Data Analytics for 5G Networks: A Complete Framework for Network Access Selection and Traffic Steering

Sokratis Barmpounakis,
Panagis Magdalinos,
Nancy Alonistioti

Dept. of Informatics and
Telecommunications,
National and Kapodistrian
University of Athens

Athens, Greece
email: sokbar@di.uoa.gr
panagis@di.uoa.gr
nancy@di.uoa.gr

Alexandros Kaloxylos

Dept. of Informatics and
Telecommunications
University of Peloponnese
Tripoli, Greece
email: kaloxyl@uop.gr

Panagiotis Spapis, Chan Zhou

Huawei German Research Center
Munich, Germany
email: panagiotis.spapis@huawei.com
chan.zhou@huawei.com

Abstract— **Data analytics has been widely accepted as a key enabler for 5G cellular networks. They can be applied in order to improve the performance of several mechanisms ranging from network management and traffic engineering to radio access network selection and traffic steering. Towards this end, this paper presents three supplementary mechanisms and provides a mapping of the required new functional entities into the latest version of the 5G architecture as specified by 3GPP. Additionally, it evaluates the performance of one of the mechanisms that targets the minimization of the information collected and used by the data analytics engine. All in all, this paper provides a holistic framework based on the use of data analytics that addresses several key 5G-oriented, resource management and traffic steering challenges.**

Keywords- 5G cellular networks; data analytics; traffic steering; mobility management.

I. INTRODUCTION

Lately, the use of data analytics to support the operation of 5G cellular networks has emerged as a hot research topic [1]. The application of data analytics covers a wide range of network operations such as network management and control as well as policy enforcement. Figure 1 illustrates how this is achieved. More specifically, data are collected from a number of network components and may include a variety of information fields such as the quality of the wireless channel, the network load, accounting information, configuration and fault indications, the profile of the subscribers, etc. This data is stored and updated regularly. When collected, it is pre-processed, and any necessary transformation, discretization, normalization, outlier detection and dimensionality reduction is applied. The result of this process is then provided to a data analysis procedure which builds a model that caters for knowledge extraction from the processed data. For example, the result of this process may be the identification of situations where the occurrence of some specific events (e.g., a significant increase of the number of high mobility users)

causes a given result (e.g., increase of the handover blocking probability). The knowledge model may also include some solutions for specific situations (e.g., force the network components to place high moving users to macro cells). The knowledge discovery results can then be communicated to either policy, management or control modules. These modules can employ this information in order to optimize the operation of the network and improve its performance. In this paper we provide a complete framework on how data analytics can be used to support Radio Access Technology (RAT) selection and traffic steering.
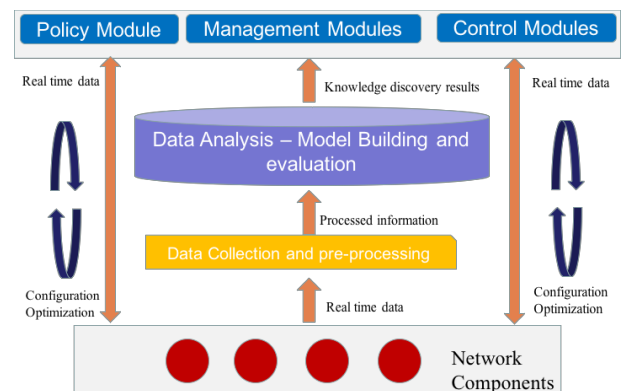

Figure 1. Big data analysis for cellular networks

To answer why data analytics are needed for RAT selection and traffic steering one has to consider the following. One of the most promising family of use cases has been named by the 3rd Generation Partnership Project (3GPP) as enhanced Mobile BroadBand (eMBB) services. These are characterized by throughput demanding services, high data volumes and significant numbers of users in

specific areas (e.g., stadiums, malls etc.). To support these services, the following architectural updates/ changes are required [2]:

- densified deployment of base stations
- support of heterogeneous access networks (e.g., operating below and above 6GHz, WiFis etc.)
- deployment of efficient integration mechanisms among heterogeneous access networks
- support of flexible and dynamic steering of users' sessions through these access networks

These design decisions create a complex heterogeneous environment. Thus, the selection of the most appropriate RAT and base station to support a user's session is no longer a simple task. Under a homogeneous network this selection is simply based on the evaluation of the signal quality as received by neighboring base stations. However, for 5G networks the number of choices is going to be considerably higher. The dense deployment of base stations will offer at any time a high number of candidate macro, micro and femto cells for a user equipment (UE). Also, different RATs have different characteristics such as the supported throughput, the tolerance to interference, the transmission power and the respective energy consumption, the tariffing policy of the operator for using an access network etc. Thus, a selection should consider, apart from the signal quality, parameters like the mobility of a UE, the Quality of Service (QoS) required by its active services, the battery status of the UE, the preferences of the users, the network access and the technologies supported by the UE.

This device/technology-wise heterogeneous environment calls for a novel and comprehensive framework, which will take into consideration the available information related to the values of the abovementioned characteristics. A plethora of approaches has been proposed in the past to address this issue [3]. These approaches are based on the dynamic collection and evaluation of contextual information. The evaluation is usually done either through the calculation of utility functions or using neural networks or fuzzy logic mechanisms to reach a decision.

However, these solutions assume that the contextual information is somehow already in place and they neglect to estimate the signaling overhead that is required for the collection of contextual information. Moreover, although they consider that the preferences of the users are known, there is up to now no automatic solution to acquire this information, apart from having the users to set their preferences manually. Finally, these approaches are reactive schemes that take decisions on the fly. However, the complexity of the 5G networks requires that when available, some decisions should be proactive, otherwise the UEs will constantly monitor the environment and switch rapidly from one base station to another. This will be an extra overhead for the UEs' computational resources and their batteries.

To address these points the current paper builds upon our prior work presented in [4] and [5]. In these papers we have presented two independent solutions:

a) a RAT selection scheme, named Compass (Context Aware RAT Selection), for 4G networks that achieves significant performance gains while keeping the signaling overhead at a minimum and b) a data analytics scheme, named Context Extraction and Profiling Engine (CEPE), that improves the performance of 4G cellular control functions (i.e., handover, call admission control and cell camping). Also, it calculates the profile of any user by monitoring its past actions. Based on this profile, the system has essentially the users' preferences and is able to predict their future behavior in terms of location, mobility and service activity. So, future decisions for serving the UEs' sessions under the most appropriate RAT can be taken essentially proactively. These two schemes can be adapted to operate under the 5G cellular systems and can be designed to operate complementary.

Very recently, the significance of these types of schemes for future cellular systems has been acknowledged by 3GPP. A new network function, called NetWork Data Analytics Function (NWDAF) has been introduced in [6]. Currently, this function has limited functionality and is employed in order facilitate operators' policy manipulation in real time. In the 5G architecture, the Policy Control Function (PCF) will be used to control the placement and service of UEs by the most appropriate cell and radio access technology. In future releases, it is expected that additional information will be provided from NWDAF to network control functions responsible to handle the mobility of active and idle UEs. Moreover, it is considered that the output of NWDAF will also feed new network functions related to Access Traffic Steering, Switching and Splitting schemes (ATSSS) [7]. This framework is currently under investigation by the 3GPP since it has been acknowledged that the traffic steering of user sessions through multiple RATs is of paramount importance for 5G networks.

This paper extends the work presented in [1]. The key contributions of our current paper are the following. At first, we illustrate the latest status of 3GPP concerning the support of data analytics and traffic steering in 5G networks. Then, we present in detail how our prior work can be adapted and mapped into 3GPP's network functions. Moreover, we provide detailed examples on their integrated operation. Additionally, we introduce and evaluate a new mechanism, called Context Information Processing (CIP) that targets the minimization of the acquisition of contextual information, which will be used by the data analytics mechanism of CEPE.

The rest of the paper is structured as follows. In Section II we present the current status of 3GPP specifications in relation to the overall architecture and the support of NWDAF and ATSSS network functions. In Section III we discuss the state of the art of research proposals for the support of RAT selection and traffic steering. In section IV we briefly describe the key functionalities of CEPE and Compass and we introduce the CIP mechanism. We also illustrate how these mechanisms can be integrated and mapped directly to the 5G network architecture. In Section V we evaluate the performance of CIP. We conclude the paper and sketch future research directions in Section VI.

II. THE 5G NETWORK ARCHITECTURE

In [6], 3GPP introduces the 5G network core architecture. Some key principles target to:

- Modularize existing network functions to enable their flexible placement into network components and potentially allow their alternate configuration for different use cases.
- Enable network functions to interact among them even directly if this is required. The architecture does not preclude the use of an intermediate function to help route control plane messages.
- Minimize dependencies between the Access Network (AN) and the Core Network (CN). The architecture is defined with a converged core network with a common AN - CN interface, which integrates different Access Types e.g., 3GPP access and non-3GPP access.
- Support "stateless" Network Functions (NFs), where the computational resources are decoupled from the storage resources.
- Support concurrent access to local and centralized services. To support low latency services and access to local data networks, user plane functions can be deployed close to the AN.

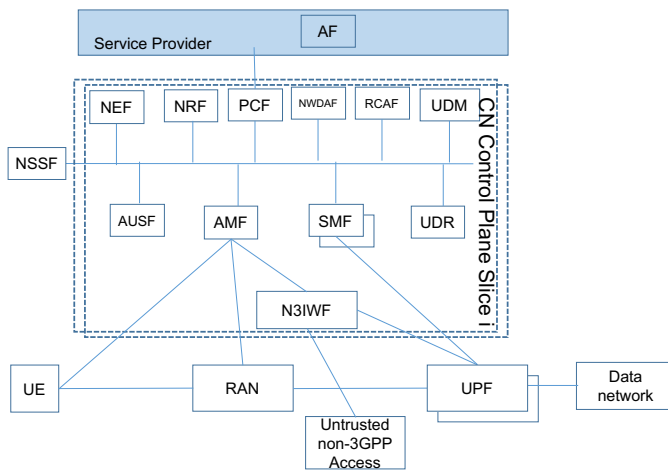The reference 5G architecture is illustrated in Figure 2.

Figure 2. 5G system architecture

The key network functions in the core networks are:

- Unified Data Management (UDM): supports the Authentication Credential Repository and Processing Function (ARPF).
- Authentication Server Function (AUSF): supports the authentication of end users.
- Policy Control function (PCF): supports unified policy framework to govern network behaviour, provides policy rules to control plane functions.
- Core Access and Mobility Management Function (AMF): supports mobility management, access authentication and authorization, security anchor functions and context management.

- Session Management Function (SMF): supports session management, selection and control of User Plane (UP) functions, downlink data notification and roaming.
- User Plane Function (UPF): is the anchor point for inter/intra RAT mobility, supports packet routing and forwarding, QoS handling for user plane, packet inspection and policy rule enforcement.
- Network Exposure Function (NEF): provides a means to securely exchange information between services and 3GPP NFs.
- Network Function Repository Function (NRF): maintains the deployed NF Instance information when deploying/updating/removing NF instances.
- Radio Congestion Awareness Function (RCAF): Informs the PCF about potential congestion issues in the radio access network.
- Network Slice Selection Function (NSSF): supports the functionality to bind a UE with a specific slice (i.e., a logical network that is configured and tailor-cut for a specific use case).

One of the novel core network functions, which are introduced is NWDAF. According to the latest standardization activities, the NWDAF will provide load level information at a network slice level. NWDAF provides slice specific network data analytics to the PCF, as well as the NSSF over their newly specified interfaces (i.e., Nnwdaf, Nnssf and Npcf). PCF will be using the NWDAF input for optimizing the policies assigned to each UE and its respective on-going sessions and data flows. NSSF will utilize the NWDAF's output to maintain the optimal UE/slice mapping for the diverse types of UEs and data flow characteristics. Note that multiple slices may be coexisting in certain areas. Besides PCF and NSSF, also TSSF (Traffic Steering Support Function) will essentially receive NWDAF's output for optimizing traffic steering decisions.

As for the support of untrusted non-3GPP access systems (like WiFi) an interworking function, called Non-3GPP Inter Working Function (N3IWF), is introduced. It supports the establishment of Internet Protocol Security (IPsec) tunnels with the UE and relays the information needed to authenticate the UE and authorize its access to the 5G CN. Note here that since Release 14 [8], 3GPP is also pursuing the tight integration of WiFis with the cellular system at the radio access network level. One such solution is the LTE-WLAN Aggregation (LWA) that capitalizes the dual connectivity split-bearer architecture, introduced in Release 12, where the aggregation of data links takes place at the Packet Data Convergence Protocol (PDCP) level. Another example is the RAN-controlled LTE-WLAN Interworking (RCLWI) mechanism that enables an eNB/New Radio gNodeB (NR gNB) to trigger a steering command to the UE to offload its traffic to a WiFi.

NWDAF and the interworking functions create the framework to define policies and allow the flexible flow handling among heterogeneous access networks. For the actual control of these capabilities 3GPP is in the process to define a new network function, namely the ATSSS, which is

responsible for the management of the different UE flows over the available access technologies [6]. The three main operations supported by the ATSSS are:

- Access Traffic Steering: The procedure that selects an access network for a new data flow and transfers the traffic of this data flow over the selected access network.
- Access Traffic Switching: The procedure that moves all traffic of an ongoing data flow from one access network to another in a way that maintains the continuity of the data flow.
- Access Traffic Splitting: The procedure that splits the traffic of a data flow across multiple access networks. When traffic splitting is applied to a data flow, some traffic of the data flow is transferred via one access and some other traffic of the same data flow is transferred via another access system.

All the three operations are applicable between 3GPP and non- 3GPP access. The ATSSS is distributed over different network entities, such as the UE, the UDR, the SMF and the PCF as shown in Figure 2.

The ATSSS logical architecture consists of the following components:

- UDR-AT3SF: The User Data Repository for Access Traffic Steering Switching and Splitting Function holds UE's subscription data.
- PC-AT3SF: The Policy Control Access Traffic Steering Switching and Splitting Function defines the ATSSS policies according to: a) the application-specific information provided by the AF, b) access information/notification provided by the AMF, c) UE subscription and user profiles provided by the UDR-AT3SF and d) network local policy. The PC-AT3SF may also take input from the NWDAF into consideration to generate or modify its policies.
- CP-AT3SF: The Control Plane Access Traffic Steering Switching and Splitting Function is the main control plane of the ATSSS architecture. It is responsible for ATSSS policy enforcement and session management of all data sessions.
- UP-AT3SF: The User Plane Access Traffic Steering Switching and Splitting Function is the user plane anchor point for all ATSSS traffic and presents a single IP address towards data networks. It is responsible for ATSSS policy rule enforcement in the user plane of the core network.
- UE-AT3SF: The UE Access Traffic Steering Switching and Splitting Function is the policy rule enforcement at the UE level for UE-initiated uplink traffic. It may also generate traffic reports to be sent to the CP-AT3SF.

From the above analysis, it becomes clear that ATSSS's functional elements will be responsible to select the access technology per each active UE data flow. 3GPP is currently defining the context parameters, on which these decisions will be made, as well as the negotiation details between the UE and the network, as far as the mapping between the data flows and the available access network resources are concerned.
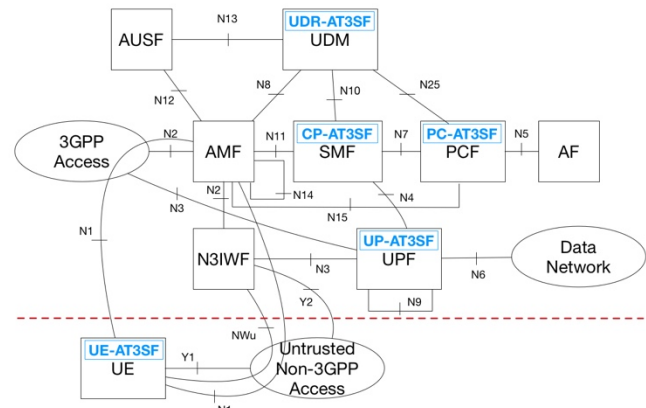

Figure 3. The ATSSS architecture

As it will be shown in detail in Section IV, the novel framework we describe in our paper is fully compatible with this latest architectural status of 3GPP.

III. STATE OF THE ART ANALYSIS

In this section, we present a state-of-the-art analysis of research proposals for a) data analytics and b) RAT selection and traffic steering in 5G networks.

A. Data analytics solutions for 5G networks

Lately, several researchers are investigating how data analytics can assist in improving the performance of 5G networks. The authors in [9] present a unified data model based on random matrix theory and machine learning. They also provide examples on how big data analytics can be used to estimate the location, select an appropriate waveform, discover coverage holes, characterize user plane traffic etc.

The work in [10] presents a generic extraction and correlation framework that reduces the collected data set through randomization and a coarse preservation of statistical relationship among data records. This scheme is very interesting but as the authors state, it remains an open issue how to filter out unrelated information so as to minimize the collection of huge amounts of data that is not really useful. We address this point with the CIP scheme that is presented in a later section.

The results of the SELFNET project are presented in [11]. The target of the authors was to provide an autonomic network management framework for 5G mobile networks. Thus, the focus lies on the analyzer module that infers data from a set of collected metrics. The paper presents in detail the operation of the analyzer, but it does not provide a detailed discussion on the information that needs to be collected to support specific 5G use cases.

Another proposal to use data analytics for network management is presented in [12]. The authors present a new framework, named Big Data SON (BSON). It takes into consideration data at a UE level (i.e., network related performance such as throughput, delay, blocking and drop

rates etc.), cell related data (Reference Signal Received Power and/or Quality - RSRP/RSRQ, of serving and neighboring cells, number of active users etc.) and core network data (alarms, configuration, security data, Call data records etc.). By applying data mining schemes upon the collected information, it is possible to transform reactive network management mechanisms into proactive ones. The paper provides an exhaustive list of information that can potentially be used. However, it does not analyze the traffic volume that needs to be collected in order for the scheme to perform the desired results. Also, they provide operational details only for a simple scenario.

The authors of [13] present the user and mobile network data that can be potentially collected and processed by a data mining scheme. They also discuss how resource management, planning, interference coordination and cache server deployment is done nowadays. They suggest that these will be optimized if data analytics mechanisms are adopted by network operators. The paper discusses the main operation principles of their framework but there are no detailed examples.

The authors of [14] present a mechanism that analyzes past information related to the load of wireless network in order to optimize the allocation of radio resources in the future. Interestingly enough, the results are based on the analysis of data collected from a commercially deployed mobile network. Another interesting idea is presented in [15]. It is proposed to use data analytics to optimize the performance of specific protocols in the radio access network. For example, the authors suggest that it is possible to reduce, for example, the overhead in Radio Header Compression in PDCP, or minimize the required signaling during the execution of a handover.

In [16], the authors introduce a solution that analyzes application layer information to identify the exact model of an end device. The authors of [17] suggest that the examination of Call Detail Records (CDR), collected from the core network, can be used to estimate moving patterns of people inside a city.

Finally, the work in [18] discusses how data analytics can be used to understand the relationship between users and services. More specifically, this approach requires to obtain real-time information about the application the users are using, analyze the information and predict the users' preferences and expectations. Then, the network resources are managed accordingly to provide the required Quality of Experience (QoE) to the end users. Based on this solution, the UEs should collect a lot of information and transfer it regularly to the network for further processing. This requires a lot of processing power in the UEs and it may affect their battery level consumption. More importantly, the exchange of a significant amount of information over the wireless link, by a large number of UEs, may cause performance issues to the real data the UEs are exchanging.

In the following sections, we will briefly present how CEPE has been designed to operate and create a dynamic profile for end users. Based on this profile several control functions can optimize their performance [5]. Also, we introduce CIP that minimizes the collection of information needed by the CEPE framework. The key idea is that information is collected and transmitted to the CEPE engine only when the UE behavior deviates from the expected one as already captured by the user profile.

B. RAT selection and Traffic Steering solutions

The challenge of optimizing traffic steering and RAT selection in heterogeneous and multi-cell layer network environments has been studied thoroughly during the past years.

The work in [19] proposes a cognitive framework, which operates on the side of the UE and targets to create a knowledge base using clustering and reinforcement learning techniques in order to create effective policies for the different user/network states. The authors present specific benefits in terms of balanced cell load; however, the evaluation is based on only 2 Long Term Evolution (LTE) nodes and 3 client nodes, which makes the model a bit simplistic. In addition, the authors restrict their model only in the downlink, and assume that there is no interference.

In [20], the authors attempt to optimize the data flow – Air Interface Variants (AIVs) mapping via dynamic traffic steering on the PDCP layer. The traffic steering framework resides in the Radio Access Network (RAN) and translates the AIV-agnostic QoS metrics to AIV-specific ones based on real-time feedback from the AIVs. Although it presents a very advanced evaluation scenario using small cells and 1000 UEs, it does not describe in detail the context information and/or the algorithmic approach that has been followed.

In [21], the authors focus on the energy perspective and propose an energy-sustainable traffic steering framework, which adjusts dynamically the traffic load to match the base stations' energy distributions in both the spatial and the temporal domains. Moreover, they identify several research challenges related to service capability with energy harvesting, low complexity operation, trade-off between system and user performance, etc. They present a comprehensive case study with realistic configurations and show a considerable optimization from the energy consumption perspective. However, no context information is analyzed and no network-related Key Performance Indicators (KPIs) are thoroughly discussed.

The work in [22] presents a context-aware framework, which uses a multi-attribute decision making scheme for optimal RAT selection in an ultra-dense network. The authors categorize the decision criteria between network- and user- oriented ones. The network-oriented criteria include the received signal strength and the communication data rate. The user-related criteria include the type of the application, jitter, packet loss and delay. In the performance evaluation of the proposed scheme, they target at the minimization of the number of handovers, while they also focus on the ranking abnormality, i.e., the condition to investigate the ranking order of the access network due to the inclusion or exclusion of a RAT. Nevertheless, the evaluation is carried out based on numerous assumptions (e.g., the results are analyzed at thirty pre-defined decision points), without deploying a realistic network environment.

In [23] the authors attempt to identify the impact of different aspects of network models, including network topology and bandwidth allocation of the performance of several RAT selection algorithms. They carry out a detailed survey of different solutions, categorizing them in reinforcement learning-based algorithms, regret matching-based, RAT selection games-based, partially/fully distributed algorithms, etc. They conclude that the performance of adaptive RAT selection algorithms is most significantly affected by the number of base stations that a user can connect to. However, they do not use any realistic/simulated environment to carry out the evaluation, they provide no details in relation to the actual evaluation environment, while the comparison is based only on "numerical tests".

Last but not least, in [24], the authors present a comprehensive survey of the main traffic steering and handover decision criteria in the literature. According to this work, some researchers focus on evaluating the RSRP/RSRQ, the user location or speed, the mobility patterns, the battery level, the mean UE transmit power and the UE power consumption, the load of the cell and the service type. Apart from the case of RSRP, typically researchers are using multiple criteria (e.g., battery lifetime, traffic type, cell load, speed) and are using different tools (e.g., cost-based functions, fuzzy logic, etc.) to reach a decision.

As it can be inferred from the above analysis, numerous efforts have attempted to provide a solution towards the heterogeneous RAT selection in dense network deployments. However, all of the reviewed papers make numerous assumptions, either on the design and architectural aspect of the solution (e.g., in relation to the available interfaces), or on the actual evaluation of the mechanism (e.g., by not implementing vital parts of the signaling procedures, simplistic network environment, etc.).

To the best of our knowledge, there is no previous work, which makes a comprehensive discussion on the specific context information to be acquired and processed and the burden this information imposes on the network. Furthermore, the work presented in the following sections combines complementary functionalities (i.e., data analytics with RAT selection schemes) and maps them into the latest 5G architecture focusing on how to acquire contextual information and minimize any related control signaling exchange.

IV. MAPPING THE PROPOSED FRAMEWORK INTO THE 5G ARCHITECTURE

In this section, we briefly describe the key points of CEPE and Compass presented in [4] and [5]. Furthermore, we present a novel mechanism, named CIP that aims to minimize the amount of contextual information that needs to be collected for CEPE's operations. Also, we provide a detailed description about how these schemes can interoperate and map their functionality into the latest 5G architecture.

A. CEPE: A Context Extraction and Profiling Engine

The work presented in [5], is able to automatically build a user profile that can be used to predict the future behavior of a subscriber. This information is used to improve the performance of network control operations. More specifically, static and dynamic information is collected about:

1. User profile related information (static): gender, device type characteristics (e.g., cpu, memory, os, device type)
2. UE and device dynamic characteristics (dynamic): location, transmission power, amount of transmitted and received data, experienced delay, loss of packets, associated cells identifiers
3. Network related measurements (static and dynamic): type of cell (e.g., macro, femto, etc.), power transmission level, available resource blocks, amount of transmitted and received, data, delay, packet loss, number of connected devices

Based on these measurements, data analytics are used to build a knowledge model, the outcome of which is essentially a dynamic profile for end users. This profile predicts their future behavior based on their location, time and day, the battery level of their devices and their monetary charging status. This way the network can use this information to place users to the appropriate cells and radio access technologies during the execution of a handover or a new session establishment. Also, this information is used by the end devices to select the most suitable cell to camp on, when they are in an idle state. Extensive simulations demonstrate significant performance improvements both for the network operator as well the end users [5]. Note here that this is exactly the information (i.e., the dynamic profile of users that captures their future behavior in terms of mobility and service consumption), that the newly introduced NWDAF can report to the PCF and other network components to improve the performance of a 5G network. Thus, CEPE is essentially a potential future evolution of NWDAF. The introduction of NWDAF in the 5G architecture and its interfacing with the PCF (Figure 4) clearly indicates that in the future, its output will be used to select the most appropriate policies for UEs or types of UEs (e.g., high/low moving terminals, terminals involved in high/low data rate exchange, etc.).
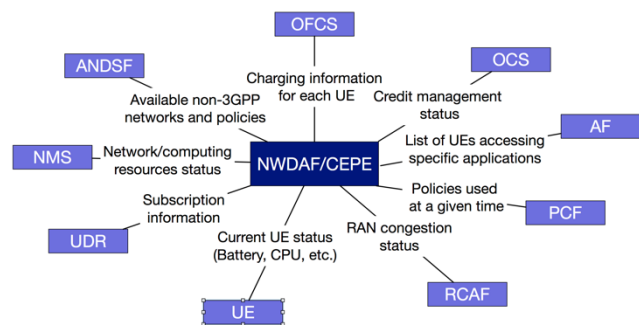

Figure 4. Provision of information to NWDAF

In the state of art research efforts (e.g., [17]), it has been identified that users tend to have the same behavior that is dependent on the terminal they use, their monetary charging status and preferences, the status of their battery and obviously their location (e.g., home office, on the road, etc.) during specific dates and hours. Their behavior also depends on the status of network components (e.g., the load of the network, the received signal strength, the experienced delay or packets losses). All this information is required for a data mining function like NWDAF/CEPE to create a realistic model and enable the selection of appropriate policies or fine tune network control functions. In Figure 3, the network components that have to feed information to the NDWAF/CEPE as well as the type of this information are illustrated. These 5G network components can potentially be UDR, PCF, RCAF and UE. Additional entities not included in Figure 1 are:

- Network Management System (NMS): reports performance indicators (e.g., bandwidth utilization, packet loss, latency, alerts, etc.).
- Access Network Discovery and Selection Function (ANDSF): reports the current policy rules shared with a UE to help it decide, to which available WiFi it may connect to.
- Offline Charging System (OFCS): passes information contained in CDRs, related to the resource usage of a UE.
- Online Charging System (OCS): informs about the current credit management status of a UE.

Based on this information, the NWDAF/CEPE can collect information that is related to the current behavior of users and create their dynamic profile that essentially is a prediction of their future actions. Table I presents some examples of behavioral profiles that can characterize a single user or a set of users that have the same behavior. To create such a list, the NWDAF/CEPE requires mainly information from the UE, the OFCS the OCS, the UDR and the AF (optionally). It also requires information from the NMS and the RCAF.

At the same time, the NWDAF/CEPE should have the information about the current policies being used by the operator as it receives the related information from the PCF and ANDSF. This way it is able to correlate the received input and identify a) the best policies to be used, b) the estimated bandwidth required for a future period of time in an area and c) what is the optimum placement of UEs in cells and radio access technologies. This is doable since the network is aware of the type of users in an area, their number (from those that are connected or those that have recently performed a location update process) and the available capacity of the network. Thus, NWDAF/CEPE can provide rules to PCF in the form of

- Profile (Home C) ^ # of users (high) ^ Network load (high) → place users in femto cells
- Profile (On the road) ^ # of users (any) ^ Network load (any) → place users in macro cells

The behavioral profiles of users can be communicated to any network functions that are responsible for managing the user mobility or the establishment and management of user sessions. Such entities are AMF, SMF, TSSF and 5G Base Stations (gNBs).

B. CIP: A Context Information Preprocessing module

One of the most critical aspects of network context-based systems is the vast volumes of data, that need to be aggregated and transmitted towards a specific network analytics function deployed in the core network. The CIP mechanism performs context data aggregation, filtering and preprocessing techniques. The module comprises aggregating and compressing mobile network-related context information, as well as techniques related to identifying and discarding user profile-redundant or unnecessary context data, before any transmission to the network data analytics function, i.e., CEPE.

TABLE I. EXAMPLES OF BEHAVIORAL PROFILES

Profile Type	Location	Day	Time	Battery Status	Charging Status	Service	Consumption level	Mobility	Network Status
Home A	Home	Mon-Fri	19:00-21:00	High	Credits available	Voice calls	Frequent and long duration	Static	Any status
Home B	Home	Mon-Fri	19:00-21:00	low	Low credits	Voice calls	Infrequent short calls	Static	Any Status
Home C	Home	Mon-Fri	21:00-24:00	High	Credits available	Video streaming	High data consumption	Static	No load
Home D	Home	Mon-Fri	24:00-08:00	Any status	Any status	No activity	No activity	Static	Any Status
Office A	Office	Mon-Fri	09:00-18:00	High	Credits available	Voice calls	Frequent and medium duration	Low mobility	Low-Medium Load
Office B	Office	Mon-Fri	09:00-18:00	low	Credits available	Voice calls	Infrequent and short duration	Low mobility	Any Status
On the road	!Home && !Office	Mon-Fri	Any time	Any status	Any status	Voice calls	Infrequent and short duration	High Mobility	Low-Medium Load

CIP operates in a distributed manner, is deployed on every context-generating network entity (UE, gNB, UPF, etc.) and performs the preprocessing locally, on the respective entity, before transmitting it "upwards" towards the core network. Each CIP aggregates context information per a unique user identifier and in the aggregated information it identifies redundant or unnecessary data and discards it. The aforementioned unique identifier may be the user IMSI (International Mobile Subscriber Identity) or any other unique identifier for the user, so as to indicate that this context information refers to a certain user. Any information item, acquired at a specific network entity is in line with the already generated user profile (in other words, any user/device behavior, in line with the predicted, profiled behavior by CEPE) is considered redundant and is discarded, as it provides no additional gain to the knowledge extraction information base. Nevertheless, we define a *Consistency Index* (CI), which is utilized by the system as a metric of the consistency of the latest generated profiles towards the next phases of the system training. Each time context information is acquired and identified as consistent with the existing profiles, CI is updated accordingly for the specific profile.

Figure 5 details the CIP process, when deployed in a network entity acting as contextual information source.
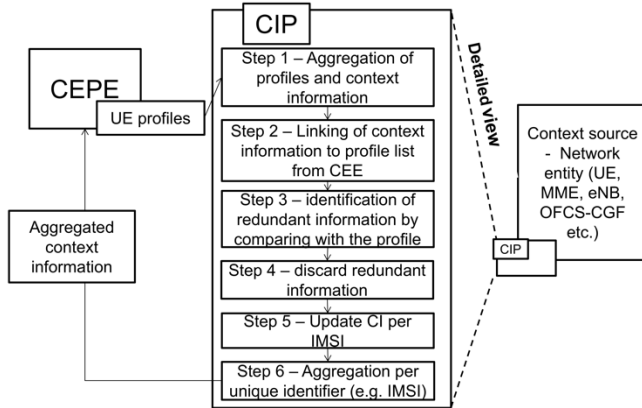
Figure 5. CIP operation

The steps it follows are:

1) Aggregation of profiles and context information: via the respective interfaces, context information is pulled/ pushed to CIP.

2) Linking of context information with the profiles provided by CEPE: for each connected IMSI, CIP makes an association with the respective profile that has already been acquired by CEPE and stored in CIP.

3) Context information comparison with the respective profile so as to identify any redundant information. If the information item that is evaluated is consistent with the respective profile, the item is identified as redundant.

4) Redundant information removal: Redundant information is discarded since it offers no additional insight to the CEPE data-mining phase.

5) Consistency Index update: every time an information item is identified as redundant, the CI is updated accordingly

6) Aggregation per unique identifier (e.g., IMSI): the filtered context information items are aggregated per identifier in order to be transmitted to CEPE.

Each user may have more than one behavioral profiles related to specific context information items (e.g., location, time periods, date, battery level, etc.); at the same time, each CIP entity may have more than one user profiles that refer to one or more users, the CIP when it has to provide context information to the CEPE (Step 1 – Figure 5), then it links the context information with the respective user behavioral profile (Step 2 – Figure 5) for the identification of the active behavioral profile. This implies that the CIP will map the context information to be transmitted with the predicted behavior of the user under certain preconditions that refer to the process that produced the context information. Then, by comparing the context information of the user, which is the actual behavior of the user with his predicted one (represented by the user behavioral profile), the CIP identifies the redundant information (Step 3 - Figure 5). Specifically, when a user complies with his predicted behavior, this information is considered redundant and may be discarded (Step 4 - Figure 5). When the user does not

comply with his predicted behavior, then this information is not redundant and has to be collected by the CEPE so as to be considered in the user behavioral profile extraction process from the "User Behavioral Profiles Extraction Function". Each time redundant data is discarded, a consistency index (CI) is updated for the associated user identifier. Only not redundant data and the Consistency Index are transmitted to the CEPE either directly or through other CIPs (Step 5 & 6 – Figure 5).

As it will be shown in the Evaluation section, CIP's operation becomes of utmost importance in large scale scenarios, with thousands of coexisting UEs and base stations. The context information escalates proportionally to the number of the users, while simultaneously, the RAT selection processing becomes even more complicated, as more UEs contend for the available attachment points.

C. Compass: A UE-oriented RAT selection scheme

The third component of the proposed framework is Compass, a light-weight RAT selection mechanism, which operates on the UE-side. Compass aggregates and processes context information in the vicinity of the UE using Fuzzy Logic. Compass' objective is to construct a Suitability list of the available RAT and cell layer options.

Compass assesses the user preferences along with five different types of context information items (Figure 6), available in the network, i.e., the Received Signal Strength or RSRQ (for non-3GPP and 3GPP networks respectively), the base station's load in terms of utilized capacity, as well as the respective load of the backhaul, the UE mobility type (static UE, low/high mobility, etc.), and finally the traffic flow characteristics (critical delay requirements, etc.). In order to avoid any signaling overhead to the network, the afore-mentioned information items have been carefully selected, as they are already available either in the UE or received from the network (e.g., in the form of broadcast messages). This approach considerably minimizes the signaling overhead of the proposed scheme.
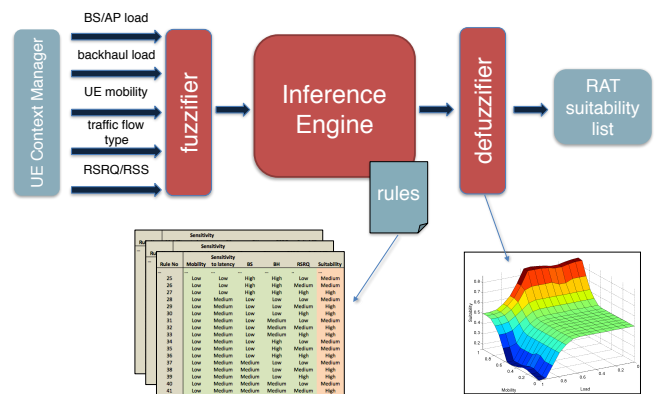

Figure 6. Compass Fuzzy Inference System

Although initially designed as a standalone scheme, Compass fits nicely with CEPE and CIP as it kicks in when

the dynamic profile of user is not yet formed or when the expected behavior of a user deviates from what is already captured by the existing profile.

D. CEPE – CIP – Compass interworking

This section highlights the complementary nature of the network-operated CEPE with the UE-oriented Compass and discusses how the assets and shortcomings of each one can provide a holistic context-based traffic steering and RAT selection framework. CIP –on the other hand- operates -with both Compass and CEPE- in a distributed manner in all involved network entities independently, towards the preprocessing of the context information before its transmission; thus, in this section we will mainly focus on CEPE and Compass.

As already mentioned earlier, the core network-based CEPE with the UE-controlled Compass are able to operate independently –and with a different approach- with regard to the RAT selection and traffic steering policies. Compass operates on the UE side, and -based on a multi-criteria fuzzy logic based scheme- calculates the suitability of the available RATs and cells/APs for each one of its active sessions (applications/services/etc.) in almost real-time. On the other hand, CEPE operates as standalone Core Network function/entity, which builds upon historical context information and patterns in relation to the UEs, the network, the consumed applications and services, etc. and generates User Profiles, on which, prediction for future situations is based.

The two mechanisms target the same objective (applying optimized radio resource management and traffic steering) from divergent perspectives, which however, are complementary and can be combined. The complementarity of the two schemes is federated by the fact, that -as happens in the majority of context-based mechanisms-, the benefits have always to confront numerous shortcomings and limitations as well.

Compass operation provides knowledge of the current situation of the network status – in the vicinity of the UE- and, based on the respective service requirements, it triggers a session setup (or a handover in case of an on-going session) to the most appropriate access technology and specific cell/AP; besides, Compass minimizes unnecessary execution of control functions based on the mobility of the user, the requirements of a particular service/application, as well as the load of the network.

Nevertheless, certain limitations refrain Compass from being capable of globally supporting the overall traffic steering and RAT selection operations in a holistic manner. The UE-assisted nature of the specific scheme, results in a solution, which on the one hand optimizes the resources provided for specific UEs, however, on the other hand, lacks any capability of an overall planning or overview of the status of the network, leading requests from the UE side impossible to address in very demanding situations, such as dense deployments. Novel traffic engineering approaches in forthcoming, challenging 5G environments will require solutions, which will act also partially from the UE side for

an efficient, real-time network probing, however, the final network-side decision making will be of utmost importance.

Furthermore, Compass acts in a reactive manner; the context acquired and processed by Compass refers to recent real-time behavior of the UE; this translates to slow convergence in cases of diverse UE behaviors in small time frames, consecutive calculations from the UE side every time the context is modified (such as mobility changes), etc. Last but not least, a context-based scheme on the UE side –no matter its efficiency- implies additional signaling information, -specifically for ultra-dense environments, where hundreds/thousands of UEs coexist among numerous base stations/Aps -, which makes the solution inefficient in terms of energy consumption – one of the most crucial aspects of UE-based solutions-, even if several energy-related optimizations have taken place in Compass design and implementation.

Contrary to Compass' reactive nature, CEPE, on the other side, is a proactive Core Network based entity/function, which aggregates context information related both to the UEs (device characteristics, behavior profile, app usage, etc.), as well as the network status, and builds prediction models and user profiles, based on patterns, which are identified in this aggregated context over a certain amount of time. The network traffic demands prediction is directly linked to one of the strongest CEPE advantages: via the holistic picture of the network over long-time frames: this provides the network administrator a framework of utmost importance, facilitating traffic engineering operations, load balancing, etc.

As already highlighted, CEPE's longer term context processing differs considerably from Compass's, which applies real-time monitoring and decision-making. One the one hand, this provides the ability to predict user future behavior, in terms of both mobility, as well as service usage; on the other hand, however, this operation implies certain weaknesses for specific/examples: a new UE connects to the network, which CEPE is not trained for, or there is a deviation from the existing profile (a UE enters a new area, in which CEPE is still not trained or an event occurs (e.g., new type of application/service launched), which deviates considerably from the profile that CEPE has built for the particular UE, etc.). Another shortcoming identified relates to the limited ability of CEPE operation to perform micro monitoring in real-time and identify the specific cell/AP, which should be selected for a specific UE –in case of multiple options-. This information is however acquired by Compass, which scans when required the UE's environment in real-time; thus, the two schemes may as well combine context information towards the optimal selection.

To sum up, certain capabilities and shortcomings of the operation of the two, primary context-based schemes (i.e. CEPE and Compass) prove that a solution, which combines their operation in a parallel and coordinated manner, would provide considerable gains (Figure 7).

- Compass operates as long as no valid user profile is currently available
- CEPE provides the main policies for UE to RAT mapping and traffic steering to the relevant network

entities, which are responsible for forwarding them down to the UEs.

- Whenever, an event occurs, which shows that a UE behavior deviates from CEPE's profiles (e.g., unexpected use of a service different mobility or location for a UE, etc.), Compass' resumes its operation and the current profile is ignored. In this case CIP will update CEPE with new and detailed information.
- Last but not least, CEPE can be exploited to fine-tune the load information request rate of Compass to ANDSF, and/or other relevant context-aware components (not shown in the figure).
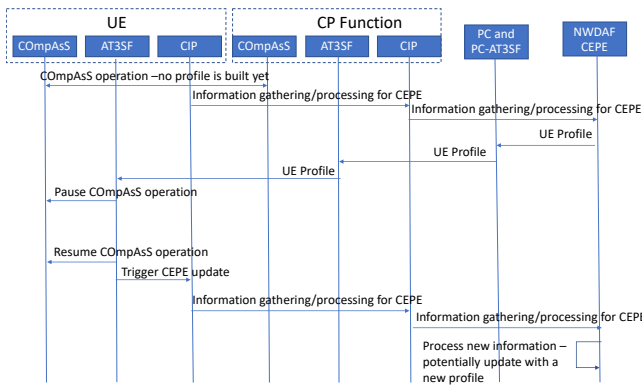

Figure 7. CEPE – Compass interworking MSC

E. The three mechanisms in 3GPP's 5G architecture

As already mentioned in the first Section of this paper, the three mechanisms we present in this work are designed in such a way, to be fully compatible with the 3GPP 5G architecture. In Section II we presented in detail the latest status in 3GPP with regard to the 5G system architecture, along with the primary system components.

One of the novel core network functions, which are introduced in [6] is a dedicated data analytics function (NWDAF). CEPE, as a network data analytics and prediction engine is essentially a module, capable of supporting fully this functionality by serving as NWDAF instance in the forthcoming 5G architecture. Apparently, different data mining engines may be deployed by different operators, focusing on specific context information items or particular verticals/business scenarios. The diverse information items aggregated and processed by CEPE have been analyzed in the respective previous section. Mapping this information, with the afore- presented 5G architecture, several 5G network components, which generate valuable context information, will forward it to CEPE/NWDAF: Unified Data Repository, Access Network Discovery and Selection Function (ANDSF), PCF, RAN Congestion Awareness Function (RCAF), etc.

Moreover, the Radio Congestion Awareness Function (RCAF) is also a vital component for Compass context

information acquisition related to one of its five core context parameters, i.e. the real-time AP/cell load information. Compass will be receiving information related to the potential congestions related to specific APs/cells –besides the ANDSF, which has already been described in previous section- and will calculate the respective Suitability of each AP/cell.

With regard to the ATSSS, UDR-AT3SF holds UE ATSSS subscription data for operator service and user profiles. This functional element is directly linked with the output of NWDAF-CEPE profiles, which receives and provides them –as will be described below- to other ATSSS functional elements. UE-AT3SF implements ATSSS policy rule enforcement at the UE for UE-initiated traffic (UL). It may also generate traffic reports to be sent to the CP-AT3SF. The traffic usage reports from UE-AT3SF are directly linked to Compass operation and fine-tuning role, described in the previous section regarding the CEPE-Compass complementarity.

Thus, it becomes clear that ATSSS' functional elements will be the responsible component/function, which will select the access technology per each active UE data flow. 3GPP is currently defining the context parameters, on which these decisions will be made, as well as the negotiation details between the UE and the network, as far as the mapping between the data flows and the available access network resources are concerned. Compass, as already described in the respective sections, could operate as an ATSSS instance for this UE flow – RAT mapping optimization, providing optimal selection with an energy- and signaling-efficient approach. CEPE-NWDAF and Compass are capable of providing this context information to the ATSSS elements. The following figure illustrates the proposed architecture, which integrates CEPE, Compass (and CIP) and maps them with the NWDA and ATSSS functionalities.
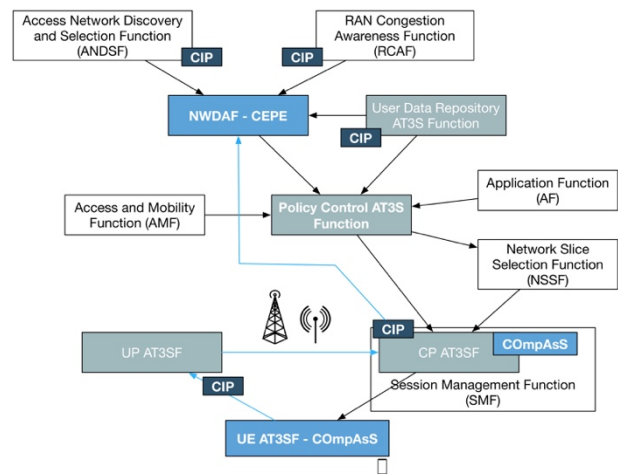

Figure 8. Logical interfaces of CEPE (as NWDAF instance) and Compass (as UE-AT3SF instance) in the 5G architecture

To sum up, according to the above analysis, the Policy Control Access Traffic Steering Switching and Splitting Function (PC-AT3SF) defines ATSSS policies according to

the application-specific information provided by the Application Function (AF), access information/notification provided by the Access and Mobility Function (AMF), and UE ATSSS subscription and user profiles provided by the User Data Repository ATSSS (UDR-AT3SF). In our proposed architecture (Figure 8) this information, –on which the ATSSS policies are generated- is federated by the NWDAF-CEPE function and is also linked to the NSSF, via the PC-AT3SF. Compass module is considered part of the UE-AT3SF instance- and is responsible for sending traffic usage and access technologies Suitability reports to UP-AT3SF, which relays them "upwards" via CP-AT3SF ultimately to NWDAF-CEPE for feedback-based fine-tuning of policies. A Compass instance is additionally deployed on the CP-AT3S function, in order to coordinate the scheme's reporting with the Control Plane.

V. CIP EVALUATION AND OVERALL SYSTEM OVERHEAD QUANTIFICATION USING AN ANALYTICAL APPROACH

For the analysis of CIP, we consider an ultra-dense network with a high number of base stations (macro, as well femto cells) and UEs within a geographically limited area. We are assuming a network area serving a million active UEs. We assume that the operator is performing profile-based data analytics using CEPE. Various parameters related to context information should be aggregated from numerous network entities, each of which is characterized by a different payload as well. For this reason, we provide Table II that shows the payload that we used in our evaluation schema; for each context item we also provide a reference, in which detailed insights are presented.

TABLE II. PAYLOAD PER CONTEXT INFORMATION PARAMETER

Parameter	Payload	Source Network Entity	Reference
IMSI	64 bits	gNB, MME	[25]
Cell ID (+MCC+ MNC)	32 bits	gNB, MME	[26]
Timestamp (TS)	32 bits	All	[27]
Mobility (Mobility State Estimation based on Handover counters as in LTE)	8 bits	MME	[28]
User Charging Data Record (CDR)	64 KB to 100 MB	OFCS (Offline Charging System) – CGF	[29]
Contract type (in SPR – Subscription Profile Repository)	4 bits	PCRF – PGW	[30]
Charging status (enough credit/no credit)	1 bit	Online Charging System (OCS) – PGW	[31]
UE Battery Status (normal/low)	2 bits	UE	[32]

The signaling overhead estimation depends on the network entities where CIP and CEPE are deployed as the

number of hops and the interface types between the source and destination of the context information transmission varies. In the evaluation example, the key assumption for CEPE and CIP is illustrated in the following figure (Figure 9).
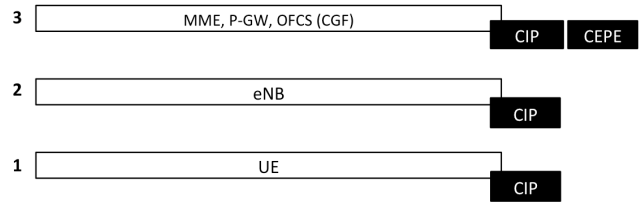

Figure 9. CEPE and CIP deployment in the network during the evaluation

The particular example (Figure 9) shows that the CIP_{UE} and the CIP_{gNB} are one hop away in terms on direct, implemented network interfaces, the CIP_{UE} and the CEPE two logical hops, etc. A similar scenario may of course be applied in the same way for N hops between the network entities, in case CEPE and CIP are deployed in different entities as well.

We compare the proposed approach against two mechanisms by quantifying the overhead induced to the basic daily signaling due to their application (Table III).

TABLE III. JUXTAPOSED SCHEMES

Proposed scheme	Pre-processing	Redundant information identification/discard
Standard	None – all information is aggregated to be transmitted	None
Semi-optimized	Basic pre-filtering (pre-aggregation per IMSI and location)	None

The format of the transmitted context information comprises all the required information items, as discussed earlier and is described as follows:

{(IMSI, location, TS), Mobility State Estimation (MSE) / Bearer ID / contract type / charging status / battery status}

The IMSI-location-TS triplet comprises a key and is always transmitted along with one or more of the remaining context parameters (e.g., MSE, bearer ID, etc.). We define equation (1) in order to calculate the overall signaling cost imposed by the transmission of the afore-described context information parameters.

$$O = \sum_{i=1}^{N} \sum_{j=1}^{M} [(1 - P_c)x(S(I_m)xFr(I_m) + S_{cdr}) + S(I_{static}) + S_{cc}] \quad (1)$$

O is the overall signaling overhead, N the number of users connected to the system, M the number of the different profiles/behaviors that the UE exhibits, P_c the consistency percentage that describes the portion of the acquired context information, which is consistent with the existing user

profile, $S(I_m)$ the payload of each one of the mth context parameter {TS / MSE / contract type / charging status / UE battery status}, $Fr(I_m)$ the transmission frequency of each one of the mth parameter {TS / MSE / CDR / contract type / charging status}, S_{cdr} the message of the varying CDR file sizes (1/10/100 MB), $S(I_{static})$ the payload of the static tuple {IMSI, location}, which equals (64 + 32 bits = 96 bits) and S_{cc} the size of the Consistency Counter.
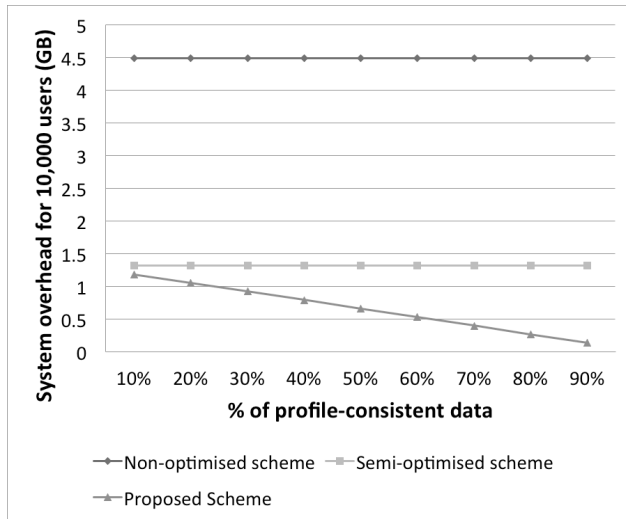

Figure 10. System overhead with varying profile consistency

The evaluation that follows illustrates the system overhead imposed during the context data acquisition and for varying scenarios of UE profile consistency. We assume that the consistency percentage - P_c - varies from 10 up to 90%, as shown in the first chart (Figure 10). The second evaluation chart (Figure 11) illustrates the gain of the proposed scheme –CIP- against the two afore-mentioned standard mechanisms. As it can be seen on the X-axis of Figure 11, we attempt to assess the CIP's gain for different cases, in terms of the CDR file sizes, which often may vary; in our evaluation we use varying sizes, from very small to considerably large: 1, 10 or 100 MB according to respective references [32]. For the evaluation, we consider some of the most challenging 5G use cases (Figure 12), which follow the METIS project specifications [33] and correspond to varying number of UEs in the network environment; traffic jam, shopping mall, stadium, open-air festival are some indicative examples. The results of the initial 3-fold evaluation that was described above are illustrated in the following charts.

As depicted in the graphs, the gains of applying CIP during the context information acquisition and transmission between the different network entities may vary from small to huge. The most important parameter that influences the extent of the signaling overhead minimization is the data consistency in terms of the already available user profiles.

In other words, as one would expect, the more consistent the context data is with the CEPE-produced user profiles, the more the overhead approaches zero when CIP is applied (Figure 10, Figure 11).
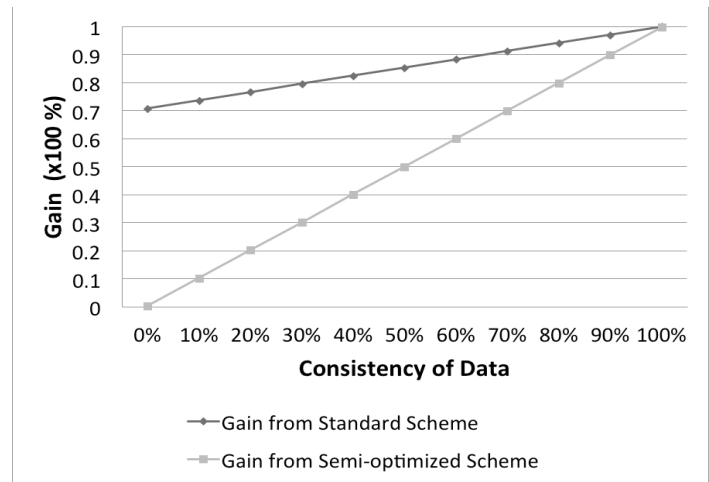

Figure 11. CIP's gain over the two selected schemes

In all cases, the semi-optimized scheme performs better already than the standard scheme, which transmits all the information towards the Knowledge Extraction engine (CEPE), while both of these schemes our outperformed by CIP.

As already discussed earlier, the number of devices to be managed by the network is one of the most challenging 5G use cases. Figure 12 illustrates the great gain that results from the application of CIP, particularly in selected, established use cases, as these have been identified and thoroughly defined in one of the first 5G EU projects, i.e. FP7 METIS 2020 [33], such as an Open Air Festival, or a Dense Urban Information Society, where the numbers of users peak.

VI. CONCLUSIONS

In this paper, a context based framework towards radio resource management was introduced, comprised of three mechanisms: Compass, CEPE and CIP. Each one of the individual mechanisms performs a complementary operation towards optimizing the RAT selection and traffic steering and switching operations in 5G network environments. The core policy extraction component, CEPE, is based upon extensive data analytics on user- and network- related information in order to generate UE profiles, while it is federated by the real-time operated, UE-oriented Compass scheme; CIP performs context information pre-processing and filtering towards the reduction of the transmitted information among the network entities.

The second part of this work presented the latest 5G system architecture, as this has been described so far by 3GPP and performed a comprehensive mapping between the former and the proposed framework.

Future directions of this work will focus on –but will not be limited to- the real deployment and evaluation of the proposed system, using customized, open source software and components, based on the Open Air Interface [34] system, as well as custom Ettus Research-based Universal Software Radio Peripheral (USRP) [35] nodes for the hardware (access) part.
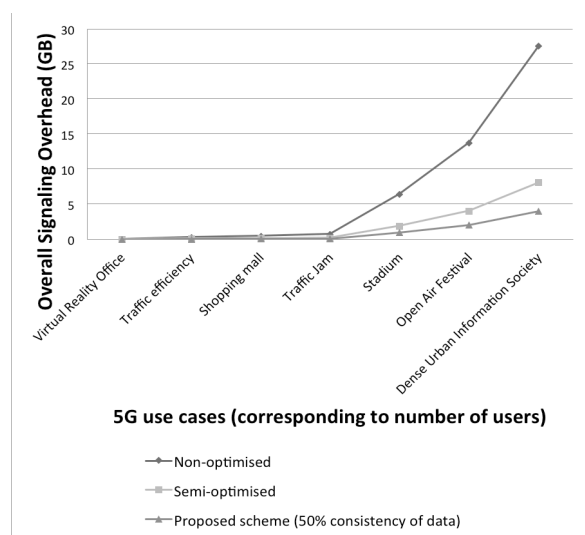
Figure 12. Signaling overhead for different
5G use cases

REFERENCES

[1] A. Kaloxylos, "Application of Data Mining in the 5G Network Architecture", the Thirteen International Conference on Digital Communications, ICDT 2018, Athens, Greece, March 2018.

[2] M. Gramaglia, A. Kaloxylos, P. Spapis, et. al, "E2E Architecture, in 5G System Design: Architectural and Functional Consideration and Long Term Research", ISBN 978-1-119-42512-0, Wiley, pp. 81-112, June 2018.

[3] K.R. Rao, Z.S. Bojkovic, and B.M. Bakmaz, "Network Selection in heterogeneous environment: a step toward always best connected and served", in Telsiks 2013, October 2013, pp. 83-92.

[4] S. Barmpounakis, A. Kaloxylos, P. Spapis., and N. Alonistioti, "Context-aware, user-driven, network-controlled RAT selection for 5G networks", Computer Networks, Elsevier, Vol. 113, pp. 124-147, 2017.

[5] P. Magdalinos, S. Barmpounakis, P. Spapis, A. Kaloxylos, G. Kyprianidis, A. Kousaridas, et al., "A context extraction and profiling engine for 5G network resource mapping", Computer Communications, Elsevier, Vol. 109, pp. 184-201, 2017.

[6] 3GPP TS 23.501, "System Architecture for the 5G System", LTE Rel. 15, v 15.1.0, March 2018.

[7] 3GPP TS 23.793, "Study on Access Traffic Steering, Switching and Splitting support in the 5G system architecture", Rel 16, v. 0.4, April 2018.

[8] 3GPP TS 23.402, "Architecture enhancements for non-3GPP accesses (Release 15)", June 2017.

[9] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, and R. Qiu, "Big Data Analytics in Mobile Cellular Networks", IEEE Access, Vol. 4, 2016.

[10] S. Robitsch, F. Zaman, S. van der Meer, J. Keeney, and G. Muntean, "Magnet: Real-Time Trace Stream Analytics Framework for 5G Operations Support", IEEE Network Magazine September/October 2017.

[11] L. I. B. Lopez, J. M. Vidal, and L. J. G. Villalba, "An Approach to Data Analysis in 5G Networks", Entropy Vol. 19, Issue 2, 2017.

[12] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G", IEEE Network Magazine, November/December 2014.

[13] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big Data-Driven Optimization for Mobile Networks toward 5G", IEEE Network Magazine, January/February 2016.

[14] Z. Feng, X. Li, Q. Zhang, and W. Li, "Proactive Radio Resource Optimization with Margin Prediction: A Data Mining Approach", IEEE Transactions on Vehicular Technology, Vol. 66, Issue 10, October 2017.

[15] S. Han, C. Lin, G. Li, S. Wang, and Q. Sun, "Big Data Enabled Mobile Network Design for 5G and Beyond", IEEE Communications Magazine, Vol 55, Issue 9, July 2017.

[16] J. Liu, F. Liu, and N. Ansari, "Monitoring and Analyzing Big Traffic Data of a Large-Scale Cellular Network with Hadoop", IEEE Network Magazine, July/August 2017.

[17] V. Kolar et. al., "People In Motion: Spatio-temporal Analytics on Call Detail Records", Sixth International Conference on Communication Systems and Networks (COMSNETS), 2014.

[18] Y. Wang et. al., "A Data-Driven Architecture for Personalized QoE Management in 5G Wireless Networks", IEEE Wireless Communications, Vol. 24, Issue 1, February 2017.

[19] J. S. Perez, S. K. Jayaweera, and S. Lane, "Machine learning aided cognitive RAT selection for 5G heterogeneous networks", IEEE International Black Sea Conference on Communications and Networking, June 2017.

[20] A. Prasad, F. S. Moya, M. Ericson, R. Fantini, and O. Bulakci, "Enabling RAN moderation and dynamic traffic steering in 5G", IEEE 84th Vehicular Technology Conference, September 2016.

[21] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and Z. Shen, "Energy-sustainable traffic steering for 5G mobile networks", IEEE Communications Magazine, Vol. 55, pp 54-60, November 2017.

[22] A. Habbal, S. I. Goudar, and S. Hassan, "Context-aware Radio Access Technology selection in 5G Ultra Dense Networks", IEEE Access, Vol.5, April 2017.

[23] D. D. Nguyen, H. X. Nguyen, and L. B. White, "Performance of adaptive RAT selection algorithms in 5G heterogeneous wireless networks", International Telecommunication Networks and Applications Conference (ITNAC), December 2016.

[24] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms", Communications Surveys & Tutorials, IEEE, vol. 16, iss. 1, pp. 64-91,1st Quarter 2014.

[25] 3GPP "Current Capacity of IMSI, MDISDN and IPv6 Identifiers", April 2010.

[26] 3GPP TS 23.003, "Numbering, addressing and identification", v. 13.3.0, Rel.13, September 2013.

[27] http://www.unixtimestamp.com/, Epoch Unix Time Stamp Converter (accessed November 2018).

[28] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification", v. 12.7.0, Rel.12, September 2015.

[29] "eLTE2.2", Offline Charging Whitepaper, Huawei Technologies, February 2014.

[30] 3GPP TR 23.803, "Technical Specification Group Services and System Aspects: Evolution of policy control and charging", v. 7.0.0, Rel. 7, October 2010.

[31] 3GPP TS 29.212, "Policy and Charging Control (PCC)", v.11.6.0, Rel.11, October 2010.

[32] 3GPP TS 25.331, "Radio Resource Control (RRC); Protocol Specification", v. 13.0.0, Rel.13, September 2015.

[33] The METIS 2020 Project – Laying the foundation of 5G, www.metis2020.com, (accessed November 2018).

[34] OpenAirInterface, 5G software alliance for democratising wireless innovation, http://www.openairinterface.org , (accessed November 2018).

[35] Ettus Research X310, https://www.ettus.com/product/category/USRP-X-Series (accessed November 2018).

APPENDIX – LIST OF ABBREVIATIONS

5G	5th Generation of Cellular Mobile Communications
3GPP	3rd Generation Partnership Project (3GPP)
RAT	Radio Access Technology
eMBB	Mobile BroadBand
UE	User Equipment
QoS	Quality of Service
CEPE	Context Extraction and Profiling Engine
NWDAF	Network Data Analytics Function
PCF	Policy Control Function
ATSSS	Access Traffic Steering, Switching and Splitting
CIP	Context Information Processing
AN/CN	Access/Core Network
NF	Network Function
UDM	Unified Data Management
ARPF	Authentication Credential Repository and Processing Function
AUSF	Authentication Server Function
AMF	Core Access and Mobility Management Function
SMF	Session Management Function
UP	User Plane
UPF	User Plane Function
NEF	Network Exposure Function
NRF	Network Function Repository Function
RCAF	Radio Congestion Awareness Function
NSSF	Network Slice Selection Function
TSSF	Traffic Steering Support Function
N3IWF	Non-3GPP Inter Working Function
IPsec	Internet Protocol Security
LWA	LTE-WLAN Aggregation
PDCP	Packet Data Convergence Protocol
RCLWI	LTE-WLAN Interworking
NR	New Radio
UDR-AT3SF	User Data Repository for Access Traffic Steering Switching and Splitting Function
PC-AT3SF	Policy Control Access Traffic Steering Switching and Splitting Function
CP-AT3SF	Control Plane Access Traffic Steering Switching and Splitting Function
UP-AT3SF	User Plane Access Traffic Steering Switching and Splitting Function
UE-AT3SF	UE Access Traffic Steering Switching and Splitting Function
BSON	Big Data SON
RSRP/RSRQ	Reference Signal Received Power/Quality
CDR	Call Detail Records
QoE	Quality of Experience
LTE	Long Term Evolution
AIV	Air Interface Variants
RAN	Radio Access Network
KPI	Key Performance Indicator
NMS	Network Management System
ANDSF	Access Network Discovery and Selection Function
OFCS	Offline Charging System
OCS	Online Charging System
IMSI	International Mobile Subscriber Identity
CI	Consistency Index
MSE	Mobility State Estimation