Evolutionary 4G/5G Network Architecture assisted Efficient Handover Signaling

Akshay Jain, *Student Member*, Elena Lopez-Aguilera, and Ilker Demirkol, *Senior Member*

***Abstract*—Future wireless networks are expected to be ultra-dense and heterogeneous not just in terms of the number and type of base stations, but also in terms of the number of users and application types they access. Such a network architecture will require mobility management mechanisms that adapt rapidly to these highly dynamic network characteristics. In particular, the optimality of the handover signaling within these future network architectures will be extremely critical given their density and heterogeneity. Here the optimality is relevant for both the total amount of signaling created and the total delay per handover process. In this article we firstly present a novel and optimized message mapping and signaling mechanism for the handover preparation and failure phases. We also develop a novel handover failure aware preparation signaling methodology, which accounts for the possibility of a handover failure and grants additional enhancements to the handover preparation and failure signaling phases. Through the analytical framework provided in this article we conduct studies to quantify the performance gains promised by the proposed mechanisms. These studies cover myriad handover scenarios as identified by 3GPP, and use the statistics from cellular network operators and vendors. We then develop the idea and analytical framework for network wide analysis, wherein the network wide processing cost and network occupation time for various handover failure rates are computed. Lastly, we propose an evolutionary network architecture that facilitates the proposed signaling mechanism as well as assists operators in maintaining a manageable Capital Expenditure (CAPEX). It combines the current day and 3GPP proposed 5G network architecture with the Software Defined Networking (SDN) approach. As a result, we argue that the proposed mechanisms are viable and outperform the legacy handover signaling mechanisms in terms of latency incurred, total network occupation time, number of messages generated and total bytes transferred.**

***Index Terms*—5G, 4G, LTE, 3GPP, SDN, Handover, Mobility Management, MME**

TABLE I
ABBREVIATIONS LIST

RSSI	Received Signal Strength Indicator
AP	Access Point
QoE	Quality of Experience
QoS	Quality of Service
HO	Handover
FHO	Frequent Handover
RAT	Radio Access Technology
SDN	Software Defined Networking
MME	Mobility Management Entity
SMF	Session Management Function
SeMMu	SDN enabled Mobility Management unit
AMF	Access and Mobility management Function
SDN-C	SDN-Controller
eNB	Evolved Node B
RACH	Random Access Channel
TDD	Time Division Duplex
FDD	Frequency Division Duplex
VLAN	Virtual LAN
CP	Control Plane
DP	Data Plane
RRM	Radio Resource Management
TEID	Tunnel Endpoint Identifiers
SGSN	Serving Gateway Support Node
UPF	User Plane Function
NG-RAN	Next Generation Radio Access Network
T-SeMMu	Target SeMMu
LTE	Long Term Evolution
EPC	Evolved Packet Core
UMTS	Universal Mobile Terrestrial System
gNB	next generation NodeB
NGC	Next Generation Core
S-GW	Serving Gateway
ITU	International Telecommunications Union
ASN.1	Abstract Syntax Notation number one
CAPEX	Capital Expenditure
RNC	Radio Network Controller
IE	Information Element
5G NR	5G New Radio
CN	Core Network
E-UTRAN	Evolved Universal Terrestrial Radio Access Network

I. INTRODUCTION

One of the most challenging requirements for a wireless network to be ubiquitous is the ability to permit mobility – whenever and wherever – without loss of quality of service and connectivity. Such mobility support allows users, with different mobility profiles, to traverse different geographical areas while accessing a myriad of mobile applications. Central to the solutions that offer the aforesaid mobility management support are handover mechanisms, which allow a user to change its physical point of attachment within the network when it is subject to a mobility event and certain pre-programmed conditions are satisfied [1]–[3]. For example, if the RSSI or the received signal power from the current serving base station goes below a particular threshold and, simultaneously if the same parameter for another base station in the vicinity goes above a certain threshold, then a decision to change the point of attachment, i.e. the base station, can be taken by the network or the user.

As compared to the current network scenario, the future 5G network scenarios will be much more complex. Contributing towards this increased network complexity, will be the burgeoning demand for high quality data by users/devices,

This work has been supported in part by the EU Horizon 2020 research and innovation programme under grant agreement No. 675806 (5GAuRA), by the ERDF and the Spanish Government through project RYC-2013-13029, and by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya.

We would like to thank OTE Group of Companies, S.A., Athens, Greece, for providing delay information in 4G architecture.

Akshay Jain, Elena Lopez-Aguilera and Ilker Demirkol are with the Department of Network Engineering, Universitat Politècnica de Catalunya BarcelonaTECH, Barcelona 08034, ES. Email: akshay.jain@upc.edu, elopez@entel.upc.edu, and ilker.demirkol@entel.upc.edu. Akshay Jain is the corresponding author.

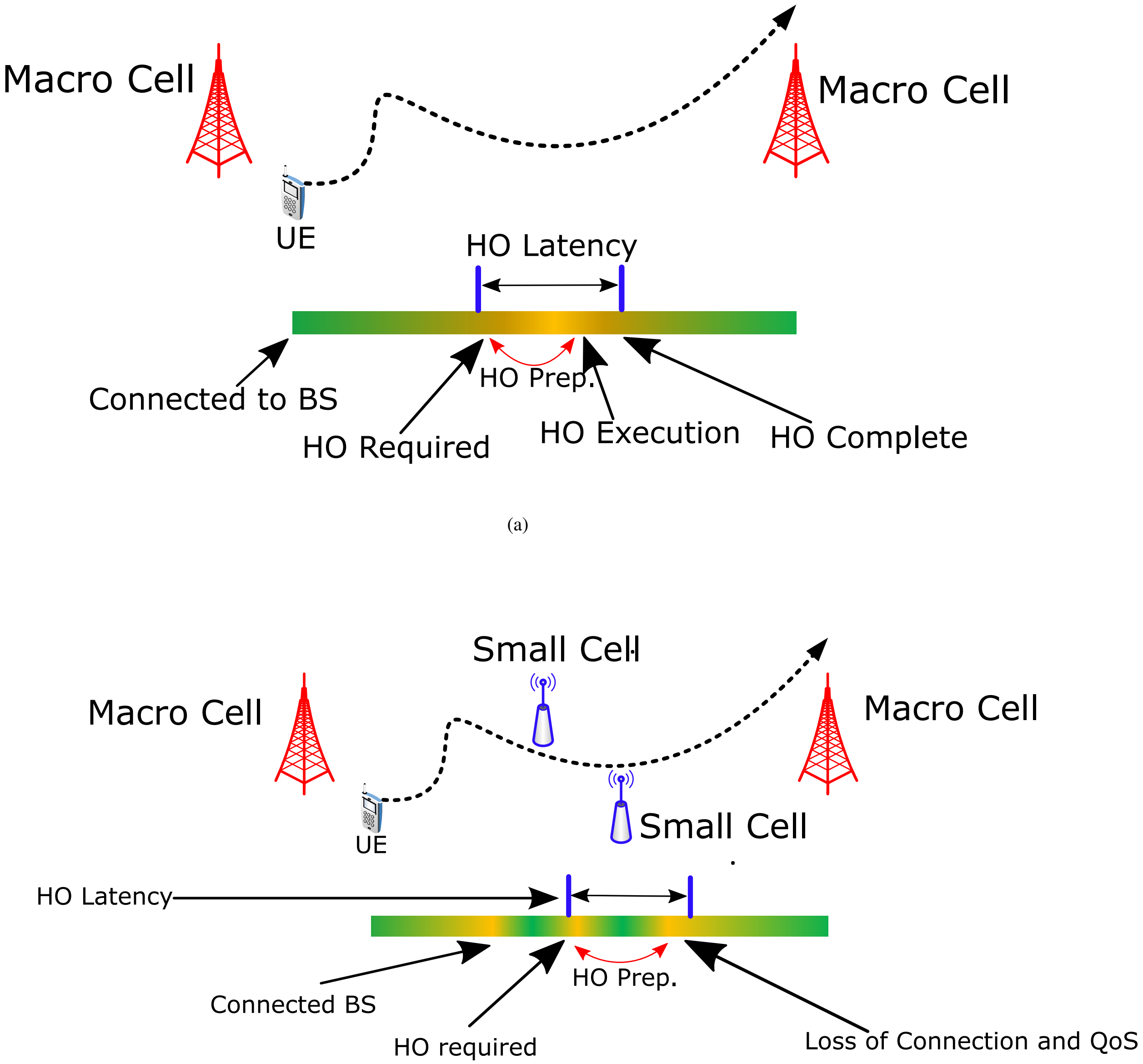
which are also expected to increase in numbers exponentially [4], [5]. Additionally, with the presence of heterogeneity of applications (Virtual Reality, Gaming, Messaging, Video playback, etc.), radio access technologies (5G NR, LTE, 3G, 2G, Wi-Fi) and mobility profiles (high speed (300-500 km/h), urban mobility (30-70 km/h), pedestrian (0-3 km/h), etc.) along side the possibility of multi-connectivity, the complexity of the 5G network is further exacerbated.

Given such a scenario, in this article we revisit the legacy handover mechanisms, which form a critical part of mobility management. These legacy handover mechanisms are composed of four phases, i.e. handover decision (parameter values, such as RSSI, etc., based decision for AP selection), handover preparation (resource negotiation and allocation involving source and target networks), handover execution/rejection/cancel (path re-routing with the user transitioning from source to target network, or issuance of a cancel/reject indication due resource allocation failure) and handover complete (release of source network resources upon successful handover to target network). Each of these stages contribute towards the overall latency and signaling cost to execute the handover. Hence, optimizing/enhancing them will facilitate in improving the QoE and QoS to the device/user.

Consequently, many current research efforts, such as [6]–[14], have provided studies and methods that will facilitate the enhancement/optimization of the aforesaid handover phases. However, the handover preparation and failure phases remain relatively unexplored in the studies referenced above and similar to them. It is during the handover preparation phase that the negotiation and allocation of resources for an impending HO is carried out. Further, during the handover failure phase, signaling that involves sending an indication to the source network and the user undergoing HO with regards to the failed HO attempt is performed. Thus, fast execution of the aforesaid signaling, in a markedly more complex network environment, will be a vital requirement for an efficient next-generation HO management framework. This requirement is further elaborated via the current and future network scenarios, and their corresponding HO phases, illustrated in Fig. 1.

In the current network scenarios, depicted in Fig. 1(a), a user has significant time to trigger, prepare, execute and complete a handover. However, the same is not true for the future scenarios, as illustrated in Fig. 1(b). In the future network scenarios, the density of access points will be high, i.e. access points with smaller coverage areas (small cells) and higher bandwidths will be packed more closely in a given area. In addition, macro-cells with significantly large coverage areas will be existent as well, to assist the small cells. Hence, if current handover management strategies are utilized, the time taken to complete the handover will be much greater than the dwell time of the user (with its mobility profile) at the desired base station whilst the conditions are still favorable to establish a link. Specifically, the time available to perform the resource allocation and negotiation process will be shorter. Such a scenario, would thus lead to loss of connectivity and hence, a poor network performance. Moreover, the HO signaling overhead will also be of critical importance for the network performance because of the FHOs caused by cell densification, and the diverse RATs used resulting in many inter-RAT HOs. Hence, an optimized handover process, where the HO latency and signaling overhead are reduced, will be an extremely vital component of future handover management strategies. Concretely, and from the discussion above, a fast and efficient handover preparation phase signaling will be important for an optimized handover process. Further, in the event that the HO has failed, the CN has to ensure a fast release of the allocated resources so that they can be reused, as well as the user under consideration is free to choose another access point. Thus the HO failure signaling process should also be optimized.

With this motivation, in this paper we have introduced a novel message mapping and parallelized control signaling methodology for the preparation and failure phase scenarios studied by the 3GPP [2], [15]. We follow this approach for both the legacy as well as the 5G networks. This approach subsequently results in a reduction of the overall transmission cost, processing cost, latency as well as the number of bytes transferred during a handover event. Further, the proposed approach has been designed not only for the Intra-RAT HO scenarios, but also for the scenarios involving inter-RAT HOs including 5G HO scenarios discussed by 3GPP. In addition, in this paper we have introduced a novel HO failure aware preparation phase signaling mechanism. The proposed mechanism accounts for the possibility of a HO failure during the design and execution of the HO preparation phase. And as will be seen in Section IV.C, the HO failure aware mechanism not only enhances the HO failure step but it also presents additional enhancement for the HO preparation signaling step. Thus, the current article significantly advances our preliminary works performed in [16], [17]. Subsequently, in this paper, we have provided a simple yet rigorous analytical methodology to validate the proposed mechanism. The aforementioned methodology enables the reader to compare the performance of the proposed mechanism with the legacy mechanisms, i.e. 3GPP standards, on the basis of of latency incurred, transmission cost (i.e., total network occupation time), processing cost (number of messages generated) and the overall amount of bytes transferred. Note that, through the message size analysis we are able to determine the reduction in the overall amount of bytes transferred through the network during a given HO preparation or failure signaling sequence. It is important to state here that the packet or system level simulations would not be able to derive realistic network parameter values, since the network topology, the transport technology used, queueing at the network elements, etc. is dependent on the specific operator scenario and cannot be modeled accurately. Hence, we have utilized real data from network operators and vendors and have attempted to provide a simplistic, realistic, and yet holistic analysis. We have also introduced a novel network wide analysis. Through this we establish the fact that, given any distribution over the number of HOs for the studied HO types and for any HO failure rate, the proposed methodology greatly improves the system performance in terms of overall processing cost and total network occupation time, as compared to the legacy mechanisms. We defer any further discussion on the details for the aforesaid network wide analysis to section V.



(a)

(b)

Fig. 1. a) Handover scenario in current wireless networks; b) Handover scenario in future wireless networks.

Next, the aforesaid message mapping and parallelized control information transfer is facilitated via an evolutionary network architecture. This network architecture establishes evolved CN entities wherein they are integrated with an SDN agent. Moreover, the MME in the 4G network and the SMF in the 5G network are evolved to SDN enabled CN entities. We refer to them as SeMMu. The reason for such an integration being that the MME and SMF are responsible for the CN signaling during a mobility event in their respective networks. Hence, this allows the SeMMu to facilitate the proposed HO signaling mechanism. Further, in this work, instead of the the AMF (which 3GPP defines as the mobility management unit in the 5G NGC), we exploit the idea of SDN-enabled SMF because it is the SMF which is involved in HO-related CN signaling whereas the AMF is connected to the access network only. Thus, the HO-related CN signaling is not influenced by the AMF. Accordingly, in this paper, we propose such evolutionary network architecture, by also describing the implementation aspects of such SeMMu entities.

And so, this paper is organized as follows: in Section II we firstly provide a brief state of the art survey, with a specific focus on handover signaling management and analysis. Then, in Section III and IV we present the legacy and proposed signaling mechanisms. We identify the drawbacks of the legacy mechanism and then present the novel message mapping and signaling sequence. In addition, in Section IV, we provide a novel handover failure aware preparation signaling design. To validate these proposed mechanisms, in Section V we elaborate the analytical framework utilized. Further, we present the latency, transmission cost, processing cost and a message size analysis for both the legacy and proposed handover preparation and failure phases. We then perform a network wide processing cost analysis and network occupation time analysis that exemplifies the superiority of the proposed method over the current methods. Lastly, in Section VI, we carry out a detailed discussion into the novel and evolutionary network architecture. We present its benefits, requirements for performing the integration between the SDN-C and the corresponding CN entity that manages mobility, and also the implementational challenges that exist. Following this discussion, the paper is concluded in Section VII.

II. STATE OF THE ART

First, we consider the current and past research efforts that provision a comprehensive study into the main stages of handover management, i.e., *handover decision*, *handover preparation*, *handover execution/failure* and *handover completion*. Notably, [10]–[14] provide sufficient background and analysis into these different stages. In [10] a detailed survey on the various aspects of handover management such as execution phase, decision phase and system information collection has been provided. Concretely, for the network discovery phase, which is the same as acquiring measurement reports from users for handover decision phase, various parameters such as network congestion, channel conditions, etc., have been discussed. Next, for the handover decision making phase, techniques involving multi-attribute decision making, user-centric decision making, etc., have been explored in detail. Following this, for the handover execution phase, methods such as mobile-assisted, network-assisted, etc., have been considered in [10]. Further, in [11] an analysis of the interruption time during the handover phase in an LTE-Advanced network has been performed. Note that, the specific stage of handover phase that has been enhanced in [11] is the handover execution stage. Further, an analysis with TDD and FDD modes has been considered for the same.

Next, in [12]–[14], SDN based approaches have been considered for mobility management. Specifically, in [12], the SDN controller along side a double V-LAN tagging approach is utilized to minimize the path switching operation which would reduce the latency and core network signaling. In [13] a policy and per-flow based mobility management approach has been presented, wherein the flow level granularity of service provision along side policies specified by network, user, applications, etc., are considered for executing the mobility management task. By policies, here we mean a collection of network, user and application parameters that are utilized in generating a handover decision. Additionally, in [14], an approach towards the seamless mobility management between LTE and W-LAN networks has been provided. This approach involves splitting the CP and DP, via SDN based approach, and migrating the CP to the cloud based infrastructure. Hence, with the help of the global view of the network, the controller can facilitate seamless mobility for the user between heterogeneous networks, i.e., LTE and W-LAN, through the specified lightweight route reconfiguration procedures.

However, most research efforts similar to [10]–[14] do not emphasize on the criticality of handover preparation and failure phase. Additionally, they do not explore their latency, transmission cost and processing cost contribution to the overall handover management operations. And hence, in this article we explore these aspects of handover management and propose enhanced signaling strategies for handover preparation and failure signaling phases. Further, while [18] proposes an integration of the MME with the SDN-C and its utility for handover management using IEEE 802.1ad CN signaling, it does not present an evolutionary mechanism such that the operator CAPEX can be manageable. In addition, the system design focuses on the 3GPP-LTE architecture and does not consider the currently proposed 5G network architecture. Further, research efforts such as [19], [20], present an SDN and NFV based evolutionary network framework for the LTE-EPC. However, like [18], they do not encompass the inter-working architecture with other 3GPP defined technologies such as 3G and the newly defined 5G architecture as well. In contrast, through this article, we propose an evolutionary architecture, considering the co-existence of the legacy networks (4G, 3G, 2G) and the newly proposed 5G networks, that enables a manageable CAPEX for the operators whilst also enhancing the handover preparation and failure phase performance. It is important to state here that, although our preliminary works such as [16], [17] utilize an evolutionary architecture that enhances the handover preparation phase for the legacy as well as the 5G networks (only when there is no N26 interface), in the current work we advance these and other research efforts, discussed in the state of the art, by:

- Introducing enhanced HO preparation and failure signaling phases for the myriad 5G and legacy networks inter-RAT HO scenarios.
- Introducing a novel HO failure aware preparation phase signaling sequence. This approach will optimize the legacy HO failure signaling step as well as enhance the HO preparation step further, as compared to the mechanism proposed in [16].
- Presenting performance analysis, based on latency, transmission cost and processing cost, of the proposed and legacy signaling mechanisms for the myriad 5G and legacy network HO scenarios specified by 3GPP.
- Introducing a novel message size analysis for the proposed as well as legacy HO scenarios.
- Introducing a novel network wide analysis in terms of the number of messages processed as well as the total network occupation time.
- Presenting a novel 5G NGC and legacy inter-working architecture with the capabilities of an N26 interface [1].
- Presenting a novel interfacing mechanism between the MME/SMF and the SDN agent.

---

[1] The N26 interface, defined by 3GPP, allows for the inter-working between the 5G and legacy networks. It allows for reduced signaling to prepare or reject a HO between 5G and LTE Evolved Packet Core (EPC) networks.

And so, in the subsequent sections of this article we present the novel enhanced HO signaling mechanism, followed by a detailed analysis of the same. We then also present an evolutionary network architecture which facilitates the execution of the proposed HO signaling mechanism.

III. LEGACY HANDOVER PREPARATION AND FAILURE SIGNALING

Erstwhile standardization efforts by 3GPP [2], [15], [21]–[30] have led to the formulation of the handover signaling mechanisms currently being utilized in cellular networks [2] and to be used in future wireless networks [15]. Specifically and according to discussions in Section I, handover preparation and failure signaling phases will be critical to the overall system performance during mobility events. Figs. 2, 3 and 4 illustrate the corresponding legacy handover preparation and failure signaling phases.

The legacy handover preparation signaling, exemplified in Figs. 2 and 4, are initiated by a *handover decision* made by the source network. This is followed by a *handover required* message (#1 in Fig. 2, and #2 in Fig. 3). Following these initial stages, the handover preparation phase is comprised of resource negotiation and allocation through the RRM operations (messages 6 and 7 in Fig. 2; messages 5 and 5a in Fig. 3), as well as CP signaling to establish GTP tunnels. These GTP tunnels require the entities at either end of the tunnel to have the TEIDs and transport layer addresses of each other. Hence, the preparation step also encompasses the creation and exchange of TEIDs and transport layer addresses between the core network entities. In order to realize a successful handover preparation, handshakes between the core network entities, i.e., messages 4, 5, 8 and 10b in Fig. 2; messages 4, 4a, 6, 6a, 8 and 8a in Fig. 3, are required. Next, the legacy handover failure phase signaling for inter-RAT HO (5G to EPS[2]) has been illustrated in Fig. 6. For the 5G NGC, the HO failure phase signaling currently only includes the HO cancel mechanism. In 3GPP specifications, the handover failure phase signaling encompasses two different types of signaling methods, i.e. Handover Cancel and Handover Rejection. We define them as follows:

- *Handover Cancel:* A handover cancel mechanism has been defined both for the 5G NGC as well as for the legacy networks, i.e. 4G, 3G and 2G. The cancel method is event based, i.e. it is initiated by a trigger event such as expiration of a timer, etc. It may be invoked only by the source network and at any point before the command to handover from the source network to the target network is sent from the MME/AMF to the source eNB/gNB.
- *Handover Reject:* A handover reject mechanism is currently defined only for the legacy networks, i.e. 4G, 3G and 2G networks. Similar to the handover cancel phase, the handover rejection method is event based, i.e. it is triggered by an event such as failure to allocate sufficient resources at the target access network. Hence, upon reception of a rejected request to reserve resources, the target MME/SGSN informs the source eNB/RNC about the rejected requested and hence, a handover reject.

And so based on the above definitions, due to certain network conditions, such as the expiration of a timer, etc., the source network may decide to cancel the HO (Fig. 4). Consequently, the source MME/AMF informs the source access point with regards to the canceled handover (message 1 in Fig. 4). Further, the source and target MME/SMF delete the sessions that had already been created with the target and source S-GWs/UPFs (messages 4, 4a, 7 and 8 in Fig. 4), respectively. The creation and deletion of these sessions with other core network entities involves handshakes, which, as we will discuss in the following subsection (III.A), are a significant source of inefficiency in CN handover signaling.

Note that the signaling schemes illustrated through Figs. 2, 3 and 4 are representative and other HO preparation and failure phase scenarios explored by 3GPP in [2], [15] are also of the same nature, wherein handshakes are utilized to accomplish the signaling procedures.

A. Signaling Inefficiency

From our discussions and Figs. 2, 3 and 4, it can be deduced that during the legacy handover preparation and failure phases, handshakes will be required to exchange the required CP information between the core network entities. For instance, in the handover preparation phase in Fig. 2, to establish a session between the Target MME and the Target S-GW a handshake, i.e., messages 4 and 5, is required between these respective entities. Such handshakes, whilst being a reliable methodology, will occupy the network for a long period as opposed to a mechanism that does not involve any handshakes. Further, it will also lead to higher latency, signaling cost, processing cost and total bytes of data transferred. And given the future network scenario depicted in Fig. 1(b), wherein the network will be dense and heterogeneous, the legacy mechanisms will be rendered inefficient. Thus, we define a new principle that is utilized to create a compressed message ensemble and an enhanced signaling method, showing increased performance in terms of latency, signaling cost, processing cost and total amount of bytes transferred. The principle is as follows:

“Identify the sequence of messages, such as the handshakes, where the performance of the 3GPP defined methods can be improved/enhanced. Then re-shuffle the information elements (IEs), if possible, to form a compressed message ensemble such that the sequence of messages under scrutiny are executed efficiently, if possible in parallel, but with the desired functionality."

And so, we next discuss the novel message mapping and signaling strategies that alleviate the deficiencies mentioned above.

---

[2] The EPS consists of EPC and E-UTRAN. Note that, the standard documents by 3GPP utilize EPS and EPC interchangeably while defining HO scenarios. Hence, in this paper we utilize the same principle.

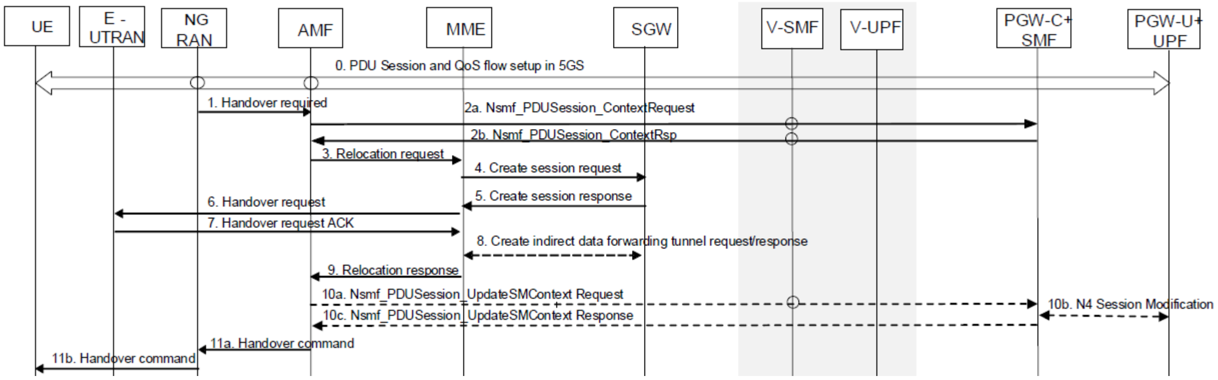


Fig. 2. Legacy handover preparation signaling for Inter-RAT HO (5G NGC to EPS) [15].

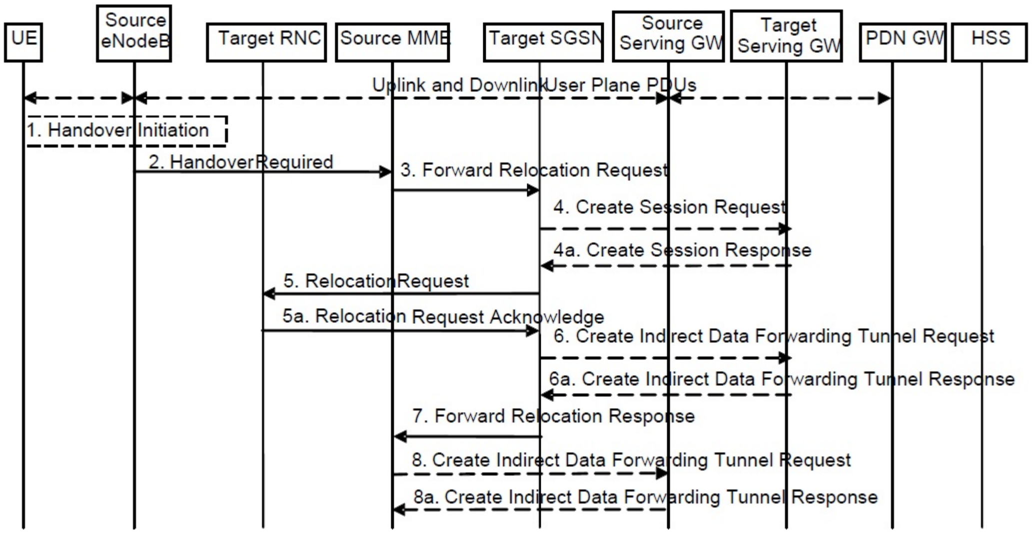


Fig. 3. Legacy handover preparation signaling for Inter-RAT HO (LTE to UMTS/2G) [2].

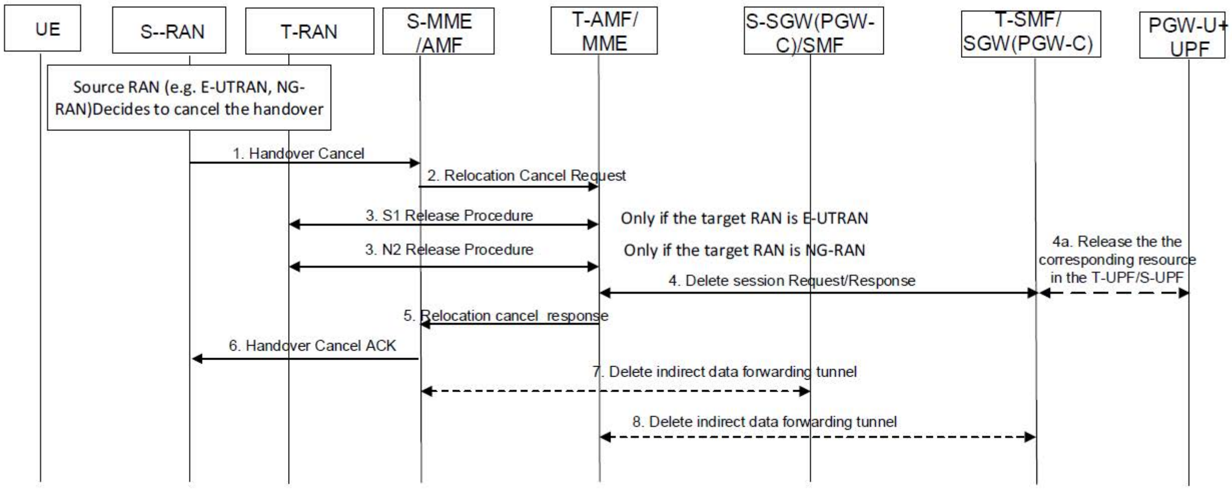


Fig. 4. Legacy handover failure signaling for Inter-RAT HO (5G NGC and EPS) [15].

TABLE II
DIFFERENT HANDOVER SCENARIOS ANALYZED

		Target Network		
		5G NGC†	**EPC**	**UMTS/2G**
Source Network	**5G NGC†**	N2 based HO: UE migrates from one NG-RAN to another	Inter-RAT HO involving an N26 interface	
			Inter-RAT HO without an N26 interface	
	EPC	Inter-RAT HO involving an N26 interface	Intra-RAT HO involving MME relocation but no S-GW relocation	Inter-RAT HO involving S-GW relocation and an Indirect tunnel
				Inter-RAT HO involving S-GW relocation and a Direct tunnel
		Inter-RAT HO without an N26 interface	Intra-RAT HO involving MME and S-GW relocation	Inter-RAT HO without any S-GW relocation but with an Indirect tunnel
				Inter-RAT HO without any S-GW relocation but with a Direct tunnel
	UMTS/2G		Inter-RAT HO involving an S-GW relocation and Indirect tunneling	
			Inter-RAT HO without any S-GW relocation but with Indirect tunneling	

†3GPP standards document only discuss 5G NGC to EPS HO and vice versa.

IV. PROPOSED HANDOVER PREPARATION AND FAILURE SIGNALING

The proposed handover preparation and failure phases consist of a novel message mapping and signaling mechanism, wherein a compact and intelligent mapping of IEs from the legacy to the proposed signaling messages has been provided. Additionally, the proposed mechanism also involves parallel transfer of CP information. To facilitate these capabilities, we utilize the SDN enabled CN entities including the SeMMu, as defined in Section I. Specifically, the SeMMu through its SDN capabilities and centralized location facilitates:

- Parallel transfer of CP information to other CN entities.
- Allocation of TEIDs and transport layer addresses for the other CN entities at the SeMMu itself.

Thus, through the use of compact message ensemble and parallelization of information transfer, the handover preparation and rejection signaling for the various 3GPP HO types are, as we will discuss in this section, optimized. The HO scenarios that have been analyzed in this work are presented in Table II. Concretely, the various networks, i.e., 2G, 3G, 4G-LTE and 5G, have been considered in this table and all the possible HO scenarios among them have been enlisted.

To evince the optimization achieved for the aforementioned 3GPP HO scenarios, we consider a representative HO scenario, i.e., Inter-RAT HO from 5G NGC to EPS network wherein the serving gateway is relocated. The optimized/enhanced message maps and signaling sequence for other scenarios have been defined in [31]. And so, for the scenario considered, a user undergoes an Inter-RAT handover with the source system being 5G and the target system being an EPS network. Further, serving gateway relocation defines that during the handover process the gateway that is serving the user is changed. In this particular scenario, the gateway in the source network is a UPF which upon handover to the EPS is switched to a target S-GW. Also, note that the considered scenario consists of an N26 interface which facilitates the inter-working between the NGC and EPS.

Note that, in addition to the proposed signaling method, in this section we also present a novel handover failure aware handover preparation method. This novel approach enhances not only the handover failure method but also enhances the handover preparation method further. For the sake of brevity, we defer the detailed discussion on this method to Section IV.C.

A. *HO Preparation: Optimal Message mapping and Signaling*

An illustration of the proposed message mapping and signaling sequence for the handover preparation phase of the representative HO scenario has been presented in Figs. 5 and 6, respectively. By message mapping (Fig. 5), here we refer to a graphical representation of how messages from the legacy message ensemble are mapped to the proposed message ensemble. Specifically, during the mapping process, IEs, which are the building blocks of these messages, are re-organized in a way that helps to reduce the message ensemble. This consequently also aids in an improved performance in terms of latency incurred, transmission cost, processing cost and the overall number of bytes transferred, as will be seen in Section V. It must be stated here that, the mapping process is performed without transforming the format and contents of the IEs, as it ensures an evolutionary approach that is easy and fast to adapt for the operators and vendors alike. Next, the signaling sequence presented in Fig. 6, is an illustration of the sequence in which the messages from the proposed ensemble are executed.

From Fig. 5, it can be observed that the proposed message mapping reduces the size of the message ensemble to 15 as compared to the 18 required during the legacy handover

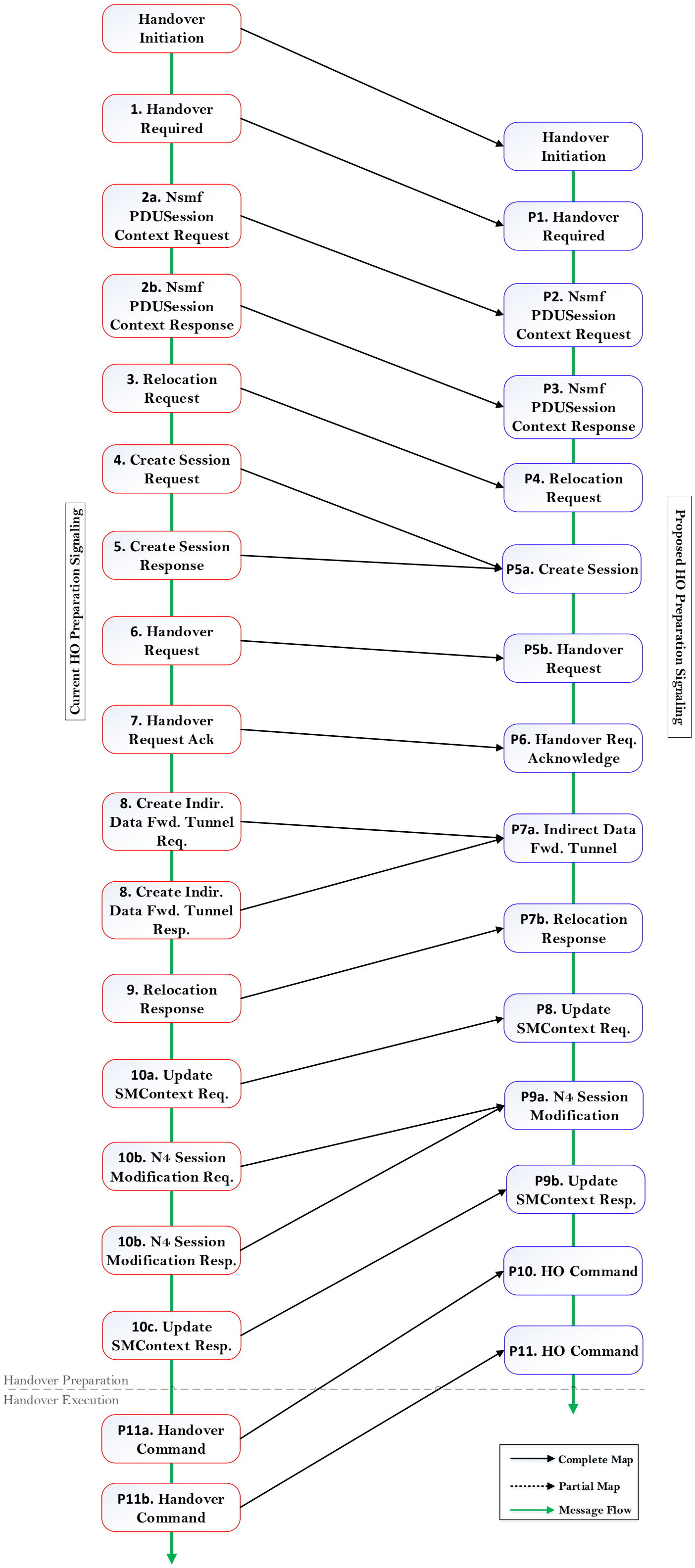


Fig. 5. Proposed Handover Signal mapping for Inter-RAT HO from 5G NGC to EPS.

preparation signaling. In the legacy and proposed message ensembles, the *Handover Command* messages (P11a-11b and P10-11 respectively)[3] have also been included instead of considering them in the handover execution phase. This is so because, unless the RRM information from the target network is delivered by the source network to the user, from the user's perspective the network is still in a handover preparatory phase. Next, in the message mapping presented in Fig. 5, the *Handover Initiation*, *Handover Required* and *HO Command* messages are left unchanged from the legacy message ensemble. Concretely, the IEs in the aforementioned messages are left unchanged from the legacy mechanism. Further, and in accordance with the discussion in Section III, the messages involving RRM operations are left unaltered as the primary aim of the proposed strategy is to reduce the CN signaling during the handover process. Hence, *Handover Request* and *Handover Request Acknowledge* messages (6 and 7) are unaltered. The messages that are modified (enhanced) have been explored below:

- *Create Session (P5a):* This message is composed of the IEs from messages 4 and 5 both. Whilst through message 4 the SeMMu provides the S-GW with the necessary information about the PDN connections that are going to be handed over, message 5 allows P5a to allocate the S-GW its own resources such as TEID and transport layer addresses. This re-arrangement of the IEs helps to eliminate the requirement of a handshake, which consequently enhances the handover preparation phase signaling.
- *Indirect Data Forwarding Tunnel (P7a):* The given message is composed of the two sub–messages that are contained within the handshake labeled message 8 in the signaling defined by 3GPP [15]. The forward message of the aforesaid handshake enables the MME to specify the TEID(s) and address(es) for the indirect data forwarding tunnel to the S-GW. Next, the S-GW specifies its own TEID(s) and address(es) for the indirect data forwarding tunnel in the response part of the specified handshake (message 8) to the MME. However, the proposed message mapping enables the SeMMu to allocate the required TEIDs and transport layer addresses, including that of the S-GW itself, for the indirect forwarding tunnel without the requirement of a handshake. Concretely, the IEs from messages involved in handshake 8 are re-organized such that the TEID(s) and address(es) of the other CN entities, as well as of the S-GW itself, for the indirect data forwarding tunnel are specified to the S-GW in a single step, i.e. through message P7a.
- *N4 Session modification (P9a):* In the signaling specified by 3GPP, the N4 modification request message (#10b in Fig. 2) permits the SMF to apprise the UPF about the TEID(s) and address(es) of the S-GW for setting up a data forwarding tunnel. Further, the UPF responds to this message with an N4 modification response message (#10b in Fig. 2) with its own CN tunnel info consisting of TEID(s) and address(es). However, in our proposed approach, the IEs of the aforesaid modification request/response messages are mapped to message P9a, such that the TEID(s) and address(es) of the S-GW and UPF for the data forwarding tunnel are specified to the UPF. This eliminates the handshake and hence, enhances the handover signaling process.

Next, in the proposed handover preparation signaling, presented in Fig. 6, the sequence and operation of *Handover Initiation, Handover Required* and *Handover Command* messages remains unaltered from the legacy signaling [15]. Further, the messages that are associated with the RRM operations, i.e., *Handover Request* and *Handover Request Acknowledge*, also remain unaltered in their operation. However, these messages (6 and 7 in the legacy signaling, i.e. Fig. 2) have been reassigned as messages P5b and P6 in the proposed signaling approach. Additionally, utilizing the already stated capability of parallel information transfer through the SDN agent on the SeMMu, messages P5a-5b, P7a-7b and P9a-9b, are executed simultaneously pairwise. Lastly, the HO command message is forwarded by the AMF to the source NG-RAN (S-NG-RAN). The S-NG-RAN then forwards it to the UE, marking the end of the HO preparation phase.

B. HO Failure: Enhanced process

Recall from our discussion in Section III that, the handover failure phase signaling consists of two methods, i.e. Handover Cancel and Handover Rejection. Considering the representative HO scenario, i.e. 5G NGC Inter-RAT HO scenario, the enhanced handover cancel phase signaling for the same has been illustrated in Fig. 7. Note that, for the sake of brevity the message mapping is not presented here. Concretely, we utilize the principle used in Section IV.A to compress the message ensemble and enhance the signaling process for the HO cancel phase as well.

In the proposed signaling presented in Fig. 7, the source RAN firstly decides to cancel the HO. Thus, following this decision, a *HO cancel* message (P1) is sent to the AMF, which then issues a *Relocation Cancel request* message to the T-SeMMu. The T-SeMMu then utilizes its SDN capabilities to simultaneously:

- Delete the sessions it created during the HO preparation phase via the *Delete Existing Session* message (P4a).
- Issue a *Relocation Cancel Response* message (P4b) to the source S-AMF (Source AMF).
- Delete the indirect forwarding tunnels that it created during the HO preparation phase via the *Delete Indirect Forwarding Tunnel* message (P4c).

Concretely, the aforementioned parallelization of messages provides the claimed optimization in the HO cancel signaling phase. Subsequently, the S-AMF sends a HO cancel acknowledgement message (P6) to the source RAN. Lastly, the S-AMF also performs a handshake (P7) with the source SeMMu for the deletion of any indirect tunnels that were setup during the HO preparation phase in the 5G NGC. Since, the AMF is not equipped with the capabilities of allocating TEID(s) and

---

[3] Legacy messages are assigned only numbers, while the proposed mechanism messages are assigned numbers beginning with letter "P". e.g. a legacy message would be numbered as *7*, while a proposed mechanism message would be numbered as *P6*.

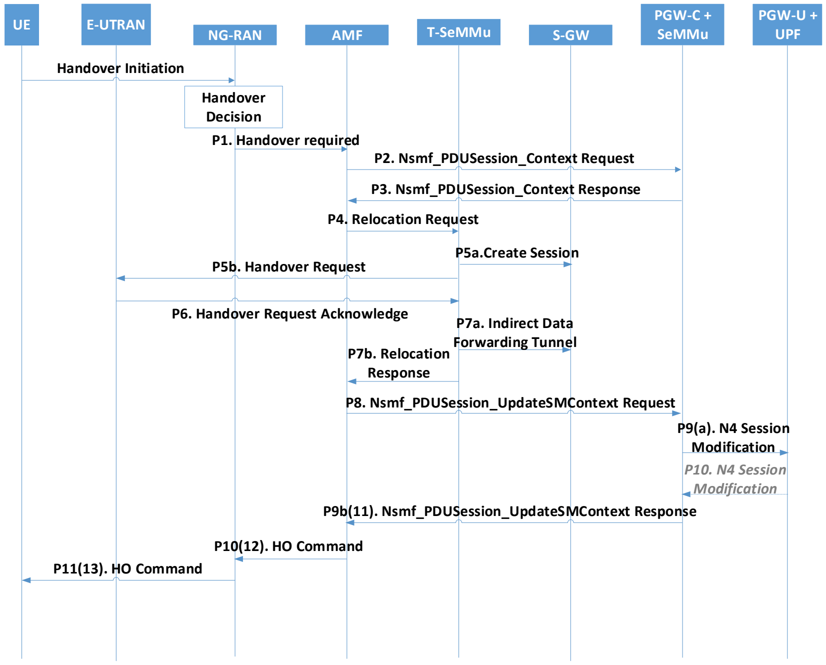


Fig. 6. Proposed Handover signaling sequence for Inter-RAT HO from 5G NGC to EPS.

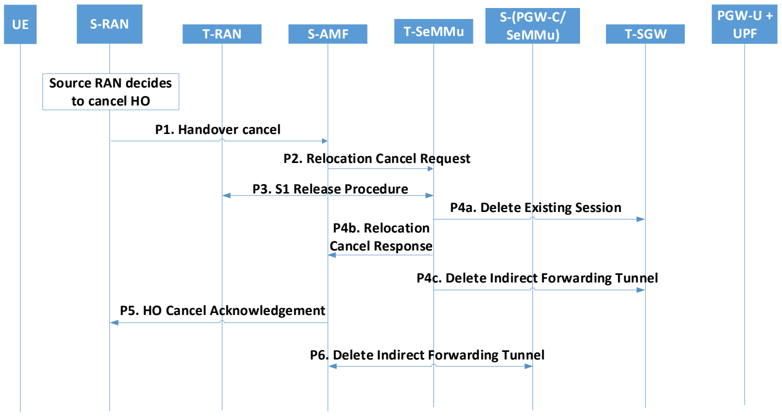


Fig. 7. Proposed Handover cancel phase signaling for Inter-RAT HO from 5G NGC to EPS.

address(es) like the SeMMu, it has to perform the handshake P7, instead of transferring just a single message with all the CP information to the source SeMMu.

Recall from our discussion in Section III that for the LTE-EPC to UMTS/2G and vice versa handover scenarios, both the handover cancel and handover rejection methods are defined [2]. However for analytical reasons, elaborated in Section V, in this work we only consider the HO rejection phase signaling for this case scenario. Further, we utilize the same principle as the HO cancel phase and HO preparation phase (Section IV.A), for the HO rejection phase signaling enhancement.

As a representative example, through Fig. 8, we present the enhanced HO rejection signaling for LTE-EPC to UMTS/2G HO scenario when there is an indirect tunnel and a S-GW relocation exists. Briefly, after the initiation of the Handover required message (P2), the SeMMu communicates with the Target SGSN for allocating resources as well as setting up the tunnel (message P3a). Simultaneously, the SeMMu also sets up tunnels with the Source and Target Serving GWs through messages P3b and P3c, respectively. Next, the Target SGSN requests the Target RNC to setup access network resources for the impending handover via the Relocation Request message (P4). However, message P4a indicates to the Target SGSN that a HO is being rejected (possibly due to lack of physical layer resources) and hence, forwards the same indication to the SeMMu in the *Forward relocation response* message. Following this message, the SeMMu deletes the sessions it had created during messages P3b and P3c with the Source and Target S-GWs respectively. The SeMMu performs this operation via messages P6a and P6b. To conclude the HO rejection phase, the SeMMu parallely also issues a *Handover Preparation Failure* to the source eNB.

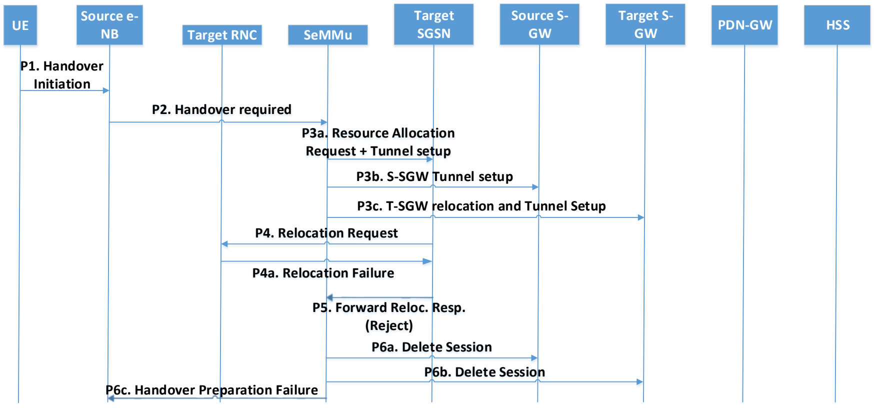


Fig. 8. Proposed Handover rejection phase signaling for Inter-RAT HO from LTE to UMTS/2G network when there is a S-GW relocation and indirect tunneling exists.

And so, based on these proposed signaling improvements, in the next section we present the novel *Handover Failure aware Preparation signaling* process.

C. Handover Failure aware preparation signaling

For the HO cancel phase presented in Fig. 7, it can be observed that the *Delete Session* and *Delete Indirect Tunneling* messages are sent to the target S-GW and source UPF by the target and source SeMMus respectively, as a consequence of the sessions that were established in messages P5a, P7a and P9a (Fig. 6). These messages release the CN resources that are reserved by the SeMMu during the HO preparation phase. Further, from [15] it is understood that the HO cancel phase can be executed at any point within the HO preparation phase before the *HO Command message*, or even after the HO preparation phase if the UE fails to attach or register to the target network.

With this background, and through Figs. 6, 7 and 10, it must be noted that if the HO cancel phase is executed before message P5a (Fig. 6) is sent, then it will be fully optimal, i.e. there will be no requirement for handshakes to delete the tunnels. This is so because, there will be no CN resources that would have been allocated during the HO preparation steps at that point, and hence, there will be no requirement of messages P4a, P4c and P7 (Fig. 7) to release those resources. However, if the HO cancel phase is executed at any time instance after the execution of message P5a (Fig. 6) within the HO preparation phase, then the cancel phase signaling will require messages that will help release the allocated CN resources. Thus, the proposed HO cancel phase signaling presented in Fig. 7 is capable of adapting itself depending on when it is initiated, and as a consequence we consider it to be near-optimal.

Upon performing a deeper analysis into the IEs that constitute messages P5a, P7a and P9a (Fig. 6), we observe that the IEs of messages P5a and P7a can be re-shuffled and mapped to create a single message P7a, i.e. *Indirect Data forwarding tunnel*, as shown in the Handover failure aware preparation phase signaling (Fig. 9). Further, the new message P7a is executed after messages P5 and P6 (Fig. 9), which facilitates the HO cancel phase with a greater chance of achieving its optimal state presented in Fig. 10. The reason being, the later the tunnels are setup, the higher is the chance of a HO cancel event not requiring to delete these tunnels as it may be initiated before they are setup. Additionally, and in the event, the HO cancel phase is executed after message P9a (Figs.6 and 9), the aforesaid enhancement would require the HO cancel signaling to delete 2 tunnels (created through messages P7a and P9a in Fig. 9) instead of the 3 (created through messages P5a, P7a and P9a in Fig. 6). However, given the dynamic nature of HO cancel phase, in the analysis we only consider the enhanced signaling specified in Fig. 7. This is also the worst case HO cancel phase scenario as it is executed after all the tunnels have been setup.

Additionally, utilizing this novel signaling approach, the handover rejection phase for the LTE to UMTS handover scenario, illustrated in Fig. 8, has been further enhanced (Fig. 12). To achieve this enhancement, the HO preparation phase for the given scenario is first modified (Fig. 11) such that all the tunnel and session creation messages are executed after *Relocation Request Acknowledge* message (5a in Fig. 3). The reason being, in the event there is a HO rejection, the *Relocation Request Acknowledge* message issues an indication of the rejection to the SGSN. The SGSN then passes this indication to the SeMMu, which instantly passes the reject indication to the source eNB, without the requirement of any session and tunnel deletion messages. Hence, the HO rejection phase signaling is further enhanced as compared to signaling proposed in Fig. 8. Further, given that the resource allocation process is successful and a positive indication is received from the *Relocation Request Acknowledge* message, the tunnel and session creation messages are executed simultaneously with the *HO command* message. This parallel execution with the *HO command* grants additional enhancement to the HO preparation phase as it helps to reduce the latency further.

D. Xn, X2 and S1 Interface based Handover Signaling

As an evolution from the EPC architecture, the 5G NGC specifies an Xn interface between two gNBs. In scenarios,

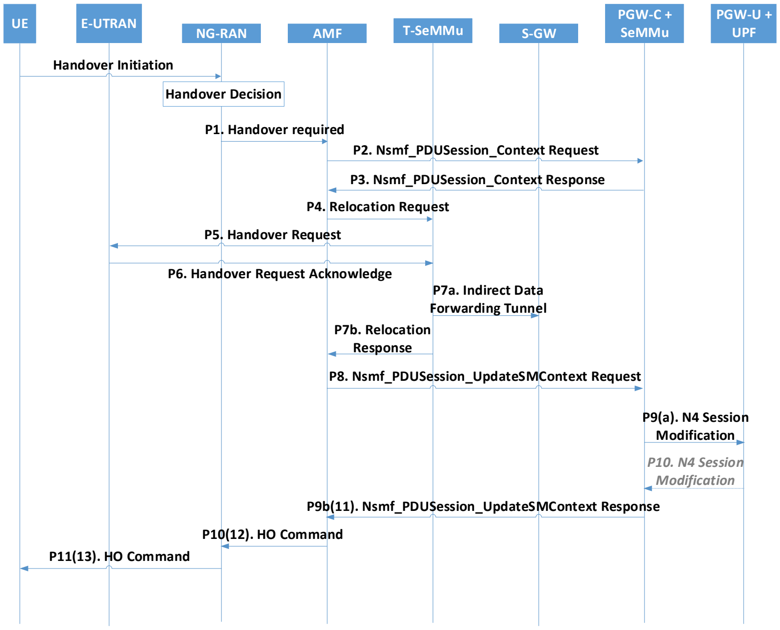


Fig. 9. Handover failure aware Handover preparation Signaling for Inter-RAT HO from 5G NGC to EPS.

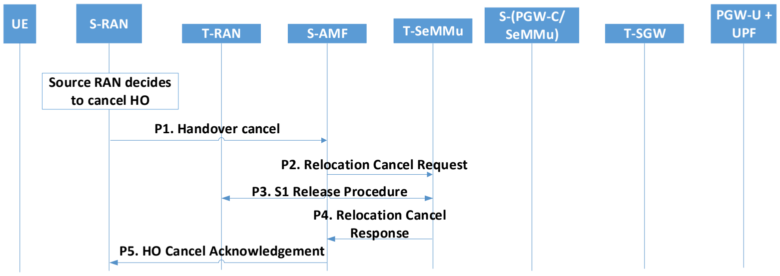


Fig. 10. Optimal proposed Handover rejection phase signaling sequence for Inter-RAT HO from 5G NGC to EPS.

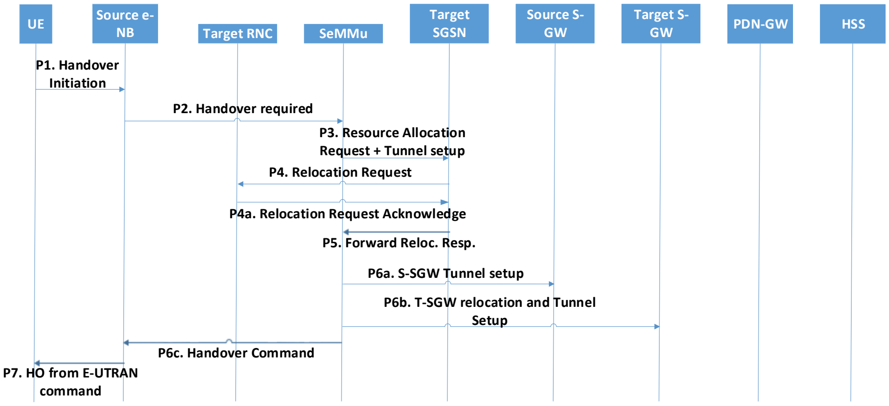


Fig. 11. Handover failure aware Handover preparation Signaling for Inter-RAT HO from LTE-EPC to UMTS/2G when there is indirect tunneling and S-GW relocation occurs.

wherein the two gNBs involved in the HO process are connected via an Xn interface, the given interface facilitates a faster handover process. Subsequently, upon deeper exploration of the handover preparation signaling mechanism for an Xn based HO from [32], it is evident that the existing mechanism is optimal. Concretely, since the signaling does not involve any handshakes and any significant interaction with the CN entities, the proposed handover preparation signaling

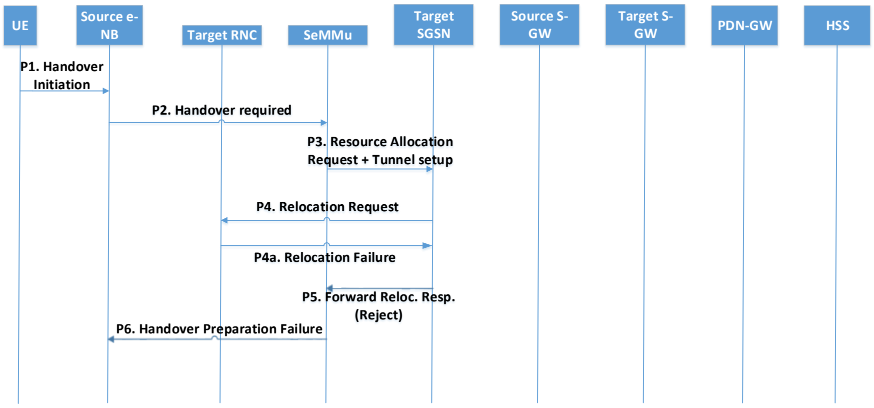


Fig. 12. Optimal proposed Handover rejection phase signaling sequence for Inter-RAT HO from LTE-EPC to UMTS/2G when there is indirect tunneling and S-GW relocation occurs.

process will neither provide any gains nor will it lead to any regressive effects on the performance of the Xn based HO mechanism.

Further, the LTE-EPC defines two specific interfaces through which the Intra-RAT handovers can be executed, i.e., the X2 and the S1 interface [2]. Whilst, the X2 interface is defined between two eNBs (like the Xn between two gNBs in 5G NGC), the S1 interface involves the CN. The legacy X2 and S1 handover preparation signaling mechanisms have been presented in [33]. Analyzing the signaling mechanisms presented in [33], it can be concluded that the existing X2 and S1 handover mechanisms, similar to the Xn handover mechanism for 5G NGC, are optimal. Concretely, while the X2 HO is similar to the Xn HO in 5G NGC (wherein the HO signaling is only at the access network), for the S1 HO the only CN signaling is that between the MME (SeMMu) and the source and target eNBs (wherein RRM operations take place). Hence, both X2 and S1 HO signaling scenarios are considered to be optimal, as stated above.

Note that, here for the S1 handover we consider the scenario wherein the user does not switch its MME (SeMMu) and S-GW. In the event either is changed, the proposed handover mechanism leads to immediate gains, which have been presented via the analysis in Section V. Additionally, it is important to state that the Xn, X2 and S1 interfaces are an integral part of the evolutionary network architecture proposed in this work, and are agnostic to the implementation of the architecture and signaling methodology.

Thus, with the aforesaid principles, processes and methodologies, in the following section we present a detailed quantitative performance improvement analysis for the myriad scenarios that have been listed in Table II.

V. PERFORMANCE ANALYSIS

To analyze the proposed handover preparation and failure signaling phases, we use latency, processing cost, transmission cost and amount of bytes transferred within CN as evaluation metrics. Note that, as in [34], utilizing these metrics for evaluating new handover strategies is standard practice.

A. Analytical Formulation

For the analysis we first define a set $S = \{s_1, ..., s_N\}$ corresponding to all the link delays encountered within a given signaling sequence. Then the set $J = \{j_1, ..., j_K\}$ is the set of all the parallel link delays, where $K \leq N$. By *Parallel link delay* we mean that, if x messages are to be executed simultaneously then the overall delay incurred will be the maximum of the delays experienced by the messages under observation. It is then computed as

$$\begin{aligned}\text{Parallel Link Delay} \quad = \quad & \max(\mathit{Link\ delay\ msg\ 1}, \ldots, \\ & \mathit{Link\ delay\ msg\ x}), \qquad (1)\end{aligned}$$

Additionally, we consider only a single processing delay for the group of messages that are being executed in parallel. Hence, the set of processing delays can be defined as $D = \{d_1, ..., d_K\}$. It is imperative to state here that the assumption for the aforesaid processing delays, mentioned in Section V.B, is conservative in nature. Hence, any SDN agent processing delay is also included within the utilized assumptions. Also, in HO procedures no routing table updates would be necessary, as they are already configured during the network setup phase, hence no signaling overhead is created by SDN agents for HOs.

And so, for the computation of latency incurred during the handover preparation and handover failure signaling phases, we consider the contributions from parallel link delays and processing delays as:

$$\text{Latency} = \sum_{i=1}^{K} \{j_i + d_i\} \qquad (2)$$

The Transmission Cost computation, on the other hand, requires that each link delay be considered for the evaluation, and is computed as

$$\text{Transmission Cost} = \frac{\sum_{l=1}^{N} s_l}{1\text{ms}}. \qquad (3)$$

Concretely, the Transmission Cost analysis represents the amount of time the CN links are occupied during the complete HO signaling process. Next, for the processing cost analysis, we utilize the analytical methodology in [35] and define it as the number of messages generated during the HO preparation/failure signaling phase. We then compute the percentage processing cost saving as

$$\text{Proc. Cost Saving} = \frac{MSG_{Legacy} - MSG_{Proposed}}{MSG_{Legacy}} * 100\%, \tag{4}$$

wherein, MSG_{Legacy} is the number of messages in the legacy approach for HO preparation/failure and $MSG_{Proposed}$ is the number of messages in the proposed approach for HO preparation/failure.

In addition, in this paper we also present an analysis for the network wide processing cost and occupation time. Whilst the network wide processing cost will reflect the network wide reduction in the processing cost through the proposed method, the network occupation time will be reflective of the reduction in the amount of time the CN links are occupied due to the handover signaling across the network.

To conduct the aforesaid analysis we introduce the formulations in Equations (5) and (6). $\text{NPC}_{DH_fS_{1p}}$ and $\text{NOcT}_{DH_fS_{1p}}$ are the Network wide processing cost and Network occupation time, respectively, given a HO distribution (distribution of percentage of total users undergoing a particular handover), HO failure (rejection/cancellation) rate and percentage of users undergoing S1 HO, respectively. Note that, we consider only the Intra-MME/S-GW scenario for S1 HOs as it is not impacted by the implementation of the proposed mechanism but it still involves CN signaling. Moreover, we do not consider the X2 and Xn handovers for the analysis, as they do not involve any significant CN signaling. The rest of the notations in (5) and (6) are as follows: $\mathbf{H_{pscost}}$ is Handover preparation processing cost vector; $\mathbf{Dist_{HO}}$ is the handover distribution vector; HO_{sperc} is the handover success percentage; $\mathbf{H_{fcost}}$ is the processing cost vector during Handover failure; HO_{fperc} is the HO failure percentage; N_{HO} is the number of users undergoing handover in the network; S_{1p} is the percentage of S1 handovers (Intra-MME/S-GW); S_{1scost} is the processing cost for a successful S1 HO preparation; S_{1fcost} is the processing cost for a failed S1 HO; $\mathbf{H_{tscost}}$ is the transmission cost vector for a successful HO preparation; $\mathbf{H_{tsfcost}}$ is the transmission cost vector during a HO failure scenario; $S_{1tscost}$ is the transmission cost for a successful S1 HO preparation; and $S_{1tsfcost}$ is the transmission cost incurred when a S1 HO fails.

B. Parameter Specification and Assumptions

As part of the analytical framework, the parameter values that will be utilized to conduct the analysis are provided in this sub-section. Firstly, the one-way delays for each CN link, necessary for the latency and transmission cost analysis, have been defined in Tables III and IV by utilizing the data from a Japanese cellular operator [36], Cisco [37] and a Greek cellular operator. Further, the delays presented for each link are considered to be symmetric, i.e., if a delay of 1ms is incurred for the link from AMF to SeMMu, then the same link delay is assumed from SeMMu to AMF.

TABLE III

LINK TYPE AND CORRESPONDING DELAYS IN PROPOSED ARCHITECTURE (DERIVED FROM A JAPANESE OPERATOR [36] AND CISCO DATA [37])

	Link Type	**Link Delay**
1.	UE to NG-RAN	1ms
2.	NG-RAN to AMF	7.5ms
3.	AMF to SeMMu (PGW-C + SMF)	1ms
4.	AMF to SeMMu	1ms
5.	SeMMu to S-GW	7.5ms
6.	SeMMu (PGW-C + SMF) to PGW-U + UPF	7.5ms
7.	SeMMu (PGW-C + SMF) to PCRF+PCF	7.5ms
8.	AMF to AMF	15ms
9.	SeMMu to PGW	7.5ms
10.	SeMMu to E-UTRAN	7.5ms
11.	E-UTRAN to UE	1ms
12.	PGW to PCRF	7.5ms
13.	S-GW to PGW	7.5ms
14.	SeMMu to SGSN	1ms
15.	SGSN to RNC	6ms
16.	SGSN to S-GW	7.5ms
17.	SeMMu to SeMMu	15ms

In Table III the link delays presented are derived from a Japanese operator deployment data [36] and CISCO data [37]. In addition, the link delays are computed considering that the MME (SeMMu in this study) and the SGSN are co-located, as specified in [38]. Utilizing this co-location principle, we also establish the link latency between AMF and SeMMu. Further, the 15 ms SeMMu-SeMMu and AMF-AMF delay is based on the premise that the delay between the SeMMus/AMFs will be greater than the largest CN delay within a SeMMu/AMF domain. Hence, for the purpose of analysis in this paper and for the data provided from the Japanese operator and Cisco, an assumption of two times the greatest link delay within a SeMMU/an AMF domain has been considered.

$$\begin{aligned}\text{NPC}_{DH_fS_{1p}} \quad = \quad & \Big\{(\mathbf{H_{pscost}} * \mathbf{Dist_{HO}}^{\mathbf{T}}) * HO_{sperc} + (\mathbf{H_{fcost}} * \mathbf{Dist_{HO}}^{\mathbf{T}}) * HO_{fperc}\Big\} * (1 - S_{1p}) * N_{HO} +, \\ & S_{1p} * N_{HO} * \Big\{S_{1scost} * HO_{sperc} + S_{1fcost} * HO_{fperc}\Big\}\end{aligned} \tag{5}$$

$$\begin{aligned}\text{NOcT}_{DH_fS_{1p}} \quad = \quad & \Big\{(\mathbf{H_{tscost}} * \mathbf{Dist_{HO}}^{\mathbf{T}}) * HO_{sperc} + (\mathbf{H_{tsfcost}} * \mathbf{Dist_{HO}}^{\mathbf{T}}) * HO_{fperc}\Big\} * (1 - S_{1p}) * N_{HO} +, \\ & S_{1p} * N_{HO} * \Big\{S_{1tscost} * HO_{sperc} + S_{1tsfcost} * HO_{fperc}\Big\}\end{aligned} \tag{6}$$

TABLE IV
LINK TYPE AND CORRESPONDING DELAYS IN PROPOSED ARCHITECTURE (DERIVED FROM A GREEK OPERATOR AND CISCO DATA [37])

	Link Type	**Link Delay**
1.	UE to NG-RAN	1ms
2.	NG-RAN to AMF	19ms
3.	AMF to SeMMu (PGW-C + SMF)	0.5ms
4.	AMF to SeMMu	0.5ms
5.	SeMMu to S-GW	1ms
6.	SeMMu (PGW-C + SMF) to PGW-U + UPF	1ms
7.	SeMMu (PGW-C + SMF) to PCRF+PCF	1ms
8.	AMF to AMF	2ms
9.	SeMMu to PGW	1ms
10.	SeMMu to E-UTRAN	19ms
11.	E-UTRAN to UE	1ms
12.	PGW to PCRF	1ms
13.	S-GW to PGW	1ms
14.	SeMMu to SGSN	0.5ms
15.	SGSN to RNC	2ms
16.	SGSN to S-GW	1ms
17.	SeMMu to SeMMu	2ms

On the other hand, the values of delays obtained from the Greek operator correspond to eNBs from two different networks and CN elements from 3 different MME domains. Consequently, for the chosen network and its MME domain, the link delays are computed as the average of all the delay values provided by the network operator for that specific link. Further, the UE-eNB and the eNB-SeMMu delay for both data sets is derived from the Cisco framework in [37]. Additionally, for the latency analysis, we consider the processing delay to be 4ms in all CN entities, as in [37].

For the network wide analysis, we consider that the number of users undergoing handover at any given time in the considered network, i.e., the parameter N_{HO} in (5) and (6), is 3 million. The analysis does not take into consideration the users that undergo an X2 or Xn based handover, i.e. they are not included amongst the 3 million users that we include in our analysis, as they do not involve any HO-related CN signaling. In addition, and based on discussions in Sections IV.B and IV.C, the HO cancel phase is considered only for the 5G networks, while for the legacy networks (4G/3G/2G) we only consider the HO rejection phase signaling. Recall that, for the 5G networks the rejection phase signaling does not exist. Further, the considered HO cancel phase for the 5G NGC is as shown in Fig. 7, which is also the worst case enhanced signaling for the same. However, for the legacy networks (4G/3G/2G) we do not consider the HO cancel phase since:

- The HO cancel signaling process for the legacy networks is fundamentally the same as that in the 5G NGC. Hence, considering the HO rejection signaling phase for the legacy networks aids in the completeness of analysis and study.
- Given the dynamic nature of HO cancel phase (Section IV), considering the HO rejection phase signaling also facilitates the ease of analysis.

We then develop five randomly distributed settings over the HO types (Table II) for the computation of network wide processing cost and network occupation time. Concretely, we define the HO distributions that will be utilized for the analysis through (5) and (6), i.e., the parameter $\mathbf{Dist_{HO}}^{\mathbf{T}}$. The distributions are generated using Algorithm 1, wherein one of the distributions is predefined to be uniform across the HO types. Through uniform we mean that the percentage of users experiencing a particular handover scenario is the same for all HO types. It is imperative to state here that, the premise behind considering random distributions over the HO types is the lack of availability of real data from network operators.

Algorithm 1 Distribution Generation

```
 1: procedure DistributionGenerator
 2:     $iter \leftarrow 5$
 3:     $i \leftarrow 1$
 4:     $mprct \leftarrow 0.2$
 5:     $NoH \leftarrow$ Number of Handover Types
 6:     for $i < iter$ do
 7:         $maxper \leftarrow mprct$
 8:         $minper \leftarrow 10^{-4}$
 9:         $j \leftarrow 1$
10:         for $j <=$ NoH do
11:             $Distper(i, j) \leftarrow U[minper, maxper]$
12:             $maxper \leftarrow min(1 - sum(Distper(i, :)), mprct)$
13:             $j \leftarrow j + 1$
14:     $Distper(5, :) \leftarrow ones(1, NoH)/NoH$
```

And so, in Algorithm 1 we first define the maximum percentage of users (*maxper*) that undergo a particular HO type to be 20%, whereas the minimum percentage (*minper*) of users that undergo a particular HO type is 0.01%. Next, to generate the random distribution, we utilize the uniform probability distribution (U), with its upper and lower bounds being specified by the maximum and minimum percentage, respectively. We continually update the maximum percentage so as to prevent any skewness in the nature of distribution. The update rule is defined as the minimum value amongst 20% (initial maximum percentage value) and the percentage of users that remain to be associated to a particular HO type. We then define the last distribution as being uniform across all the HO types (Algorithm 1: Line 14).

C. Performance Analysis

In this section, utilizing the formulation presented in Section V.A, we present and discuss the analytical results for the latency, processing cost and transmission cost of the new signaling framework for the handover preparation and failure phases presented in Section IV. For the analysis, we utilize the link latency data shown in Tables III and IV. The analytical methodology undertaken here is used to compare the performances of the proposed approach and the current 3GPP defined approach.

1) Latency analysis: Utilizing (2), as well as the cellular operator data from Section V.B, we present the analytical results for the latency improvement for the handover preparation phase in Tables V and VI.

We show through this analysis that the proposed mechanism reduces the latency as compared to the legacy mechanism for both sets of operator data and all HO types considered. Note that, while the proposed mechanism helps reduce the latency

TABLE V
Preparation Phase: Handover Latency Improvement Analysis (Cisco and Cellular Operator-Japan)

Handover Type	Legacy Mechanism	Proposed Mechanism	Percentage Latency Reduction
$1.\mathrm{U}^{\rho}$	155 ms	95 ms	38.71%
$1.\mathrm{U}^{\triangle}$	138.5 ms		31.41%
$1.\mathrm{V}^{\rho}$	181 ms	89 ms	50.82%
$1.\mathrm{V}^{\triangle}$	171.5 ms	138.5 ms	19.24%
1.W	179 ms	123 ms	31.28%
$1.\mathrm{X.a}^{\dagger}$	128 ms	65.5 ms	48.83%
$1.\mathrm{X.b}^{\dagger}$			
$1.\mathrm{Y.a}^{\dagger}$	82 ms	65.5 ms	20.12%
$1.\mathrm{Y.b}^{\dagger}$		58 ms	29.27%
$1.\mathrm{X.a}^{*}$	129.5 ms	65.5 ms	49.42%
$1.\mathrm{Y.a}^{*}$	82 ms	65.5 ms	20.12%
2.y	113 ms	90 ms	20.35%
2.x	159 ms	90 ms	43.40%

1: Inter-RAT HO; **2:** Intra-RAT (LTE) HO; **a:** Indirect Tunnel; **b:** Direct Tunnel; **U:** with N26 interface
V: without N26 interface; **X:** with T-SGW; **Y:** without T-SGW; $^{\rho}$5GS to EPS; $^{\triangle}$EPS to 5GS
y: inter-MME and intra-SGW; **x:** inter-MME and S-GW; *UMTS/2G to LTE; †LTE to UMTS/2G
W: Intra-NG-RAN N2 based HO in 5G NGC

TABLE VI
Preparation Phase: Handover Latency Improvement Analysis (Cisco and Cellular Operator-Greece)

Handover Type	Legacy Mechanism	Proposed Mechanism	Percentage Latency Reduction
$1.\mathrm{U}^{\rho}$	155 ms	126 ms	18.71%
$1.\mathrm{U}^{\triangle}$	155.5 ms		18.97%
$1.\mathrm{V}^{\rho}$	162 ms	104 ms	35.80%
$1.\mathrm{V}^{\triangle}$	151 ms	132 ms	12.58%
1.W	157 ms	129 ms	17.83%
$1.\mathrm{X.a}^{\dagger}$	103 ms	73.5 ms	28.64%
$1.\mathrm{X.b}^{\dagger}$			
$1.\mathrm{Y.a}^{\dagger}$	83 ms	73.5 ms	11.45%
$1.\mathrm{Y.b}^{\dagger}$		73 ms	12.05%
$1.\mathrm{X.a}^{*}$	103 ms	73.5 ms	28.64%
$1.\mathrm{Y.a}^{*}$	83 ms	73.5 ms	11.45%
2.y	120 ms	110 ms	8.33%
2.x	140 ms	110 ms	21.43%

1: Inter-RAT HO; **2:** Intra-RAT (LTE) HO; **a:** Indirect Tunnel; **b:** Direct Tunnel; **U:** with N26 interface
V: without N26 interface; **X:** with T-SGW; **Y:** without T-SGW; $^{\rho}$5GS to EPS; $^{\triangle}$EPS to 5GS
y: inter-MME and intra-SGW; **x:** inter-MME and S-GW; *UMTS/2G to LTE; †LTE to UMTS/2G
W: Intra-NG-RAN N2 based HO in 5G NGC

by more than 19% for all HO types over the Japanese operator data (Table V), the latency reduction over the Greek operator data (Table VI) ranges from 8.3% to 35.80%. Such differential behavior is a consequence of the varied deployment scenarios for different operators dependent on their requirements.

From the analytical results it is evident that the gains obtained for the 5G HO scenarios ($1.\mathrm{U}^{\rho}$, $1.\mathrm{U}^{\triangle}$, $1.\mathrm{V}^{\rho}$, $1.\mathrm{V}^{\triangle}$, 1.W) is significant. The reason being, the prevalence of handshakes that involve the exchange of tunnel setup information and their consequent optimization by the proposed mechanism. Specifically, the gains obtained for the scenarios that do not involve the N26 interface (scenarios $1.\mathrm{V}^{\rho}$, $1.\mathrm{V}^{\triangle}$) are significant not only due to quantitative reasons, but also because scenarios without an N26 interface will be prevalent during initial deployment scenarios. For the sake of brevity, we refer the reader to our earlier work in [17] wherein a detailed discussion and corresponding analysis for these specific scenarios has been presented. Further, we also show that the LTE-UMTS/2G HO scenarios, wherein a S-GW is being relocated (scenarios $1.\mathrm{X.a}^{\dagger}$, $1.\mathrm{X.b}^{\dagger}$), the percentage reduction in latency via the proposed mechanism is the highest amongst any other LTE-UMTS/2G HO scenarios. The aforesaid characteristic is observed because the number of messages that can be optimized through parallel message transfer and intelligent IE mapping is higher as compared to other scenarios. Concretely, during S-GW relocation process more handshakes are performed as compared to the scenario where there is no relocation, which consequently results in more avenues for optimization of the signaling messages.

Next, the handover latency improvement analysis for the handover failure phase, corresponding to the data from both operators, has been presented in Tables VII and VIII. Whilst for the Japanese operator data, the latency improvement ranges from 34.07% to 44.23%, for the Greek operator data the latency reduction is between 15.50% to 28.50%. The differential performance behavior, as mentioned earlier, is representative of the variable deployment scenarios dependent on operator requirements. Specifically, for the 5G NGC network, the HO cancellation phase signaling between NGC and EPS (scenarios $1.\mathrm{Z}^{\rho}$, $1.\mathrm{Z}^{\triangle}$) observes a latency reduction by upto 40.57%. The

TABLE VII
FAILURE PHASE: HANDOVER LATENCY IMPROVEMENT ANALYSIS (CISCO AND CELLULAR OPERATOR-JAPAN)

Handover Type	**Legacy Mechanism**	**Proposed Mechanism**	**Percentage Latency Reduction**
$1.Z^{\rho}$	112 ms	72.5 ms	35.27%
$1.Z^{\triangle}$	122 ms		40.57%
$1.X.a^{\dagger}$	104 ms	64.5 ms	37.98%
$1.X.b^{\dagger}$			
$1.Y.a^{\dagger}$	58 ms	64.5 ms	-11.20%
$1.Y.b^{\dagger}$		58 ms	0.00%
$1.X.a^{*}$	104 ms	58 ms	44.23%
$1.Y.a^{*}$	58 ms	58 ms	0.00%
2.y	89 ms	89 ms	0.00%
2.x	135 ms	89 ms	34.07%

1: Inter-RAT HO; **2:** Intra-RAT (LTE) HO; **a:** Indirect Tunnel; **b:** Direct Tunnel; $^{\rho}$5GS to EPS
$^{\triangle}$EPS to 5GS; †LTE to UMTS/2G; *UMTS/2G to LTE; **X:** with T-SGW; **Y:** without T-SGW
x: inter-MME and S-GW; **y:** inter-MME and intra-SGW; **Z:** 5G HO Cancel

TABLE VIII
FAILURE PHASE: HANDOVER LATENCY IMPROVEMENT ANALYSIS (CISCO AND CELLULAR OPERATOR-GREECE)

Handover Type	**Legacy Mechanism**	**Proposed Mechanism**	**Percentage Latency Reduction**
$1.Z^{\rho}$	130 ms	110.5 ms	15.00%
$1.Z^{\triangle}$	139 ms		28.50%
$1.X.a^{\dagger}$	92 ms	72.5 ms	21.20%
$1.X.b^{\dagger}$			
$1.Y.a^{\dagger}$	72 ms	72.5 ms	-0.69%
$1.Y.b^{\dagger}$		72 ms	0.00%
$1.X.a^{*}$	92 ms	72 ms	21.73%
$1.Y.a^{*}$	72 ms	72 ms	0.00%
2.y	109 ms	109 ms	0.00%
2.x	129 ms	109 ms	15.50%

1: Inter-RAT HO; **2:** Intra-RAT (LTE) HO; **a:** Indirect Tunnel; **b:** Direct Tunnel; $^{\rho}$5GS to EPS
$^{\triangle}$EPS to 5GS; †LTE to UMTS/2G; *UMTS/2G to LTE; **X:** with T-SGW; **Y:** without T-SGW
x: inter-MME and S-GW; **y:** inter-MME and intra-SGW; **Z:** 5G HO Cancel

aforesaid improvement is as a consequence of the presence of handshakes, whose composition and execution have been enhanced by the proposed mechanism. Recall that, as per discussions in Sections IV.B, IV.C and V.B, the worst case scenario for the HO cancel phase signaling has been considered, i.e., the HO cancel phase (Fig. 7) is executed after all the resources have been setup in the HO preparation phase.

Further, the rejection phase signaling in LTE and UMTS/2G networks for scenarios that do not involve S-GW relocation (scenarios $1.Y.a^{\dagger}$, $1.Y.b^{\dagger}$) is already optimal and does not incur any improvement or degradation through the implementation of the proposed mechanism. However, for the scenario where there is no S-GW relocation but an indirect tunnel is utilized during the LTE to UMTS/2G handover ($1.Y.a^{\dagger}$), the proposed mechanism results in a degraded performance for the handover rejection phase as compared to the legacy approach. A deeper analysis (through Fig. 3 but without the presence of T-SGW) reveals that while in the legacy mechanism the source S-GW tunnel (message 8 and 8a in Fig. 3) is not setup until the resource negotiation phase (message 5 and 5a in Fig. 3) is accomplished, the proposed mechanism, in order to obtain the advantages of parallelization, performs the source S-GW tunnel setup (message P3b in the signaling scenario specified by Fig. 8 without target S-GW) before resource negotiation. Consequently, when the RRM operation results in a handover failure, the extra source S-GW tunnel setup message in the proposed setup leads to the aforementioned performance degradation. It is important to state here that, for the corresponding scenario the handover preparation phase signaling incurs an improvement of 20.12% over the legacy approach.

Further, to de-register the resources setup by the source S-GW setup message in the proposed mechanism, a delete session request message (P6a in the signaling scenario specified by Fig. 8, but without target S-GW) from the SeMMu will also be required. And as will be seen in the next subsection (V.C.2), this will lead to a degraded performance in terms of incurred transmission cost as compared to the legacy mechanism. However, through the analysis in Section V.C.4, wherein the novel handover failure aware method is used, we will observe that this performance degradation is alleviated whilst also benefiting the handover preparation phase further.

2) Transmission Cost Analysis: Utilizing (3), we present the transmission cost analysis for the handover preparation and failure signaling phases for the different HO types specified by 3GPP (Table II) and the different deployment scenarios presented by the Japanese and Greek telecom operators (Tables III and IV). Concretely, a comparative performance analysis between the legacy and proposed mechanism for the handover preparation phase signaling has been presented in Figs. 13 and 14. Further, a similar comparative analysis for the handover failure phase signaling has also been provided through Figs.

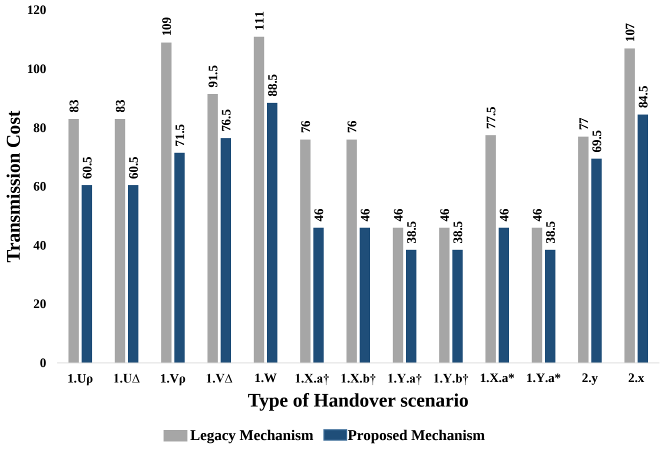


Fig. 13. Handover preparation scenario: Transmission cost analysis for the Japanese operator deployment (X axis notations have been re-utilized from Tables V-VIII).

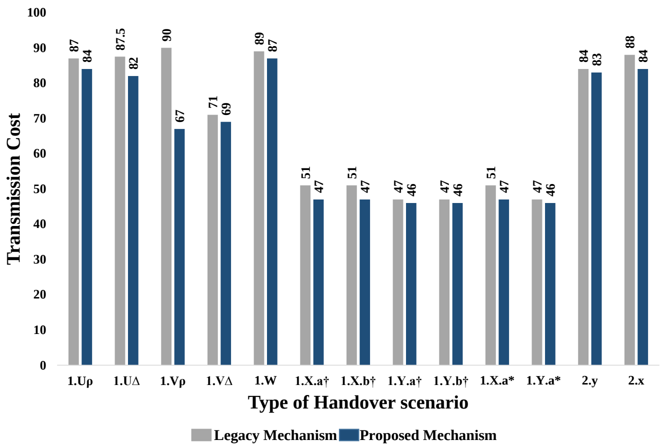


Fig. 14. Handover preparation scenario: Transmission cost analysis for the Greek operator deployment (X axis notations have been re-utilized from Tables V-VIII).

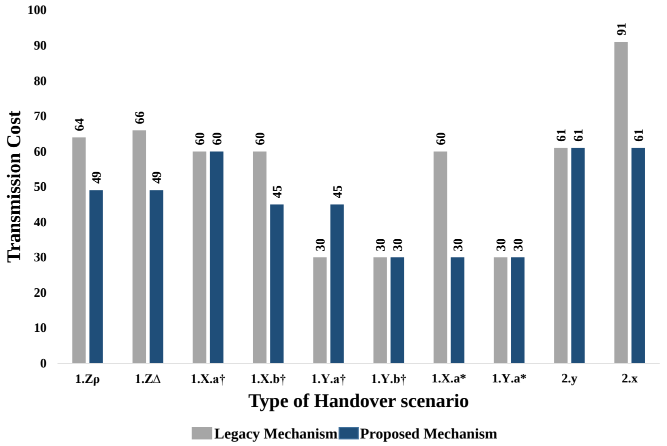


Fig. 15. Handover failure scenario: Transmission cost analysis for the Japanese operator deployment (X axis notations have been re-utilized from Tables V-VIII).

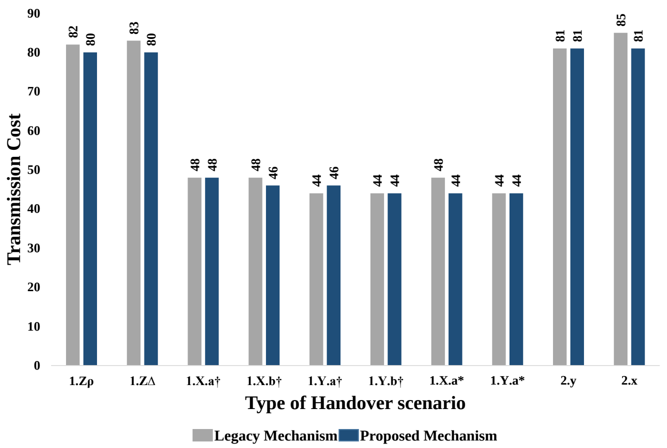


Fig. 16. Handover failure scenario: Transmission cost analysis for the Greek operator deployment (X axis notations have been re-utilized from Tables V-VIII).

15 and 16.

From the analytical results presented in Figs. 13 and 14, it is established that the proposed mechanism enhances the handover preparation phase signaling compared to the legacy mechanism, by reducing up to 40.67% in transmission cost incurred to complete the signaling process for all the considered HO scenarios. For the scenarios involving 5G NGC, the gain characteristic, i.e., the trend in performance gains, is similar to that observed for the latency improvement analysis in Section V.C.1. Further, in scenarios involving 4G/3G/2G networks, where S-GW relocation occurs (scenarios 1.X.a†, 1.X.b†, 1.X.a*), the transmission cost reduction obtained is higher than that obtained in the other legacy HO scenarios. However, as per our discussions in Section V.C.1, the gains obtained for the Greek and Japanese operator deployments are different due to the difference in the resources and requirements presented by the operators.

Next, through Figs. 15 and 16, it can be observed that the proposed mechanism either improves or does not degrade the performance of the HO scenarios considered for the HO failure phase signaling, except when there is no S-GW relocation with an indirect tunnel for a LTE to UMTS/2G Inter-RAT HO (scenario 1.Y.a†). The reason being, for the purpose of parallelization of CP information transfer from the SeMMu, the Source S-GW tunnel setup message is executed before the RRM procedure (message 3b in Fig. 10 but without the T-SGW). And, since the RRM procedure at the target network results in a handover failure, an extra message to de-register the allocated resources is required (Message 6a in Fig. 10 but without the T-SGW). Hence, these extra messages contribute towards the aforesaid degradation in performance. However recall that in Section V.C.4, we discuss the analysis where the novel HO failure aware preparation signaling has been utilized. Through the analysis we establish that the concerns of degraded performance are alleviated by this novel strategy.

3) Processing Cost Analysis: Unlike the transmission cost and latency, the processing cost is unaffected by the change in operator deployment scenarios as it solely depends on the number of messages that will be processed within the CN. Hence, in this subsection, utilizing the formulation in Section V.A as well as the proposed and legacy signaling sequences, a comparative analysis with regards to the processing cost savings offered by the proposed and legacy mechanisms has been presented via Tables IX and X, respectively.

Through the analytical results in Table IX, it can be observed that the proposed algorithm reduces the processing cost for the handover preparation phase signaling for all the considered handover scenarios. Quantitatively, the performance enhancement for the 5G NGC HO scenarios ($1.U^{\rho}$, $1.U^{\triangle}$, $1.V^{\rho}$, $1.V^{\triangle}$) ranges from 10% to 27.77%, with the scenarios without an N26 interface ($1.V^{\rho}$, $1.V^{\triangle}$) also showing improvement. In addition, the savings offered over legacy scenarios where a S-GW relocation occurs ($1.X.a^{\dagger}$, $1.X.b^{\dagger}$, $1.X.a^{*}$) is above 20%, while that offered in other legacy scenarios is 10%. Next, for the handover failure phase signaling (Table X), the proposed mechanism enhances the signaling for both the HO cancel phases in 5G NGC ($1.Z^{\rho}$: 16.67% and, $1.Z^{\triangle}$: 28.57%) as well as in two other specific scenarios, i.e., HO rejection in LTE to UMTS/2G Inter-RAT HO with Target S-GW and direct tunnel ($1.X.b^{\dagger}$: 18.18%), and HO rejection in UMTS/2G to LTE Inter-RAT HO with a Target S-GW ($1.X.a^{*}$: 36.36%). The proposed mechanism neither enhances nor degrades the performances of the failure phase signaling for other HO scenarios, except when there is an Inter-RAT HO from LTE to UMTS/2G involving an indirect tunnel and without S-GW relocation (scenario $1.Y.a^{\dagger}$). The degradation in performance for the aforesaid scenario stems from the reasons discussed in Section V.C.1 and V.C.2, i.e., the execution of source S-GW tunnel setup message before the RRM process requires an extra delete session message from the SeMMu towards the S-GW to de-register the allocated resources. Hence, this increases the number of messages to be processed by the CN, which consequently leads to the degraded performance.

4) Handover Failure aware Preparation Signaling: The analytical evaluation presented in Sections V.C.1 to V.C.3 reveals that the proposed mechanism, while enhancing the handover preparation phase, can slightly underperform for certain handover scenarios during the handover failure phase. And so, utilizing the discussions as well as the novel handover failure aware signaling from Section IV.C, we show that at least eight HO scenarios can be optimized further in terms of latency, transmission cost and processing cost. The rest of the scenarios are already optimal and hence, are not impacted by the proposed enhancement. The analytical results, presented in Table XI, show that the HO failure aware approach not only alleviates the performance degradation issue in the handover failure phase, but also enhances the handover performance phase signaling too.

Note that for the sake of brevity, the analytical results presented in Table XI utilize only the delay values from the Japanese operator deployment. Quantitatively, for the HO preparation scenarios in 5G NGC, a reduction of up to 15.38% in the processing cost over the values in Table IX is observed. However, for the 4G/3G/2G (legacy) HO scenarios, the number of messages required to complete the entire signaling process does not change, and thus, the processing cost remains unchanged. Further, for scenarios $1.X.a^{\dagger}$, $1.X.b^{\dagger}$, $1.Y.a^{\dagger}$, the number of messages required to execute the handover failure phase, i.e. HO rejection phase, is reduced significantly as compared to that specified in Table X. In addition, for the 5G NGC scenarios ($1.U^{\rho}$, $1.U^{\triangle}$, $1.V^{\rho}$, $1.V^{\triangle}$, 1.W), we consider the handover failure phase, i.e. HO cancel phase, to be near optimal owing to its sensitivity to the time at which it is initiated during an ongoing HO preparation phase, as discussed in Section V.B.

Next, the added enhancement over the proposed mechanism improves the processing cost saving for HO scenarios $1.X.a^{\dagger}$ and $1.X.b^{\dagger}$ during a handover failure phase by 36.36%, as compared to the values in Table X. Further, for the HO scenario $1.Y.a^{\dagger}$ in Table X, the processing cost saving performance is no longer degraded. Instead, the novel handover failure aware method reduces the number of messages required from 9 to 7, i.e. by 22.22%, for the proposed mechanism.

Further, the latency analysis presented in Tables V, VII and XI establishes that the HO preparation and failure phases can be enhanced further with the novel handover failure aware method proposed here. Quantitatively, the HO preparation phase corresponding to the first eight HO scenarios are further enhanced by up to 9.92% with the maximum gains being obtained for the 4G/3G/2G HO scenarios. The HO failure phase signaling for the corresponding 4G/3G/2G HO scenarios are also enhanced further by 10.07%. Consequently, the improvement in the handover failure phase signaling also alleviates the drawback of degraded performance.

Lastly, for the transmission cost, the analytical evaluation reinforces the trend of added enhancement to the performance of the handover failure phase. The transmission cost for the HO failure phase signaling in scenario $1.X.a^{\dagger}$ of Table XI is halved compared to the cost presented in Fig. 15. Further, the scenarios $1.X.b^{\dagger}$ and $1.Y.a^{\dagger}$ also experience an improvement of 33.33% in the transmission cost for their corresponding HO failure phase signaling. As a consequence, the drawback of performance degradation is also mitigated. Additionally, for the handover preparation signaling, the added enhancements facilitates an improvement ranging from 8.47% to 20.97% for scenarios $1.U^{\rho}$, $1.U^{\triangle}$, $1.V^{\rho}$, $1.V^{\triangle}$ and 1.W in Table XI, whilst the transmission cost performance of the remaining scenarios remains unaffected as compared to that presented in Fig. 15.

It is important to state here that, in Table XI, scenarios $1.Y.b^{\dagger}$, $1.X.a^{*}$, $1.Y.a^{*}$, 2.x and 2.y are not impacted by the Handover failure aware method and, hence, are referenced as *Optimal*. Concretely, the proposed mechanism without the handover failure aware methodology is already optimal for the aforementioned scenarios.

D. Message Size Analysis

The mechanism that has been proposed in this work utilizes the fact that the IEs can be intelligently re-packaged to create lesser number of messages and hence, enhance the signaling that is performed at the CN. This restructuring of the messages will also alter the size of the messages, i.e., the number of bytes carried per message, as well as the overall bytes transferred per signaling sequence. Thus, through Tables XII

TABLE IX
PROCESSING COST ANALYSIS FOR HANDOVER PREPARATION PHASE

$^{\alpha}$**Handover Type**	**Processing Cost**		
	Legacy Mechanism	**Proposed Mechanism**	**% Saving**
1.U$^{\rho}$	18 messages	15 messages	16.67%
1.U$^{\triangle}$	20 messages		25.00%
1.V$^{\rho}$	18 messages	13 messages	27.77%
1.V$^{\triangle}$	20 messages	18 messages	10.00%
1.W	18 messages	15 messages	16.67%
1.X.a†	14 messages	10 messages	28.57 %
1.X.b†			
1.Y.a†	10 messages	9 messages	10.00 %
1.Y.b†			
1.X.a*	14 messages	10 messages	28.57 %
1.Y.a*	10 messages	9 messages	10.00 %
2.y	10 messages	9 messages	10.00 %
2.x	14 messages	11 messages	21.43 %

$^{\alpha}$The notations have been re-utilized from Tables V-VIII

TABLE X
PROCESSING COST ANALYSIS FOR HANDOVER FAILURE PHASE

$^{\gamma}$**Handover Type**	**Processing Cost**		
	Legacy Mechanism	**Proposed Mechanism**	**% Saving**
1.Z$^{\rho}$	12 messages	10 messages	16.67%
1.Z$^{\triangle}$	14 messages		28.57%
1.X.a†	11 messages	11 messages	No Change
1.X.b†		9 messages	18.18%
1.Y.a†	7 messages	9 messages	-22.22%
1.Y.b†		7 messages	No Change
1.X.a*	11 messages	7 messages	36.36 %
1.Y.a*	7 messages	7 messages	No Change
2.y	7 messages	7 messages	No Change
2.x	7 messages	7 messages	No Change

$^{\gamma}$The notations have been re-utilized from Tables V-VIII

TABLE XI
HANDOVER FAILURE AWARE SIGNALING DESIGN ANALYSIS

$^{\beta}$**Type of Handover**	**HO preparation**						**HO Failure**					
	Proc. Cost$^{\eta}$	% Saving$^{\gamma}$	Latency	% Saving	Trans. Cost	% Saving$^{\gamma}$	Proc. Cost$^{\eta}$	% Saving$^{\gamma}$	Latency	% Saving$^{\gamma}$	Trans. Cost	% Saving$^{\gamma}$
1.U$^{\rho}$	14	6.67%	95 ms	0.00%	53	12.39%	The HO cancel phase, as discussed in section IV.C, is near optimal owing to its adaptability depending on when it is invoked.					
1.U$^{\triangle}$			88.5 ms	7.34%								
1.V$^{\rho}$	11	15.38%	89 ms	0.00%	56.5	20.97%						
1.V$^{\triangle}$	17	5.56%	132 ms	4.69%	69	9.80%						
1.W	14	6.67%	116.5 ms	5.28%	81	8.47%						
1.X.a†	10	0%	59ms	9.92%	46	0.00%	7	36.36%	58ms	10.07%	30	50.00%
1.X.b†								22.22%				33.33%
1.Y.a†	9	0%	59ms	9.92%	38.5	0.00%	7	22.22%	58ms	10.07%	30	33.33%
1.Y.b†	Optimal											
1.X.a*	Optimal											
1.Y.a*	Optimal											
2.y	Optimal											
2.x	Optimal											

$^{\beta}$The notations have been re-utilized from Tables V-VIII; $^{\eta}$The processing cost, defined in Section VI.1, is the number of CN messages
$^{\gamma}$ The percentage saving over the values obtained with the proposed mechanism in Tables V, VII, IX, X and Figs.15 and 17.

and XIII we provide a comparative analysis between the legacy and proposed mechanisms for the message sizes and the overall bytes transferred in a single sequence of handover signaling.

Note that for the analysis, we do not consider the 5G HO scenarios as the message sizes for 5G HO signaling are still not completely defined. Hence, we consider four representative scenarios from the LTE-EPC and UMTS/2G HO signaling, shown in Table XIII. The chosen scenarios encompass Inter- and Intra-RAT HO, relocation/no relocation of S-GW/S-GW and SeMMu, and indirect and direct tunneling, thus ensuring completeness to the analysis. Further, it must be stated that the current analysis is independent of the operator deployment scenario, and hence, it is valid for both the operator deployments considered in this work.

Table XII presents a detailed breakdown of the messages

TABLE XII

MESSAGE SIZE COMPUTATION: INTER-RAT HO FROM LTE TO UMTS/2G WHEN S-GW IS RELOCATED AND INDIRECT TUNNELING EXISTS

Msg. num.	**Legacy messages**	**Size (Bytes)**	**Msg. num.**	**Proposed messages**	**Size (Bytes)**
1	Handover Initiation	62	P1	Handover Initiation	62
2	Handover Required	302	P2	Handover Required	302
3	Forward Relocation Request	762	P3a	Resource Allocation Request + Tunnel Setup	838
4	Create Session Request	288	P4	Relocation Request	335
4a	Create Session Response	117	P4a	Relocation Request Acknowledge	130
5	Relocation Request	335	P5	Forward Relocation Response	128
5a	Relocation Request Acknowledge	130	P6a	S-SGW Tunnel Setup	105
6	Indirect Data Forwarding Tunnel Req. T-SGW	86	P6b	T-SGW Tunnel Setup	345
6a	Indirect Data Forwarding Tunnel Resp. T-SGW	111	P6c	Handover Command	166
7	Forward Relocation Response	147	P7	HO from E-UTRAN Command	63
8	Indirect Data Forwarding Tunnel Req. S-SGW	86			
8a	Indirect Data Forwarding Tunnel Resp. S-SGW	111			
9	Handover Command	166			
10	HO from E-UTRAN Command	63			
Total bytes		2766	Total bytes		2474

and their sizes in bytes for the HO preparation signaling corresponding to the scenario when there is an Inter-RAT HO from LTE to UMTS/2G and S-GW relocation along side indirect tunneling occurs. For the analysis, the message sizes corresponding to the legacy and proposed mechanism were constructed utilizing the data provided in 3GPP specifications [2], [21]–[23], [25], [26], [28], [39], [40], wireshark traces [41] and ITU ASN.1 specifications [42] (for the data types and sizes of the IEs). Through the analysis, it was deduced that since the number of messages in the proposed mechanism is reduced as compared to the legacy mechanism, the number of bytes for message headers is also reduced. Concretely, since each message that is passed through the network consists of a message header, specifying source and destination addresses/identifiers, etc., a reduction in the number of messages will also mean that there is a corresponding decrease in the amount of header that traverses through the network.

Quantitatively, the largest message in the proposed mechanism, i.e., message P3a, is 838 bytes long, while the largest message in the legacy mechanism (message 3) is 762 bytes long. Thus, the proposed restructuring process maintains the message sizes near the range of message sizes in the legacy mechanism. Consequently, it can be said that the proposed mechanism does not present any significant challenge for the reliable transmission and processing of CP messages within the network. In addition, the total amount of bytes transferred within the proposed mechanism will be 2474 as compared to 2766 bytes in the legacy mechanism, to complete the HO signaling. And hence, through the non-repetitive and intelligent repackaging of IEs into the proposed messages, the number of bytes that have to be transported across the CN for the HO scenario under observation is reduced by 10.56%. Next, and for the sake of brevity, we present the analysis for the total message bytes transferred during the legacy and proposed mechanism for the scenarios under observation (Table XIII).

TABLE XIII

MESSAGE SIZE ANALYSIS

$^{\zeta}$Type of Handover	**Total bytes for Legacy Messaging**	**Total bytes for Proposed Messaging**	**Percentage Reduction**
1.X.a†	2766	2474	10.56%
1.Y.b†	2164	2072	4.25 %
1.X.a*	2766	2493	9.87%
2.x	2817	2544	9.69%

$^{\zeta}$The notations have been re-utilized from Table V-VIII

The analytical results in Table XIII reinforce the fact that HO scenarios in which S-GW or S-GW and SeMMU relocation occurs are optimized more than the other scenarios. While scenario 1.Y.b† in Table XIII has a 4.25% reduction in the total bytes that would be transferred over the CN to complete the signaling sequence, scenarios 1.X.a†, 1.X.a* and 2.x register a reduction of 10.56%, 9.87% and 9.69% respectively. The aforementioned results also illustrate that the proposed mechanism, irrespective of the HO scenario, reduces the number of bytes that would be transferred over the CN, thus enhancing the network performance as it will have lesser bytes to transfer across the network as well as to process.

E. Network Wide Analysis

In this subsection, we present an analysis for the network occupation time and network wide processing cost savings by utilizing (5) and (6) from Section V.A, and the parameter framework presented in Section V.B. Figs. 17 and 18 illustrate the network wide occupation time and processing cost performance, respectively, for the legacy and proposed mechanisms.

Note that, the processing cost analysis is independent of the operator deployment scenario. However for the network wide occupation time analysis, for the sake of brevity, we only

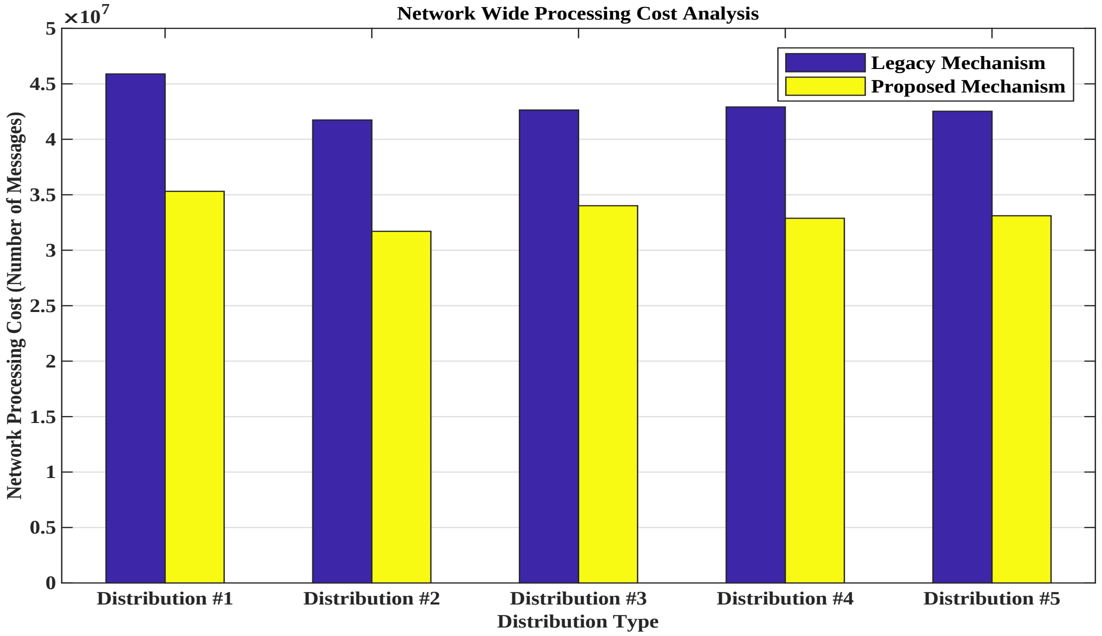


Fig. 17. Network wide processing cost analysis.

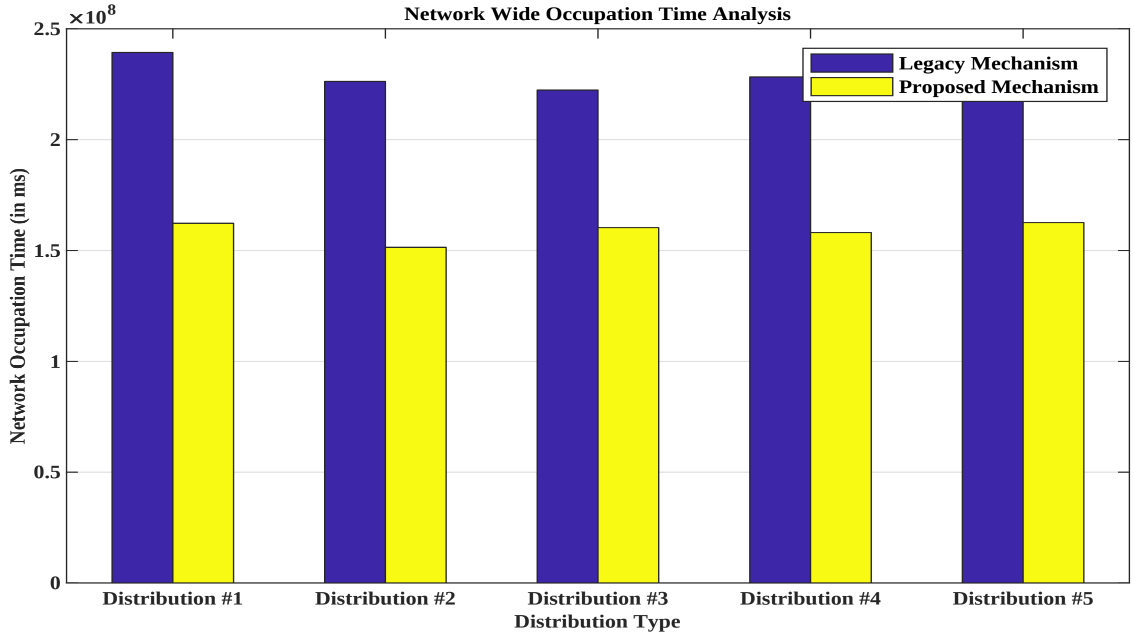


Fig. 18. Network wide occupation time analysis.

consider the delay values as obtained from the Japanese cellular operator (Table III). Further, we consider HO failure rates from 0.1%-0.5% and also vary the percentage of S1 HO (intra-MME/S-GW) from 10%-50%. Given the lack of availability of real data from the telecom operators, we randomly select a particular HO failure rate and S1 HO percentage alongside a distribution, and then compute the two metrics utilizing (5) and (6). Such an evaluation process helps to eliminate any possible bias in specifying the prevalent handover scenario, and thus aids in the completeness of the analysis.

Figs. 17 and 18 show that given any prevalent HO scenario and distribution of HO types, the proposed mechanism outperforms the legacy mechanism. Concretely, the proposed mechanism provisions a saving of 27.90%-33.06% over the legacy mechanism for the network occupation time, while for the network wide processing cost, the proposed mechanism provides a saving of 20.24%-24.05% over the legacy mechanism. And, given these significant savings in the processing cost and link occupation time, it will help the future networks, such as 5G, to be more time and resource efficient. By resource efficient here we mean that, the network will be more scalable in terms of computational and physical resources.

VI. EVOLUTIONARY 4G/5G NETWORK ARCHITECTURE

Given the performance analysis results, and specifically the network wide analysis, the benefits offered by the proposed HO mechanisms, utilizing the SeMMu, are compelling. Henceforth, in this section we present an exemplary evolutionary network architecture that not only facilitates the execution of the proposed mechanism, but also provides the operator with an avenue to have a manageable CAPEX towards evolving their networks to being fully softwarized. Thus, through Fig. 19 we illustrate the proposed evolutionary core network architecture. Note that the presented network architecture is an enhancement of the architecture utilized in [17], given that here we have also considered the N26 inter-working interface.

The proposed network architecture is evolutionary with respect to the fact that, it firstly introduces an evolved core network entity, namely, the SeMMu. The SeMMu combines the functionalities of the MME/SMF and the SDN-C. Recall that, in the proposed architecture we consider the SMF as the main 5G mobility management unit instead of the AMF, as done by 3GPP. The reason being, the SMF is involved in the CN signaling during a HO whilst the AMF is only limited to the access network resource management. Moreover, the functional integration is carried out such that the SeMMu only modifies the CP between the network entities and itself, while avoiding any impact on other core network operations (such as the DP). Additionally, while in the 3GPP defined network architecture specific interfaces are utilized to connect the network entities, in the proposed network architecture, for the SeMMu to communicate with the other core network entities, an SDN agent needs to be integrated with these other entities (such as SGSN, S-GW, UPF, etc.). Such an integration, whilst maintaining the smooth inter-working between 5G and legacy networks, also enables the operators to evolve their legacy networks towards a completely softwarized architecture. As a consequence of this evolutionary framework, the proposed architecture will help to facilitate a reduction in the CAPEX for the operators, which is a major 5G objective. The SDN capabilities also enable the proposed architecture to execute the optimized handover signaling, discussed in Sections IV and V, as it allows the SeMMu to push the required CP information to other CN entities. It is imperative to state here that, although we introduce an SDN agent overlay, we only transform the 3GPP defined functionalities of the MME/SMF whilst preserving the functionalities of all the other CN entities. Further, given any of the proposed signaling sequences and mapping, the network architecture remains the same. Concretely, the proposed evolutionary network architecture is consistent for any HO scenario. And, given the results in Section V as well as the fact that distinct flow rules per user do not require separate SDN agent threads, the network will be scalable.

Moreover, the proposed architecture is designed such that the RRM interactions are left unaltered. The reason being, if the RRM procedures are handled at the SeMMu, then while it would enable enhanced decision making given the global view the SeMMu has, it will introduce additional delays, and hence, increased latency for the handover process. Subsequently, the SeMMu is neither connected directly to the Radio Network Controller (RNC) nor to the NG-RAN. Instead, the SeMMu communicates with the SGSN/AMF, which is responsible for managing the session as well as the CP signaling with the RNC/NG-RAN. Further, within the EPC, the SeMMu allows the eNB to perform the RRM operations, even though it is directly connected to it. Lastly, the interworking framework, presented in Fig. 19, is facilitated by the presence of an N26 interface between the AMF in the NGC and the SeMMu in the EPC. Note that the interworking between 5G NGC and EPC can be established even in the absence of the N26 interface [15]–[17].

A. Benefits and Challenges

The aforesaid integration has multiple benefits as well as certain design and implementation challenges. The benefits of the SeMMu based network architecture include:

- The ability to access system parameters, which will allow the SeMMu to establish optimized MM solutions through the virtualized functions in a fully SDN architecture, via a global or locally global view of the network domain. Here domain refers to the geographical area of the network that is administered/controlled by the SeMMu.
- Introduction of SDN agents to the CN entity is a first step towards the fully SDN architecture that is envisioned for the future networks. Given the ability of an SDN controller based entity to decouple the CP from DP (and implement the rules on the DP entities), the SDN agents are utilized to push the CP information necessary for the handover related signaling to the CN entities.
- The given framework establishes an evolutionary path towards a fully softwarized network architecture. Thus, the given framework assists in reducing the CAPEX for the operators as it helps them evolve their current architecture towards a fully softwarized architecture.
- With the SDN based architecture presented in this section, the handover preparation and failure signaling phases can be optimized (Section IV) as compared to the legacy mechanisms, i.e., 3GPP standards. The optimizations obtained via signaling re-sequencing and message mapping have been elaborately discussed in Section V.

Next, the main implementation challenge arising as a consequence of the SeMMu based evolutionary network architecture, is the integration of the SDN agent to the CN entities. On one hand it will include an initial CAPEX to integrate the SDN agents and, on the other hand, new interfaces need to be defined so as to allow the MME/SMF and the SDN agent to communicate with each other. Additionally, advanced software mechanisms to identify and pack the IEs into the proposed message ensemble, discussed in Section IV, will be required. Whilst the CAPEX incurred will not be significant given the benefits offered by the SeMMu solution, in the subsequent

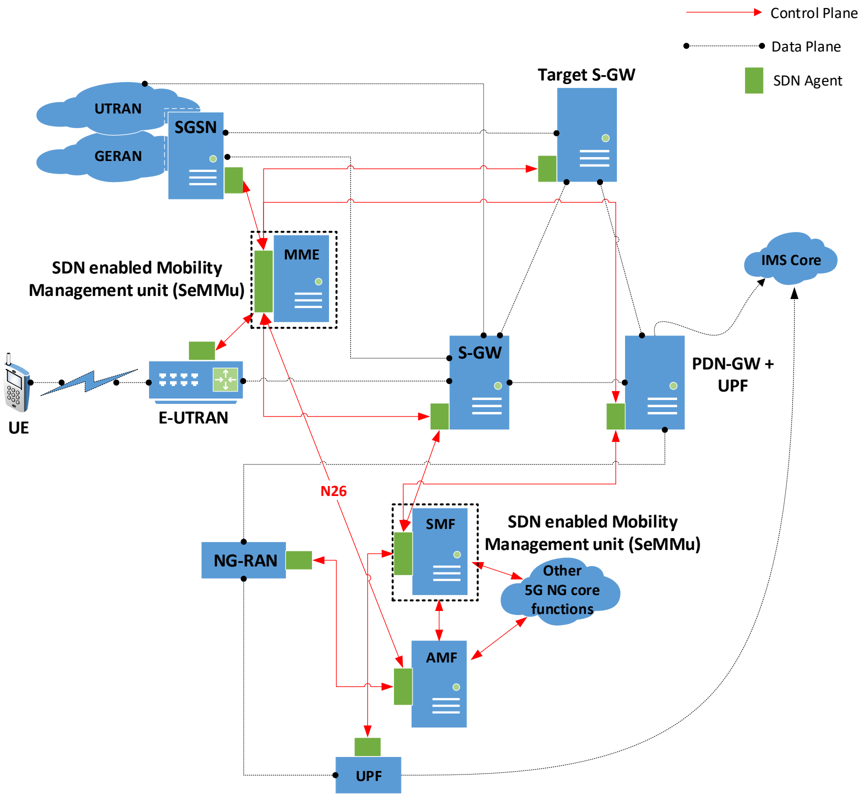


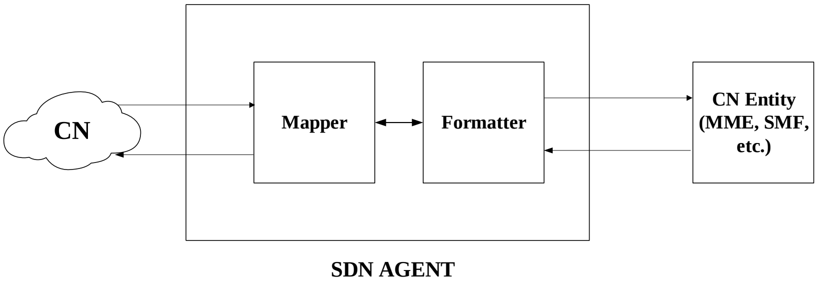
Fig. 19. Proposed evolutionary network architecture.

discussion we provide a brief insight into the approaches that can be utilized to overcome this implementation challenge.

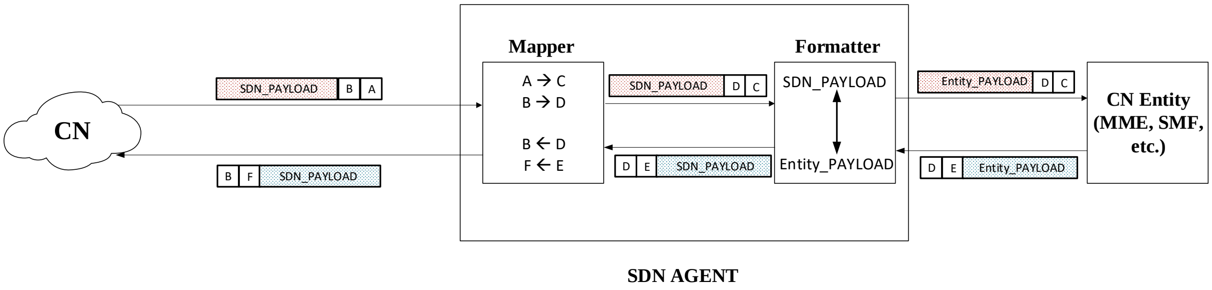
B. SDN agent integration

The integration process of the SDN agent should not disrupt the overall network functioning, design and architecture. Further, the DP operations should be agnostic to the proposed integration process. In order to realize this seamless integration, we introduce a novel setup wherein the MME/SMF entail a software modification and the SDN agent is composed of two components (illustrated in Fig. 20(a)). Concretely, the two components that constitute the SDN agent are the *Mapper* and the *Formatter*. Given that, there is an SDN agent overlay on top of the 3GPP defined network architecture, the SDN agents will view the messages and destination address in a different format as compared to the CN entities defined by 3GPP. Note that, the mapper is connected to the external network through a communication interface through which the SDN agent transmits/receives the data. In addition, the formatter is connected to the CN entity through a bidirectional communication interface for exchanging the CP information messages. The detailed functioning of both these components is provided as follows:

- *Mapper*: The mapper essentially performs a mapping and de-mapping of the address that the CN entities would observe without the SDN agent overlay to the addresses as observed by the SDN agents on the CN, and vice versa. Thus, when the mapper receives a message frame from the CN, it first removes the frame header. During this process, it identifies the message source and destination, i.e. SDN-enabled CN entity addresses, and then maps these addresses to the address of the source and destination as would be seen by the CN entities, if there were no SDN agents integrated with them. Next, it transfers the message payload along with the source and destination address to the formatter. On the other hand, when the mapper receives the messages from the formatter, it identifies the type of message, the source address and its destination address. It then maps these addresses, i.e. address that would be observed by the CN entities in the absence of the SDN agent overlay, to the address in the external CN (observed by SDN agents) and transmits it to the intended CN entity. Lastly, the application level scheduling of the messages to be sent to other CN entities is done by the scheduler present in the transformed MME/SMF CN entity, discussed later in this section.
- *Formatter*: The main task of the formatter is to transform the format of the incoming and outgoing messages according to the format expected by the SDN agent and the CN entity, respectively. For a message coming from a CN entity, the formatter changes the formatting applied by the CN entity to the one understood by the SDN agent and then passes it to the mapper. Conversely, when a message arrives at the formatter from the mapper, it formats the payload along with the source and destination address



(a) SDN agent architectural framework.



(b) SDN agent operational framework.

Fig. 20. SDN agent for the evolutionary network architecture.

into a format that can be deciphered by the CN entity.

Next, a graphical illustration of the entire message processing chain within the SDN agent has been presented in Fig. 20(b). Upon the reception of a message from another CN entity, it is passed onto the mapper. Here, the mapper firstly resolves the source and destination SDN-enabled CN entity address A and B, respectively. Concretely, the source address A is mapped to the actual source CN entity address C and, similarly, the destination address B is mapped to the actual destination CN entity address D. Upon performing this mapping, the message is then sent to the formatter. The formatter converts the message payload alongside the source and destination addresses to a format that is understood by the MME/SMF. This is then passed to the modified MME/SMF modules. On the other hand, for an outgoing message, the formatter is the first entity of the SDN agent to process it. The aforesaid processing involves transforming the outgoing message to a format that is understood by the SDN agent. It is then passed onto the mapper wherein the actual source and destination CN entity address, i.e. D and E, is mapped and replaced by its SDN-enabled CN entity address, i.e. B and F, respectively.

This discussed SDN agent architecture can be implemented as a software within the existing CN entities (in which case the mapper in the SDN agent would not be required as the address of both the SDN agent and the CN entity will be the same) or on a generic hardware platform which is interfaced with the existing CN entity hardware. While the former process can be accomplished as a software upgrade at the CN entities, the latter will require additional hardware interfacing and CAPEX for installation. Next, the MME/SMF will entail an additional software upgrade irrespective of the type of SDN agent integration. Note that, we only introduce a software upgrade on the MME/SMF since, it is one of the components of the SeMMu and hence, it will be required to execute the proposed signaling mechanism that involves transformed and compressed (in terms of number of messages) message ensemble, as discussed in Section IV. Thus, a message analyzer-generator and a scheduler component have been introduced within the MME/SMF. Fig. 21 illustrates a block diagram of the transformed MME/SMF. The message analyzer-generator component performs the function of analyzing the type of message received as well as its IEs, and then generates the appropriate response to the received information in the form of messages from the new message ensemble (Section IV). It also generates metadata that informs the scheduler about the possibility of parallelization with a given set of outgoing messages. Subsequently, the scheduler at the MME/SMF determines whether a certain set of messages have to be parallelized or not, depending on the metadata received from the message analyzer-generator block. Here parallelization refers to the fact that messages to multiple CN entities can be executed simultaneously. Hence, the scheduler in the MME/SMF determines the possibility of parallelization, and accordingly passes the set of messages to the formatter entity of the SDN agent. Given the aforesaid functionality, architecturally we define the scheduler in an MME/SMF to perform a bi-directional exchange of information with the message analyzer-generator within that MME/SMF as well as the formatter of the SDN agent integrated with its MME/SMF.

Note that, for the sake of brevity, we have not provided a graphical illustration of the message processing chain, similar to the SDN agent, for the SeMMu. The reason being, the

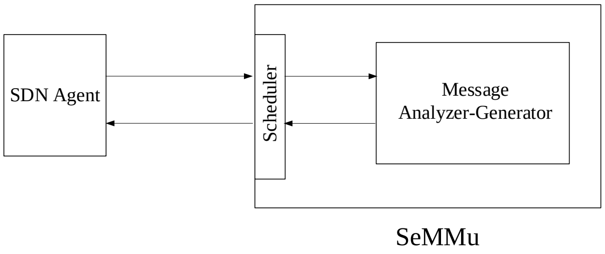


Fig. 21. SDN-enabled Mobility Management unit (SeMMu) architectural framework.

discussions in Section IV with regards to the compressed message ensemble creation and parallel transfer of HO-related CP messages, essentially presents the main functionalities of the message analyzer-generator and scheduler components, respectively, of the modified MME/SMF in the SeMMu. And so, when a message is received from the SDN agent at the MME/SMF, it is first processed by the scheduler. For the incoming message, the scheduler simply removes the headers within the received frame and passes its payload to the message analyzer-generator module. The message analyzer-generator module then:

- Analyzes the IEs of the received message.
- Determines the response message(s) and generates the required IEs.
- Generates the metadata to be forwarded to the scheduler indicating whether the outgoing message(s) can be parallelized or not.
- Formats the IEs into a message payload.
- Passes the payload along with the destination address to the scheduler.
- Passes the metadata to the scheduler.

The scheduler then forwards the messages accordingly to the SDN agent, where they are further processed according to the process illustrated in Fig. 20(b). Thus, as a consequence of this integration process, the proposed enhanced HO signaling approach can be executed, while the DP remains agnostic to these transformations.

To conclude this section we note that the SDN capabilities, provisioned to the CN entities for enhancing CP signaling, can be extended to DP functionalities such as data forwarding, path switching, etc. The provision of such an extension enables the proposed architecture to be evolutionary in nature, acting as a bridge between current and envisioned future networks.

VII. CONCLUSION

In this paper, we have firstly proposed the enhanced messaging mechanism, wherein we transform the critical HO preparation and failure signaling phases for the various 5G NGC and LTE-EPC Inter- and Intra-RAT HO (involving 5G, 4G, 3G and 2G networks) scenarios. We establish a set of principles that allows us to restructure the messages corresponding to the aforesaid signaling phases. This restructuring helps in compressing the message ensemble and in enabling parallel execution of the messages. Further, a latency, transmission cost, processing cost and message size analysis is conducted, which concludes that the proposed mechanism enhances the legacy handover signaling significantly. We also provision a novel HO failure aware signaling methodology, which accounts for the possibility of a HO failure in the design of the HO preparation signaling. The aforesaid novel strategy is proven to enhance both the preparation as well as the failure phase signaling. Further, and as a means to exemplify the superiority of the proposed mechanism, we present a network wide analysis. Through this analysis we have demonstrated that, for large number of users, the proposed mechanism outperforms the legacy mechanism both in terms of the total processing cost as well as the amount of time the network is occupied to transfer the HO preparation/failure messages.

Lastly, we have proposed an exemplary novel evolutionary architecture that consists of an evolved CN entity, namely, the SeMMu. The evolutionary characteristic of the proposed mechanism helps to maintain a manageable CAPEX. It also facilitates the execution of the aforementioned enhanced HO signaling.

Thus, to conclude, in this article we have advanced the work in the area of handover signaling by accomplishing, and verifying analytically, strategies that enhance the process of handover management in terms of latency, processing and transmission overhead. Given the fact that handover management is a critical component of mobility management, this article provisions enhanced mobility management mechanisms, that can cater to future network requirements, for the operators and vendors.

REFERENCES

[1] K. Alexandris, N. Nikaein, R. Knopp, and C. Bonnet, "Analyzing X2 handover in LTE/LTE-A," in *14th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wirel. Networks (WiOpt)*, May 2016, pp. 1–7.

[2] 3GPP, "LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (3GPP TS 23.401 version 13.9.0 Release 13)," Tech. Rep., 2015.

[3] K. Dimou *et al.*, "Handover within 3GPP LTE: Design principles and performance," *IEEE Veh. Technol. Conf. (VTC 2009)*, pp. 1–5, 2009.

[4] Ericsson, "Ericsson Mobility Report," Tech. Rep., 2016. [Online]. Available: https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf

[5] Ericsson, "Ericsson Mobility Report," Tech. Rep., 2018. [Online]. Available: https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf

[6] A. F. Almutairi, M. A. Landolsi, and A. O. Al-Hawaj, "Weighting Selection in GRA-based MADM for Vertical Handover in Wireless Networks," in *UKSim-AMSS 18th Int. Conf. on Comput. Model. Simul. (UKSim)*, Apr. 2016, pp. 331–336.

[7] S. Barmpounakis, A. Kaloxylos, P. Spapis, and N. Alonistioti, "Context-aware, user-driven, network-controlled {RAT} selection for 5G networks," *Comput. Networks*, vol. 113, pp. 124–147, 2017.

[8] A. Habbal, S. I. Goudar, and S. Hassan, "Context-Aware Radio Access Technology Selection in 5G Ultra Dense Networks," *IEEE Access*, vol. 5, pp. 6636–6648, 2017.

[9] M. Lahby, L. Cherkaoui, and A. Adib, "An enhanced-TOPSIS based network selection technique for next generation wireless networks," *Int. Conf. Telecommun. (ICT 2013)*, pp. 1–5, 2013.

[10] M. Zekri, B. Jouaber, and D. Zeghlache, "A review on mobility management and vertical handover solutions over heterogeneous wireless networks," *Comput. Commun.*, vol. 35, no. 17, pp. 2055–2068, Aug. 2012.

[11] D. Singhal *et al.*, "LTE-advanced: Handover interruption time analysis for IMT-A evaluation," *Int. Conf. Signal Process. Commun. Comput. Netw. Technol. (ICSCCN-2011)*, pp. 81–85, 2011.

[12] Q. Wang, S. Zhao, and C. Hou, "UE assisted mobility management based on SDN," in *11th Int. Conf. Comput. Sci. Educ.(ICCSE)*, Aug. 2016, pp. 689–695.

[13] M. Kantor, R. State, T. Engel, and G. Ormazabal, "A policy-based per-flow mobility management system design," in *Proc. Princ. Syst. Appl. IP Telecommun. - IPTComm '15*, 2015, pp. 35–42.

[14] L. Wang *et al.*, "An SDN-based seamless convergence approach of WLAN and LTE networks," in *IEEE Inf. Technol. Networking, Electron. Autom. Control Conf.*, vol. 12, May 2016, pp. 944–947.

[15] 3GPP, "TS 23.502: Procedures for the 5G System (Stage 2, Release 15)," 2017.

[16] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "Enhanced Handover Signaling through Integrated MME-SDN Controller Solution," in *87th IEEE Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–7.

[17] A. Jain, E. Lopez-aguilera, and I. Demirkol, "Improved Handover Signaling for 5G Networks," *PIMRC Workshop 2018*, 2018.

[18] J. Costa-Requena, J. Llorente Santos, J. Manner, and R. Kantola, "Enhanced Mobility Management (US 2017/0041824 A1)," 2015.

[19] A. Jain, Sadagopan N S, S. K. Lohani, and M. Vutukuru, "A comparison of SDN and NFV for re-designing the LTE Packet Core," in *IEEE Conf. Netw. Funct. Virtualization Softw. Defin. Networks*, 2016, pp. 74–80.

[20] A. Mahmoud *et al.*, "Software-defined networking approach for enhanced evolved packet core network," *Int. Jour. of Commun. Syst.*, vol. 31, no. 1, pp. 1–15, 2018.

[21] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3 (3GPP TS 29.274 version 12.6.0 Release 12)," pp. 0–37, 2014.

[22] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3 (3GPP TS 24.301 version 10.3.0 Release 10)," 2011.

[23] 3GPP, "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Mobile radio interface Layer 3 specification; Core network protocols; Stage 3 (3GPP TS 24.008 version 8.6.0 Release 8)," 2009.

[24] 3GPP, "Universal Mobile Telecommunications System (UMTS); Radio Resource Control (RRC); Protocol specification (3GPP TS 25.331 version 13.1.0 Release 13)," 2016.

[25] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (3GPP TS 36.331 version 13.0.0 Release 13)," 2016.

[26] 3GPP, "Universal Mobile Telecommunications System (UMTS); UTRAN Iu Interface RANAP Signalling (3GPP TS 25.413 version 3.4.0 Release 1999)," 2000.

[27] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Evolved Packet System (EPS); 3GPP Sv interface (MME to MSC, and SGSN to MSC) for SRVCC (3GPP TS 29.280 version 8.3.0 Release 8)," 2010.

[28] 3GPP, "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Numbering, addressing and identification (3GPP TS 23.003 version 10.2.0 Release 10)," 2011.

[29] 3GPP, "TS 23.501: System Architecture for the 5G System (Stage 2, Release 15)," 2017.

[30] 3GPP, "3rd Generation Partnership Project TR23.799; Technical Specification Group Services and System Aspects; Study on Architecture for Next Generation System (Release 14)," 2016.

[31] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "Optimized Handover Preparation Signaling-Other Scenarios," pp. 1–38, 2018. [Online]. Available: https://wireless.upc.edu/en/research/documentation/hosig5g4g_otherscenarios.pdf

[32] 3GPP, "3rd generation partnership project;technical specification group radio access network; nr; nr and ng-ran overall description; stage 2 (release 15)," 2017.

[33] V. S. Rao and R. Gajula, "Interoperable UE Handovers in LTE," pp. 1–11, 2011.

[34] S. Chourasia and K. M. Sivalingam, "SDN based Evolved Packet Core architecture for efficient user mobility support," in *Proc. 2015 1st IEEE Conf. Netw. Softwarization (NETSOFT 2015)*, Apr. 2015, pp. 1–5.

[35] S. Oh, B. Ryu, and Y. Shin, "EPC signaling load impact over S1 and X2 handover on LTE-Advanced system," *2013 3rd World Congr. Inf. Commun. Technol. WICT 2013*, pp. 183–188, 2014.

[36] S. Abe, G. Hasegawa, and M. Murata, "Design and performance evaluation of bearer aggregation method in mobile core network with C/U plane separation," in *2017 IFIP Netw. Conf. (IFIP Networking) Work.*, Jun. 2017, pp. 1–8.

[37] Z. Savic, "LTE Design and Deployment Strategies," *Cisco*, pp. 1–79, 2011.

[38] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Domain Name System Procedures; Stage 3 (3GPP TS 29.303 version 10.1.0 Release 10)," 2008.

[39] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification (3GPP TS 36.323 version 12.4.0 Release 12)," 2015.

[40] 3GPP, "Digital cellular telecommunications system (Phase 2+) (GSM); Mobile radio interface layer 3 specification; Radio Resource Control (RRC) protocol (3GPP TS 44.018 version 13.1.0 Release 13)," 2016.

[41] ng4t GmBH, "NG4T project wireshark traces." [Online]. Available: http://www.ng4t.com/wireshark.html

[42] O. Dubuisson, *ASN.1 Communication between Heterogeneous Systems*, 2000.