Context-based Resource Management and Slicing for SDN-enabled 5G Smart, Connected Environments

Sokratis N. Barmpounakis*[1]
National and Kapodistrian University of Athens, Greece
Department of Informatics and Telecommunications
sokbar@di.uoa.gr

Abstract. 5G mobile communication systems will address unprecedented demands in terms of system capacity, service latency and number of connected devices. Future 5G network ecosystems will comprise a plethora of 3GPP and non-3GGP Radio Access Technologies (RATs), such as Wi-Fi, 3/4/5G, etc. Deployment scenarios envision a multi-layer combination of macro, micro and femto cells where multi-mode end devices, supporting diverse applications, are served by different technologies. This thesis focuses on the radio resource management (RRM) from the perspective of the primary RAT and cell layer selection processes (i.e., cell (re)selection, handover, admission control); afterwards, it goes one step beyond, in order to link the RRM with Network Slicing approaches, as introduced in Software Defined Networking (SDN)-enabled environments, which creates smaller, virtual “portions” of the network, adapted and optimized for specific services/requirements. Towards this end, this thesis introduces a context-based radio resource management (RRM) framework, comprised of three distinct mechanisms: Two out of the three mechanisms exploit contextual information, while the third mechanism acts with an augmenting role to the former two, by pre-processing the contextual information required by such, context-based mechanisms and –thus- by limiting the signaling cost required for communicating this contextual information among network entities. Last but not least, a comprehensive analysis has taken place in relation to the architectural aspects, in the context of the forthcoming 5G network architecture and by mapping them with the latest 5G network components – as these were introduced in the latest 3GPP work-.

Keywords: 5G, Radio Resource Management, RAT selection, User Profiling, Handover, Context, SDN, Network Slicing.

1 Dissertation Summary

In the following years, the number of wireless and mobile devices is expected to increase considerably. Along with it, a huge increase of mobile traffic [1] will also take place. More specifically, the mobile traffic in 2016 was nearly 30 times the size of the entire global Internet in 2000. Almost half a billion mobile devices and connections were added in 2016, while at the same time, average smartphone usage grew 20% in

* Dissertation Advisor: Athanasia (Nancy) Alonistioti, Assistant Professor

the same year. In addition, the actual traffic volume per subscriber increases 25-40% per year, thus exceeding the expectations set by ITU [2]. The deployment of 5G cellular networks targets to support this vast number of devices, while at the same time existing 3GPP specifications will keep on supporting legacy cellular access networks (e.g., GSM, HSPA, LTE, LTE-A), as well as alternative radio access technologies (e.g., Wi-Fi). In this environment, the end users will have access to a diverse set of services (high definition video and audio, web browsing, games, etc.). It is worth pointing out in parallel that the high penetration of smartphones and tablets on the market [3] will enable end users to make use of all these services while on the move.

5G networks are expected to support billions of small end devices (e.g., sensors, actuators, etc.) as well as communicating vehicles [4] in the context of Machine Type Communication (MTC). The vision is that 5G networks will manage to materialize the Internet of Things (IoT) ecosystem. This realization will unveil new requirements to the network operators and telecom manufacturers. The aforementioned requirements should also be taken into consideration by the underlying network infrastructure. Overall, existing mechanisms used for the communication of end terminals are inadequate to support the future needs. Towards improving efficiency, well- established mechanisms have to be redesigned. One of the most important areas 5G networks have to improve is the mapping of smart devices and services to different RATs and layers (i.e., macro, micro, pico, femto cells). This mapping affects the Key Performance Indicators (KPIs) of a network in relation to the experienced grade of service (e.g., blocking probabilities, throughput, delay, jitter, etc.). The placement of a UE to a RAT or a cell layer, that is either in idle or a connected mode, is primarily realized via three vital RRM mechanisms, namely: a) Cell-(re)Selection (CS), b) Access/Admission Control (AC) and c) Handover (HO).

The contributions of this dissertation move towards the following major directions:

. a) COmpAsS, a user-oriented context-based scheme for RAT selection and traffic steering/switching, which processes context in real-time and produces a RAT suitability list to be used for handover management reasons,

. b) CEPE, a knowledge extraction engine based on data mining techniques towards user profiling and network policies formulation,

. c) CIP, a context information pre-processing scheme, which acts in an augmenting manner to the former two, in order to minimize the high context acquisition signaling overhead.

. d) A study on CEPE-COmpAsS interworking in an SDN-enabled, 5G architecture, capable of applying network slicing approaches.

Supplementary contributions, which enforce the research carried out in the aforementioned primary four directions and also comprise parts of the next steps to be made in the context of this research domain are:

- the architectural perspective of the proposed schemes, which takes into account the latest 3GPP standardization guidelines and attempts to provide a valid and viable solution towards the forthcoming 5G architecture,
- a study on network traffic engineering policies, which can exploit CEPE and user profiling methodology is included,

- an attempt to describe from a common point of view 3 primary RRM mechanisms, i.e. cell (re)selection, handover and call admission control via a comprehensive categorization of the existing approaches, both from the academic area, as well as from the industry, by incorporating the available patents as well, and
- a 5G use case application related to IoT and Precision Farming, which highlights specific requirements related to industrial applications, ultra-low delay requirements, etc.
-

The first major contribution of this thesis is COmpAsS, a context-aware, user-oriented RAT selection mechanism, which operates on the User Equipment (UE) side and ultimately produces a list of the most suitable RATs per active traffic flow/session, towards QoS optimization. One of the greatest advantages of the UE-based solution is the minimization signaling overhead over the air interface, as well as the computation load on the base stations. COmpAsS collects information related to the network status, such as the load of the base stations, the load of the backhaul link, the Reference Received Signal Quality (RSRQ), user mobility information, such as the velocity of the UE, as well as the specific QoS requirements of the type of traffic to be transmitted, in order to assess -in real-time- the most suitable RAT and/or cell layer, which should serve the UE's active sessions. COmpAsS mechanism adopts Fuzzy Logic (FL) as one the core logic modules, responsible for the perception of the network situation and, in combination with a set of pre-defined rules, calculates a list of the most suitable available access network options. Furthermore, we propose an evolution of 3GPP's Access Network Discovery and Selection (ANDSF) function, as one of the primary Evolved Packet Core (EPC) network functions collaborating with COmpAsS for the exchange of the required context information. The merits of COmpAsS are showcased via an extensive series of simulation scenarios, as part of 5G ultra dense networks (UDN) use cases. The results prove how the proposed mechanism optimises Key Performance Indicators (KPIs), when juxtaposed to a well-established LTE handover algorithm.

The second major contribution of the current thesis the Context Extraction and Profiling Engine (CEPE), a resource management framework, which collects diverse types of context information and performs data mining techniques in order to extract meaningful knowledge. The context information, which is aggregated, primarily relates to four categories: network operation data, user behavior information, terminal capabilities and application/service data. CEPE analyzes this information, extracts meaningful knowledge and performs user profiling in order to apply it for optimal resource planning, as well as prediction of resource requirements. CEPE collects information about users, services, terminals and network conditions and –based on offline processing– derives a knowledge model, which is subsequently used for the optimization of the primary RRM mechanisms, i.e. handover, cell selection and call admission control. From a methodological point of view, initially the KPIs that will be employed are identified in order to assess the efficiency of the mechanism. Next, the types of data that should be monitored are identified (network operation data, user behaviour information, etc.). Then, the extracted context information is translated into user profiles and is finally applied as input for enhanced cell (re)selection, handover or admission control. CEPE's operation is tightly connected to the scheme, which follows, CIP, and focuses on the pre-processing of the vast amount of information, which is collected, towards minimizing the signaling overhead. The viability and validity of CEPE is demonstrated via an extensive set of simulation scenarios.

The third major contribution is CIP, a Context Information Pre-processing scheme, aiming to identify and discard redundant or unnecessary data before knowledge extraction. CIP could be considered as an integral part of the afore described profiling schemes, i.e. COmpAsS and CEPE. CIP comprises a framework that primarily relies upon data aggregation and pre-processing techniques. Context information processing and knowledge extraction is considered a great tool towards the optimisation of several network functions; nevertheless, the acquisition of the context is often a very costly process –in terms of signaling burden imposed on the network. The module comprises aggregating and compressing mobile network-related context information per unique identifier, such as the end device's International Mobile Subscriber Identity (IMSI), as well as techniques related to identifying and discarding user profile-redundant or unnecessary context data, before any transmission to CEPE. CIP gains are illustrated via a detailed analytical approach, guided by well-established 5G use case requirements.
The fourth major contribution of this thesis is a mapping of the proposed scheme in a Software Defined Networking-enabled 5G architecture, as proposed by the latest 3GPP standardization, capable of applying Network Slicing approaches for further optimizing the network resources distribution and sharing and addressing the challenging 5G use cases, such as massive IoT.

2 Results and Discussion

2.1 COmpAsS experimental evaluation

The performance of COmpAsS is demonstrated via a series of simulation scenarios. In this section, 3 rounds of experiments will be presented, along with the respective assumptions, topologies and outcomes. All simulations were carried out using the open-source NS-3 discrete-event network simulator (versions NS-3.19 and NS-3.23).

Experiment 1: Simplified Shopping Mall use case.

The first scenario presents a simplified version of a shopping mall test case; two rows of femto cells (assuming they are inside the mall's shops) and a pedestrian corridor in the middle, while macro cells (LTE eNBs) co-exist at a distance of around 1km which is a typical range for an urban – suburban location (**Fig. 1**). In our simulation scenarios, we included 2 eNBs and 5 HeNBs. For simplicity's sake, no WiFi APs were used in the scenario, as the large-small cell handling evaluation of our mechanism was achieved by using LTE macro and femto cells. It is furthermore assumed that inside the mall area, several UEs are either static or moving at pedestrian's speeds, i.e. 0 – 1.5 m/s. These UEs –being attached to the mall's HeNBs- contribute as well in the creation of the load that needs to be taken into account for selecting the appropriate RAT from UE.

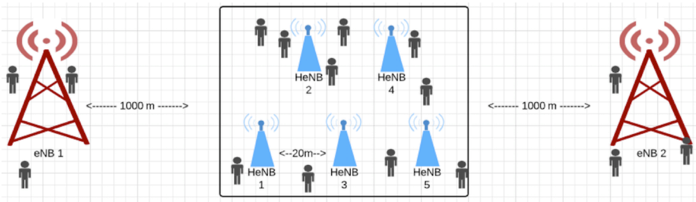


Fig. 1 Simplified Shopping mall use case

The simulations were made using three moving UEs at pedestrian speed: low, medium and high. In order to evaluate the performance of the algorithms, we increased the load of two out of the five overall HeNBs gradually, reaching from zero load to very high load. The background traffic in all rest (H)eNBs causes them to be in a medium load state during simulation time. By load, we are referring to both Load and Backhaul Load –as they were presented in the previous sub-sections-, which is calculated by the number of the rest static UEs, associated to each (H)eNB, as well as the number of bearers per UE. The KPIs, which are assessed, are the overall number of handovers that took place, the average throughput as well as the average experienced delay for the three moving UEs.

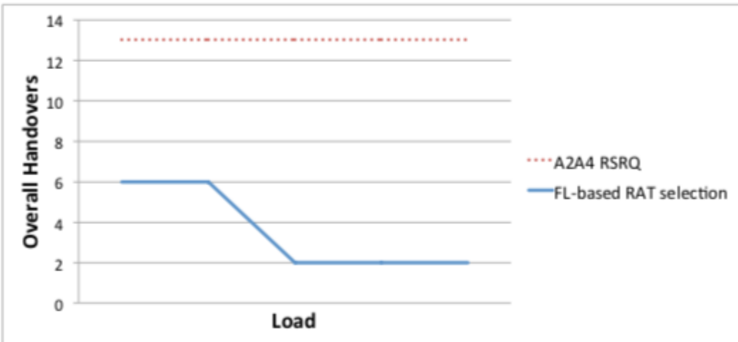


Fig. 2 Overall number of handovers

The above figure illustrates the performance of the two algorithms in terms of the overall number of handovers that took place from the 3 UEs. Noticeably, the A2A4 RSRQ algorithm's decisions are not influenced neither by the higher mobility of the UEs, nor by the increasing load of the HeNBs. On the contrary, the proposed mechanism tends to minimize handovers in the afore-mentioned cases. When the load is low, the number of handovers is reduced by 53.8% -due to reduced number of executed handovers of the high mobile UEs-, while when the load increases, the overall handovers are reduced by 84.6% since the suitability factor of candidate (H)eNB is low.

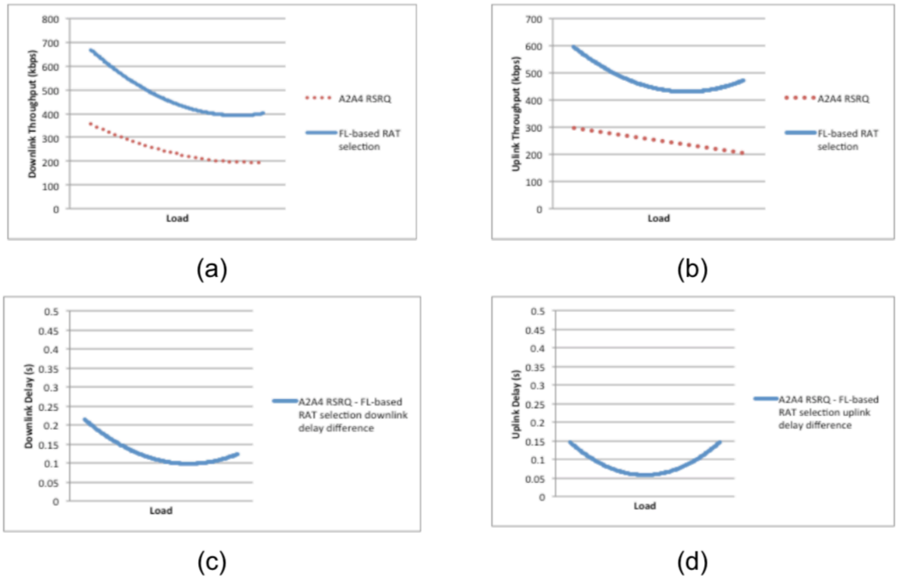


Fig. 3 (a) Downlink Throughput, (b) Uplink Throughput, (c) Downlink Delay, (d) Uplink Delay

Fig. 3 illustrates the performance of the algorithms with regard to the calculated average throughput both in the downlink and the uplink. The FL-based RAT selection outperforms the A2A4 RSRQ in terms of throughput; both in the downlink, as well as the uplink case by an offset of 300 kbps roughly as load increases. An interesting observation in the case of the proposed mechanism's performance is the increase of the throughput when the load of the HeNBs is extremely high; this advantage results from the fact that as load increases the suitability of the femto cells is constantly decreasing, tending to retain the UE from doing a handover to them. As a result, throughput and delay performance are directly related to the handover decisions presented earlier.

Similarly, the difference in the delay (measured as A2A4 RSRQ average packet delay minus the FL-based average packet delay) remains positive in all scenarios. Once more, as load increases radically, the FL-based mechanism tends to increase its performance since the UE keeps its connection to medium loaded eNB.

Experiment 2: Realistic Shopping Mall use case.

The 2nd experiment presents a realistic business case scenario of a shopping mall comprising 3 floors (ground floor, 1st and 2nd floor), and 20 shops per floor (Figure 31). The UEs are either static or moving, and are roaming around the shopping mall rooms (shops, cafes, etc.). Several HeNBs are deployed in the three floors. In addition, two macro cells (eNBs) exist outside the mall area in a distance of 200m to different directions. Due to the fact that COmpAsS handles Wi-Fi APs and HeNBs in a similar way, with regard to the pre-defined rules of the Fuzzy Inference Engine, for the sake of simplicity, in the simulations only macro and femto-cells are deployed.

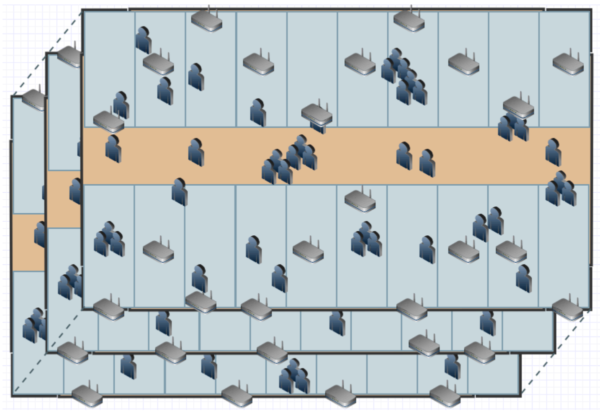
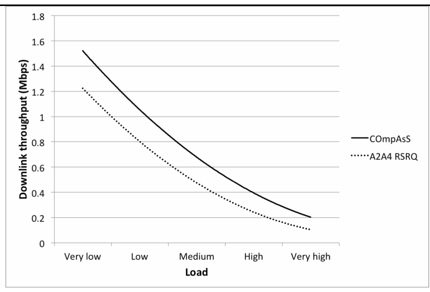


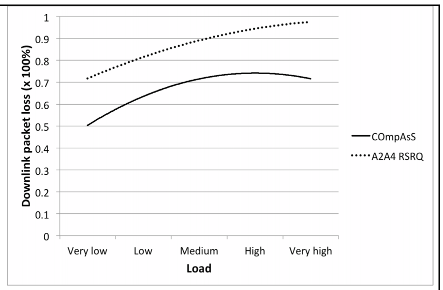
Fig. 4 Experiment 2: Shopping mall with 3 floors and 20 shops per floor

Besides the several UEs, which are roaming inside the mall area and creating respective traffic to the HeNBs, we use one "test UE", in which COmpAsS is deployed. Different simulations were carried out to test the UE at different velocities (low, medium, high), in each one of the scenarios in order to evaluate the proposed scheme for varying UE mobility, as mobility is one of the inputs, which are taken into consideration for the decision. The test UE is moving with linear velocity between the rows of the shops, on the 1st floor.

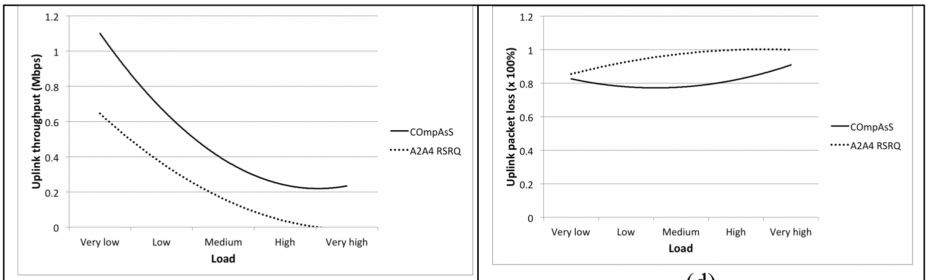
Indicatively, in the following figures we illustrate some of the measured KPIs, which resulted from the two mechanisms with regard to the number of overall handovers which took place during the simulation, the throughput of the test UE, the experienced delays, as well as the packet loss during the measurements. Variable load of the femtocells of the shopping mall was tested, calculated in relation to the overall associated users per base station and traffic that is generated. In particular, the load of the base stations varies from 10% up to 90% of their available resources (horizontal axis).



(a)



(b)



(c) (d)

Fig. 5 Experiment 2 results

Experiment 3: Advanced Shopping Mall use case.

The final COmpAsS evaluation experiment is the most advanced one, among the three, which are included in this evaluation section. We evaluate this 3rd experiment on the basis of 4 different scenarios, which we describe in detail below. Overall, the network deployment allows seamless handling of services across different domains, e.g. mobile/fixed network operators, real estate/shop owners and application providers. The environment is similar to Experiment 2, presented earlier. Each floor's dimensions are 200x100m, containing 20 rooms/shops per floor, with several LTE Femto cell placed on each floors, depending on the scenario. Outside, two LTE eNBs are placed, 150m north and west of the mall respectively.

The proposed framework's algorithm uses two parameters, i.e. *Suitability Threshold and Hysteresis*. Different parameter values may alter radically COmpAsS's responsiveness and functionality, primarily in terms of triggering events frequency. Different network "states" (e.g., denser or scarcer deployments) would require different configurations of these two control parameters. Towards this fine-tuning process, hence, in the first two scenarios, we incorporate in our experimentation a range of values, both for Threshold and Hysteresis. Overall, the evaluation of COmpAsS moves along 4 axes-scenarios, each one of which focuses on a different varying parameter of the experiment's setup, in order to simulate -in the most realistic extent possible- all the radio conditions and network "states" that the proposed framework may encounter.

The following scenarios were evaluated:

- Per suitability threshold
- Per suitability margin
- Per deployment density (number of femto cells)
- Per network load (number of traffic bearers/UE)

For each one of the 4 scenarios, we ran 75 similar experiments in order to maximize the validity of our experimentation results. In order to define the number of runs per sub- scenario, as well as the experiment duration, we initially carried out some test scenarios; each one of the different runs incorporates a random generation of mobility patterns for the UEs, as well as slightly varying traffic models. We defined our confidence level at 95% in order to be able to demonstrate a satisfactory and statistically valid outcome. Indicatively, we provide the results for the 4th scenario (i.e. per network load):

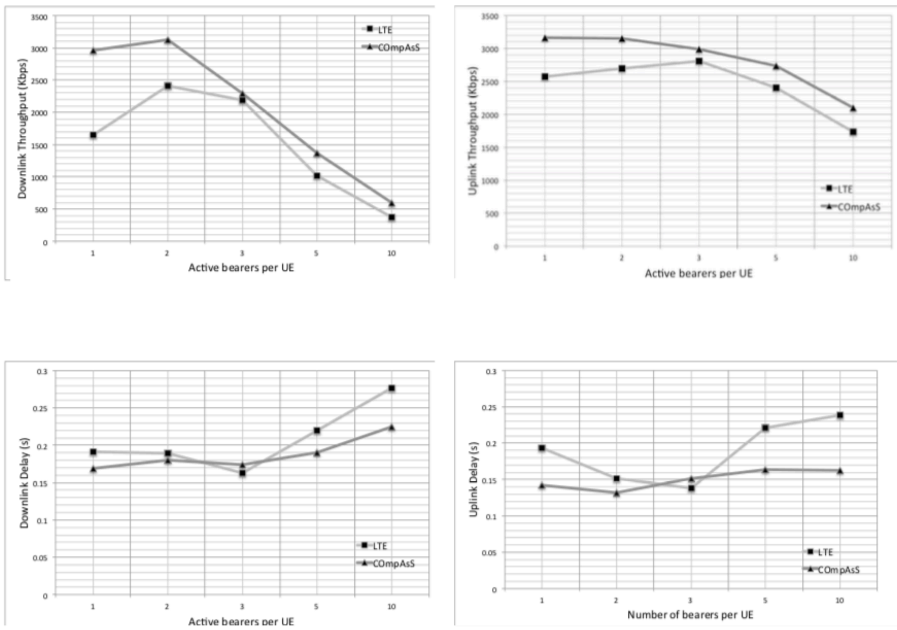


Fig. 6 Experiment 3 indicative results

2.2 CEPE experimental evaluation

The evaluation of CEPE followed the same simulation environment and principles with the earlier COmpAsS's evaluation steps. Indicatively, below the performance evaluation is presented in relation to the type of the Radio Access Technology, associated with the measured UE.

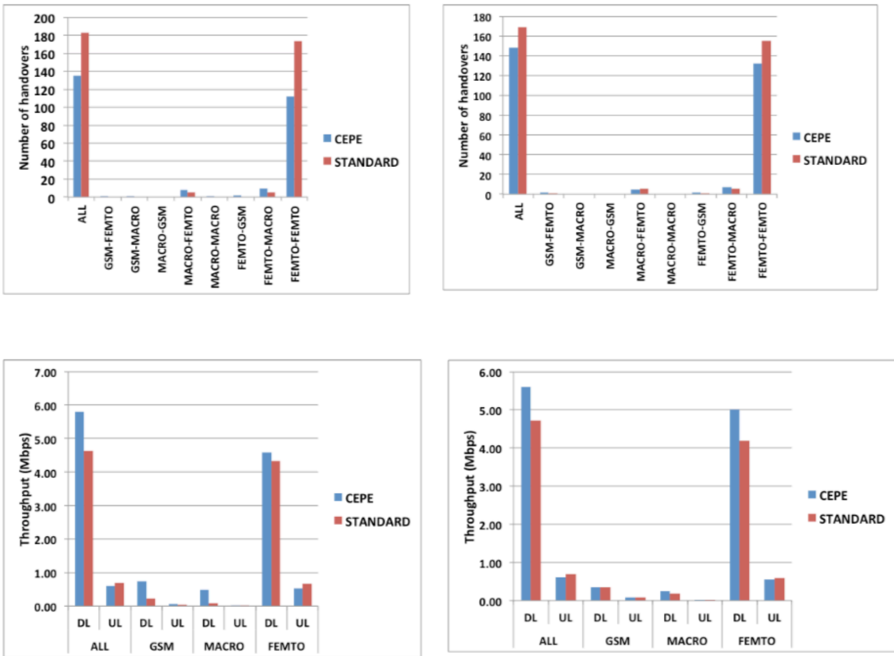


Fig. 7 CEPE evaluation results per type of RAT

3 Conclusions

Context awareness comes at a cost. The more information is acquired and processed, the higher the granularity of the context awareness, however the larger the burden, which is placed –both on the network, as well as the computing entities-; one of the crucial topics, which was analyzed in the context of this thesis was the signaling overhead evaluation for each one of the proposed mechanisms, as well as the type of the information acquired, which should be within the available information items and network entities, and in line with the latest standardization efforts towards 5G.

In the context of this thesis, the focus has been placed on three distinct context-aware mechanisms, which attempt to address the resource scarcity issues, which will be faced in the forthcoming ultra dense network deployments. All three mechanisms, focus on a different aspect of the context-aware approach: COmpAsS is a UE-based scheme, which attempts to select the most suitable radio access technology and point of access for the UE; CEPE, is an offline profiling engine, which is used to generate user and device profiles, and –based on these profiles- optimize the resource allocation- attempting to map in an optimal manner the mapping between the available resources and user/device profiles; CIP –the 3rd core mechanism- is a context information pre-processing and filtering scheme, which acts in a complementary manner to the aforementioned schemes, targeting to minimize the signaling and processing burden, which is posed by the diverse and numerous context information items –especially in dense 5G deployments- with massive number of users, devices and coexisting access technologies.

This thesis provided a holistic study, which comprised comprehensive analyses from diverse aspects: architectural, algorithmic, experimental, analytical, etc. for all mechanisms. Besides the evaluation of the integrity and validity of each one of the three core schemes, a novel architecture is proposed, in line with the latest 3GPP standardization steps, comprising COmpAsS, CEPE and CIP as instances of the proposed novel network entities of the new 5G-EPC.

One of the crucial matters, on which focus was given, was the context acquisition process. In order to design COmpAsS's, as well as CEPE's system parameters, a comprehensive analysis on the network resources, respective interfaces and context information item types was made. In addition, an analytical approach was presented in the case of COmpAsS, which provided detailed insights on the information items, which are used, along with the signaling overhead required to aggregate them. To the best of our knowledge, there is no previous work, which attempts to quantify the signaling overhead of the proposed context-based mechanism, and juxtapose it with the gains measured in the network-related KPIs part.

The validity of each one of the mechanisms was showcased via an extensive set of experimental scenarios, carried out in line with the 5G Ultra Dense network scenarios and requirements, in realistic simulated topologies and with diverse access technologies and layers (macro cells, femto cells, Wi-Fi APs, etc.). The flexibility of the open source NS3 simulator, -which was used throughout all the experimentation-, enabled us to customize our environment according to very specific requirements, and –thus- achieve the realistic models we targeted.

This extensive demonstration proved a number of gains with regard to primary network KPIs, such as the maximization of the achieved throughput, the minimization of unnecessary handovers, as well the reduction of the latency measurements, particularly for delay-critical services, as described in the 5G verticals' requirements. The performance of the proposed schemes was juxtaposed to well established handover and RAT selection mechanisms, already deployed in 4G/LTE. This comparison highlighted numerous outcomes, both as far as the network and QoS KPIs are concerned (throughput, latency, packet loss, etc.), as well as the signaling overhead evaluation, since we compared with baseline –already deployed in the market- solutions, and not theoretical solutions found in the literature.

References

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021, url: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile- white-paper-c11-520862.html, last accessed 2017/07/11.
2. ITU report, ICT Facts and Figures 2016, url: http://www.itu.int/en/ITU- D/Statistics/Documents/facts/ICTFactsFigures2016.pdf, last accessed 2017/07/11.

3. Henry Blodget, The Number Of Smartphones In Use Is About To Pass The Number Of PCs, https://www.businessinsider.com.au/number-of-smartphones-tablets-pcs-2013-12, last accessed 2017/07/11.
4. 5G Vision Brochure, 5G-PPP, The 5G Infrastructure Public Private Partnership: The next generation of communication networks and services.