Received: 15 May 2017 | Revised: 11 June 2017 | Accepted: 13 June 2017
DOI: 10.1002/dac.3379

RESEARCH ARTICLE

WILEY

Software-defined networking approach for enhanced evolved packet core network

Ashraf Mahmoud | Ahmad Abo Naser | Marwan Abu-Amara | Tarek Sheltami | Nidal Nasser

Department of Computer Engineering, KFUPM, KSA

Correspondence
Ashraf Mahmoud, Department of Computer Engineering, KFUPM, KSA.
Email: ashraf@kfupm.edu.sa

Summary
The evolved packet core (EPC) network is the mobile network standardized by the 3rd Generation Partnership Project and represents the recent evolution of mobile networks providing high-speed data rates and on-demand connectivity services. Software-defined networking (SDN) is recently gaining momentum in network research as a new generation networking technique. An SDN-based EPC is expected to introduce gains to the EPC control plane architecture in terms of simplified, and perhaps even software-based, vendor independent infrastructure nodes. In this paper, we propose a novel SDN-based EPC architecture along with the protocol-level detailed implementation and provide a mechanism for identifying information fields exchanged between SDN-EPC entities that maintains correct functionality with minimal impact on the conventional design. Furthermore, we present the first comprehensive network performance evaluation for the SDN-based EPC versus the conventional EPC and provide a comparative analysis of 2 networks performances identifying potential bottlenecks and performance issues. The evaluation focuses on 2 network control operations, namely, the S1-handover and registration operations, taking into account several factors, and assessing performance metrics such as end-to-end delay (E2ED) for completion of the respective control operation, and EPC nodes utilization figures.

KEYWORDS
EPC, experiment design, LTE, simulation, software-defined network

1 | INTRODUCTION

Mobile networks are increasingly pushed to the limits with exponential increase of data flow across the network. In addition, new types of services are being deployed with strict latency constraints that motivate researchers to investigate new network technologies and paradigms in an attempt to improve network performance. However, network migration to conceptually new technologies always presents new technical challenges that demand careful analysis and evaluation to prove efficiency and optimality of the new proposals.

The evolved packet core (EPC) network is the core network for the fourth generation (4G) mobile network.[1] Its wide adoption by operators globally signifies its central role for determining the performance and quality of the services provided by the overall 4G mobile network. Therefore, for operators as well as subscribers, further enhancements to the EPC network is a critical and enduring requirement. On the other hand, software-defined networking (SDN) is recognized as one of the new generation networking technologies that aim to enhance network scalability and simplify its management and operations.[2,3] Therefore, it is natural to find recent proposals in the literature that propose various SDN-based EPC architectures similar to those found by several authors.[4,5,6] Very recently, SDN is also being considered for the core of the fifth generation (5G) mobile networks as well.[7,8] In addition to the gains in terms simplified and perhaps even software-based infrastructure nodes, a cost saving in capital expenditure is expected when SDN architecture is employed in the mobile network.[9] However, most of the reviewed proposals did not provide an adequate protocol-level analysis of the proposed architecture, and most of these studies presented a descriptive analysis of an SDN-based EPC network without addressing the modifications required on the EPC signaling necessitated by such proposals. Furthermore, only a few of the proposed approaches attempted to provide a quantitative evaluation of the SDN-based realization of the EPC network, and none of the few quantitative evaluations found were for a mature SDN-based EPC approach, as detailed in Section 3.

In this work, we provide (1) a detailed qualitative analysis and comparison of EPC and SDN architectures, (2) a detailed proposal for a novel SDN-based EPC network design, (3) the first thorough quantitative performance evaluation of conventional EPC network versus our proposed network, and (4) a reflection and discussion of the findings to aid in the deployment of such solutions. The detailed performance study follows a systematic engineering approach where factors expected to affect the performance metric of interest are numerated, accounted for, and then the importance of each of these factors is quantified.

The remaining of this paper is organized as follows: Section 2 provides a brief background, while Section 3 describes the related work in the area of SDN-based EPC network. Section 4 provides a conceptual overview of our SDN-based EPC architecture interpretation and the proposed SDN-based EPC. The quantitative evaluation of the proposed network and the main results are provided in Section 5. Finally, Section 6 presents the conclusion.

2 | BACKGROUND

2.1 | Mobile network architecture

The fourth generation long-term evolution (LTE) network architecture includes an EPC as the core network and an evolved UMTS terrestrial radio access network (E-UTRAN), which consists of evolved NodeBs (eNBs) that serve as the radio access points. The core network and E-UTRAN are connected through backhaul transport network. Figure 1 shows an abstract view of an LTE-EPC network. The main functionality of an EPC network is supported by 3 principle entities[1]: a serving gateway (SGW), a packet data gateway (PGW), and a mobility management entity (MME). The users' equipment (UE) connects to eNBs that forward data traffic to the SGW. At any given instance, the UE is connected to only 1 SGW that serves as the local mobility anchor for handovers between eNBs, and forward the data traffic toward the PGW. Note that UEs are allowed to be simultaneously connected to multiple PGWs that are possibly connected to different service providers. The PGW provides connectivity for an EPC to external packet data networks such as the Internet. Moreover, it performs policy enforcement, per user packet filtering, and charging support. On the other hand, the MME is the key control entity and is responsible for registration signaling, bearer establishment, and performs packet data network and SGW gateway selection for UEs at initial attachment and handover operations.

2.1.1 | GPRS tunneling protocol (GTP)

The GTP protocol collections are used for communication and management among the major EPC nodes such as SGW, PGW, and MME.[1] Tunneling mechanisms, seamless mobility procedures, and users' data forwarding in a 3rd Generation Partnership Project (3GPP) network depend on the GTP protocol. The GTP protocol stack operates on top of the user datagram protocol (UDP) protocol, making the EPC network an overlay network with respect to the underlying IP network. GTP processing capability in network devices are required to participate in EPC network operations. Therefore, conventional regular Internet protocol (IP)–based platforms, eg, routers, cannot be part of such network. The GTP protocol comprises a control part (GTP-C) and a user part (GTP-U) as depicted in Figure 2. The GTP-C is utilized for signaling among MME, SGW, and PGW

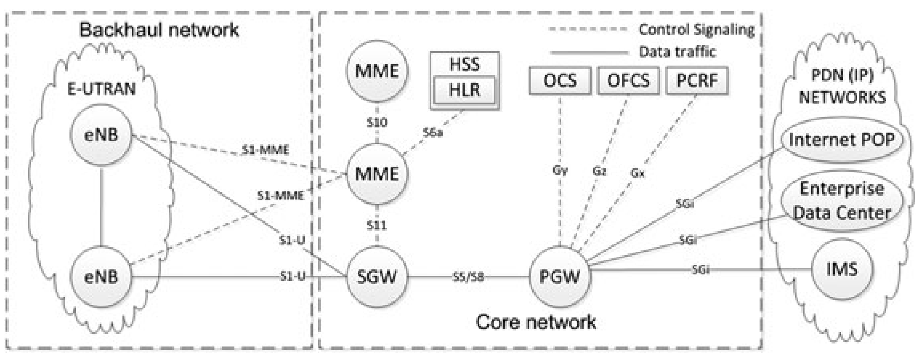

FIGURE 1 LTE-EPC architecture[1]

FIGURE 2 GTP protocol stacks

and supports the establishment of tunnel end points, bearer contexts for user specific information, and quality-of-service (QoS) management. On the other hand, the GTP-U is the data plane part responsible for encapsulation and tunneling user's IP packets.

2.2 | SDN and OpenFlow

The principle concept of SDN is to separate the control plane from the data plane and concentrate the control plane functionality of network devices in a logically centralized controller.[3,10] This control plane architecture introduces new perspectives into network design. Recently, many proposals emerged for applying the concept of SDN to the EPC network. An SDN-based EPC architecture is investigated to introduce benefits provided by the SDN paradigm into the EPC. These benefits are dependent on engineering design decisions made for the new scheme. Simplification of network devices is a natural outcome of the SDN design because a substantial part of the control operation is no longer performed in network devices but rather concentrated in a controller in a central location. This centralization also leads to simplification of network management and operations. Simplified network devices promote cheaper, easier to configure, more distributed, and possibly better performing network infrastructure.[9.11] In addition, the SDN is sought as an enabler for network function virtualization, which allows the migration from hardware-based to software-based network devices and thus accelerating service innovation and creation. Examples of recent efforts to influence the design of 5G mobile networks utilizing network function virtualization can be found.[12,13,14,15]

Control operations, typically carried out in network elements, decide the type of packet processing that is needed such as forwarding, modifying, or dropping packets. The SDN principle modifies the control plane by separating the network control plane functionality from the forwarding elements such as routers and switches. The SDN then relocates these operations into a centralized entity responsible for the network intelligence called the SDN controller. Hence, in the SDN realm, the control plane is decoupled from the data plane and shifted into a centralized controller as shown in Figure 3. This centralization of the control architecture can reduce network complexity, introduce flexibility, and exploit centrality of control information.[2,3]

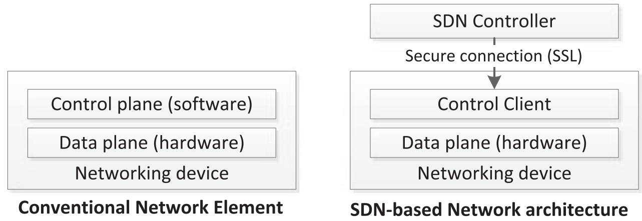

FIGURE 3 SDN architecture

The most widespread open SDN standard that has industry adoption is the OpenFlow (OF) protocol.[16] The OF protocol can be described as a forwarding table management protocol. A group of forwarding tables is maintained in each OF-based network forwarding element, which consists of forwarding/matching rules, called flow rules. Flow rules dictate the operation performed on each packet that matches a flow rule upon arrival. An OF controller supplies flow table rules to OF-based devices such as a hub, a switch, or even a router.

3 | SDN-EPC IN RELATED WORK

In this section, we present the previous views of SDN realizations in the EPC network and address points that differ from our interpretation. It should be noted that SDN may be introduced in the 2 main network areas for mobile network: the backhaul network and the core network. In the backhaul network, traffic engineering is enhanced through SDN's centralized control of the backhaul. Also, SDN can potentially aid in the mobility procedures of users' traffic.[17] However, an SDN-based backhaul does not necessarily imply an SDN-based EPC architecture, but it certainly has an overall impact on the performance of the backbone transport network.

Previous works[4,18,19] model the GTP protocol as an additional layer to the OF protocol, where a new extension to OF is proposed to accommodate EPC operations in the OF paradigm. The OF controller is then co-located with the MME. This architecture transforms an OF device to a GTP-capable device for participation in EPC network operations. However, this change does not necessarily exploit the inherent SDN features of control plane centrality. In addition, the proposed method does not justify the use of lower network layers, ie, layers 2 to 4, nor the use of the flow-based forwarding process by OF functionality in EPC nodes.

Pentikousis et al[6] provided 2 strata for SDN in the EPC architecture. First, an SDN-enabled mobile network accommodating EPC nodes controlled by a custom controller that is supported by custom interfaces. Second, an OF-enabled transport network, which is controlled by another, but OF-based, controller that cooperates with the SDN mobile network's custom controller. Forwarding elements in the mobile network stratum are composed of custom forwarding elements that meet carrier grade functionality, which also advertise a flat architecture. Their work in essence is conceptually similar to this work interpretation of an SDN-EPC. However, details of EPC nodes control functionality transformation toward a centralized controller are not provided. Although the work provides an actual implementation, it does not provide any quantitative evaluation of the new system.

Gurusanthosh et al[17] presented a semidistributed mobility scheme based on OF. They argue for dynamically delegating part of mobility management such as anchor points to the backhaul network. As a result, a large part of the routing path in the backhaul network is unchanged when inter-eNB handovers occur. The backhaul network is realized through OF-enabled switches controlled by OF controllers. The proposed architecture focuses on the SDN-based backhaul network elements and how these elements can participate with the EPC network in traffic management and mobility procedures. The proposal does not capitalize on the inherent EPC architecture and only provides an SDN-assisted EPC network. Moreover, it suggests the removal of the GTP tunnels used in the data path and relying solely on OF operations. Their proposal to dispose of the well-established GTP tunnel is non-3GPP compliant.

Most previous efforts focus on OF as the control plane architecture for the SDN-based EPC network. This perspective in our view is not necessary for the evolution of the EPC network, as shall be presented in the following sections. An SDN-based EPC network can capitalize on the SDN architecture without direct utilization of the OF protocol.

4 | SDN-EPC—CONCEPTUAL OVERVIEW

The motivation behind an SDN-based EPC is to leverage the central control entity with network wide overview while maintaining the functionality of distributed data plane devices utilizing simpler platforms. In accordance with this goal, the central entity should take charge of all control operations in the distributed gateways, ie, the controller directly controls the gateways. This requires a proper analysis of the control plane in the EPC architecture and derivation of a proper mechanism for control plane transformation that maintains EPC functionality and 3GPP compliance. This section presents analysis of the conventional EPC architecture along with a comparison with an OF architecture in terms of control plane architecture, underlying transport protocol, protocol layers required for functionality, and platform requirements. The section also provides this work's interpretation of an SDN-based EPC network.

4.1 | Analysis and comparison of EPC and OF architectures

4.1.1 | Control plane architecture (control plane nodes' role)

The concept of centralization ideally conveys the meaning of the presence of a central entity that directly communicates with distributed forwarding entities. Such entities align with the OF controller and the OF switches, respectively. Reflecting on this centrality concept for the EPC architecture, it may be observed that centrality is already intrinsic in this architecture and its operations. For example, the MME is responsible for establishment, maintenance, and removal of users' connections in the EPC gateways. Thus, the MME acts as the controller and the EPC gateways act as the forwarding nodes. However, 1 major difference between the EPC architecture and a true SDN architecture is that the PGW node is not directly controlled by the MME but rather through the SGW node. This creates a trombone path for the control traffic of the PGW as seen in Figure 4. That is the control traffic for the PGW must pass through the SGW node first. On the other hand, in an SDN-based EPC the MME should control the PGW directly as depicted in Figure 5.

4.1.2 | Control plane communication protocol (TCP vs UDP)

Communication between the controller and the forwarding nodes in an SDN architecture is based on reliable communication. This mechanism is used to ensure the reliability of network operations and to isolate network control operations from the unreliability of the underlying: physical layer L1 and media access layer L2. This reliable communication in the OF protocol is carried over secure TCP connections in a one-to-one communication pattern between the controller and the OF switches. In the EPC architecture, the control plane communication is carried using the GTP protocol over the UDP protocol, and using the S1 application protocol (S1AP) over the stream control transmission protocol (SCTP)

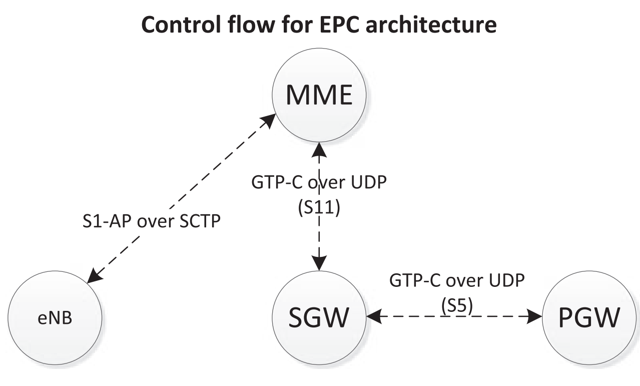

FIGURE 4 Control plane architecture in conventional EPC

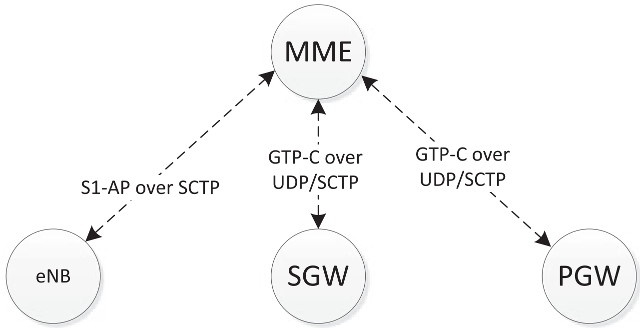

FIGURE 5 Control plane architecture in SDN-based EPC

protocol as shown in Figure 4. The reliability of the control communication to-and-from the SGW and the PGW does not depend on the underlying UDP transport layer but rather on the application layer of the MME. The SDN-based EPC should take into consideration that reliable communication is essential to mitigate the effects of network errors, particularly when adopting distributed network deployment strategies of the EPC gateways.

4.1.3 | Protocol layers required for functionality

The OF forwarding depends on processing packets fields related to layers 2, 3, and 4, which matches typical network devices protocol stacks except for layer 4, the transport layer. Hence, the OF protocol stack is targeted for processing traffic originating and destined to network devices that use the same protocol stack. The EPC, on the other hand, has a different network protocol stack that is based on the GTP protocol as presented in Figure 2. It acts as an overlay network to the typical network devices. As such, all EPC traffic is encapsulated in UDP packets except for control packets between the MME and the eNBs which uses SCTP packets for increased reliability. Thus, users' traffic processed in the SGW and the PGW use IP addresses that are only accessible by the EPC overlay network. This indicates that the default OF protocol stack is not capable of directly running EPC operations. Nonetheless, the OF can be seen as a complementary extension to EPC nodes that increases its functionality.

There is a direction in the research community to explore the GTP extensions for the OF protocol to allow it to process GTP packets and effectively making the OF switches part of the overlay network.[4,19] The advantages of this approach include establishing a homogeneous control of the network in the case of a complete migration of a network to an OF-based realm and benefiting from OF features such as flexible control of traffic flows and QoS. However, it is still early in the industry to decide for a complete migration to OF networks as some of the OF features such as flexible control of traffic flows are already available in the EPC through user's connection bearer contexts and traffic flow templates processing.

4.1.4 | Platform requirement

Costumed IP-based router platforms are often used to implement SGW and PGW nodes. These platforms are characterized by a hardware-optimized architecture that is typically composed of separate and dedicated control processors for control plane processing. In addition, the hardware architecture has a large quantity of network processors dedicated for high-speed and high-throughput packet processing and forwarding. These gateways hold large volumes of bearer contexts that are required to provide UEs with multiple connections with different QoS parameters. On the other hand, OF switches have different paradigm for control plane processing. They depend on the process of matching fields in a group of look up tables. There has been hardware and software based OF platforms investigated by the research community, each with its pros and cons.[20] The OF hardware based switches, such as NetFPGA,[21] have shown high performance in terms of processing control operations, while the OF software based switches use standard industry commercial off-the-shelf (COTS) servers.[20] The use of look up tables in the OF switches puts constraints on memory requirements and processing time in the gateways. Subsequently, the overall performance might not be optimal when compared to the conventional EPC platforms.

4.2 | Proposed SDN-based EPC architecture

Based on analysis of the EPC and the OF architecture and comparison between them, a novel SDN-based EPC architecture is designed that captures the essence of the SDN architecture taking into account the inherent properties of EPC. The proposed architecture covers the main topics discussed in previous subsection to reach a mature proposal.

4.2.1 | Control plane architecture (control plane nodes' role)

The SDN-based EPC should adhere to the concept of control centrality. Therefore, EPC gateways should be controlled directly by the controller, which is a task that may be performed by the MME as pointed out earlier. The SGW and the eNBs are directly controlled by the MME. On the other hand, the PGW is dependent on the SGW operations since the MME communicates with the SGW that in turn communicates with the PGW, if required. Therefore, the PGW should be controlled directly by the MME by relocating the control plane operations occurring between the SGW and the PGW to be managed by the MME directly. To achieve this objective, it should be noted that the control operations between the SGW and the PGW are referenced in the S5 interface in the 3GPP networks.[1] The S5 interface specifies information elements in GTP-C messages exchanged between the SGW and the PGW gateways on each control operation event that are required for the PGW functionality. Hence, the new proposed architecture affects fundamentally the S5 interface as the control signaling would be initiated from the MME rather than from the SGW.

The information fields exchanged between gateways are classified based on their final destination in the conventional EPC into 2 groups: information required at both the SGW and the PGW and information required at the SGW only or the PGW only. The information required at PGW can be characterized as tunnel end points, users' bearer contexts, and state information such as charging identifications. In the SDN-based EPC, all information fields are generated and managed by the MME, and the information fields required at both gateways such as tunnel end points and user traffic identifications are sent to both gateways to maintain consistent and correct functionality. On the other hand, information fields required at only 1 gateway, such as charging identification that is required at the PGW only, is sent only to the respective gateway. It should be noted that these changes do not violate the GTP protocol as the protocol allows for additional optional fields to be defined in the existing GTP-C messages. However, this would require changes to the protocol processing procedure to adapt to the new source of information fields.

Imposing these modifications should affect all operations of the control plane. As such, this will affect the sequence of message exchanges between EPC nodes, in addition to affecting the information fields contained in these messages. Of particular interest to this study are the following 2 prominent control operations: the initial attachment operation and the S1-based handover operation. The initial attachment operation, also referred to by the registration operation, occurs when the user equipment first connects to the mobile network, whereas the S1-based handover operation occurs when a relocation of the anchor SGW node is required.[1] Below we show how these 2 representative control operations are transformed when implemented by the proposed SDN-based EPC network.

Figure 6 shows the conventional registration operation as prescribed by 3GPP, which involves direct signaling between the SGW and the PGW.[1] Applying the above-mentioned transformation to this operation, the "Create Session Request" is now sent directly from the MME toward the PGW where all the required information for PGW processing is provided by the MME and not the SGW. Thus, the PGW does not depend on the SGW to create its bearer entries. Furthermore, the SGW would depend on the MME in creating its bearer entries because the MME is now responsible for bearer management as opposed to the gateways themselves. The proposed message flow for SDN-based registration operation is as shown in Figure 7. A related control operation is the bearer modification operation. This operation is required when users modify their service requirement or initiate new service requests. The modification message flow is exactly the same as registration operation message flow but has the GTP messages "Modify Bearer Context" instead of "Create Session."

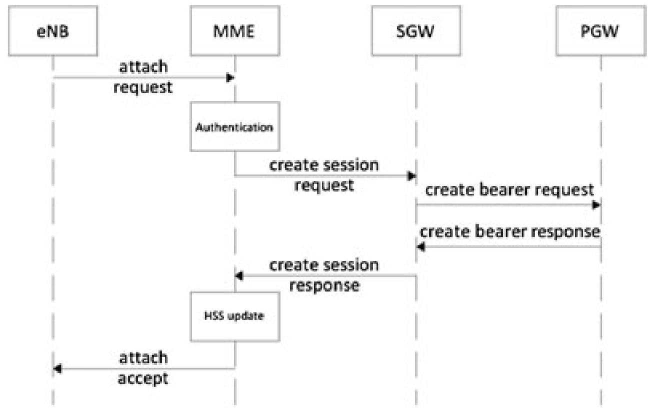

FIGURE 6 Registration operation in conventional EPC

In a similar manner, we also transform the message flow for the conventional S1-handover operation specified by 3GPP, not shown here for the sake of brevity, to produce the SDN-based procedure shown in Figure 8. The core changes are when the "Handover Notify" is received at the MME and the MME sends a "Modify Bearer Request" to the target SGW, which in turn passes the signal to the PGW. The corresponding responses flow in the opposite direction. Applying the centralized concept of SDN, we identify the information required at the PGW and send it simultaneously with the "Modify Bearer Request" message to the PGW. The PGW in turn sends its response directly toward the MME without the SGW intervention.

Naser[22] detailed and presented the information fields in the affected GTP messages required for the 2 proposed SDN-based control operations: the registration operation and the S1-based handoff operation. Therein, we highlight the proper modification needed to carry out the needed signaling and specify the respective fields that should be added or removed from modified GTP messages.

4.2.2 | Control plane communication protocol (TCP vs UDP)

The control plane communication in the conventional architecture is operating using the GTP-C protocol over UDP and the S1AP protocol over SCTP. The S1AP interface is kept unchanged because there is direct reliable communication between the eNBs and the MME and it already uses standardized interfaces. For the proposed architecture, we investigate 2 possible transport protocols to encapsulate

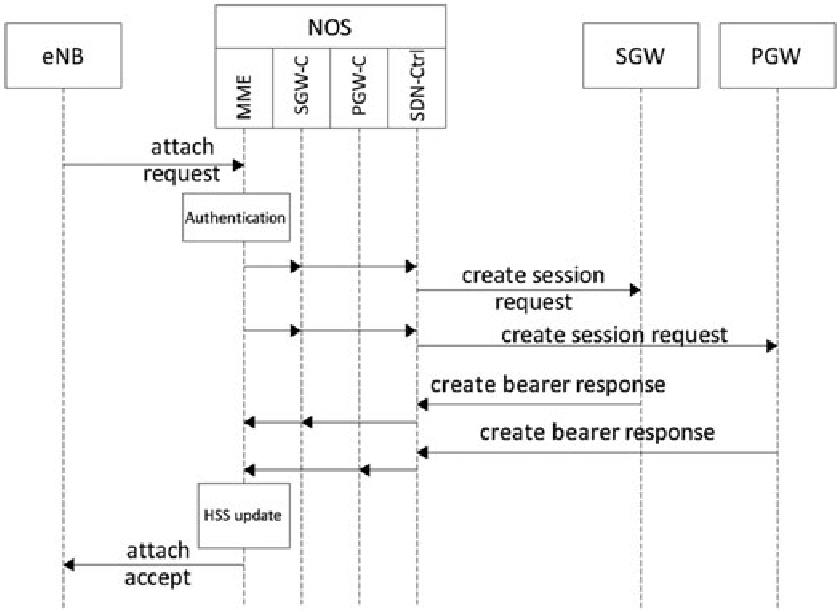

FIGURE 7 Modified registration operation for SDN-EPC

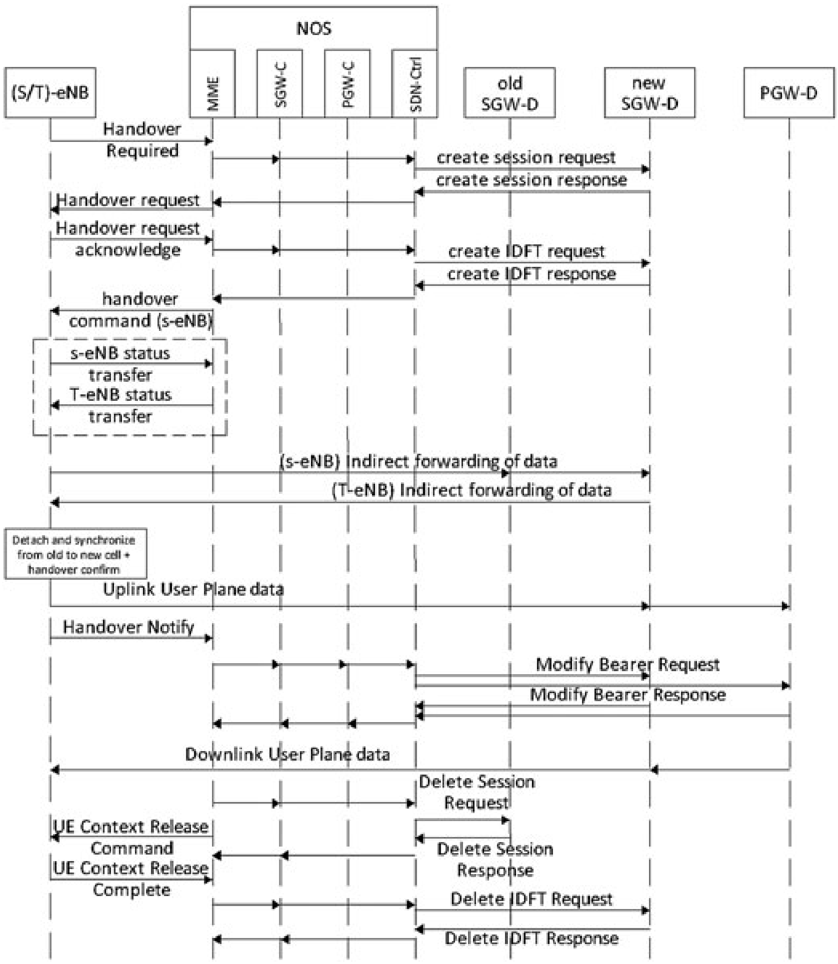

FIGURE 8 Modified S1-based handover operation for SDN-EPC

control packets of the GTP-C interface for the communications between the MME and the gateways. The first option is the UDP protocol, and this choice would be in accordance with existing gateway's transport stack and does not add any overhead in terms of processing resources when compared to the conventional architecture. Despite the limitations pointed out by Wallace and Shami,[23] the SCTP protocol represents the second logical choice. This choice fits well the primary motive for an SDN-based EPC of promoting distributed deployment of EPC nodes. With distributed deployment of nodes, the probability of lost packets increases, and thus, we consider the SCTP protocol as the encapsulation layer for control packets to introduce reliability in packet delivery at the transport layer. Another advantage for an SCTP-based communication is the flow control capabilities of SCTP. For example, in case of signaling storms, the SCTP can use throttling techniques to control the requests arrival rate.[24] The SCTP stack is not used in the SGW and the PGW in the conventional architecture, which means that additional firmware additions are required at these nodes. Moreover, there is an overhead caused by acknowledgement packets and packet headers introduced by the SCTP that are typically larger than those of UDP packets.

4.2.3 | Protocol layers required for functionality

Being an overlay network, the EPC architecture operates exclusively on the GTP protocol stack. It should be noted that layers 2 and 3 of the GTP protocol stack are not included directly in the EPC's mobile network operation as users are isolated from these layers. Therefore, the proposed architecture is based on centralization in the overlay network only. By contrast, layers 2 and 3 can continue to operate as they normally do without inclusion in the SDN realm. The latter will not make the new architecture any less SDN-like nor will change the overall network functionality.

4.2.4 | Platform requirement

The relocation of the control operations from the SGW to the MME means less resource capacity at the SGW is required and more resource capacity is required at the MME. This effect may be desirable where now less resources can be deployed at the network edge, and software-based platforms may be appropriate for use instead of custom-made or hardware-based platforms. In addition, more resources are required at the central location where higher processing capacity is available and dynamically manageable through data centers architectures and horizontal scaling techniques. The data plane processing requirements in both the conventional and the SDN-based architectures are unaltered, as the gateways still need to deliver high throughput operations.

The new direction in processing platforms is seeking to optimize software for the COTS platforms to support dynamic programmability of platforms and higher packet processing capacity.[20] This directly supports the proliferation of SDN-based software controlled platforms. For example, Data Plane Development Kit and signal processing development kit are released by Intel targeting programmable network functions on COTS.[25,26] Thus, in the current state of platforms' resources, the choice of software-based platforms is available, but further platforms testing is required to meet carrier-grade gateway performance specifications.

5 | SIMULATION AND QUANTITATIVE ANALYSIS

A simulation tool is developed to evaluate the conventional EPC and the proposed SDN-based EPC, designated by CONV-EPC and SDN-EPC, respectively, and validate their respective performances. In addition to supporting both architectures, the tool allows the simulation of various configurations and considers numerous factors that are deemed influential. A comprehensive array of performance results is produced and used for comparing the performance of the proposed architecture against that of the conventional one. Although the output of this study may serve as an indicator to the real performance values as we try to utilize performance value from the field and conform to the standardized operations, it is the relative performance between the 2 architectures and the effect of various factors that is the focus of quantitative results.

5.1 | Simulation test bed

In this work, network simulation environment OMNeT++[27] is utilized to build the simulation tool. The rest of this subsection describes the main nodes included in the simulation tool, and the factors considered and their levels.

The simulation tool implements the MME node as a single-queue–multiserver model, whereas the EPC gateways are implemented using single-queue–single-server models. The processing time of the MME for each GTP message is assumed to be linearly proportional to the message size. The capacity of the MME node, denoted herein by MME-CAP, is the first and most important factor in this modeling exercise. The study by Brown[28] reports a minimum capacity of 5 K requests per second for a typical MME node in a 10 Gbps data center. In this study, we adopt 2 levels of capacity, namely, 30 Mbps and 60 Mbps, which translate to the approximate capacities of 6.5 K and 13 K requests per second.

The processing time for the EPC gateways, SGW and PGW, designated by gateway processing time (GWPT), is the second factor to consider, and it should be based on the performance figures for the OF switch platforms reported in the literature. The processing time for a request at OF switches can be as low as 10 μs for hardware-based platforms[29] and can be as high as 150 μs for software-based platforms.[30] The processing time of 75 μs is also used to represent the expected performance of an enhanced software based platform. Therefore, for our experiments, we adopt the 3 values, 10, 75, and 150 μs, for the EPC GWPT factor.

The implemented simulation tool supports 3 distinct control operations: the registration operation, the bearer modification operation, and the S1-handover mobility operation. These control operations were described in Section 4. The experiments are designed to evaluate the performance of the registration and S1-based handoff control operations, while the traffic resulting from the bearer modification operation is utilized as control-plane background traffic. Because the majority of control signaling in the MME is reported as service requests operations,[31] the bearer modification operation request rate is varied in the experiments to levels that amount to 20%, 50%, and 80% of the MME's maximum request service rate. On the other hand, the foreground request rates of the registration operation or the S1-handover operation are configured to amount to 10% of the MME's capacity; this is considered appropriate taking into consideration the low frequency of registration and S1-handover requests relative to overall MME signaling. Therefore, the MME's control load factor, designated by MME-CL, is assumed to take on the levels of 30%, 60%, and 90% of the overall MME capacity. The GTP messages used in both registration and S1-handover mobility operations are utilized to implement the required signaling for the SDN-based architecture as explained previously.

5.2 | Network set up

The network set up shown in Figure 9 is considered for simulation tests where it contains a core center and a local center. This network topology allows the consideration of various strategies of network deployment adopted in the industry. The interested reader is referred to the NEC Corporation[32] and more specifically to Figure 7 of Simha[33] for a review of modern deployments adopted in the field. The core location is considered as the de facto location for mobile operator services such as voice-over IP, short messaging service, and multimedia messaging service, in addition to housing the MME node itself, a practice that is widely adopted for network deployment. On the other hand, the local location is considered as a local offload data center, which serves as an Internet breakpoint, ie, access point, closer to users' location. SGW and PGW are deployed in both core and local locations, as shown in Figure 9. This design leads to different scenarios of EPC mobile control procedures with respect to the specific gateways involved in the control operations, as shown in Tables 1 and 2. The different scenarios are intended to investigate the effect of gateways location distribution on performance metrics of the EPC network, assuming the control-plane background traffic operations are distributed evenly between core and local centers to simulate distributed services destinations.

All data links in the setup are configured to 10 Gbps, and propagation delays of backhaul links "R1-R2" and "R2-R3" are configured to values of 0.1, 0.5, and 1.0 ms corresponding to distances of 20, 100, and 200 km from the radio access network to the core center. The rest of the links are configured only to 10 μs; these configurations reflect appropriate different distances of gateways locations from core center and in turn should influence the overall performance. The propagation delay, designated by DEL, represents the fourth factor considered in this evaluation.

Based on physical interaction of the experiment factors, it is observed that the DEL factor is tightly coupled with EPC

TABLE 1 Registration procedure scenarios configurations

Case name	PGW location	SGW location
Reg-1	Core	Core
Reg-2	Core	Local
Reg-3	Local	Core
Reg-4	Local	Local

TABLE 2 S1-based handover procedure scenarios configurations

Case name	PGW location	Old SGW location	New SGW location
MOB-A	Local	Local	Core
MOB-B	Local	Core	Local
MOB-C	Core	Local	Core
MOB-D	Core	Core	Local

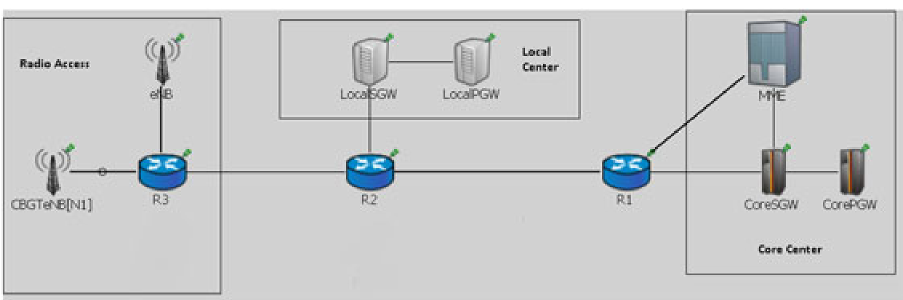

FIGURE 9 Simulated network setup

gateways location, whereas the GWPT and the MME control load are tightly coupled together. This leads to formulation of 2 groups of experiments combinations. The first is designed to quantify relation between simulation scenarios and the DEL levels, while the second is designed to quantify effect of MME control load (MME CL) and GWPT. Table 3 summarizes the factors, their levels, and the considered scenarios for each of the 2 experiments groups. It should be noted that for each experiment, the 2 EPC architectures, ie, CONV-EPC and SDN-EPC, are considered, and for each the 2 transport protocols, UDP and SCTP are tested. This results in 96 different experiment setups for experiments in the first group and 144 different experiment setups for experiments in the second group. In addition, 10 simulation runs are executed for each of the experiment setup and the 90% confidence intervals for the computed means are also generated.[34]

Finally, the reported results in this work focus on the following performance metrics: (1) the end-to-end delay of control operation (E2ED), (2) the MME resource utilization, and (3) the SGW resource utilization. The end-to-end delay metric represents the duration of time it takes to complete all the transactions between the network entities needed to execute the corresponding control operation, whereas resource utilization refers to the fraction of time the resource is busy in processing the messages corresponding to the respective control operation.

5.3 | Simulation results

5.3.1 | Effect of transport layer on E2ED results

The experiments simulated considered the 2 options for transport layer: the UDP and the SCTP protocols. The difference in the E2ED of control operations between the SCTP-based and the UPD-based SDN-EPC is quantified. The obtained results indicate that there is no perceivable effect for the transport protocol for the same simulation configuration, and the 2 protocols have almost the same E2ED in the same scenario. This finding is explained by the abundant link capacity, ie, 10 Gbps, and the small packets exchanged for control communication. In addition, the acknowledgment packets, for the case of SCTP, do not have substantial contribution to E2ED compared to overall E2ED contributed by other factors such as DEL. It should be noted that simulation model does not take into account the protocol stacks of UDP and SCTP (ie, below the overlay network), which, if considered, may affect the E2ED results. However, this part is left for future work.

TABLE 3 Factors configuration in experiment sets

Factors	Experiments—group 1	Experiments—group 2
Scenarios	All (4 scenario)	REG-1 and MOB-C
DEL (ms)	DEL-1, DEL-2, and DEL-30.1, 0.5, and 1.0	DEL-2 = 0.5
MME CL (%)	MME-CL2 = 60	MME-CL1, MME-CL2, and MME-CL3 30, 60, and 90
MMECAP (Mbps)	MMECAP-1 = 30	MMECAP-1 and MMECAP2 30 and 60
GWPT (μs)	GWPT-1 = 10	GWPT-1, GWPT-2, and GWPT-3 10, 75, and 150

5.3.2 | Group 1 experiment results

Table 4 and Table 5 show the E2ED results for group 1 experiments for the S1-handover and registration operations, respectively. The tables show the E2ED results in milliseconds for SDN-EPC in the upper half of the respective table and the relative E2ED of SDN-EPC compared to CONV-EPC at the same simulation scenario in the lower half of the table. In addition to the absolute SDN-EPC E2ED figures in the upper half, the table also shows these figures normalized to the highest E2ED figures. Therefore, the former relative E2ED figures serve to provide comparison between the EPC network type (ie, SDN vs CONV), whereas the latter relative E2ED figures for the SDN-EPC serve to provide comparison between the different mobility scenarios considered in these experiments.

Comparison of SDN-EPC to CONV-EPC: MOB-A, B, C, REG-1, 3, and 4 are showing slight increase of E2ED in SDN-EPC compared to CONV-EPC. The increase is restricted to 1%–2% and 1%–4% for S1-handover and registration procedures, respectively. For MOB-D and REG-2, the E2ED figures are showing a decrease of 2%–6% and 6%–20%, respectively, with greater decreases for increased DEL levels. The reason for this decrease is attributed to the elimination of trombone path in control operation between MME and PGW, as explained previously. For the SDN-based EPC, the MME communicates directly with PGW, whereas a pipeline path exists for the case of CONV-EPC in the form of MME-SGW-PGW. This leads to a decrease in E2ED that is most evident when the distance between the MME and PGW is substantial.

The fact that only MOB-D and REG-2 have perceivable change in E2ED indicates that effect of SDN-EPC is coupled with the location of EPC gateways involved in the control operation. In particular, the farther the distance between gateways, ie, MME-SGW-PGW route, the more SDN-EPC have perceivable reduction in E2ED figures

TABLE 4 E2ED results in group 1 experiment sets for S1-handover operation

	MOB-A		MOB-B		MOB-C		MOB-D	
	ms	MOB-A/MOB-D (%)	ms	MOB-B/MOB-D(%)	ms	MOB-C/MOB-D(%)	ms	MOB-D/MOB-D(%)
DEL-1	6.76	91	7.4	100	6.6	100	7.4	100
DEL-2	14	82	17	100	13	76.5	17	100
DEL-3	23	79	29	100	21	72.4	29	100
	SDN-EPC/CONV-EPC (%)							
DEL-1	101		102		102		98	
DEL-2	101		101		101		95	
DEL-3	100		100		101		94	

TABLE 5 E2ED results in group 1 experiment sets for registration operation

	REG-1		REG-2		REG-3		REG-4	
	ms	REG-1/REG-2(%)	ms	REG-2/REG-2(%)	ms	REG-3/REG-2(%)	ms	REG-4/REG-22(%)
DEL-1	2.1	91	2.3	100	2.3	100	2.3	100
DEL-2	3.7	79	4.7	100	4.7	100	4.7	100
DEL-3	5.7	74	7.7	100	7.7	100	7.7	100
	SDN-EPC/CONV-EPC (%)							
DEL-1	104		94		103		103	
DEL-2	102		83		101		102	
DEL-3	101		80		101		101	

relative to CONV-EPC. This is the case for the considered simulation scenarios when local SGW and core PGW are involved.

Effect of DEL on E2ED: increasing DEL levels leads to a proportional increase in E2ED as expected regardless of simulation scenario; however, the relative percentage of increase is slightly varying according to simulation scenarios, which dictate the location of gateways. Comparing the performance for DEL-2 relative to DEL-1, we can observe relative increases ranging from $17/6.6-1$ or 97% to $17/7.4-1$ or 130% for the S1-handover operation. For the registration operation, the relative increases ranges from $3.7/2.1-1$ or 76% to $4.7/2.3-1$ or 104%. Similarly, the increase from DEL-2 to DEL-3 increases E2ED by approximately 62%–71% and 54%–64% for the S1-handover and registration operations, respectively. The relative increase of E2ED is observed to be higher when the route of message exchange includes more traversal of backhaul links, and therefore the highest increase is found in MOB-B, MOB-D, REG-2, REG-3, and REG-4, ie, the scenarios that include local SGW in S1-handover and local SGW or PGW in the registration operation. It is observed that the difference between these scenarios and the rest of the scenarios increases with increasing DEL. This is attributed to increased contribution of DEL in overall E2ED.

5.3.3 | Group 2 experiment results

Table 6 and Table 7 show the E2ED results in group 2 experiment sets for the S1-handover and registration operations, respectively. Again, the tables show E2ED figures for SDN-EPC in the upper half of the table, whereas the lower half lists the relative E2ED of SDN-EPC compared to CONV-EPC at the same simulation scenario.

Comparison of SDN-EPC and CONV-EPC: At MMECAP-1 for configurations of MME-CL1 or MME-CL2 coupled with GWPT-1 or GWPT-2, SDN-EPC is showing approximately similar E2ED performance to CONV-EPC with differences not exceeding 1% for the S1-handover operation and less than 2.5% for the registration operation. Still at MMECAP-1 but for GWPT-3 coupled with MME-CL1 or MME-CL2, the E2ED is approximately 2.5%–4.3% and 6.6%–8.6% less than CONV-EPC for the respective control operations. Henceforth, with low and average control load level and high gateway capacity, SDN-EPC and CONV-EPC have very similar E2ED performance, whereas at GWPT-3, which is the case of low gateway capacity, the SDN-EPC provides E2ED enhancement, up to 8.6%, compared to CONV-EPC.

At MMECAP-1 and MME-CL3 coupled with GWPT-1 or GWPT-2, SDN-EPC is observed to provide higher E2ED

TABLE 6 E2ED results in group 2 experiment sets for S1-handover operation

	MMECAP-1			MMECAP-2		
MOB-C	**MME-CL1 ms**	**MME-CL2 ms**	**MME-CL3 ms**	**MME-CL1 ms**	**MME-CL2 ms**	**MME-CL3 ms**
GWPT-1	12.9	13.0	15.8	10.5	10.5	12.5
GWPT-2	13.1	13.2	16.1	10.7	10.8	12.7
GWPT-3	13.4	13.6	16.3	11.1	11.5	19.6
SDN-EPC/CONV-EPC (%)						
GWPT-1	100	100	117	100	100	114
GWPT-2	99.0	99.5	115	98.5	97.6	100
GWPT-3	97.5	95.7	97.2	94.8	2.7*	1.5*

*E2ED for CONV-EPC very large (unstable).

TABLE 7 E2ED results in group 2 experiment sets for registration operation

	MMECAP-1			MMECAP-2		
MOB-C	**MME-CL1 ms**	**MME-CL2 ms**	**MME-CL3 ms**	**MME-CL1 ms**	**MME-CL2 ms**	**MME-CL3 ms**
GWPT-1	3.64	3.68	4.45	2.83	2.85	3.33
GWPT-2	3.71	3.74	4.52	2.9	2.92	3.39
GWPT-3	3.79	3.83	4.56	2.98	3.05	3.82
SDN-EPC/CONV-EPC (%)						
GWPT-1	101	102	119	100	101	115
GWPT-2	97.5	97.9	114	95.6	94.0	91.9
GWPT-3	93.4	91.4	91.5	88.7	1.9*	3.33*

*E2ED for CONV-EPC very large (unstable).

than CONV-EPC, whereas when coupled with GWPT-3, SDN-EPC provides less E2ED than CONV-EPC by 2.8% and 8.5% in S1-handover and registration operations, respectively. This is because for the same number of control operations processed by the MME and EPC gateways, the SDN-EPC consumes more processing resources than CONV-EPC in the MME whereas CONV-EPC consumes more resources at the SGW. This leads to increased processing time at the MME in SDN-EPC and increased processing time at the SGW for low gateway capacity, ie, GWCAP-3, in CONV-EPC. However, the increase in E2ED at GWCAP-3 due to SGW processing in CONV-EPC is higher than the increase in E2ED due to MME processing in SDN-EPC; therefore, the SDN-EPC performs better than the CONV-EPC when processing time of the EPC gateway is high even when the MME control load is at its highest level.

Comparing the SDN-EPC E2ED figures for MMECAP-2 relative to those with MMECAP-1 for the same configuration, reductions of 15%–23% and 16%–29% are observed for S1-handover and registration operations, respectively, with the except for the case of S1-handover operation at MME-CL3 and GWPT-3 where E2ED actually increased by 20%. This decrease is attributed to the increase in the processing speed of MME resources. It should be noted that although the MME capacity have doubled, the decrease in E2ED is minor since the portion of processing time in the MME is small compared to effect of the DEL factor. The delay factor DEL remains the main contributor to the E2ED figure as indicated results of group 1 experiments, and the contribution of MME processing has limited effect on the overall E2ED.

It is also noted that at when comparing SDN-EPC to CONV-EPC at MMECAP-2, E2ED is suddenly much lower in SDN-EPC than CONV-EPC at MME-CL2 and MME-CL3 for GWPT-3. Simulation results show that SDN-ECP E2ED are at most 2.7% of CONV-EPC E2ED. Further inspection into the obtained results and setup indicates that CONV-EPC have large degradation in E2ED when SGW is configured with large processing time, ie, set at GWPT-3, while receiving average to high load, ie, MME-CL2 and MME-CL3. As previously noted, the CONV-EPC requires more resources than SDN-EPC at SGW, and at MMECAP-2, 60% and 90% of MME system capacity, ie, MME-CL2 and MME-CL3, provides offered load that is larger than the SGW system processing capability. Therefore, congestion at SGW occurs for the CONV-EPC case leading to severe degradation in E2ED. In comparison, because the SDN-EPC configuration requires less resource at SGW, it has not encountered this degradation as the MME control load increases. These results show that software-based platforms may be adopted for the gateways for the case of SDN-EPC more than that for CONV-EPC and, thus, promotes the use of cheaper platforms in the EPC network.

5.3.4 | MME resource utilization

The MME control load points defined in our study have been represented as a percentage of the MME capacity. Typically, the control load is proportional to requests arrival rate and is estimated to its respective MME resource utilization. Both CONV-EPC and SDN-EPC will serve the exact same number of registration or handoff requests at each control operations settings but the MME resource utilization is different between them due to different message sizes and exchange paths between the two architectures.

The actual MME resource utilization has been measured for both architectures as depicted in Figure 10. It shows that SDN-EPC MME utilization is 29.8/24.9 − 1 or 20% to 89.5/77 − 1 or 16% higher relative to CONV-EPC utilization, this increase translates to a range of 5% to 13% of actual MME capacity across control load points, ie, MME-CL1 to MME-CL3. Therefore relocating control operations to the MME in SDN-EPC leads to an average of 18% increase in MME resource utilization depending on control load levels in the network. However, this increase is exchanged for less resource utilization in SGW compared to that for the CONV-EPC network as mentioned earlier. This evaluation hints that the anticipated increase in MME utilization is moderate and still allows for software-based implementations, as will be discussed in the conclusions section.

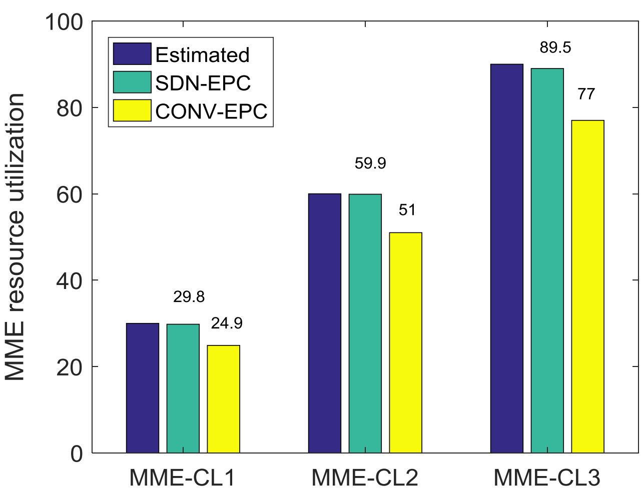

FIGURE 10 MME resource utilization versus MME-CL factor

5.3.5 | SGW resource utilization

The SGW resource utilization has been measured for both architectures and summarized in Table 8. It is clear that resource utilization is increasing linearly with the control plane load MME-CL as expected. Furthermore, it may be noted that resource utilization at MMECAP-2 is twice that at MMECAP-1 for any EPC network type due to the doubling requests rate. It is observed that the utilization of SGW gateway in SDN-EPC is approximately half the utilization in CONV-EPC, regardless of GWPT. This is due to control operations being shifted from SGW into the MME in SDN-EPC.

Considering the cases for MMECAP-2, the utilization for GWPT-1 did not exceed 10% in most configurations, whereas for GWPT-2, it reaches 88% and 44% at MME-CL3 for CONV-EPC and SDN-EPC, respectively. However, for GWPT-3, resources are 100% fully utilized at highest load for CONV-EPC and are at 90% for SDN-EPC. Again, it is asserted that CONV-EPC is more resource consuming at the gateways compared to SDN-EPC. For example, at MME-CL3, the utilization increases from 12% to 88% and to 100% as the GWPT increases. This effect of larger resources requirement at the gateways for CONV-EPC leads to degradation in E2ED performance at very high control load as seen previously.

These results are consistent with the qualitative analysis that SDN-EPC reduces utilization of EPC gateway resources, and thus SDN-EPC enables the usage of software-based

TABLE 8 SGW resource utilization

		MMECAP-1		MMECAP-2	
GWPT	MME-CL	CONV-EPC	SDN-EPC	CONV-EPC	SDN-EPC
GWPT-1	MME-CL 1	2	1	4	2
(10 μs)	MME-CL 2	4	2	8	4
	MME-CL 3	6	3	12	6
GWPT-2	MME-CL 1	16	8	32	16
(75 μs)	MME-CL 2	30	15	60	30
	MME-CL 3	44	22	88	44
GWPT-3	MME-CL 1	31	16	62	32
(150 μs)	MME-CL 2	60	30	100	60
	MME-CL 3	88	45	100	90

gateways even at high loads. This is in contrast with CONV-EPC where software-based platforms are only appropriate for low or may be average control plane loads. The rate of increase in gateway resource utilization indicates that further increase in requests rate would saturate the resources, which stresses the importance of careful engineering design and network dimensioning.

6 | CONCLUSION

The SDN technique is not entirely new in the network literature. However, the application of its concept to the EPC network has gained momentum recently. The advantages of an SDN-based EPC should outweigh its disadvantages in order for manufacturers and operators to adopt this change in the control plane architecture. An SDN-based EPC, in our view, is best suited to be an overlay architecture rather than controlling the lower layers as it is the case in OpenFlow architecture.

This study proposes an SDN-based EPC architecture along with the protocol-level details required for communicating between SDN entities while maintaining the correct functionality with minimum impact on the conventional EPC network. The study provided a comprehensive quantitative and comparative analysis of the 2 networks: SDN-based EPC versus conventional EPC by considering several factors that influence performance and adopting factor levels from the field. The evaluation focuses on 2 control-plane operations, namely, the S1-handover operation and the registration operation, and evaluating performance metrics such as E2ED and nodes utilization.

The quantitative evaluation has provided evidence that performance of the SDN-based EPC compared to the conventional one is dependent on multiple factors, primarily gateway locations. When the path between MME-SGW-PGW route is maximal, SDN-EPC has shown decrease in control operations overall delay, whereas in the rest of the cases considered a slight increase in the overall delay is observed in SDN-based EPC. It is observed that in the case of the overall delay reduction, the reduction increases with increasing backhaul propagation delay, which promises more enhancements for distributed gateways.

In terms of resource utilization, the SDN-based EPC have shown less resource demand in SGW which promotes the use of software-based gateways in the EPC network, whereas more resource demand in the MME is required. However, the MME is located in core centers which typically harbors abundant machine resources with the availability to scale based on resource demands. Therefore, the increase in the MME resource is considered acceptable for a mobile network operator.

ACKNOWLEDGMENT

The authors would like to acknowledge the support provided by King Fahd University of Petroleum and Minerals (KFUPM) and the Deanship of Scientific Research (DSR).

REFERENCES

1. "General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access," 3GPP, 2014.
2. Lara A, Kolasani A, Ramamurthy B. Network innovation using OpenFlow: a survey. *Comm Surv Tutor*. 2013;16(1):493-512.
3. Nunes B, Mendonca M, Nguyen X-N, Obraczka K, Turletti T. A survey of software-defined networking: past, present, and future of programmable networks. *Comm Surv Tutor*. 2014;16(3):1617-1634.
4. J. Kempf, B. Johansson, S. Pettersson, H. Lüning and T. Nilsson, "Moving the mobile evolved packet core to the cloud," in Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference, Barcelona, 2012.
5. S. Ben Hadj Said, M. Sama, K. Guillouard, L. Suciu, G. Simon, X. Lagrange and J. Bonnin, "New control plane in 3GPP LTE/EPC architecture for on-demand connectivity service," in 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet), San Francisco, 2013.
6. Pentikousis K, Wang Y, Hu W. Mobileflow: toward software-defined mobile networks. *Comm Mag*. 2013;51(7):44-53.
7. Trivisonno R, Guerzoni R, Vaishnavi I, Soldani D. SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility. *Trans Emerging Telecom Technol*. 2015;26(1):82-92.
8. "Technical specification group services and system aspect—study on architecture for next generation system," 3GPP TR 23.799-V14.0.0, 2016.
9. B. Naudts, Kind, M., F. J. Westphal, S. Verbrugge, D. Colle and M. Pickavet, "Techno-economic analysis of software defined networking as architecture for the virtualization of a mobile network," in Software Defined Networking (EWSDN), 2012 European Workshop, 2012.
10. Open Networking Foundation, "Software-defined networking: the new norm for networks," 2012.
11. J. Andersson, E. Bizouarn, M. Boc, A. Petrescu, Ç. Edemen, S. Ergut, Z. Faigl, J.-L. Lafragette, C. Larsson, R. Purushothama, A. S. Tan, E. Zeydan, L. Bokor, K. Daoud and W. Hahn, "Architectural EPC extensions for supporting heterogeneous mobility schemes," Mobile Networks Evolution for Individual Communications Experience (MEVICO CELTIC/CP7-011) project, 2013.
12. B. Ahmad, T. Taleb, A. Vajda and M. Bagaa, "Dynamic cloud resource scheduling in virtualized 5G mobile systems," in 2016 IEEE Globecom (GLOBECOM), Washington, USA, 2016.
13. Taleb T, Mada B, Corici M, Nakao A, Flinck AH. PERMIT: network slicing for personalized 5G mobile telecommunications. *IEEE Comm Mag*. 2017;55(5):88-93.
14. Taleb T, Dutta S, Ksentini A, Muddesar I, Flinck H. Mobile edge computing potential in making cities smarter. *IEEE Comm Mag*. 2017;55(3):38-43.

15. Taleb T, Ksentini A, Jantti R. Anything as a service for 5G mobile Systems. *IEEE Net Mag*. 2016;30(6):84-91.
16. Open Networking Foundation, "OpenFlow switch specification," 2011.
17. P. Gurusanthosh, A. Rostami and R. Manivasakan, "SDMA: a semi-distributed mobility anchoring in LTE networks," in Mobile and Wireless Networking (MoWNeT), 2013 International Conference on Selected Topics in, Montreal, QC, 2013.
18. A. Basta, W. Kellerer, M. Hoffmann, K. Hoffmann and E. D. Schmidt, "A virtual SDN-enabled LTE EPC architecture: a case study for S-/P-gateways functions," in Future Networks and Services (SDN4FNS), Trento, 2013.
19. Liu B. *Software Defined Networking and Tunneling for Mobile Networks*. Stockholm: KTH Royal Institute of Technology; 2012.
20. Greenhalgh A, Huici F, Hoerdt M, Papadimitriou P, Handley M, Mathy L. Flow processing and the rise of commodity network hardware. *ACM SIGCOMM Comp Comm Rev.* April 2009;39(2):20-26.
21. J. Fong, X. Wang, Y. Qi, J. Li and W. Jiang, "ParaSplit: a scalable architecture on FPGA for terabit packet classification," in High-Performance Interconnects (HOTI), 2012 IEEE 20th Annual Symposium, Santa Clara, CA, 2012.
22. A. A. Naser, "Performance evaluation of a software defined network based architecture for a 4G EPC network," King Fahd University of Petroleum and Minerals, Dhahran, 2015.
23. Wallace TD, Shami A. A review of multihoming issues using the stream control transmission protocol. *IEEE Comm Surv Tutor*. 2012;14(2):565-578.
24. R. Mishra and S. Papa, "Signaling storm reduction from radio networks". United States of America Patent US20160029246 A1, 28 January 2016.
25. Intel, "Data-plane development kit," [Online]. Available: http://http://dpdk.org.
26. G. Pongracz, L. Molnar and Z. Kis, "Removing roadblocks from SDN: OpenFlow software switch performance on Intel DPDK," in Software Defined Networks (EWSDN), 2013 Second European Workshop, Berlin, 2013.
27. ZoltanBojthe, "INET framework for the OMNeT++ discrete event simulator," 2012. [Online]. Available: https://github.com/inet-framework/inet.
28. G. Brown, "LTE/SAE & the evolved packet core: technology platforms & implementation choices," Alcatel-Lucent, 2009.
29. M. Jarschel, S. Oechsner, D. Schlosser, R. Pries, S. Goll and P. Tran-Gia, "Modeling and performance evaluation of an OpenFlow architecture," in Teletraffic Congress (ITC), 2011 23rd International, San Francisco, CA, 2011.
30. A. Basta, W. Kellerer, M. Hoffmann, H. J. Morper and K. Hoffmann, "Applying NFV and SDN to LTE mobile core gateways, the functions placement problem," in the 4th workshop on All things cellular: operations, applications, & challenges, New York, NY, 2014.
31. Alcatel-Lucent, "The impact of small cells on MME signaling, methods to reduce and optimize MME core signaling caused by small cells," Alcatel-Lucent, 2013.
32. NEC Corporation, "NEC virtualized evolved packet core—vEPC; design concepts and benefits," 2015.
33. Simha A. *NFV Reference Architecture for Deployment of Mobile Networks*. Raleigh: Red Hat Reference Architecture Series; 2017.
34. Jain R. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. New York: Wiley; 1991.

How to cite this article: Mahmoud A, Abo Naser A, Abu-Amara M, Sheltami T, Nasser N. Software-defined networking approach for enhanced evolved packet core network. *Int J Commun Syst*. 2017;e3379. https://doi.org/10.1002/dac.3379