
**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCE**
**DEPARTMENT OF INFORMATICS AND TELECOMMUNICATION**


**BSc THESIS**

# Context awareness and related challenges: A comprehensive evaluation study for a context-based RAT selection scheme towards 5G networks


**Nikiforos K. Leonidakis**

**Supervisors:** **Nancy Alonistioti**, Assistant Professor
**Sokratis Barmpounakis**, PHD candidate



**ATHENS**

**MARCH 2017**

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

# Επίγνωση πλαισίου και σχετικές προκλήσεις: Μια εκτενής μελέτη αξιολόγησης για έναν μηχανισμό επιλογής τεχνολογίας ασύρματης πρόσβασης με αξιοποίηση πλαισίου για τα 5G δίκτυα

**Νικηφόρος Κ. Λεωνιδάκης**

**Επιβλέποντες:** **Νάνσυ Αλωνιστιώτη**, Επίκουρος Καθηγητής
**Σωκράτης Μπαρμπουνάκης**, Υποψήφιος Διδάκτορας

**ΑΘΗΝΑ**

**ΜΑΡΤΙΟΣ 2017**

**BSc THESIS**

# Context awareness and related challenges: A comprehensive evaluation study for a context-based RAT selection scheme towards 5G networks

**Nikiforos K. Leonidakis**

**S.N.:** 1115200900085

**Supervisors:** **Nancy Alonistioti**, Assistant Professor
**Sokratis Barmpounakis**, PHD candidate

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

# Επίγνωση πλαισίου και σχετικές προκλήσεις: Μια εκτενής μελέτη αξιολόγησης για έναν μηχανισμό επιλογής τεχνολογίας ασύρματης πρόσβασης με αξιοποίηση πλαισίου για τα 5G δίκτυα

**Νικηφόρος Κ. Λεωνιδάκης**

**Α.Μ.:** 1115200900085

**ΕΠΙΒΛΕΠΟΝΤΕΣ:** **Νάνσυ Αλωνιστιώτη**, Επίκουρος Καθηγητής
**Σωκράτης Μπαρμπουνάκης**, Υποψήφιος Διδάκτορας

# ABSTRACT

Effective network planning is essential to cope with the increasing number of mobile internet subscribers and bandwidth-intensive services competing for limited wireless resources. Additionally, key challenges for the constantly growing LTE networks is increasing capabilities of current mechanisms, reduction of signaling overhead and the utilization of an effective Radio Access Technology (RAT) selection scheme. There have been various proposals in literature regarding these challenges, some of which are discussed here.

The purpose of this work is to research the current advances in LTE networks regarding EPC - WiFi integration and context awareness in mobility management, and propose the COmpAsS algorithm, a mechanism using fuzzy logic to select the most suitable Radio Access Technology. Furthermore, we quantify the signaling overhead of the proposed mechanism by linking it to the current 3GPP specifications and performing a comprehensive analysis. Finally, we evaluate the novel scheme via extensive simulations in a complex and realistic 5G use case, illustrating the clear advantages of our approach in terms of handover frequency and key QoS metrics, i.e. the user-experienced throughput and delay.



**SUBJECT AREA**: Wireless Communications

**KEYWORDS**: LTE, 3GPP, handover, signaling, mobility, ns3

# ΠΕΡΙΛΗΨΗ

Ο αποτελεσματικός σχεδιασμός των δικτύων είναι απαραίτητος για να αντιμετωπιστεί ο αυξανόμενος αριθμός των συνδρομητών κινητού διαδικτύου και των απαιτητικών υπηρεσιών δεδομένων, που ανταγωνίζονται για περιορισμένους ασύρματους πόρους. Επιπλέον, οι βασικές προκλήσεις για τα συνεχώς αναπτυσσόμενα δίκτυα LTE είναι η αύξηση των δυνατοτήτων των υφιστάμενων μηχανισμών, η μείωση της υπερβολικής σηματοδότησης (signaling) και η αξιοποίηση ενός αποτελεσματικού μηχανισμού επιλογής τεχνολογίας ασύρματης πρόσβασης (RAT). Υπάρχουν ποικίλες προτάσεις στην βιβλιογραφία σχετικά με αυτές τις προκλήσεις, μερικές από τις οποίες παρουσιάζονται εδώ.

Ο σκοπός της εργασίας αυτής είναι να ερευνήσει τις τρέχουσες εξελίξεις στα δίκτυα LTE σχετικά με την ενσωμάτωση EPC και WiFi και την επίγνωση πλαισίου (context awareness) στην διαχείριση κινητικότητας, και να προτείνει τον αλγόριθμο COmpAsS, έναν μηχανισμό που χρησιμοποιεί ασαφή λογική (fuzzy logic) για να επιλέξει την πιο κατάλληλη τεχνολογία ασύρματης πρόσβασης για τα κινητά. Επιπλέον, έχουμε ποσοτικοποιήσει το κόστος σηματοδότησης του προτεινόμενου μηχανισμού σε σύνδεση με τις σημερινές προδιαγραφές του 3GPP και εκτελέσαμε μια ολοκληρωμένη ανάλυση. Τέλος, αξιολογήσαμε τον αλγόριθμο μέσω εκτεταμένων προσομοιώσεων σε ένα πολύπλοκο και ρεαλιστικό σενάριο χρήσης 5G, που απεικονίζονται τα σαφή πλεονεκτήματα της προσέγγισής μας όσον αφορά τη συχνότητα μεταπομπών (handover) και τις μετρήσεις βασικών QoS τιμών, όπως ρυθμός μετάδοσης και καθυστέρηση.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ**: Ασύρματες Επικοινωνίες

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ**: LTE δίκτυα, 3GPP, μεταπομπή, σηματοδότηση, κινητικότητα, ns3

## CONTENTS

**PREFACE** ..... **10**

**1. INTRODUCTION** ..... **11**

**2. THE EPC NETWORK** ..... **15**

**2.1 Cellular networks intro** ..... **15**

**2.2 EPC, SAE, LTE** ..... **16**
2.2.1 LTE – SAE ..... 16
2.2.2 EPS Architecture ..... 17
2.2.3 EPC – WiFi integration ..... 18

**2.3 Ultra Dense Networks** ..... **25**

**2.4 Context Awareness in Mobility Management for Cellular Networks** ..... **27**
2.4.1 State of the art – current approaches ..... 27
2.4.2 Challenges – signaling overhead ..... 31

**3. IMPLEMENTATION** ..... **33**

**3.1 COmpAsS mechanism** ..... **33**
3.1.1 The network architecture perspective ..... 33
3.1.2 Overview of the proposed solution ..... 34
3.1.3 Description of the algorithm ..... 35

**3.2 Our signaling analysis** ..... **38**

**4. EVALUATION** ..... **45**

**4.1 Intro – scope of the experiments** ..... **45**

**4.2 The NS-3 simulator** ..... **45**

**4.3 Topology – Scenarios** ..... **46**

**4.4 Results – figures & analysis** ..... **48**

**5. CONCLUSION** ..... **57**

**ABBREVIATIONS - ACRONYMS** ..... **58**

**REFERENCES** ..... **61**

# LIST OF FIGURES

Figure 1: EPS elements overview[18].....18
Figure 2: Evolution to Ultra Dense Networks [20].....25
Figure 3: Fuzzy Logic Controller for the extraction of the RAT Suitability metric.....35
Figure 4: The algorithm of our proposal [59].....36
Figure 5: X2 Signaling Bearer Protocol Stack.....39
Figure 6: COmpAsS Signaling.....44
Figure 7: Simulation environment: Shopping mall comprised of 3 floors and 20 shops per floor.....46
Figure 8: Suitability Threshold Scenario Handovers.....49
Figure 9: Average Handovers per Suitability Threshold.....49
Figure 10: Average Uplink Delay per Suitability Threshold.....50
Figure 11 Suitability Margin Scenario Handovers.....51
Figure 12: Average Handovers per Suitability Margin.....51
Figure 13: Average Uplink Delay per Suitability Margin.....52
Figure 14: Deployment Density Scenario Handovers.....53
Figure 15: Average Handovers per Deployment Density.....53
Figure 16: Average Uplink Delay per Deployment Density.....54
Figure 17: Active Bearers Scenario Handovers.....54
Figure 18: Average Handovers per Active Bearer Scenario.....55
Figure 19: Average Uplink Delay per Active Bearer Scenario.....55
Figure 20: COmpAsS Handover Types.....56

## LIST OF TABLES

Table I $N_F$ x $N_R$ *Suitability* Calculation example for a UE with 4 active IP flows .....35
Table II Mobility management functions .....40
Table III: Signaling per message type .....40
Table IV: BSS load / Backhaul load request .....41
Table V: Backhaul load response .....42
Table VI: BSS Load response .....43
Table VII: NS3 Simulations' Configuration .....47
Table VIII: Scenarios details .....48

# PREFACE

The following thesis “Context awareness and related challenges: A comprehensive evaluation study for a context-based RAT selection scheme towards 5G networks” was completed to fulfill the graduation requirements of Department of Informatics and Telecommunications in National and Kapodistrian University of Athens. It was a subject that I took great interest into as 5G networks is an upcoming highly promising technology that will be vital for enhancing mobile networks, a field in constant need for development. Throughout my work for this thesis I was able to gain extensive knowledge in regards of LTE networks, how they operate and what are the future plans for improvement.

I would like to thank my supervisors for their help and especially Sokratis Barmpounakis for his invaluable assistance and input during the course of this work.

Nikiforos Leonidakis

Athens, 14 March 2017

# 1. INTRODUCTION

Wireless communication networks is a field in need of constant development due to the exponential growth in data traffic on wireless cellular infrastructure in the latest years. It is expected that in the very near future, cellular networks will have to deal with a massive data traffic increase, as well as a vast number of devices. Optimal placement of the end devices to the most suitable access network is expected to provide the best Quality of Service (QoS) experience to the users but also the maximum utilization of the scarce wireless resources by the operators. Effective and efficient network planning is essential to deal with the constantly increasing number of mobile users and bandwidth-intensive services.

Operators have met this challenge by increasing capacity with new radio spectrum, adding multi-antenna techniques and implementing more efficient modulation and coding schemes. However, these measures alone are insufficient in the most crowded environments and at cell edges where performance can significantly degrade. One way to expand an existing macro-network, while maintaining it as a homogeneous network, is to “densify” it by adding more sectors per eNB or deploying more macro-eNBs, leading to the Ultra Dense Networks (UDNs).

According to network traffic data analysis and projections [1], one of the greatest challenges in the forthcoming wireless networks era is that 5G networks will have to cope with a huge increase both in terms of data traffic as well as number of end devices (e.g., smartphones, tablets, sensors etc.). Besides the tremendous growth, which is expected in terms of number of devices, due to an increasingly diverse set of new and yet unforeseen services, users and applications (including machine-to-machine modules, smart cities, industrial automation, etc.), novel and less predictable mobile traffic patterns are also expected to emerge [2].

In order to address this issue the research community designs solutions to improve the spectral efficiency, to increase the network cell density and to exploit the underutilized radio spectrum resources [3]. One of the main trends suggests the exploitation of the available small cells, i.e. primarily femto-cells (Home eNBs) or WiFi Access Points (APs) to efficiently distribute the network load [4],[5] via intelligent dynamic steering of the network traffic. An introduction of the small cells can be through the addition of low-power base stations (eNBs, HeNBs or Relay Nodes (RNs)) or Remote Radio Heads (RRH) to existing macro-eNBs. Site acquisition is easier and cheaper with this equipment, which is also correspondingly smaller. Small cells are primarily added to increase capacity in hot spots with high user demand and to fill in areas not covered by the macro network – both outdoors and indoors. They also improve network performance and service quality by offloading from the large macro-cells. The result is a heterogeneous network with large macro-cells in combination with small cells providing increased bitrates per unit area.

It is envisaged that the aforementioned trend will inevitably result in very dense deployments, in which on the one hand, Long Term Evolution (LTE) base stations (BSs) will co-exist with their 5$^{th}$ generation evolution, while in addition, 3GPP networks will co-exist with the non-3GPP ones (primarily WiFi), creating thus a multi-tier architecture consisting of heterogeneous radio access technologies. Some of the greatest challenges in such dense wireless environments are the efficient inter-working between

the legacy with the latest cellular systems, as well as with WiFi APs, the optimization of the UE placement - RAT selection procedures, as well as the minimization of the unnecessary handovers -and ping-pong effect-related events- between adjacent RATs and cells, which inevitably deteriorate the provided QoS to the users: The handover procedure in the current Evolved Packet Core (EPC)/4G system involves latency overheads, even in limited coverage areas, over the GPRS Tunnelling Protocol (GTP) tunnel [6]. In order to enable seamless UE mobility when moving across the different (H)eNBs, the S-GW (at the network core) communicates with the eNBs (at the network edge) to perform handover management; QoS allocation, traffic condition monitoring, user terminal mobility management and security tasks are also forwarded to the Packet Gateway (P-GW). At the same time, the eNB, the S-GW, and the P-GW perform several signaling procedures to handle the session setup at different levels. Such an approach decreases considerably the network performance by increasing the latency and thereby reducing the QoS required for the future real-time applications. Thus, it becomes of utmost importance that frequent or unnecessary handovers in such ultra-dense network environments are minimized; latency overheads should be minimized and the optimal RAT options for the UEs should be available in an efficient way via a viable RAT selection approach.

Lately, new directions have been presented by 3GPP's specification groups [7] towards the network capacity issue optimization and the so called tight interworking of 3GPP and non-3GPP access technologies, with plenty of these novel directions and standards planned to be integrated in the forthcoming releases. In relation to the efficient interworking between heterogeneous wireless systems (e.g., LTE and WiFi), although during the last decade there has been considerable progress in terms of specifications and standards, still a successful demonstration of a seamless integration of WiFi APs with commercial cellular networks, and in realistic 5G business cases is missing [8]-[13]. This is because of a number of reasons. WiFi suffers from interference issues since it operates on the unlicensed spectrum.

Most importantly however, switching from a cellular network to a WiFi access point has not properly yet evolved to a fully transparent process -from different perspectives-; for the end users the authentication process had to take place manually –thus, deteriorating the QoE-; furthermore, the mobility of multiple flows (even of the same service) among different PDN connections and access technologies was only recently standardized and described [14]. In addition to the first point –and as this is described by Hotspot 2.0-, all the "islands" of hotspots should be also interconnected into larger "footprints" via further roaming agreements between WiFi operators. Finally, there are still diverse strategies in the way non-3GPP networks are handled by different devices and operating systems, meaning that the software that handles the active UE connections (e.g. the "Connectivity Manager" in Android) has not been standardized. In some cases, even the same OS handles differently the connections depending on the version of the OS, e.g. [15].

In parallel with the steps being made in terms of the heterogeneous networks tight integration and interworking, RAT selection optimization for active UE sessions, i.e. handover procedure, is of paramount importance. Due to the fact that the selection of the access technology influences on a great extent both the resource allocation per user, as well as the generated interference among cells, the experienced quality of the respective services may vary greatly. Numerous schemes have been proposed to optimize the RAT selection procedure. These schemes extend from very simple solutions that often do not attempt to acquire a holistic picture of the network environment context (in order to avoid signaling overhead issues) to complex frameworks, which however require major modifications in the core network

components, proposing diverse parameters as inputs towards the assessment of the available RAT choices. So it is of great importance to optimize the tradeoff between context acquisition and signaling overhead and provide a feasible and realistic approach for efficient RAT selection in 5G. In addition to the previous, it must be also highlighted that with latency being one of the most crucial QoS metrics for next generation networks, any mechanism that will be finally deployed –either on the network or the UE side- should demonstrate exceptional performance in terms of algorithm execution/computation delays.

In order to overcome these challenges we propose a novel mechanism, which follows closely the latest 3GPP directions and guidelines and attempts to cover the aforementioned gaps. More specifically:

(a) This work concentrates on the context acquisition process: A comprehensive analysis on the network sources, respective interfaces and context information item types is made. In addition, an analytical approach is presented, which provides detailed insights on the information items, which are used, along with 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, as we will present in subsection 3.1.

(b) The mechanism is extensively evaluated from the performance perspective: The proposed scheme is a lightweight module, the core of which is based on a fuzzy-logic inference system. The validity of our proposal is evaluated via numerous simulation scenarios and diverse traffic flow types, in a realistic 5G configuration and set-up, comprising an ultra-dense heterogeneous network environment.

(c) The novelty of this work is further reinforced by the fact that the proposed scheme is based solely on assumptions in line with the latest standardization efforts in terms of context information acquisition, attempting this way to highlight the realistic and viable aspect of the solution for next generation wireless networks. To the best of our knowledge, no research proposal has attempted to limit its assumptions totally in line with the standardization guidelines; on the contrary, the vast majority of solutions make numerous assumptions, which often lead non-realistic proposals.

In this work, the algorithm of COmpAsS is presented in detail, the signaling cost perspective of the mechanism has been thoroughly investigated taking also in account the current (3GPP's) available information items as well; from the algorithmic perspective, we have optimized the mechanism of the triggering events: an optimized *Threshold* and *Hysteresis* mechanism (see section 3), that optimize its functionality from the energy consumption, as well as network signaling perspective; furthermore, the rules set that we apply on the fuzzy inference system has been progressively fine-tuned and optimized after numerous simulations and feedback loops assessing pre-defined Key Performance Indicators (KPIs); next, the environment of the simulation has become more realistic and sophisticated, having also added a building propagation model, as well as a shadowing loss model, -which were previously missing-; last but not least, in this comprehensive round of simulations, we assign multiple traffic types to the UEs (VoIP, FTP, etc.), studying the behavior of the scheme and the fine-tuned rules set for completely diverse different types of IP flows (and QoS requirements respectively).

The rest of the paper is organized as follows. In Section 2 we describe the basic components and mechanisms of the EPC network, a research in literature regarding EPC – WiFi integration with reference to the proposed schemes, and a study in context awareness in mobility management. In Section 3 we present the COmpAsS mechanism in detail as well as our signaling analysis of the algorithm during the handover

procedure. In Section 4 we demonstrate our experiments in ns3 simulator as well as the results of said experiments and our analysis.

# 2. THE EPC NETWORK

## 2.1 Cellular networks intro

A cellular network is a radio network distributed over land through cells where each cell includes a fixed location transceiver known as base station. These cells together provide radio coverage over larger geographical areas. User equipment (UE), such as mobile phones, is therefore able to communicate even if the equipment is moving through cells during transmission.

Cellular network technology supports a hierarchical structure formed by the base transceiver station (BTS), mobile switching center (MSC), location registers and public switched telephone network (PSTN). The BTS enables cellular devices to make direct communication with mobile phones. The unit acts as a base station to route calls to the destination base center controller. The base station controller (BSC) coordinates with the MSC to interface with the landline-based PSTN, visitor location register (VLR), and home location register (HLR) to route the calls toward different base center controllers.

Cellular networks maintain information for tracking the location of their subscribers' mobile devices. In response, cellular devices are also equipped with the details of appropriate channels for signals from the cellular network systems. These channels are categorized into two fields:

- Strong Dedicated Control Channel: Used to transmit digital information to a cellular mobile phone from the base station and vice versa.
- Strong Paging Channel: Used for tracking the mobile phone by MSC when a call is routed to it.

A typical cell site offers geographical coverage of between nine and 21 miles. The base station is responsible for monitoring the level of the signals when a call is made from a mobile phone. When the user moves away from the geographical coverage area of the base station, the signal level may fall. This can cause a base station to make a request to the MSC to transfer the control to another base station that is receiving the strongest signals without notifying the subscriber; this phenomenon is called handover. Cellular networks often encounter environmental interruptions like a moving tower crane, overhead power cables, or the frequencies of other devices.

The size of a cell can vary according to the number of users that have to be served in a certain area and the amount of traffic per user. If there is much traffic in an area the cell size will be smaller than in rural areas.

In a cellular system, as the distributed mobile transceivers move from cell to cell during an ongoing continuous communication, switching from one cell frequency to a different cell frequency is done electronically without interruption and without a base station operator or manual switching. This is called a handover or handoff.

Typically, a new channel is automatically selected for the mobile unit on the new base station which will serve it. The mobile unit then automatically switches from the current channel to the new channel and communication continues. The exact details of the mobile system's move from one base station to another varies considerably from system to system.

## 2.2 EPC, SAE, LTE

### 2.2.1 LTE – SAE

LTE (Long Term Evolution) is a standard in mobile radio communications developed by 3GPP (3rd Generation Partnership Project), introduced in 3GPP R8. LTE is accompanied by an evolution of the non-radio aspects of the complete system, under the term 'System Architecture Evolution' (SAE) which includes the Evolved Packet Core (EPC) network. Together, LTE and SAE comprise the Evolved Packet System (EPS), where both the core network and the radio access are fully packet-switched.

GSM was developed to carry real time services, in a circuit switched manner, with data services only possible over a circuit switched modem connection, with very low data rates. This means that circuits are established between the calling and called parties throughout the telecommunication network (radio, core network of the mobile operator, fixed network).

The first step towards an IP based packet switched solution was taken with the evolution of GSM to GPRS, using the same air interface and access method, TDMA (Time Division Multiple Access. With this technology, data is transported in packets without the establishment of dedicated circuits. This offers more flexibility and efficiency. In GPRS, the circuits still transport voice and SMS (in most cases). Therefore, the core network is composed of two domains: circuit and packet.

To reach higher data rates in UMTS (Universal Mobile Terrestrial System) a new access technology WCDMA (Wideband Code Division Multiple Access) was developed. The access network in UMTS emulates a circuit switched connection for real time services and a packet switched connection for datacom services. In UMTS the IP address is allocated to the UE when a datacom service is established and released when the service is released. Incoming datacom services are therefore still relying upon the circuit switched core for paging.

The Evolved Packet System is purely IP based. Both real time services and datacom services are carried by the IP protocol. This allows operators to deploy and operate one packet network for 2G, 3G, WLAN, WiMAX, LTE and fixed access (Ethernet, DSL, cable and fiber). The IP address is allocated when the mobile is switched on and released when switched off. LTE is based on OFDMA (Orthogonal Frequency Division Multiple Access) and in combination with higher order modulation (up to 64QAM), large bandwidths (up to 20 MHz) and spatial multiplexing in the downlink (up to 4x4) high data rates can be achieved. The highest theoretical peak data rate on the transport channel is 75 Mbps in the uplink, and in the downlink, using spatial multiplexing, the rate can be as high as 300 Mbps.

### 2.2.2 EPS Architecture

As we mentioned before LTE and SAE comprise the Evolved Packet System (EPS). EPS uses the concept of EPS bearers to route IP traffic from a gateway in the PDN to the UE. A bearer is an IP packet flow with a defined Quality of Service (QoS). The E-UTRAN and EPC together set up and release bearers as required by applications. EPS natively supports voice services over the IP Multimedia Subsystem (IMS) using Voice over IP (VoIP), but LTE also supports interworking with legacy systems for traditional CS voice support. Below we present the overall EPS network architecture giving an overview of the functions provided by the Core Network and E-UTRAN.

The key components of EPS are:

- Mobility Management Entity (MME) - The MME is the control node which processes the signaling between the UE and the CN. The protocols running between the UE and the Core Network are known as the Non-Access Stratum (NAS) protocols. It is responsible for the establishment, maintenance and release of the bearers, as well as the establishment of the connection and security between the network and UE.
- Serving Gateway (SGW) - All user IP packets are transferred through the S-GW, which serves as the local mobility anchor for the data bearers when the UE moves between eNodeBs, and also when inter-working with other 3GPP technologies such as GPRS3 and UMTS4
- Packet Data Node Gateway (PGW) - The P-GW is responsible for IP address allocation for the UE, as well as QoS enforcement and flow-based charging according to rules from the PCRF (Policy and Charging Rules Function). Also responsible for the filtering of downlink user IP packets into the different QoS-based bearers.
- Home Subscriber Server (HSS) – HSS is a database that contains user-related and subscriber-related information. It also provides support functions in mobility management, call and session setup, user authentication and access authorization.
- ANDSF (Access Network Discovery and Selection Function): The ANDSF provides information to the UE about connectivity to 3GPP and non-3GPP access networks (such as WiFi). The purpose of the ANDSF is to assist the UE to discover the access networks in their vicinity and to provide rules (policies) to prioritize and manage connections to these networks.
- ePDG (Evolved Packet Data Gateway): The main function of the ePDG is to secure the data transmission with a UE connected to the EPC over an untrusted non-3GPP access. For this purpose, the ePDG acts as a termination node of IPsec tunnels established with the UE.

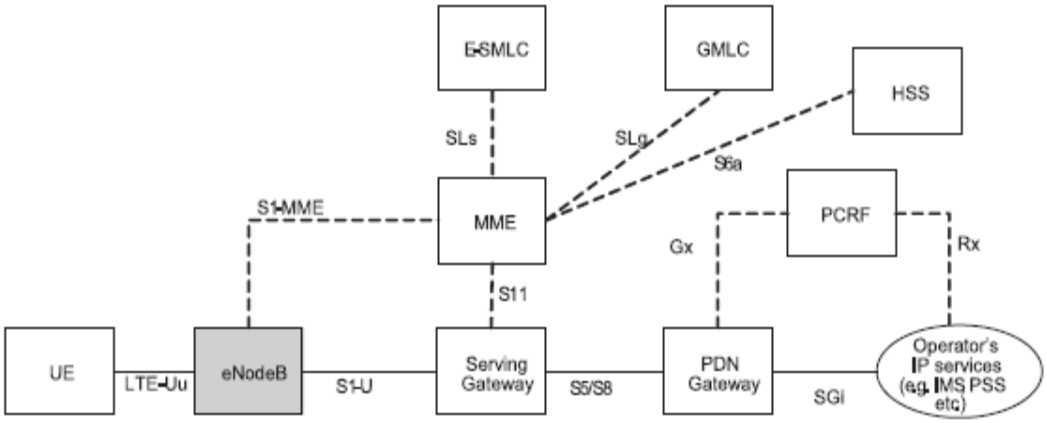


**Figure 1: EPS elements overview[18]**

### 2.2.3 EPC – WiFi integration

With the ever increasing rate of mobile broadband subscription, operators are compelled to explore new options to meet the high demands for mobile Internet. One such option is integrating WiFi with the core network as its relatively low cost, simple architecture and usage of non-licensed spectrum makes it an attractive data solution. However, one of the key challenges that operators consider when integrating WiFi into the mobile core is maintaining session continuity when realizing a handover between WiFi and other access technologies, as well as maintaining security throughout. In this section we examine some of the proposed ways to manage this integration seamlessly regarding the network architecture, stability, ease of deployment and other parameters.

As far as integrating Wifi and Cellular networks there are two models of WiFi/Cellular interworking with multiple variations possible for each. These models are generally referred to as tightly coupled and loosely coupled networks. The following is an overview of the network aspects related to coupling between 3GPP and WiFi:

- **Loosely coupled networks**: In a loosely coupled network, the WiFi network performance is usually not within the 3GPP operator's control, or has not been integrated by the 3GPP operator into a common converged wireless solution (e.g., when a mobile operator partners with a Wireless Internet Service Provider (WISP) or Multiple System Operator (MSO) who has deployed a WiFi network). End user experience may include loss of IP session continuity, and break in data connectivity when reselection occurs between networks. Typically, this type of solution is used to provide offload of best-effort traffic to WiFi while freeing up resources on constrained cellular networks. Given the potential impacts on user experience, intelligent network selection can play an important role in this model. In this case, ANDSF can be used by operators to distribute policies that guide traffic steering decisions that maximize user experience.
- **Tightly coupled networks**: In a tightly coupled network, the WiFi network performance is usually within the 3GPP operator control. This may also include integration between the 3GPP and WiFi RAN networks, with common core infrastructure. Integration between the networks is designed to provide IP session continuity and seamless end user experience, so the end user is agnostic of wireless network type. As such, carriers can start making decisions

based on which RAT will provide the greatest QoE for a given subscriber/service at a given time/location. In some cases, Trusted carrier WiFi may provide that best experience, other times the cellular RATs might, and intelligent operator controlled network selection will play a vital role in making sure that the decision to move traffic to/from WiFi is done in a way that maximizes QoE. Just as in the loosely coupled model, ANDSF gives operators a means of guiding WiFi network selection within a tightly coupled network. In addition, discussions are also ongoing in 3GPP RAN standardization group on network centric solutions supporting WiFi and 3GPP interoperability, whereby offloading and traffic steering decisions should consider not only the concerned user's experience but also the other active users in both the 3GPP network and WiFi.

Many of the challenges facing WiFi/Cellular integration have to do with realizing a complete intelligent network selection solution that allows operators to steer traffic in a manner that maximizes user experience and addresses some of the challenges at the boundaries between RATs (2G, 3G, LTE and WiFi). Four of those key challenges are the following:

- Premature WiFi Selection: As devices with WiFi enabled move into WiFi coverage, they reselect to WiFi without comparative evaluation of existing cellular and incoming WiFi capabilities. This can result in degradation of end user experience due to premature reselection to WiFi. Real time throughput based traffic steering can be used to mitigate this.
- Unhealthy choices: In a mixed wireless network of LTE, HSPA and WiFi, reselection may occur to a strong WiFi network, which is under heavy load. The resulting 'unhealthy' choice results in a degradation of end user experience as performance on the cell edge of a lightly loaded cellular network may be superior to performance close to a heavily loaded WiFi AP. Real time load based traffic steering can be used to mitigate this.
- Lower capabilities: In some cases, reselection to a strong WiFi AP may result in reduced performance (e.g. if the WiFi AP is served by lower bandwidth in the backhaul than the cellular base station presently serving the device). Evaluation of criteria beyond wireless capabilities prior to access selection can be used to mitigate this.
- Ping-Pong: This is an example of reduced end user experience due to ping-ponging between WiFi and cellular accesses. This could be a result of premature WiFi selection and mobility in a cellular environment with signal strengths very similar in both access types. Hysteresis concepts used in access selection similar to cellular IRAT, applied between WiFi and cellular accesses can be used to mitigate this.

Some key aspects of the Wifi/ Cellular integration are:

**Seamless service continuity between 3GPP and WIFI**

Session mobility between WiFi and cellular networks (with IP address preservation) has been a desirable vision for many years. The demand for data traffic and pressure on macro networks is accelerating and driving the need to make this possible. It is highly desirable for both the end user and the service provider to have the ability to seamlessly move an IP session between cellular access and WiFi access. This functionality also

enables more sophisticated mobility scenarios such as Multi-Access PDN Connectivity (MAPCON) and IP Flow Mobility (IFOM) as defined in 3GPP [19]. There are several protocols Proxy Mobile IPv6, Dual Stack Mobile IPv6, GPRS Tunneling Protocol (PMIP, DSMIPV6, GTP) and interfaces (S2a, S2b, S2c) that can be used to achieve this objective. WLAN access may be used by mobile operators to provide mobile network access. It allows an end-user to use their mobile device's WLAN access interface and a "connection manager" client to route traffic back into the mobile network operator's packet core network and hence to both obtain access to mobile operator services and in-direct access to the public Internet via mobile operator. The mobile operator role involves both user plane routing and control plane functions including backend support for the Authentication, Authorization and Accounting chain to provide access control and billing for WLAN service. In this case the end-user's device is assigned an IP address by the mobile operator and any requirement for legal interception of user traffic would fall on the mobile operator.

- **Service layer session continuity**

  Service layer session continuity refers to the solution where the application ensures the continuity of the service even though the IP address used to access the service has changed (due to a mobility event). For applications (e.g. web browsing, e-mail client) where the UE is a client, when the IP address of the UE has changed, the application can issue further requests on a new Transmission Control Protocol (TCP) connection using the new IP address of the UE.

  However for a mobility event too frequent changes of IP address can result in the UE continuously being interrupted, for example during the filling of an application form information need to be re-entered, or the streaming of a video file the UE might not be able to get a portion of the video file.

  So certain service types are likely to be impacted by the IP changes. The user experience depends on the type of service, and relying on service layer session continuity is not a generic solution.

- **3GPP Methods for mobility between 3GPP and WIFI networks**

  The Evolved Packet Core (EPC) architecture has been designed to provide support to both legacy (2G/3G) and LTE access and to provide support for access to mobility with non-3GPP access (e.g. WiFi). Support of non-3GPP access is described in 3GPP TS 23.402. Two kinds of non-3GPP access networks to EPC are defined by 3GPP TS 33.402: un-trusted and trusted non-3GPP access networks. As defined in clause 4.3.1.2 of TS 23.402, it is the home operator policy decision if a non-3GPP access network is treated as trusted non-3GPP access network. When all of the security feature groups provided by the non-3GPP access network are considered sufficiently secure by the home operator, the non-3GPP access may be identified as a trusted non-3GPP access for that operator. However, this policy decision may additionally be based on reasons not related to security feature groups. When one or more of the security feature groups is considered not sufficiently secure by the home operator, the non-3GPP access is identified as an un-trusted non-3GPP access for that operator. In this case, the UE has to establish an IPsec tunnel to the Enhanced Packet Data Gateway (ePDG) by conducting Internet Key Exchange (IKEv2) with Extensible Authentication Protocol Method-Authentication and Key Agreement

(EAP-AKA) for UE authentication, (refer to the description on S2b access to EPC later in this section for additional details). When IP address preservation during mobility between 3GPP and non-3GPP access is required, the EPC relies on the P-GW acting as an anchor point between these two kinds of accesses and hiding the mobility to the entities of the Packet Data Network.

Depending upon the nature of the interworking solution the following options may be offered:

- Non-seamless mobility of all packet connections: The UE gets different IP addresses (and possibly different services) over WLAN and over normal mobile network access. This feature has been defined in 3GPP Release 6 as “I-WLAN / 3GPP IP Access”.
- Seamless mobility of all packet connections: At mobility between 3GPP and WLAN access, all PDN connections are handed-over and have their IP address preserved. This feature has been defined in 3GPP Release 8 as “Un-trusted non-3GPP access”.
- Seamless mobility of individual PDN connections: At mobility between 3GPP and WLAN access, the UE determines which Public Data Network (PDN) connections are handed-over (with IP address preservation). For example, an Access Point Name (APN) for best effort Internet moves between cellular and WLAN access as soon as WLAN is available while a second APN for IMS service remains on the cellular access. This feature has been defined in 3GPP Release 10 as MAPCON.
- Seamless mobility of individual IP flows on specific PDN connections: At mobility between 3GPP and WLAN access the UE determines which IP flows of a PDN connection are handed-over (all PDN connections have their IP address preserved). For example, best effort Internet traffic on the default APN move between cellular and WLAN access as soon as WLAN is available while a dedicated video streaming flow on the same APN and a second APN for IMS service remains on the cellular access. This feature has been defined in 3GPP Release 10 as Internet Protocol Flow Mobility and seamless WLAN Offload (IFOM).

**Supporting Real-time Services & QOS over trusted WIFI**

A key service requirement is to be able to carry real time services such as VoIP or two-way video over a Trusted WLAN (S2a). In such a scenario, the following features are desirable:

- The use case where an UE can simultaneously access to IMS and to Internet is important: IMS support requires a dedicated PDN connection (per GSMA IR 92) and thus over Trusted WLAN requires the multi-homing capability brought by SaMOG phase 2 (3GPP Rel12). The UE can simultaneously benefit from NSWO (for access to Internet) and from a PDN connection dedicated to IMS.
- IMS signaling: it is desirable that after mobility between 3GPP and WiFi coverage, the UE does not need to re-REGISTER (or to issue re-INVITE). The preservation of UE’s IP address after mobility between 3GPP and WiFi coverage and the capability for the same Proxy Call Session Control Function (P-CSCF) to serve the UE over both 3GPP and Trusted WiFi coverage are key pre-requisites for this goal.

- QoS: Making sure that a Trusted WLAN Access Network delivers packets carrying VoIP with the relevant QoS. Both DL and UL directions should be considered. This is discussed in greater detail in the next section.
- Charging and Location: e.g. making sure that a Trusted WLAN Access Network can at the set-up and release of a VoIP call provide to the service layer (e.g. IMS) location Information (such as Cell-Id and PLMN-Id) that is similar to the information a 3GPP access can provide. For a 3GPP access, this has been defined in 3GPP Rel11 as part of the "NetLoc" Work Item (which defines how the IMS can get the identity of the 3PP Cell serving the UE at the start and at the release of an IMS session).
- Roaming where TWAN and 3GPP are controlled by the same operator.

### Security and Authentication

Making secure connectivity to WiFi access networks transparent for the end user is clearly a service requirement. A key obstacle to user experience of seamless connectivity over WiFi has been a lack of appropriate air link security and access authentication mechanisms.

To provide network access to subscribers on an integrated WiFi cellular network they would first have to be uniquely identified and authenticated by the network. An integrated network environment based on 3GPP Evolved Packet Core (EPC) will have dual mode devices supporting both WiFi and cellular technologies. Such devices would include a UICC module with (U)SIM application having user subscription information and authentication credentials stored in a tamper proof manner. These (U)SIM based credentials are for authentication with cellular networks but their existence on dual mode devices makes it easier to reuse them for authentication over WiFi accesses that are integrated with cellular networks. EAP-SIM, EAP-AKA and EAP-AKA' are WiFi access authentication mechanisms that make use of (U)SIM credentials.

EAP-SIM is the EAP based mechanism defined for authentication based on SIM credentials. EAP-AKA is an improvement on EAP-SIM and is based on USIM symmetric keys allowing for mutual authentication, integrity protection and replay protection. EAP-AKA' is a minor revision to EAP-AKA method with a new key derivation function. It should be noted that while all three of these mechanisms make use of (U)SIM credentials, based on existing 3GPP specifications only the EAP-AKA and EAP-AKA' methods can provide access to the EPC via non-3GPP accesses6 . It is therefore recommended that both EAP-AKA and EAP-AKA' be supported for authentication purposes in an integrated network environment

### Intelligent network selection

As it's now possible for a given UE to be simultaneously in range of a variety of different networks: traditional cellular networks (i.e., 3G/LTE (e)NodeB's), integrated small cells (i.e., with 3G, LTE, and WiFi), and a variety of standalone WiFi AP's (i.e., ranging from private consumer-grade AP's to carrier-grade AP's that are tightly integrated with existing cellular networks), selecting the best network for a given user at a given time in a given location is critically important for optimizing user experience.

### ANDSF

Access Network Discovery and Selection Function [10] as mentioned earlier in EPS elements, is a primary enabler of intelligent network selection between 3GPP and non-3GPP access networks able to provide UE's with useful information and operator-defined policies to guide network selection decisions. ANDSF -closely coupled with the Policy and Charging Rules Function (PCRF) - implements dynamic data offload for the User Equipment (UE) in a structured method, while in addition, enables the operator to store its policies for discovery and selection of RATs on a server. The UEs are updated with these polices by the server. The policies within ANDSF contain information on which of the available WiFi hotspots are preferable during a specific time or day, and at a specific location as well, based on indications from past measurements.

### SaMOG

SaMOG [13] allows UEs to seamlessly handover between cellular and WiFi network. According to SaMOG specification, the WiFi gateway does not connect directly to the EPC via the Packet Gateway (PGW). Another network entity, the Trusted Wireless Access Gateway (TWAG) is used, acting as the perimeter security entity of the EPC network and connects to the PGW over a secure GTP tunnel.

### TWAG - ePDG

The telecommunications industry has defined two network elements to serve as a secure gateway between a service provider's core network—the evolved packet core in the case of mobile service providers—and both trusted and untrusted WiFi networks. For access to trusted WiFi networks such as those deployed by or in partnership with the service provider, the industry has defined the Trusted WLAN Access Gateway/Proxy (TWAG/TWAP) as this secure entry point. For access to untrusted WiFi networks such as those operated independently or in connection with another service provider, the appropriate network element to secure WiFi access would be the evolved Packet Data Gateway (ePDG).

The ePDG can be used for interworking between the EPC and untrusted non-3GPP networks that require secure access, such as a WiFi, LTE metro, and femtocell access networks. It can use either IPSec/IKEv2 or proxy mobile IPv6 (in case the mobile subscriber is roaming in an untrusted non-3GPP system) for highly secure access to the EPC network. The ePDG builds strength and security into the network using:

- Tunnel authentication and authorization
- Transport level packet marking in the uplink
- Policy enforcement of Quality of Service (QoS) based on information received via Authorization, Authentication, Accounting
- Lawful interception, and other functions

Once the ePDG function is integrated within their packet core, mobile operators can take back control of the user experience and protect brand perception. For subscribers, this means assurance of a managed user experience along with secure transactions and session continuity.

### GTP and PMIP approach

GTP [11] and PMIP [12] are two network-based IP-level mobility protocols that can help operators support IP mobility in low-latency, higher data-rate, all-IP core networks that support real-time packet services over multiple access technologies. They support uninterrupted handoff by maintaining the same UE IP address when moving from one network to another. They rely on an all-IP core network to enable interworking mobility, while other standards and solutions heavily depend on clients' implementation with additional hardware and software. The client-based approach requires coordinated and lockstep efforts from both operators and device vendors, making it more difficult to arrive at a short-term solution. As defined in the EPC architecture, GTP and PMIP support IP-session continuity.

GTP was originally developed by ETSI for GPRS packet core architectures in late 1990s. It has become one of the fundamental protocols of 3GPP packet core and is very widely deployed. The GTP-based mobility mechanism requires entities in the network to communicate via GTP-based interfaces. New tunnels are built and the same IP address for the UE is maintained to support mobility. All packets sent to a home network are routed to the UE via the home GGSN and the TTG/PDG in a visited network.

PMIP is a more inclusive MIP (Mobile IP) -based network mobility protocol defined by the IETF in late 2000s. It relies on the network, as does GTP, to track the host movement and initiate the mobility signaling to the mobile core. Since the standard's finalization, PMIP has been established as the mobility protocol to accommodate various non-3GPP access technologies, such as WiFi, CDMA, and WiMAX. The PMIP-based mobility mechanism requires entities in the network to communicate via PMIP-based interfaces. When the UE travels from one network to another, it doesn't notice the movement due to the unchanged IP address and the mimic point of attachment in the visited network.

Two key roles are involved to support mobility — the Mobile Access Gateway (MAG) in the access network and the Local Mobility Anchor (LMA) in the mobile core.

### Hotspot 2.0

HotSpot 2.0 (HS 2.0), also called WiFi Certified Passpoint, is a new standard for public-access WiFi that enables seamless roaming among WiFi networks and between WiFi and cellular networks. Hotspot 2.0 was developed by the WiFi Alliance and the Wireless Broadband Association to enable seamless hand-off of traffic without requiring additional user sign-on and authentication. A hot spot (or hotspot) is a wireless lan (local area network) node that provides Internet connection and virtual private network (VPN) access from a given location for users of devices with wireless connectivity. Hotspot 2.0 enables compatible mobile devices to automatically and silently discover WiFi access points that have roaming agreements with the user's home network, then automatically and securely connect. The HS 2.0 specification is based on a set of protocols called 802.11u, which facilitates cellular-like roaming, increased bandwidth, and service on demand for wireless-equipped devices in general. When a subscriber's 802.11u-capable device is in range of at least one WiFi network, the device automatically selects a network and connects to it. Network discovery, registration,

provisioning, and access processes are automated, so that the user does not have to go through them manually in order to connect and stay connected. The benefits of HS 2.0 are:

- public hotspots become easier and more secure as the process is automatic and you don't have to manually pick a network
- network providers can band together as the networks are designed to work better when service providers partner with other providers
- encryption is mandatory, hotspots 2.0 require enterprise-grade WPA2 encryption

## 2.3 Ultra Dense Networks

To meet the wireless traffic volume increment of the next decade the fifth generation (5G) cellular network is becoming a hot research topic in telecommunication companies and academia. 5G Ultra dense networks are being proposed to deploy in overall cellular scenarios. They are based on the massive multiple-input multi-output antennas and the millimeter wave communication technologies. Two challenges regarding UDN deployment are the way Access Nodes (AN) density should scale in order to accommodate increasing traffic load requirements and if densification alone is sufficient to accommodate future network requirements.

Today's networks use macro cells to provide capacity and coverage, supplemented by Distributed Antenna Systems (DAS) and WiFi for indoor coverage and capacity respectively. The macro layer uses HSPA with LTE overlay and small cells are deployed mainly to extend coverage in hard to reach places. The next phase of network densification includes upgrading the macro installation with multicarrier, carrier aggregation and advanced antenna solutions. Capacity hot zones will be served via dedicated small cells and WiFi offload. Indoor DAS solutions will be upgraded to support LTE. Very dense networks will require dense small cell deployment for hot spots and indoor sites. DAS will be supplemented by indoor small cells, while LTE on unlicensed bands will be deployed as an integrated indoor solution for LTE.

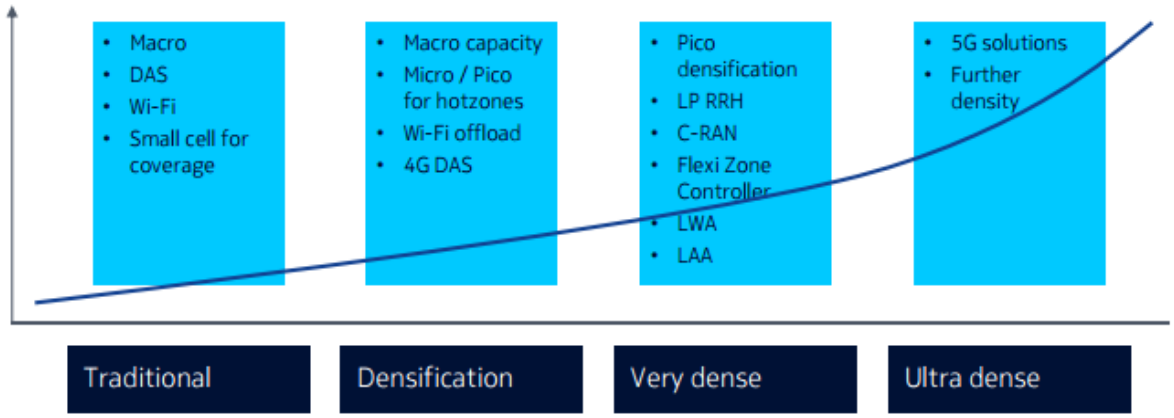


**Figure 2: Evolution to Ultra Dense Networks [20]**

The density of access/serving nodes is expected to increase up to the point where it is comparable to or even surpass the (also increasing) density of user equipments (UEs), thus introducing the ultra-dense network (UDN) paradigm. Realization of such a disruptive network topology, with respect to the evolution path of previous cellular generations, will be achieved by utilizing, in addition to densified traditional access nodes such as small cells, user deployed ANs (e.g., WiFi, femtocells), as well as “infrastructure prosumer” UEs, i.e., (mobile) devices with computational and storage capabilities allowing them to act as infrastructure ANs.

Another research involves the operation of UDNs in mm-Wave bands, due to the availability of vast amount of frequency resources compared to traditional bands, allowing to realize 5G use cases, such as amazingly fast multi-Gbps speed. Potential OFDM-based physical layer structures were proposed for realizing UDN mm-Wave PHY, the importance of beamforming was highlighted, and self-backhauling along with spectrum sharing were proposed as major technology enablers.

Centralized and distributed wireless backhaul network architectures were also compared. Simulation results suggested that the distributed wireless backhaul network architecture is more suitable for future 5G networks employing massive MIMO antennas and millimeter-wave communication technologies. It is noteworthy that the distributed wireless backhaul network architecture was also discussed for IEEE 802.16 mesh networks. Considering that the radius of IEEE 802.16 BSs is typically 1500 m, which is much larger than the 50–100 m radius of small cells, IEEE 802.16 mesh networks are not ultra-dense wireless networks. Therefore, the small cell density deployment bottleneck is not a problem for IEEE 802.16 mesh networks.

To solve the mobile user frequent handover problem in small cells, the macrocell BS is configured only to transmit the management data to control the user handover in small cells, and the small cell BS takes charge of the user data transmission. Therefore, the small cell network is not a complement for the macrocell network. 5G ultra dense cellular networks are jointly composed of small cells and macrocells. Based on the backhaul gateway configuration, two distribution architectures of ultra-dense cellular networks were proposed as follows:

- UDN networks with a single gateway - Without loss of generality, the gateway is configured at the macrocell BS, which usually has enough space to install massive MIMO millimeter-wave antennas for receiving the wireless backhaul traffic from small cells in the macrocell. The backhaul traffic of a small cell BS is relayed to the adjacent small cell BS by millimeter-wave links. All backhaul traffic of small cells is finally forwarded to the macrocell BS by multihop millimeter-wave links. In the end, the backhaul traffic aggregated at the macrocell BS is forwarded to the core network by fiber to the cell (FTTC) links.
- UDN networks with multiple gateways - In the distribution architecture of ultradense cellular networks, multiple gateways deployment is flexible for forwarding the backhaul traffic into the core network. In this case, gateways are deployed at multiple small cell BSs according to the requirement of backhaul traffic and geography scenarios. Different from the single gateway configuration, the backhaul traffic of small cells will be distributed into multiple gateways in the macrocell. The backhaul traffic aggregated at the specified small cell BS, that is, the gateway, is finally forwarded into the core network by FTTC links.

Based on comparison tests between conventional cellular networks, single and multiple gateway architecture, we see the following results: the architecture of conventional cellular networks is a centralized network architecture, and some microcells are densely

deployed in partial areas (e.g., urban areas) to satisfy crowded communication requirements. When 5G small cell BSs are equipped with massive MIMO antennas and millimeter-wave communication technologies, the coverage of a small cell is obviously reduced. To realize seamless coverage, 5G cellular networks must be densely deployed by a large number of small cells. In this case, 5G ultra-dense cellular networks can provide high bit rates in all cellular coverage regions. Moreover, the architecture of ultra-dense cellular networks is distributed, considering cost and geographic deployment requirements. Every BS in conventional cellular networks has the same function, and the coverage of macrocells and microcells overlaps. For 5G ultra-dense cellular networks, macrocell BSs transmit the management data, and small cell BSs take charge of the user data transmission. There is no overlap of the function and coverage between macrocell BSs and small cell BSs. 5G ultra-dense cellular networks with single gateways are cost efficient, but the backhaul capacity bottleneck may exist at the single gateway. 5G ultra-dense cellular networks with multiple gateways experience high cost of small cell deployment. Compared to conventional cellular networks, 5G ultra-dense cellular network performance will provide graceful degradation as the degree of mobility increases.

## 2.4 Context Awareness in Mobility Management for Cellular Networks

Mobile communication is arguably the most ubiquitously used technology in contemporary world, evolving towards its fifth generation (5G). The key challenges being faced by present day mobile communication are growing number of mobile users and subsequent high traffic volume posed by them. Providing uniform service quality and best quality of experience (QoE) in such dense scenarios is a major motive of 5G. Context information is the information that enables the perception of states and situations of network entities (e.g., network nodes, terminals, users, etc.) and their interactive relations. The radio network becomes context aware when context information is utilized to assist and optimize the operation of the network. Context awareness is recognized as one of the key pillars in enabling uniform quality of experience for mobile users and it can be utilized it to optimize user performance. For instance, predicting the next cell for user transition, predicting the crowd formation in a cell etc., will assist the base station to reserve or manage resources and prepare the cell well in advance for a future event, targeting to provide uninterrupted and uniform QoE.

### 2.4.1 State of the art – current approaches

In the recent years, numerous novel efforts found in journal papers, which address the handover mechanism, traffic steering and RAT selection procedures for future networks have been published. Some of them are mentioned below.

In [22], the authors focus on the handover delay challenges, from the handover security and user authentication perspective. Ultra Dense deployments may result in frequent handovers, which may subsequently introduce high delay overheads. They propose the Software Defined Networking (SDN) enabler as one of the most promising solutions; through its centralized control capability, user-dependent security context may be exchanged between related access points and enable delay-constrained 5G communications. The context is shared between nodes and APs based on UE path prediction. A redesign on the “intra-macrocell” handover procedure is described in [23],

focusing on the control/user plane split Heterogeneous Networks (HetNets) of future systems; the handover optimization is realized by predicting the received signal quality of the UE, triggering as a result the handover decision in a more efficient way. The authors focus on the challenges of the handover between macro cell at high speed scenarios (railway, highway, etc.). The respective evaluation shows that by predicting the forthcoming UE measurement reports, the handover execution takes place in advance and the handover performance is enhanced.

Similarly, in [24], the authors also focus on the -control and user plane separated- future HetNets and more specifically, on the signaling latency reduction in cases of macro cell base station fail-over periods. The proposed solution is based on a small cell controller scheme for controlling and managing small cells boundaries in a clustered fashion, during the corresponding macrocell's fail-over period. The evaluation of the proposed scheme on the UE side demonstrates reduced signaling latency, particularly for high user velocities; however, at the same time, the data delivery latency increases comparing to the legacy scheme; the authors conclude that the application of the proposed scheme can be selected on the specific signaling and data delivery latency requirements of each use case.

The RAT selection and handover procedure have been also studied from lower levels' perspectives as well. The very recent work in [25] focuses on the high frequency bands above 6 GHz, which will provide considerably larger bandwidths than the legacy systems; the challenge that the authors identify relates to the modifications that need to take place in the design of some key functions, such as the handover in order to support future deployments. They propose a novel frame structure, flexible and scalable to support various numbers of beams/antennas, users, or traffic conditions. The evaluation that was conducted involved static, as well as high velocity UEs; the authors conclude that the proposed enabler succeeds at satisfying all the throughput and delay requirements of the forthcoming 5G and beyond use cases.

Handover management in ultra dense heterogeneous small cell networks is studied in [26], focusing on the cell edge users. The authors describe an architecture comprising a cloud radio access network (C-RAN), as well as base band unit (BBU) pools, in which resource management and control capabilities are co-located, such as handover decision function and admission control. The proposed handover is realized between the BBU pools. The evaluation of the proposed scheme showed that the capacity of the small cells is increased, without increasing however the QoS of the users as well. In [27], the authors outline the main challenges that come with the UDNs. Among numerous challenges, such interference mitigation, backhaul issues and energy consumption, the authors tackle the mobility and handover challenge as well. Among the enabling technologies they propose is cell and receiver virtualization, self-backhaul solutions and user-centric control of user information to minimize signaling.

Next we present some approaches focused on mobility prediction. The mobility framework for LTE was originally developed and analyzed by the 3rd generation partnership project (3GPP) for macro-only networks, and was therefore not explicitly optimized for HetNets. In LTE Rel. 11, mobility enhancements in HetNets have been investigated through a dedicated study item [16]. Mobility prediction has been a key component in building context awareness. By anticipating/learning mobility behavior of the user, it is possible to design various context aware resource management schemes, handover procedures, cell activation/deactivation schemes etc. In literature, there are several related schemes that exploit user mobility behavior to build context awareness.

Mobility prediction can be briefly classified into two groups:

1) History based: The user's next cell or path is predicted based on the statistics of user's mobility. The mobility history of the user is recorded and probability of user transition into next cell is derived. The common methods of deriving probability of transition into next cell involve Markov chain model [28], Hidden Markov model [29], neural networks and machine learning [30], route clustering [31] etc.

2) Measurement based: These schemes do not rely on the user mobility history rather they derive probability of user transition to next cell based on real time measurements (e.g., RSSI, geometry, user angle, distance, etc.). [32], [33] make use of signal strength (RSSI, geometry) in dB to predict next cell, whereas, [34] relies on user angle and distances to predict the next cell.

These mobility prediction schemes are then used to proactively reserve resources [28] [29], trigger load balancing (LB) [34] or activate/deactivate small cells [33]. Thus, mobility prediction can be seen as a driving force for context awareness in cellular networks.

Based on commonalities among various context aware schemes discussed, a general framework to support mobility context awareness is proposed in [35]. The set of required information, involved signaling and interfaces are outlined. Further, a context aware resource allocation scheme is presented that makes use of information arising from vehicular infrastructure

The common information set required as input by the mobility context aware schemes are listed below.

- **User position**: This information can be in the form of (x,y,z) co-ordinates and could be obtained from global positioning system (GPS) or network assisted positioning.
- **User velocity**: This information can be obtained by Doppler measurements or can be obtained from speedometer of vehicle by using proper interfacing from vehicular infrastructure to cellular network.
- **User geometry (dB)**: Measured at the user terminal.
- **Neighboring cells list (NCL)**: Maintained by operations support system.
- **Route maps**: Similar to NCL but has information about roads, cross roads, possible coverage holes in them.

In addition it is proposed to extract and exploit the following information from vehicular infrastructure to enhance the mobility prediction and context awareness.

- **Origin**: The initial position (location) from where the user started his journey.
- **Destination**: The final position (location) to where the user intends to travel.

In addition in [35], the probability of transition of a user from base station $BSn$ to $BSn$+1 based on Markov based next cell prediction [28] can be obtained as the ratio between; number of times user in $celln$ transited to $celln$+1 and number of times user was found in $celln$. The Markov model can be further enhanced by exploiting context information about user's origin and destination. It could be observed that prediction accuracy increases when additional context information about user origin and destination are used. The next cell prediction accuracy is enhanced to around 85% and route prediction accuracy is improved to around 90%, indicating that extracting and exploiting information about user origin and destination from vehicular infrastructure is valuable for mobility prediction.

Mobility context aware procedures can be functionally decomposed into three blocks:

1) Context extraction: The functionality of obtaining user positions, signal strength/geometry measurements, obtaining information about origin and destination from vehicular infrastructure etc., are inclusive to this block. Velocity estimate is required additionally to set the sampling rate of aforementioned information.

2) Communication: This block deals with signaling various information between user and base stations. Two major parts are: a) Signaling user positions, geometry etc., to the base station. b) Signaling context message (e.g., trigger for load balancing/resource management, cell activation/deactivation message etc.) from serving base station to target (predicted) base station.

3) Prediction/Decision: The prediction of next cell, route and other event predictions making use of extracted context information, take place in this block. The resource management, cell activation/deactivation decisions etc., are also made here.

In [38], a joint MM and context-aware UE scheduling approach is proposed, by using tools from reinforcement learning. Hereby, each base station (BS) individually optimizes its own strategy (REB, UE scheduling) based on limited coordination among tiers. Both macro- and picocells learn how to optimize their traffic load in the long-term and the UE association process in the short-term by performing history and velocity based scheduling. Multi armed bandit (MAB) and satisfaction based MM learning approaches are proposed, aiming at improving the overall system performance and reducing the HOF and PP probabilities.

In the classical MM approach, there is no information exchange among tiers in case of UE handover and traffic offloading might be achieved by picocell range expansion. In the proposed MM approaches, instead, each cell individually optimizes its own MM strategy based on limited coordination among tiers. The major difference between MAB and satisfaction based learning is that MAB aims at maximizing the overall capacity while satisfaction based learning aims at satisfying the network in terms of capacity. In both cases, macro and pico BSs learn on the long-term how to optimize their REB, which results in load balancing. On the short-term, based on these optimized REB values, each cell carries out user scheduling by considering each UE's velocity and average rate, through coordinated effort among the tiers. In the proposed MM approaches, the focus is on both short-term and long-term solutions. In the long-term, a traffic load balancing procedure in a HetNet scenario is proposed, while in the short-term the UE association process is solved. To implement the long-term load balancing method, two learning based MM approaches are proposed by using reinforcement learning techniques: a MAB based and a satisfaction based MM approach. The short-term UE association process is based on a proposed context-aware scheduler considering a UE's throughput history and velocity to enable fair scheduling and enhanced cell association.

There are two approaches to the problem in [38]:

**The Multi-Armed Bandit Based Learning Approach**. The objective of the MAB approach is to maximize the overall system performance. MAB is a machine learning technique based on an analogy with the traditional slot machine (one armed bandit). When pulled at time *tn*, each machine/player provides a reward. The objective is to maximize the collected reward through iterative pulls, i.e. learning iterations. The player selects its actions based on a decision function reflecting the well-known exploration exploitation trade-off in learning algorithms

**The Satisfaction Based Learning Approach**. Satisfaction based learning approaches guarantee to satisfy the players in a system. Here, the player is considered to be satisfied if its cell reaches a certain minimum level of total rate and if at least 90% of the UEs in the cell obtain a certain average rate. The rationale behind considering these satisfaction conditions is to guarantee each single UE's minimum rate while at the same time improving the total rate of the cell.

### 2.4.2 Challenges – signaling overhead

In order for a context aware mobility management algorithm to be efficient we need to take into consideration the signaling overhead during the UEs movement and handovers. UEs may send small infrequent data, which suppose a challenge for cellular networks not optimized for such traffic, where signaling load could increase significantly and cause congestion over the network.

There have been efforts to combine the context-based knowledge extraction with context preprocessing for signaling minimization between involved network entities. In [39] authors describe a framework that comprises preprocessing primitive context information and transferring it into knowledge via tools like categorization, correction or compression. However, the particular framework remains purely at the data management level, as it does not cope with the problem from the perspective of the network signaling optimization, since the actual information exchanged between network entities is never referenced. Similarly, authors in [40] propose a model-based autonomic context management system that can dynamically configure its context information gathering and pre-processing functionality in order to provide fault tolerant provisioning of context information. The approach aims at increasing openness, interoperability of context-aware systems; however it does not manage to present an overall solution from the network signaling perspective.

In addition to the academic research efforts, several patents have been claimed related to context information preprocessing mechanisms, compression techniques, redundant traffic reduction, routing optimization methods etc. Due to space limitations, in the context of this paper, we present the most relevant, advanced and indicative solutions. Authors in [41] consider bandwidth limited communication links and evaluate the efficiency of the compression of a communication protocol. Packet header compression is another method ([42]) that minimizes the signaling overhead, however, no context awareness is taken into account, resulting only in minor improvement. Besides compression, information transfer optimization can be achieved by enhancing routing techniques as well. Optimizing mobile traffic data management via optimized polling intervals ([43]) is a paradigm that attempts to enhance the information sharing and minimize signaling overhead in mobile networks. Such methods include batching data that are directed to a mobile device received over multiple transactions, so that a connection is established only once and not for every transaction. However, it does not apply redundant information identification and removal. Another approach to implement redundant traffic reduction in wireless networks ([44]) – especially when a device requests data download from the network – is by identifying another device in the same network that has at least a portion of the requested data. In this way, the initiating device eventually requests the remaining portion, offloading this way the wireless network. Such solutions are based on caches residing on each computing system for making the portion comparison. Finally, as described in [45], the pre-fetching and preparation of certain content may serve as a means for optimizing the information

sharing process among network entities. The idea is to store content data replicas in two or more network locations and perform prediction regarding context requests by users and devices. The processing step includes content transcoding in order to ensure compatibility with the predicted user and his device. The method, however, claims no context consistency evaluation or redundant data identification and removal.

An approach to reduce signaling overhead outside the handover procedure was made in [46], focused on idle UEs. In LTE, when a UE, registered and connected at the network, becomes inactive because is not using any service, the network releases some of the allocated resources performing the S1 release procedure [8]. This procedure is used by the eNB to release the UE from connected state into idle. The inactivity timer that controls when this procedure is triggered is not standardized. It is rather defined as a vendor implementation choice. A typical value for this timer is 10ms [47]. When an idle UE wants to send a data packet, it has to perform the service request procedure to get activated and reallocated resources. Therefore, each data transmission in LTE from idle state implies a reactivation of data bearers released before, that is, it requires a new bearer setup. So in order to reduce this signaling, a new mechanism is proposed, which simplifies the transmission of a data packet using RA procedure to get enough resources to send the data packet, avoiding the need of a RRC connection and reducing the signaling load generated per transmission. It is based on 5G SDN-based architecture, and could be implemented in LTE by adding some changes, such as the increase of MME functionalities or the removal of RRC connection for small transmissions.

Another approach to reducing overall signaling in LTE networks can be seen at [48]. The paper focuses on Tracking Area List (TAL) which is a logical area-partitioning of the network, with each partition consisting of a subset of cells. While the UEs are in idle mode, their location is known to the network to the granularity of their last registered TA. Whenever a UE passes a TA boundary, an uplink signaling message is sent from the UE to the Mobility Management Entity (MME) to update its TA. This procedure is referred to as Tracking Area Update (TAU). On the other hand, when a UE is being called, a downlink message is sent from the MME to the cells inside the UE's registered TA in order to find the cell in which the UE is located. This procedure is called paging. One key parameter for designing the TA configuration of a network is the total signaling overhead from the TAU and paging signaling messages. Large-size TAs virtually eliminate TAU and cause excessive paging, whereas small TAs have the opposite effect. The proposed mechanism in [48] is a linear programming model to configure overlapping tracking area lists (TALs). It is shown that the optimum overlapping TAL solution for this problem is that each site assigns one specific TAL to all UEs being registered in that site.

As it is made clear, it is vital to reduce signaling overhead in all phases of LTE networks to minimize traffic. With this challenge in mind we designed our proposed mechanism, which will be presented in the following section.

# 3. IMPLEMENTATION

## 3.1 COmpAsS mechanism

This section focuses on the comprehensive presentation of our proposed mechanism. The section is divided into two sub-sections; the first one provides some initial insights in terms of inputs, the decision making process and outputs and next, the algorithm of the solution is illustrated and each step is discussed in a thorough way.

### 3.1.1 The network architecture perspective

The mechanism –as already mentioned earlier- is user-oriented, i.e. deployed on the UE side, rather than one of the main LTE network entities. Nevertheless, its functionality is not completely independent as the architecture of the network is directly influenced in the sense that the context information, which is required to be aggregated by the UE, is available in specific network components. Moreover, the mechanism is directly associated with the network policies, in the sense that the UE-oriented decisions are being forwarded to the central decision-making entity in the network core, which will make the final assessment. As a result, some minor adaptations need to be realized in order to enable the required context information acquisition. In this subsection, we address the requirements in terms of the data sources, message types, as well as interfaces, which are required to support the proposed mechanism.

In relation to the input parameters that have been taken into consideration for COmpAsS scheme, certain inputs are already available using the existing standards (thus no further assumptions are required), while for the rest, some additional assumptions regarding the applied protocol (message type, etc.), as well as the respective interfaces are required. The UEs monitor the following contextual information items:

- the traffic load of the cellular base stations and/or WiFi APs (in terms of available bandwidth)
- the backhaul load of the available access networks
- the mobility characteristics of the UE (speed, etc.)
- the type of the traffic flow (mapped to a specific sensitivity to latency for each flow type)
- the RSS (or Reference Signal Received Quality - RSRQ for 3GPP networks) of the available RATs/cells/APs.

More specifically:

- The *RSRQ value* is already part of the UE measurements report used in LTE for evaluating the quality of the signal of the neighbor base stations. Similarly for WiFi, RSS metric is already included in the existing IEEE 802.11 reporting metrics, even for the end-user devices.
- As indicated in [54] the *mobility state* of the UE (high-mobility state, medium, etc.) is considered and is sent via the system information broadcast from the serving cell.
- The *traffic flow type*, mapped to the respective sensitivity to latency for each application/service type executed by the UE: the different application categories and respective flow QoS requirements are extracted by the UE connection manager, from the well-established port numbers of the applications/services.

- The *traffic load* of the base stations and the *backhaul load* of the network: In the proposed mechanism, the UEs collect information from a local instance of ANDSF (Local ANDSF - L-ANDSF) as suggested in [55] about the policies of the operator for accessing WiFis in the area, as well as information from Hotspot 2.0 protocols to evaluate the status of WiFi APs (e.g., number of users associated to the AP, the load of the backhaul link of the AP etc.). The main functionality and role of the ANDSF has already been discussed. We extend this concept by assuming that local ANDSF entities (L-ANDSF) contain similar information (i.e., number of associated users, load of the network link, etc.) for every (H)eNB in a specific area . This distributed model radically decreases the information exchange delays between the nodes in a limited area, comparing to a scenario, in which one central ANDSF entity of the operator serves thousands or millions of devices. This requires appropriate logical interfaces from the (H)eNBs to the ANDSF. Last but not least, the nodes update L-ANDSF in a coarse manner (e.g., Load is Low, Medium or High) only when thresholds are violated, so as to further minimize the signaling overhead in the network.

The obtained information is aggregated by a *Context Manager* entity, part of the proposed scheme, which resides inside the UE and processes the information in order to forward it to the Fuzzy Inference Engine, which is described in the following sub-sections.

### 3.1.2 Overview of the proposed solution

The proposed RAT selection mechanism presented in this work aims at enabling the UEs to identify in an intelligent way the most suitable RAT to associate with in a specific urban area, where a cellular operator (with deployed macro, pico or femto cells) co-exists with Wireless Internet Service Providers (WISPs) with whom it has roaming agreements as suggested in [56]. The mechanism is applicable for users of 5G smartphones that support a number of RATs.

The framework is using pre-defined, customizable and fine-tuned rules for all the possible combinations of the different aforementioned scheme's inputs. The rules that are applied are policies, based on objective network parameters, KPIs and general principles, derived from the state of the art of the domain, as it was presented in the previous section. More specifically, according to these rules/policies, a RAT, which is characterized by low (backhaul) load and high RSS/RSRQ, is advantageous for the UE choice. In addition, the higher the sensitivity to latencies (traffic flow type input type), the higher impact the mobility metric has on the *Suitability*; high mobility UEs are preferably placed in larger cells to avoid unnecessary handovers and/or ping-pong effects. Using Fuzzy Logic Controllers (FLC) each UE evaluates the available RATs and identifies the most suitable one, which optimizes the Quality of Service (in terms of pre-defined KPIs) for each application (or type of traffic); afterwards it performs a session initiation or a per flow-handover using existing 3GPP mechanism described in the introductory section. The KPI, which is utilized to describe the selection prioritization among heterogeneous cells and access technologies is denoted as *Suitability* in Figure 3.

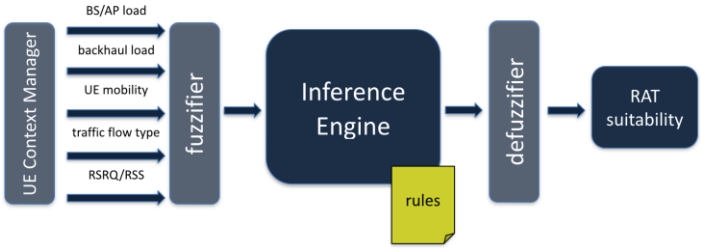


**Figure 3: Fuzzy Logic Controller for the extraction of the RAT Suitability metric**

All in all, the *Suitability* metric relates separately to each one of the active UE's traffic flows; in other words, for each active traffic flow *F* and for each available RAT *R* there is a different value, resulting thus, in $N_F$ x $N_R$ overall values, where $N_F$ and $N_R$ are the number of the flows or the available RATs respectively (see example in Table I for $N_F$=*4* and $N_R$=*5.*).

**Table I $N_F$ x $N_R$ *Suitability* Calculation example for a UE with 4 active IP flows**

| **UE active flow #** | **RAT Suitability list** |
|---|---|
| 1: browsing (downlink) | *eNB3, WLAN SSID1, eNB2, eNB1, WLAN SSID2* |
| 2: VoIP (uplink) | *WLAN SSID1, WLAN SSID2 eNB3, eNB2, eNB1* |
| 3: VoIP (downlink) | *WLAN SSID1, WLAN SSID2 eNB3, eNB2, eNB1* |
| 4: background cloud syncing (uplink) | *eNB1, eNB2, eNB3, WLAN SSID2, WLAN SSID1* |

The proposed scheme's decision making process selects for each one of these active flows the RAT, for which *Suitability* is maximized; afterwards, the UE makes a handover request to the respective (H)eNB or AP in order to transfer the flow to the optimal access technology. The process is running both on a pre-defined time interval basis, as well as upon pre-defined trigger events, which are described in detail in the following sub-section. If for any reason, the handover to the highest-ranking RAT is not possible, the 2nd choice in the *Suitability* list is selected, etc

### 3.1.3 Description of the algorithm

In this section we present the algorithm we designed for the optimal RAT selection per application/session (Figure 4). Detailed presentation has been also published in [57], [58] and [59].

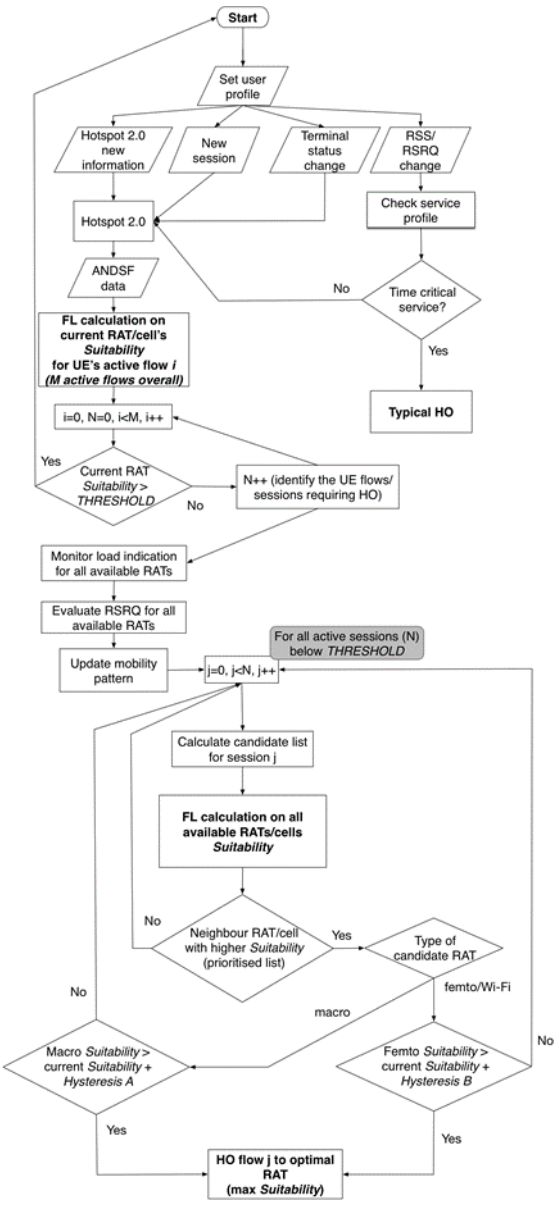


**Figure 4: The algorithm of our proposal [59]**

Initially, and prior to proceeding in each one of the algorithm steps description, it is required at this point to provide some insights regarding two parameter types, which need to be pre-set prior to the algorithm deployment on the UE. Although the FL computational requirements are minimum, in order to further optimize the energy consumption of COmpAsS scheme inside the UE, as well as to minimize the unnecessary handovers, the algorithm is evaluating two types of parameters, namely:

- a *Suitability Threshold T (0%<T<100%)*: the UE evaluates the current *Suitability* of its currently associated RAT/cell and compares it with a pre-defined algorithm parameter, namely the *Threshold*, above which the current RAT is considered as satisfactory for serving the UE requirements. For example, if the *T=90%*, no FL computation is performed (for the particular IP flow) if the associated (current) RAT's *Suitability* is above *90*% (implying that the current UE's RAT is satisfactory enough to attempt any new handover).
- a *Suitability Hysteresis* (Margin) value *(0%<H<100%)*: it describes the required advantage difference between the candidate cell's *Suitability* when compared to

the current one's in order to consider it as preferred choice. Multiple *Hysteresis* values may be used for different target RATs, according to the planning of the network administrator, for example: if $H_{MACRO}$*=10%* and $H_{FEMTO}$*=3%,* examined RAT's *Suitability* must be at least *10*% higher than the current RAT's, -if a neighbor RAT is a macro cell-, or at least 3% higher than the current RAT, -if neighbor RAT is a femto cell-, in order to trigger a handover towards the respective candidate RAT. The higher *Hysteresis* in the case of macro neighbor RAT may be chosen aiming to impel the handover to smaller RATs for offloading reasons. From a broader perspective, the customizable $H_{MACRO}$ and $H_{FEMTO}$ values as far as the *Hysteresis* is concerned provide the network administrator a wide range of options, being able to control the interworking balance between the macro and small cells, as well as dynamically route the offloaded traffic flows. Both the Suitability *Threshold*, as well as the *Hysteresis* parameter evaluation follow in the next algorithm steps.

The values of the *Threshold* and the *Hysteresis* may be configured according to the specific needs of a particular network environment by the network administrator before the mechanism is deployed on the UE. An extension of this feature that could also be accommodated in the future is the enablement of an automatic adaptation of the two control parameters, by defining the different possible "states" of the network and the respective *Threshold-Hysteresis* configuration for each one of these states. For example, for a more dense, -in terms of network deployment- environment, the solution performs better for higher *Threshold* and *Hysteresis* values. It should be also noted that the network "state" sensing is already enabled via the available context information, which is being aggregated by the UE.

The algorithm is described thereinafter step by step: initially, the user defines either on a per session basis (e.g., HTTP traffic to be handled only by free WiFi) or collectively (e.g., use always the RAT that minimize the energy consumption) his preferences (i.e., "user profile" in the algorithm flowchart). For the user profile generation numerous solutions have been proposed, i.e. either manually or automatically using data analytics solutions; for our algorithm, a solution that suggests the profile creation in an automated manner is selected. The mechanism algorithm may be triggered only if there is at least one session active in the UE. Thus, as long as there is at least one session, pre-defined events trigger the algorithm initiation, i.e.: new information from Hotspot 2.0 is received (e.g., a WiFi AP is now unloaded), a new session is initiated on the UE side, a significant change in the terminal status (e.g. battery level is falling below a certain threshold) or a significant change in the monitored RSS/RSRQ values is identified, etc.

By the time the process is triggered, and only if there is no time critical service to be served, (-in which case RSS/RSRQ has fallen below a threshold and a typical HO must urgently take place-), the UE proceeds to the information aggregation phase, which is being supported by Hotspot 2.0 and ANDSF servers, in relation to all the available neighbor RATs (3GPP or non-3GPP) and cell layers (macro, pico, femto cells, etc.), without any direct association with them. This information relates to the available 3GPP or non-3GPP cells' and APs' load, number of associated UEs, quality of received signal, etc. As already discussed, in 3GPP Rel-12 and beyond, the ANDSF enhancement creates a sufficient RAT context source. This is achieved by incorporating additional information items, better granularity for the existing for traffic steering conditions [60], as well as integrating information from Hotspot 2.0. This information is updated on the UE side, triggered on a per trigger-event basis.

Having aggregated the updated context information from the aforementioned sources, the UE performs an updated calculation on the *Suitability* of the currently associated RAT/cell, i.e., whether it is over the preset *Threshold* value. If yes, then no further calculation or signaling is required and then algorithm reverts to the starting point.

However, if the *Suitability* of the currently associated RAT/cell is below the *Threshold*, the UE continues with the aggregation of the context information of all available RATs/cell layers (i.e. load monitoring, RSRQ indicators, mobility patterns). Then, for each one of the active sessions in the UE (i.e. all active flows, including applications download/upload, background services, etc.), the Suitability KPI is calculated for the available (candidate) RATs/cells. That results in a *Suitability* –based prioritized list for each one of the active sessions/flows. Starting from the top RAT/cell, and following a top-down approach throughout the priority list, an evaluation of the *Suitability* value takes place: for macro cells (LTE, GSM, etc.), the candidate RAT's *Suitability* must be higher than the current's RAT *Suitability* by *Hysteresis* A in order for it to be selected for session handover; for small cells respectively (i.e., LTE femto cell/WiFi AP), the candidate RAT's Suitability must be higher than the current RAT's *Suitability* by *Hysteresis* B. As already discussed above, the network administrator/traffic engineer is able to control –without altering any other policies/rules- the offloading flow routes via dynamically adapting these two different *Hysteresis* values; by increasing A value comparing to B for example, the traffic engineer may target to induce session handovers to smaller RATs for offloading reasons.

The RAT/cell with the highest *Suitability* is being selected for starting the procedure of the handover; in case of a rejection the second in the list is being selected for initiating the same procedure, etc. It must be noted, that in case the priority RAT list is exhausted without satisfying the Hysteresis conditions (e.g., due to the fact that the *Hysteresis* value has been set too high), –and as a result, no handover has been decided and triggered-, the list is once more traversed without the *Hysteresis* values, in order to facilitate the handover realization.

## 3.2 Our signaling analysis

In this section we present our work regarding the signaling analysis of the handover procedure using COmpAsS. The purpose of this work is to demonstrate in detail the signaling that is occurring when a UE goes through a handover with COmpAsS algorithm in comparison to the signaling without it. A handover taking place with current mechanisms will go through the X2 interface. X2 is a transport interface used to connect eNBs together in a LTE/4G network. It supports the exchange of signaling information between two eNBs, in addition the interface shall support the forwarding of PDUs to the respective tunnel endpoints; - from a logical standpoint, the X2 is a point-to-point interface between two eNBs within the E-UTRAN. A point to-point logical interface should be feasible even in the absence of a physical direct connection between the two eNBs. X2 signaling bearer provides the following functions:

- Provision of reliable transfer of X2-AP message over X2 interface.
- Provision of networking and routing function
- Provision of redundancy in the signaling network
- Support for flow control and congestion control

The protocol stack for X2 Signaling Bearer is shown in the following figure.

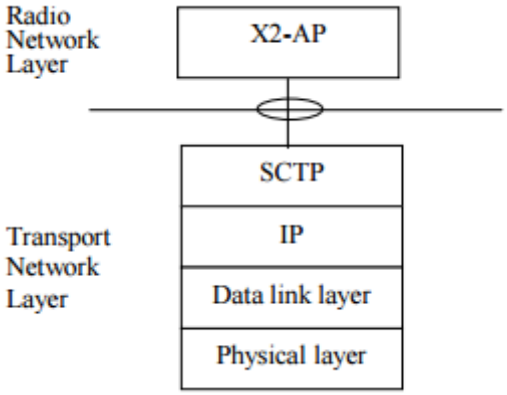


**Figure 5: X2 Signaling Bearer Protocol Stack**

We analysed extensively the information from 3GPP documents related to the X2 interface [36][37] , its functions and procedures to define the total signaling that is required during a handover using X2. We then added our COmpAsS related information that are needed for our mechanism execution and were missing from the X2 procedure.

The X2 interface X2AP procedures are divided into two modules as follows:

- X2AP Basic Mobility Procedures;
- X2AP Global Procedures;

The X2AP Basic Mobility Procedures module contains procedures used to handle the UE mobility within E-UTRAN.

The Global Procedures module contains procedures that are not related to a specific UE. The procedures in this module are in contrast to the above module involving two peer eNBs.

The X2AP function that we will mostly deal with is Mobility Management. This function allows the eNB to move the responsibility of a certain UE to another eNB or request another eNB to provide radio resources for a certain UE while keeping responsibility for that UE. Forwarding of user plane data, Status Transfer and UE Context Release function are parts of the mobility management. The Mobility management function entails the following elemental procedures

a) Handover Preparation
b) SN Status Transfer
c) UE Context Release
d) Handover Cancel

Each of the elemental procedures includes an initiating message and a response message for the occasion of successful or unsuccessful outcome. So by combining the above information we can see the mobility management functions that are important for the handover procedure and our calculations in the table below.

**Table II Mobility management functions**

| Elementary Procedure | Initiating Message | Successful Outcome | Unsuccessful Outcome |
|---|---|---|---|
| | | Response Message | Response Message |
| Handover Preparation | HANDOVER REQUEST | HANDOVER REQUEST ACKNOWLEDGE | HANDOVER PREPARATION FAILURE |
| SN Status Transfer | SN STATUS TRANSFER | | |
| UE Context Release | UE CONTEXT RELEASE | | |
| Handover Cancel | HANDOVER CANCEL | | |

By collecting information from [36] about the contents of the above messages we can add up the total signaling cost per message in the form of integers, enumerated types and bit strings. We took into consideration the minimum and maximum cases as defined in [36] with the mandatory and optional fields.

**Table III: Signaling per message type**

| Message | Integers | | Enums | | Bits | |
|---|---|---|---|---|---|---|
| | Min | Max | Min | Max | Min | max |
| **Handover Request** | 9 | 1834 | 21 | 856 | 460 | 177164 |
| **Handover Request Acknowledge** | 4 | 772 | 1 | 1027 | 8 | 81928 |
| **Handover Preparation Failure** | 2 | 259 | 2 | 516 | 0 | 0 |
| **SN Status Transfer** | 8 | 2307 | 1 | 1 | 0 | 5242880 |
| **UE Context Release** | 3 | 3 | 1 | 1 | 0 | 0 |

So the total signaling of the handover procedure using X2AP interface is:

- At minimum 24 Integers, 24 Enums and 468 bits
- At maximum 4916 Integers, 1885 Enums and 5501972 bits

In the following section we will add to the above procedure our custom messages with fields necessary for COmpAsS execution and we will present the flow of the algorithm.

As we mentioned before COmpAsS needs the following information

- the traffic load of the cellular base stations and/or WiFi APs (in terms of available bandwidth)
- the backhaul load of the available access networks
- the mobility characteristics of the UE (speed, etc.)
- the type of the traffic flow (mapped to a specific sensitivity to latency for each flow type)
- the RSS (or RSRQ for 3GPP networks) of the available RATs/cells/APs.

From the above only the *traffic load* of the base stations (or BSS load) and the *backhaul load* of the network is missing from the current message sequences of X2AP handover. The UEs will have to collect this information from a local instance of ANDSF as well as from Hotspot 2.0 protocols to evaluate the status of WiFi Aps. The ANDSF periodically updates the BSS load and Backhaul load by requesting the information from each eNB and WiFi AP that it is connected with. The UE then periodically (we define a time interval of 60 seconds) requests this information from the ANDSF to use in the COmpAsS calculations. When The COmpAsS algorithm is executed a suitability list report is generated. Afterwards the UE sends the suitability list along with the suitability threshold to the eNB that it's connected and the most suitable RAT for handovering is selected.

We calculate the signaling for messages BSS load and Backhaul load from [36]. The Backhaul load is the sum of the Composite Available Capacity element in X2 from each eNB. The BSS load is the result of the S1 transport load from each eNB. The signaling for the requests and responses can be seen in the tables below.

**Table IV: BSS load / Backhaul load request**

| Information Element (IE)/Group Name | Criticality | IE/Group Name | Presence | IE type and reference | Semantics |
|---|---|---|---|---|---|
| Message Type | YES | Procedure Code | M(andatory) | INTEGER (0..255) | "0" = Handover Preparation<br>"1" = Handover Cancel<br>"2" = Load Indication<br>...(BSS load request)<br>…(Backhaul load request)<br>etc. |
| | | Type of Message | M | ENUM(Initiating Message, Successful Outcome, Unsuccessful Outcome, …) | |
| Target Access Network type | YES | | M | ENUM (2G, 3G, 3G+, LTE, WLAN, LTE_WLAN, ALL) | |

**Table V: Backhaul load response**

| Information Element (IE)/Group Name | Criticality | IE/Group Name | Presence | IE type and reference | Semantics |
|---|---|---|---|---|---|
| Message Type | YES | Procedure Code | M(andatory) | INTEGER (0..255) | "0" = Handover Preparation<br>"1" = Handover Cancel<br>"2" = Load Indication<br>...(Backhaul load response)<br>etc. |
| | | Type of Message | M | CHOICE (Initiating Message, Successful Outcome, Unsuccessful Outcome, …) | |
| Target Access Network type | YES | | M | ENUM (2G, 3G, 3G+, LTE, WLAN, LTE_WLAN, ALL) | |
| Backhaul load value | | | If one network type | 4 Int (2 int for uplink and downlink according to Composite available capacity group in [36] | |
| | | | If LTE_WLAN | 2 * 4 Int | |
| | | | If ALL | 5 * 4 Int | |

**Table VI: BSS Load response**

| Information Element (IE)/Group Name | Criticality | IE/Group Name | Presence | IE type and reference | Semantics |
|---|---|---|---|---|---|
| Message Type | YES | Procedure Code | M(andatory) | INTEGER (0..255) | "0" = Handover Preparation<br>"1" = Handover Cancel<br>"2" = Load Indication<br>...(Backhaul load response) etc. |
| | | Type of Message | M | CHOICE (Initiating Message, Successful Outcome, Unsuccessful Outcome, …) | |
| Target Access Network type | YES | | M | ENUM (2G, 3G, 3G+, LTE, WLAN, LTE_WLAN, ALL) | |
| Backhaul load value | | | If one network type | 2 Enum (1 Enum for uplink and downlink according to S1 Transport load Indicator in [36] | |
| | | | If LTE_WLAN | 2 * 2 Enum | |
| | | | If ALL | 5 * 2 Enum | |

So with the above information we created a diagram that shows the signaling taking place between the LTE elements with the COmpAsS mechanism, which can be seen below.

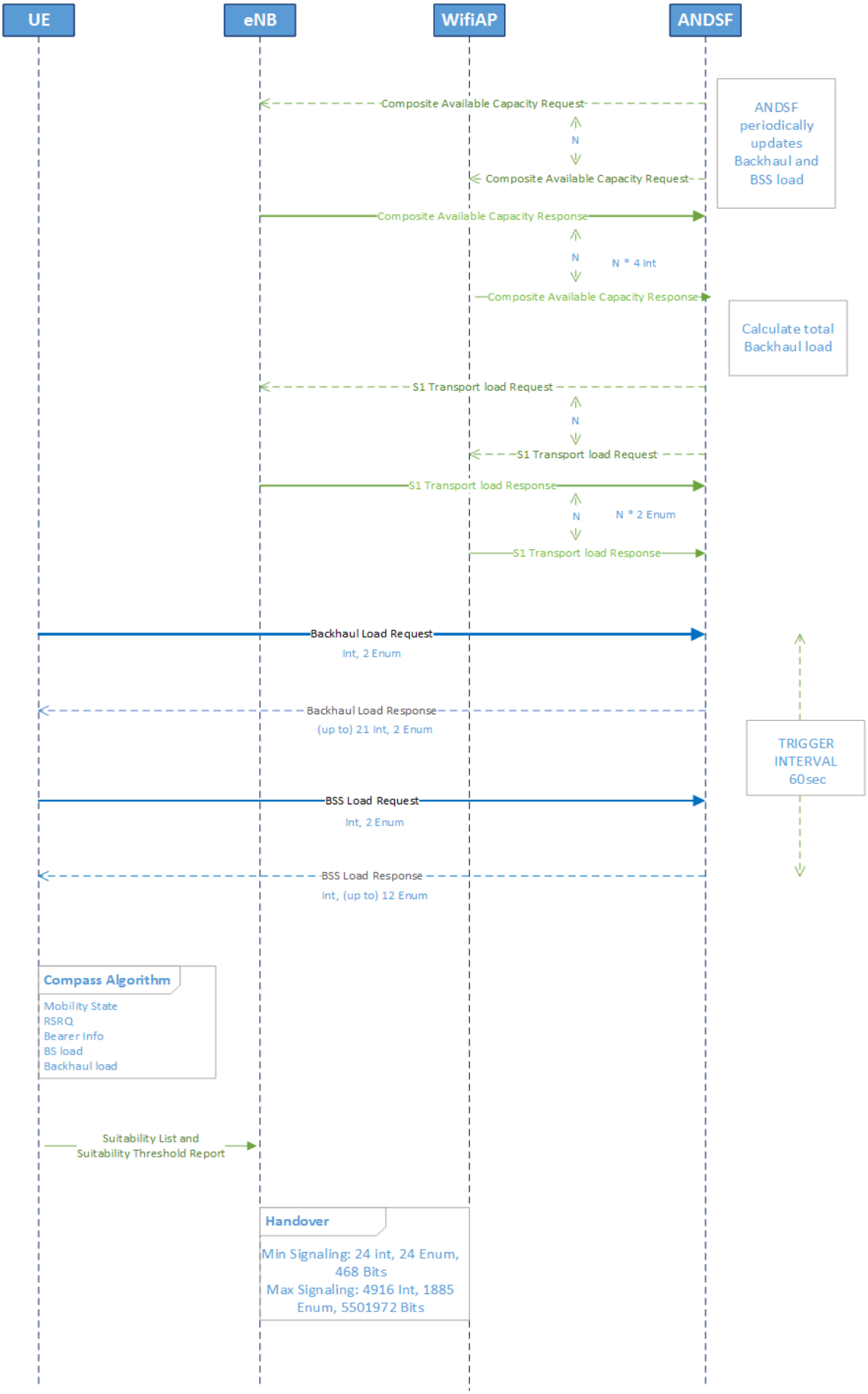


**Figure 6: COmpAsS Signaling**

# 4. EVALUATION

## 4.1 Intro – scope of the experiments

The purpose of our experiments is to determine the efficiency of our proposed mechanism in a simulated environment similar to real life conditions. In order to assess the validity and viability of our approach, we performed extensive experiments using the NS3 network simulator and -customized for the particular evaluation needs- Python modules. We considered a usage scenario from the METIS project [51], which we implemented in NS3 and evaluated the application of COmpAsS on complex RAT environments. Through the experiments we attempted to replicate –to the best possible extent and taking into account the simulator's limitations– a real life situation.

## 4.2 The NS-3 simulator

The *ns-3* simulator is a discrete-event network simulator targeted primarily for research and educational use. It has been developed to provide an open, extensible network simulation platform, for networking research and education. In brief, *ns-3* provides models of how packet data networks work and perform, and provides a simulation engine for users to conduct simulation experiments. Some of the reasons to use *ns-3* include to perform studies that are more difficult or not possible to perform with real systems, to study system behavior in a highly controlled, reproducible environment, and to learn about how networks work. Users will note that the available model set in *ns-3* focuses on modeling how Internet protocols and networks work, but *ns-3* is not limited to Internet systems; several users are using *ns-3* to model non-Internet-based systems.

Many simulation tools exist for network simulation studies. Below are a few distinguishing features of *ns-3* in contrast to other tools.

- *ns-3* is designed as a set of libraries that can be combined together and also with other external software libraries. While some simulation platforms provide users with a single, integrated graphical user interface environment in which all tasks are carried out, *ns-3* is more modular in this regard. Several external animators and data analysis and visualization tools can be used with *ns-3*. However, users should expect to work at the command line and with C++ and/or Python software development tools.
- *ns-3* is primarily used on Linux systems, although support exists for FreeBSD, Cygwin (for Windows), and native Windows Visual Studio support is in the process of being developed.
- *ns-3* is not an officially supported software product of any company. Support for *ns-3* is done on a best-effort basis on the ns-3-users mailing list.

The general process of creating a simulation can be divided into several steps:

1. ***Topology definition*:** To ease the creation of basic facilities and define their interrelationships, ns-3 has a system of containers and helpers that facilitates this process.
2. ***Model development*:** Models are added to simulation (for example, UDP, IPv4, point-to-point devices and links, applications); most of the time this is done using helpers.

3. ***Node and link configuration*:** models set their default values (for example, the size of packets sent by an application or MTU of a point-to-point link); most of the time this is done using the attribute system.
4. ***Execution*:** Simulation facilities generate events, data requested by the user is logged.
5. ***Performance analysis*:** After the simulation is finished and data is available as a time-stamped event trace. This data can then be statistically analysed to draw conclusions.
6. ***Graphical Visualization*:** Raw or processed data collected in a simulation can be graphed using tools like Gnuplot [49] or XGRAPH [50].

## 4.3 Topology – Scenarios

The main scenario considers one of the established 5G use cases, -as these were documented in [51]-, i.e., a large shopping mall with high density of customers and service staff (Figure 7: Simulation environment: Shopping mall). We selected this set-up, as a typical setting for a future extended rich communication environment, involving both “traditional” radio networks, as well as wireless sensor networks, where customers access mobile broadband communication services while they are directly addressed by personalized location-based services of the shopping environment. We evaluate this setting 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. Based on this description, we use the NS3 and model a 3-floor shopping mall. 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.

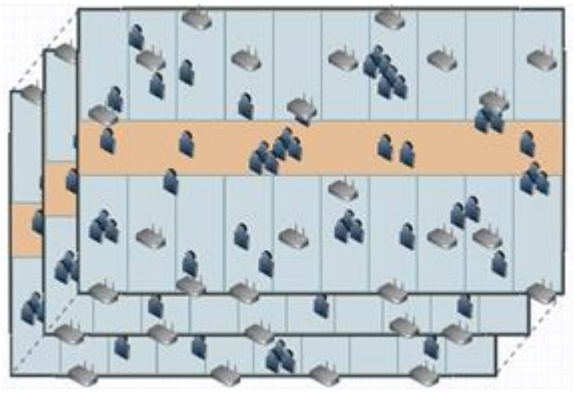

**Figure 7: Simulation environment: Shopping mall comprised of 3 floors and 20 shops per floor**

In order to evaluate the proposed framework, using LTE femto cells and macro cells is sufficient, as the rules that apply for WiFi are almost identical with the ones applied for femto cells. In addition, the IP flow mobility between LTE and WiFi networks is not available in the NS3 simulator that was selected. Our simulation scenarios are based on 3GPP Specifications

[52] and [53]. In details, the transmission mode is SISO (Single Input Single Output) and the scheduler is the NS-3 implementation of the Proportional Fair MAC scheduler. We use the Hybrid Buildings Propagation Loss Model for path loss implemented in NS3 with Internal Wall Loss at 10.0 db Shadow, Sigma Indoor at 10.0 db. The network node configuration appears in

Table VII. Services are implemented using NS3's UDP client-server application model and the desired data rates are achieved through configuration of the packet size and the inter-packet interval parameters. The service schedule for every user is pseudo-randomly generated at the beginning.

**Table VII: NS3 Simulations' Configuration**

| **NS3 Network Node** | **Tx Power (dBm)**[52] | **Downlink (DL) Earfcn (MHz)** [52] | **Bandwidth (RBs)** [52] [53] | **Antenna Type** [52] |
|---|---|---|---|---|
| Macro cell | 35 | 2120 | 50 (10 MHz) | Parabolic, 15 dBi |
| Femto cell | 20 | 2120 | 15 (3 MHz) | Isotropic |
| UE | 20 | - | - | Isotropic |
| **Other parameters** | | | | |
| Number of eNBs | 2 | | | |
| Number of HeNBs | 50 (max.) | | | |
| Number of UEs | 50 | | | |
| Simulation time | 100 s | | | |
| Time unit | 0.1 s | | | |
| Transmission mode | SISO (Single Input – Single Output) | | | |

## Scenarios Details

As mentioned before, 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. In the following table, we present in more detail the 4 different scenarios.

**Table VIII: Scenarios detailS**

| Scenario # | Scenario Parameter | Value range | Number of experiments | |
|---|---|---|---|---|
| | | | **COmpAsS** | **A2A4-RSRQ** |
| 1 | *Suitability threshold* | [0.99, 0.7, 0.1] | 9 experiments (3 different executions per threshold value) | 3 experiments (*Suitability threshold* does not apply) |
| 2 | *Suitability margin* | [0.3, 0.1, 0.001] | 9 experiments | 3 experiments (*Suitability margin* does not apply) |
| 3 | *Deployment density (number of HeNBs)* | [2, 10, 50] | 9 experiments | 9 experiments |
| 4 | *Network load (number of traffic bearers/UE)* | [1, 3, 10] | 9 experiments | 9 experiments |

### Handover Failure

In addition to the above scenarios and topology used in our ns3 project, it was necessary to implement a function for handover failure as it was missing. All handovers complete successfully by default in ns3. As a result in order to have a more realistic scenario and be able to measure the handover failure probability, we simulate a handover failure. A handover failure occurs when:

a. the handover is initiated but the target network does not have sufficient resources to complete it, or when

b. the mobile terminal moves out of the coverage of the target network before the process is finalized.

In case (a), the handover failure probability is related to the channel availability of the target network

In case (b), it is related to the mobility of the user.

## 4.4 Results – figures & analysis

In this section we present the results of our simulations in ns3. In our experiments we tested for handover calls, handovers successfully done, the type of handover (macro to femto etc.) and various throughput stats and delay. The results accompanied with the related diagrams can be seen below.

### Scenario 1: Varying Suitability Threshold

In our first evaluation scenario the *Suitability threshold* ranges between 0.99, 0.7 and 0.1; taking into consideration that in the proposed scheme, context evaluation and *Suitability* calculation procedures are performed only when the current RAT's *Suitability* has fallen below the *threshold,* the behavior of COmpAsS varies significantly, primarily as far as the triggering frequency of the mechanism is concerned.

As we can see in the figure below, as the *threshold* value decreases, the handover frequency in COmpAsS case decreases as expected. It generally falls below the LTE case except from a single case, which can be considered that results as an outlier of the particular simulation run ("rng"). So we can see that with COmpAsS we can fine-tune the *threshold* value in order to considerably minimize the handover frequency, while –most importantly- preserving –and in several cases optimizing- the network related KPIs, such as throughput and delay.

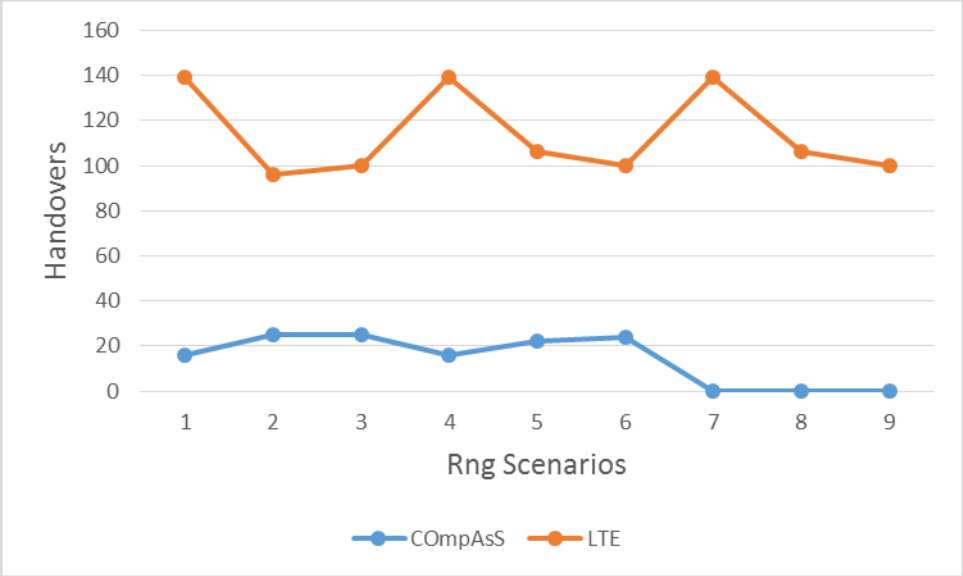


**Figure 8: Suitability Threshold Scenario Handovers**

The effect can be seen clearly if we take the average from the 3 simulation runs of each *threshold* value (Figure 9)

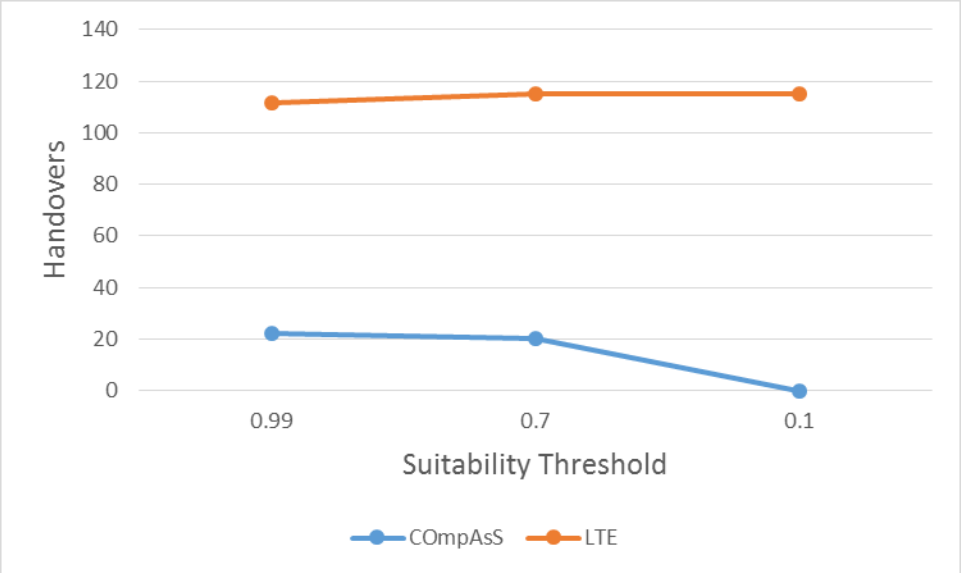


**Figure 9: Average Handovers per Suitability Threshold**

In the next part of our analysis, we illustrate how the Delay KPI varies in relation to the *Suitability Threshold*. The figure, which follows proves how COmpAsS may simultaneously minimize the number of handovers, while at the same time optimize specific KPIs of utmost importance, such as the uplink delay. Below we can see the

Uplink Delay for varying *Suitability Threshold*. We can see that COmpAsS has lower Uplink Delay than the LTE algorithm in all cases.

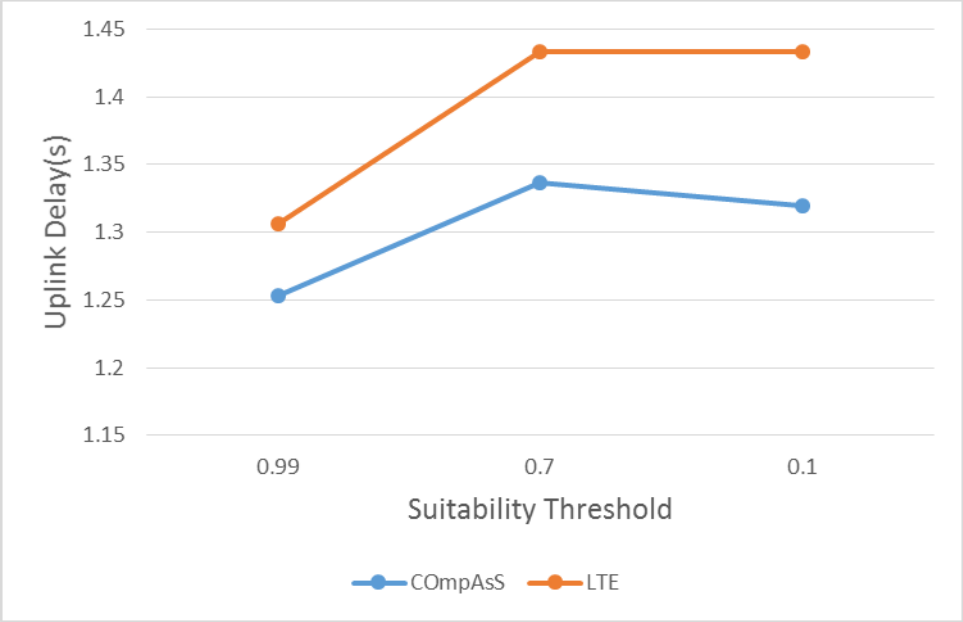


**Figure 10: Average Uplink Delay per Suitability Threshold**

### Scenario 2: Varying Suitability Margin

The second evaluation perspective illustrates the simulation outcomes in relation to the *Suitability margin,* in a similar pattern with the first scenario; the *margin* ranges between: 0.3, 0.1, and 0.001. As it can be inferred, the higher the *margin (*i.e. the difference between the current RAT's and the candidate target RAT's *Suitability* value)*,* the "stricter" requirements of COmpAsS in terms of *Suitability* of the successor RAT for handover.

As we can see in the figures below, as the *suitability margin* value decreases, the number of handovers in COmpAsS case increases at a small but steady pace, which shows the flexibility of COmpAsS regarding handover frequency, which is already low compared to the LTE case.

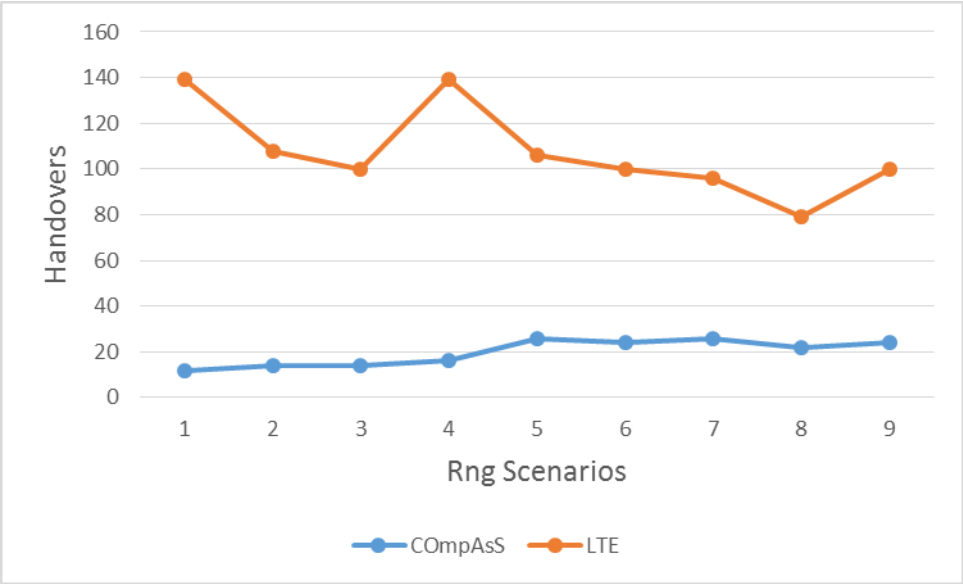


**Figure 11 Suitability Margin Scenario Handovers**

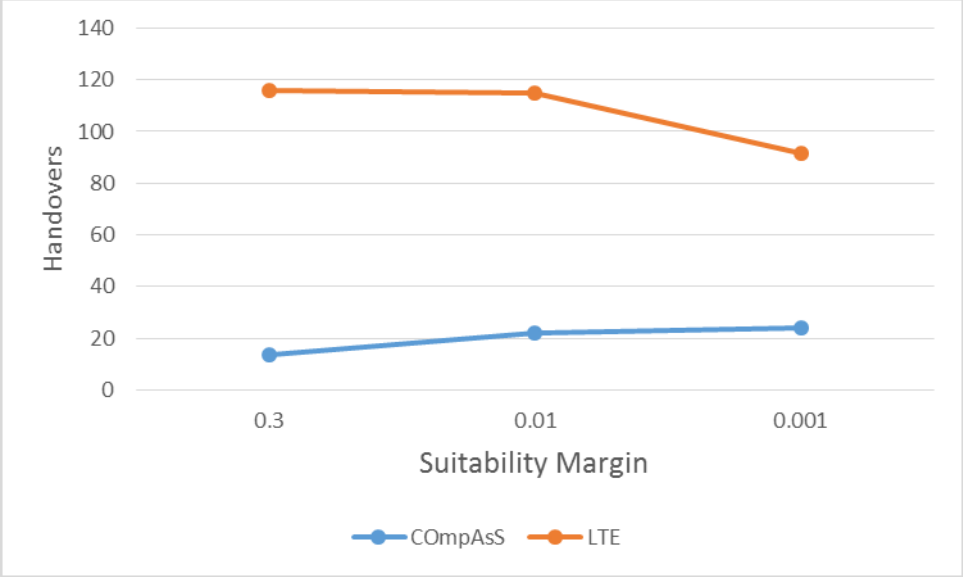


**Figure 12: Average Handovers per Suitability Margin**

In Figure 13 we can see the Uplink for varying *Suitability Threshold*. It is clear once again that COmpAsS has lower Uplink Delay, thus better performance, than the LTE algorithm in all experiments.

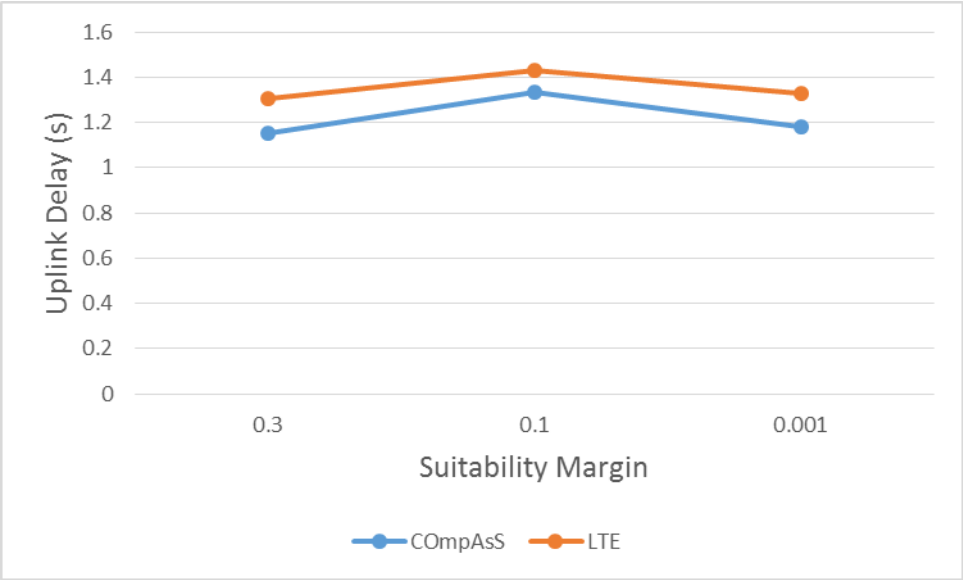


**Figure 13: Average Uplink Delay per Suitability Margin**

## Scenario 3: Deployment Density

In the context of the third evaluation scenario, we compare the two mechanisms in a varying environment, in terms of HeNBs' deployment density. The number of the femto cells ranges from 2 (sparse deployment), 10 and 50 (ultra-dense deployment); the latter resembles a typical 5G scenario as already discussed earlier. In these scenarios the *Suitability threshold* of COmpAsS algorithm is set to 0.7, while the *hysteresis* parameter at 0.1.

First we test for handover frequency. We can see that COmpAsS algorithm is more efficient in the increasing femto cell deployment, as the handovers remain quite stable whereas the LTE algorithm has a rapid increase in handover frequency as the HeNBs reach Ultra Dense Environment numbers, resulting thus, in excessive signaling; that would make the legacy LTE mechanisms inappropriate for such UDN scenarios, -on the contrary, COmpAsS seems capable of maintaining a stable performance-.

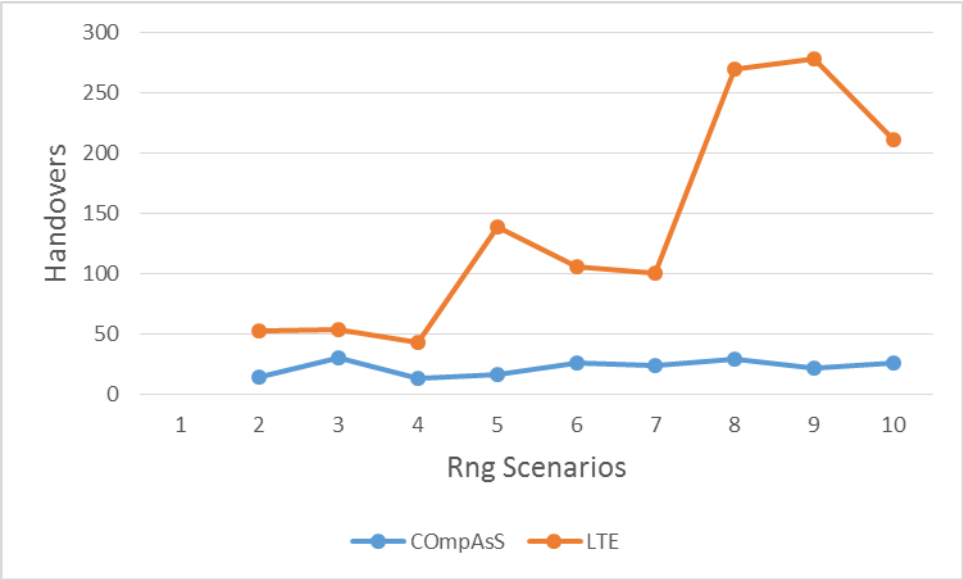


**Figure 14: Deployment Density Scenario Handovers**

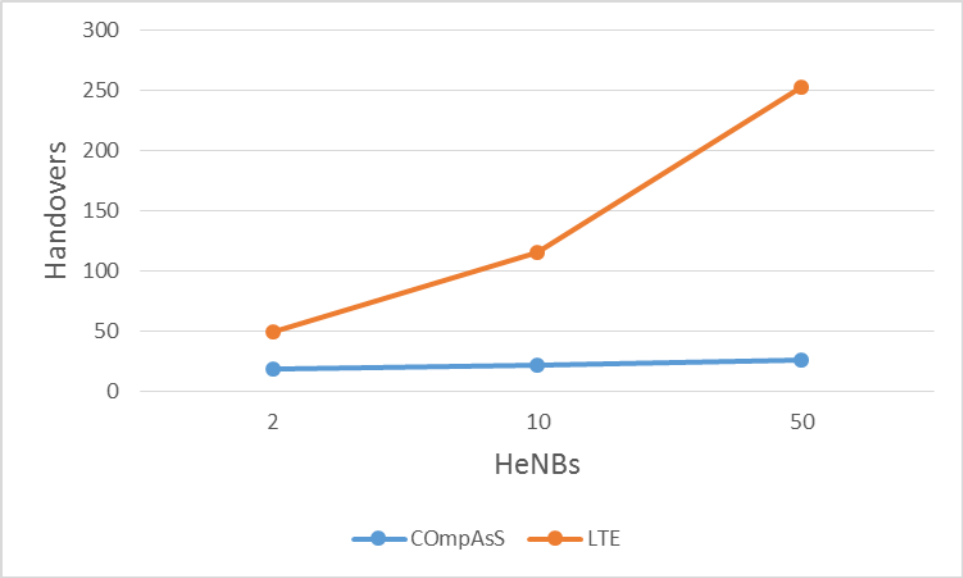


**Figure 15: Average Handovers per Deployment Density**

In terms of delay measurement the two algorithms were found to have similar uplink delay (Figure 16). Both mechanisms show a decreasing trend in terms of delay; especially, the denser the environment deployment is, the higher COmpAsS's gain is.

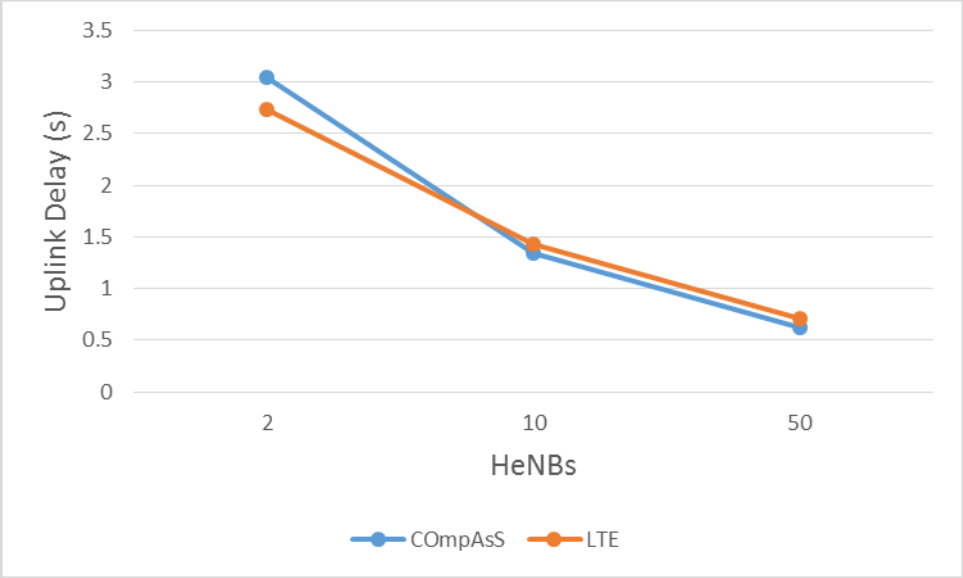


**Figure 16: Average Uplink Delay per Deployment Density**

## Scenario 4: Network Load

In the final round of experiments and in the context of the last scenario, we gradually increase the network load in terms of active bearers (active traffic flows) per UE, aiming at comparing the performance of our scheme in extreme load and interference conditions. In the particular set-up, we deploy 10 femto cells (co-existing with the fixed – throughout all experiment scenarios- number of macro cells).

By observing the handover results below we can say that both handover algorithms remain pretty unaffected by the increase in active bearers. However the COmpAsS algorithm as in most previous cases has a desirable much smaller handover frequency than the LTE algorithm.

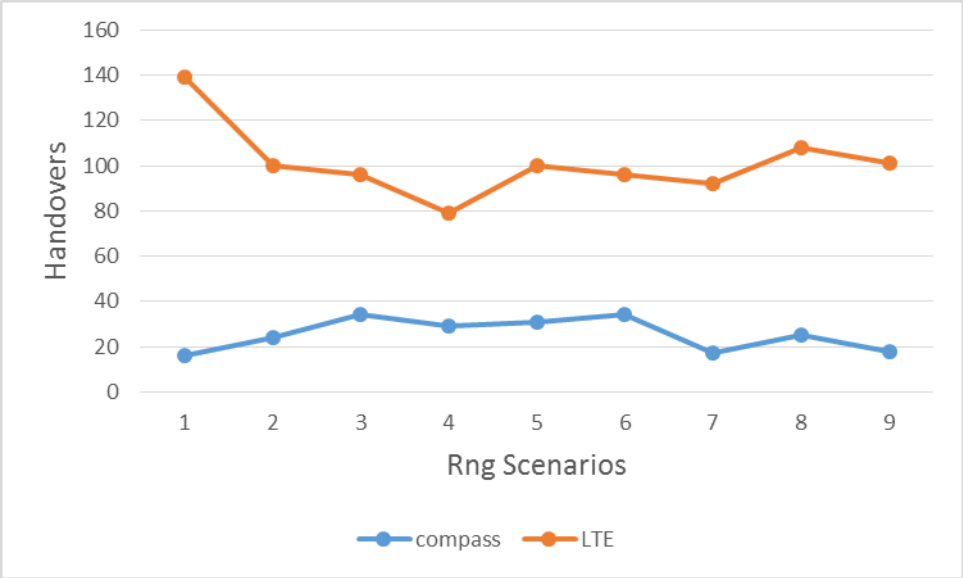


**Figure 17: Active Bearers Scenario Handovers**

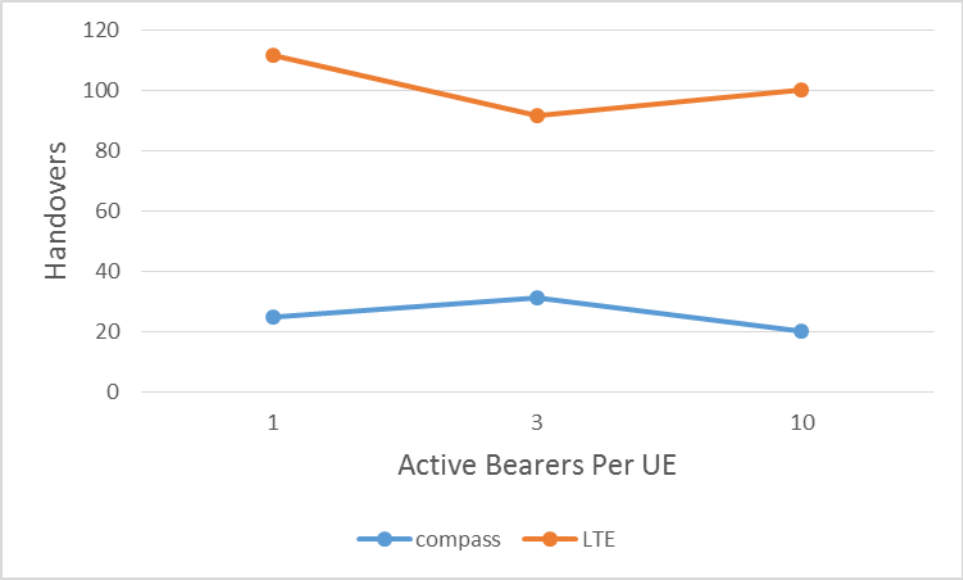


**Figure 18: Average Handovers per Active Bearer Scenario**

Regarding Uplink Delay for the increasing Network Load scenario the two algorithms have similar performance (Figure 19). Nevertheless, as the load gradually increases, there is a tendency for COmpAsS towards stable delay performance; on the contrary, LTE shows an increasing trend.

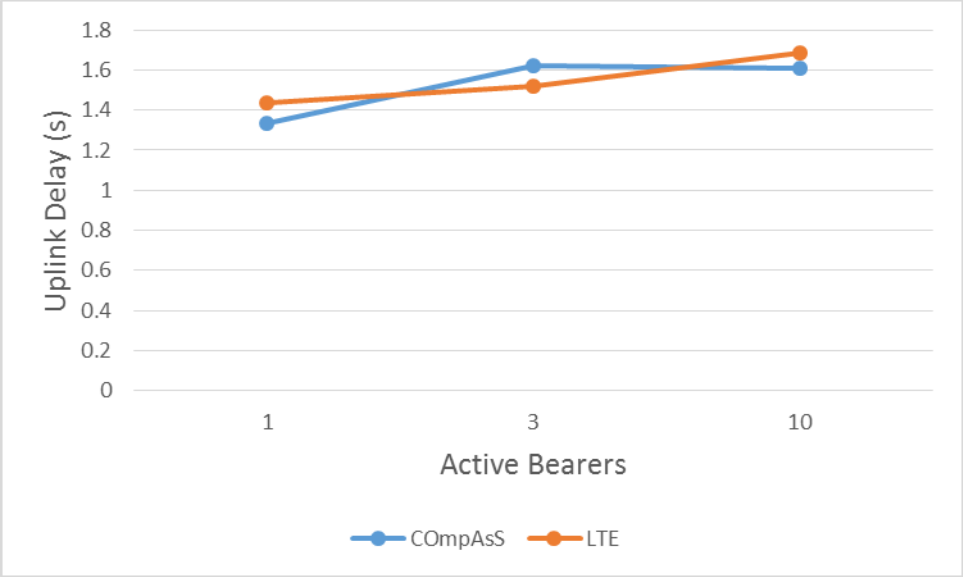


**Figure 19: Average Uplink Delay per Active Bearer Scenario**

## Handover Flexibility

Besides the aforepresented measurements we provide some supplementary results, which illustrate the types of handover that took place in the above scenarios. We logged the source and the target of the handovers in order to study the frequency of the handovers types that were taking place (i.e., macro to macro, macro to femto, femto to femto, etc.). Our final aim is to identify whether the handover types are distributed evenly between the femto cells and macro cells, or one of them was more prevalent.

The results for the four scenarios with COmpAsS can be found below (Figure 20). As we can see the COmpAsS experiments UEs are found to swap “smartly” between both femto and macro cells, selecting the most suitable cell each time. On the other hand in the LTE experiments the handovers were at an overwhelming degree between the femto cells (that’s why we omitted the diagrams). This displays the tendency of the LTE algorithm to ignore the larger macro cells and instead swap back and forth between the smaller and closer femto cells, in contrast to the more flexible COmpAsS algorithm.

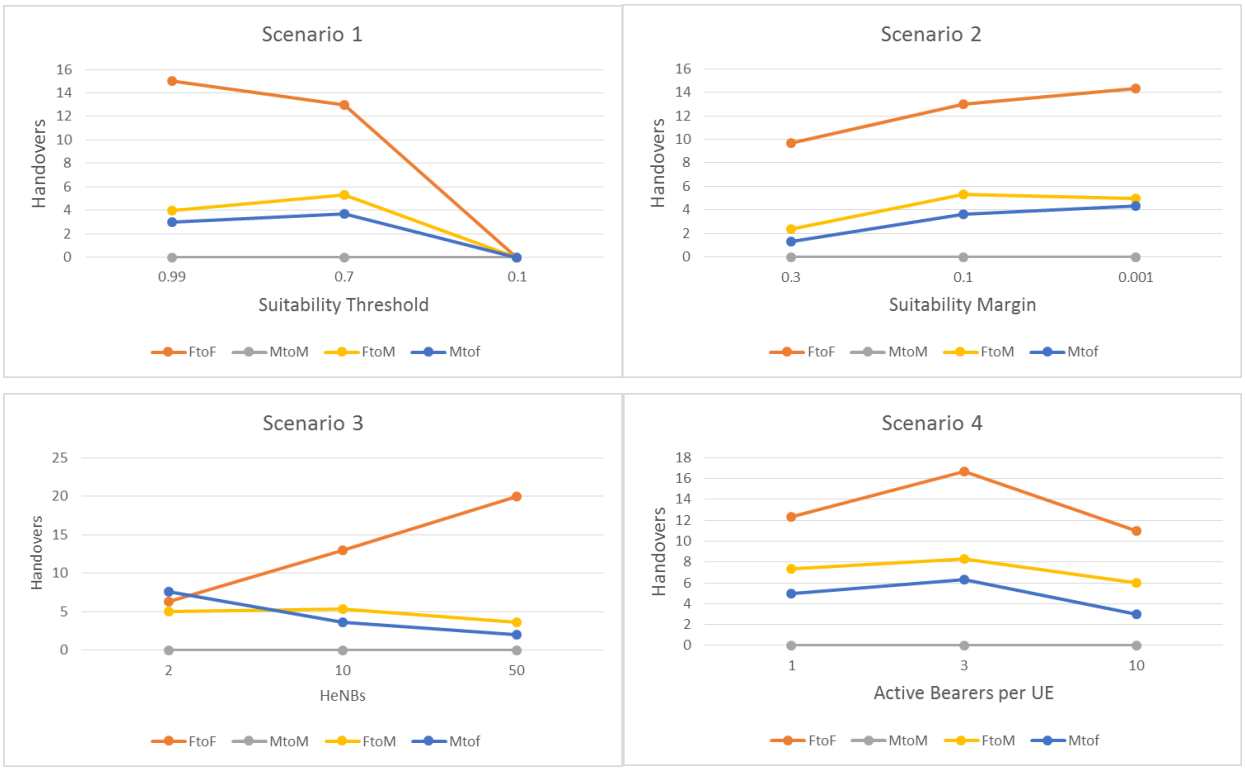


**Figure 20: COmpAsS Handover Types**

# 5. CONCLUSION

In this paper we presented a comprehensive study in terms of network performance for COmpAsS, a framework for RAT selection in 5G ultra-dense network environments based on a context-aware scheme. By collecting the necessary context information, taking advantage of latest advancements and 3GPP trends in the LTE-EPC architecture (e.g. ANDSF functionality) and using fuzzy logic, the UE is able to evaluate the available RATs and identify the most suitable one in any case.

Through an extensive literature research we demonstrated the latest advancements in the EPC – WiFi integration, Ultra Dense Networks and state of the art proposals for Context Awareness in Mobility Management in LTE networks.

We carried out and presented a detailed analysis of the signaling overhead of the proposed mechanism by linking it to the current 3GPP specifications. Experimentation in a simulated realistic 5G network environment was provided in order to assess the viability of our proposal compared to the LTE algorithm. We displayed the superiority of our mechanism regarding handover frequency and fundamental network KPIs resulting in higher service quality and -eventually- higher quality of experience for the end-user.

# ABBREVIATIONS - ACRONYMS

| | |
|---|---|
| LTE | Long-Term Evolution |
| RAT | Radio Access Technology |
| EPC | Evolved Packet Core |
| SAE | System Architecture Evolution |
| QoS | Quality of Service |
| QoE | Quality of Experience |
| 3GPP | 3rd Generation Partnership Project |
| eNB | E-UTRAN Node B |
| HenB | Home eNodeB |
| UDN | Ultra Dense Network |
| AP | Access Point |
| RN | Relay Node |
| RRH | Remote Radio Head |
| BS | Base Station |
| UE | User Equipment |
| S-GW | Serving Gateway |
| P-GW | Packet Data Network Gateway |
| PDN | Packet Data Network |
| GPRS | General Packet Radio Service |
| OS | Operation System |
| KPI | Key Performance Indicator |
| IP | Internet Protocol |
| VoIP | Voice over Internet Protocol |
| FTP | File Transfer Protocol |
| Ns | Network Simulator |
| BTS | Base Transceiver Station |
| MSC | Mobile Switching Center |
| PSTN | Public Switched Telephone Network |
| BSC | Base Station Controller |
| VLR | Visitor Location Register |
| HLR | Home Location Register |

| | |
|---|---|
| EPS | Evolved Packet System |
| GSM | Global System for Mobile Communications |
| TDMA | Time Division Multiple Access |
| WCDMA | Wideband Code Division Multiple Access |
| UMTS | Universal Mobile Telecommunications System |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| MME | Mobility Management Entity |
| CN | Core Network |
| HSS | Home Subscriber Server |
| ANDSF | Access Network Discovery and Selection Function |
| ePDG | Evolved Packet Data Gateway |
| WISP | Wireless Internet Service Provide |
| RAN | Radio Access Network |
| HSPA | High Speed Packet Access |
| MAPCON | Multi-Access PDN Connectivity |
| IFOM | IP Flow Mobility |
| PMIP | Proxy Mobile IP |
| WLAN | Wireless Local Area Network |
| TCP | Transmission Control Protocol |
| Ipsec | Internet Protocol Security |
| IKE | Internet Key Exchange |
| EAP-AKA | Extensible Authentication Protocol Method-Authentication and Key Agreement |
| APN | Access Point Name |
| PLMN | Public Land Mobile Network |
| DL | Downlink |
| UL | Uplink |
| SIM | Subscriber Identity Module |
| TWAG/TWAP | Trusted WLAN Access Gateway/Proxy |
| GTP | GPRS Tunneling Protocol |
| ETSI | European Telecommunications Standards Institute |
| MIP | Mobile IP |
| MAG | Mobile Access Gateway |

| | |
|---|---|
| LMA | Local Mobility Anchor |
| HS | Hotspot |
| VPN | Virtual Private Network |
| WPA | WiFi Protected Access |
| AN | Access Node |
| DAS | Distributed Antenna Systems |
| FTTC | Fiber To The Cell |
| HetNet | Heterogeneous Network |
| NCL | Neighboring Cell List |
| REB | Range Expansion Bias |
| MAB | Mult-Armed Bandit |
| MM | Mobility Management |
| RRC | Radio Resource Control |
| TAU | Tracking Area Update |
| SDN | Software Defined Networking |
| RSRQ | Reference Signal Received Quality |
| FLC | Fuzzy Logic Controller |
| SSID | Service Set Identifier |
| | |
| | |
| | |
| | |

# REFERENCES


[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] Hakiri, Akram and Pascal Berthou. "Leveraging SDN for The 5G Networks: Trends, Prospects and Challenges." Book chapter in "Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture", 2015 John Wiley & Sons, Ltd.

[3] J. Wannstrom, “LTE-Advanced”, May 10, 2012, http://www.3gpp.org/IMG/pdf/lte_advanced_v2.pdf [accessed 06/2016].

[4] Nokia, “Smart WiFi Traffic Steering”, December 2015

[5] BT & Alcatel Lucent White paper, WiFi Roaming building on ANDSF and HOTSPOT 2.0, October 2012.

[6] 3GPP TS 29.060, V11.5.0, “General Packet Radio Service (GPRS) Tunnelling Protocol (GTP) across the Gn and Gp interface”, Release 11, December 2012.

[7] 3GPP Specification Groups, http://www.3gpp.org/specifications-groups

[8] 3GPP TS 23.401, V13.6.0, “GPRS enhancements for E-UTRAN access “, Release 13, March 2016.

[9] 3GPP TS 23.402, V13.5.0, “Architecture enhancements for non-3GPP accesses”, Release 13, June 2013.

[10] 3GPP TS 24.312 Access Network Discovery and Selection Function (ANDSF) Management Object (MO)

[11] 3GPP TS 29.060 General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface

[12] 3GPP TS 29.275 Proxy Mobile IPv6 (PMIPv6) based Mobility and Tunnelling protocols

[13] 3GPP TS 23.852, “Study on S2a Mobility based on GPRS Tunnelling Protocol (GTP) and Wireless Local Area Network (WLAN) access to the Enhanced Packet Core (EPC) network (SaMOG)”, September 2013.

[14] 4G Americas White Paper, “Mobile Broadband Evolution Toward 5G: Rel. 12 & Rel.13 and Beyond”, June 2015

[15] Apple iOS Connection Manager, https://support.apple.com/en-us/HT202831 [accessed June 2016]

[16] 3GPP TR 36.839 “Evolved Universal Terrestrial Radio Access (EUTRA); Mobility Enhancements in Heterogeneous Networks

[17] S. Barbera, P. H. Michaelsen, M. SÃd’ily, and K. Pedersen, “Mobility Performance of LTE Co-Channel Deployment of Macro and Pico Cells,” in Proc. IEEE Wireless Comm. and Networking Conf. (WCNC), France, Apr. 2012.

[18] LTE – The UMTS Long Term Evolution From Theory to Practice Second Edition S.Sessia

[19] 3GPP TR 23.861 Network based IP flow mobility V13

[20] Ultra Dense Networks Nokia White paper, 2016

[21] D. Lopez-Perez, I. Guvenc, and X. Chu, “Mobility Management Challenges in 3GPP Heterogeneous Networks,” IEEE Comm. Mag., vol. 50, no. 12, Dec. 2012

[22] X. Duan and X. Wang, "Authentication handover and privacy protection in 5G hetnets using software-defined networking," in IEEE Communications Magazine, vol. 53, no. 4, pp. 28-35, April 2015.

[23] H. Song, X. Fang and L. Yan, "Handover Scheme for 5G C/U Plane Split Heterogeneous Network in High-Speed Railway," in IEEE Transactions on Vehicular Technology, vol. 63, no. 9, pp. 4633-4646, Nov. 2014.

[24] J. S. Thainesh, N. Wang and R. Tafazolli, "A scalable architecture for handling control plane failures in heterogeneous networks," in IEEE Communications Magazine, vol. 54, no. 4, pp. 145-151, April 2016.

[25] Y. Kim et al., "Feasibility of Mobile Cellular Communications at Millimeter Wave Frequency," in IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 589-599, April 2016.

[26] H. Zhang, C. Jiang, J. Cheng and V. C. M. Leung, "Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks," in IEEE Wireless Communications, vol. 22, no. 3, pp. 92-99, June 2015.

[27] H. Peng, Y. Xiao, Y. N. Ruyue and Y. Yifei, "Ultra dense network: Challenges, enabling technologies and new trends," in China Communications, vol. 13, no. 2, pp. 30-40, Feb. 2016.
[28] S. Bellahsene and L. Kloul A New Markov-Based Mobility Prediction Algorithm for Mobile Networks, Conference on Computer Performance Engineering, Bertinoro, Italy, 2010.
[29] A. Hadachi, O. Batrashev et al., Cell Phone Subscribers Mobility Prediction Using Enhanced Markov Chain Algorithm, Intelligent Vehicles Symposium, Michigan, USA, 2014
[30] S.Michaelis, N. Piatkowski and K. Morik Predicting next network cell IDs for moving users with Discriminative and Generative Models, Mobile Data Challenge Workshop in conjunction with International Conference on Pervasive Computing, Newcastle, UK, 2012
[31] K. Laasonen, Clustering and Prediction of Mobile User Routes from Cellular Data, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD),Porto, Portugal, 2005
[32] Z. Becvar, P. Mach and B. Simak, Improvement of handover prediction in mobile WiMAX by using two thresholds, The International Journal of Computer and Telecommunications Networking, 2011.
[33] N.P. Kuruvatti; A. Klein; H.D. Schotten, Prediction of Dynamic Crowd Formation in Cellular Networks for Activating Small Cells, VTC-Spring, Glasgow,UK,2015
[34] N.P. Kuruvatti, A. Klein, J. Schneider and H.D. Schotten, Exploiting Diurnal User Mobility for Predicting Cell Transitions, IEEE Globecom workshops, Atlanta, USA, 2013
[35] N.P. Kuruvatti, H.D. Schotten Framework to Support Mobility Context Awareness in Cellular Networks University of Kaiserslautern 2016
[36] 3GPP TS 36.423 E-UTRAN X2 application protocol (X2AP) Release 14, 2017
[37] 3GPP TS 36.331 E-UTRA Radio Resource Control (RRC) protocol specification Release 14, 2016
[38] M. Simsek, M. Bennis, and I. Güvenç, Context-Aware Mobility Management in HetNets: A Reinforcement Learning Approach 2015
[39] S.S. Liao, J. W. He, T.H. Tang, “A framework for context information management”, JIS journal, vol.30, no.6, pp. 528-539, December 2004
[40] H. Peizhao, J. Indulska, R. Robinsion, “An Autonomic Context Management System for Pervasive Computing”, PerCom 2008, 6th Annual IEEE international Conference on Pervasive Computing and Communications, pp. 213-223, Hong Kong, March 2008
[41] K. Svanbro, J. Allden, “Communication System and Method for shared context compression”, Patent No US6950445 B2, Publication date: September 27th, 2005
[42] M. C. Chuah, “Header compression for general packet radio service tunneling protocol (GTP)-encapsulated packets”, Patent No US 6839339 B1, Publication date: January 4th, 2005
[43] M. Luna, M. Tervahauta, “Mobile network background traffic data management with optimized polling intervals”. Patent No US 8539040 B2, Publication date: September 17th, 2013
[44] A. Anand, J. Subtamanian, “Redundant traffic reduction in wireless networks”, Patent No WO 2014146755 A1, Publication date: September 25th, 2014
[45] G. S. Banavar, M. R. Ebling, G. D. H. Hunt, et al., “Method and Apparatus for Content Pre-Fetching and Preparation”, Patent No US 20090287750 A1, Publication date: November 19th, 2009
[46] P. Andres-Maldonado, P. Ameigeiras, Reduced M2M Signaling Communications in 3GPP LTE and Future 5G Cellular Networks 2016
[47] Nokia, “What is going on in Mobile Broadband Networks? Smartphone Traffic Analysis and Solutions,” White paper, 2014.
[48] S. M. Ravazi Reducing Signaling Overhead by Overlapping Tracking Area List in LTE
[49] https://en.wikipedia.org/wiki/Gnuplot
[50] http://www.xgraph.org/
[51] METIS EU Project, https://www.metis2020.com
[52] 3GPP TR 36.931 v13.0.0, “Radio Frequency (RF) requirements for LTE Pico Node B”, Release 13, January 2016
[53] 3GPP TS 36.921 v13.0.0, “FDD Home eNode B (HeNB) Radio Frequency (RF) requirements analysis”, Release 13, January 2016
[54] 3GPP 36.304, V.13.1.0, “User Equipment (UE) procedures in idle mode”, Section 5.2.4.3 “Mobility states of a UE”, Release 13, March 2016.
[55] IEEE 802.11u-2011, “IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and Metropolitan networks-specific requirements” - Amendment 9: Interworking with External Networks.
[56] 3GPP, TR 23.865, V12.1.0, “Study on Wireless Local Area Network (WLAN) network selection for 3GPP terminals; Stage 2”, Release 12, December 2013.

[57] A. Kaloxylos, S. Barmpounakis, P. Spapis, N. Alonistioti, “An efficient RAT selection mechanism for 5G cellular networks”, International Wireless Communications and Mobile Computing Conference, 4-8 August 2014, Nicosia, Cyprus

[58] S. Barmpounakis, A. Kaloxylos, P. Spapis, N. Alonistioti, “COmpAsS: A Context-Aware, User-Oriented RAT Selection Mechanism in Heterogeneous Wireless Networks”, Mobility 2014, Fourth International Conference on Mobile Services, Resources, and Users, July 20-24 – 2014, Paris, France

[59] S. Barmpounakis, A. Kaloxylos, P. Spapis, N. Alonistioti, “Context-aware, user-driven, network-controlled RAT selection for 5G networks”, Computer Network Journal, Elsevier, Vol. 113, February 2017, pp. 124–147

[60] 3GPP 23.890, V 12.0.0, “Optimized offloading to Wireless Local Area Network (WLAN) in 3GPP Radio Access Technology (RAT) mobility”, September 2013.