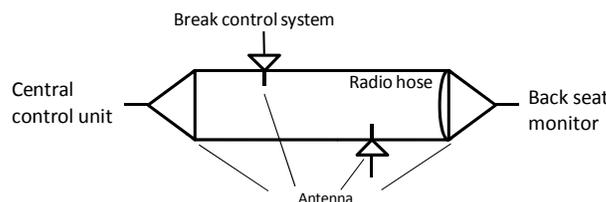


Fig.10 Multi channel communications

TABLE III. SLC TRANSMISSION PERFORMANCE

Hose length[m]	Transmission loss[dB]	Delay spread [ns]
0.25	22.6	0.32
0.5	25.4	0.27
0.75	30	0.28

TABLE IV. WLC TRANSMISSION PERFORMANCE

Hose length[m]	Transmission loss [dB]	Delay spread[ns]
0.75 (no-SLC antenna)	13.8	0.14
0.25	14.2	0.14
0.5	14	0.14
0.75	14.1	0.13

An Analysis of the Need for Dedicated Recovery Methods and Their Applicability in Wireless Sensor Networks Running the Routing Protocol for Low-Power and Lossy Networks

Anne-Lena Kampen

Bergen University College, Bergen
ITEM NTNU, Trondheim
Norway

Anne-Lena.Kampen@hib.no
anneleka@stud.ntnu.no

Knut Øvsthus

Bergen University College, Bergen
Norway

knut.ovsthus@hib.no

Øivind Kure

NTNU, Trondheim
Norway

okure@item.ntnu.no

Abstract—Wireless Sensor Networks Running (WSN) functionality depends critically on the network connectivity. The connectivity is generally determined by the node density and the nodes' transmission range. However, the applied routing protocol decides the routing path topology. A failing node may disrupt the current path topology such that dedicated recovery methods are needed to ensure a loop-free reconnection of the disconnected nodes. In this article, we estimate the probability that disconnected nodes need dedicated recovery methods in networks where the nodes are randomly located and which use RPL as routing protocol. We further calculate the success rate and overhead cost for different RPL fitted recovery protocols to better judge where the different methods should be used.

Keywords—WSN; Recovery; Disconnection; Energy; Looping.

I. INTRODUCTION

A common requirement for Wireless Sensor Networks (WSN) is marginal need for human support during operation. Hence, the networks should be able to autonomously handle common error conditions, such as loss of connectivity due to failing nodes [1]. Connectivity loss negatively affects the data throughput and may lead to network partitioning. Thus, nodes should be reconnected without any unnecessary delay. The reconnection process should further expend limited amount of energy to minimize its influence on network longevity.

Reconnection of nodes located such that several neighboring nodes are at the same routing distance from the sink as a failing next-hop node (parent) introduces insignificant delay and energy consumption. However, nodes located such that all neighbors are at a routing distance further away from the sink than a failing parent node cannot make an immediate reconnection. The reason is that the reconnection process may create routing loops if not controlled. Routing loops are created if the disconnected nodes choose their own directly, or indirectly, connected successors as new parent nodes. Dedicated global or local recovery methods are means to ensure against the formation of routing loops during the reconnection process.

Global recovery processes generally postpone reconnection of nodes in loop-prone topologies until the next global network update. Local recovery processes make nodes in the vicinity of a disconnection communicate routing information to enable fast, loop-free reconnections. The most suitable recovery method is decided by the network characteristics and the requirements of the running application.

Our contribution is twofold and relates to recovery in randomly deployed network running Routing Protocol for Low-Power and Lossy Networks (RPL) [2], which is one of the recommended protocols for WSNs. First, we present calculations and simulations to assess the need for dedicated recovery methods to reconnect disconnected nodes. The result can be used as a base to decide whether to introduce dedicated recovery management in applied networks. We further suggest one on-demand recovery method that combines the global and local approaches. The suggested method, along with two additional local recovery methods, is analyzed to better judge where the different recovery methods should be used.

The rest of the article is organized as follows: the related work is introduced in Section 2. The probability that dedicated recovery methods are needed to mend disconnection is presented in Section 3. An analysis of two local recovery methods are presented in Sections 4 and 5. The on-demand method is presented in Section 6. The methods are compared in Section 7. Section 8 comprises the conclusion.

II. RELATED WORK

Network connectivity calculations are presented in several papers, such as [3][4][5]. Zhu et al. [6] conclude that a network with satisfying coverage is connected if the communication range is twice the coverage range. Topology controls methods to maintain k-connected networks are investigated and suggested in [7][8] and [9]. Kleinroch [10] discusses network connectivity based on cited works and presents the node degree needed to achieve network connectivity. However, the analysis performed in these papers focuses on connectivity without considering the

applied routing protocol or reconnection of disconnected nodes. Our analysis is based on the functionality of the applied routing protocol.

Many of the presented recovery protocols suggest movable nodes to reconnect disconnected nodes [11][12]. Nodes are proactive moved to prohibit disconnections, or reactive moved to mend disconnections. However, we assume that the nodes location is static and the goal is to discover all alternative possible recovery paths.

All recovered paths are required to match the routing protocol's path construction method and avoid formation of routing loops. Global methods fulfil this requirement by making disconnected nodes in loop-prone positions postpone reconnection until the next global update. This approach is used in RPL [2].

The disconnected node initiates the recovery process in local recovery methods. The affected node signals its state to adjacent nodes. Depending on the recovery method used, the signaling may be relayed further on to reach nodes eligible to offer new loop-free paths to the disconnected node. Sequence numbers, as used in [13][14], is a common mean applied in such local recovery methods to discover new paths. Other local recovery processes include avoiding any duplicates in the address field during source routing [15] and caching alternative feasible successors paths in case the current route is broken [16]. The feasible successor paths are guaranteed loop free as they report a distance to the destination that is shorter than the current path from the source to the destination.

III. PROBABILITY OF DEDICATED RECOVERY

This section presents an analysis of the need for dedicated recovery methods to mend routing path disconnections. The routing protocol that is used as the basis for our analysis is RPL.

A. Short presentation of RPL

RPL is a soft state routing protocol that creates routes that are directed toward the sink. The overall topology of the routing entries creates Destination Oriented Directed Acyclic Graph (DODAG). A node's logical location in the routing graph is defined by the nodes rank and selected parent, which are two strongly interrelated properties of a node. A node calculates its own rank based on its selected parent rank and the metric-based cost-of-path between itself and its parent.

To prevent against routing loops are nodes running RPL

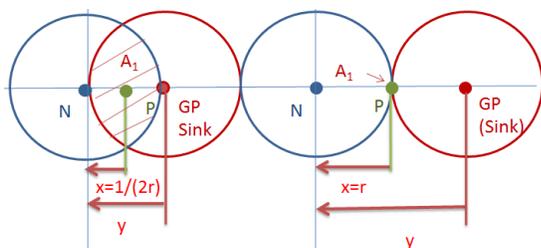


Figure 1. Node arrangement for a node at two-hop distance from the sink.

prohibited from increasing their rank in between global DODAG updates. Performing a rank increase means that nodes make a logically move away from the sink in the routing graph, an action which may results in routing loops. Global DODAG updates are initiated by the sink by distributing updated Destination Information Option messages (DIO) that flow like a wave throughout the whole network.

B. Presentation of extreme popints for probability calculations

To estimate the probability that a dedicated recovery method is needed is complex and depends on the relative location of all possible parents' next-hop node (grandparent). However, the highest and lowest probability limits may be calculated by studying the difference between dedicated recovery need for nodes located at the extreme points. The extreme location for the nodes is at the border of the routing graph. One of the extreme points is represented by the nodes located at a two-hop distance from the sink. The nodes next-to-the-leaf nodes are the highest rank nodes that may require dedicated recovery, and represent the second extreme point. All other nodes that may require dedicated recovery lie between these two borderline cases, so do their average probability.

1) Extreme location one : Two-hop distance node

Figure 1 illustrates a general node arrangement for nodes at two-hop distance from the sink. The blue dot labeled N and the blue circle is the node under consideration and its transmission distance, respectively. The red dot is the grandparent node, and its transmission range is defined by the red circle. The grandparent node is the sink as the node under consideration, N, is a node at two-hop distance from the sink. The parent node is represented by the green dot labeled P. All nodes choose parents that minimize their own rank. Thus, the node and its grandparent cannot communicate directly. Further, if the parent node dies, the node N needs to find a new parent node to maintain the path toward the sink.

According to the loop-avoiding rules of the RPL method, a node is never allowed to increase its rank unless a global update is performed. Hence, node N needs to maintain or improve its rank if the parent node in Figure 1 dies. The only way node N can improve its rank is to achieve a direct connection to the sink, which it cannot. Hence, node N needs to maintain its rank. It follows that it needs to get connected to a node that is directly connected to the sink, i.e., it must keep the sink as its grandparent node after recovery. Thus, the alternative new parent node must reside in the overlapping area of the transmission circle of the grandparent and the transmission circle of the node N (area A in the figure).

2) Extreme location two: Node next-to-the-leaf node

Figure 2 is used as a reference to calculate the permitted-area A for a node next-to-the-leaf node. The multiple circles centered at the sink represent the location for nodes at a specific hop distance from the sink. The area between the red dot representing the sink, and the inner red circle, represents the location for the one-hop nodes. The area

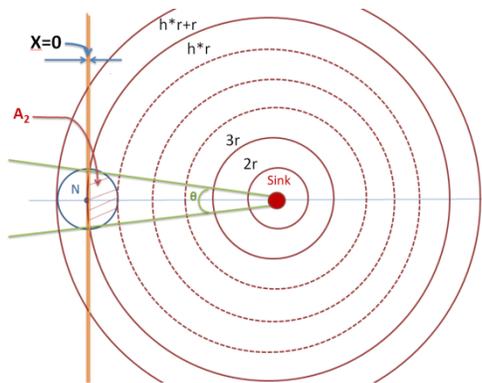


Figure 2. Disconnected node next-to-the-leaf node.

between the inner red circle and the second inner red circle represent the permitted-area for the two-hop nodes and so forth. As we are considering a node next-to-the-leaf node, we assume that the distance between the sink and the node under consideration is so far apart that the curvature of the sink's h-hop circle line cutting through the node N's circle approaches a straight line.

Figure 3 is a segment of Figure 2. The orange vertical lines in Figure 3 illustrate the sink's outer h-hop circle lines as straight lines based on the explanation above. The red shaded area named 'A' illustrate the permitted-area for a parent node of node N.

To find the probability that dedicated recovery is needed we derive the expression for the expected value of the probability that there exists more than one node inside the area A. If there is more than one node in area A, it means that there exists a recovery node after the current parent node dies. To find the wanted expectation we need an expression for the probability that there is another node in A, as well as an expression of the probability density function for location of node N.

C. Probability that there is a recovery node in area A

We assume a uniform node distribution, thus, the number of nodes in an area is given by the Poisson distribution. λ is defined to be the node density, which corresponds to the expected number of nodes in a circular shaped area with radius equal to the transmission range. All nodes have equal transmission range, r .

The probability that there is another node in A in Figures 1 and 2 is given as Prob(more than 1 node in area A | given

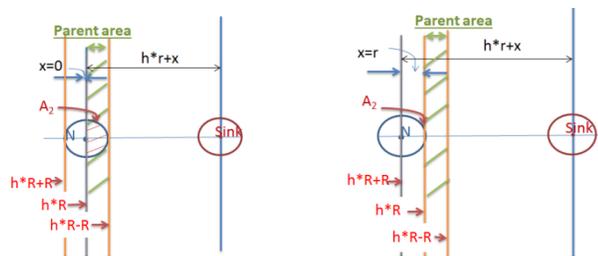


Figure 3. Node arrangement for a node far from the sink.

that there is at least 1 node in area A). $\text{Prob}(1 \text{ or more } \cap \text{ at least } 1) / \text{prob}(\text{at least } 1) = \text{Prob}(2 \text{ or more}) / \text{Prob}(\text{at least } 1)$:

$$p(\text{more than one in } A | \text{at least one in } A) = \frac{1 - P(1 \text{ node in area } A) - P(0 \text{ nodes in area } A)}{1 - P(0 \text{ nodes in area } A)} = 1 - \frac{\lambda A(x)}{e^{\lambda A(x)} - 1} \quad (1)$$

We assume that node N is at a distance y from a grandparent node, as shown in Figure 1 for the two-hop node. y is in the range $r, 2r$. The area of A_1 is symmetric around $y/2=x$. Then the area A_1 as a function of x is:

$$A_1(x) = 2r^2 \left(\cos^{-1} \left(\frac{x}{r} \right) - \sqrt{1 - \left(\frac{x}{r} \right)^2} * \frac{x}{r} \right) \quad (2)$$

According to the node next-to-the-leaf node, the permitted-area is one half of the area A_1 in (2), using $y=x$.

Notice that the area A_2 in Figure 3 is bigger than area A_1 in Figure 1 when the node N is in its closest position to the sink (left hand side of the figures). Hence, with node N in this position, the permitted-area for the recovering node next-to-the-leaf node is bigger than the permitted-area for the two-hop recovering node.

D. Node next-to-the-leaf-node calculations

We use Figure 2 as reference to calculate the probability density function for the node N location. The sector θ in Figure 2 defines the sector where a recovery parent node N may be located. The cumulative distribution function for the node N's location and the probability density function of node N's location are respectively given by:

$$F_1(y) = \frac{\frac{\pi((h^*r)+y)^2}{\theta} \cdot \frac{\pi(h^*r)^2}{\theta}}{\frac{\pi(h^*r+r)^2}{\theta} \cdot \frac{\pi(h^*r)^2}{\theta}} = \frac{y(2hr+y)}{(1+2h)r^2} \quad (3)$$

$$f_1(y) = \frac{2(hr+y)}{(1+2h)r^2} \quad (4)$$

Using the presented equations, we can derive the expression for the expected value of the probability that there exists a recovery node in area A_2 for the node next-to-the-leaf node. The expression is found by combining (1) with (4) and (2). The expected value of $P(\text{there exist a recovery node inside area } A_2)$ is:

$$E[P(y)] = \int_r^{2r} \{ (P(\text{more than } 1 \text{ in } A_2(y) | \text{at least } 1 \text{ in } A_2(y))) f(y) \} dy = \int_r^{2r} \left\{ \left(1 - \frac{\lambda A_2(y)}{e^{\lambda A_2(y)} - 1} \right) * \frac{2(hr+y)}{(1+2h)r^2} \right\} dy \quad (5)$$

E. Two-hop node calculations

We will now derive the expectation of the probability that there exists a recovery node in area A_1 for the two-hop node, ref. Figure 1. First we need the expression for the probability density function for the location of node N. This is found using Figure 4. According to Figure 4 are the expression for the cumulative distribution and probability density function of the node N given by:

$$F_2(y) = \frac{\frac{\pi(y)^2}{\theta} \cdot \frac{\pi r^2}{\theta}}{\frac{\pi(2r)^2}{\theta} \cdot \frac{\pi r^2}{\theta}} = \frac{y^2 \cdot r^2}{3r^2} \quad (6)$$

$$f_2(y) = \frac{2y}{3r^2} \quad (7)$$

Combining (1), (2) and (7), gives the following expected value of the probability for the two-hop node:

$$E[P(y)] = \int_r^{2r} \left\{ \left(1 - \frac{\lambda A_1(y)}{e^{\lambda A_1(y)} - 1} \right) \left(\frac{2y}{3r^2} \right) \right\} dy \quad (8)$$

To summarize; the expression for the expected value of the probability that there exist recovery nodes for a node next-to-the-leaf node given by (5), and the corresponding expression for a two-hop node given by (8). In other words, this is the probability that dedicated recovery methods are *not* needed to mend routing path disconnections. The calculations performed are based on numerically calculations of the equations.

F. Simulations - dedicated recovery

Simulations are conducted in Java to validate the calculated results for the expected value of the probability that dedicated recovery methods are not needed to mend routing path disconnections.

The simulation for the two-hop node is initialized by placing a node N in a fixed position. The next node is randomly paced in the donut shaped area between r and $2*r$ from the fixed node. The second node becomes the fixed node's grandparent (the sink). A varying number of nodes are subsequently randomly distributed with average density λ inside the simulation area.

The sought probability for the two-hop node is estimated

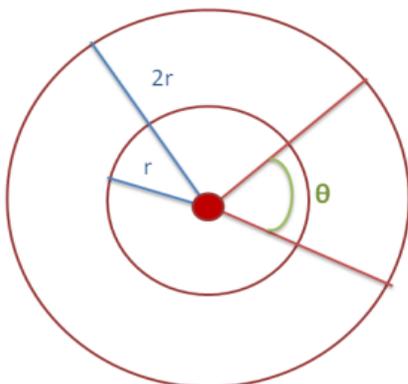


Figure 4. The sector where a possible node N may be located.

based on the percentage of the simulation runs resulting in two or more nodes located inside the overlapping area defined by the node's transmission range and the grandparent's transmission range. Dedicated recovery is needed if the number of nodes in the overlapping area is less than two. The reason is that if there is only one node in the overlapping area, it is definitely the parent node and there are no nodes left in the area when it dies. Simulation runs resulting in zero nodes inside the area is discarded. The simulated result is averaged over 1000 runs for each node density.

According to the node next-to-the-leaf node the simulations is performed by placing two nodes at a distance $h*r + x$ apart, in the same manner as for the two-hop node. $0 < x < r$. The two nodes represent the node under consideration, N, and the sink. The number of nodes located both inside node N's transmission range and inside a radius of $h*r$ from the sink are counted. The investigated probability is further performed following the same procedure as when calculating the two-hop node probability.

G. Results – dedicated recovery

In this section, we present and discuss the simulated and calculated results of the expected value of the probability that dedicated recovery methods are *not* needed to mend routing paths. The curves in Figure 5 show the expected value of the probability for the extreme points, i.e., the lowest and highest average probability values. The dashed red curve represents the simulated probability of the nodes at two-hop distance from the sink, and the blue curve shows the calculated probability for the two-hop node. The red and the dashed green curve show respectively the calculated and simulated results for the node next-to-the-leaf node. The curvature of the simulation results conform to the curvature of the calculated results validating each other.

The curves in Figure 5 show that the disconnected nodes next-to-the-leaf nodes have lower need for dedicated recovery than the two-hop nodes.

The difference between the curves in Figure 5 is caused by the unequal characteristics of the two extreme points in the routing graph topology. Both the probability density function for the node N's location and the permitted-area are different in the two extreme points. The probability density function for the location of a leaf node N approaches a uniform distribution as the number of hop gets high. This is illustrated in Figure 6. The reason for the uniform distribution is the straightening of the $h*r$ curvature and the related small difference between the circumference of the $h*r$ and $h*r+r$ circle when h is high. This is easily seen looking at Figure 3. On the contrary, the probability density for the location of the two-hop node N is increasing toward the outer circumference. The reason is the increased circumference which increases available deployment area for the node N. This can be observed in Figure 4. Figure 7 shows how the permitted-area varies with distance between the node N and the sink node. The figure shows that the area decreases with increased distance, and also illustrates the slightly bigger permitted-area of the node next-to-the-leaf nodes. Combining the information given in Figures 6 and 7,

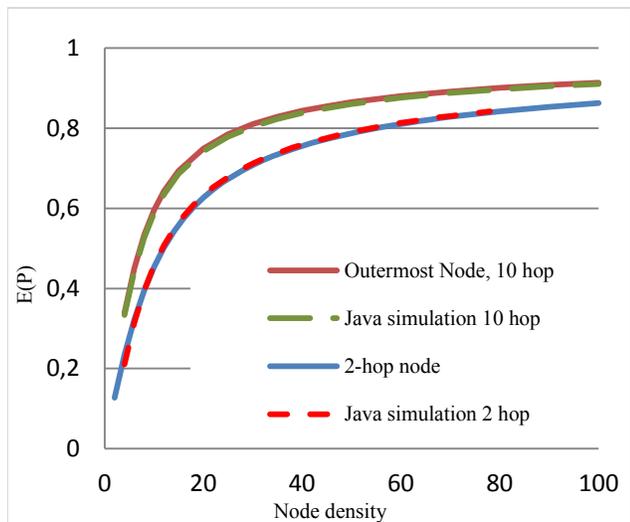


Figure 5. The expected value of the probability that dedicated recovery methods are not needed to mend routing path disconnections.

shows that the location probability of the two-hop node favors the smallest permitted-area size, while the node next-to-the-leaf node gives equal priority to all permitted-area sizes. A smaller area means that the probability that it contain more than one node is lower. Thus, the probability that it contains a candidate recovery parent node is lower.

As expected, and illustrated by the graphs in Figure 5, is the need for a dedicated recovery method decreasing with increased node density. The reason is simply that the probability that more than one node is located inside a defined area increases with node density. However, the probability never reaches 1 although the node density gets high. The reason is due to the explanation given related to Figures 6 and 7: the permitted-area for the recovery nodes is very small for some of the locations of node N. Hence, there is always a probability that some nodes do not have available recovery nodes.

The graphs in Figure 5 show that if all the nodes in a network are required to stay connected some kind of special repair method is needed. According to Takagi and Kleinrock [10], eight is the magic number of neighbors regarding network throughput, and four neighbors are needed to maintain a connected network. Hence, we may assume that an average well design network has a node density between 8 and 20. We define the node density as the number of nodes inside a circular area with radius equal to the transmission range, thus is equals the number of neighbors plus one. The graphs show that the probability that dedicated recovery is not needed is between 40% and 75% at node densities between 8 and 20. Hence, between 25% and 60% of the disconnected nodes needs a dedicated recovery method to get properly reconnected.

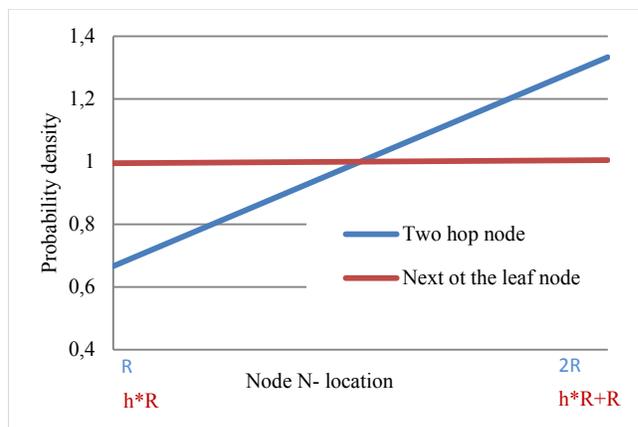


Figure 6. Probability density function of node N's location.

IV. ANALYSIS OF GUO ET AL.'S METHOD

This section presents an analysis of a local recovery method suggested by Guo et al. [17]. The method forces intermediate nodes on potential recovery paths to adjust their rank to make the path feasible for a disconnected node.

Their method [17] is activated and runs as follows. A poisoned node, which is a node that needs dedicated recovery to reconnect, initiates the recovery process by broadcasting a request. The request is further relayed to the receivers' parent nodes. The process lasts until the requests reach a node with better rank than the requesting node. Receivers with better rank than the requesting node generate a reply and forward it toward the requesting node using the same path as the associated request. The nodes along the path adjust their rank such that a new, valid path for the requesting node is made.

However, the method is not able to find a new valid path for all kind of topologies. The reason is twofold. Requests received from a parent node are silently discarded. Hence, paths pointing through child nodes are never found. In addition, a race condition occurs when siblings of a dying parent node simultaneous enter poisoning state. Nodes with pending requests silently discard received requests. The result may be that paths pointing toward sibling nodes remain undiscovered.

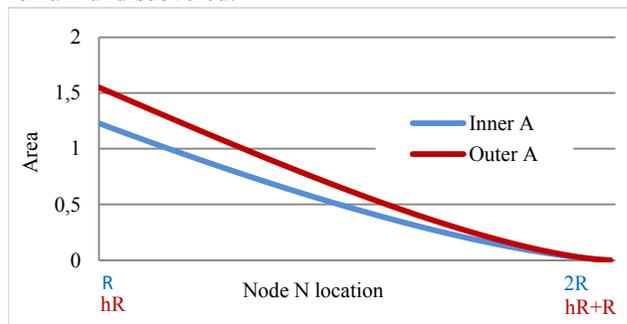


Figure 7. Permitted-area for recovery node.

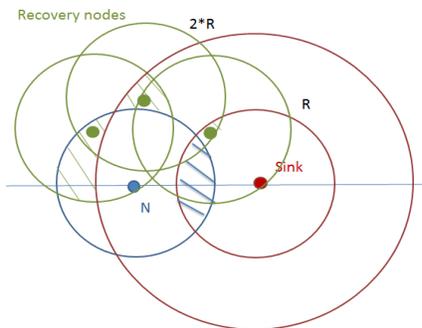


Figure 8. Recovery path goes through higher level node.

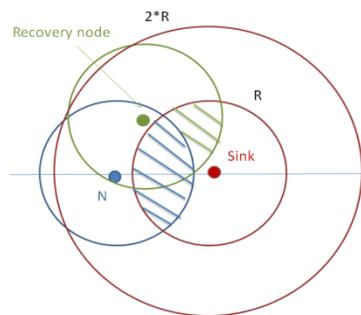


Figure 9. Recovery path goes through sibling node.

A. Simulation of Guo et al.'s method

Based on the layout in Figures 8 and 9, we simulate the probability that a poisoned node running Guo et al.'s method discovers a new valid path. The red dot is the sink node. The area between the sink and the inner red circle represents the localization of one-hop (rank one) nodes, the two-hop (rank two) nodes are located between the red circles, and so forth. The blue node N represents a node that is poisoned and need dedicated recovery to get reconnected to the DODAG. The blue circle represent node N's transmission range. The figures illustrate the two scenarios that make the [17] method local recovery succeed.

Figure 9 illustrates the scenario that a reachable node with equal rank as the poisoned node, has a parent node outside the poisoned node's transmission range. Expressed according to the figure, it means that there exists an equal-rank green node in the overlapping area made of the blue circle and the donut shaped area made of the red circles. Further, this equal-rank-node has a parent inside the green shaded area.

The other scenario that makes [17] succeed is if a lower-level node of N (node lying outside the outer red circle) have a path toward the sink that does not include N, or N's parent. According to Figure 8, it means that the node N has a neighbor in the leftmost green shaded area. This neighbor has a parent in the upper green shaded area, which further has a parent in the rightmost green shaded area.

The simulation is implemented in Java. A varying number of nodes are randomly deployed inside a circle shaped area with radius that is three times the transmission range. All the nodes are supposed to have equal transmission range. 5000 runs with different node densities, and node N locations, are performed. The numbers of runs which satisfy one or both of the scenarios discussed above, and indicated in Figures 8 and 9, are counted. This number is normalized by the number of runs where recovery is needed to reconnect N, i.e., the number of runs where only one node reside inside the blue shaded area.

B. Results on Guo et al.'s method

Figure 10 shows the simulated probability that a poisoned node gets reconnected after performing Guo et al.'s local recovery procedure. The x-axis shows the node density, λ . As expected is the success probability increasing

rapidly with node density. When the node density is 4 the probability is about 40%. When the node density approaches 20, the probability approaches 100%. Thus, the approach of [17] works best in high density networks. The probability that a dedicated recovery method is needed to reconnect disconnected nodes is highest at low node densities; Figure 5. Hence, the lowest probability of solving the problem is in the scenarios where the problem is most likely to occur.

V. ANALYSIS OF THE ACK LOCAL RECOVERY METHOD

This section presents a local recovery method that is based on reliable poisoning of successors (sub-DAG) nodes. We call this method the ACK-method. Reliability is achieved by letting nodes be aware of their children, and make all children acknowledge reception of poisoned information transmitted by the poisoned node. Information about children is achieved by making all nodes inform about their parents in regular transmitted DIO messages.

Receiving ACK form all children enables the poisoned node to increase rank to reconnect to the DODAG. No loop is created because sub-DAG nodes with no alternative recovery parent inform about their poisoned state in the transmitted ACK messages.

The ACK method will mend disconnections as long as the poisoned node receives ACK messages from all its children. Hence, the probability of success using this method depends on probability of successful reception of transmitted packets. We name the probability of successful transmission P_{rec} . A poisoned message is retransmitted once if the

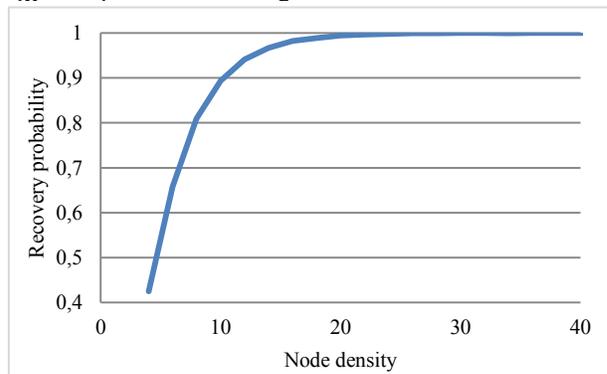


Figure 10. Probability that Guo et al.'s method succeed

poisoned node does not receive ACK from all children. Thus, recovery will not succeed if the two poisoned messages are lost, or if the ACK message is lost.

Thus, assuming one child gives the following success probability of the ACK method: $P(\text{ACK succeed one child}) = p(\text{First Poisoning messages succeed}) * P(\text{ACK succeed}) + p(\text{First Poisoning messages do not succeed}) * p(\text{Second Poisoning messages succeed}) * P(\text{ACK succeed})$:

$$P(\text{ACK succeed one child}) = P_{rec} * P_{rec} + [1 - P_{rec}]P_{rec}P_{rec} \quad (9)$$

Assuming that either all or none of the child nodes receive the poisoning message, the expression for the ACK success probability for c child becomes:

$P(\text{ACK succeed for } c \text{ child}) = p(\text{First Poisoning messages succeed}) * P(\text{ACK succeed})^c + p(\text{First Poisoning messages do not succeed}) * p(\text{Second Poisoning messages succeed}) * P(\text{ACK succeed})^c$:

$$P(\text{ACK succeed } c \text{ child}) = P_{rec} * P_{rec}^c + [1 - P_{rec}]P_{rec}P_{rec}^c \quad (10)$$

A. Results ACK method

The probability that the ACK method succeeds is shown in Figure 11. The x-axis represents the probability that a transmitted packet is received. The blue graph illustrates the success probability for a disconnected node with one child, and the red graph shows the success probability for a disconnected node with five child nodes. As expected is the success probability increasing with increased probability of receiving transmitted messages and with reduced number of child nodes.

VI. ON-DEMAND METHOD

In this section, we present our proposed combination of local and global recovery that may be used to guarantee recovery for all node densities while keeping the network energy consumption as low as possible. We call this method the on-demand method.

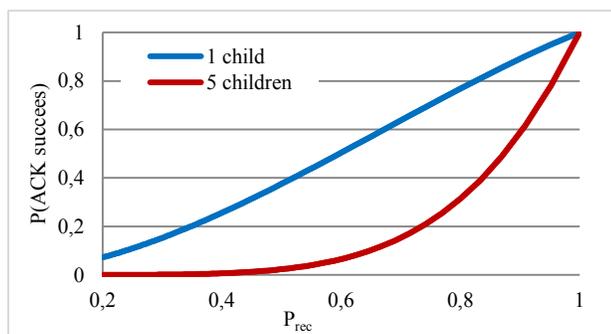


Figure 11. Probability that ACK succeed vs. Prec.

The method functions as follows. A node entering the poisoning state broadcasts an increase-sequence-number request. The request is broadcasted throughout the whole network, which means that it will eventually reach the sink if there exists a path between the poisoned node and the sink. When the sink receives the request it initiates the global recovery algorithm.

The message overhead, hence the energy cost, of running one iteration of the on-demand method is about twice the cost of running one iteration of periodic global update. The reason is that the request is broadcasted throughout the whole network in a manner similar to DIO message during global network update.

Running the update only when nodes are poisoned means that the network wide broadcast is run only when needed, and no periodic global network update is in fact ever needed.

The recovery time using the on-demand method decrease compared to the periodic approach, as the global update is run immediately after the request reach the sink.

VII. ANALYSIS OF RECOVERY OVERHEAD COST

We perform calculations to estimate the overhead cost difference between Guo et al.'s, ACK, on-demand, and periodic recovery methods. The overhead is calculated as the total number of transmitted and received management messages during the recovery process. The overhead cost is proportional to the network energy consumed, which should be as small as possible to limit the recovery process' impact on the network lifetime.

The Guo et al.'s method overhead relates to the transmission of requests and replies. We assume a uniformly distributed network where the average number of neighbors is n . The fraction of neighbors forwarding the request is α , and the fraction of neighbors that replies the request is β . Thus, the number of nodes transmitting the request is $(1 + \alpha n)$. The digit 1 in the expression refers to the poisoning node initiating the request transmission. Each transmitted request is, on the average, received by n neighboring nodes. Hence, the total request cost is $(1 + \alpha n) * n$. Further, we assume that the request and replies are relayed once. The reply is answered by βn nodes and relayed once by βn nodes. Each transmission is received by n nodes. Hence, the reply cost is $(2\beta n) * n$, and the total overhead cost becomes:

$$Guo \text{ et al. } 's_{overhead} = 2 * \beta n^2 + (1 + \alpha n) * n \quad (11)$$

The ACK overhead cost relates to the poisoning message and ACK message transmission. The poisoned node transmits a poisoning message, which is received by all neighboring node giving a total cost of $1+n$. Further, we assume that the fraction of neighbors that are child of the poisoning node is Δ . Hence, the ACK messages is transmitted by Δn nodes and all messages are received by n nodes, giving a cost of $\Delta n * n$. In addition, all nodes in the neighborhood transmit a DIO concluding the recovery process. The DIO is received by all neighbors giving a cost

of $n*n$. Hence, the total cost of the ACK recovery process is:

$$ACK_{overhead} = (1 + \Delta) * n^2 + n + 1 \quad (12)$$

The on-demand overhead relates to the total number of nodes in the network, N , transmitting requests and DIO messages. All transmitted messages are received by the average number of neighbors.

$$(On - Demand)_{overhead} = 2 * Nn \quad (13)$$

The overhead according to one run of global recovery relates to all nodes N transmitting DIO messages, which are received by the average number of neighbors.

$$OneRunGlobalRecovery_{overhead} = N * n \quad (14)$$

A. Results comparing methods

Figure 12 shows the overhead for the different methods using $\Delta=0.4$ (share of neighbors being child of poisoned node), $\alpha=0.6$ (share of neighbors relaying request) and $\beta=0.2$ (share of neighbors replying request). The value of Δ is chosen looking at the right-hand side of Figure 1: We assume that approximately all nodes located inside an area about the same size as the red shaded area are children of a node N . The rest of the neighboring nodes relay the requests received, hence the value of $\alpha = 1 - \Delta$. β is chosen assuming that only a small fraction of nodes receiving the relayed request are able to answer. These values clearly changes according to the network topology. However, the mutual relation between the parameters will generally remain unchanged. Hence, the information given by the figure is valuable. The total number of nodes in the network is 100.

There is a big difference between the local recovery approach methods' overhead and the periodic update, as shown in Figure 12. However, as the local recovery methods cannot guarantee reconnection they require periodic global update to coexist to guarantee full network connectivity.

The figure shows a substantial overhead cost difference between on-demand method and periodic update. However, the great advantage of using the on-demand method is that the method is only triggered by a disconnection. Thus, the overhead cost will be lower than the periodic update method in network with low disconnection probability.

The significance of our findings is the statistical analysis of the need for dedicated recovery presented in Section 3, in addition to the overhead cost for the recovery methods presented in this Section. The statistical analysis showed that dedicated recovery is needed especially in low density networks. In addition, nodes in the vicinity of the sink are most vulnerable and require dedicated recovery. These nodes are critical for sustaining network connectivity. Our overhead cost findings show that cost analysis should be performed as part of real networks' deployment methodology to select an appropriate recovery method. The selected method should either be the on-demand method, or

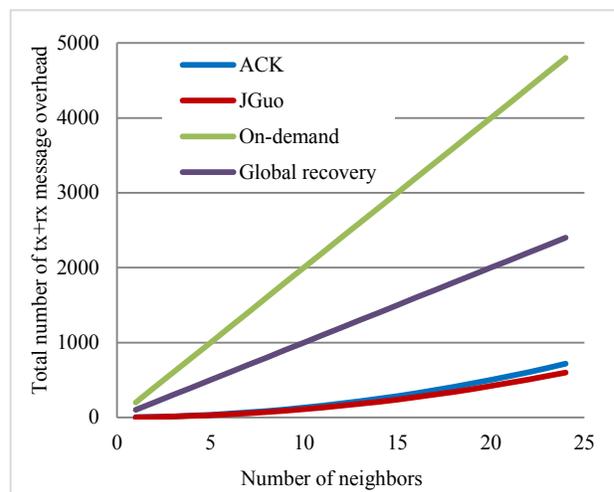


Figure 12. TX+RX message overhead. Number of nodes is 100.

adapting of the periodic global update frequency, as these methods are reliable.

Our results can be used to perform overhead cost calculations for network design. The overhead associated with the on-demand method is calculated combining (13) with the expected number of nodes that need dedicated recovery during a time span. The expected number of nodes is found combining information about the total number of nodes, the nodes failure probabilities, and the probability for disconnected nodes' recovery need, presented in Figure 5.

The overhead cost of adapting the global update frequency is calculated combining (14) with the recovery delay requirement. The delay requirement decides the network update frequency.

VIII. CONCLUSION

Disconnections in WSNs need to be resolved to sustain total network availability and avoid destructive data loss. Whether disconnections needs dedicated methods to regain connectivity depends on the topology in the vicinity of the disconnection.

In this article, we calculated and simulated the probability that dedicated recovery methods are needed to reconnect disconnected nodes in randomly deployed networks. The findings are that dedicated methods are needed in 25% to 60% of the cases when a node is disconnected. These findings demonstrate the significance of including dedicated recovery methods as a part of the network management in real scenarios where the network's availability is crucial. If dedicated recovery is not included, the periodic global update frequency should be adjusted according to the networks' recovery delay requirements. In addition, the findings demonstrate that increased node density may be used as deployment methodology to improve connectivity stability in critical areas of a network.

The findings further show that disconnected nodes close to the sink most often need dedicated recovery. These nodes are critical to sustain network connectivity. Hence, it may be

wise to adapt the recovery method according to routing graph location.

The failure frequency increases with network size assuming equal failure probability for the nodes. Hence, the periodic update frequency has to increase with network size in network without any dedicated recovery method. However, increased update frequency increases the nodes' energy consumption causing reduced network lifetime.

Introducing a dedicated recovery method may reduce the load caused by periodically updates. In this article, we calculate the overhead and success rate for two local recovery methods, and one suggested global on-demand recovery method. The two local recovery methods have lowest overhead, but they cannot guarantee reconnection success. The global on-demand method is reliable as reconnections are established if possible. But, it has high overhead. However, using an unreliable recovery method that cannot guarantee connectivity requires a simultaneous periodic global update mechanism to assure the total network connectivity, while a reliable recovery method makes periodic global updates superfluous.

Thus, networks requiring reliable network connectivity should either include an on-demand recovery method, or adjust the global network update frequency. The on-demand method may greatly reduce the long-term network energy consumption. Overhead cost analysis presented in this article may be used for real scenarios to choose between the two methods.

REFERENCES

- [1] V. C. Gungor and G. P. Hancke, "Industrial Wireless Sensor Networks: Challenges,," IEEE Transactions On Industrial Electronics, Oct. 2009, pp. 4258-4265.
- [2] T. Winter, et. al., "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks," RFC 6550, Mar. 2012.
- [3] W. Jia, Y. Fu, and J. Wang, "Analysis of Connectivity for Sensor Networks Using Geometrical Probability," Springer, Lecture Notes in Computer Science, Aug. 2004, pp. 601-611.
- [4] Y. Wang, B. M. Kelly, and X. Li, "On the Network Connectivity of Wireless Sensor Networks following a Random and Non-uniform Distribution," IEEE, Sixth International Workshop on Selected Topics in Mobile and Wireless Computing, Oct. 2013, pp. 69-74.
- [5] X. Xingt, G. Wangt, J. Wu, and J. Li, "Square Region-Based Coverage and Connectivity Probability Model in Wireless Sensor Networks," IEEE, International Conference on Collaborate Computing, Nov. 2009, pp. 1-8.
- [6] C. Zhu, C. Zheng, L. Shu, and G. Han, "A Survey on Coverage and Connectivity Issues in Wireless Sensor Networks," ELSEVIER, Journal of Network and Computer Applications, Mar. 2012, pp. 619-632.
- [7] I. Saha, et al., "Distributed Fault-Tolerant Topology Control in Static and Mobile Wireless Sensor Networks", IEEE, 2nd International Conference on Communication System Software and Middelware, Jan. 2007, pp. 1-8.
- [8] N. Li and J. Hou, "FLSS: A Fault Tolerant Topology Control Algorithm for Wireless Networks", ACM, Proceedings of the 10th annual international conference on Mobile computing and networking, MobiCom, 2004, pp. 275-286.
- [9] R. Ramanathan and R. Rosales-Hain, "Topology Control of Multihop Wireless Networks using Transmit Power Adjustment," IEEE, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, Mar. 2000, pp. 404-413.
- [10] H. Takagi and L. Kleinrock, "Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals," IEEE Transactions on communications, vol. 32, no. 3, Mar. 1984, pp. 246-257.
- [11] K. Vaidya and M. Younis, "Efficient Failure Recovery in Wireless Sensor Networks through Active Spare Designation," IEEE, International Conference on Distributed Computing in Sensor Systems Workshops, June. 2010, pp. 1-6.
- [12] M. F. Younis, S. Lee, and A. A. Abbasi, "A Localized Algorithm for Restoring Internode Connectivity in Networks of Moveable Sensors," IEEE, Transactions On Computers, Dec. 2010, pp. 1669-1682.
- [13] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC 3561, July. 2003.
- [14] J. Garcia-Luna-Aceves and H. Rangarajan, "A New Framework for Loop-Free On-Demand Routing Using Destination Sequence Numbers," IEEE, International Conference on Mobile Ad-hoc and Sensor Systems, Oct. 2004, pp. 426 - 435.
- [15] D. Jhonson, U. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," RFC 4728, Feb. 2007.
- [16] J. Raju and J. J. Garcia-Luna-Aceves, "A New Approach to On-demand Loop-Free Multipath Routing," IEEE Eight International Conference on Computer Communication and Networks, Oct. 1999, pp. 522-527.
- [17] J. Guo, C. Han, P. Orlik, and J. Zhang, "Loop-Free Routing in Low-Power and Lossy Networks," IARIA, International Conference on Sensor Technologies and Applications, Aug. 2012, pp. 59-66.

Domain Specific Modeling Language for Object Localization in Marine Observatories

Charbel Geryes Aoun^{*‡}, Iyas Alloush^{*},
Yvon kermarrec^{*}, Oussama Kassem Zein^{*}

^{*}Université européenne de Bretagne

^{*}Telecom Bretagne, Institut Mines-Telecom

^{*}UMR CNRS 6285 Lab-STICC

Bretagne, France

Email: charbel.aoun@telecom-bretagne.eu

iyas.alloush@telecom-bretagne.eu

yvon.kermarrec@telecom-bretagne.eu

oussama.zein@telecom-bretagne.eu

Joel Champeau[‡]

[‡]Ecole Nationale Supérieure de Techniques

[‡]Avancées Bretagne

Bretagne, France

Email: joel.champeau@ensta-bretagne.fr

charbel.aoun@ensta-bretagne.fr

Abstract—Marine observatories (MO) based on sensor networks provide a continuous observation of the ocean. The logical and physical components that are used in these observatories provide data exchanged environment between different devices (Smart Sensor, Data Fusion). These components provide new functionalities or services due to the stable running of this network. In this paper, we present our approach in extending the modeling languages to include new domain-specific concepts and constraints. Thus, we propose a meta-model that is used to generate a new design tool (ArchiMO). We illustrate our proposal with an example from the MO domain. Additionally, we generate the corresponding simulation code using our self-developed domain-specific model compiler. Our approach helps to reduce the complexity and time of the design activity. It provides a way to share the different viewpoints of the designers in the domain of MO.

Keywords—Underwater Object Localization; Marine Observatories.

I. INTRODUCTION

Sensor network is a group of specialized sensors with a communications infrastructure designed to monitor and record terms at various locations. MO (Based on Sensor Networks) provide new opportunities to sea surveys, such as a continuous observation of the sea [1]. Our research scope is in the first phase of a MO project: Marine e-Data Observatory Network (MeDON) [2]. MeDON contains different elements (Hydrophones, Fusion Servers, Object Localization Algorithms), and different communication protocols (e.g., REST, SOAP) [2][3]. The implementation of this information system is considered as a complex distributed system [3]. We distinguish two sources of complexity: the complexity of the system, and the design. The complexity of the system under study [2][3] is related to: (1) the architecture of the system (Distributed) which contains different elements from different sub systems (Underwater Sensor Network and the rest of the information system); (2) the interactions between the different elements of the information system and the core network that relies on standard protocols and transactions; (3) the large number of sensors (Hydrophones) and servers existing on the networks.

Our scope in MeDON project is in the design of the Smart Sensor Network and the system to localize the underwater objects. Designing complex distributed systems consumes considerable time. According to [2][3], the complexity of the

design is a result of: (1) the different domains of experience (Business Process Modeling, Information System Modeling, and the Underlying Infrastructure Modeling) that are required from the designer(s) to be able to model and describe such system; (2) distributed Software Structure of MeDON Information System (see Figure. 1) since each component (e.g., Data Fusion Server, Smart Sensor, etc.) is responsible to perform set of specific tasks.

Our global objective is to help the designers of MO to reduce: the complexity of the design and the time of the design activity. The deployment of set of sensors (Sensor Network) is an costly operation due to: the necessary equipments such as specific boats, marine cables, Sensors (Hydrophones), Data Fusion Servers, and experts in diving, etc. Additionally, we cannot ignore that the deployment operation is risky and the placement of sensors and servers should be in the right position where an error in meters may cause larger bit-error rates in the communication channel (Cables). Thus, an integration between the information system (Sensors, Servers) and the communication system (e.g., IMS) [4] is needed.

The large number of sensors that are communicating with set of fusion servers results in more complex design [5]. We consider that the time to obtain useful results from the MeDON system is the resultant of: the time of the operations of deployment and the design time. Thus, our research question is: how to improve the time of the design phase and reduce the complexity of the deployment and maintenance phase?

Our objective is to provide a design tool to the designers of MO that helps them to model their design taking into consideration reducing the time of development process, and managing the complexity.

In this paper, we propose a modeling design tool (ArchiMO) that helps to manage the complexity and prevents design modeling errors during the design time. This tool provides the designer a set of reusable graphical elements and concepts that respect ArchiMate [6] and the MO concepts. Our approach is based on the concept of domain specific modeling languages (DSMLs), which relies on Model Driven Engineers (MDE) fundamentals [7]. In order to model MO systems, we choose ArchiMate modeling language as it relies on Enterprise Architecture (EA) framework [8][9] that allows describing a

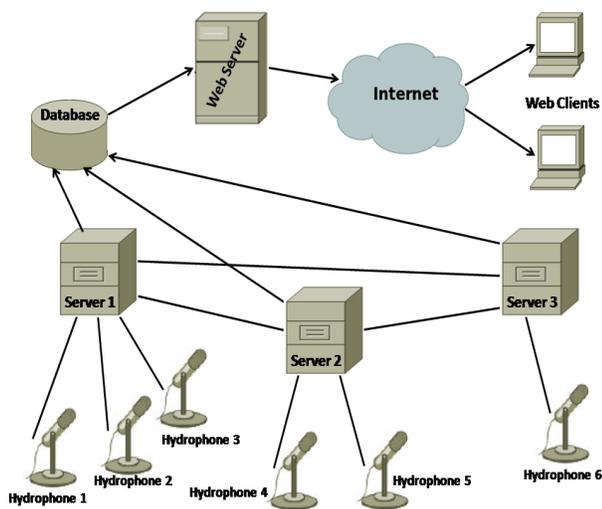


Fig. 1. Structure of MeDON- An Example: N=6, Y=3

wide range of domains [10]. We use meta-models to generate the tools that belong to different development activities using Eclipse Modeling Framework (EMF) [11].

ArchiMate is proper to model systems from the IT domain [6]. Our proposal extends the ArchiMate meta-model (Abstract and Concrete Syntax) to add new concepts and constraints of MO to ArchiMate. We add specific constraints to the grammar of the design tool according to the meta-model proposal. On one hand, a main feature of (EA) frameworks is sharing the multiple viewpoints [10]. This reduces complexity of one view to a manageable size. EA frameworks introduces interoperability issues between views and their dedicated software [10]. On the other hand, our proposed DSML is extensible, where the developers may extend it and add new concepts and standards according to the progress and needs in MO domain.

Linking our MO meta-model to the IP Multimedia Subsystem (IMS) one (proposed previously in [12] helps to integrate the different smart sensors of the sensor network to the rest of the information system through the core network [4]. We apply our design model to a model compiler to generate simulation code that runs directly in NS-3 network simulator [13].

The paper content is organized as follow: in Section II, we present the related work that is connected to the design tools. Section III presents MO project. In section IV, we present MDE fundamentals, DSMLs, ArchiMate, and our proposal meta-model for the MO/MeDON. Section V explains the abstract syntax, concrete syntax and semantics of the proposed DSML. In Section VI, we present the generated design tool and the simulation approach. In section VII, we conclude and discuss our future work.

II. RELATED WORK

In this section, we present the related work in connection with the design tools.

In relation with the concept of Architectural Description Languages (ADLs) [14] and their design tools; we are in-

terested in the following concerns that we shall specify and analyze in this section: (C1) preventing errors during design by invoking grammar or syntax of language; (C2) multiple viewpoints that are represented in the architectural description [15] since a viewpoint is a work product establishing the conventions for the construction, interpretation and use of architecture views to frame specific system concerns; (C3) extensibility of design tool; (C4) heterogeneity of components and communications; (C5) testing/execution platform.

According to the preventing errors concern, the design tool prevents errors during design activity that may be made by the designer, rather than correct them after the fact. This error prevention is available in [16][17][18]. Like in our approach, it's avoided by invoking the abstract syntax (Our Proposed Meta-Model) where we have defined and added our specific constraints and relations.

Concerning the multiple viewpoints concern, the design tool provides different viewpoints for the designers according to their specialties and domains of experience. In [16][17][18], the design tool provides only one viewpoint in order to fit software development tasks. This design tool does not provide the ability to share the design between different designers. Our approach considers this issue thanks to the different layers of EA standard that separates between perspectives.

Regarding the extensibility concern, the extension of a meta-model allows the extension of a design tool by adding new concepts and constraints to it [16][17]. It's realized in our approach by extending the ArchiMate meta-model by new elements and constraints, then generating a new design tool that contains the concrete syntax inside the palettes. These palettes contains the new added components like in [4][10].

Concerning the heterogeneity concern, the existence of different components and communications that are related to different contexts and activities. We are facing this heterogeneity in the software components and models in [16][17][18]. In our approach, we are facing this heterogeneity (e.g., Smart Sensor different than Data Fusion).

According to the execution test platform concern, the designer in [16][17][18] is not able to test and verify his models or instances on an executable platform (e.g., IMS). Relying on [4] in our approach, we are able to test our proposed model (see section VI). For example, IMS can be used to exchange messages between terminals (e.g. Smart Sensors, and Fusion Servers).

III. MARINE OBSERVATORIES

Underwater Sensor Networks that aims to environmental data acquisition will play an essential role in the development of future large data acquisition systems [19]. They allow the data to be exchanged and treated between the different devices (Servers, Sensors). On all these devices, we can have software components to treat and store the data. An example about MO is the project Marine e-Data Observatory Network (MeDON).

In this context, the designer should be able to include N acoustic sensors that are connected to the Y fusion servers as shown in (see Figure. 1). These servers treats the acoustic data

acquired by the hydrophones then diffuse them on the network. Servers store their data on the same database. The Database server provides the treated data to the web server where the configuration of a web application is done. Thus, the web server diffuses the information detected by the hydrophones such the voice of the dolphin to the web clients through a graphical interface.

IV. MODEL DRIVEN ENGINEERING (MDE) AND DOMAIN SPECIFIC MODELING LANGUAGES (DSML)

MDE [15] is "a software development method which focuses on creating and exploiting domain models. It allows the exploitation of models to simulate, estimate, understand, communicate, and produce code". MDE helps to manage complexity thanks to the modeling concept and model transformations. Modeling helps to describe the design in a high abstract way and model transformation helps to have a generated design tool.

A meta model defines by itself a language for describing a Specific Domain of interest [7]. In our approach, modeling tools follow the constraints and represent the concepts that are defined in the meta-model. It permits to instantiate large number of models that conform to it like in programming languages [20]; numerous of programs can be implemented relying on a specific programming language (e.g., C, C++, Java, etc.).

Eclipse IDE provides a powerful environment that relies on EMF which facilitates the modeling/meta-modeling activities, it supports many model transformation languages as well. Model transformations help us to generate design tools and simulation programs directly and automatically considering meta-models and model instances. Every model transformation depends on a set of rules that describe and control the transformation process. The transformation rules may map models that conform to different meta-models (on the same abstraction level), such as ATL [21], or map between different domains using one meta-model for the source model to generate texts/codes (e.g., XPAND [22]).

In our case (see Figure. 2), the input model represents the design of highly abstract level, and the meta-model is the extended ArchiMate meta-model which represents the abstract syntax [15][12]. Our code generation is an automated process that links directly the design model to the simulation scripts [13]. Thus, it helps to reduce the time of the implementations for large simulation programs, and it minimizes the implementation errors.

A. Domain-Specific Modeling Languages

Domain-Specific Modeling Languages (DSMLs) [23] enable designers from different domains and backgrounds to participate in software development tasks and to specify their own needs using domain concepts. A DSML [24] is comprised of three components: abstract syntax, concrete syntax, and semantics. The abstract syntax defines modeling concepts and their relationships. There are several kinds of concrete syntaxes: visual, XML-based, textual, etc [25]. The concrete

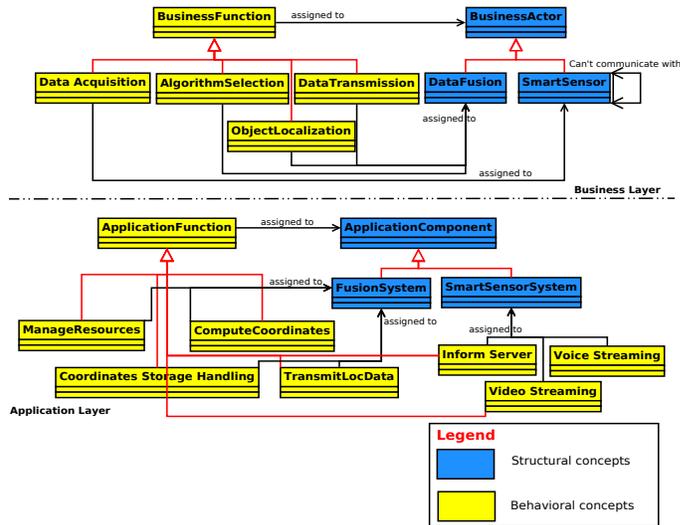


Fig. 2. Extending business and application layers of ArchiMate: proposal of MO Meta-Model

syntax is associated with a set of rules which defines the representation of the abstract syntax. Semantics describe the meaning of a model and are related to the abstract syntax. They are well-formed rules for the model and are used to constrain the concrete syntax [24].

Historically, data fusion methods were developed primarily for military applications (e.g., radars tracking a variable object) since fused data from multiple sensors provide several advantages over data from a single sensor [5]. We resume, such methodology as combining set of observations would result in an improved estimate of the target position. Concepts such information fusion and sensors networks have perforated the research and specially the military research. We distinguish different architecture for data fusion as follows [5][26]: (1) centralized fusion; (2) hierarchical fusion without feedback; (3) hierarchical fusion with feedback; (4) distributed fusion. According to our context, we have selected the most complex architecture (distributed) to model it, then simulate it in section VI. During the design activity, set of constraints and restrictions should be respected by the designer in order to model such architecture [5]. We will present them in the contribution section.

In general, errors caught during the design cycle are much less time consuming to identify and correct than those found during testing. In order to avoid errors in the design activity, we have implemented constraints that are defined in the abstract syntax of the language (Meta-Model) (see Figure. 2). The concrete syntax that is associated with these added constraints can be implemented in the design tool such as 'ArchiMO' tool in our context. This tool is generated relying on Eclipse-EMF (Tool Generation Concept thanks to Model Transformations).

Modeling languages are used to describe a system with high level of abstraction (e.g., UML 2.0) [25]. For MeDON/MO, and in relation with our objectives, we describe distributed sys-

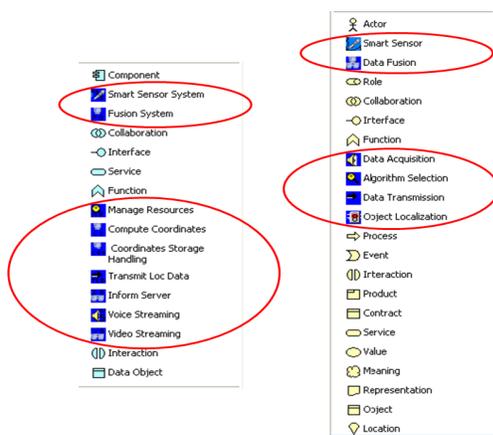


Fig. 3. Business and Application Layers (Palettes)

tems. UML is not enough to cover our needs, as it has only one layer that contains all of the concepts of the design, and these concepts are too general [27]. Thus, we selected ArchiMate modeling language that meets UML in some concepts, but it can describe the systems from IT domain and share multiple viewpoints during the design as it relies on TOGAF framework [15].

ArchiMate relies on Enterprise Architecture (EA) framework [9][15]. It decomposes the system design into three layers: business, application, and technology. In our approach, we present these layers in the following way:

- 1) Business layer: specifies the end-user functions and actors. It describes the service activities as perceived by the end-user, and the flow between them;
- 2) Application layer: specifies the functions and software components of the service. It describes the capability of the system under study, and the way of performing its tasks;
- 3) Technology layer: specifies the functions, topology, hardware elements, and signaling protocols of the underlying platform. It describes the execution platform that offers functions to be used by the functions of the application layer.

V. CONTRIBUTION

In general, a meta-model of DSL represents the concepts/operations and constraints that belong to the domain specificities (MO in our case). In this section, we present our contribution of a new meta-model (Abstract Syntax), concrete syntax, and design tool. We extend the concepts of ArchiMate meta-model to represent the domain specifications of MO. The new meta-model enables us to generate and develop design tools that are coherent with Archi [28]. They contain additional concepts, elements, constraints and relations that are specific to the MO domain and for data fusion concepts [5].

Relying on the distributed fusion architecture (DFA) in [5], our meta-model (see Figure. 2), and according to [2], we distinguish the following constraints: for SmartSensor:(I)

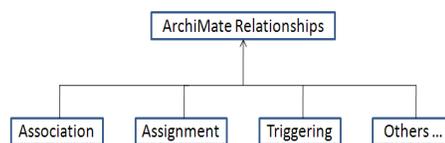


Fig. 4. ArchiMate Relationships

communication between two Smart Sensor elements is not allowed; (2) communication between Smart Sensor and Data Fusion element is allowed; (3) Smart Sensor is only allowed to be related to the Data Acquisition function. For DataFusion: (1) communication between two Data Fusion elements is allowed; (2) Data Fusion is only allowed to be related to Algorithm Selection, Data Transmission and Object Localization functions.

Like in ArchiMate [9][15] our proposed meta-model is composed of two views: one for the business layer, and another for the application layer. Regarding the technology layer, we rely on a meta-model for IMS that provides an underlying platform in [4] to integrate the information system with the core network.

For each extended concept or element, a graphical view (belonging to the concrete syntax) should be defined [10]. Our proposed concrete syntax are shown in the palettes of the business (the red circles on the right of Figure. 3) and application (the red circles on the left of Figure. 3) layers. These palettes are coherent with MO specific concepts and relations from which the designer can select, drag and drop the desired ones.

ArchiMate contains different types of relationships such association, assignment, etc (see Figure. 4). We have specialized the definition of the relationships regarding the new added concepts. In our context, we have defined the association relationship for the smart sensor according to the constraints of DFA (e.g. smart sensor could be only associated to the data fusion). Furthermore, we have defined the assignment relationship for the smart sensor according to the constraint of MO [2] (e.g., Smart Sensor could be only assigned to the Data Acquisition).

For each ArchiMate element, we can define the relationship type that are allowed with this related element. The encoding ArchiMate relationship is based on an enumeration for all the possible types. The key values are for example 'o' for *association* and 'i' for the *assignment* relationship. For the business and application layers, we have implemented the keys relative to the selected relationships and mainly the associated constraints in java code regarding to our proposed extended meta-model (see Figure. 2). This implementation is the grammar of the new proposed DSML.

In order to have a graphical view for the added constraints and elements, we have generated the design tool ArchiMO relying on eclipse EMF.

This design tool helps the designer to model the system in a highly abstract way by drag and drop the elements and

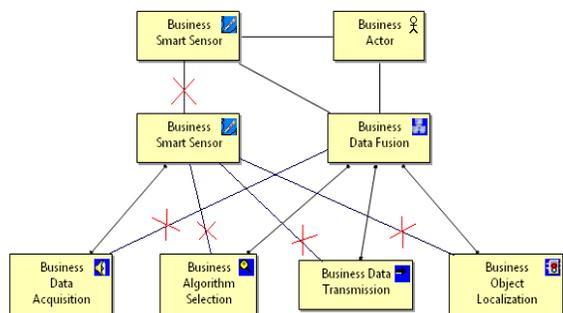


Fig. 5. Association and assignment relationships

relations from the palette. During the model edition, all the constraints specified for the MO extension are checked: (1) forbid the designer to associate two Smart Sensor elements together; (2) the designer is able to associate a Smart Sensor element to Data Fusion, Business Actor or other actors (see Figure. 5); (3) the assignment is only allowed from SmartSensor to the DataAcquisition function (see Figure. 5). Concerning the Data Fusion element: (1) the association between two Data Fusion elements is allowed; (2) the designer is able to associate Data Fusion element to Smart Sensor element (see Figure. 5); (3) the designer is able only to assign the Data Fusion to the Algorithm Selection, Data Transmission and Object Localization functions (see Figure. 5).

ArchiMO tool considers different domains of experience, each domain expert works in a specific layer (Business, Application or Technology) as the model created in section VI. Our contribution replies to the concerns that we have mentioned in II as it: (C1) prevent syntax and relation errors that can be made during the design activity; (C2) provides three layers according to each domain specificity; (C3) extends an open, standard, and classical design tool to have a specific one like ArchiMO; (C4) deploys different physical components (Sensors and Servers), and logical components such acquisition/localization algorithms.

VI. OBJECT LOCALIZATION CASE STUDY

In order to validate our proposed tool, we use it to model the application of Object Localization using the different new elements that are proposed in the meta-model (see Figure. 2). Then we apply the design model to a model compiler (see Figure. 6) that we have developed to perform some error checks and generate automatically simulation code for NS-3. This simulation code runs in NS-3 tool that is a standard and classical simulator in the networking domain.

A. Design Model

We have modeled a system that localizes an underwater object using our generated design tool ArchiMO. In order to localize this object, sensors should be connected to data fusion servers. We have applied the distributed fusion architecture (DFA) [5] for this design.

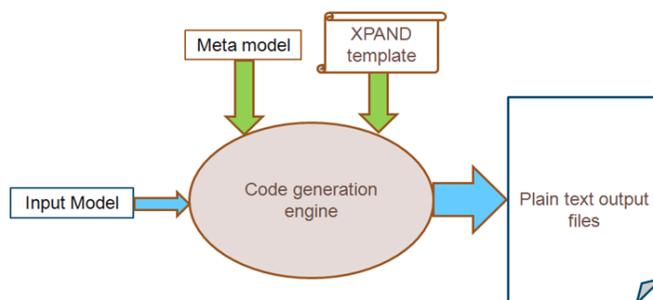


Fig. 6. The code generator workflow in XPAND language

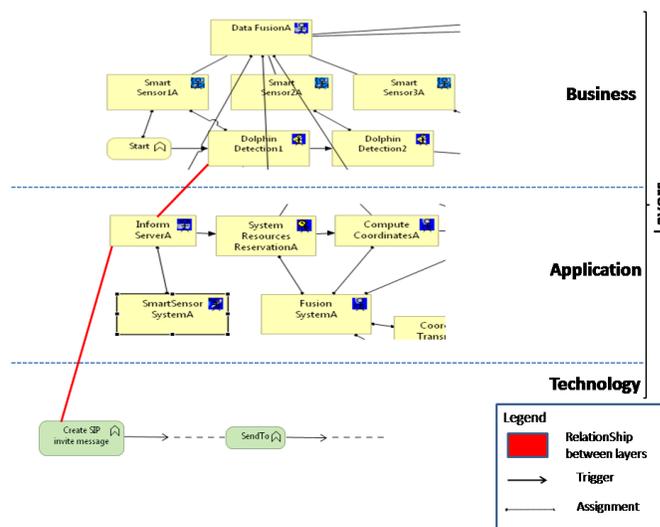


Fig. 7. Object Localization Underwater

The design model is composed of three views regarding to the layers of ArchiMate (see Figure. VI-A): Business, Application, and Technology. In Figure. VI-A, we present parts of the large model that is designed by ArchiMO. The model contains behavioral elements, in the business layer (see Figure. VI-A) shows the first activity of the smart sensor which is the dolphin detection1, etc. These activities are assigned to their proper smart sensors and these smart sensors are associated with the different data fusion servers and smart sensors that are required in the DFA [5]. Concerning the application layer, the behavioral elements are such the compute coordinates function that is triggered by the resources reservation function, and so on. Since ArchiMate allows the association between the layers, (see Figure. VI-A) shows this association. For example, the InformA Application Function aims to inform the fusion server A by the detection of an object through the smart sensor1A. Regarding the technology layer, a large series of functions are associated in it (e.g., *sendto*) to execute this application function. The *sendto* function forwards/sends a message of type SIP or Diameter from one node to another.

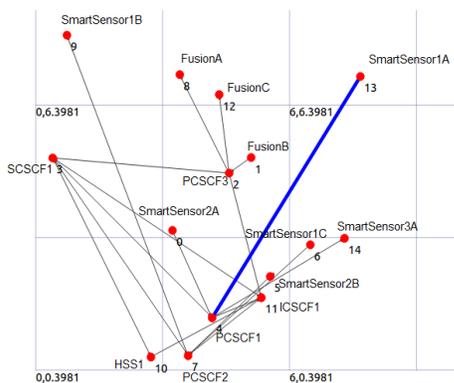


Fig. 8. Snapshot from the animation through NetAnim tool after running NS-3 simulation

B. Compilation and Simulation

The design tool ArchiMO generates an XMI file to represent the graphical design. This helps to conduct the design model to other tools. We use the XMI file as an input to our self-developed domain-specific model compiler to generate the simulation code (see Figure. 6). This hides complexity of constructing simulation programs from the designer and saves considerable time of the development process. The code generator needs both the meta-model including the abstract syntax of DSML for MO, and the input model that is generated from the design tool.

The XPAND template in (Figure. 6) contains the mapping rules between the model elements and their representations in NS-3 [13].

We have run the generated code in NS-3 (version 3.13), and the results of compilation and running shows no errors. Traces and logs (e.g., PCAP files) were generated to analyze the simulation outputs.

(Figure. 8) shows the architecture of the system design that is generated by NS-3 for the mentioned design model. NS-3 generated a hardware representations (Nodes, Interfaces, Wires) for the elements of the design model and the blue colored stream represents a message that is exchanged between two nodes in a fixed moment. This confirms that the behavioral elements were mapped as well.

We have used our approach in different application domains and network simulators (Video Conferencing System [12][13], and MO context). The common design concept between all cases is the underlying platform (IMS) that represents the Platform Specific Model (PSM) [25].

In other words, considering using one tool (e.g., NS-3), we could change the application domain relying on ArchiMate and our extensions (DSMLs) by fixing the underlying platform that is represented in the technology layer. This confirms that our proposed design tool (ArchiMO) creates models that follow the same meta-model and domain-specific concepts/constraints.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have presented a Domain Specific Modeling Language (DSML) for MO context. Our approach is based

on extending the ArchiMate meta-model relying on MDE fundamentals. We have proposed a new design tool (ArchiMO) that is generated from the extended MO meta-model and respects the domain-specific concepts and constraints.

ArchiMO protects the designer from making design errors earlier than the other design activities and the code generation step. We are relying on a standard and open tool (Archi) and developing it by extending the modeling language and Java implementations. Another advantage is the extensibility of our proposed meta-model/tool. The developers may extend it and add new concepts and standards according to the progress in MO domain. ArchiMO provides the reusability of the added MO and Data Fusion concepts (e.g., Smart Sensor, Data Fusion, etc.) in different applications, activities, models or instances. ArchiMO reduces the time of the design activity as well, by having the specific elements and constraints in the palette of this tool. Additionally, we conserve the standard constraints in the abstract syntax (Meta-Model) of ArchiMate since the new added elements inherits concepts from standard ArchiMate elements.

On the other side, representing and meta-modeling the domain knowledge is itself a hard job that needs experience and high level of accuracy, especially when setting the grammar of the DSML according to the meta-model constraints.

As perspectives, we will extend our meta-model in order to satisfy and cover the most possible required operations, concepts and activities in the context of MO.

REFERENCES

- [1] O. Zein, J. Champeau, D. Kerjean, and Y. Auffret, "Smart sensor metamodel for deep sea observatory," in *OCEANS 2009 - EUROPE*, May 2009, pp. 1–6.
- [2] *MeDON - Acoustic Data*. URL: <http://www.medon.info/>, Last visited 03-November-2014.
- [3] J.-P. Schneider, J. Champeau, and D. Kerjean, "Domain-specific modelling applied to integration of smart sensors into an information system," in *International Conference on Enterprise Information Systems (ICEIS 2011)*, Lille, France, Jun. 2011.
- [4] V. Chiprianov, I. Alloush, Y. Kermarrec, and S. Rouvrais, "Telecommunications service creation: Towards extensions for enterprise architecture modeling languages," in *6th Intl. Conf. on Software and Data Technologies (ICSOFT)*, vol. 1, Seville, Spain, 2011, pp. 23–29.
- [5] M. E. Liggins, D. L.Hall, and J. Llinas, *Multisensor Data Fusion, Theory and Practice*, S. edition, Ed. Taylor & Francis Group, LLC, 2009.
- [6] *The Open Group, ArchiMate 1.0 Specification*. <http://www.opengroup.org/subjectareas/enterprise/archimate>, Last visited 03-November-2014.
- [7] J.-L. Pérez-Medina, S. Dupuy-Chessa, and A. Front, "A survey of model driven engineering tools for user interface design," in *Proceedings of the 6th International Conference on Task Models and Diagrams for User Interface Design*, ser. TAMODIA'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 84–97.
- [8] O. Noran, "An analysis of the zachman framework for enterprise architecture from the {GERAM} perspective," *Annual Reviews in Control*, vol. 27, no. 2, pp. 163 – 183, 2003.
- [9] D. Quartel, W. Engelsmanb, H. Jonkersb, and M. van Sinderenc, "A goal-oriented requirements modelling language for enterprise architecture," in *Enterprise Distributed Object Computing Conference, 2009. EDOC '09. IEEE International*, University of Twente. IEEE, 2009, pp. 3 – 13.
- [10] V. Chiprianov, Y. Kermarrec, and S. Rouvrais, "Extending enterprise architecture modeling languages: Application to telecommunications service creation," in *The 27th Symposium On Applied Computing*. Trento: ACM, 2012, pp. 21–24.
- [11] *Eclipse Modeling Framework*. <http://www.eclipse.org/modeling/emf/>, Last visited 03-November-2014.

- [12] I. Alloush, V. Chiprianov, Y. Kermarrec, and S. Rouvrais, "Linking telecom service high-level abstract models to simulators based on model transformations: The IMS case study," in *Information and Communication Technologies (EUNICE 2012)*, ser. Lecture Notes in Computer Science, R. Szabó and A. Vidócs, Eds., vol. 7479. Springer Berlin Heidelberg, August 2012, pp. 100–111.
- [13] I. Alloush, Y. Kermarrec, and S. Rouvrais, "A generalized model transformation approach to link design models to network simulators: Ns-3 case study," in *International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2013)*. SciTePress Digital Library, July 2013, pp. 337–344.
- [14] N. Medvidovic and R. Taylor, "A classification and comparison framework for software architecture description languages," *J*, vol. 26, pp. 70–93, Jan 2000.
- [15] V. Chiprianov, "Collaborative construction of telecommunications services. an enterprise architecture and model driven engineering method," Ph.D. dissertation, Telecom Bretagne, France, 2012.
- [16] L. Touraille, M. K. Traoré, and D. R. C. Hill, "A model-driven software environment for modeling, simulation and analysis of complex systems," in *Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium*, ser. TMS-DEVS '11, San Diego, CA, USA, 2011, pp. 229–237.
- [17] K. Y. A. Achilleos and N. Georgalas, "Context modelling and a context-aware framework for pervasive service creation: A model-driven approach," *Pervasive and Mobile Computing*, vol. 6, no. 2, p. 281–296, 2010.
- [18] J.-L. Bakker and R. Jain, "Next generation service creation using xml scripting languages," vol. 4, pp. 2001–2007, 2002.
- [19] J. Sorribas, A. Barba, E. Trullols, J. Del Rio, A. Manuel, and M. de la Muela, "Marine sensor networks and ocean observatories. a policy based management approach," in *Computing in the Global Information Technology, 2008. ICCGI '08. The Third International Multi-Conference on*, July 2008, pp. 143–147.
- [20] J. Bezivin, "In search of a basic principle for model driven engineering.," *Novatica Journal*, vol. vol. 2, p. pp. 21–24, 2004.
- [21] *Atlas transformation language*. <http://www.eclipse.org/atll/>, Last visited 03-November-2014.
- [22] *Eclipse Modeling*. <http://www.eclipse.org/modeling/>, Last visited 03-November-2014.
- [23] M. M. T. Zekai Demirezen, Barrett R. Bryant, "Dsml design space analysis," in *UAB, Birmingham, AL 35294, USA*, 2011.
- [24] H. Cho, J. Gray, and E. Syriani, "Creating visual domain-specific modeling languages from end-user demonstration," in *Modeling in Software Engineering (MISE), 2012 ICSE Workshop on*, June 2012, pp. 22–28.
- [25] I. Kurtev, J. Bézin, F. Jouault, and P. Valduriez, "Model-based DSL frameworks," in *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*, ser. OOPSLA '06. New York, NY, USA: ACM, 2006, pp. 602–616.
- [26] I. Liggins, M.E., C.-Y. Chong, I. Kadar, M. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 95–107, Jan 1997.
- [27] I. Sommerville, *Software Engineering, Ninth Edition*, M. Horton, Ed. Pearson, 2011.
- [28] *Archi tool*. <http://archi.cetis.ac.uk/developer/model-new-element.html>, Last visited 03-November-2014.

A Formal Graph-Based Model Applied to Cluster Communication in Wireless Sensor Networks

Laura M. Rodríguez Peralta, Eduardo Ismael
Hernández
Engineering Department,
UPAEP University
Puebla, México
lauramargarita.rodriguez01@upaep.mx,
eduardo.ismael@upaep.mx

Lina M. P. L. Brito
M-ITI (Madeira Interactive Technologies Institute)
Exact Sciences and Engineering Competence Center,
University of Madeira (UMa),
Madeira, Portugal
lina@uma.pt

Abstract— In several application scenarios, like the case of structural monitoring, it is important to model and represent the Wireless Sensor Networks (WSN) to be deployed. To model the components and properties of an entire WSN types of collaboration in WSNs, we have created the Collaborative Wireless Sensor Networks (CWSN) model. This model also models the different types of collaboration that occur in a WSN. Our main goal is to provide a theoretical mathematical foundation that can model and analyze WSNs. Our approach is based on graph theory and propositional logic. The main contribution of this paper is applying the CWSN model to describe and represent a WSN. The use of the CWSN model brings several advantages, such as enabling the graphical representation of the state of the network and of several important properties: i) the network topology; ii) the transmission between nodes considering a multi-hop communication; iii) the transmission hierarchy; iv) the evolution of the networks through a succession of graphs; and so on.

Keywords: *wireless sensor networks; cluster; graphs; communication; classical proposition logic, network modeling.*

I. INTRODUCTION

WSNs are a special case of wireless ad hoc networks, but characterized by specific constraints. Besides energy restrictions, sensor nodes suffer from other resource limitations: they have reduced memory and processing capabilities; and, due to short transmission range, nodes can only communicate locally, with a certain number of local neighbors [1] [2] [3] [4]. In many cases, these networks are subject to highly dynamic conditions, caused by nodes' mobility, hardware failures, lack of battery, or other factors. To overcome these limitations, nodes have to collaborate in order to accomplish their tasks: sensing, signal processing, computing, routing, localization, security, etc. Therefore, WSNs are, by nature, collaborative networks [5].

There are quite a few works in the literature concerning collaboration in WSNs; however, they only focus a specific type of collaboration, which is associated with the accomplishment of a specific task. In [6], we proposed a formal and hierarchical model of cooperative work, the Collaborative Wireless Sensor Networks (CWSN) model, which is designed specifically for WSNs. It allows not only

the modeling of collaborative work (based in CSCW - Computer Supported Cooperative Work [7] - concepts), but also the modeling and visual representation of all the entities that can compose a WSN, as well as its properties. Moreover, CWSN is a generic model since it can be applied to heterogeneous networks. We have used first-order logic to formalize and describe the proposed model. We further employ graph-theory to describe how communication occurs within a cluster. This formal description of clusters' communication is based on directed graphs; nodes are labeled with the signal +/-, respectively indicating if a node is active or not [6].

In this paper, we focus on using the graphical representation of the CSWN model to represent the state of the network and its properties. This graphical representation allows the user to comprehend what is occurring in a certain moment of the WSN lifetime, by easily visualizing the communication interactions, the state of the nodes (active or not), the state of the links, etc. This will allow enriching the proposed model and giving the user a better understanding of the components and the state of the WSN as well, through a more complete visual representation of the WSN. This is the main contribution of this paper.

This paper is organized as follows. In Section 2, we briefly describe the related work. In Section 3, the CWSN model and its entities are briefly presented. Then, a comparison between CWSN and other models is presented. The advantages of the model are, also, outlined. Section 4 presents the SAVER project, an application example of our model. Section 5 provides some conclusions and future work perspectives.

II. RELATED WORKS

There are several works that try to model some aspects of WSNs. We have observed that the great majority of works focus on modeling of connectivity or mobility problems, or even on both problems. Moreover, we have identified other modeling concerns, such as: communication models, interference models, data aggregation models, coverage models, and signal processing models. On the contrary, the CWSN model intends to model a whole WSN, i.e., it tries to consider the most complete set possible of entities that can exist in a WSN, and their respective attributes.

Regarding the works focusing collaboration in WSNs, the great majority of them covers a specific type of collaboration, which is associated with the accomplishment of a certain task, such as: signal processing [8], sensing [9], computing [10], routing [11], localization [12], security [13], task scheduling [14], heuristics [15], calibration [16], resource allocation [17], time synchronization [18], transmission [19], etc., and also works concerning collaboration between wireless sensor nodes and other kind of devices (heterogeneous groupware collaboration) [20], [21] to support some specific applications (for example, collaboration between sensor nodes and PDAs, in a fire fighting scenario).

According to the literature, the only work that presents a model for collaborative work, in sensor networks, was proposed by Liu et al. [22]. It is the Sensor Networks Supported Cooperative Work (SNSCW) model, a hierarchical model that essentially divides cooperation in sensor networks in two layers; the first one relates to cooperation between humans and sensor nodes; the second one relates to cooperation between the sensor nodes. This model was designed for sensor networks.

However, the SNSCW model only allows the modeling of collaboration itself. On the contrary, the CWSN model, which has been presented in [6], is a formal model that was created specifically to describe WSNs. However, the CWSN model allows not only the modeling of collaborative work (based in CSCW concepts), but also the modeling, formalization and visual representation of the entities that can constitute a WSN (different types of nodes, clusters, relationships, sessions, obstacles, etc.), as well as its attributes. Moreover, it allows the representation of the WSN's hierarchy and of the network evolution.

The CWSN model formalizes the most significant properties of each entity through first-order logic. Even though the CWSN model is a graph-based model, it includes other objects [6] in order to make the modeling of the various entities of a WSN possible. This is of paramount importance to completely represent a WSN.

WSNs are extremely dynamic systems, both in the sense that their characteristics change over their lifetime and for the fact that sensor networks' technology (hardware and software) is subject to fast changes. To overcome this issue, the CWSN model can be updated or extended, through the introduction of new entities and/or new attributes. Therefore, another key point of this model is its scalability, since it can easily evolve.

III. THE CWSN MODEL

The CWSN model is a formal model of collaborative work that was specifically created to describe WSNs. This model allows the representation of the entities (different types of nodes, clusters, relationships, sessions, obstacles, etc.) and properties of a WSN, of its hierarchy, and of the network evolution; therefore, it allows not only the modeling of collaborative work (based in CSCW concepts), but also the modeling, formalization and visual representation of a whole WSN.

The CWSN model formalizes all the properties of each entity through first-order logic. Also, CWSN is a graph-based model; however, it includes other objects in order to make possible the modeling and visual representation of all the entities that can compose a WSN. This is of paramount importance to completely represent a WSN.

A. CWSN Model Definitions

We define entities as all the components that might exist in a WSN. The symbol, the concept and the description of all the entities included in the proposed model are illustrated in Table I.

A WSN can have different types of nodes: ordinary wireless *sensor nodes*, anchor nodes, one or more sink nodes (also known as base stations) and a gateway. The *sink node* and the *anchor node* are wireless sensor nodes with special functions.

A *cluster* is a group of nodes, created according to: geographical area, type of sensor nodes, type of phenomenon, task to be performed, etc., providing the WSN with a hierarchical structure. If nodes are grouped in clusters, one of the members of each cluster becomes the cluster head (there is only one cluster head per cluster). In this case, all nodes in the cluster have to send collected data to the cluster head (for instance, the more powerful node or the router, in case of a ZigBee-based WSN), which, in turn, is responsible for sending data to a sink node.

If two nodes collaborate, there is a *relationship* between them. Associated with a relationship there is always an exchange of data, which corresponds to the *data flow* entity. Collected data (temperature, humidity, light, etc.) can be sent to other nodes using one or more types of signals (radio, ultrasound, acoustical, etc.).

Obstacles are objects (for, e.g., building, tree, rock, etc.) that may obstruct the line-of-sight between two or more nodes, not allowing for direct communication between them. So, they can influence the relationships created.

The *user* is the entity who interacts with the WSN, defining the application, querying the network, visualizing data, customizing the work of the sensor nodes, etc.

Several collaborative *sessions* can be established when monitoring a WSN, and they can exist simultaneously or not. Basically, new sessions may be established based on new queries a user makes on the WSN.

And, last but not least, as the *battery* is the most critical resource of a sensor node, it is really important that the user knows the state of the battery of each sensor. That is why the battery is also an entity of our model.

3.1 CWSN Model Formalization

In this section, we formalize the model's entities and their main properties, using both first-order logic and graph theory.

1) Definitions: We can formulate the sensor network as a graph $G(V,E)$. V (vertices) represents the set of sensor nodes, and E (edges) describes the adjacency relation between nodes. That is, for two nodes $u, v \in V$:

$(u, v) \in E$, if, and only if, v is adjacent to u . If a node u is within a node v 's transmission range, we say that u is adjacent to v , or equivalently, that u is a neighbor of v . In the absence of interference, this relation is typically symmetric (or bidirectional), i.e., if a node u can hear a node v , also v can hear u .

An arrow between two nodes represents a relationship between them. A relationship can be established based on: localization, phenomenon, type of sensor node, etc. The arrow represents a producer-consumer relationship. Let us consider two nodes A and B ; the arrow $\overset{\textcircled{A}}{\longrightarrow} \overset{\textcircled{B}}{}$ means that node A transmits data to node B . So, node B consumes information from node A . The transmission of data between both nodes follows the format $\text{TypeOfSignal.Data (TypeOfSignal,Data)}$ \longrightarrow), verifying the consumer-producer property.

So, according to the specifications of the CWSN model, a WSN can be represented using labeled and directed graphs; the labels are associated to the edges and are designated by data flow. TypeOfSignal.Data identifies these labels. The labels inform the user about the type of signal that is being used by the sensor nodes for transmitting data (for, e.g., radio frequency, ultrasounds, acoustical, etc.), and about the type of data that is being collected and sent to the sink node (for, e.g., temperature, humidity, light, acceleration, etc.). These labels are important because they allow the user to become more aware of the state and the behavior of the WSN, since the labels add information that goes beyond the mere representation of the communication interactions between nodes.

Let's represent the total number of sensor nodes that constitute the WSN by N_r , with $N = \{1, 2, \dots, N_r\}$ and a wireless sensor node by N_i , with $i \in N$. The WSN has a limited lifetime, which can vary from some hours to several months or years. Let us denote the lifetime of the network (in seconds) by LT , with $T = \{1, 2, \dots, LT\}$, and the j^{th} second of life of the network by t_j , with $j \in T$.

2) Sensor Node (N_i): A sensor node (N_i) is defined by:

$$N_i = \{ID, TS, CM, CT, R, PS, L, TM, S, PD, CN\}$$

Table II defines and formalizes the properties that are important to identify a sensor node (N_i). This Table serves as an example for the type of formalization that has been proposed for the remaining entities, presenting a formal description of their most important properties [6].

3) Sink Node (S_K): The sink node is the node to which data collected by ordinary sensor nodes is sent. It is responsible for sending data to the gateway being the only node that can do it, what verifies the flow control property. Regarding mobility, two cases must be distinguished: the Stationary

Sink Node (StS_K), with the localization of the sink being well-known and independent of time; and the Mobile Sink Node (MbS_K), where the localization of the sink node varies as it moves along the WSN. .

4) Anchor Node (A): If the localization (L) of wireless sensor nodes is unknown (usually, due to an ad hoc deployment), it may be necessary to have some anchor nodes to help these sensor nodes to determine their own localization. So, an anchor node differs from a sensor node because its localization is always well known. This can be achieved either by equipping the anchor node with a GPS receiver or by manually configuring its position prior to deployment. Regarding mobility, an anchor node (A) can be:

- Stationary (StA): In this case: $TM(StA) = St$
- Mobile (MbA). In this case: $TM(MbA) = \{ContMb, Des\}$ or $TM(MbA) = \{OcMb, Des\}$

5) Network (WSN): So, a WSN can be defined by the following properties:

$$WSN = \{To, M, H, Nr, A, C, D, Hi, NS_K, NA, NC, NO, LT\}.$$

6) Session (Se_i): A session is the essential unit of a collaborative activity, which can be created based on different queries posed by the user. Depending on the WSN specific application, sessions can take place in parallel or in sequence; or they can be synchronous or asynchronous. Thus, in a certain moment, there may be several collaborative sessions in a WSN. A session (Se_i) can also be formulated as a subgraph, g , of the WSN, with $g(V,E) \subseteq G(V, E)$. Accordingly, some properties of the entities network and the sensor node are inherited.. Similarly to a sensor node (N_i), a session (Se_i) can be in one of two states: Active (Ac), or Inactive (In) when its objective is fulfilled. So:

$$S(Se_i) = Ac \text{ or } S(Se_i) = In.$$

Besides, similarly to the entity network, each session can have a group of active sensor nodes, a group of inactive sensor nodes and a group of relationships and data flows. So, a session (Se_i) is defined by the following properties:

$$Se_i = \{Se_{iID}, Se_{iObj}, S, DTx, Se_{iTLife}, To, M, H, A, C, D, Hi, Nr, NS_K, NA, NC, NO\}.$$

Note that the topology (To) has the same definition as in Table II; however, considering a specific instant of time, the topology of the session (Se_i) may be different from the topology of the WSN.

Also, considering the number of nodes, number of anchor nodes, number of sink nodes, number of clusters and number of obstacles, note that:

$$Nr(Se_i) \leq Nr(WSN), NA(Se_i) \leq NA(WSN), NS_K(Se_i) \leq NS_K(WSN), NC(Se_i) \leq NC(WSN), NO(Se_i) \leq NO(WSN), \text{ and } Se_{iTLife}(WSN) \leq Se_{iTLife}(Se_i)$$

TABLE I. DEFINITION OF THE ENTITIES CAN CONSTITUTE A WIRELESS SENSOR NETWORK.

Symbol	Concept	Description
	Sensor node	Nodes can be either stationary or mobile. Also, they can be in one of three possible states: active, sleep mode (in order to save energy), or inactive.
	Sink node/ Base Station	Node to which data collected by ordinary nodes is sent; being responsible to send data to the gateway. If there is only one sink node, all data collected by sensor nodes has to be sent to it. Otherwise, data may be sent to any sink node and, in this case, sink nodes must be able to communicate to each other
	Anchor node	Node with known localization, which support the other sensor nodes in the localization process
	Cluster	Group of nodes, created according to: geographical area, type of sensor, type of phenomenon, task, etc.
	Cluster Head	Sensor node to whom all sensor nodes in the cluster send the collected data; it is responsible for sending the received data to the Sink node.
	Relationship	The arrow represents a relationship between nodes A and B. It also represents an adjacency relation between nodes A and B; nodes A and B are neighbors. A relationship can be established based on: localization, phenomenon, type of sensor node, etc.
	Data flow	This label identifies both the type of signal being used (radio frequency, ultrasound, acoustical or light) and the type of data being transmitted between nodes (temperature, humidity, light, sound, video, internal voltage, etc.).
	Gateway	Device responsible to send the data to the user, through the Internet or satellite.
	Obstacle	An object (building, tree, rock, etc.) which may obstruct the line-of-sight between two or more nodes; depending on the type of signal that is being used by nodes (radio frequency, optical, acoustical, etc.), the obstacles may even not allow for communication between nodes.
	Session	In a certain moment, there may be several collaborative sessions in a WSN. A session can be established based on the objective (type of phenomenon to monitor, geographical area to monitor, etc.) of the WSN.
	Battery	It represents the percentage of the sensor node's remaining battery.
	User	Person that interacts with the WSN, querying the network, visualizing data, etc. The user customizes the work of the sensor nodes; the data collected by sensor nodes is used by the users' application.

7) Cluster (C): If a clustering algorithm is applied [23], clusters will be formed. Sensor nodes are grouped into clusters, mainly to support scalability (for managing a high number of nodes). But, besides supporting scalability, clustering can have several different objectives, such as: load balancing, fault tolerance, network connectivity, maximal network longevity, etc. Each cluster has a leader, the cluster head (CH). So, a cluster (C) is defined by:

$$C = \{C_{ID}, CH, Stb, NrC, IaC-To, IeCH-Con, CMet\}.$$

Even though clustering is influenced by the network and link layer protocols, some attributes can be identified.

8) Cluster Head (CH): The cluster head (CH) can be elected by the sensors in a cluster or pre-assigned by the network designer. Also, CHs may be the richest nodes in resources of the whole network. So, a cluster head (CH) is defined by [23]:

$$CH = \{CH_{ID}, TM, TN, Ro\}.$$

B. Main Properties Represented by the CWSN Model

The CWSN enables the graphical representation of several important properties, like nodes' mobility, connectivity and communication, network heterogeneity and

stability, network coverage, consumer-producer and flow control, as well as the graphical representation of other important aspects, like, the occurrence of failures, the network topology, the established routing paths, or the communication modality used by nodes and the type of sensed data. However, some aspects like signal interference was not considered in the CWSN model. The representation of some of these network properties, like mobility of nodes or topology changes, is possible through the representation of the network evolution. In other words, the model represents a screenshot of a WSN in a specific moment of time. As time goes by, several aspects of a particular WSN can change: (i) the state of the nodes can change; (ii) new nodes can be deployed; (iii) the topology can suffer modifications; (iv) new clusters can be created; (v) new obstacles can appear; etc. In the CWSN model, these network changes are naturally represented through a succession of figures as exemplified in Figure 1,

Figure 1 represents a modification in the topology of the network caused by the elimination of an obstacle. The obstacle was located between nodes N_4 and N_1 , impeding direct communication between these two nodes. For some reason (the obstacle may be static, like a tree or building, or it may move, like an animal or a car), the obstacle

disappeared. Consequently, connectivity and, therefore, communication was established between the referred nodes.

Thus, the CWSN model also addresses the analysis of temporal information. This can be used as an input for

creating a real-time tool that allows visualization and representation of a WSN, as we have demonstrated in [25], where the CWSN model was translated into XML language, which was used as an input for the tool.

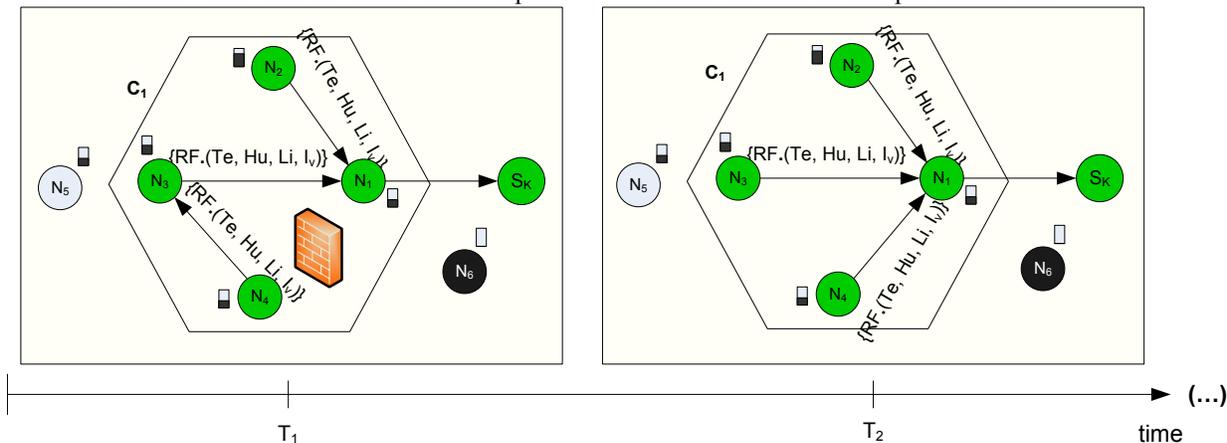


Figure 1. Modelling a change in the network topology, applying the CWSN Model.

C. Comparing the CWSN Model With Other Models For WSNs

Table III allows a quick overview of the main differences between the CWSN model and other models that have been applied to model these networks.

The remaining state-of-the-art models mostly address the modelling of a single issue of WSNs, addressing up to two issues (like mobility and connectivity) in the most complete modelling cases. Contrarily to these models, instead of focusing on modelling some specific problem of WSNs, the CWSN model is used to systematically describe and represent the features and properties of all the components that might constitute a WSN. So far, no other model has covered so many aspects of a WSN as the CWSN model does.

For example, unlike the SNSCW model [22], our model does not only model the cooperation within the network, but also the entire WSN. The CWSN model also allows the representation of the network hierarchy, from the collected data to the user (passing through the clusters, the session and the WSN). Moreover, the CWSN model is generic, in the sense that it can model heterogeneous networks and that it can be applied to any type of wireless sensors, (regardless its size, its hardware characteristics, the types of signals it can measure, etc.). It can also be applied to any WSN despite its specific application. So, it is possible to use all the entities defined in the model to represent a specific scenario of any application (monitoring a forest, a vineyard, a volcano, a museum, a natural catastrophe, etc.).

Besides, any changes that might occur on a certain application scenario (new collaborative sessions, new clusters, nodes moving, etc.) can be represented by a

sequence of figures; hence, the CWSN model allows the representation of each state of the network and its evolution.

Regarding collaboration, the model includes some fundamental CSCW [7] concepts (for, e.g., session, relationship, data flow and groups) and properties (for, e.g., consumer-producer and flow control). Thus, analyzing Table III, it is possible to conclude that the CWSN model presents important advantages over the state-of-the-art models presented in section II.

D. Contributions of the CWSN model

WSNs are extremely dynamic systems, both in the sense that their characteristics change over their lifetime and for the fact that sensor networks' technology (hardware and software) is subject to fast changes. To overcome this issue, the CWSN model can be updated or extended, through the introduction of new entities and/or new properties. Therefore, another key point of this model is scalability, since it can easily evolve.

The CWSN model itself models the most important components of the WSN. As such, their advantages are:

- The CWSN model provides a grammar for formally modeling an entire WSN, i.e., all the entities that can exist in a WSN, and their respective attributes.
- It also allows to visually and graphically representing a whole WSN, including not only its entities, but a very important aspect to users and network managers, the network topology. Basically, the CWSN model provides a common framework for describing and representing any WSN.
- Moreover, the CWSN model allows representing the network hierarchy, from the collected data to the user,

passing through the sensor nodes, the clusters, the sessions and, finally, the whole WSN.

- Besides, any changes that might occur on a certain application scenario (mobility of nodes, failure of nodes, topology changes, new collaborative sessions, new clusters, etc.) can be represented by a sequence of sub-graphs; hence, the CWSN model allows the representation of each state of the network and its evolution.
- Furthermore, the CWSN model is generic, since it can model heterogeneous networks; it can be applied to any type of wireless sensor nodes (regardless their size, their hardware characteristics, the types of signals they can measure, etc.), and to describe and represent any WSN, despite of its specific application. So, it is possible to use all the entities defined in the model to represent a specific network scenario of any application (monitoring a forest, a vineyard, a volcano, a museum, a natural catastrophe, etc.).

IV. THE CWSN MODEL APPLIED TO STRUCTURAL HEALTH MONITORING

The use of WSNs have brought several advantages in structural monitoring and the establishment of structural health compared to conventional methods where computers connected to accelerometers are used. In conventional methods, it is necessary to install cables through the structure, disturbing its normal operation and generating maintenance cost. Compared with conventional methods, WSNs provide the same functionality at a much lower price and a more flexible monitoring. Another problem is the high equipment and wiring installation and maintenance cost.

The cost of a conventional system with a computer and a piezoelectric accelerometer is about USD 40000 per sampling point. The estimated cost of the proposed system, in this work is less than USD 200 per point. In WSNs no wiring is required, making installation and maintenance much easier and inexpensive. Moreover, the use of WSNs allows Structural Analysis of Vulnerability of buildings through wireless sensor networks (SAVER) platform to be deployed and operate even if the building is in operation. It does not cause further visual impact due to its small size, low power consumption and installation flexibility. The advantage of structural health monitoring based on WSNs can be extended if the MEMS acceleration sensor type is used. The MEMS accelerometer is a silicon chip, which is very compact in size, low power consumption and cheap. Without MEMS, a small WSN, even low-power and low-cost accelerometer, would be degraded.

Thus, the SAVER platform will aim at gathering information to establish the vulnerability level of structural health of buildings. Such information will be used in

decision making for both schemes and prevention programs, and for post-seismic evaluation.

The SAVER platform will be able to monitor and display information in real-time. It will determine from the implementation of several methods for estimating seismic response and damage detection, the level of structural vulnerability of buildings. In addition, our platform will offer several services that will notify users about potential risks of the structure through alarms, email and SMS. Besides, it will have a Web based monitoring platform and a mobile app for Android and I-Phone. Also, this platform will generate graphs, reports and statistics. Some preliminary results of the SAVER project was published in [26].

The expected results, in SAVER project, intend to give the basis for the analysis of buildings and gather instrumental data that can be useful for decision-making of institutions and users that are responsible for infrastructure and buildings. Furthermore, in this project, we pretend to provide the necessary information to implement methods of vulnerability analysis and therefore, to estimate the seismic risk of buildings, such as hospitals or schools.

SAVER project will be validated in the building B (Figure 2) of UPAEP University, located in Puebla city in México. This building is structured based on reinforced concrete rigid frames. Furthermore, it is regular in plan and consists of four levels with a height of 3.15 m each one, so it has a total height of 12.60 m. In the transversal direction, it has a bay of 10.50 m. In the longitudinal direction, the building has eight bays of 6 m each one, so that it has a total length of 48 m. The structural elements are composed of beams and columns. The beams, in transversal direction, have square cross section at all levels.

In the longitudinal direction at level 1 and 4, they have variable prismatic section, while in level 2 and 3 are rectangular. All columns in each level have variable hexagonal section. Floor system has 0.25 m thickness and is prefabricated. The building has masonry walls with 0.15 m thickness. This building was built in 1984. In recent studies, we have determined a high level of structural vulnerability. But, these studies were made using only three wired sensors. In order to obtain a better vulnerability estimation, we intend to instrument this building using the SAVER platform. The proposed topology for this building is shown in Figure 3.

We are planning to install 12 sensor nodes. Each node has two sensors, a temperature sensor and an acceleration sensor. Using CWSN model, we can visualize the interaction among the sensors and their relationship. Figure 4 shows the WSN of the Building B.

TABLE II. DEFINITION OF THE PROPERTIES OF THE ENTITY A SENSOR NODE (N_i).

○	Properties	Description / Formalization
Sensor Node (N_i)	Identifier (I_D)	Each sensor node has a unique identifier (I_D) $I_D(N_i) = i, i \in \mathcal{N}$
	Types of sensors (TS)	A sensor node (N_i) can have several types of sensors, each one measuring a different phenomenon: light (Li), temperature (Te), humidity (Hu), sound (Sd), internal voltage (Iv), acceleration (Ac), pressure (Pr), vibration (Vb), received signal strength indicator (RSSI), etc. $So, TS(N_i) \subseteq \{Li, Te, Hu, Sd, Iv, Ac, Pr, Vb, RSSI, \dots\}$
	Communication modality (CM)	A number of communication modalities can be used, such as: radio (RF), light (Li), ultrasound (US), acoustical (Ac), optical (Opt), hybrid (Hy). $So, CM(N_i) \subseteq \{RF, Li, US, Ac, Opt, Hy\}$
	Communication Technology (CT)	A sensor node (N_i) can use different communication technologies. So far, three technologies have been proposed: ZigBee (ZB), which corresponds to IEEE 802.15.4; Bluetooth (BT); IEEE 802.11 (Wi-Fi). $So, CT(N_i) \subseteq \{ZB, BT, Wi-Fi\}$
	Transmission Range (R)	Let P_i be the nominal transmission power of a node. $P_{R,j} - i$ is the received power of a signal propagated from node i to node j . A received power $P_{R,j} - i$ above a given threshold P_{th} will provide sufficient SNR (<i>Signal to Noise Ratio</i>) in the receiver to decode the transmission. The nominal transmission range for successful communication can be defined as [19]: $R = P_i/P_{th}$ Note that due to the instability in the transmission range, the area a wireless sensor node can reach is not necessarily a circle and the range can vary between $r = (1-\epsilon).R$ and $R, \epsilon > 0$ [19].
	Power Supply (PS)	Energy can be supplied by batteries (that are, usually, of very limited capacity), solar cells or an external and unlimited power supply (only possible if nodes are stationary and in indoor applications). <ul style="list-style-type: none"> ▪ Battery (B); ▪ Solar cells (SC); ▪ External and unlimited power supply (VDC); ▪ Hybrid (Hy) – for, e.g., battery and solar cells; ▪ Etc. $PS(StS_k) \subseteq \{B, SC, VDC, Hy, etc.\}$ However, the great majority of sensor nodes are equipped with batteries. The lifetime of a sensor node (N_i) is limited by its battery, depending on its capacity and type. The battery can be defined by: <ul style="list-style-type: none"> ▪ Type of battery: T_B, with $T_B(N_i) \in \{\text{lithium, alkaline, li-ion, AA, external power supply, solar cells, electromagnetic and piezoelectric transducers, etc.}\}$ ▪ Capacity (voltage): $C_B(N_i)$ [V] ▪ Remaining capacity at time t_j: $P_{BN_i}(t_j)$ [%] $BN_i(t_j) = \{TB(N_i), CB(N_i), PBN_i(t_j)\}$
	Localization (L)	Let $L_{N_i}(t_j)$, with $i \in \mathcal{N}$ and $j \in \mathcal{T}$, denote the location of node N_i at time t_j . The type of deployment affects important properties of the network (node density, node locations, etc.). The deployment of sensor nodes may be: <ul style="list-style-type: none"> ▪ Random (ad hoc deployment, for, e.g. dropped by an aircraft). In this case, the localization of a node is unknown: $L_{N_i}(t_j) = (x, y, z)$, where $x, y, z \in \mathbb{R}$ are unknown. ▪ Manual: sensor nodes are deployed in pre-determined positions. In this case, the localization of a node is well-known: $L_{N_i}(t_j) = (a, b, c)$, where $a, b, c \in \mathbb{R}$ are known.
	Product Description (PD)	Alphanumeric that identifies the brand and the model of the sensor node. $PD(N_i) = \{\text{Brand}(N_i), \text{Model}(N_i)\}$
	Consumer Node (CN)	Often referred to as parent node, is the node to which the sensor node N_i is sending data at time t_j . Considering for example a relationship between nodes A and B; if node A is transmitting to node B, then B is the consumer node since it is receiving the data.
	Type of Mobility (TM)	A sensor node (N_i) can be: <ul style="list-style-type: none"> ▪ Stationary (St): $L_{N_i}(t_1) = L_{N_i}(t_2) = \dots = L_{N_i}(t_{L,T})$ ▪ Mobile (Mb): The period of mobility can be occasional or continuous: <ul style="list-style-type: none"> Occasional (Oc), when long periods of immobility occur: $\exists j, l \in \mathcal{T}: L_{N_i}(t_j) \neq L_{N_i}(t_l), \text{ and } j \neq l \wedge \exists r, s \in \mathcal{T}: L_{N_i}(r) = L_{N_i}(r+1) = \dots = L_{N_i}(s), \text{ and } s \gg r$ Continuous (Cont): $\forall j \in \mathcal{T} \setminus \{LT\} L_{N_i}(t_{j+1}) \neq L_{N_i}(t_j)$ Mobility can still be classified in: <ul style="list-style-type: none"> ▪ Incidental (Inc), for, e.g., due to environmental influences \approx Occasional ▪ Desired (Des), whether active or passive, which can be applied to any period of mobility (occasional or continuous). $So, TM(N_i) \in \{St, \{OeMb, Inc\}, \{OeMb, Des\}, \{ContMb, Inc\}, \{ContMb, Des\}\}$
State (S)	Depending on its power mode, the node N_i can be in one of two states (S): <ul style="list-style-type: none">  Active (Ac): Node that is in the active state. Its color will depend on the cluster it belongs to, since each cluster will be represented by a different color. By default, the green color will be used.  Sleep Mode (Sm): Node that is in the sleep mode, in order to save energy. Colored in grey color.  Inactive (In): Node that is damaged, or has some failure or has run out of energy. Colored in black color. $So, S(N_i) \in \{Ac, Sm, In\}$	

TABLE III. COMPARATIVE ANALYSIS OF THE MODELS CREATED FOR WSNs.

Model	Modelling issues	Graph-based	CSCW concepts	Collaboration levels	Graphical representation of the WSN	Formal model
Kenniche and Ravelomananana (2010)	Topology	YES	NO	NO	YES (topology)	NO
SNSCW (Liu et al., 2006b)	Cooperation	NO	YES	YES (two)	NO	YES
Wang (2011)	Coverage	no	no	no	no	Deployment and topology control
Bonaci et al. (2010)	Behaviour of the WSN under attack	yes	no	no	no	Network security
Wu and Chung (2009)	Sensing and Coverage	no	no	no	no	Deployment and topology control
Ni et al. (2009)	Sensor faults	no	no	no	no	Fault detection
Wüchner et al. (2010)	Performance	no	no	no	no	Evaluation of performance and energy efficiency
CWSN	The whole WSN and collaboration hierarchy	YES	YES	YES (five)	YES (whole WSN)	YES

The spatial distribution of the sensors is established from the geometry of each building. But it is necessary to deploy each sensor at least one in the geometric center of each level, and one sensor on the corner of the roof. If the longitudinal dimension of the building is large, it is suggested to deploy some sensors in one border of the building. It is important to monitor also the ground response using a free field sensor.



Figure 2. Building B of the UPAEP University

Figure 5 shows the WSN deployed in Building B, but represented using the entities proposed by the CWSN model. Basically, this representation clearly shows that only one session is established, but more importantly, it shows the structure of the WSN and the interactions that will take place between the different types of nodes that compose it. For instance, it shows that nodes are grouped in two clusters.

Using the CWSN model, it becomes evident the type of data collected by each node.

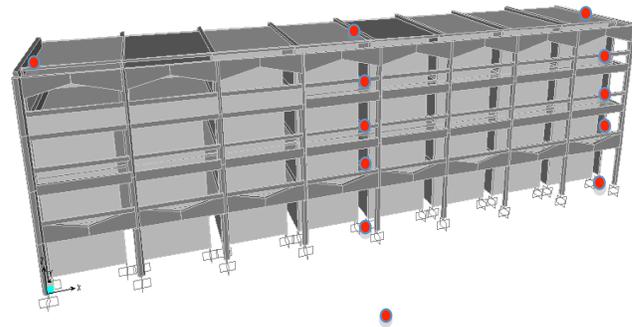


Figure 3. Topology of the WSN installed in the Building B

The problem to establish the structural health on buildings will be reduced if we use the CWSN model. The model can provide a tool for getting instrumental information about the structural properties, like acceleration on each storey as well as some structural dynamic parameters. If we use a sensor that can measure the acceleration (A_c) at each storey and considering that each sensor can provide the record of (A_c), in three orthogonal directions, respect to the building (longitudinal, transversal and vertical), the CWSN model can be implemented in order to estimate the structural vulnerability level. For this reason, we need to compute first the lateral displacement and then the structural damage parameter ψ , which is defined as the inter-storey drift. Lateral displacement can be estimated if we use a procedure that involves two integration steps of each acceleration record for longitudinal and transversal directions. In this implementation we can define the specific

properties for each Sensor Node (N_i), for example, we can use three Types of sensors (TS): Ac, Te and Iv; Communication modality (CM) can be defined as radio; for the Communication Technology (CT) we can be use ZigBee (ZB); the Power Supply (PS) can be a Hybrid type (Hy); the Localization (L) can be defined as Manual.

In order to show the advantages of the SAVER platform, we present a structural vulnerability function for building B (Figure 4). The vulnerability function describes the damage level $d(\psi)$ in terms of inter-storey drift (ψ). The values of damage are from 0 to 1. A damage equals to 0 indicates fully health condition and damage equals to 1.0 indicates collapse of the building. The vulnerability function was estimated using only one accelerometer located in three different points on the building. We recorded three acceleration records (ambient vibration), but these records were in three different intervals of time. This is a big limitation that can be covered by the SAVER platform, because the acceleration records must be at the same interval of time. The CWSN model can represent this easily, since it defines the evolution of the network, which can be represented through a succession of graphs.

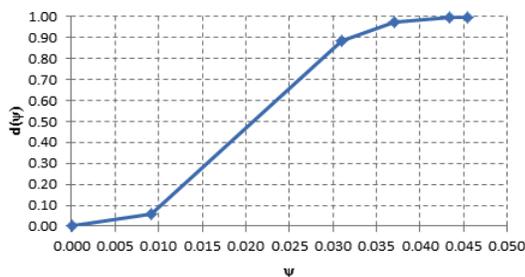


Figure 4. Typical vulnerability function to establish the structural health condition on Building B

The modelling and representation of the deployed WSN using the CWSN model can bring several advantages from the user and network manager’s points of view. The main contribution of the CWSN model is to standardize ways to model a WSN and provide a unified view of such a network regardless of what aspects are considered. Moreover, it allows the user and the network manager to become more aware of the composition and state of the whole network. That is, the CWSN model allows for visually representing several details about the WSN that has been deployed, what provides them with a more intuitive and prompt understanding of the WSN.

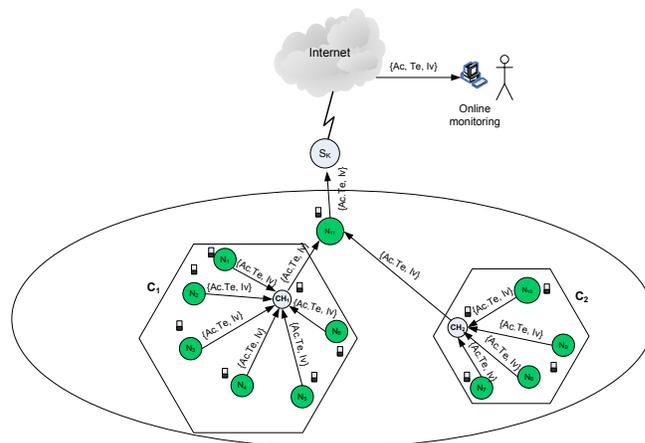


Figure 5. Modeling Building B’s WSN

V. CONCLUSIONS

The CWSN model also defines the evolution of the network, which can be represented through a succession of graphs.

In this paper, we presented the CWSN model, which is based on the CSCW methodology and specifically designed for WSNs. The great advantage of using this model to represent WSNs is that, besides modeling collaboration, it can also model the entire WSN. Also, this model allows for the representation of each state of the network and its evolution. Moreover, the CWSN model is generic, in the sense that it can be applied to heterogeneous WSNs, and scalable, as it can be updated if any modifications need to be introduced. The CWSN model was formalized in first-order logic. This attribution allows knowing which edges are active and which are not. In other words, we are able to identify which nodes are transmitting information. If some failure occurs on the process of transmission, our approach allows identifying this situation.

We consider that some advantages of the CWSN model arise from the fact of being formal and based on graphs.

The attributes defined for each entity of the CWSN model cover several dimensions. Thus, this model can be used as a framework for developing more generic software solutions for WSNs. Given the fact of being a broad and generic model, also confers the CWSN model with the ability of being applied to describe any WSN regardless of its application scenario.

We also believe that this model can assist network designers in making better decisions regarding the organization and management of the network. This contribution becomes more significant given that the CWSN model was used as a basis to implement an awareness tool and a sessions’ managing tool for WSNs, which will be described in the next chapter.

Moreover, the CWSN model can be used to automatically generate some graphs of the WSN that will allow for identifying routing paths, detecting damaged/failed nodes or links, etc.

REFERENCES

[1] I., Su, W. Akyildiz, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensors networks: a survey", *Computer Networks*, 38, 2002, pp. 393-422.,

[2] L. Brito and L. M. Rodríguez Peralta, "Wireless Sensor Networks: Basic Concepts", in *Encyclopedia of Networked and Virtual Organizations*, Idea Groups, vol. I, 2008, pp. 57-64, ISBN: 978-1-59904-885-7.,

[3] S. Tilak, N. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensors network models", *Mobile Computing and Communication Review*, vol. 6, no. 2, April 2002, pp. 28-36.

[4] M. Tubaishat and S. Madria, "Sensor Networks: an Overview", *IEEE Potentials*, May 2003, vol. 22, n. 2, pp. 20-23.

[5] K. Römer and F. Mattern, "The Design Space of Wireless Sensor Networks", *IEEE Wireless Communications*. 11, 6, December 2004 , pp. 54-61.

[6] L. Brito and L. M. Rodríguez Peralta, "A Collaborative Model for Wireless Sensor Networks Applied to Museums' Environmental Monitoring", *Proc. of the 5th International Conference on Cooperative Design, Visualization and Engineering (CDVE'2008)*, 5220/2008, Mallorca, Spain., *Lecture Notes in Computer Science*, Springer Verlag, September 2008, pp. 107-116.

[7] K. Mills, "Computer-Supported Cooperative Work Challenges", *Encyclopedia of Library and Information Science*, 2003.

[8] P. Ramanathan, K. Saluja and Y. Hu, "Collaborative Sensor Signal Processing for Target Detection, Localization and Tracking", *23rd Army Science Conference*. Orlando, USA 2002.

[9] K. C. Wang and P. Ramanathan, "Collaborative Sensing Using Sensors of Uncoordinated Mobility", *Int. Conference on Distributed Computing in Sensor Systems*, LNCS, vol. 3560, Springer, Marina del Rey, USA (2005), pp. 293—306.

[10] L. Iftode, C. Borcea and P. Kang, "Cooperative Computing in Sensor Networks", *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. Mohammad Ilyas (ed.), CRC Press, 2004.

[11] G. Chen, T. D. Guo, W. G. Yang and T. Zhao, "An improved ant-based routing protocol in Wireless Sensor Networks", *Int. Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2006)*, Atlanta, USA, 2006, pp. 1—7.

[12] D. Dardari, and A. Conti, "A Sub-Optimal Hierarchical Maximum Likelihood Algorithm for Collaborative Localization in Ad-Hoc Networks", *IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2004)*, Santa Clara, USA 2004, pp. 425—429..

[13] A. Chadha, Y. Liu, and S. Das, "Group Key Distribution via Local Collaboration in Wireless Sensor Networks", *IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005)*, Santa Clara, USA, 2005, pp. 46—54.

[14] H. Sanli, R. Poornachandran and H. Cam, "Collaborative Two-Level Task Scheduling for Wireless Sensor Nodes with Multiple Sensing Units", *IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005)*, Santa Clara, USA, 2005, pp. 350-361.

[15] R. Reghelin, and A. Fröhlich, "A Decentralized Location System for Sensor Networks Using Cooperative Calibration and Heuristics", *9th ACM Int. Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM'06)*, Torremolinos, Spain, 2006, pp. 139-146.

[16] V. Bychkovskiy, S. Megerian, D. Estrin and M. Potkonjak, "A Collaborative Approach to In-Place Sensor Calibration", *2nd Int. Workshop on Information Processing in Sensor Networks (IPSN'03)*, Palo Alto, USA, 2003, pp. 301—316.

[17] S. Giannecchini, M. Caccamo, and C. S. Shih, "Collaborative resource allocation in wireless sensor networks", *16th Euromicro Conference on Real-Time Systems (ECRTS'04)*, Catania, Italy, 2004, pp. 35—44.

[18] A. Hu and S. Servetto, "Algorithmic Aspects of the Time Synchronization Problem in Large-Scale Sensor Networks", *Mobile Networks and Applications*, vol. 10, Springer Science + Business Media Inc., 2005, pp. 491--503.

[19] A. Krohn, M. Beigl, C. Decker, T. Riedel, T. Zimmer and D. Varona, "Increasing Connectivity in Wireless Sensor Network using Cooperative Transmission", *3rd Int. Conference on Networked Sensing Systems (INSS)*. Chicago, USA, 2006.

[20] L. Cheng, T. Lian, Y. Zhang and Q. Ye, "Monitoring Wireless Sensor Networks by Heterogeneous Collaborative Groupware". *Sensors for Industry Conference (Sicon/04)*. New Orleans, USA, 2004.

[21] Z. Chaczko, F. Ahmad, and V. Mahadevan, "Wireless Sensors in Network Based Collaborative Environments", *6th Int. Conference on Information Technology Based Higher Education and Training (ITHET 2005)*, 2005, pp. F3A-7-F3A-13.

[22] L. Liu, H. Ma, D. Tao and D. Zhang, "A Hierarchical Cooperation Model for Sensor Networks Supported Cooperative Work", *10th Int. Conference on Computer Supported Cooperative Work in Design (CSCWD'06)*, 2006, pp. 1-6.

[23] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of Data Aggregation in Wireless Sensor Networks", *Proceedings of IEEE 22nd Int. Conf. on Distributed Computing Systems Workshops*, 2002, pp. 575-578.

[24] S. Hussain, U. Farooq, K. Zia, and M. Akhlaq, "An Extended Topology for Zone-Based Location Aware Dynamic Sensor Networks", *Proc. National Conference on Emerging Technologies (NCET)*, Szabist Karachi, Pakistan, Dec. 2004.

[25] L. M. Rodríguez Peralta, L. M. P. Leão Brito, Teixeira Gouveia B.A., Sousa D.J.G., and Alves C.S., "Automatic monitoring and control of museums' environment based on Wireless Sensor Networks", *Electronic Journal of Structural Engineering (EJSE)*, Special Issue: Wireless Sensor Networks and Practical Applications, 2010, pp. 12 – 34, ISSN 1443-9255.

[26] L. M. Rodríguez Peralta E. Ismael Hernández, S. A. Cardeña Moreno, D. Martínez Jiménez, and A. E. Muñoz Guarneros, "Towards to a platform of monitoring based in WSN to estimate the structural health of buildings", *Second European Conference on Earthquake Engineering and Seismology (2ECEES)*, Istanbul, Turkey, August, 2014, pp. 24-29.

High Deployability of IEEE 802.15.4k DSSS Systems in Interference Dominated Bands

Yasutaka Tada Yosuke Sato Shuzo Kato
 Research Institute of Electrical Communications, Tohoku University
 Sendai, Japan
 yasutaka.tada@riec.tohoku.ac.jp

Abstract— This paper clarifies that 802.15.4k direct sequence spread spectrum (DSSS) systems are suitable for wide sensor networks with star topology. According to our evaluation, the robustness of 802.15.4k DSSS systems under the co-channel interference (CCI) is 60 dB stronger than that of 802.15.4g due to its DSSS (1023 spreading factor) and FEC. In addition, this paper clarifies the area where 802.15.4g cannot survive under 802.15.4k CCI and vice versa. The latter area is 50 times larger the former. Furthermore, we show that directive base station antennas can reduce the interference impacted area by 80% compared to omnidirectional base station antennas.

Keywords—LECIM, IEEE802.15.4k, IEEE802.15.4g

I. INTRODUCTION

One of the latest standardization projects of 802.15 Working Group is P802.15.4k Low Energy Critical Infrastructure Monitoring (LECIM) Networks [1]. The standard is an amendment to the 802.15.4 standard to enable wide area sensor networks. Communication systems based the standard combine very low power consumption with long communication ranges with order of kilometers. There are two PHY layer options in the standard, the first one is based on direct sequence spread spectrum (DSSS PHY), the second one is based on frequency shift keying (FSK PHY). The focus of this paper is the DSSS PHY of 802.15.4k with following three reasons: (i) DSSS PHY deploys direct sequence spreading and this process gain is mandatory for longer transmission range – if 1023 spreading code length is used, the system can enjoy 30 dB processing gain. (ii) DSSS PHY standard mandates the use of strong FEC (rate 1/2 convolutional encoding with a constraint length of 7) while FSK PHY regulates it as an option and there has been none of FSK PHYs deploying FEC in market as of today. This may generate about 5 dB advantage of DSSS PHY against FSK PHY in AWGN channels and more than 8 dB difference in interference dominated environments, (iii) DSSS PHY regulate the use of PSK modulation which can be demodulated coherently while FSK PHY can't be demodulated coherently. This creates another 3 dB advantage to DSSS PHY. Thus, the total advantage of DSSS PHY against FSK PHY is more than 40 dB practically (with a spreading code length of 1023) and this is why this paper selected DSSS PHY as a targeted system

Although there will be a lot of different sources of interferences in 900 MHz bands, this paper focuses on how 802.15.4k DSSS systems can survive against interference from another standard of the 802.15 family, namely 802.15.4g Smart Utility Networks (SUN). This is important assumption since we need to identify the interferer to analyze and in Japan, 802.15.4g is supposed to be a major interferer to IEEE802.15.4k systems. Both systems operate in unlicensed bands, between 920 and 928

MHz in Japan [2]. Our analyses assume 802.15.4k systems operating in the presence of IEEE 802.15.4g co-channel interference (CCI) as well as 802.15.4g systems experiencing 802.15.4k CCI.

For the interference analysis, we consider two parameters that affect the performance of both systems under CCI. The first is a relative location of an 802.15.4g sub-channel with respect to an 802.15.4k channels. The second one is minimum desired signal to undesired signal ratios (D/U) of each system in AWGN environments. By using these two parameters, we will deduct “Installation inhibited area” in which the interferer destroys communications channels in each system. Further, the “installation inhibited area” is re-evaluated with a directive base station antenna for 802.15.4k DSSS systems showing how the “inhibited area” will be reduced by adopting high gain base station antennas instead of Omni antennas.

The outline of the paper is as follows. In the next section, we introduce 802.15.4k and 802.15.4g systems and illustrate their mutual interference. Section III provides the permissible D/U ratios of both systems under CCI. In Section IV, we show the Installation inhibited areas of 802.15.4k and 802.15.4g systems. Section V elucidates “installation inhibited areas” reduction by adoption of directional BS antennas. Conclusions are drawn in Section VI.

II. SYSTEM MODELS AND CO-CHANNEL INTERFERENCE

As mentioned in the introduction, we focus on the DSSS PHY of the 802.15.4k standard. It can operate in many unlicensed bands throughout the world including 920-928 MHz band in Japan with up to 1 MHz bandwidth as shown in Table I. Here, one sub-channel is defined as a channel with a bandwidth of 200 kHz and the maximum bandwidth is 1 MHz by bonding 5 sub-channels. The other system examined here, the 802.15.4g system, has a smaller bandwidth of 200 KHz or 400 KHz. Its details are shown in Table I as well.

The channelization of these standards enables 802.15.4g with a bandwidth of 200 KHz to interfere with 802.15.4k systems in one of its 5 sub-channels. In Figure 1, we illustrate the spectra of 802.15.4g operating at 50 kbps (in 200 kHz) and 802.15.4k systems operating in 1 MHz bandwidth, where a D/U ratio is 0 dB. Here, D/U ratio is defined by (1), as a ratio of transmission powers of desired and undesired signals.

$$DUR = \frac{\text{Transmit Power}_{\text{desired signal}}}{\text{Transmit Power}_{\text{undesired signal}}} \quad (1)$$

Sub-channel 3 of 802.15.4g is located at the exact center of the 802.15.4k signals with a bandwidth of 1 MHz.

In Figure 2, there are three channels of 802.15.4g systems with a bandwidth of 400 kHz, each interfering with 802.15.4k

systems operating with 1MHz bandwidth. In this case, sub-channel 2 of 802.15.4g systems is located at the exact center of the 802.15.4k signals.

Fig. 3 shows a block diagram of the interference simulation. “Gain” enables us to set the transmit power of 802.15.4k and set expected D/U ratios. “Freq. offset” is used to set the location of 802.15.4g channels relative to the 802.15.4k bandwidth. Using this system, we evaluate BER performance of each system, and derive the minimum D/U ratios to meet the target PER (Packet error rate).

Fig. 4 compares simulated and theoretical BER performance of 802.15.4k and 802.15.4g systems. The theoretical performance is evaluated by (2) and (3) as follows [7].

Equation (2) provides bit error probability of BPSK modulation with FEC (R=1/2, k=7, Soft decision), and equation (3) yields that of Gaussian filtered FSK (GFSK) modulation. As indicated by the simulated and theoretical performance, the 802.15.4k and 802.15.4g simulation systems work correctly. In next section, we use the simulation system to evaluate the interference of both systems.

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{R \cdot k \cdot \frac{E_b}{N_0}} \right) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{7}{2} \cdot \frac{E_b}{N_0}} \right) \quad (2)$$

$$P_b = \frac{1}{2} \exp \left(-\frac{E_b}{2N_0} \right) \quad (3)$$

TABLE I. MAJOR SYSTEM PARAMETERS FOR 802.15.4K AND 802.15.4G

Parameter	802.15.4k	802.15.4g
Bandwidth	200 kHz to 1 MHz	200 / 400 kHz
Chip / Bit Rate	Up to 800 kcps	50/100 kbps
Modulation	BPSK, OQPSK	GFSK
Filter	Roll off factor = 0.2	Gaussian filter BT=0.5
FEC	Convolutional coding Rate=1/2, K=7	No *
Spreading factor	16-32,768	No
Information bits per packet	32 octets	32 octets
Target PER	0.01	0.01

* -Optional (Convolutional coding Rate=1/2, K=4)

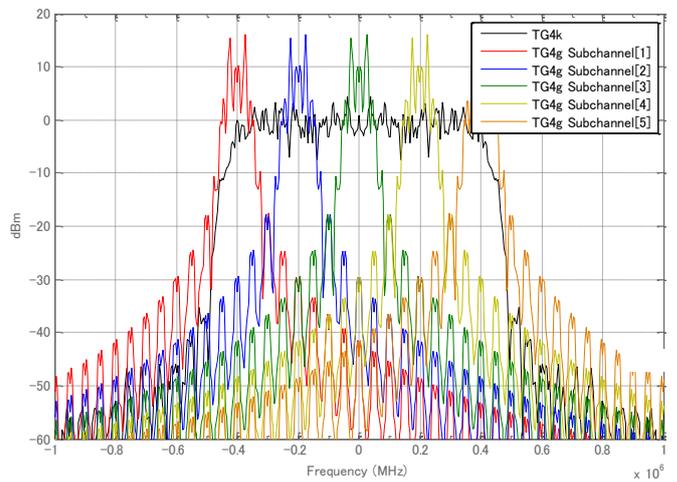


Fig. 1 Interference example of 802.15.4g signals with a bandwidth of 200 kHz and 802.15.4k signals with a bandwidth of 1 MHz

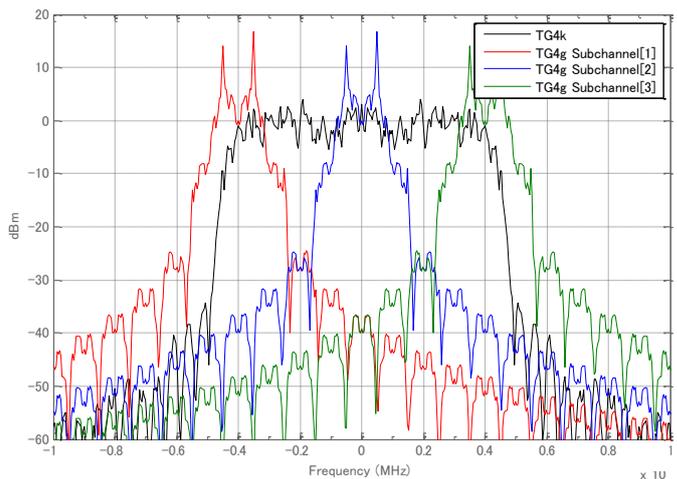


Fig. 2. Interference example of 802.15.4g signals with a bandwidth of 400 kHz and 802.15.4k signals with a bandwidth of 1 MHz

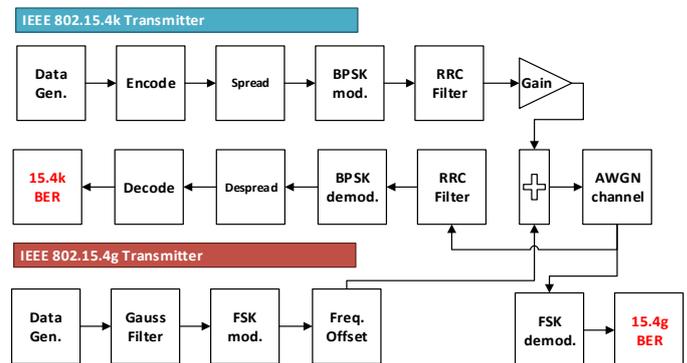


Fig. 3 A block diagram of the performance evaluation under CCI

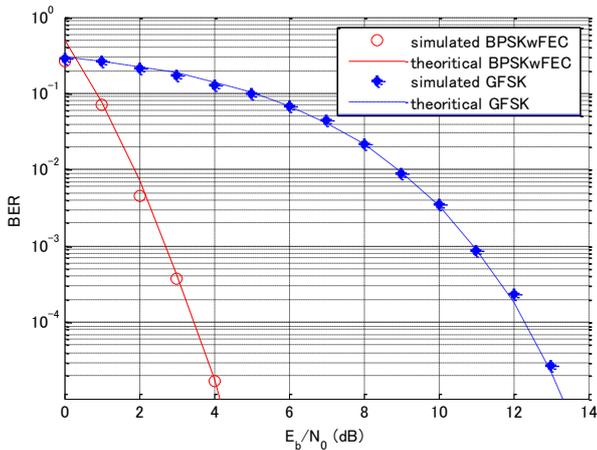


Fig. 4 BER performance comparison of 802.15.4k(BPSKwFEC) and 802.15.4g(GFSK) systems (theory and simulation)

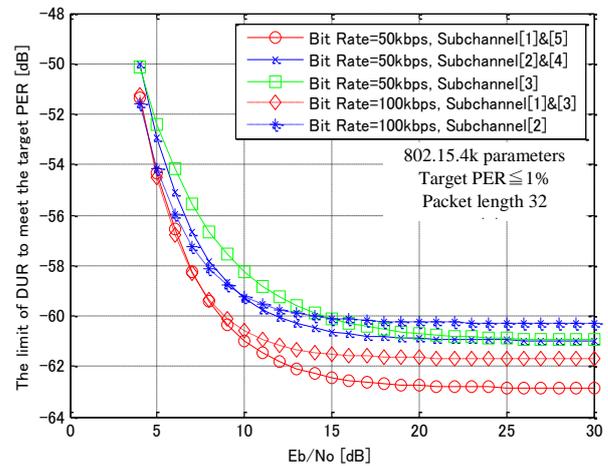


Fig. 5 Minimum D/U ratio vs. Eb/N0 of 802.15.4g under CCI from 802.15.4k with a 1023 spreading factor

III. THE MINIMUM D/U RATIOS OF BOTH SYSTEMS

A. 802.15.4k DSSS system performance

We evaluate first the minimum D/U ratios of 802.15.4k systems with a spreading factor of 1023 to meet the target PER under CCI of 802.15.4g operating at different data rates (bandwidths). Fig. 5 plots the minimum D/U ratios versus Eb/N0 to meet the target PER with parameters of bandwidth and relative channel locations of 802.15.4g signals against 802.15.4k signals. For 802.15.4k systems, the target bit error rate (BER) is calculated as 3.9×10^{-5} to transmit 32-byte packets with less than 1 % packet error rate.

The minimum D/U ratio of 802.15.4k (with 1023 spreading factor) under CCI from 802.15.4g tends to be less than -60 dB for Eb/N0 values higher than 15 dB, regardless of 802.15.4g bandwidths and sub-channels. In the case of narrower bandwidth interference, 200 kHz bandwidth 802.15.4g interference, the interference on the center of the 802.15.4k channel generates the largest degradation in the low Eb/N0 region. In the case of 400 kHz bandwidth interference, both non-centered 802.15.4g signals (sub-channels 1 and 3) give less degradation due to the drop in interference power created by passage through a 802.15.4k receiver filter.

B. 802.15.4g GFSK system performance

Fig. 6 plots Eb/N0 versus minimum D/U ratio of 802.15.4g under CCI from 802.15.4k with spreading factor of 1023. The target BER is the same value as the previous one. The minimum D/U ratios of 802.15.4g show up to 4 dB difference depending on the bandwidths and relative interfering channels (sub-channels) in the high Eb/N0 region.

From these two figures, 802.15.4k DSSS systems show much stronger interference-resistant performance than 802.15.4g FSK systems such as a -60 dB D/U ratio requirement for DSSS systems and 2-4 dB D/U ratio requirement for 802.15.4g FSK systems to realize a PER of 1%. This huge difference is due to the DSSS processing gain & de-spreading of the interference and FEC coding gain.

Using the performance results obtained above, the next section we will estimate the range / area in which 802.15.4g systems can disturb 802.15.4k communications and vice versa.

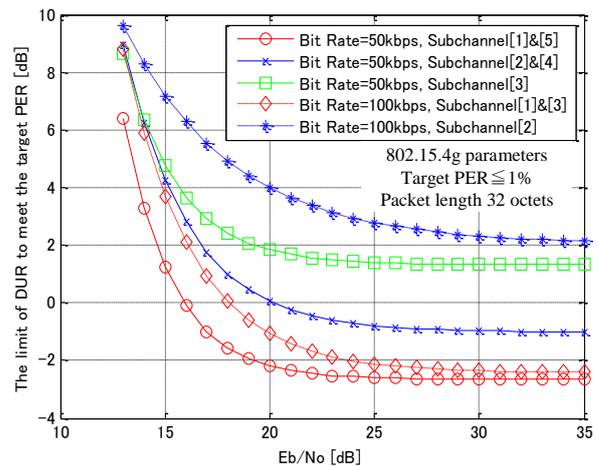


Fig. 6 Minimum D/U ratio vs. Eb/N0 of 802.15.4k (with spreading factor of 1023) under CCI from 802.15.4g operating with different bandwidths and sub-channels

IV. INSTALLATION INHIBITED AREAS OF 802.15.4K AND 802.15.4G SYSTEMS

In this section, we evaluate “installation inhibited area” of 802.15.4g and 802.15.4k systems. We assume two things. First is that both systems operate as one-to-one communications. Second is that the 802.15.4k system uses a spreading factor of 1023. Simulation parameters are shown in Table II. Fig. 7 shows the “installation inhibited area” from the Sensor Node of 802.15.4g to the BS or Sensor Node (SN) of 802.15.4k as an interferer. Fig. 8 shows the “installation inhibited area” from the SN of 802.15.4k to the BS or SN of 802.15.4g as an interferer.

In this simulation environment, the receiver location is shown by $(x, y) = (0, 0)$ meters. The transmitter is located at $(x, y) = (0, d)$ meters. “d” means the communication range in Table II. We assume that if the D/U ratio is less than the permissible D/U ratios (defined in section III), communication is successful.

Areas inside the contours in each figure indicate locations where the interferer transmitter can disturb communication, namely, “installation inhibited area”.

Figures 7 and 8 show that the 802.15.4g system is weaker against the interference. The “installation inhibited area” is 210 meters range, 105 percent of the communication distance. Fig. 8 indicates that spreading can greatly suppress 802.15.4g interference on 802.15.4k systems. As a result, “installation inhibited area” is a maximum of 100 meters, 2 percent of the communication distance. Finally, we conclude 802.15.4k systems have high tolerance to the interference due to spreading and FEC. The next section clarifies how the directive antenna reduces the interference compared with the omnidirectional one.

TABLE II. SIMULATION PARAMETERS

Parameter	802.15.4k	802.15.4g		
Center frequency	922.4 MHz	922.4 MHz		
802.15.4k Bandwidth	1 MHz	200 kHz		
EIRP	16dBm	16dBm		
Transmit power	13dBm	13dBm		
Base station antenna gain	3dBi	3dBi		
Base station antenna height	Urban	30m	Urban	30m
	Suburban	10m	Suburban	10m
	Rural	10m	Rural	10m
Sensor node antenna gain	3dBi	3dBi		
Sensor node antenna height	1m	1m		
Spreading factor	1023	No		
Communication range	5000m	200m		
Path loss model	Extended-Hata model[4]	Extended-Hata model[4]		

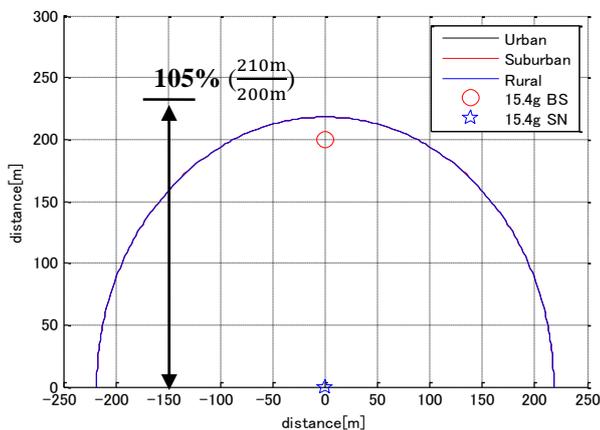


Fig. 7 “installation inhibited area” of 802.15.4g SN with interference from 802.15.4k SN or BS to victim 802.15.4g SN

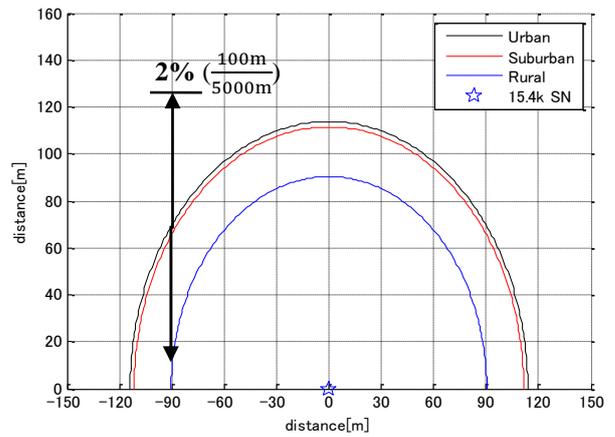


Fig. 8 “installation inhibited area” of 802.15.4k SN with interference from 802.15.4g BS or SN

V. INSTALALTION INHIBITED AREA REDUCTION BY DIRECTIONAL BS ANTENNA

In this section, we investigate the reduction of the interference from 802.15.4k to 802.15.4g and the one from 802.15.4g to 802.15.4k when 802.15.4k BS employs a directional antenna. The BS antenna gain pattern of the 802.15.4k system is based on a 4-element linear beam forming antenna as shown in Figure 9. We use this antenna with a gain of 10dBi and an omnidirectional antenna with a gain of 3dBi. Simulation parameters are the same as in Table II, except the transmit power for directive BS antenna is set the total transmit eirp does not exceed 16 dBm as regulated in Japan. Fig. 10 shows the “installation inhibited area”—of 802.15.4k under 802.15.4g interference for the two types of antennas. The “installation inhibited area” is evaluated by formulas (4) and (5) as follows. Formula (4) yields a continuous value, while the formula (5) yields a discrete value.

$$S = \frac{1}{2} \int_0^{2\pi} r^2 d\theta \tag{4}$$

$$= \frac{\pi}{N} \sum_{k=0}^{N-1} r_k^2 \tag{5}$$

With the omnidirectional antenna, the area is 0.04 km² but with the directive antenna it is 0.01 km². The directional antenna reduces the area by 75 percent relative to the omnidirectional antenna. Fig. 11 shows the “installation inhibited area” of 802.15.4g under 802.15.4k interference. The omnidirectional antenna offers an area of 0.15 km² while the directional antenna reduces this to 0.03 km². The directional antenna reduces the area by 80 percent relative to the omnidirectional antenna. Therefore, BS directional antenna can well suppress the interference and increase spectrum efficiency.

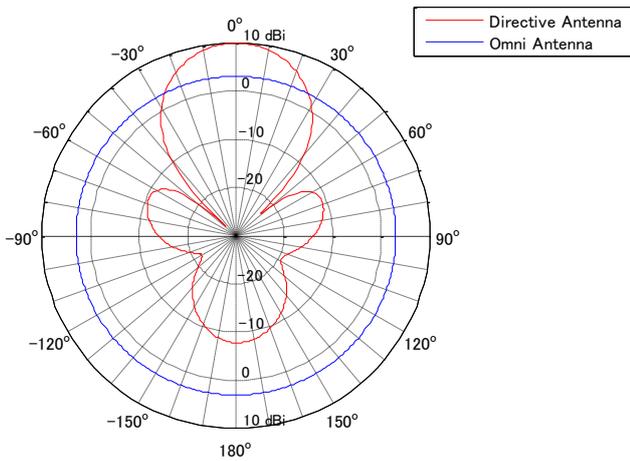


Fig. 9 802.15.4k BS antenna directivity pattern and omnidirectional antenna pattern

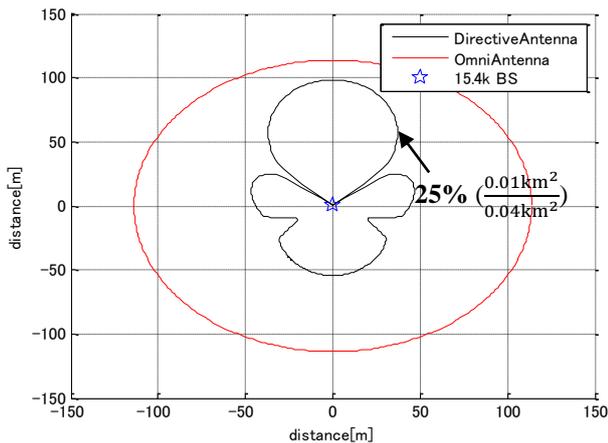


Fig. 10 The reduction of "installation inhibited area" by using directional antenna: interference from 802.15.4g

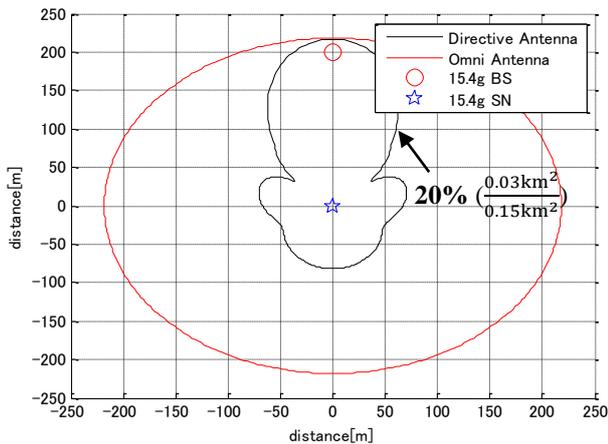


Fig. 11 802.15.4g "installation inhibited area" reduction by using directional BS antenna: interference from 802.15.4k BS

VI. CONCLUSIONS

In this work, we have investigated the performance of 802.15.4k and 802.15.4g systems under CCI. We confirmed that 802.15.4k DSSS Systems under CCI is 60 dB more robust than 802.15.4g. We have clarified "installation inhibited area" in which 802.15.4g or 802.15.4k cannot survive under CCI of the other system. The "installation inhibited area" of 802.15.4k turned out to be 50 times smaller than that of 802.15.4g. Furthermore, we have showed deployment of directional antennas can reduce the "installation inhibited area" (from other system) by 80 percent and by 75 percent (to other system). Therefore, 802.15.4k DSSS PHY and directional 802.15.4k BS antennas can reduce the interference much smaller and increase the spectrum efficiency a lot. 802.15.4k DSSS systems are suitable for wide area sensor networks with star topology and are much more widely deployable in interference dominated ISM bands than 802.15.4g systems with higher spectrum efficiency.

ACKNOWLEDGEMENT

This project is partially supported by Ministry of Internal Affairs and Communications, SCOPE and also partially supported by Japan Science and Technology Agency, A-STEP.

REFERENCES

- [1] "IEEE Draft Standard for Local and metropolitan area networks – Part 15.4: Low-Rate Wireless Personal Area Networks (WPANs) Amendment – Physical Layer Specifications for Low Energy, Critical Infrastructure Monitoring Networks (LECI-M)," IEEE P802.15.4k/D3, November 2013, pp.1-133, Jan, 2013.
- [2] "IEEE Standard for Local and metropolitan area networks--Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 3: Physical Layer (PHY) Specifications for Low-Data-Rate, Wireless, Smart Metering Utility Networks," IEEE Std 802.15.4g-2012 (Amendment to IEEE Std 802.15.4-2011), vol., no., pp.1-252, April 27 2012, pp.1-252
- [3] Rec. ITU-R F.1245-1, "Mathematical model of generalized radiation patterns of point-to-point fixed-service antennas for use in statistical interference assessment." [Online]. Available from: http://www.itu.int/dms_pubrec/itu-r/rec/f/R-REC-F.1245-1-200005-S!!PDF-E.pdf
- [4] Rep. ITU-R SM.2028-1, "Monte Carlo simulation methodology for the use in sharing and compatibility studies between different radio services or systems." [Online]. Available from: http://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-SM.2028-1-2002-PDF-E.pdf
- [5] Q. Li and S. Jillings, "TG4k Coexistence Document," IEEE 802.15-12-0314-01-004k, August 2012, pp.1-69
- [6] Y. Tada, T. Baykas and S. Kato, "Performance Evaluation of 802.15.4k DSSS System in the Presence of 802.15.4g Co-Channel Interference," Global Symposium on Millimeter Wave 2013, April 2013.
- [7] Proakis, J. G., Digital Communications, 4th ed., McGraw-Hill, 2001.

A Real-Time Bridge Pier Scouring Monitoring System Based on Hall-Effect Sensors

Chen-Chia Chen, Ssu-Ying Chen, Wen-Ching Chen, Gang-Neng Sung, Jin-Ju Chue, Chih-Ting Kuo,
Yi-Jie Hsieh, Chih-Chyau Yang, Chien-Ming Wu, and Chun-Ming Huang
Applied Research Laboratories, National Chip Implementation Center,
7F, No. 26, Prosperity Rd. I, Science Park,
Hsinchu City, Taiwan.
Email: ccchen@cic.narl.org.tw

Abstract—Scour around bridge pier is the major cause of bridge failure such as collapse resulted in loss of life and property. Most of available bridge pier scour sensors and approaches are very expensive, which is a challenge for mass deployment of numerous bridges. Our proposed scour monitor system utilized low-cost commercial sensors, hall-effect sensors (unit price < \$1) that are capable of real-time measuring bridge pier scour with resolution of ca. 2.5 cm, and overall cost for single sensor node of my proposed work is at least 40% less expensive than existing work. After scour event, the typical output voltage difference of ~ 10 mV and the signal-to-noise ratio of ~ 10 were observed. After simple modified setup, the output voltage difference could be reached to ~300 mV. Moreover, the master-slave architecture of bridge pier scour monitoring system has scalability and flexibility for mass deployment. This technique has the potential for further widespread implementation in the field.

Keywords-bridge; piers; scour; senso; hall-effect

I. INTRODUCTION

In the past few decades, global warming has increased dramatically in rainfall intensity, duration, and frequency, which resulted in harsh floods in Taiwan. Nevertheless, most mountains in Taiwan are very steep with slope gradients, so rivers in Taiwan are usually short and steep. When typhoons come and bring intensive rainfall, resulted in serious floods, even disaster flow [1]. It usually causes tremendous damages and loss of life and property. According to 2012 annual report by directorate general of highways, MOTC, there are ~ 9699 bridges of highway in Taiwan area, total length ~ 502021.8 m. Some of the crossing river bridges face serious challenges of bridge foundation scouring problem during harsh floods and disaster flow. Bridges lose their piers due to excessive pier scour and high flow velocities, which is one of the major causes for partial bridge collapse [2].

For more than a hundred years, bridge pier scour has been extensively studied in the world. Many methodologies and instruments have employed to measure and monitor the local pier scour depth, such as sonar, radar, and Time-Domain Reflectometry (TDR), Fiber Bragg Grating (FBG) sensor [3]. The sonar and radar sensor provide contactless measurement of streambed scouring near bridge pier and abutments, and usually used to show the final status of streambed after a flood. One of disadvantages of the sonar and radar is that they have limit for measuring status of streambed in real time as rush water contained sands, even rocks during a flood. A method of TDR measures the reflections that results from a fast-rising step pulse travelling

through a measurement cable. The depth of soil-water interface is determined by counting the round trip travel time of the pulse. However, the major drawback of TDR is that accuracy of TDR is strongly depended on environment temperature and humidity. Moreover, monitoring scour depth by FBG is depended on number of FBG elements. However, the cost of monitoring of the scour depth by FBG technique is higher than that of existing methods [4]. The costs of Radar and TDR are expensive due to high-speed hardware requirement. For example, a commercial TDR (Campbell Scientific Inc., TDR100) was used to real-time monitor scour evolution [5], and its price is about \$250. For FBG, optical devices such as laser, photo detectors and the optical fibers are very expensive. In addition, most of the existing methods used for scour detection are expensive and complicated, which is a major challenge for mass deployment to a lot of bridge piers.

In this study, we develop and verify proposed real-time bridge pier scouring monitoring system which has a gateway and sensor nodes as master-slave configuration. Each sensor node has a hall-effect sensor module; overall cost for single sensor node is about ~\$100, the cost includes components and Printed Circuit Boards (PCB) manufacturing and assembly. Our proposed solution is 40% less expensive than existing work (TDR). Furthermore, we also developed a Bridge Surveillance program for remotely accessing raw sensor data. The experimental results show that our proposed monitoring system can monitor pier scour process in real-time.

We start with an overview of our proposed architecture and experimental setups in related work in Section II. Section III presents the results provided from our experiments. Finally, we conclude this paper in Section IV.

II. DEVELOPMENT OF REAL-TIME BRIDGE PIER SCOUR MONITORING SYSTEM

Architecture of bridge pier scouring monitoring system

The architecture of real-time bridge pier scour monitoring system is shown in Figure 1. The architecture is based on master-slave configuration. A master sends commands to slave for controlling sensor node and accessing sensor data. The host controller communicates with gateway through Power Over Ethernet (POE) and Ethernet switch. The host controller sends a command to the gateway. When the gateway receives command, the gateway converted Ethernet command to RS485 command. After converting command, the gateway broadcasts to sensor

nodes. Since the command packet contains unique sensor ID, only specific sensor node returned sensor data to the host controller. We adopt both accelerometer and hall-effect sensor modules in our sensor node. In order to reduce number of wires in this study, POE is utilized. Simply connected 48V battery (3 packs in series for 48V with individual 16V lithium iron phosphate battery) to the POE adaptor (Cerio, POE-PE03), and used an Ethernet cable to connect the POE adaptor to the gateway and sensor node in series.

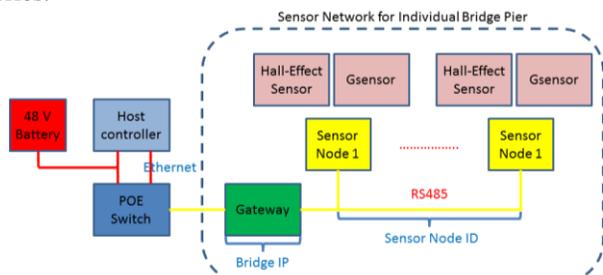


Figure 1: The architecture of real-time bridge pier scouring monitoring system

Gateway and Sensor nodes

The gateway is comprised of two stacked PCBs – a power module and a core module (see Figure 2). The top board is the power module, which operates as a DC-DC converter for creating 1.2~5V outputs from the 48V input. An Ethernet PHY (TI, DP83640) is used to receive and send Ethernet data, and also to send signals and power to sensor nodes through RS485 interface (ADI, ADM2682E). The core module is composed of a Cortex-R4 MCU (TI, RM48L952) and a FPGA (Xilinx, Spartan-6). Ethernet data and RS485 data are processed by the Cortex-R4 MCU and the FPGA, respectively. The FPGA mainly translates the sensor data from serial format to parallel format, between the FPGA and the Cortex-R4 MCU have three control signals (Int, Rdy, En) and 8-bit data signals. The FPGA receives the sensor data in 8-bit as a unit, after the FPGA collected 8-bit data, the FPGA will deposit to register, then send Int signal = 1 to the Cortex-R4 MCU, notify the Cortex-R4 MCU can receive sensor data. After the Cortex-R4 MCU receives 8-bit data, the Cortex-R4 MCU will set the Rdy signal to send a plus for the FPGA, followed by cycle until the FPGA En Signal = 0, on behalf of the sensor data has been transferred.



Figure 2: The photos of printed circuit boards of gateway (left) and sensor node (right).

The configuration of sensor node was similar to the gateway. The Cortex-R4 MCU is used to access sensor data through SPI (Serial Peripheral Interface) interface, and the FPGA is used to process RS485 data. The block diagram of the FPGA in sensor node is shown in figure 3.

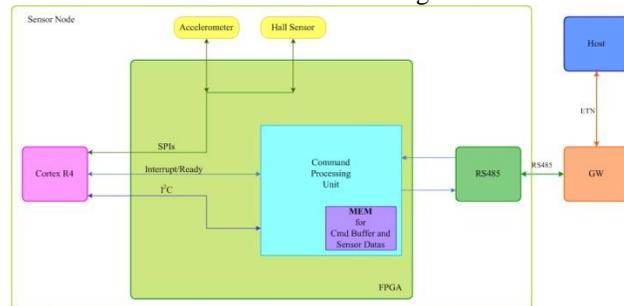


Figure 3: Block diagram of FPGA in sensor node

The FPGA parses receives commands, executes part of commands, and responses to the host. The Cortex-R4 MCU takes charge of collecting sensor data. Figure 4 describes processing sequence of the FPGA and the Cortex-R4 MCU. In Figure 4, the steps in blue are tasks of the FPGA, those in purple are memory related tasks, and those in red are tasks of the Cortex-R4 MCU. In the case that the host requests sensor data, the FPGA will receive a Read command. The FPGA then parses and decodes the command and is aware that cooperation with the Cortex-R4 MCU is necessary. The FPGA puts this command in memory and notifies the Cortex-R4 MCU with an interrupt. The Cortex-R4 MCU reads command from memory via I2C, collects sensor data and stores them in memory. After collection is done, the Cortex-R4 MCU notifies the FPGA with a General-Purpose Input/Output (GPIO) signal. Then the FPGA reads data from memory and generates response to the host.

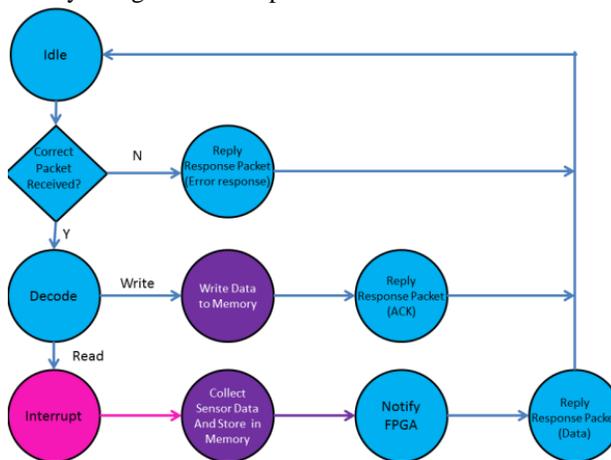


Figure 4: Processing sequence of FPGA and MCU

Accelerometer and hall-effect sensor modules

The core module of the sensor node which is connected to an accelerometer (ADI, ADXL345) and a hall-effect sensor (Allegromicro, A1301) which is used in this study are widely available online. But in this study, we only focus on the monitoring scour event by using the hall-effect sensors.

The output voltage of the hall-effect sensor is returned to an ADC (TI, ADS 1258) to digitize its analog signals, as shown in Figure 5. Digitizing output voltage of the hell-effect sensor sends back to the Cortex-R4 MCU via the SPI interface.



Figure 5: Top-view and bottom-view photos of the hall-effect sensor

Setup of the experimental tests

A Neodymium magnet with a diameter of 8 mm and thickness of 3 mm is fixed on thin metal strip with thickness of 0.3 mm, as shown in Figure 6. The hall-effect sensor and Neodymium magnet are aligned well, and are separated by a distance of 10 mm. Both types of sensor modules are installed along the pier model, as shown in Figure 7.

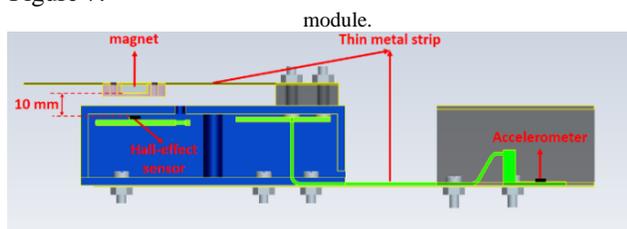


Figure 6: The drawing of house for accelerometer and hall-effect sensor module.

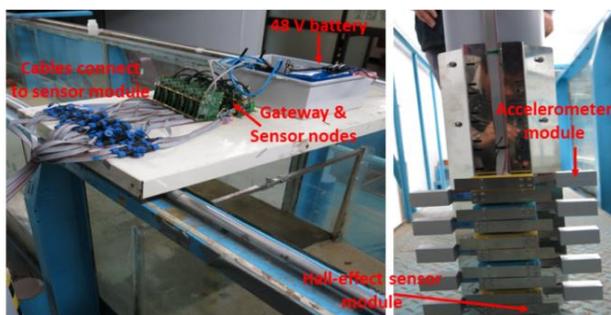


Figure 7: The photos of setup of real-time bridge pier scouring monitoring system.

The monitoring bridge scour erosion detection is carried out in a recirculating laboratory flume (length = 36m, width = 1 m, depth = 1.1 m) at Hydrotech Research Institute of National Taiwan University, Taiwan [6]. The layout of the flume and experimental setup are shown in Figure 8. A false test bed has a sediment recess (length = 2.8 m, width = 1 m, depth = 0.3 m) which is filled by nearly uniform sediment. A 15-cm-diameter hollow cylindrical pier made of plexiglass is located at the middle of the recess. An inlet valve and a

tailgate are used to regulate depths of flow and flow speed.

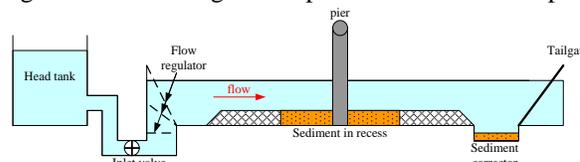


Figure 8: Partial layout of recirculating laboratory flume.

Gateway Initialization

The procedure including gateway initialization and sensor data access are shown in Figure 9. A Bridge Surveillance program is ran in the host controller, and it sends command (Request Packet) to specific gateway for making TCP connection through Ethernet. When the gateway receives ACK from the host controller and returned SYN/ACK, the TCP connection is made successes. The host controller sends a command to disconnect an active connection when the busy gateway does not return SYN/ACK. After making TCP connection, the gateway firstly initializes GPIO of the FPGA. Secondly, it converts Request Packet to serial data, and then it broadcasts serial data to sensor nodes. Until FPGA signal = 1, the FPGA receives sensor data from sensor node, and then the gateway sends the Response Packet to the host controller.

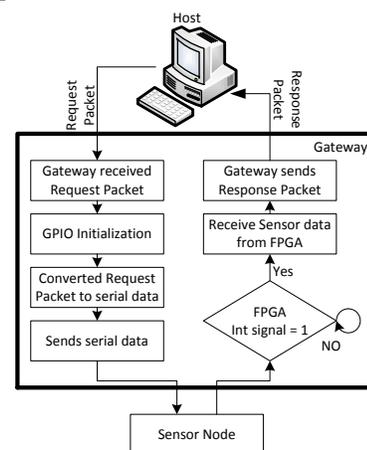


Figure 9: The flow chart of Gateway initialization

Figure 10 shows the Bridge Surveillance program with friendly graphical user interface, which automatically collects data from accelerometer and hall-effect sensor modules.

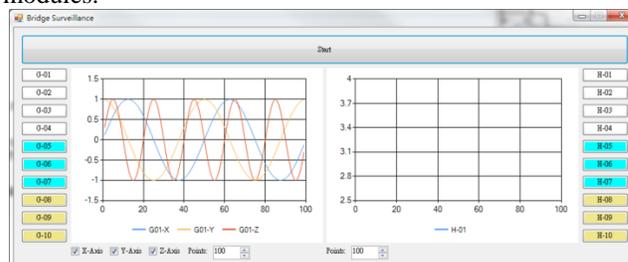


Figure 10: A Bridge Surveillance program with friendly graphical user interface (simulation data).

The host controller sends command every 20 ms, and the host controller will receive the Response Packet. The proposed software extracts the sensor data from payload of the Response Packet and then it displays them in real-time on charts (left: accelerometer, right: hall-effect sensor) of GUI as shown in Figure 10. Therefore, remote users can access raw data of each sensor nodes by WiFi communication, and they can analyze them in-situ. If pier scour status reaches critical condition of bridge collapse, the next version of the Bridge Surveillance program will alarm people away from the bridge.

III. RESULTS AND DISCUSSIONS

A simple method for calibrating the measuring gap between the hall-effect sensor and the magnet will be shown. A permanent magnet is moved away from the hall-effect sensor, and the dependence between corresponding output voltage of the hall-effect sensor and the distance between the magnet and the hall-effect sensor is shown in Figure 11. The output voltage of the hall-effect sensor as a function of the distance is nonlinear, but can be expressed by the following equation,

$$y = 7.513 - 1.055x + 0.079x^2 - 0.002x^3 \quad (1)$$

where x is the distance between the magnet and the hall-effect sensor output voltage in millimeter and y is the corresponding output voltage of hall-effect sensor distance in volt. We can use look-up table or transfer function to quickly find out the gap between the magnet and the hall-effect sensor. Besides, performance of a hall-effect sensor was guaranteed over an extended temperature range from its datasheets. We did not make temperature calibration for the hall-effect sensor prior to starting experiments.

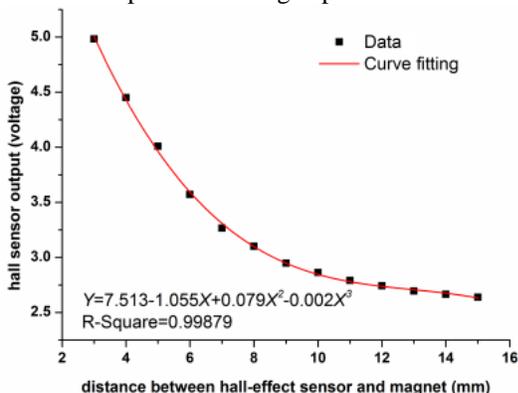


Figure 11: The data and fitting curve of the output voltage of the hall-effect sensor versus distance between hall-effect sensor and magnet.

When the maximum number (1024) of sensor data is reached, the host controller creates a new file to a disk. In order to save disk space, a gzip utility is used for compression of the sensor data. After uncompressed file, the sensor data were shown in Figure 12, each line represents a

data and time stamp, sensor ID, output voltage of a hall-effect sensor and ends with a newline character.

hall-effect sensor			
Date and time stamp	sensor ID	output voltage of sensor	
2014-01-06 16:19:06.969	@ H-2	: 2.72722752888997	
2014-01-06 16:19:07.269	@ H-2	: 2.72717793782552	
2014-01-06 16:19:07.419	@ H-2	: 2.72728665669759	
2014-01-06 16:19:07.439	@ H-2	: 2.72714233398438	
2014-01-06 16:19:07.469	@ H-2	: 2.72726631164551	
2014-01-06 16:19:07.499	@ H-2	: 2.72732543945313	
2014-01-06 16:19:07.529	@ H-2	: 2.72725931803385	
2014-01-06 16:19:07.559	@ H-2	: 2.72722053527832	

Figure 12: Hall-effect sensor record file.

Figure 13 shows the evolution of signals of the hall-effect sensors during the monitoring scour experiment. A cylinder pier model is used for these tests. Some of the hall-effect sensor modules (sensor module 4-10) are completely submerged under sand; the others (sensor module 1-3) are left in the air.

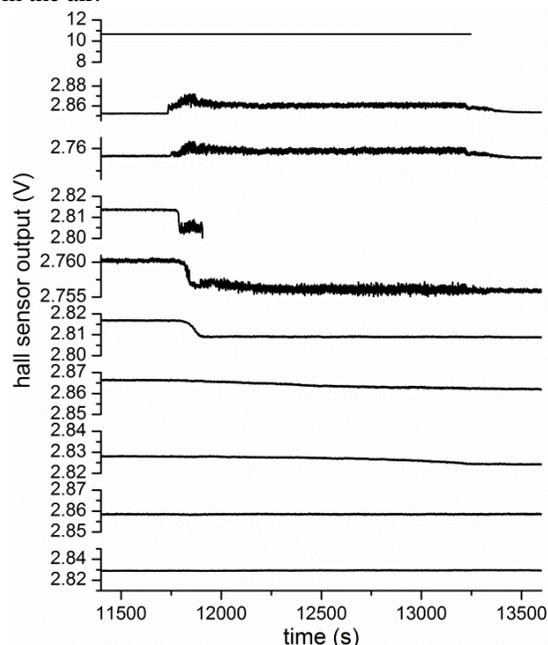


Figure 13: Real-time monitor output voltages of all of hall-effect sensor modules during scour experiment. Plots from top to bottom were corresponded from hall-effect module 1 to 10.

At ca. 10000 s, the inlet valve is ON. At 11733s, water flow starts striking the thin metal strips of the hall-effect sensor module 2 and 3, and their output voltage abruptly grows up due to bending of the thin metal strips caused by the water flow. It is worth knowing that quiver of voltage of the hall-effect sensors were strongly depended on the rate of flow water (data was not shown in here). The output voltages of the hall-effect sensor module 4, 5, and 6 are suddenly dropped down at 11787, 11834, and 11868, respectively. The out voltage of the hell-effect sensor drops so quickly meaning that the sand around the sensor node is totally removed by scour process. However, the rate of change of the output voltage of the hall-effect sensor module 7 and 8 are slower than that of module 4, 5, and 6. It reflects that the scour processes are slow and sand around the hall-effect

sensor module is just partially removed during scour process. The output voltages of the hall-sensor module 9 and 10 almost have no clear drop, so the sand still covers them well. Moreover, the hall-sensor module 1 and 4 are broken due the leaking water from imperfect silicon package sites. Figure 14 shows the pier model before, under, after scour experiment.

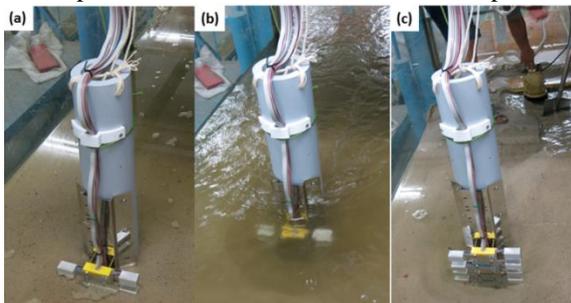


Figure 14: Photos of pier model (a) before, (b) under, and (c) after scour experiment.

In order to sense the scour event more sensitivity by the hall-effect sensor, we try to amplify the change of output voltage of the hall-sensor module after scour event, as shown in Figure 15..

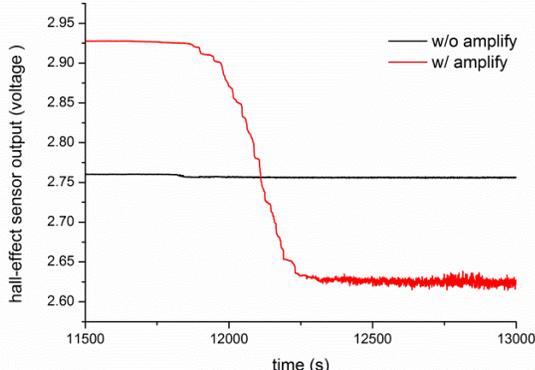


Figure 15: Real-time monitor output voltages of the hall-effect sensor modules with and without amplified signals during scour experiment.

We know that the output voltage of the hell-effect sensor strongly depends on the distance of the magnet and the hall-effect sensor. Therefore, we make the distance between the magnet and the hall-effect sensor get closer during install sensor node. The change of voltage of the hall-effect sensor is enhanced ca. thirty fold that of before original setup, as shown in Figure 15.

IV. CONCLUSION

Comparison of existing system for monitoring bridge scour is shown in Table I. In this study, we have proposed the architecture of real-time bridge pier scouring monitoring system featuring hall-effect sensors. Our proposed real-time monitoring system offers the advantage of low-cost and easy mass deployment. The overall cost for single sensor node of our proposed work is at least 40% less expensive than existing solutions (TDR).

TABLE I. COMPARISON OF EXISTING SYSTEM FOR MONITORING BRIDGE SCOUR

	TDR	Radar	FBG	This work
Cost (\$)	>250	>3500	240 ^a	~100
Resolution (cm)	2.5	10	10	2.5
Temperature effect	Yes	Yes	Yes	NO
work environment	Water; Sand	Air	Water; Sand	Water; Sand
Real-time	Yes	No	Yes	Yes

a. This price is only for sensor.

Our bridge pier scour monitoring system with a solution of 2.5 cm is demonstrated. Furthermore, we also have developed the Bridge Surveillance program with friendly GUI, which automatically collects data of each sensor node. Therefore, it is practical to use our proposed real-time scour monitoring system for diagnosis bridge pier scour events. Next step, we will further widespread implementation in the field.

REFERENCE

- [1] J. C. Chen, C. D. Jan, and W. S. Huang, "Characteristics of Rainfall Triggering of Debris Flows in the Chenyulan Watershed, Taiwan," *Natural Hazards and Earth System Sciences*, vol. 13, 2013, pp. 1015-1023.
- [2] H. Wang, S. C. Hsieh, C. Lin, and C. Y. Wang, "Forensic Diagnosis on Flood-Induced Bridge Failure. I: Determination of the Possible Causes of Failure," *Journal of Performance of Constructed Facilities*, vol. 28, Feb. 2014, pp. 76-84.
- [3] L. J. Prendergast, "A Review of Bridge Scour Monitoring Techniques," *Journal of Rock Mechanics and Geotechnical Engineering* vol. 6, 2014, pp. 138-149.
- [4] J. Y. Lu, J. H. Hong, C. C. Su, C. Y. Wang, and J. S. Lai, "Field Measurements and Simulation of Bridge Scour Depth Variations During Floods," *Journal of Hydraulic Engineering-Asce*, vol. 134, Jun. 2008, pp. 810-821.
- [5] X. B. Yu, and X. Yu, "Laboratory Evaluation of Time-Domain Reflectometry for Bridge Scour Measurement: Comparison with the Ultrasonic Method," *Advances in Civil Engineering*, vol. 2010, 2010, pp. 1-12.
- [6] Y.-B. Lin, J.-S. Lai, K.-C. Chang, W.-Y. Chang, F.-Z. Lee, and Y.-C. Tan, "Using MEMS Sensor in the Bridge Scour Monitoring System," *Journal of the Chinese Institute of Engineers*, vol. 33, Jan, 2010, pp. 25-35.

Building Automation: Experience with Dynamic Reconfiguration of a Room

Maxime Louvel, François Pacull, Safietou Raby Thior, Maria Isabel Vergara-Gallego, Oussama Yaakoubi
 Univ. Grenoble Alpes, F-38000 Grenoble, France
 CEA, LETI, MINATEC Campus, F-38054 Grenoble, France
 firstname.lastname@cea.fr

Abstract—This paper details a case study in the domain of building automation systems. This case study is the reconfiguration of a room that can be split in two or merged in one according to the current needs. The building is equipped with the LON system, a standard in building automation. Such a reconfiguration is normally done manually by a skilled technician. Thanks to our approach, it can now be autonomous and triggered by various external events such as sensor readings, a remote controller or information from an agenda. To achieve dynamic reconfiguration in this context we rely on LINC, a coordination middleware providing an abstraction layer allowing the encapsulation of hardware and software components. Thus, we can not only integrate sensors and actuators but also legacy systems such as the LON configuration tool. The coordination is provided by coordination rules which can be statically defined at application deployment or dynamically added or removed to reconfigure the system.

Keywords—Coordination; middleware; building automation.

I. INTRODUCTION

Advances in consumer electronics and embedded systems have leveraged the emergence of connected smart devices for a huge number of applications. These devices communicate using different protocols and physical medium. Such explosion of heterogeneous devices and services, and the necessity of interaction between them, have yielded architectural concepts based on a common software layer called *middleware*. A middleware provides an abstraction view of heterogeneous hardware, services, and protocols. It hides heterogeneity to the application and facilitates software re-utilisation, application deployment, and coordination between different devices to achieve a common goal.

Building Automation Systems (BAS) are a typical example of systems requiring a middleware. Indeed, a BAS contains a huge number of devices and services that use different protocols and that need to interact between them. A BAS typically has the following requirements: i. The *necessity of a common software layer* to access heterogeneous devices, ii. The *compatibility with legacy systems and protocols* to communicate with existing infrastructures and extend BAS capabilities and iii. The *possibility to dynamically reconfigure the system* to allow special operating modes, adaptation to changing devices and space redistribution. Existing middlewares [1] [2] [3] only partially fulfil these needs. They are either too complex to allow dynamic reconfiguration or they focus on a specific part of the BAS needs.

In this paper, we use LINC [4] to answer these challenges. LINC is a resource-based middleware that provides interactions between connected devices or software components through coordination rules. Several technologies have already been encapsulated for building automation. For instance, in [5] LINC was used to control interactions between different devices. In the present paper, we describe another angle of LINC

where it is used as an administration layer where devices can be dynamically reconfigured and interactions between entities belonging to the same technology may be changed according to a given context. In this case, LINC is required only for the reconfiguration phase.

This paper presents an example where a BAS, that uses the Local Operating Network (LON) [6] platform, is dynamically reconfigured to allow space reconfiguration. In the case study we consider a room that is split in two (or merged in one) when a removable wall is closed (or opened). Then, all the devices present in the room are automatically reconfigured according to the position of the wall. The rules describing all the actions to perform for reconfiguration are automatically generated, compiled and executed. These actions include removing the current configuration and putting in place the new one. Additionally, LINC provides the possibility to extend the capabilities of the BAS thanks to the encapsulation of new components.

The structure of this paper is as follows. Section II presents some related work; then, Section III describes the LINC middleware and Section IV describes the architecture put in place for dynamic reconfiguration. The case study is described in Section V along with some discussions and results. Finally, Section VI concludes the paper.

II. RELATED WORK

A BAS concerns the control of building services. Extending or modifying the behaviour of an existing building automation infrastructure requires dealing with heterogeneity of communication protocols and legacy systems [7]. There are lots of communication protocols and many of them coexist on the same building.

Therefore, a middleware or abstraction layer [8] is needed in such scenario. Some characteristics as scalability, fault tolerance and flexibility are desired when choosing the proper middleware. For instance, scalability allows extending the capabilities of the BAS by introducing new communication technologies. Wireless technologies are becoming very attractive given its flexibility and easiness to install and deploy [9] [10]. Some efforts trying to integrate legacy BAS technologies and wireless sensor networks technologies can be found in the literature [11] [12]. Besides, some middleware solutions for BAS have been proposed [2] [3]. However, since they target a specific technology or application, these solutions lack of flexibility. Scalability may also be an issue given the complexity of such approaches.

Regarding reconfiguration, it can be performed at different levels: i. the *nodes or devices level*, where the behaviour of the node is modified, i.e., sampling period, routing mechanism, sensing precision, transmission power and ii. the *system level* where relationships and coordination of devices are modified according to a new defined scenario. In a Wireless Sensor Network scenario, much effort has been put at the node

level; in this case, reconfiguration is mostly related to reliable communication between nodes and energy efficiency. Some examples of reconfiguration in Wireless Sensor Networks are presented in [13] [14]; furthermore, a middleware may facilitate this task [15]. In this paper, we focus on the system level reconfiguration which allows to adapt an existing system to a new defined scenario so that new coordination rules are generated and executed. Such reconfiguration may be triggered by an external event. Similarly, a context-aware middleware, allows adapting the application, based on information retrieved from sensors or events. Several context-aware middlewares can be found in the literature [16] [1] each of them focusing on a given set of applications. For the specific case of building automation Istvan et al. [17] propose a mechanism to reconfigure dynamically the relationships between devices and services in a building. This proposition remains as a conceptual approach and no real implementation is presented.

III. OVERVIEW OF LINC

Full description of the LINC middleware may be found in [4]. This section describes the very basic information to make the paper self-contained.

LINC provides a uniform abstraction layer to encapsulate the different software and hardware components. This layer simplifies the integration of legacy components and their coordination. This abstraction layer relies on the associative memory paradigm implemented as a distributed set of bags containing resources (tuples). Following the Linda [18] approach, bags are accessed only through three operations:

- `rd()`: takes a partially instantiated tuple as input parameter and returns a stream of fully instantiated tuples from the bag, where the fields match the given input pattern;
- `put()`: takes a fully instantiated tuple as input parameter and inserts it in the bag;
- `get()`: takes a fully instantiated tuple as input parameter, verifies if a matching resource exists in the bag and consumes it in an atomic way.

A. Examples of components encapsulated as bags

A typical sensor, measuring a physical quantity, can be modelled through a **Sensor** bag containing tuples, that are formed as $(sensorid, value)$. The measured quantity is then retrieved by accessing this bag. To differentiate sensors capabilities (temperature, humidity, and so on), a **Type** bag is added, containing tuples formed as $(sensorid, type)$.

Actuators may be modelled with a bag **command** designed as $(actuatorid, function, parameter1, parameter2)$. Then, when a resource is inserted with a `put()` operation, the bag triggers the awaited action over the targeted actuator adapted with the appropriate parameters. In a similar manner, any other component or service which provides a given protocol or API, such as SOAP, RPC, CORBA, or REST can be encapsulated in a bag or set of bags.

B. Object

Bags are grouped within objects according to application logic. For instance, all the bags used to control a network with a given technology are grouped in the same object.

C. Coordination rules

The three operations described above, i.e., `rd()`, `get()` and `put()`, are used within production rules [19] A pro-

duction rule is composed of a precondition phase and a performance phase.

1) *Precondition phase*: The precondition phase is a sequence of `rd()` operations which detect or wait for the presence of resources in several given bags. The resources are for instance values from sensors, external events or results of service calls. In the precondition phase:

- the output fields of a `rd()` operation can be used to define input fields of subsequent `rd()` operations;
- a `rd()` is blocked until at least one resource corresponding to the input pattern is available.

2) *Performance phase*: The performance phase of a production rule combines the three `rd()`, `get()` and `put()` operations to respectively verify that some resources (e.g., the one(s) found in the precondition phase) are present, consume some resources and insert new resources.

In this phase, the operations are embedded in one or multiple *distributed transactions* [20], executed in sequence. Each transaction contains a set of operations that are performed in an atomic manner. Hence, we can guarantee that actions that belong to the same transaction, are either all executed or none. This ensures properties such as:

- Some conditions responsible for firing the rule (precondition) are still valid at the time of the performance phase completion;
- All the involved bags are effectively accessible. For instance, for a bag encapsulating a remote service we can determine if such service can be actually accessed.

These properties are very important as they ensure that the set of required objects, bags and resources, are actually available “at the same time”. More properties offered by LINC in the BAS context are detailed in [5].

IV. APPLICATION ARCHITECTURE FOR DYNAMIC RECONFIGURATION

Figure 1 presents the architecture of an application based on LINC. The top layer is the application layer and it defines how devices and services interact with each other through the use of bags and coordination rules.

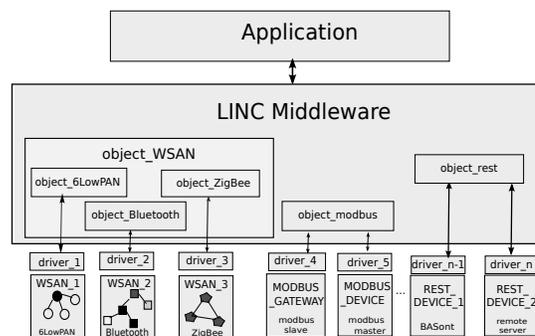


Figure 1: Example of application

The bottom layer corresponds to the hardware or service specific layer. The API and actions provided by each device are encapsulated so that they can be accessed through bags defined in LINC. To facilitate software re-use and evolution, and to reduce complexity, we have defined frameworks defining common features. For instance, all wireless sensor networks share a common set of bags for controlling the sensors and the actuators. Only the communication part which is dependent

on the technology is different. We have thus defined a WSAN (from Wireless Sensors and Actuator Network) object which is inherited by all the object responsible for a technology. Adding a new protocol is basically integrating the driver for the new technology. Hence, when the driver part has been written the object can be coordinated with the rest of the system. More details on this software architecture may be found in [5] [21].

In several applications, reconfiguration of devices at run time may be needed. For instance, we may need to change device parameters and the relationships between devices, or we may need to perform a software update.

Usually with such architecture, the coordination of devices of different technologies is done through coordination rules. These rules define the different configurations and interactions between devices. For instance, what action has to be done on the light when the button is pressed twice. Thus, reconfiguring is removing some coordination rules and replacing them by new ones, taking into account the new behaviour.

Alternatively, when all the concerned devices use a technology that integrates one controller, i.e., LON or KNX, we can use LINC to reconfigure these controllers. To do so, we extended the application described above with the encapsulation of the legacy system used to reconfigure the controllers of the devices. In the case of LON it is the LNS220 tool [22]. Then, the reconfiguration is done by dynamically inserting new coordination to reconfigure the different controllers.

The reconfiguration procedure will be started by the application as a response to a given event, i.e., a request from the user, the presence of a resource, or the value of a given sensor. We now detail how it is possible to dynamically generate rules and how the activated rules can controlled.

A. Dynamic rules generation

When performing reconfiguration, coordination rules must be consistent with the new configuration or scenario. To ensure this consistency, rules can be automatically generated corresponding to contextual information. In LINC, rules are seen as resources in bags. Hence, they can be added, enabled or disabled at run-time. To add a new rule, a resource is added in a dedicated bag of the object called `AddRules`. This bag receives resources of the form `(package, source)` where `package` is the logical name of the group of rules and `source` is the actual code of the rules.

When the reconfiguration is triggered, the device is reconfigured and new coordination rules are generated and added in the `AddRules` bag. When a resource is added in this bag, the rule is dynamically compiled. This compilation includes syntax verifications, and various checks to prevent potential issues at execution time. If the rule contains no detectable error, the object starts to execute it right away. At this point, the new scenario is set-up.

B. Control of rules execution

In a rule-based system it might be difficult to know which rule is executed at any time. Ensuring that a rule is really not active may be even more difficult. To overcome such issues, LINC uses a specific bag, called `RulesId`, containing the rules enabled. When a rule is compiled (when the system is started, or dynamically added as explained previously), an operation `rd(ruleId, "ENABLED")` is added in the beginning of the precondition and in every transaction of the rule. Hence, removing a rule only requires to remove the corresponding resource in the bag `RulesId`. Indeed, no new

instance of the rule may be started. If an instance reaches the performance phase, it will be aborted because the resource `rd(ruleId, "ENABLED")` is not in the bag.

Obviously, the same pattern may be applied to a group of rules by using a specific bag mapping the rules id to some other information that is meaningful for the application. A typical example in a BAS is to have two set of rules. The first set are the rules for daily life (e.g., managing lights, heating and comfort). The second set of rules are dedicated to emergency situation (e.g., to coordinate the lights with the evacuation procedure). In case of an emergency, the comfort rules must be disabled to prevent unexpected situation. This can be done with a bag containing the current mode i.e., `comfort` or `emergency`. A `rd("comfort")` (respectively a `rd("emergency")`) on the this bag is added in every precondition and every transactions ensuring that only one mode is used at a time.

Another important aspect of rules execution is that they can be run on a distributed way. This characteristic provides scalability and the possibility to integrate a considerable number of technologies. For the particular case of building automation, several BAS can be easily integrated. Moreover, rules can be replicated on several machines so that we guarantee its execution even if there is a faulty machine.

V. CASE STUDY: RECONFIGURATION OF A BUILDING AUTOMATION SYSTEM

The context of this case study is the SCUBA (Self-organising, Co-operative and robUst Building Automation) project [23]. It aims to provide novel architectures, services, and engineering methodologies for robust, adaptive, self-organising, and cooperating monitoring and control systems. The case study presented in this paper consists in reconfiguring a LON system when a room is split in two (or merged in one). This implies to change the binding between devices to adapt the system to the current situation: Which button trigger what and which sensor control what.

This case study illustrates a requirement more and more important in BAS where people want to optimise the room usages. To better explain the scenario, we first introduce some basic concepts and definitions regarding building automation and the LON technology, which is the imposed legacy technology. Then, we describe the reconfiguration procedure and we discuss the obtained results.

A. Building Automation

A BAS is a network of software and hardware components that sense, control, and act on the environment and communicate between them. It ensures the operational performance of the facility as well as the comfort and safety of building occupants.

Nowadays, it is possible to monitor and control several systems of a building. Typically, these systems work independently and can communicate between them thanks to a given interconnection technology which can be a standard or a proprietary solution. In general, several communication technologies are present in a building.

At deployment time, the interconnections between devices are defined according to the disposition of the building. Configuring such interconnections is complex, time demanding, and requires the intervention of qualified personal. Reconfiguration of the BAS may be required in order to respond to the needs of occupants, or to replace or upgrade some equipments.

Typically, these modifications are planned in advance and they are done on purpose by a technician.

Changes in the space configuration is something more dynamic, unplanned and that can be triggered by any occupant. Physical reconfiguration of a room needs to be synchronised with the involved equipments, i.e., sensors, actuators, controllers. For example, consider a button to switch on or off the lights in a room. If this room is combined with another room by removing a wall, the lights of the second room must also be connected to the button of the first room.

B. LON

LON [6] is a control network system designed by Echelon Corporation. LON is widely used in existing building automation systems and it allows communication between devices coming from different manufacturers using a common protocol called LONtalk. LON devices are physically linked together and they embed a special controller called the Neuron Chip. The latter is associated with a transceiver that allows its reconfiguration and communication with other devices. A unique identifier permits to identify inputs and outputs of each device on the network. Although devices are physically connected, they are not able to exchange messages unless a logical connection, called a Binding, is created between them. Figure 2 shows an example of two bindings between three devices. In the example, one of the Network Variable Outputs (NVO) of the Device_1 is associated to the Network Variable Inputs (NVI) of two devices: the Device_2 and Device_3.

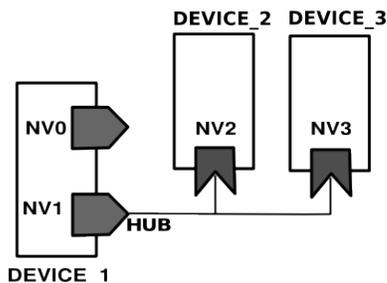


Figure 2: Binding between devices

LON networks are designed, monitored, and managed through a dedicated MS-Windows software stack. This stack is based on the LON Network Services (LNS) tool [22]. The tool gives access to a database storing LON designs and provides a complex API to interact with the network. An LNS Server is provided on top of the LON Networks as an ActiveX component. Usually, these tools allow reconfiguring the network connections by manually creating or removing bindings between network variables. However, the user can only act on one binding at a time. A classical room configuration typically contains dozens of bindings and reconfiguration needs to be done by a skilled technician. Hence, space reconfiguration with the LON technical tool rapidly becomes too costly and time demanding.

C. Reconfiguration of a Room

The platform used for the case study is illustrated in Figure 3. It is called T1 and is located in the Schneider premises in Grenoble, France and is one of the use case sites of the SCUBA project. It consists of two separate rooms (Room A and Room B) with a removable wall. The wall may be open

to combine both offices to form a single office, or it may be closed to obtain two different offices.

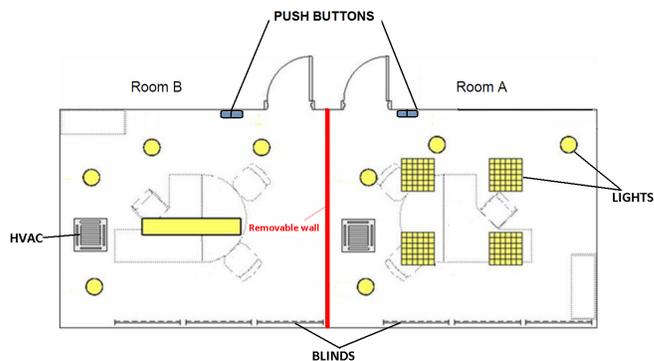


Figure 3: T1 Platform

Each office is equipped with sensors (temperature, luminosity, and presence), a Heating, Ventilation and Air Conditioning (HVAC) system, dimmable spotlights, motorised venetian blinds, and push buttons. All these devices use the LON technology.

To manage the space reconfiguration of the platform, we consider two configurations:

- **configuration 1:** The wall is open, i.e., the platform consists of a single room. In this configuration, LON bindings are created between push buttons (Room A and Room B) and actuators (lights, blinds) of both rooms. In addition, a single temperature sensor is used to control both HVACs. The two presence sensors are used with an "OR" logical and any of them can signal the room occupancy.
- **configuration 2:** The wall is closed i.e., the platform consists of two separated rooms. Both rooms have an individual control; then, push buttons of Room A only manage actuators of Room A, and the same applies for Room B. HVACs are controlled separately by their respective temperature sensor. The two presence sensors are independent.

Detecting the presence of the wall, in order to trigger reconfiguration, is done thanks to an external device that may use an arbitrary communication protocol encapsulated by LINC. In our case study, we have used a simple wireless magnetic sensor, that operates in the 433Mhz frequency band.

Changing the room configuration requires reconfiguring LON bindings. We now detail how the LON tools have been encapsulated in LINC, and how automatic reconfiguration is made possible. The approach significantly decreases the time and cost, compared to a manual reconfiguration.

1) *Encapsulation of LON in LINC:* For encapsulation, a software layer (driver) has been built on top of the LNS Server. This driver permits to access LON administrative services (for reconfiguration). The LONobject has been placed just on top of the driver layer and it contains bags associated to a specific LON network service. Then, requesting a given service is reduced to a simple standard operation on a bag (`rd()`, `get()` or `put()`).

The most important bags in our scenario are:

- *CreateBinding* A `put()` operation on this bag, with the binding information, creates the binding;

- *RemoveBinding*: A `put()` operation on this bag, with the binding identification, removes the binding.

Internally, when a `put()` operation is done on one of these bags, the corresponding APIs in the LON ActiveX layer are called, implementing the new binding configuration.

2) *Automatic Reconfiguration of LON*: Listing 1 gives an example of a generated rule to switch between two configurations. The rule starts with the removal of bindings and then, it creates bindings for the new configuration. The precondition and performance of a rule are separated by the symbol “::”. Hence, this rule has no precondition¹ and it is always triggered once.

This rule is automatically generated from predefined configuration stored in a dedicated component called BASont [21] provided by one of the SCUBA project partners. Alternatively, this information could be stored in a database.

```

::
{
# remove bindings
["LonObject", "removeBinding"].put("Locations.Room 2 , R2
  Push Button.Switch[1] ,
  nvoSWsetting[1]@@@Locations.Room 2 , R2 Room
  Box.ConstLightCtrl[0] , nviCLsetting") ;

#...
# create the new bindings
["LonObject", "createBinding"].put("Locations.Room 1 , R1
  Push Button Left.Switch[0] ,
  nvoSWsetting[0]@@@Locations.Room 1 , R1 Light
  Controller (LRC 5141).LightCotroller[0] ,
  nviChannel_1") ;
#...
}
    
```

Listing 1: Generated Rule

Once generated, the rule is inserted in the `AddRules` bag of the `LONObject` upon a condition concerning the presence of the wall. To do so, we use three rules detailed in Listing 2 and 3.

```

["LonObject", "Sensors"].rd("magnetic_1", "opened") &
::
{
["LonObject", "Sensors"].rd("magnetic_1", "opened") &
["LonObject", "TriggerConfig"].put("wall_opened") &
}.

["LonObject", "Sensors"].rd("magnetic_1", "closed") &
::
{
["LonObject", "Sensors"].rd("magnetic_1", "closed") &
["LonObject", "TriggerConfig"].put("wall_closed") &
}.
    
```

Listing 2: LINC rules that trigger reconfiguration

The first set of rules (Listing 2) takes the value of the magnetic sensor, that detects the position of the removable wall. The wall is typically moved by the user of the room. When a new value is sensed, the rule inserts the required configuration in the bag `TriggerConfig` of the `LONObject`. As observed in the rule, the information obtained from the magnetic sensor is used to trigger the reconfiguration process. Thus, we can also trigger very easily the change with an event coming from a remote controller or a time-based order from an agenda. We only need to add a rule for each case.

The second rule (Listing 3) generates the rules presented in Listing 1 when the wall position changes. For each configuration, there is a list of corresponding bindings which

are stored in the bag `Bindings`. This bag belongs to the object `BindingDB` and associates the list of bindings to the configuration: "wall_closed" or "wall_open". This bag also contains the list of all bindings to erase, associated to "clear_bindings", in order to have a clean configuration.

```

["LonObject", "TriggerConfig"].rd("T1", wallPosition) &
["BindingDB", "Bindings"].rd("clear_bindings",
  allBindings) &
["BindingDB", "Bindings"].rd(wallPosition,
  wallPositionbinding) &
COMPUTE generatedRule, Rulesname =
  lonFunctions.generateConfigRule(allBindings,
  wallPositionbinding, wallPosition) &
::
{
["LonObject", "AddRules"].put(Rulesname, generatedRule); ;
}
    
```

Listing 3: LINC rule that generates reconfiguration rules

The rule executes as follows:

- 1) A `rd()` operation on the bag `TriggerConfig` retrieves the configuration (wall position) when it changes (line 25);
- 2) the next `rd()` operation on the bag `Bindings` with the variant name "clear_bindings" retrieves the list of all possible bindings on the T1 platform independently of the actual configuration (line 26).
- 3) the next `rd()` operation is performed on the bag `Bindings` with the wall's position read from the bag `TriggerConfig` to get the list of bindings to be created (line 27).
- 4) the `generateConfigRule` method generates the LINC rule for the new room configuration following the format of the rule presented in (Listing 1). The `COMPUTE` operation permits calling an external method, in this case `generateConfigRule`.
- 5) the generated rule is put in the bag `AddRules` of `LONObject` in order to be executed to switch to the new configuration (line 31).

Note that, once the rule has been generated and executed, the middleware is not used anymore; the BAS continues to work autonomously thanks to the LON controller.

After the encapsulation of LON, the reconfiguration mechanism can be applied to any number of configurations involving any number of rooms. Thanks to LINC, the approach is not restricted to a given technology and it allows extending the BAS functionalities using new devices and services. Considering another standard would only require to develop the appropriate driver following the same architecture than for the LON system. Even if these types of systems are based on a quite heavy API, using several software levels, the process is repeatable. Once the driver is done, the mechanism for reconfiguration presented in this case study can be applied.

The possibility to add and remove rules dynamically, along with encapsulation of multiple technologies, provides a way to integrate an existing infrastructure with new sensors and actuators on a dynamic way. For instance, as soon as a new sensor is detected, we can execute new rules to allow its interaction with existing devices. In addition, we can also bridge two or more existing infrastructures that use different BAS technologies, i.e. LON, BacNet, KNX, Modbus. This bridging provides communication between different buildings or different rooms on a building for example.

Besides, LINC provides several characteristics that are

¹

desirable for applications such as building automation. Firstly, it allows scalability thanks to its distributed nature. LINC also provides graceful degradation so that alternative actions are executed when there is a system fault. In addition, LINC uses transactions guaranteeing the execution of all the actions defined in a rule, more details can be found in [5]. When reconfiguration is performed, the latter characteristic is fundamental, since it guarantees that the system is on a consistent state after reconfiguration.

VI. CONCLUSION

In this paper, we have presented our approach for dynamic reconfiguration in the context of building automation. We have relied on the coordination middleware LINC to provide a non intrusive reconfiguration of a legacy system. Once the BAS have been encapsulated in LINC, it is possible to reconfigure connected devices. Reconfiguration allows changing relationships between devices or software changes as a response to an event.

Given that LINC allows adding and generating new rules on a dynamic way, reconfiguration and adaptation to new environmental conditions can be performed easily. As soon as encapsulation is done, reconfiguration requires putting a resource on a bag (send the reconfiguration command) and generating the rules according to the new scenario.

This paper has described a case study presenting the reconfiguration of a building automation system based on the LON technology. The case study is a room that can be split in two or merged in one with a removable wall. This reconfiguration concerns a significant set of equipments: temperature, luminosity and presence sensors, an HVAC system, dimmable spotlights, motorised Venetian blinds, and push buttons. In LON, such reconfiguration is normally done manually by a skilled technician. With our approach, a dynamic reconfiguration can be triggered by an external device that determines the position of the removable wall. This mechanism provides a fully automatic system, where the only action required from the user is to remove/put in place the wall. Here, the middleware is only used to reconfigure the LON controller. After reconfiguration, the system continues to run autonomously.

The approach presented in this paper opens the way to new trends in building automation at a larger scale. Indeed, based on a middleware such as LINC, it will be possible to provide such automatic reconfiguration across several buildings. Future work will focus on automatic reconfiguration through several buildings, using different BAS.

ACKNOWLEDGEMENT

This work has been partially funded by the FP7 SCUBA project under grant nb 288079 and Artemis ARROWHEAD project under grant agreement number 332987.

REFERENCES

[1] A. Saeed and T. Waheed, "An extensive survey of context-aware middleware architectures," in *Electro/Information Technology (EIT)*, 2010 IEEE International Conference on, May 2010, pp. 1–6.

[2] M. Sarnovský, P. Kostelník, P. Butka, J. Hřeňo, and D. Lacková, "First demonstrator of hydra middleware architecture for building automation," in *Proceedings of the scientific conference Znalosti 2008*, 2008.

[3] C. Aghemo et al., "Management and monitoring of public buildings through ICT based systems: Control rules for energy saving with lighting and HVAC services," *Frontiers of Architectural Research*, 2013, pp. 147 – 161.

[4] M. Louvel and F. Pacull, "Linc: A compact yet powerful coordination environment," in *Coordination Models and Languages*, ser. Lecture Notes in Computer Science, 2014, pp. 83–98.

[5] L.-F. Ducreux, C. Guyon-Gardeux, S. Leseq, F. Pacull, and S. R. Thior, "Resource-based middleware in the context of heterogeneous building automation systems," in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. Montreal, Canada: IEEE, 2012, pp. 4847–4852.

[6] U. Rysssel, H. Dibowski, H. Frank, and K. Kabitzsch, *IEHHandbook LONWorks*. Berlin: Vde-Verlag, 2010.

[7] W. Kastner, G. Neugschwandtner, S. Soucek, and H. Newmann, "Communication Systems for Building Automation and Control," *Proceedings of the IEEE*, June 2005, pp. 1178–1203.

[8] S. Krakowiak, "Middleware architecture with patterns and frameworks," 2007.

[9] W. Guo and M. Zhou, "An emerging technology for improved building automation control," in *Systems, Man and Cybernetics*, 2009. SMC 2009. IEEE International Conference on. IEEE, June 2009, pp. 337–342.

[10] V. C. Gungor and G. P. Hancke, "Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches," *IEEE Transactions on Industrial Electronics*, 2009, pp. 4258–4265.

[11] O. F. et al., "Integrating building automation systems and wireless sensor networks," in *Emerging Technologies and Factory Automation*, 2007. ETFA. IEEE Conference on, Sept 2007, pp. 1376–1379.

[12] M. Jung, C. Reinisch, and W. Kastner, "Integrating building automation systems and ipv6 in the internet of things," in *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, July 2012, pp. 683–688.

[13] S. Marcel et al., "Proactive reconfiguration of wireless sensor networks," in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '11, 2011, pp. 31–40.

[14] M. Szczodrak, O. Gnawali, and L. P. Carloni, "Dynamic reconfiguration of wireless sensor networks to support heterogeneous applications," in *Proc. of IEEE DCOSS Conf.*, may 2013, pp. 51–61.

[15] P. Grace, G. Coulson, G. S. Blair, B. Porter, and D. Hughes, "Dynamic reconfiguration in sensor middleware," in *MidSens*, ser. ACM International Conference Proceeding Series. ACM, 2006, pp. 1–6.

[16] K. E. Kjær, "A survey of context-aware middleware," in *Proceedings of the 25th Conference on IASTED International Multi-Conference: Software Engineering*, 2007, pp. 148–155.

[17] P. Istoan, G. Nain, G. Perrouin, and J.-M. Jezequel, "Dynamic software product lines for service-based systems," in *Proceedings of the 2009 Ninth IEEE International Conference on Computer and Information Technology - Volume 02*, ser. CIT '09, 2009, pp. 193–198.

[18] N. Carriero and D. Gelernter, "Linda in context," *Commun. ACM*, vol. 32, 1989, pp. 444–458.

[19] T. Cooper and N. Wogrin, *Rule-based Programming with OPS5*. San Francisco: Morgan Kaufmann, 1988, vol. 988.

[20] P. A. Bernstein, V. Hadzilacos, and N. Goodman, *Concurrency control and recovery in database systems*. New York: Addison-wesley, 1987, vol. 370.

[21] P. François et al., "Self-organisation for building automation systems: Middleware linc as an integration tool," in *IECON 2013-39th Annual Conference on IEEE Industrial Electronics Society*. Vienna, Austria: IEEE, 2013, pp. 7726–7732.

[22] E. Corporation, *LNS TM for Windows® Programmer's Guide*. USA: Echelon Corporation, 1996-2000.

[23] S. F. Project, 2011-2014, <http://www.aws.cit.ie/scuba/>.

An Integrated Ambient Intelligence System in the Monitoring and Rehabilitation of the Disorder of Consciousness.

Francesco Riganello, Luigi Flotta, Giuliano Dolce

Institute S. Anna - RAN (Research in Advanced Neuro-rehabilitation),
Crotone, Italy;
f.riganello@istitutosantanna.it;
giulianodolce@libero.it

Luigi Piscitelli, Calogero Pace

Department of Informatics and Modeling Electronics and Systemistics, University of Calabria,
Cosenza, Italy
lui.piscitelli@gmail.com;
calogero.pace@unical.it

Walter G. Sannita

Department of Neuroscience, ophthalmology and genetics, University of Genova, Genova, Italy/Department of Psychiatry, State University of New York, Stony Brook, NY, USA.
wgs@dism.unige.it

Abstract—Ambient Intelligence (AmI) collectively refers to a family of sensitive electronic systems responsive to humans and mediating in the human interaction with devices and environment. This novel paradigm of information technology complies with the international standards for the functional integration of biomedical domotics and informatics in hospital and homecare. We have designed and implemented an AmI system with sensor networks for the continuous automatic monitoring of subjects with severe brain damage and disorder of consciousness hospitalized in the S. Anna-RAN Institute for medical care and rehabilitation. The system was designed to allow real-time analyses of relevant environmental parameters and the subjects' vital signs. Main purposes are to identify: 1- partially preserved or recovered circadian/ultradian rhythms; 2) functional changes potentially associated to prognostic indicators; 3) momentary subject/environment interactions or functional changes possibly indicative of residual/recovered responsiveness; 4) predictive models of responsiveness. The system also supports the clinician in decision making. In this respect, AmI should be regarded as equivalent to a traditional laboratory for data collection and processing, with substantially reduced dedicated equipment and staff and limited costs. Moreover, the AmI should provide an accurate system of observation of the patient-ambient interaction, offering a better support to the clinical decision in the rehabilitation phase.

Keywords-Monitoring Systems; Decision Making; Disorders of Consciousness; Ambient Intelligence; Sensory Device .

I. INTRODUCTION

The label Ambient Intelligence (AmI) collectively indicates a family of electronic systems that are sensitive and responsive to humans and mediate in the human interaction with devices and environment. It provides pervasive but unobtrusive sensing and computing devices and ubiquitous networking for human/environment interaction [1]. This novel paradigm of information technology complies with the international Integrating Healthcare Enterprise board (IHE) [2] and eHealth HL7 technological standards [3][4] for the

functional integration of biomedical domotics and informatics in hospital and home care. The European Commission charted a path for research on AmI in 2001 [5]. The use of biomedical robotics and informatics in hospitals is also involved in the processes for medical decision making - a complex aspect of the patient's care whereby a high level of cognitive processing is required to manipulate large datasets of disparate kind, origin and significance [6]–[8]. Subjects with disorder of consciousness due to severe brain injury who are in the vegetative (VS/UWS) or minimally conscious (MCS) states [9]–[12] need constant monitoring; a continuous stream of clinical/neurobiological/ behavioral information is required for appropriate care by the medical and nursing staff and to optimize the rehabilitation therapy [13]. A dedicated AmI system has been designed and developed for this purposes, to provide an accurate system of observation of the patient-ambient interaction and to have a better support to the clinical decision in the rehabilitation phase. In section II of the paper is described the method of data acquisition; the section III is dedicated to the biometric and environmental nodes, while the section IV is dedicated to the data analysis; the section V contains a brief description of the Heart Rate Variability and the importance of its analysis in the rehabilitation therapy; finally the section VI reports the conclusions.

II. DATA ACQUISITION

Daily medical and nursing care and the rehabilitative protocols in hospital require multi- and cross-disciplinary activities. Data acquisition and analyses are core facilities in the rehabilitation process, with both scientific and applicative implications; computational environments capable of integrating heterogeneous representations from different fields are also mandatory in this context. Such environments should integrate heterogeneous formalisms in the same model and assist the modeller in designing and implementing

new models based on new knowledge acquired in the rehabilitation environment.

We have designed and implemented an AmI system with *ad hoc* sensor networks for the continuous monitoring of subjects in VS/UWS or MCS hospitalized in the S. Anna - RAN Institute for medical care and rehabilitation. The system can be implemented to allow exporting the facility to homecare with control in remote. The acquisition and integration procedures are implemented to record, store and process both biometric and environmental data from dedicated nodes. The monitoring system is designed to allow real time analyses of the subject's vital signs and of the environmental parameters in which he/she lives (Figures 1 and 2). Samples of recorded data are shown in figure 3.

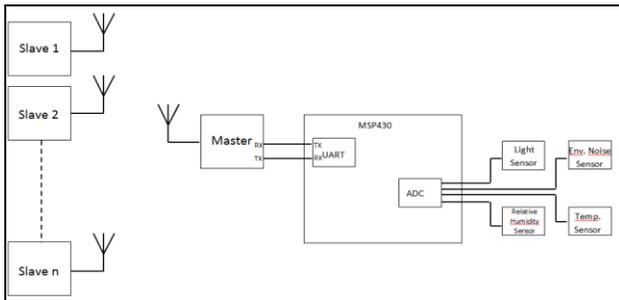


Figure 1: Block diagram of the monitoring system

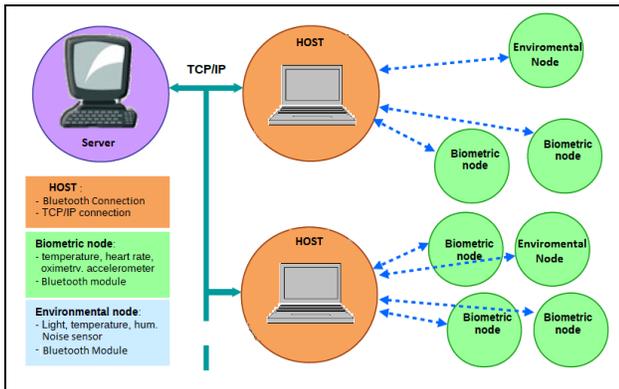


Figure 2: System overall architecture.

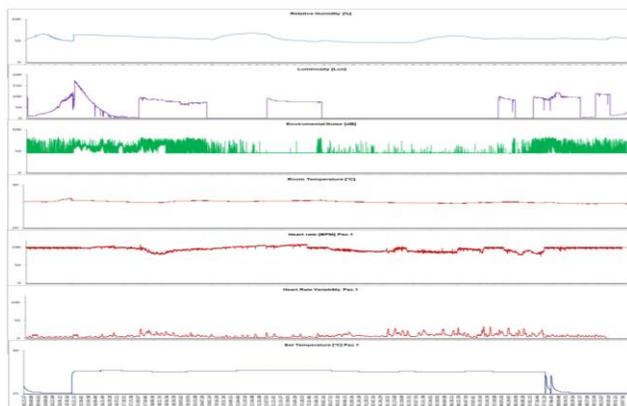


Figure 3: Sample plot of data from the patient and ambient parameters. From top to bottom: humidity, luminosity, noise, room temperature, HR, HRV and Patient temperature.

III. BIOMETRIC AND ENVIRONMENTAL NODES

Data transmission from the subject to the biometric node is realized via Bluetooth by mean of Zephyr Bioharness 3 devices [14]–[16] positioned on a chest strap. Electrodes are built to monitor heart rate, heart rate variability, respiration rate, temperature, movement and oxygen saturation. Specifically:

- Two electrodes for heart rate and heart rate variability;
- A sensor on the strap right side for respiration;
- A triaxial accelerometer incorporated in the device to detect the posture angle and activity indexes;
- An infrared temperature sensor to measure skin temperature.
- A photoplethysmographic sensor to measure PO2.

Four sensors detect environmental parameters:

- Temperature (Texas Instruments LM 35);
- Relative Humidity (Honeywell HIH 3605);
- Noise (Omnidirectional condenser electret microphone);
- Light intensity (Light Dependent Resistor).

The Texas Instrument LM35 is a high-precision integrated sensor for detecting ambient temperature. It has a linearly proportional output voltage-temperature characteristic, with scale factor: +10 mV/°C and a range of measurements from -55°C to 150 °C, with ± 0,5°C accuracy at 25 °C in addition to the low current draw. The system core is based on a microcontroller. The Honeywell HIH 3605 Environmental Relative Humidity sensor is an integrated sensor with linear voltage-relative output, low current draw, fast response in time and high accuracy. It works, with a 5 V voltage supply in the 0-100 % range; accuracy is ± 2%. The datasheet provides formulas to calculate the relative humidity and compensate for temperature in respect to the relative humidity. The device has three terminals. Light intensity is detected by mean of a light sensor using a light dependency resistor (LDR) that through its conditioning circuit returns a voltage signal that is converted into light intensity. It is a passive sensor relay for the variation in a semiconductor electric resistance due to the incidence of electromagnetic radiation with wavelength from 1mm to 10 nm. The LDR used in the prototype has a range of operating temperature between -30°C and 70°C. At dark, it offers a resistance of 2 MΩ and response time of 30 ms. The characteristic of the spectrum response is similar at human eye, with peak value at 560 nm. The Condenser Electret Microphone is an omnidirectional microphone to detect environmental noise. Its operative sensitivity of -38 dB ± 3 dB at f = 1 kHz, with operating 2-10V range offers a signal-to-noise ratio of 58 dB and a frequency spectrum similar to that of the human ear (20 ÷ 20000 Hz). The MSP430 family [17][18] is a series of mixed-signal processors for ultra-low power signals, with 16-bit Von Neumann architecture, RISC-based. In our project, the Texas Instruments MSP430F2274 was used to handle, compute and make a first filtering of the signals; low power microcontrollers will allow implement a battery-powered device. Each sensor requires a conditioning

circuit making the output signals from each sensor suitable for transmission, display or recording meeting the requirements of the device following in the line. A conditioning circuit comprises electronic circuits performing each of the following functions: amplification, level shifting, filtering, impedance matching, modulation, demodulation. The accuracy and reliability of environmental sensors and measures have been tested in the Nexus Laboratories of the Department of electronic engineering, University of Calabria, Italy.

Msp430 used in this prototype module starts through its UART (Universal Asynchronous Receiver Transmitter)[19] a serial communication at 9600 baud/sec in configuration 8N1 (8 data bit, no parity bit, 1 stop bit). The microcontroller pins for signal transmission and reception are linked to those of the Bluetooth module, which serves as the master interacting with the slave sensor node networks (environmental and biometric). The bit sent contains digital information of parameters monitored from sensors. An analogic/digital converter is integrated in microcontroller for data transmission; it is a SAR [17], 10 bit resolution, with internal reference, set at 1,5 V for greater signal accuracy.

A. Architecture of the system

The structure of the system (server) (Figures 2 and 3) has been designed to collect data from smaller structures (nodes and hosts) distributed in the hospital ambient suitable of, and requiring, monitoring. Specifically, each ambient will be equipped by a host (personal computer) collecting data to be sent to the server and a series of biometric sensors worn by the patients, that will send the biometric data to the host. Host collects data information from sensor nodes of network by Bluetooth communication. Servers will collect all data from the network by TCP/IP protocol communication.

B. Application Software

A Labview National Instruments environment [20] is installed in the host to communicate with biometric and environmental devices and to acquire physiological and environmental information useful for monitoring. The Labview application front panel (Figure 4) serves as a graphic user interface for the user to display (in graphs, waves or numbers) any measured physical quantity.



Figure 4: Front panel Labview application

The block Visual Instrument Software Architecture (VISA) [21], that allow the control of data stream of serial port necessary to control in a specific time, is used for handling serial communication by means of the G language of Labview. The application communicates with the two sensor nodes of networks by two similar Bluetooth protocols for the procedures of communication and data transfer to share the same core characteristics. The information collected by Labview is stored in real time in a .txt file. The weight of .txt files of the environmental parameters is 2160 bytes/min, while the biometric parameters .txt file weight is composed by a linear coefficient characterized by mean weight of one parameter (4 bytes) multiplied for the number of parameters chosen and the time of recording calculated in seconds. The reliability of acquired information (status info) and battery status are shown in Table I. In a chosen example of example 4 biometric parameters, the weight of the .txt file would be 11.84 Mb/day.

TABLE I: DAILY WEIGHT STORAGE

Info	Weight(byte)s	Weight(byte)m	Weight(byte)h	Weight(byte)day
Date	10	600	3600	0.864
Time	8	480	28800	0.691
Environment Parameters	9x4	2160	129600	3.11
Biometric Parameters	4xn	240xn	14400xn	0.346xn
Battery Info	4	40	1400	0.346
Formatt.Byte	13	780	46800	1.123

C. Information about the rehabilitation process

These data are obtained through clinical observations ranked according to established rating scales filled in by the attending physician and nursing staff. The scales in use are:

- The Coma Recovery Scale Revisited – CRS [22];
- The Wessex Head Injury Matrix – WHIM [23];
- The Levels of Cognitive Functions – LCF [24];
- The Nociception Coma Scale – NCF [25].

Dedicated web applications have been developed to enable medical and nursing staff to fill in each scale item with the appropriate values. Data are stored in the central database and integrated with the information for further processing and correlative studies.

IV. DATA ANALYSIS

The traditional biomedical model treats disease as a binary variable, whereby a subject is or is not ill. However, most chronic disorders are gradual processes and cannot be classified as binary, while the threshold determining severity can be ambiguously set (the evolution from VS/UWS to MCS is an example in this regard) [11][26]. Many biological variables, such as cholesterol, blood pressure, and blood glucose are normally distributed in the population and a convenient and reliable cutoff can be determined to separate disease from health. In the definition of the disease and its severity and prognosis, judgment needs to be based on knowledge and experience for a correct interpretation of clinical and laboratory information [27]. In our project, measurements and analyses of datasets with information

from environmental and biometric sensors follow a straightforward strategy, with major focus on variability, either spontaneous or in response to environmental changes [28]. Sources of variation are manifold and can reflect changes in the functional status or in responsiveness as well as the existence of residual endogenous mechanisms of self-regulation or circadian/ultradian cycles. Functional changes in VW/UWS or MCS subjects can also depend on interaction with the staff or relatives, result from the nursing or rehabilitation procedures, or reflect endogenous mechanisms [29].

The main lines of research are: 1) identification of correlations between sets of stimulus-conditions administered to subjects with disorder of consciousness and changes in the recorded variables [30]–[32], and 2) analysis of the two-way interaction between the cardiovascular system, the central nervous system and the central autonomic network model [33]–[36]. The goals in these lines of investigation are: 1) to verify whether circadian/ultradian rhythms are partially preserved or have recovered to a significant extent in individual VS/UWS or MCS subjects; 2) to identify functional processes potentially associated to prognostic indicators; 3) to detect momentary interactions between the subject and the environment or other functional changes possibly indicative of residual/recovered responsiveness; 4) to develop predictive models of responsiveness in patients with disorder of consciousness; 5) support the clinician in decision making.

The Decision Support System implemented to help the attending physician is based on *R*, an Open Source statistical package expanded with custom modules [37]; traditional and advanced techniques of statistical analysis are also used. Among these are: the regression analysis (to identify relationships among variables), the Neural network (a sophisticated pattern detection algorithm using machine learning techniques to generate predictions), the Clustering/Segmentation processes (to create groups for applications), the Association Rules techniques (to detect related items in a dataset), Bayesian statistics, Data mining, Neural network, etc.) [38][39].

V. HEART RATE VARIABILITY

Systematic investigation on the heart rate variability (HRV) (i.e. the heart rate fluctuations around the mean value over the time sample) is a major scientific and applicative approach to the functional understanding of VS/UWS and MCS at the Institute S. Anna - RAN [30][35][40][41]. HRV reflects in time the momentary function of the cardio-respiratory control system and is regarded as a reliable index of the sympathetic/parasympathetic functional interplay [34] and intrinsic influence on heart rate. HRV is also thought to provide independent information on the autonomic nervous system and its two-way functional integration with the central nervous system, to express physiological factors modulating the heart rhythm and homeostatic adaptation to the changing conditions. It is anatomically and functionally described by the central autonomic (nervous) network (CAN) model [42][43]. The autonomic system influences

heart rate adaptation through multiple connections (inputs from sensory and baroreceptors within the heart and great vessels, respiratory changes, vasomotor regulation, thermoregulatory system and changes in endocrine function and neuroendocrine interaction) [34][44]. HRV indirectly reflects the organization of affective, physiological, cognitive, and behavioural elements and is emerging as a possible descriptor of the brain functional organizations contributing to homeostasis and homeostatic responses [35][45][46]. Research on the possible patterns of correlation between HRV measures and the functional models available today is crucial in the diagnosis and prognosis of disorder of consciousness and in the online monitoring of patients with severe brain injury.

VI. CONCLUSIONS

The medical care, daily management and rehabilitation of subjects with severe brain damage and disorder of consciousness require inter- and cross-disciplinary activities. This is the basic approach in research on consciousness and rehabilitation. The extensive monitoring of all aspects of the subject's functional condition and reactivity in his/her daily activity is a prerequisite to understand the real efficacy of ambient conditions/changes potentially contributing to the recovery of the consciousness in the *rehabilitative milieu*. In this context, we need a reliable computational environment capable of integrating the heterogeneous representations from different rehabilitative and scientific approaches into heterogeneous formalisms in a useful model. To this end, our AmI system has been contrived to contribute in the efficient representation of the rehabilitative environment, by studying and correlating biometric parameters with environmental variables. The current development in electronic systems and pervasive intelligent devices and computing in the surrounding environment is the new trend and a new frontier of research in the rehabilitative science. New scenarios of application are supported by the ongoing miniaturization of electronic circuits and increasing computational power. New devices can be implemented at low cost to help researchers and clinicians.

A major advantage of our AmI system is the integration of artificial intelligence technology [6][28][47]–[49] with traditional or advanced data acquisition systems such as those in use in the monitoring of clinical or functional parameters of inpatients or of subjects under remote medical control. In this respect, our AmI platform complies with the standards of the international IHE board and the eHealth HL7 format as hedge technological approach in the eHealth functional integration of biomedical and traditional domestic equipments and informatics in hospital and home care. Our AmI system has been designed and implemented to monitor and help treat and rehabilitate subjects with disorder of consciousness. It is a multipurpose hw/sw tool suitable of extensive application in patients' monitoring as well as in medicine and neuroscience when large biomedical datasets are acquired and measures of spontaneous or condition-dependent variability are needed. In this respect, AmI should

be regarded as equivalent to a traditional laboratory for data collection and processing, with substantially reduced dedicated equipment and staff, and limited costs, providing an accurate system of observation of the patient-ambient interaction and a better support to the clinical decision in the rehabilitation phase.

REFERENCES

[1] E. Aarts, H. hawig, and S. Schuurmans, "Ambient Intelligence," in *The Invisible Future*, Denning J., McGraw Hill, 2001.

[2] K. Anyanwu, A. P. Sheth, J. Cardoso, J. A. Miller, and K. J. Kochut, "Healthcare enterprise process development and integration," 2003.

[3] R. Gajanayake, R. Iannella, and T. Sahama, "Sharing with Care: An Information Accountability Perspective," *IEEE Internet Comput.*, vol. 15, no. 4, pp. 31–38, Jul. 2011.

[4] D. M. López and B. Blobel, "Architectural Approaches for HL7-based Health Information Systems Implementation:," *Methods Inf. Med.*, vol. 49, no. 2, pp. 196–204, Mar. 2010.

[5] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J.-C. Burgelman, *Scenarios for ambient intelligence in 2010*. Office for official publications of the European Communities, 2001.

[6] B. W. Pickering, J. M. Litell, and O. Gajic, "Ambient Intelligence in the Intensive Care Unit: Designing the Electronic Medical Record of the Future," in *Annual Update in Intensive Care and Emergency Medicine 2011*, P. J.-L. Vincent, Ed. Springer Berlin Heidelberg, 2011, pp. 793–802.

[7] B. W. Pickering, O. Gajic, A. Ahmed, V. Herasevich, and M. T. Keegan, "Data Utilization for Medical Decision Making at the Time of Patient Admission to ICU*," *Crit. Care Med.*, vol. 41, no. 6, pp. 1502–1510, Jun. 2013.

[8] W. W. Stead, J. R. Searle, H. E. Fessler, J. W. Smith, and E. H. Shortliffe, "Biomedical Informatics: Changing What Physicians Need to Know and How They Learn," *Acad. Med.*, vol. 86, no. 4, pp. 429–434, Apr. 2011.

[9] G. G. Celesia and W. G. Sannita, "Can patients in vegetative state experience pain and have conscious awareness?," *Neurology*, vol. 80, no. 4, pp. 328–329, Jan. 2013.

[10] G. Dolce, W. G. Sannita, and for the European Task Force on the, "The vegetative state: A syndrome seeking revision?," *Brain Inj.*, vol. 24, no. 13–14, pp. 1628–1629, Dec. 2010.

[11] S. Laureys, G. G. Celesia, F. Cohadon, J. Lavrijsen, J. León-Carrión, W. G. Sannita, L. Sazbon, E. Schmutzhard, K. R. von Wild, A. Zeman, G. Dolce, and \$author firstName \$author.lastName, "Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome," *BMC Med.*, vol. 8, no. 1, p. 68, Nov. 2010.

[12] The Multi-Society Task Force on PVS., "Medical aspects of the persistent vegetative state.," *N Engl J Med*, vol. 330, pp. 1499–508, 1572, 1994.

[13] L. Flotta, F. Riganello, and W. G. Sannita, "Intelligent Monitoring of Subjects with Severe Disorder of Consciousness," in *SENSORDEVICES 2013*, The Fourth International Conference on Sensor Device Technologies and Applications, 2013, pp. 135–138.

[14] J. A. Johnstone, P. A. Ford, G. Hughes, T. Watson, and A. T. Garrett, "Bioharness™ Multivariable Monitoring Device: Part I: Validity," *J. Sports Sci. Med.*, vol. 11, no. 3, pp. 400–408, Sep. 2012.

[15] J. A. Johnstone, P. A. Ford, G. Hughes, T. Watson, and A. T. Garrett, "Bioharness™ Multivariable Monitoring Device:

Part. II: Reliability," *J. Sports Sci. Med.*, vol. 11, no. 3, pp. 409–417, Sep. 2012.

[16] E. Jovanov, D. Raskovic, A. O. Lords, P. Cox, R. Adhami, and F. Andrasik, "Synchronized physiological monitoring using a distributed wireless intelligent sensor system," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003, 2003, vol. 2, pp. 1368–1371 Vol.2.

[17] J. H. Davies, *MSP430 Microcontroller Basics*. Elsevier, 2008.

[18] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Information Processing in Sensor Networks*, 2005. IPSN 2005. Fourth International Symposium on, 2005, pp. 364–369.

[19] Thakare, "A Review on Implementation of Serial Communication by Universal Asynchronous Receiver Transmitter," *Int. J. Res. Comput. Eng. Electron.*, vol. 2, no. 1, 2013.

[20] R. Bitter, T. Mohiuddin, and M. Nawrocki, *LabView: Advanced Programming Techniques*, Second Edition. CRC Press, 2006.

[21] H. Zhao and Z. Y. Dong, "Design and Implementation of USB-based Microwave Power Sensor," *Appl. Mech. Mater.*, vol. 347, pp. 1039–1042, 2013.

[22] J. T. Giacino, K. Kalmar, and J. Whyte, "The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility," *Arch. Phys. Med. Rehabil.*, vol. 85, no. 12, pp. 2020–2029, Dec. 2004.

[23] A. Shiel, S. A. Horn, B. A. Wilson, M. J. Watson, M. J. Campbell, and D. L. Mclellan, "The Wessex Head Injury Matrix (WHIM) main scale: a preliminary report on a scale to assess and monitor patient recovery after severe head injury," *Clin. Rehabil.*, vol. 14, no. 4, pp. 408–416, Jan. 2000.

[24] C. Hagen, D. Malkmus, and P. Durham, "Rancho Los Amigos levels of cognitive functioning scale," *Downey CA Prof. Staff Assoc.*, 1972.

[25] C. Schnakers, C. Chatelle, A. Vanhaudenhuyse, S. Majerus, D. Ledoux, M. Boly, M.-A. Bruno, P. Boveroux, A. Demertzi, G. Moonen, and S. Laureys, "The Nociception Coma Scale: a new tool to assess nociception in disorders of consciousness," *Pain*, vol. 148, no. 2, pp. 215–219, Feb. 2010.

[26] A. Bosco, G. E. Lancioni, M. O. Belardinelli, N. N. Singh, M. F. O'Reilly, and J. Sigafos, "Vegetative state: efforts to curb misdiagnosis," *Cogn. Process.*, vol. 11, no. 1, pp. 87–90, Feb. 2010.

[27] R. M. Kaplan and D. L. Frosch, "Decision Making in Medicine and Health Care," *Annu. Rev. Clin. Psychol.*, vol. 1, no. 1, pp. 525–556, Apr. 2005.

[28] L. Pignolo, F. Riganello, G. Dolce, and W. G. Sannita, "Ambient intelligence for monitoring and research in clinical neurophysiology and medicine: the MIMERICA* project and prototype," *Clin. EEG Neurosci.*, vol. 44, no. 2, pp. 144–149, Apr. 2013.

[29] F. Riganello, G. Dolce, M. D. Cortese, and W. G. Sannita, "Responsiveness and prognosis in the severe disorder of consciousness," *Brain Damage Causes Manag. Progn. Schäffer AJ Müller J Eds Pp*, pp. 117–135, 2010.

[30] A. Candelieri, M. D. Cortese, G. Dolce, F. Riganello, and W. G. Sannita, "Visual Pursuit: Within-Day Variability in the Severe Disorder of Consciousness," *J. Neurotrauma*, vol. 28, no. 10, pp. 2013–2017, 2011.

[31] F. Riganello, M. D. Cortese, G. Dolce, and W. G. Sannita, "Visual pursuit response in the severe disorder of consciousness: modulation by the central autonomic system and a predictive model," *BMC Neurol.*, vol. 13, no. 1, p. 164, 2013.

[32] P. Urbenjaphol, C. Jitpanya, and S. Khaoropthum, "Effects of the sensory stimulation program on recovery in

- unconscious patients with traumatic brain injury,” *J. Neurosci. Nurs.*, vol. 41, no. 3, pp. E10–E16, 2009.
- [33] V. Napadow, R. Dhond, G. Conti, N. Makris, E. N. Brown, and R. Barbieri, “Brain correlates of autonomic modulation: Combining heart rate variability with fMRI,” *NeuroImage*, vol. 42, no. 1, pp. 169–177, 2008.
- [34] F. Riganello, G. Dolce, and W. Sannita, “Heart rate variability and the central autonomic network in the severe disorder of consciousness,” *J. Rehabil. Med.*, vol. 44, no. 6, pp. 495–501, 2012.
- [35] F. Riganello, S. Garbarino, and W. G. Sannita, “Heart Rate Variability, Homeostasis, and Brain Function: A Tutorial and Review of Application,” *J. Psychophysiol.*, vol. 26, no. 4, pp. 178–203, 2012.
- [36] R. Lane, K. Mcrae, E. Reiman, K. Chen, G. Ahern, and J. Thayer, “Neural correlates of heart rate variability during emotion,” *NeuroImage*, vol. 44, no. 1, pp. 213–222, Jan. 2009.
- [37] J. Marques de Sá, “Directional Data,” *Appl. Stat. Using SPSS Stat. MATLAB R*, pp. 375–401, 2007.
- [38] A. Candelieri, G. Dolce, F. Riganello, and W. G. “Data Mining in Neurology,” in *Knowledge-Oriented Applications in Data Mining*, K. Funatsu, Ed. InTech, 2011.
- [39] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Third Edition, 3 edition. Burlington, MA: Morgan Kaufmann, 2011.
- [40] G. Dolce, F. Riganello, M. Quintieri, A. Candelieri, and D. Conforti, “Personal interaction in the vegetative state: A data-mining study,” *J. Psychophysiol.*, vol. 22, no. 3, pp. 150–156, 2008.
- [41] F. Riganello, A. Candelieri, M. Quintieri, D. Conforti, and G. Dolce, “Heart rate variability: An index of brain processing in vegetative state? An artificial intelligence, data mining study,” *Clin. Neurophysiol.*, vol. 121, no. 12, pp. 2024–2034, Dec. 2010.
- [42] E. E. Benarroch, “The central autonomic network: functional organization, dysfunction, and perspective,” *Mayo Clin. Proc.*, vol. 68, no. 10, pp. 988–1001, 1993.
- [43] E. Benarroch, W. Singer, and M. Mauermann, *Autonomic Neurology*. Oxford University Press, 2014.
- [44] E. E. Benarroch, “The Autonomic Nervous System: Basic Anatomy And Physiology” *Contin. Lifelong Learn. Neurol.*, vol. 13, pp. 13–32, 2007.
- [45] E. E. Benarroch, “Pain-autonomic interactions,” *Neurol. Sci.*, vol. 27, no. 2, pp. s130–s133, May 2006.
- [46] B. H. Friedman and J. F. Thayer, “Autonomic balance revisited: panic anxiety and heart rate variability,” *J. Psychosom. Res.*, vol. 44, no. 1, pp. 133–151, 1998.
- [47] C. Ramos, J. C. Augusto, and D. Shapiro, “Ambient intelligence—the next step for artificial intelligence,” *Intell. Syst. IEEE*, vol. 23, no. 2, pp. 15–18, 2008.
- [48] G. Riva, “Ambient intelligence in health care,” *Cyberpsychol. Behav.*, vol. 6, no. 3, pp. 295–300, 2003.
- [49] W. Weber, C. Braun, R. Glaser, Y. Gsottberger, M. Halik, S. Jung, H. Klauk, C. Lauterbach, G. Schmid, X. Shi, T. F. Sturm, G. Stromberg, and U. Zschieschang, “Ambient intelligence - key technologies in the information age,” in *Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International*, 2003, pp. 1.1.1–1.1.8.

Accurate Sled Velocity on a Short-Inclined Track Using Accelerometer Data

Part of an Automated Bob-Skeleton Push-Start Performance Monitoring System

Mark Gaffney, Dr. Michael Walsh, Brendan O’Flynn and Dr. Cian Ó Mathúna
 Tyndall National Institute, University College Cork,
 Cork, Ireland.
 e-mail: mark.gaffney@tyndall.ie

Abstract—Wireless Inertial Measurement Units (WIMUs) are increasingly used to gather data and improve understanding of various human performance and complex motion scenarios. The Bob-Skeleton Push-Start features a stooped sprint from a crouch while pushing a heavy sled. Maximizing velocity during this brief period is considered crucial to performance, however it is poorly understood. An adjustable sled Push-Start training tool was instrumented with custom WIMUs, and a test subject performed 36 runs, with 12 combinations of 3 Incline and 4 Weight settings. A developed algorithm automatically identified, extracted, and integrated Pushing-Phase Acceleration data to Velocity and Displacement at hundreds of samples per second. Drift correction methods improved accuracy; while additional checks rejected problematic data-files. WIMU derived Average Velocities were within -0.005 ± 0.074 meters per second ($0.319 \pm 4.214\%$) of an existing Light-Gate system. Such an accurate, automatic, WIMU-based system could supplement or replace Light-Gate or other performance monitoring methods, while being more portable and readily usable by coaches or athletes. This would enable consistent, low-cost and high-fidelity, performance monitoring from the gym to the ice-track for improved candidate selection, comparison and training in Bob-Skeleton and other ice-track sled sports.

Keywords; WIMU, Accelerometer, Bob-Skeleton, Sled, Error Correction, Performance Monitoring.

I. INTRODUCTION

Bob-Skeleton is an ice-track sledding sport – similar to Bob-sleigh, Luge and Toboggan – with a single athlete riding an open sled in a face-forward, prone manner. Each run begins with the Push-Start (PS), which requires the athlete to sprint from stationary, in a crouched position, accelerating to maximum velocity, over a short distance (~30-45m), while pushing a heavy sled (~30-40kg), before transitioning to riding the sled through a series of turns for the remainder of the up to 1.5km long track.

The sport is highly competitive, with the top times over the roughly 90 second run duration often within a fraction of a second of each other. While low PS time is generally believed to be the most crucial aspect of final race time [1], [2], this motion is poorly understood. A combination of the sport’s small size and difficulty in accessing ice-tracks, as well as the lack of available data are likely responsible for the lack of detailed PS studies. Relevant publications often rely on problematic data sources, such as: official timing

(which ignores the first 15 meters) [1], [3-5]; alternative single interval timing (which hampers direct comparison, or understanding of subtle changes) [6], [7]; or use complex and costly data gathering systems (preventing more widespread use) [8-10].

As such, we set out to develop an easy-to-use, portable system that can provide high-quality sled velocity data. Ideally “On-Ice” performance would be investigated, however “Dry-Land” methods are more likely to be used for selection, comparison and training of Bob-Skeleton athletes [11-13] – especially for new athletes or in countries without a well-established amateur Bob-Skeleton system or easy access to ice-tracks – so initial system development and data gathering used such facilities at the University of Bath.

The “Assassin – Horizontal Power Trainer” is a PS training tool. It consists of a sled which runs along a pair of parallel straight rails, allowing a 3 meter free travel length before impacting attached buffers, with a Light-Gate pair covering the majority of this; additional weights and adjustable track incline can be used to change pushing resistance (See Fig. 1 and 2).

Wireless Inertial Measurement Units (WIMUs) are small electronic devices containing sensing elements, similar to those in smart-phones (e.g., Accelerometers, Gyroscopes,

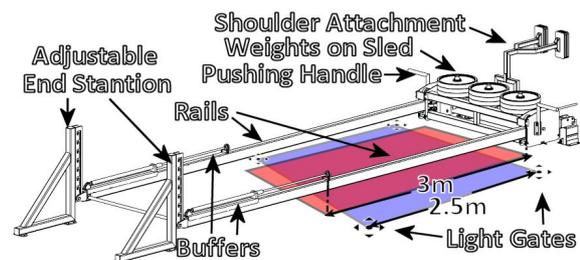


Figure 1. Labelled Diagram of Assassin.

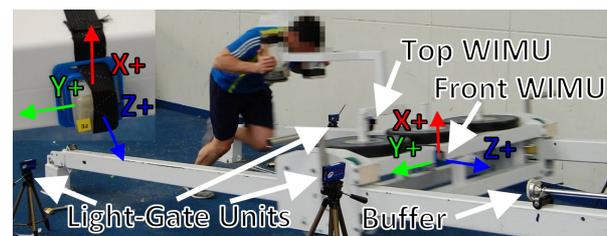


Figure 2. Attached WIMU (inset) and Assassin during Pushing-Phase.

and Magnetometers), along with supporting components, which can act as un-tethered motion sensors. A custom WIMU system was built on the Tyndall 25mm Mote Micro-System platform as part of on-going work on human motion capture for health and sports applications [14-19]; laboratory calibration was performed during assembly [20]. These were attached to the Assassin Sled for PS data gathering.

An automatic process allowing for performance to be accurately quantified using WIMU data recorded from an Assassin run was developed. High sampling rates provided detailed information on how velocity develops from standstill, which could be crucial in determining the subtle effect of changes to training, warm-up and pushing technique and giving a competitive edge.

WIMU derived results were validated by comparing sled Average Velocities to an existing Light-Gate system. An initial target of accuracy within 0.1 meters per second was set as this was considered the threshold for indicating notable differences in performance levels and effectiveness of coaching interventions.

In Section 2, “Data Sources and Method”, the equipment, setup, subject and procedure are described. Section 3, “Analysis”, contains the acceleration features of a run, segmentation and processing of WIMU data, estimating the full run duration, rejecting runs and the adaptive integration process. Section 4, “Results”, explains and provides, individual and combined contour graphs as well as the quantification method used to compare and assess the accuracy of final integrated WIMU data against the Light-Gate and a brief estimation of Light-Gate accuracy. Sections 5 and 6 contain a brief “Discussion” and “Conclusion”.

II. DATA SOURCES AND METHOD

A. WIMUs, Location and Orientation

Several identical WIMUs were configured to provide wireless Accelerometer data at the maximum sample rate and sensor range, of up to 256 Hertz and $\pm 16g$ respectively, reducing the likelihood of saturation and under-sampling of large-magnitude or high-frequency acceleration features. The effective sampling rate varied, being dependant on unpredictable events such as wireless packet loss. Data was streamed via 802.15.4 compatible radio at 2.45GHz to a Base-station connected to a notebook computer. APIs and scripts – written in Python – enabled WIMU configuration as well as sensor data gathering, processing and storage. All sensor data were converted to real world units – using previously gathered on-board laboratory calibration values – and written to file.

Two WIMUs (Front and Top) were placed into 3D-printed holders before being secured to the metal spars of the moveable sled, with similar orientation, using Velcro-elastic straps and tape as shown in Fig. 2. This provided some measure of redundancy and allowed for investigation of the effect of WIMU placement.

B. Other Equipment and Data Sources

A Brower “Timing Centre” Light-Gate system [21] – consisting of 2 emitter and receiver beam sets – positioned to

cover the central 2.5 meter portion of each run (See Fig. 1 and 2). Light-Gate ground separation and rail heights were measured using a surveyor’s tape to the nearest centimeter. Nominal inclination angles and free travel length were taken from technical drawings of the Assassin.

C. Subject

A fit male was used as the test subject, representing a potential Bob-skeleton athlete undergoing selection. He was familiar-with and trained-in the use-of the Assassin, and was part of on-going sports science and performance research programs at the University Of Bath and UK Sports which these tests were a part of. The purpose, procedures and equipment were explained to him and he had opportunity to ask questions or suggest changes to the procedure. He was also allowed to warm-up, take breaks, perform practice runs or stop the testing at his discretion.

D. Procedure

36 test runs were planned, with 3 runs at each combination of 3 nominal rail angles (0, 4 and 7°), and 4 weight settings (0, 20, 40 and 60 Kg). This allowed for adjustment of the effort required of the test subject. The centrally located padded shoulder pushing attachment was used and the buffers were positioned for the maximum sled free travel length of 3m to provide WIMU datasets with the largest number of samples possible. The test procedure was as follows:

1. Sled is at rest at starting point
2. Change Weight and Inclination Settings if needed
3. Reset WIMUs and Light-Gates
4. Test subject proceeds when ready
5. Stop WIMU recording after the end of the run

III. ANALYSIS

Initial manual review of data established appropriate processing strategies and identified consistent events or features of interest as described below and in Fig. 3.

A. Overview of an Assassin Run

- Pre-Push-Off (PPO): Region with sled at rest at the start of the track, mostly quiescent with occasional motion artifacts due to the athlete addressing the sled. Quiescent

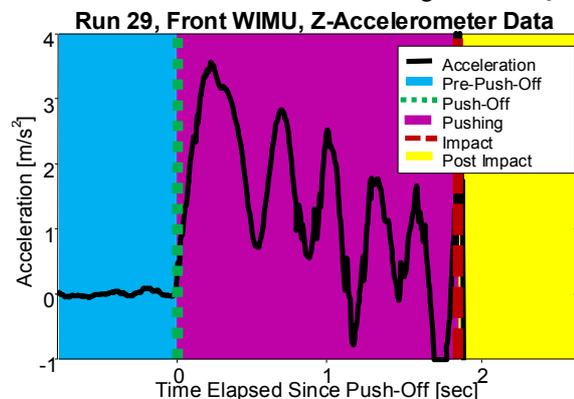


Figure 3. Labelled Main Features of Assassin Run.

sensor level is affected by track inclination.

- Push-off (PO): The start of Pushing-Phase, located at the beginning of a sudden rise in acceleration from PPO quiescent levels.
- Pushing-Phase (PP): Region lasting roughly 2 seconds, with large, cyclical, acceleration features likely due to individual steps.
- Impact Point (IP): A sudden large acceleration feature when the sled contacts the buffers, sensor saturation is common.
- Post Impact (PIP): The remaining data, often beginning with saturated severe oscillations which damp down as the sled comes to a stop, possibly followed by quiescent data and motion artifacts.

B. Real-World Data Processing Considerations

Using the equations of motion, it should be possible to integrate all the recorded Accelerometer samples over time to yield sled Velocity and Displacement. However, the recorded WIMU data is not perfect, due to limited sampling rates and sensor range, with additional errors due to sensor noise and quantization further contributing to this. When such data is integrated it tends to drift further from the actual values as these errors compound, increasing greatly over time. These issues are often encountered with WIMU sensors and mitigation strategies have been developed to compensate. A new methodology applicable in Bob-Skeleton and in the more general case is proposed here involving application specific adjustments and integration limits during processing which leads to significantly improved results.

By performing integration only within regions of interest where sensor data is not saturated – in this case the PP between PO and IP – the potential for drift caused by orientation changes, motion artifacts and saturation is reduced. Knowledge of the track inclination angle or the average value of quiescent sections of the Pre-Push-Off region can be used as a Sensor Offset to improve results.

Using known physical limits as integration constraints can further increase the accuracy of the results. In the case of the system described: Initial Velocity and Displacement values are 0, negative Velocity or Displacement values are not possible and Displacement at impact should equal the sled’s free travel length (3 meters).

Re-estimation of the integration period to account for data loss etc. causing differences between requested and effective sampling rates can also be used – essentially acting as Time Warping – although this requires an estimate of the duration over which a known number of samples were recorded.

While segmentation and identification of the previously described run features and adjustments to improve the accuracy of results could be performed “by-eye” or manually, an automated method is desirable to reduce subjective human variability and enable development of a self-contained high-accuracy performance monitoring system suitable for use by athletes and trainers.

C. Automated Data Analysis Methodology

An automated analysis system was implemented in the Python programming language. It consists of several stages: Pre-Processing; Impact Detection; Run Segmentation; Start Detection; Integration and an additional stage of Evaluation versus the Light-Gate, as described below and illustrated in the flowchart in Fig. 4.

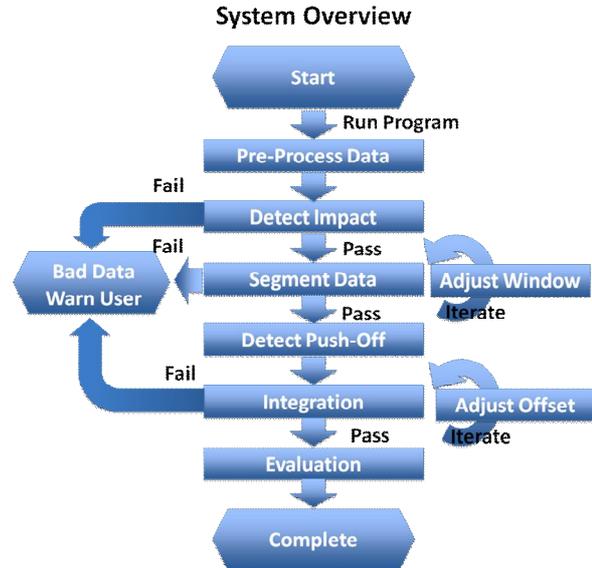


Figure 4. Simplified Flowchart of Assassin Data Analysis Algorithm.

1) Pre-Processing

Data are prepared for subsequent analysis. WIMU sensor data are converted to more convenient units, filtered to remove outliers and smoothed to reduce noise, simplifying subsequent integration, segmentation, and feature detection stages. Data from other sources such as Light-Gate timing, Assassin settings and physical measurements; are used to estimate useful values.

2) Impact Detection

The largest magnitude Acceleration features are identified as Impact Candidates. A threshold is used to check if these are suitably large.

3) Segmentation

Contiguous Active and Passive regions of sensor data are identified using detection thresholds estimated from the most quiescent region of the filtered sensor data. From these, the Active region that contains sufficient data between its start and an Impact Candidate is identified as the PP. If segmentation is unsuccessful, an iterative process attempts to determine a new threshold.

4) Push-Off Detection

The region around the PPO-PP transition is searched for a characteristic peak in the smoothed Acceleration data, the beginning of this feature being the Push-Off.

5) Integration

Initial conditions are set, the offset is applied and integration is performed using the previously decided integration period and standard equations of motion to yield

Sled Velocity and Displacement at each PP Accelerometer sample. Comparing the estimated and known Displacement at IP allows iterative refinement of the Sensor Offset to yield improved integration results.

6) Evaluation

The integrated data corresponding to the region between the Light-Gates can be extracted from the WIMU Sled Displacement data and the known Light-Gate positions. The WIMU derived Sled Average Velocity values within this region can then be calculated and compared to Light-Gate derived values.

D. Estimation of Full Run Duration

Ideally, the integration period would be the inverse of requested sample rate. However, variable on-device sampling rate and wireless data loss can make this a poor estimate of the system’s effective sample period. Improved estimates can be made using per-sample times, or known sample counts and durations; however the system used lacked accurate time-stamping, preventing such direct estimations of effective integration period.

Light-Gate durations and preliminary WIMU integration data were instead used. From the 52 valid WIMU data-files, the average sample count in the Light-Gate region was estimated at approximately 67.7% (standard deviation of 1.9%) of the full run sample count (Table 1). An initial estimate of the expected Full Run Duration for each Assassin setting could then be provided by dividing the Light-Gate timing value by 0.677, dividing PP sample count by this gave an estimate of the Integration Period. Using the final integrated data, the validity of the timed region fractional duration estimate was checked (See Fig. 5), with a best-fit line showing similar results to the initial estimate as can be seen in Table 1. An improved WIMU system with improved time-stamping would allow direct determination of sample period, removing the current implementation’s reliance on the Light-Gates for estimating these.

TABLE I. TIMING REGION DURATION ESTIMATES

Run Dur. [%]	Pre-LG 1	Post-LG 2	Un-Timed	Timed
Initial Est.	27.3	5.1	32.3	67.7
Std. Dev.	2.24	0.50	1.90	1.90
Final Check	25.72	5.32	31.05	68.95

a. Estimates of Region Average Durations as Percentage of Pushing-Phase Duration

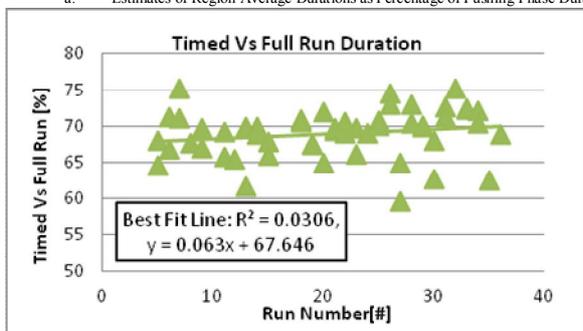


Figure 5. Timed Region Duration Estimates Based on WIMU Results.

E. Identification of Valid Data

Not all recorded data-sets were of sufficient quality to yield reliable integration results. A fully automated system should be able to distinguish good and bad data to ensure valid results are generated. Several poor data rejection conditions were identified, with suitable tests performed during analysis and warnings provided as follows:

1) Missing Events in WIMU Data

Recording started too late or finished too early, cutting off PO or IP, causing failure during event detection stages.

2) Missing Data from Other Sources

Other essential data was un-available (i.e. Light-Gate).

3) Excess Data Loss

PP had less than 50% of the data samples expected.

F. Integration Process

An initial estimate of sensor offset is made based on the average value of passive sensor samples in the PPO region (1). The initial offset is applied to each PP sensor sample but will be iteratively refined later.

$$acc_{offset} = \frac{\sum_{k=0}^N acc_k}{N} [m / s^2] \quad (1)$$

For N Quiescent raw sample values, acc_k , in PPO Region

An integration period that accounts for wireless data loss, removal of bad sensor samples or WIMU internal issues that change the effective sample rate is needed. From the PP sample count and full run duration we can estimate an effective sample rate for each data-file and hence integration period (2). However, without accurate sensor data time-stamping we must use Light-Gate timing data estimations of the full run duration.

$$t = \frac{T_{PP}}{i_{PP}} = \frac{T_{IP} - T_{PO}}{i_{IP} - i_{PO}} \approx \frac{T_{LG} / 0.677}{i_{IP} - i_{PO}} \quad (2)$$

For Integration Period t , Time T and Sensor Sample i

Then integration of PP WIMU data can begin, converting offset adjusted Acceleration a , to Velocity v (3) and Displacement s (4) for the n^{th} sensor sample since PO.

$$v_n = v_{n-1} + a_n t \quad (3)$$

$$s_n = s_{n-1} + v_{n-1}t + \frac{1}{2} a_n t^2 \quad (4)$$

Known and WIMU estimated Displacement at Impact are compared to each other and used to refine the Offset value in an iterative binary search manner. This process is explained in the C-style pseudo-code in Fig. 6.

```

WHILE( !complete && i<max_iteration ){
  IF( ABS(displ_error) >= target_accuracy ){
    IF(displ_error > 0){
      test_offset = offset - offset_step;
    }ELSE{ // displ_error < 0
      test_offset = offset + offset_step;
    }
  }
}
    
```

```

z_vel = integrate(z_acc, test_offset, period);
z_displ=integrate(z_vel, test_offset, period);
impact_displ = z_displ[-1];
new_error = impact_displ - target_displ;
IF( ABS(new_error) < ABS(displ_error) ){
  offset = test_offset; // update offset
  displ_error = new_error;} // update error
offset_step /= 2;
i++;
}ELSE{ // ABS(displ_error) < target_accuracy
  complete=True; } }
    
```

Figure 6. Iterative Sensor Offset Refinement

IV. RESULTS

Of the 70 data-files processed, 52 were determined to contain valid PP data, all of which were segmented successfully on the first attempt, with 12.17 (Standard Deviation 0.98) sensor offset refinement iterations required. Samples of output shows the WIMU and Light-Gate derived average velocity over the timed region (indicated by height of cyan shaded region and horizontal cyan dashed line respectively) are very similar, having a difference of 0.06m/s² (see Fig. 7).

Contour graphs of mean Light-Gate and WIMU derived Sled Average Velocity for each Weight and Inclination setting (See Fig. 8 and 9) show similar magnitudes and a trend for reduced velocity with increasing resistance (i.e. additional Weight and steeper Inclination) with average difference of -0.005±0.074 meters per second. The Root Mean Squared Error (RMSE) between the two (See Fig. 10) better illustrates this high similarity, with a maximum error of 0.105 m/s, most results are well within the target accuracy level of 0.1 m/s across a wide range of speeds and equipment settings used.

A. Light-Gate Un-Certainty

The Light-Gate derived Sled Average Velocity is not

exact, as both the timing and distance measurements required have an associated uncertainty. The time is given in seconds to two places of decimals so estimated un-certainty is 0.01 seconds. The un-certainty in the distance travelled by the sled, due to errors in positioning the Light-Gates on 1m tall tripods, was estimated at 0.02 meters. Additionally, inclination affecting the rail length between the Light-Gates was trigonometrically estimated as approximately 0.02 meters (2.5 meters at 0° versus 2.519 meters at 7°). Adding these gives an overall maximum distance error estimate of 0.04 meters. By combining the lower time with upper distance estimates and vice-versa, the un-certainty in Light-Gate Derived Sled Average Velocity was estimated as ranging from 0.066 to 0.115 meters per second (±2.3% on average) (See Table II).

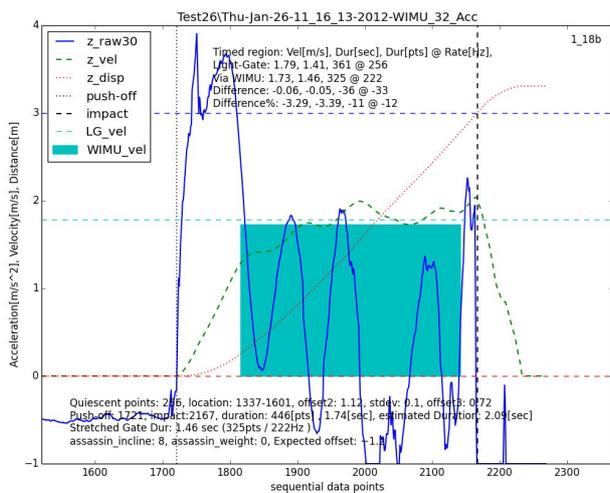


Figure 7. Processed Run showing Good Average Velocity Agreement.

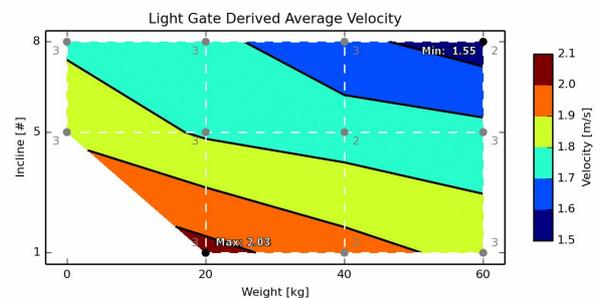


Figure 8. Timed region Light-Gate derived Average Sled Velocity.

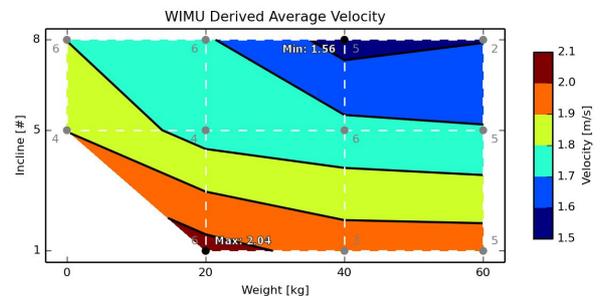


Figure 9. Timed region WIMU derived Average Sled Velocity

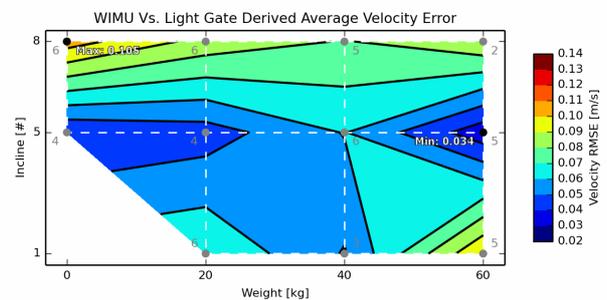


Figure 10. RMSE of Sled Average Velocity for the 2 Methods. White regions indicates no data available, gray numbers denote number of samples used to produce value.

TABLE II. LIGHT-GATE UNCERTAINTY ESTIMATES

For 2.5±0.04 m	Duration [sec]	Velocity [m/s]	Un-Certainty
Slowest	1.67±0.01	1.497±0.066	±2.199%
Fastest	1.09±0.01	2.294±0.115	±2.518%
Average	1.43±0.01	1.748±0.080	±2.300%

V. DISCUSSION

WIMU and Light-Gate differences are low, generally within the target and often within the Light-Gate uncertainty levels. Additionally WIMU data provides a more complete picture of the Pushing-Phase, yielding velocity and displacement at each sensor sample. This enables the creation of arbitrary virtual timing intervals; analysis of the development of velocity; the detection of individual step features, etc. Future WIMU based systems could be even more low-cost, small and self-contained than that developed here; allowing use across gym, test-track and on-ice sleds, without requiring trained users, extensive sled modifications or costly installation of trackside equipment. Such a system holds great potential for: improving the understanding of the Push-Start; identifying good athletes and determining the effectiveness of coaching and training interventions.

VI. CONCLUSION

Although the Bob-Skeleton Push-Start is considered crucial to performance, it is poorly understood due to a lack of detailed data or accessible methods for gathering such data. Using WIMUs to instrument Assassin a method for automatic segmentation, drift correction and integration of Accelerometer data to Velocity and Displacement was developed. Sled Average Velocity results were similar to Light-Gate with Root Mean Squared Error within or similar to the target accuracy and un-certainty levels. The system’s accuracy, low-cost, ease-of-use and portability, could provide greater access to such quantitative performance data, with its highly detailed data enabling improved understanding of the Push-Start. These could lead to improved methods for Selection, comparison and training, potentially providing a valuable competitive edge.

ACKNOWLEDGMENT

We would like to acknowledge the support of Science Foundation Ireland (SFI) in funding the CLARITY Centre for Sensor Web Technologies under grant 07/CE/I1147, and the technical and financial assistance of the University Of Bath and UK Sports in this deployment.

REFERENCES

[1] C. Zanoetti, A. La Torre, G. Merati, E. Rampinini, and F. M. Impellizzeri, "Relationship between push phase and final race time in skeleton performance," *The Journal of Strength & Conditioning Research*, vol. 20, Aug. 2006, pp. 579-583.

[2] W. A. Sands, et al., "Skeleton: Anthropometric and physical abilities profiles: US national skeleton team," *Sports Biomechanics*, vol. 4, Jul. 2005, pp. 197-214.

[3] FIBT Bob-Skeleton Results. Available: www.fibt.com/races-results/results.html, Retrieved: Sep. 2014.

[4] FIBT International Skeleton Rule 16.9 - Starting Area: Bobsleigh and Skeleton, FIBT, 2009. Available: http://www.fibt.com/fileadmin/Rules/Reg_SKELETON-2009_-_E.pdf, Retrieved: Sep. 2014.

[5] N. Bullock, W. G. Hopkins, D. T. Martin, and F. E. Marino, "Characteristics of performance in skeleton World Cup races," *Journal of Sports Sciences*, vol. 27, Feb. 2009, pp. 367-372.

[6] C. Cook, D. Holdercroft, S. Drawer, and L. P. Kilduff, "Designing a warm-up protocol for elite bob-skeleton athletes," *International Journal of Sports Physiology & Performance*, vol. 8, Mar. 2013, pp. 213-215.

[7] S. Colyer, "Consecutive Days of Push Start Testing May Mask the Effect of Starting Style In Skeleton - a Pilot Study," presented at the BASES Biomechanics Interest Group, University of Ulster, Belfast, Northern Ireland, Apr. 2012.

[8] N. Bullock, et al., "Characteristics of the start in women's World Cup skeleton," *Sports Biomech*, vol. 7, Sep. 2008, pp. 351-360.

[9] S. Lemberg, O. Schachner, and C. Raschner, "Development of a measurement and feedback training tool for the arm strokes of high-performance luge athletes," *Journal of Sports Sciences*, vol. 29, Nov. 2011, pp. 1593-1601.

[10] F. Braghin, F. Cheli, M. Donzelli, S. Melzi, and E. Sabbioni, "Multi-body model of a bobsleigh: comparison with experimental data," *Multibody System Dynamics*, vol. 25, Aug. 2010, pp. 185-201.

[11] USA-Bobsled-Skeleton-Federation. Combine Test Protocol. Available: www.teamusa.org/USA-Bobsled-Skeleton-Federation/, Retrieved: Sep. 2014.

[12] N. Bullock, et al., "Talent identification and deliberate programming in skeleton: Ice novice to Winter Olympian in 14 months," *Journal of Sports Sciences*, vol. 27, 2009, pp. 397-404.

[13] C. Valle, "Think you want to be a bobsledder?" Available: <http://www.freelapusa.com/think-you-want-to-be-a-bobsledder/>, Retrieved: Sep. 2014.

[14] B. O'Flynn, et al., "The development of a novel miniaturized modular platform for wireless sensor networks," *Proceedings 4th International Symposium on Information Processing in Sensor Networks (IPSN)*, Apr. 2005, pp. 370-375.

[15] J. Barton, et al., "Miniaturised inertial measurement units for wireless sensor networks & novel display interfaces," *Proceedings 55th Electronic Components & Technology Conference (ECTC)*, Jun. 2005, pp. 1402-1406.

[16] J. Barton, A. Gonzalez, J. Buckley, B. O'Flynn, and S. C. O'Mathuna, "Design, Fabrication and Testing of Miniaturised Wireless Inertial Measurement Units (IMU)," *Proceedings 57th Electronic Components & Technology Conference (ECTC)*, May 2007, pp. 1143-1148.

[17] C. Rodde, "A Wireless Inertial Measurement System for Tennis Swing Dynamics," *MEngSc (Microelectronic design) Taught Masters*, Department of Microelectronic Engineering, University College Cork (UCC) and Tyndall National Institute, Cork, 2009.

[18] M. Gaffney, et al., "Wearable wireless inertial measurement for sports applications," *Proceedings 33rd IMAPS-CPMT, Gliwice-Pszczyna, Poland*, Sep. 2009, pp. 138-141.

[19] M. Gaffney, et al., "A Smart Wireless Inertial Measurement Unit System," Presented at Pervasive Health Conference, UCD, Dublin, Ireland, May 2011.

[20] M. Gaffney, M. Walsh, B. O'Flynn, and C. O Mathuna, "An automated calibration tool for high performance Wireless Inertial Measurement in professional sports," *Proceedings IEEE Sensors Conference, Limerick, Ireland*, Oct. 2011, pp. 262-265.

[21] Brower Timing Systems, Test Centre (TC) Timing System. Available: www.browertiming.com, Retrieved: Sep. 2014.