Secure Publication Subscription Framework for Reliable Information Dissemination

Shugo Yoshimura
Graduate School of Information Sci.
and Electrical Eng., Kyushu Univ.
Fukuoka, Japan
yoshimura.shugo.822@s.kyushu-u.ac.jp

Kouki Inoue
Graduate School of Information Sci.
and Electrical Eng., Kyushu Univ.
Fukuoka, Japan
inoue.kouki.882@s.kyushu-u.ac.jp

Dirceu Cavendish
Graduate School of Eng.
Kyushu Institute of Tech.
Iizuka, Japan
Dirceu_cavendish@yahoo.com

Hiroshi Koide
Research Institute of Info. Tech.,
Kyushu Univ.
Fukuoka, Japan
koide@cc.kyushu-u.ac.jp

***Abstract*—In this study, to make it easy for everyone to distinguish the right information from the wrong information, we suggest a new framework (Secure Publication Subscription Framework) that defines the reliability of publishers and provides it to subscribers. Nowadays, services like blogs and social media make available large amounts of information easily. On the other hand, there is a lot of unreliable information on the Internet. It is difficult to distinguish between true and false information. This problem is known as fake news and has become a serious problem. To solve this problem, we suggest a new framework for publishers and subscribers. The framework allows subscribers to easily confirm the authenticity of information by registering publishers and subscribers, and tracking publishers' reputation via a reputation score, guaranteeing the quality of the information that subscribers view. In this study, we show a proof of concept of a simple Secure Publication Subscription Framework and confirm that it is possible to implement a framework with the proposed functionality. We also confirm that the reputation score can be used as an indicator of the reliability of the information by using 1000 randomly generated articles within the framework.**

Keywords-dissemination; publication; social networking; authenticity of information; reputation score.

I. Introduction

In recent years, Internet technologies have made great progress, with the population of Internet users increasing rapidly. Thanks to services like blogs and social media, anyone can get a large amount of information easily. Nowadays, we can see what is happening around the world, no matter where we are.

On the other hand, there is a lot of unreliable information on the Internet. It is difficult to distinguish between true and false information. This problem is known as fake news and has become a serious problem. Fake news is fabricated information that mimics news media content in form but not in organizational process or intent [1]. It is not just a prank, but a serious problem. As an example, during the 2016 United Status presidential election, fake news was highly used and had a big impact on Twitter [2] [3].

To solve this problem, we suggest a new framework for publishers and subscribers. This framework allows subscribers to easily confirm the authenticity of information by registering publishers and subscribers, guaranteeing the publisher of the information that subscribers view, checking the information challenge from subscribers, and providing the publisher's reputation score that increases or decreases as a result of the authenticity of the information.

This framework consists of three parts, Publisher, Subscriber and Arbitrator. The main role of the Publisher is publishing articles or news. The Subscriber registers with the Publisher and subscribes for publications. The Arbitrator provides the Publisher's reputation and verifies the information challenge from the Subscriber.

The paper is organized as follows. Related work is included in Section II. Section III describes our proposed secure publication/subscription reference model. Section IV describes a proof of concept implementation of the reference model. Section V describes two experiments used to track the performance of the proposed publication/subscription model. Section VI presents the performance results and discussions. Section VII summarizes our studies and addresses directions we are pursuing as follow up to this work.

II. Related work

Previous research on publication/subscription systems have covered various areas, such as security, confidentiality and scalability.

Nakamura and Enokido [4] focused on a peer to peer publication/subscription model where multiple topics are supported. In that work, they propose a subscription initialization protocol to ensure that peers not authorized to have access to topics do not have access to them. They do not address the quality of the information exchanged within topics. In contrast, our framework addresses information quality on a

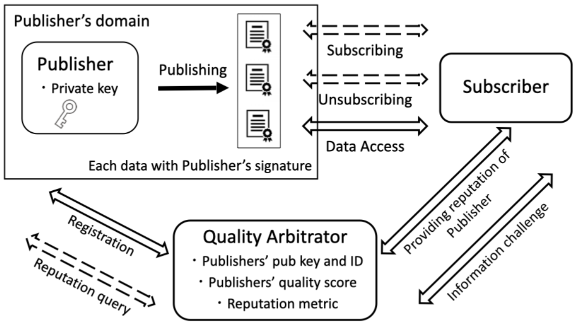

Figure 1. Secure Publication/Subscriber Architecture

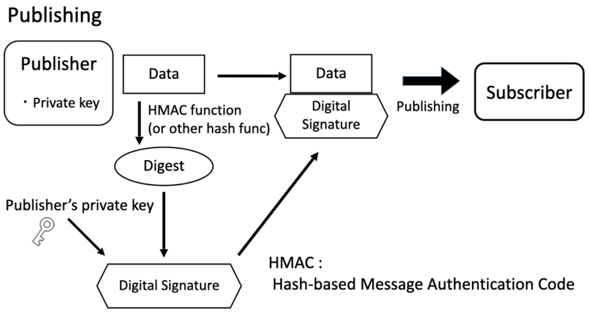

Figure 2. Signed publishing

generic publication/subscription architecture, not necessarily requiring a peer to peer model.

Salem [5] addresses the problem of authenticating users of a pub/sub system containing a message broker in a privacy-preserving way. The proposal supports mutual authentication in a scalable way, and may be adopted by pub/sub systems with a broker. In contrast, our work does not focus on anonymity of publishers/subscribers, although our pub/sub model could be adapted to include a broker, if necessary.

In Srivatsa [6], a secure event dissemination protocol is proposed where encryption and authorization keys are used on top of an IP network that does not provide confidentiality nor integrity of data. In contrast, although our pub/sub model supports integrity verification of data, our focus is on the control of the quality of data published.

Bovet and Makse [3] describe an information ranking mechanism to fight unreliable (spam) data in a pub/sub system model with a broker reference architecture. They propose to rank information as a way to avoid blacklisting. However, their ranking system is still based on participants voting. Although the purpose of the research is similar to ours, our solution to control quality of disseminated data is based on an arbitrator that is supposed to be able to verify data quality on specific domains, rather than relying on voting.

III. SECURE PUBLICATION/SUBSCRIPTION

This section describes the operation of the Secure Publication Subscription Framework in detail.

Figure 1 describes our proposed secure publication/subscription system architecture. Multiple publishers provide signed data contents to consumers, or subscribers. Data content quality is tracked by an independent quality arbitrator. The quality arbitrator provides publishers' reputation to subscribers. Also, the arbitrator may receive data truthfulness challenges from subscribers.

A. Sec Pub/Sub Components

Figure 2 illustrates how Publishers provide signed data contents. Publishers also produce a digest of the data content using standard asymmetric cryptography, using their private key to ensure data integrity.

Figure 3 illustrates publisher/subscriber interfaces. The subscriber requests subscription services from a publisher and receives the publisher public key used to verify data authenticity. Once the subscription service has been agreed upon, an information retrieval interface is used to request signed data from the publisher.

Figure 4 illustrates the subscriber's data processing of published data. Data processing includes data integrity verification and confirmation authorship. The subscriber verifies the digital signature and the digest of the data, using the publisher public key. In this process, the subscriber verifies the integrity of the received data and confirms the data's authorship.

Figure 5 illustrates publisher reputation tracking feature of the secure pub/sub framework. Each publisher registers first with the quality arbitrator, upon which its public key is passed to the arbitrator. The arbitrator then tests the publisher's possession of the corresponding private key as part of the registration. Each successfully registered publisher is associated with a reputation score metric, which can be queried by both the publisher itself as well as subscribers.

Figure 6 illustrates the subscriber/quality arbitrator interfaces. Subscribers can request publisher's reputation score from the arbitrator. In addition, subscribers can challenge publisher's trustfulness for each data received. The quality arbitrator, upon receiving the challenge, verifies data truthfulness, and adjusts the publisher reputation score according with data verification status.

B. Reputation Algorithm

The reputation score of a publisher is defined as $score =$ (the number of correct data) / (the number of all published data). However, as the quality arbitrator may not estimate correctly every and all data published, we introduce a noise model for data verification, as per Figure 7. In the model, p is the probability that a true piece of data be recognized as false, whereas q represents the probability of a false piece of information be admitted as true. In the experimental section, we exemplify the arbitrator score reputation tracking on two publisher scenarios: i- trusted publisher (all data is truthful);

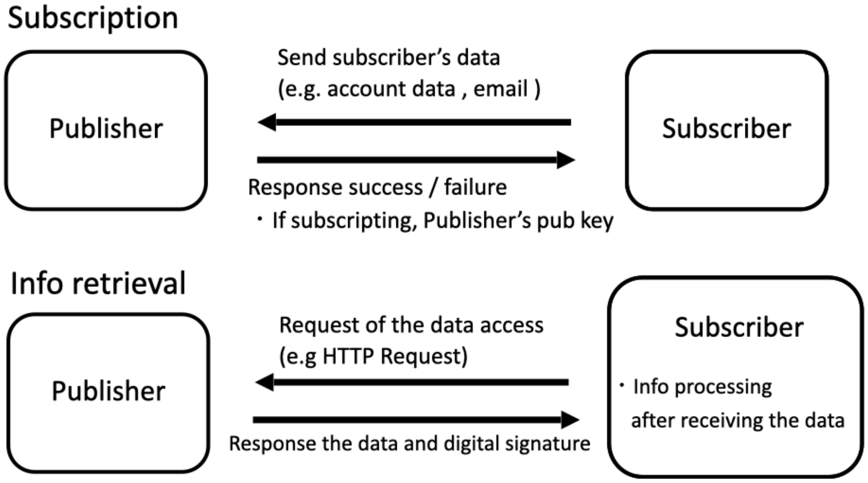
Subscription
Publisher
Send subscriber's data
(e.g. account data , email)
Subscriber
Response success / failure
・If subscripting, Publisher's pub key
Info retrieval
Publisher
Request of the data access
(e.g HTTP Request)
Subscriber
・Info processing
after receiving the data
Response the data and digital signature

Figure 3. Subscription and Information Retrieval

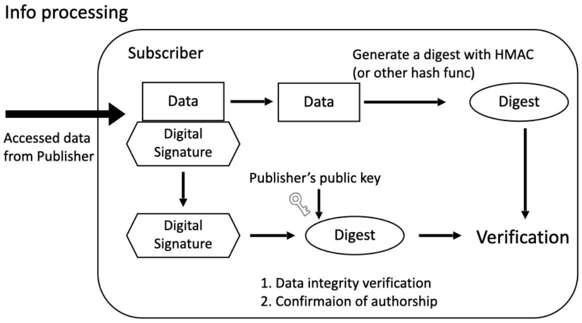

Figure 4. Data Integrity Verification

ii- untrusted publisher; Publisher produces up to 1000 data pieces (the data can be right or wrong).

IV. IMPLEMENTATION

In this section, we describe an overview of the implementation of Publisher, Arbitrator, Subscriber. We implemented the Publisher and the Arbitrator with Node.js and Express that is a JavaScript Web framework, and we implemented the Subscriber with Python3. The Publisher and the Arbitrator operate like a Web server, independently, and the Subscriber accesses them according to the scenarios. The versions used in the implementation are summarized in Table I.

A. Publisher

The Publisher is implemented with Node.js and Express, and it operates as a Web server. Figure 8 describes the implementation. The Publisher has subscriber registration, login, some data pages and digital signatures. In addition, it has a MySQL database that saves the Subscriber's name and hashed password. If it receives an HTTP Request from the Subscriber, it replies with an HTTP Response and sends the data.

TABLE I
IMPLEMENTATION

Application	Version
Node.js	12
MySQL	5.7
Python	3.9.12

B. Arbitrator

The Arbitrator is also implemented with Node.js and Express, and operates as a Web server. Figure 9 describes the implementation of the Arbitrator. The Arbitrator receives the Publisher's registration, reputation query, as well as information challenge and request for publisher's pub key. Additionally, the Arbitrator supports a MySQL database, which saves the Publisher's name, password, public key and Publisher reputation score. Firstly, the Publisher registers its name, password and public key. In our experiment scenarios, the Publisher's information is saved in initial state, so this step is omitted. If the Subscriber requests the Publisher's public key, the Arbitrator responds to it. If the Subscriber requests the Publisher's reputation score, the Arbitrator sends the Publisher's score. If the Arbitrator receives an information challenge from the Subscriber, it verifies data truthfulness, updating the score of the Publisher.

C. Subscriber

The Subscriber is implemented with Python3. It accesses the Publisher and the Arbitrator according to the different scenarios. During information processing, it verifies the integrity of received data and confirms data authorship (Figure 10).

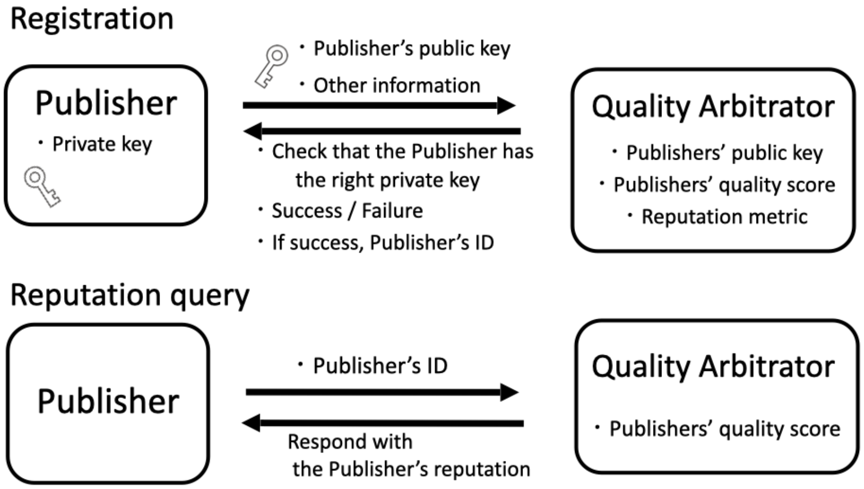

Figure 5. Publisher registration and Reputation Tracking

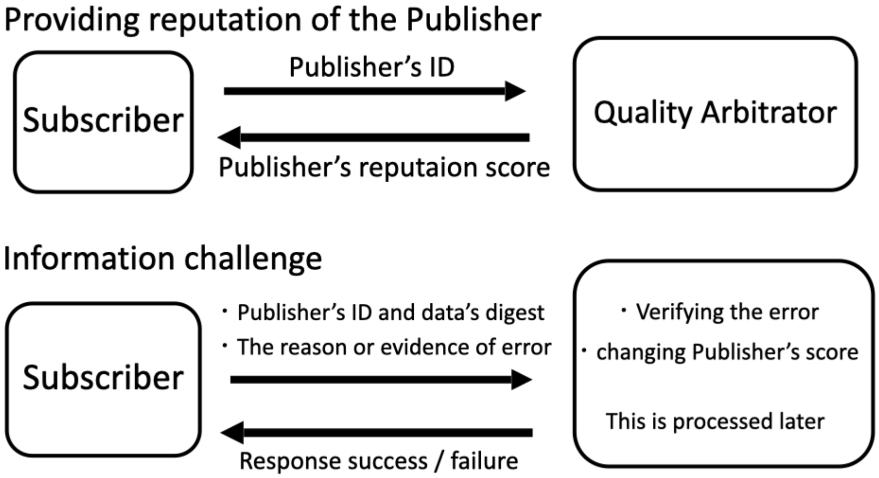
Providing reputation of the Publisher
Subscriber
Publisher's ID
Publisher's reputaion score
Quality Arbitrator
Information challenge
Subscriber
・Publisher's ID and data's digest
・The reason or evidence of error
Response success / failure
・Verifying the error
・changing Publisher's score
This is processed later

Figure 6. Reputation service interface

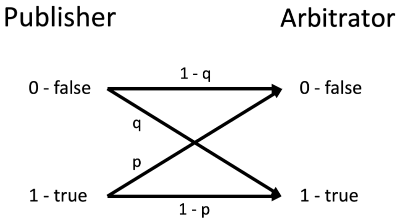

Figure 7. Noisy Channel Model

V. Experiment

This section demonstrates the evolution of the reputation estimator and reputation score for the Secure Publication Subscription Framework using 1000 randomly generated true and false data.

The resulting graph shows 3 lines:

- Actual reputation score: the reputation score actually obtained after going through the Secure Publication Subscription Framework,
- Expected reputation score : the expected value of the reputation score obtained from the actual truth of the data, p and q,
- True reputation : proportion of data that is actually true.

We exemplify the secure publication/subscription model with the following scenarios:

A. Scenario 1

1) Subscribers register and login in with the Publisher

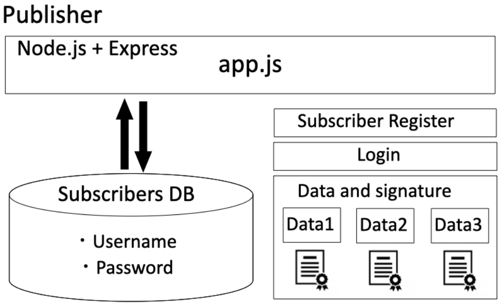

Figure 8. Publisher

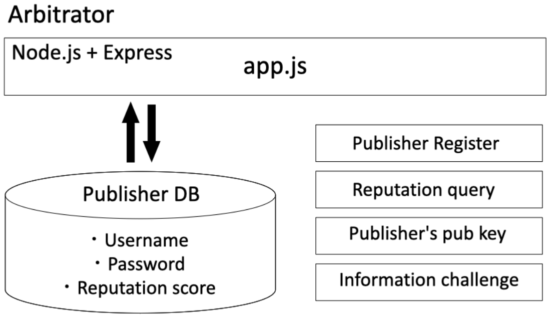

Figure 9. Arbitrator

2) Subscribers subscribe to data from the Publisher
3) Subscribers retrieve the data
4) Subscribers send a query about the Publisher's reputation to the Arbitrator

In Scenario 1, the credibility of the Publisher's data is 100%, hence the Publisher's true reputation is 1. However, the expected reputation score is (1 - p) because there is a possibility that the Arbitrator will judge it to be false. In this experiment, the values of the p and q are set to 0.3 to check the reputation scores. To show that the accuracy of the reputation score does not drop even if the accuracy of the true/false discrimination is not so high, p and q were set to fairly low values. We think that there is still room for further study on this value.

Figure 11 shows the graph of the results for Scenario 1.

B. Scenario 2

In scenario 2, Publisher's data is not always true.

1) Subscribers register and login in with the Publisher
2) Subscribers subscribe to data from the Publisher
3) Subscribers retrieve the data
4) Subscribers issue an information challenge
5) The Arbitrator decides the data as false, and updates the Publisher's reputation
6) Subscribers query the reputation of the Publisher from the Arbitrator

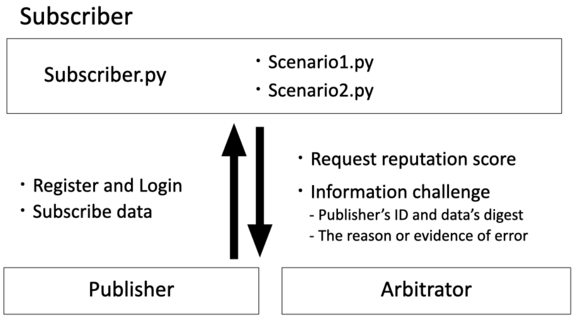

Figure 10. Subscriber

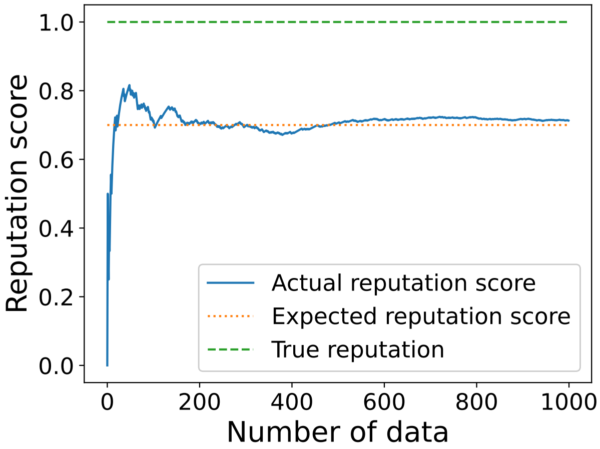

Figure 11. scenario 1

Let a be the probability that the publisher's data is false. Then, the expected value of the true reputation is (1 - a), while the expected reputation score is a * q + (1 - a) * (1 - p). In Scenario 2, step 1, 2, 3 are the same as in Scenario 1. However, the Subscriber carries out an information challenge in steps 4 and 5. The probability of judging the data to be correct was varied between 0.8 and 0.6, and p and q were 0.3 to check the reputation scores for each case.

The experimental results are shown in Figures 12 and 13.

VI. Performance Analysis

In this section, we present the reputation tracking results of our secure pub/sub system. In scenario 1, the final three scores obtained from the 1000 data points are shown in Tables II.

TABLE II
SCENARIO 1

Actual reputation score	0.713
Expected reputation score	0.700
True reputation	1.000

In scenario 2, the final three scores obtained from the 1000 data points are shown in Table III and IV.

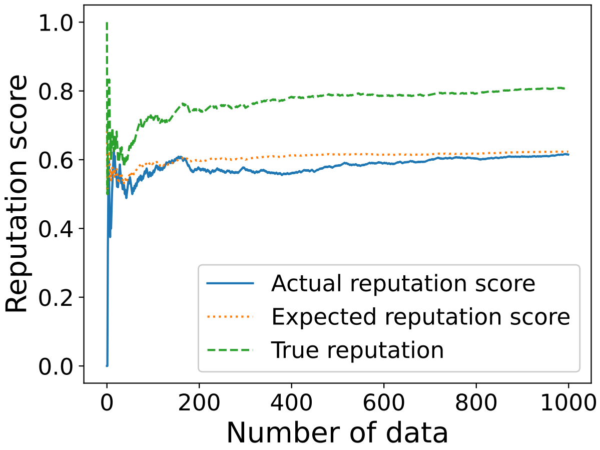

Figure 12. scenario2 data accuracy = 0.8

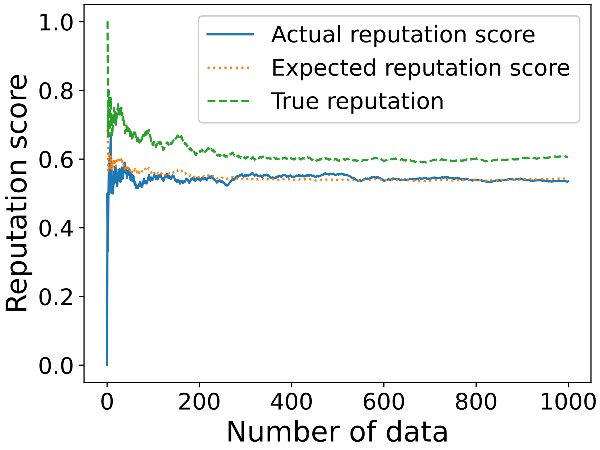

Figure 13. scenario2 data accuracy = 0.6

From these experimental results, with a sufficient number of data points and a certain degree of accuracy in determining the truth of the data, we see that the actual reputation score converges to the expected reputation score.

Moreover, we use a noise model for data verification, and we define the expected reputation to be a * q + (1 - a) * (1 - p). So, if p and q are known, the Publisher's true reputation can be estimated from the actual score.

VII. Conclusion and Future work

In this study, we proposed a new framework (Secure Publication Subscription Framework) that allows subscribers to check the accuracy of information based on the authenticity of the publisher's historical data by checking the reputation score. In this framework, subscribers can check the reputation score of the publisher and challenge data reliability if the information is suspected to be unreliable. We also conducted experiments on the publisher's reputation score, and found that the actual reputation score approximates the expected value calculated from the probability of correctly judging the reliability of information.

TABLE III
SCENARIO 2 DATA ACCURACY = 0.8

Actual reputation score	0.615
Expected reputation score	0.623
True reputation	0.808

TABLE IV
SCENARIO 2 DATA ACCURACY = 0.6

Actual reputation score	0.535
Expected reputation score	0.543
True reputation	0.607

With fake news becoming a major problem, it is important to have a system that allows subscribers to easily verify the authenticity of information. As such a system, our framework can be one of the promising options.

As future research, integration of AI(Artificial Intelligence) algorithms to automatically identify fake news with expert arbitrators is a promising path. Although the accuracy of discriminating fake news has been a challenge for AI technologies, our expert framework can aid by using AI algorithms to improve false positives/negatives. Combined with these technologies, we believe that a robust data reliability framework for publication/subscription platforms can emerge.

References

[1] D. M. J. Lazer *et al.*, "The science of fake news." *Science*, pp. 1094–1096, 2018.

[2] N. Grinberg *et al.*, "Fake news on Twitter during the 2016 US presidential election." *Science*, pp. 374–378, 2019.

[3] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 US presidential election." *Nature communications*, pp. 1–14, 2019.

[4] S. Nakamura, T. Enokido, and M. Takizawa, "Subscription Initialization (SI) Protocol to Prevent Illegal Information Flow in Peer-to-Peer Publish/Subscribe Systems," *19th International Conference on Network-Based Information Systems*, pp. 42–49, Sept. 2016.

[5] F. M. Salem, "A Secure Privacy-Preserving Mutual Authentication Scheme for Publish-Subscribe Fog Computing," *14th International Computer Engineering Conference*, pp. 213–218, Dec. 2018.

[6] M. Srivatsa and L. Liu, "Secure Event Dissemination in Publish-Subscribe Networks," *27th International Conference on Distributed Computing Systems (ICDCS '07)*, pp. 22–22, June 2007.