PERFORMANCE EVALUATION OF A SOFTWARE DEFINED NETWORK BASED ARCHITECTURE FOR A 4G EPC NETWORK

BY

AHMAD A. ABO NASER

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

COMPUTER ENGINEERING

APRIL 2015

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Ahmad Azmi Abd Alfatah Abo Naser** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER ENGINEERING.**

Thesis Committee

Dr. Ashraf Mahmoud
(Advisor)

Dr. Tarek Sheltami
(Member)

Dr. Marwan Abu-Amara
(Member)

Dr. Ahmad Al Mulhem
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

8/6/15

Date

ACKNOWLEDGMENTS

All praise and thanks are due to Almighty Allah. He bestowed upon me health, knowledge and patience to complete this work.

I will take this opportunity to acknowledge, with deep gratitude and appreciation, the encouragement, valuable time and continuous guidance given to me by my thesis advisor, Dr. Ashraf Mahmoud. I am also highly grateful to my thesis committee members: Dr. Tarek Sheltami and Dr. Marwan Abu-Amara for their valuable guidance, suggestions and motivation. Thanks to all KFUPM, lab's technicians, engineers for their help in using university's facilities related the study specially computer engineer department. Acknowledgement is due to KFUPM for the support extended towards my research and for granting me the opportunity to pursue graduate studies. My heartfelt thanks are due to my Family (mother, sisters and brothers) for their prayers, guidance, and moral support throughout my academic life. Finally, special thanks are due to my senior colleagues at KFUPM for their help and prayers.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATIONS

3GPP	Third-Generation Partnership Project
AP	Application Protocol
APN	access point name
EPC	Evolved Packet Core
EPS	Evolved Packet System
eNodeB	Evolved NodeB
E-UTRAN	Evolved UTRAN
GPRS	general packet radio service
GTP	GPRS Tunneling Protocol
GTP-U	GPRS Tunneling Protocol-User plane
HSS	Home Subscriber Server
IMS	IP Multimedia Subsystem
IP	Internet Protocol
LTE	Long Term Evolution
MME	Mobility Management Entity
NAS	Non Access Stratum
PCC	Policy Control and Charging
PCEF	Policy Control Enforcement Function
PCRF	Policy Control and Charging Rules Function
PDN	packet data network
PDU	protocol data unit
PGW	PDN Gateway
PMIP	Proxy Mobile Internet Protocol
QCI	QoS class identifier
QoS	quality of service
RAN	radio access network
RAT	Radio Access Technology
SAE	System Architecture Evolution

SCTP	Stream Control Transmission Protocol
SGW	Serving Gateway
TCP	Transmission Control Protocol
TEID	Tunneling End ID
TFT	Traffic Flow Template
TNL	Transport Network Layer
UE	user equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP

ABSTRACT

Full Name Ahmad Azmi Abd-Alfatah Abo Naser

Thesis Title Performance Evaluation of software defined network approach of EPC network

Major Field Computer Engineering

Date of degree April.2015

The evolved packet core (EPC) network is a mobile network standardized by the 3GPP. EPC is the recent evolution of mobile networks providing high speed data rates and on-demand connectivity services. However, new business models and service criteria in Telecom industry pushes the need to investigate new technologies and architectures. Software defined networking (SDN) is recognized as one of the new generation network technologies where the principle concept is to separate the control plane from the data plane, and concentrate the control plane functionality of network devices in a logically centralized controller. This control plane architecture introduces new perspectives into network design. Appling the SDN concept in the evolved packet core (EPC) network has been investigated by some researchers with emphasis on its qualitative advantages to EPC network. But very few studies have tackled the quantitative analysis and evaluation of an SDN-based EPC network. Furthermore, the protocol-level details of applying SDN concept to an EPC network has not been specified thoroughly before.

This motivated us to study EPC network architecture and its functionality, the concept of SDN and its benefits, and the application of SDN concept on EPC network; all for the purpose of reaching a mature understanding of an SDN-based EPC network. In the light of this motivation, in this study, we present the following: (1) A brief

background of EPC and software defined networking (SDN) architectures. (2) The potential role of SDN in mobile network operator, particularly in the EPC architecture. (3) Detailed qualitative analysis and comparison of EPC and SDN architectures taking into consideration the OpenFlow architecture which is viewed as the most mature SDN standard. (4) A proposal for a novel approach of an SDN-EPC. (5) A quantitative study of control operations in conventional EPC and proposed SDN-EPC, where the following are addressed: a) factors controlling operations performance, b) simulation-based testbed for performance evaluation, and c) an engineered investigation for measuring the effect of each factor on performance metrics and their interactions with other factors. Finally, (6) we provide perspective to the findings that serve in the big picture of an SDN-EPC architecture.

The control operations addressed are registration procedure and S1-based handover mobility procedure. Both procedures are described according to 3GPP standards. In addition, SDN-derivatives are presented taking into consideration the proper functionality of the control procedure.

The study evaluates control operations metrics; mainly end to end delay of operation, resource utilization of network devices, and bandwidth utilization. The assessment of the metrics will provide insight to their dependence on factor levels which can serve as means for performance prediction of the EPC network and guidelines for proper engineering design of EPC network.

The results show the detailed end to end delay of control operations of SDN-EPC compared to conventional EPC (CONV-EPC). It finds that SDN-EPC produces less delay than CONV-EPC when serving gateway (SGW) is located in local center, the anchor packet gateway (PGW) is located in core center, and the mobility management entity (MME) is under low to average resources utilization. The reductions are observed to be 1-7% and 6-23% in S1-handover and registration procedures, respectively, in comparison to CONV-EPC. It finds that SDN-EPC produces worse end to end delay when the MME is under high resource utilization regardless of EPC gateways location. SDN-EPC requires 10-22% more processing resources at the MME due to centralization of control operations, whereas it requires 35-50% less resources at the SGW for the same previous reason. Finally, the increase in bandwidth utilization due to SDN-EPC is found negligible.

ملخص الرسالة

الاسم الكامل: أحمد عزمي عبد الفتاح أبو ناصر

عنوان الرسالة: تقييم أداء شبكة الحزم المطورة القائمة على مبدأ الشبكات المبرمجة

التخصص: هندسة الحاسب الألي

تاريخ الدرجة العلمية: أبريل 2015

شبكة الحزم المتطورة (EPC) هي شبكة للهاتف المحمول موحدة المعايير من قبل مؤسسة 3GPP. شبكة الحزم المتطورة (EPC) هي التطور الأخير في عالم شبكات المحمول التي تتميز بتوفير معدلات نقل البيانات بسرعة عالية وخدمات الاتصال حسب الطلب. و مع ذلك, فان نماذج الأعمال الجديدة ومعايير الخدمة في مجال صناعة الاتصالات تدفع الحاجة للتحقيق في التقنيات والأبنية جديدة . الشبكات المبرمجة (SDN) يتم اعتبارها كواحدة من تقنيات الشبكات الجيل الجديد حيث يقوم مبدأها على مفهوم الفصل بين مستوى التحكم و مستوى نقل البيانات, و يركز وظائف مستوى التحكم في أجهزة الشبكة في كيان مركزي. هذه البنية لمستوى التحكم تقدم آفاقا جديدة في تصميم الشبكات. تطبيق مفهوم الSDN في شبكة الحزم المتطورة (SDN) تم التحقيق فيه من قبل بعض الباحثين، مع التركيز على مزاياه نوعية في شبكة الEPC. ولكن عدد قليل جدا من الدراسات قد تناولت التحليل الكمي وتقييم أداء شبكة الEPC المبنية على SDN. وعلاوة على ذلك، لم يتم تحديد مستوى التفاصيل في البروتوكولات عند تطبيق مفهوم SDN على شبكة EPC جيدا من قبل.

هذا حفّزنا لدراسة بنية شبكة EPC و وظائفها، ومفهوم SDN وفوائدها وتطبيق مفهوم SDNعلى شبكة EPCو ذلك من أجل الوصول فهم ناضج عن شبكة الEPC المبنية على SDN.في ضوء هذه الحافز, في هذه الدراسة، قمنا بعرض التالي: (1) خلفية ملخصة عن هيكلية الEPC و الSDN, (2) دورها المحتمل في مشغلي شبكات الهاتف المحمول، لا سيما في شبكة الEPC؛ (3) تحليل نوعي مفصل و مقارنة ما بين هيكلية الEPC و الSDN؛ (4) رؤيتنا الفريدة من نوعها لشبكة EPC المبنية على الSDN؛ (5) دراسة كمية عن شبكة ال EPCالتقليدية و الEPCالمبنية على الSDN المقترحة, حيث نقوم بأخذ اعتبار

التالي: أ) العوامل التي تأثر على الأداء, ب) بيئة المحاكاة المستخدمة لدراسة الأداء, ت) تحقيق هندسي في كيفية قياس تأثير كل عامل على الأداء و التفاعلات ما بين العوامل. و أخيرا, (6) نقدم منظورنا على النتائج التي توعي بالصورة الكبرى شبكة ال EPC المبينة على الSDN.

نقدم حالتين من عمليات مستوى التحكم، وخاصة إجراءات التسجيل (Registration) و اجراء S1-Based Handover للتنقل. يتم شرح الاجراءين و تطوير مشتقات من كليهما مبنيان على الSDN. نقوم بالاشارة إلى التعديلات المطلوبة في البروتوكولات والعمليات الأساسية في أجهزة شبكة الEPC لصياغة نظام EPC مبني على SDN فعال.

الدراسة تقيم مقاييس أداء عمليات التحكم؛ بشكل رئيسي الزمن الكامل لاتمام العملية, واستخدام موارد المعالجة في أجهزة الشبكة, واستخدام نطاق البيانات. التقييم يُظهر اعتماد المقاييس على قيم العوامل المختلفة و الذي بدوره يقدم وسيلة لتوقع الأداء في شبكة الُEPC وتوجيهات للتصميم الهندسي السليم.

النتائج تظهر تفاصيل الزمن اللازم لاتمام عمليات التحكم في شبكة ال SDN-EPC مقارنة بشبكة الCONV-EPC. الSDN-EPC تتطلب زمنا أقل من الCONV-EPC عندما يكون الSGW متواجد في المركز المحلي, و الPGW متواجد في المركز المركزي, و تكون ال MME تحت استخدام للموارد منخفض أو متوسط. نسبة التقليل من زمن العملية يتراوح ما بين 1-7% و 6-23% في عمليتا ال-S1 handover و Registration على التوالي مقارنة بالCONV-EPC. النتائج أيضا تُظهر أن ال SDN-EPC تتطلب زمن أعلى من الCONV-EPC عندما تكون الMME تحت استخدام عالي بغض النظر عن أماكن تواجد الgateways. الSDN-EPC تتطلب 10-22% موارد معالجة أكثر من الCONV-EPC في الMME و ذلك بسبب التمركز في عمليات التحكم, بينما تتطلب 35-50% أقل في موارد المعالجة في الSGW لنفس السبب السابق. وأخيرا, تُظهر النتائج أن الزيادة في استخدام نطاق البيانات غير مهمة على الاطلاق.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION AND PROBLEM DESCRIPTION

The network of mobile operator is increasingly pushed to the limits with exponential increase of data flow across the network and new types of services being deployed that has strict latency constraints. Operators, ISPs and other parties are investigating new technologies to improve bandwidth utilization and latency requirement to deploy their services in an effective manner. However, before network migration to any conceptually new technology, technical challenges and performance metrics faced must be studied to prove the efficiency and features of the new proposals. Software defined networking (SDN) is recognized as one of the new generation network technologies[2][3]. SDN, in general, separates the control plane from the data plane and concentrate the control plane functionality of network devices in a logically centralized controller. This control plane architecture introduces new perspectives into network design.

The concept of SDN in the evolved packet core (EPC) network has been investigated by some researchers. In this work, we provide an insight into the EPC and OpenFlow (OF) architectures. We examine different aspects of the EPC network with regard to control plane, and compose a comprehensive comparison with OF standard. We show the benefits of an SDN-based EPC network and provide our unique realization for an SDN-based EPC. None of the studies in the literature, to the best of our knowledge, have included performance evaluation of a full EPC core network based on SDN implementation. In addition, no study thus far provides a quantitative comparison between SDN-based and conventional EPC architecture where major control operations are accounted for. Therefore, the main contributions of this work are: (1) a novel analysis and proposal of SDN-based EPC architecture in different variations while maintaining 3GPP compliance with EPC functionality; (2) A framework for performance evaluation of a carrier-grade mobile network that accounts for various factors affecting performance and network design. The study models EPC under three variations: (1) a conventional EPC architecture, (2) SDN-based EPC architecture with SCTP protocol as transport layer, and (3) SDN-based EPC architecture with UDP protocol as transport layer.

Furthermore, two cases of control plane operations are presented, particularly Initial attachment (registration) procedure and S1-based handover mobility procedure. The two procedures are explained and derivatives of the same procedure are developed in the SDN-based architecture. Modifications required in underlying protocol and message formats are identified out along with impact on EPC nodes to formulate a functional SDN-based EPC system.

A simulation based performance evaluation of the conventional EPC architecture and SDN-based EPC architecture in the control plane is designed to match closely real network conditions. A fractional factorial experiment design is depicted to determine the effect of each factor under investigation for their role in performance metrics variation. Factors examined in the study for performance evaluation are: data and control plane background traffic, distributed deployment of EPC gateways, processing capacity of main control entity and EPC gateways, and propagation delays of backhaul link. The experimental framework is suited to test any network architecture and provide metrics characteristics with their dependence on design and performance factors. The test framework will be used to demonstrate the performance evaluation of SDN-EPC compared to conventional EPC with regards to the following metrics: (a) end to end delay of control operation under test, (b) bandwidth utilization of communication links connecting main control entity (MME), (c) EPC network main control entity (MME) resource utilization, (d) EPC gateways resource utilization. The results in this study will show the relation between metrics performance and factor levels and provide perspective to expected outcome of the network architecture performance metrics under variable factor level, these results can be used as a benchmark for expected performance and design criteria of an SDN-based EPC network for future experimental prototypes.

CHAPTER 2

BACKGROUND

This chapter gives an introduction that discusses and explains the main concepts of mobile network architecture and software defined networking which aims to reasonably facilitate understanding of the study fundamentals. The mobile network architecture is addresses in the following topics: evolved packet core (EPC) network architecture and its elements functions, GPRS tunneling protocol and its role in EPC network, an overview of control plane procedures in the EPC, and EPC's gateway architecture. The software defined networking (SDN) is addressed in the following topics: the general concept, potential benefits of SDN for networks, and OpenFlow protocol as an example of SDN standard.

2.1 MOBILE NETWORK ARCHITECTURE

2.1.1 EVOLVED PACKET CORE NETWORK

The 4^{th} generation network architecture includes an Evolved Packet Core (EPC) as the core network and an Evolved UMTS Terrestrial Radio access network (E-UTRAN) which contains evolved NodeBs (eNBs), i.e. radio access points. Figure 2-1 shows an abstract view of the 4th generation network architecture [1]. The main functionality of an EPC network is supported by three principle entities [1]: Packet Data Gateway (PGW), Serving Gateway (SGW) and Mobility management Entity (MME). Other specialized nodes include the Home Subscriber Server (HSS), Policy Charging and Rules Function

(PCRF) and Online Charging system (OCS). A brief description of these entities is given below:

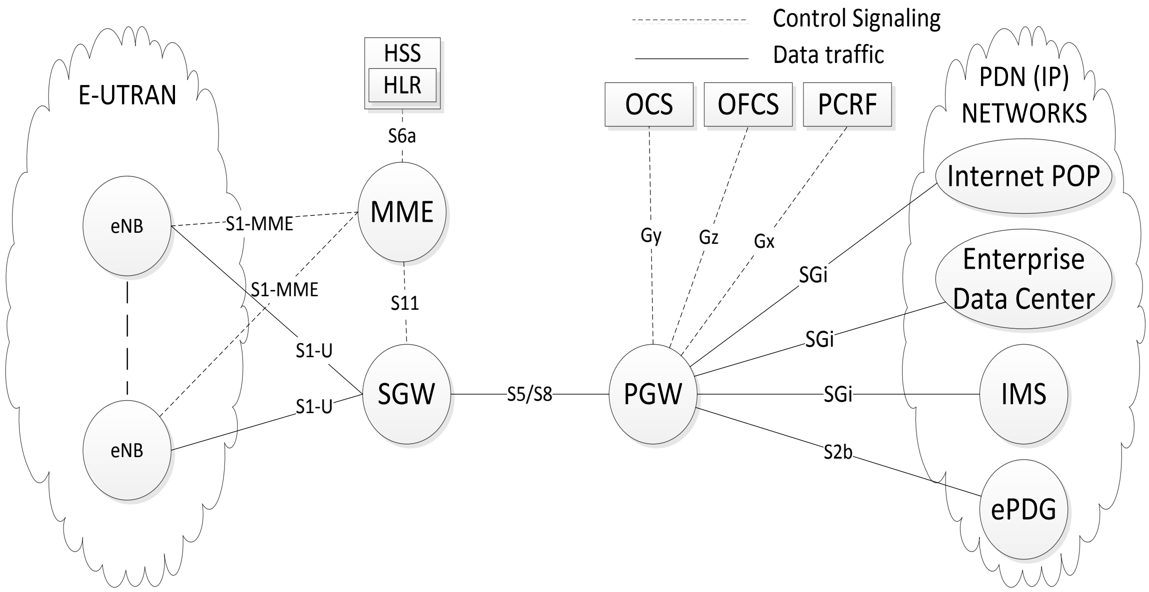


Figure 2-1: 4th generation network architecture

Packet Data Gateway (PGW): This gateway provides connectivity of EPC to external packet data networks (PDNs), e.g. internet. All user equipment (UE) IP address assignments are performed by PGWs, each from its allocated address space. The UE is allowed to connect via multiple PGWs simultaneously and obtain distinct IP address from each according to the recent 3GPP Network architecture. The PGW performs policy enforcement (e.g. rate enforcement), packet filtering per user, transport level packet marking for downlink, charging support and lawful interception. The PGW also serves as an anchor for mobility between 3GPP and non-3GPP access networks for the allocated address of the UE, such as in the case of where WiMAX is used as the radio access network.

Serving Gateway (SGW): User data traffic is forwarded by eNBs to SGWs. Each UE can connect to only one SGW and it serves as the mobility anchor for handovers between eNBs (intra-E-UTRAN). It also serves as mobility interface to other 3GPP access networks, e.g. 2G/3G networks. The SGW generates paging requests when downlink data arrives for UEs while in idle state. In addition it performs packet buffering and initiation of network triggered service request procedures. UE context information maintenance and user traffic replication, for case of lawful interception, are also performed by the corresponding SGW. Finally the SGW may perform transport level packet marking (uplink and downlink), packet routing/forwarding, and QoS support.

Mobility management Entity (MME): The MME is the key control entity as it is responsible for Non-Access Stratum (NAS) signaling (e.g. registration), bearer establishment (e.g. activation/deactivation), and PDN and SGW gateway selection for UEs at initial attachment and handovers. It is also involved in reachability procedures: tracking and paging UEs in the idle state. Furthermore, the MME is responsible for security procedures and for allocating temporary identities to UEs. Finally, the MME provides control plane functionality for mobility between LTE and non-3GPP access networks.

The E-UTRAN is the air interface of 3GPP Long Term Evolution (LTE) architecture that provides radio access to UEs. The backhaul network is a layered access network that provides connectivity between E-UTRAN and EPC. It provides required capacity and traffic differentiation to maintain quality of service (QoS) requirement.

2.1.2 GPRS TUNNELING PROTOCOL (GTP)

GTP [1] is a collection of protocols used for communication between the major EPC nodes, e.g. SGW, MME, and PGW. It is important to take in consideration of GTP protocol when dealing with EPC network design as EPC network, standardized by 3GPP, is largely dependent for its core operations on GTP. Therefore, any discussion of EPC design without referencing GTP protocol operations can be considered an abstract, if not even shallow, work. Moreover, GTP protocol stacks comprise the protocol stack for EPC nodes based on its role, which is vital for this study of a new network design.

GTP is central for IP mobility within 3GPP networks and uses a tunneling mechanism to support seamless mobility procedures. GTP protocol stack operates on top of UDP protocol making EPC network an overlay network with respect to IP networks. An IP-based network device cannot process and forward GTP packets without decapsulation. Thus, for conventional IP-based network devices to participate in the EPC architecture, they need to have GTP capability, which can be either by GTP hardware line cards or software processing of GTP packets. GTP is comprised of GTP-C, GTP-U and GTP variants. Figure-2 depicts GTP-C component utilized in for signaling amongst MME, SGW and PGW. It also supports the establishment of tunnels for mobility management and bearers for QoS management. The GTP-U component is the user part/bearer and is responsible for encapsulation and tunneling user’s IP packets.

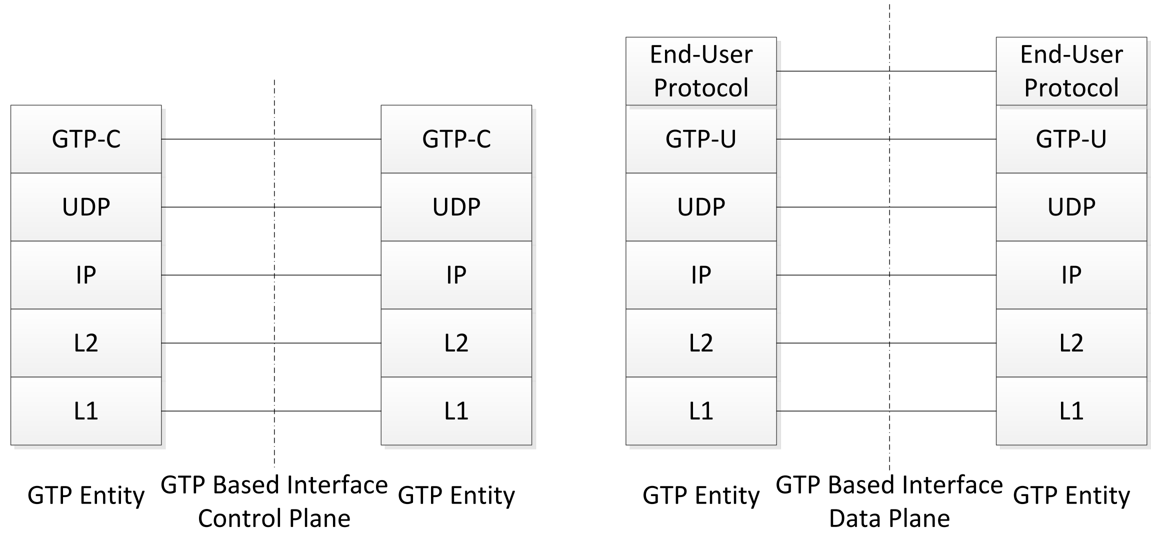


Figure 2-2: GTP protocol stacks

2.1.3 EVOLVED PACKET CORE CONTROL PLANE PROCEDURES

The EPC is comprised of two different planes: control plane and data plane. The control plane is responsible for UEs network attachment, session management, and mobility management. The procedures for control plane operations are clearly defined in 3GPP standards [1]. The data plane is responsible for UE traffic tunneling, QoS enforcement, etc. The main functionality of an EPC network that provides connectivity and seamless mobility to users can be viewed as:

Network attachment: the process by which the UE registers in the network and obtains IP connectivity for performing communication session.

Communication sessions: this contains the different aspects of sessions, including session setup, QoS negotiation, security procedures, etc. Various procedures are defined for these operations including: registration procedure, user-initiated session setup and dedicated bearer activation.

Mobility in active/idle mode: procedures required to enable EPC network track the location of UE while idle/moving in the network. Handover procedures are required to maintain seamless mobility of UE services while moving during active sessions. UE mobility involves different cases depending on EPC nodes involved in the handover. These cases include: inter-E-UTRAN mobility, intra-E-UTRAN mobility with EPC node relocation, mobility between EPC network, and other 3GPP access network.

2.1.4 GATEWAYS ARCHITECTURE

The gateways of EPC system, namely MME, SGW and PGW, as described in previous section provide border between mobile network and the fixed backhaul network, which is the reason for being named gateways. The EPC main entities have distinctive features that differentiate them from regular IP-based platforms, and it is essential to understand these aspects especially with regard to logical configuration, hardware architecture, and deployment locations. EPC nodes could assume various engineering designs in these areas which dictate how its functionality and possibly performance are executed. In this section, characterization of various properties of EPC gateways that are of interest to this study is presented, which will provide clearer understanding of network design choices later in the study.

A. Logical Configuration

In EPC network, the functionality of EPC nodes can be compounded together in different combinations resulting in four distinct architecture options [32], as shown in

Figure **2-3**. It should be noted that different logical arrangements leads to different handling of control signaling and of data plane traffic. The integration of more functionality might result in less delay and increased resource efficiencies, but the functional complexity and the cost of introducing new services would also increase.

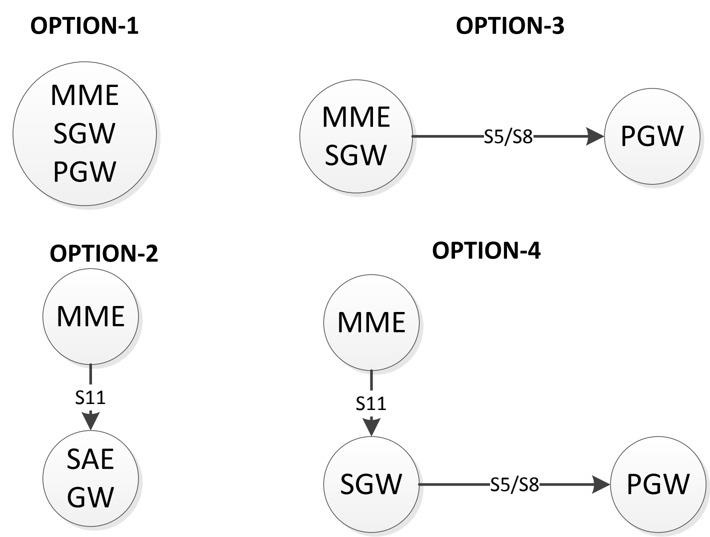


Figure 2-3: EPC network nodes logical architecture combinations

Option 1 shows the choice of grouping all functionality in one node which is the most stringent option. The performance in this option is maximized and efficiency is increased due to less communication delays and the minimal interfaces between logical entities. The downside of this configuration is that scalability into larger operator capacity is quite difficult and more complex. Moreover, the interoperability of this configuration is not possible [4].

Option 2 shows the choice of grouping nodes responsible of data plane handling, SGW and PGW, into a single entity and control plane functions, i.e. MME, is placed into another entity. This option is term as “flat architecture” and the combined entity is termed as system architecture evolution (SAE) gateway in 3GPP standards [21]. The advantage is that the optimum number of nodes is involved in processing data plane traffic, eNB and SAE gateway only. The reduction of hardware is achieved by 50%. However, the combination of SGW and PGW into a single entity eliminates availability of previously existing interface between them externally (which is called S5/S8 interface) and restricts its functionality into internal operations only. Whereas the presence of S5/S8 interface is much needed for certain reasons that force the split architecture of SGW and PGW. The reasons for the split architecture can be summarized in: a) multiple PDN connections is enabled, b) topologically distributed SGW and centrally located PGW, and c) inter-PGW mobility is not defined in 3GPP procedure thus for seamless mobility between different SGW and same PGW the separation is required.

Option 3 combines the control entity, i.e. MME, with the SGW. This option is not widely adopted in EPC networks; however, it is similar to serving GPRS support node and Gateway GPRS support node (SGSN/GGSN) allocation in UMTS network.

Option 4 is a typical operator configuration of an EPC network. Each node, SGW, PGW and MME, is located separately with its own logical space and defined interfaces. This case enables the highest horizontal scaling through native 3GPP mechanisms but introduces more signaling costs and delays than other configurations.

B. Hardware Architecture and Processing Capacity

The mobility management entity (MME) is the dedicated control plane element which is designed to process large control plane signaling received directly from cells into MME. In [6], they report that for large LTE network deployment the MME can experience in normal peak busy hours a signaling load of 500 to 800 messages per user equipment and up to 1500 per user in adverse conditions. They also report that during peak busy hour usage service requests can reach 45 requests per user per hour. The key dimensioning parameters for MME nodes is transaction capacity and subscribers density. The MME is expected to handle large surges of signaling requests per second at peak hours with high metrics performance. It is also expected to retain large amount of subscriber's information in its platform. These two parameters are expected to scale independently of each other. Due to the nature of general processing required for control plane operations, the MME platform is generally based on standard industry servers such as Advanced Telecom Computing Architecture platforms [5].

The SGW and PGW are optimized for high bandwidth data plane processing. IP-based router platforms are often used; these platforms are characterized by hardware optimized architecture which comprises of separate and dedicated control processors for control plane processing as well as a large quantity of network processors dedicated for high speed and high throughput packet processing and forwarding. The key dimensioning parameters for SGW and PGW nodes is data plane processing throughput, transaction capacity, and subscribers' density. These gateways hold large capacity of bearer contexts which is required to provide UEs with multiple connections with different QoS parameters.

C. Deployment locations

The location of deploying EPC nodes can affect performance in the control plane and data plane and it could optimize certain aspects, such as traffic delay, and services the operators may provide to its users, such as content distribution. The deployment location is categorized into two locations: core (central) location and local (remote) location. A core location is where a central infrastructure, data centers and other service provider facilities located to serve as the highest level in the hierarchy of the mobile network. The local location is a distributed location far from the central location where the operator has networking devices and possibly data centers that may be used as point of presence for operator network.

Each of the main nodes in the EPC, i.e. MME, SGW, and PGW, has a different interaction with regards to location. With the MME node, distributing MME nodes in remote locations is possible with the native 3GPP mechanism called MME pooling. A distributed/pooled MME architecture would lead to higher signaling load between MMEs. It requires as well communication with central location for fetching subscribers profiles into remote location, this is considered as less secure from the case where MME is centrally located and subscribers information are exposed into external network.

In case of PGW, a central PGW is required to provide services located within the core location such as IP multimedia subsystem (IMS) services. Bearers directed into services in central location face a certain supported QoS compared to services located within the internet. However, in certain services the traffic destination is terminated in internet locations that are closer to the UE location rather than the core PGW location, this could lead into non-optimal routing of traffic, and hence performance degradation occurs. A local PGW would enable local internet offload that is nearer to the user through data offloading mechanism defined in 3GPP standards such as selective IP traffic offloading [31]. This leads to a reduction of traffic load towards core network and also across the backhaul network. It also facilitates content distribution networks.

In case of SGW, a distrusted deployment of SGW is widely adopted as SGW acts as mobility anchor for UE connections between eNBs. The closer the anchor to the user the better performance the network can provide in mobility cases when the user connection is relocated from one SGW to another. In addition a local SGW can provide better forwarding of traffic between local cells.

2.2 SOFTWARE DEFINED NETWORK AND OPENFLOW

Software Defined Network (SDN) follows the concept of programmable network, where the network control plane is separated from forwarding elements [2][3]. Typical network elements, e.g. routers and switches, are developed with control plane and forwarding logic tightly coupled in the same entity. Control plane protocols perform numerous functions such as: interface state management, connectivity management, and topology information exchange such as IP/IPv6 routing and spanning tree protocols. The control plane processes decide the behavior of the data-plane such as forwarding, modifying, or dropping packets. SDN aims at decoupling control plane and data plane from network elements. The control plane is shifted from network elements toward a centralized entity, as shown in Figure 3. The network intelligence mechanisms, i.e. the control plane, are transformed from a distributed architecture to a logically centralized controller, referred to by the SDN controller. This centralization of network control can reduce network complexity and introduce more flexibility. The SDN controller resides in a network operating system (NOS) which harbors interfaces to various network applications that enforce a networking functionality on the network devices using network-wide view enabled by the control centralization. In this approach, the SDN controller receives information from forwarding elements and pushes rules to them that instruct them on how to handle traffic according to these rules. The most widespread open SDN standard with industry adoption is the OpenFlow (OF) protocol [8].

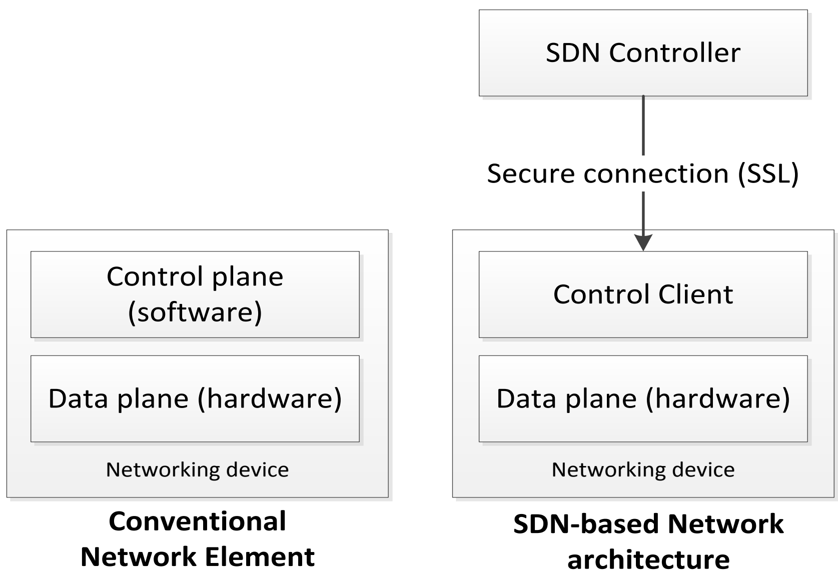


Figure 2-4: SDN architecture

2.2.1 SDN BENEFITS

The advantages of SDN concept can be seen in different aspects, the following is a depiction of main SDN benefits to networks in general:

Network innovation and centralized control of the network [2][3]: Designing distributed algorithms is complex, especially when defined at the protocol level. It needs to account for various constraints and network layers heterogeneity in the design to assure convergence. It also requires implementation of its own mechanisms irrespective of other algorithms, consuming more resources from network element. The SDN approach refactors functionality of control plane. Instead of closed boxes and distributed elements, an open interface to the hardware of network elements is defined to control these devices by a separate entity that household the intelligence of network control alongside network-wide view. The impact of introducing new feature is much simplified in the centralized entity as opposed to the distributed architecture.

Flexible control of traffic and flow abstraction [2][3]: By redefining network control to an abstraction layer, the control-plane can be more programmable which could solve architectural problems, reduce complexity and promote evolution. A network control program operates on a global view of network and desired goals of functionality, and generates configuration of each network device. Network device configuration can behave according to flow rules and set-up features by controller which can exhibit micro-flows or aggregated forwarding, reactive or proactive flow processing, virtual or physical resources handling or even can behave in a hybrid configuration. For example, An SDN-based network can be leveraged to perform load balancing, L3/L4-based or source/destination-based forwarding (policy-based routing), synchronized distributed policies (security, QoS, etc.), Software-based traffic analysis and others.

2.2.2 OPENFLOW (OF)

OpenFlow (OF) is a forwarding table management protocol. Each OpenFlow-based (OF-based) network forwarding element (FE) maintains a group of flow tables. Each flow table consists of forwarding/matching rules, called flow rules. Each flow rule consists of match fields, actions and statistics. Actions in flow rules dictate the operation performed on each packet that matches the flow rule upon arrival. Thus, a packet can be forwarded, dropped, encapsulated, etc. based on these rules. Flow table rules are acquired from OF-controller (SDN controller). The OpenFlow 1.0 specification defines twelve key fields including ingress port number, virtual LAN identifier (VLAN), and Layer 2, Layer 3, and Layer 4 information. The forwarding behavior can be based on any field supported in OpenFlow specifications and controlled by the OF-controller. The OF-controller decides the flow rules based on processes within the NOS that aim for a particular network forwarding behavior for each networking device as a hub or a switch or even acting as a router. It should be noted that OpenFlow, or SDN in general, has its constraints as well, such as limited table sizes, energy consumption used in matching process, new failure modes to handle, limited functionality of initial versions among others which are out of the scope of this work. The performance of OpenFlow under various conditions is studied in [25][26].

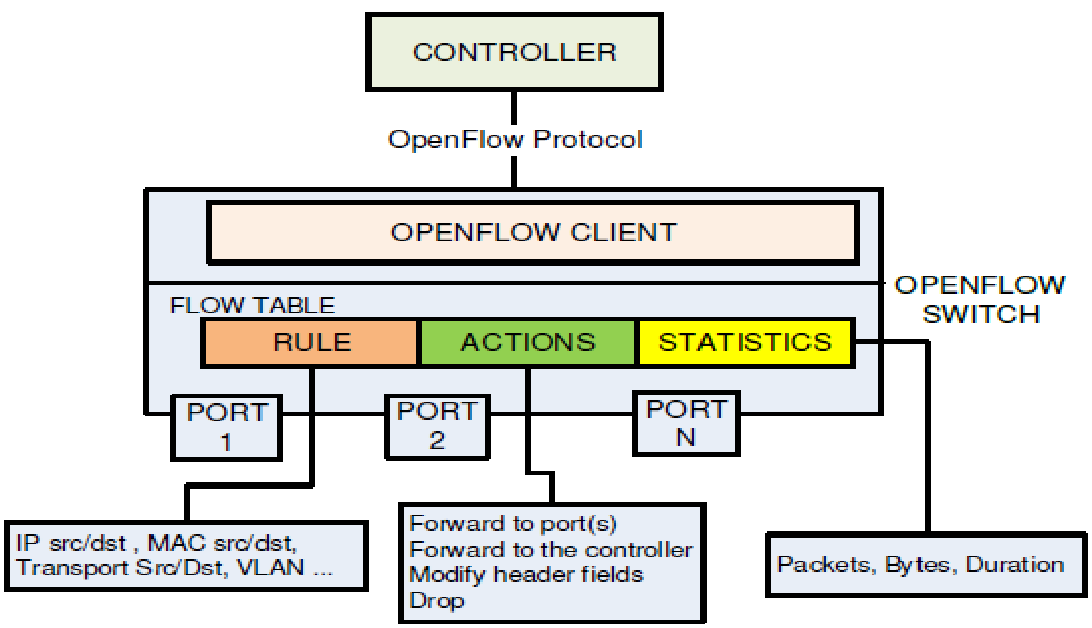


Figure 2-5: OpenFlow Architecture

CHAPTER 3

LITERATURE REVIEW

The concept of SDN-based mobile networks is gaining interest in the research community recently, and several researchers have presented their perspective of an SDN-EPC with a focus on specific properties to gain. The majority of previous attempts to tackle an SDN-EPC spanned around OpenFlow approach to SDN or an abstract modeling of SDN in the EPC. Few provided subtle analyses to provide a proof of concept for an SDN-EPC and its potential benefits. Therefore in this chapter, the characteristics of EPC that are the point for exploring SDN-EPC are presented, potential benefits gained from SDN in the EPC are enumerated in its different forms, a state-of-the-art survey is conducted that covers most attempts' concepts and ideas for an SDN-EPC and a general trend is analyzed and used to reach an understanding of where to explore SDN in the EPC.

3.1 EPC CHARACTERISTICS

The centralized architecture of core network which reflects on many dimensions: centralized monitoring, access control, and quality-of-service functionality at the PGW leads to concentrated data-plane traffic at PGW. This centralization introduces latency in services and content delivery and might induce congestion which is expected to increase due to significant growth in data traffic [10][19]. The centralized architecture can in some cases introduce un-optimized routing and resources inefficiency due to centralized architecture where traffic is tunneled over backhaul network towards the core [10][28].

Thus, network resources are over-provisioned to maintain busy hour traffic load, although these resources are underutilized in normal conditions. Moreover, the use of highly specialized and complex nodes, i.e. SGW/PGW/MME, increases the cost of the network and restricts the capability of adding new services. This centralized architecture is inflexible in the face of traffic dynamics and new services introduction [16][23]. Furthermore, different operators cannot easily interoperate different vendors' capabilities due to specialized interfaces [18][23].

The complexity of core network equipment requires vendor-specific configuration interfaces and communication through complex control protocols. This feature gives the operator less control over the core equipment in their mobile network [16]. Not to mention, the carriers might acquire core network equipment with functionalities that are not needed or be missing functionalities that cannot be added to the equipment [18][23].

3.2 SDN-BASED EPC SYSTEM

SDN introduces flexibility in network operation and provides simplified management of the network by removing the different control plane component from the distributed network nodes and concentrating those in a centralized control plane [19]. Several recent studies have proposed an SDN-based EPC architectures focusing on some of the gained benefits. The focus of these studies can be summarized as follows:

A common control-plane that can integrate different cellular technologies [16][30]. The common plane could: (1) reduce mobility management signaling and (2) reduce session establishment latency by installing flow rules in multiple switches simultaneously instead of performing hop-by-hop signaling [16]. However, this latency reduction is very dependent on the chosen network design.

SDN promotes the capability to distribute core network functionality over multiple, possibly cheaper, network nodes which in turn increases the scalability of the network and enables positioning of these nodes closer to the users.

Routing and Traffic Engineering: SDN can simplify routing by removing distributed routing architecture and substituting it with a centralized control plane aligned with rest of EPC control elements [19][23]. Furthermore, SDN can be used to enhance traffic management and steering [23] such as selective flow routing for in-line services [19], and for data-offloading techniques.

It enables flexible and fine-grained handling of traffic by directing traffic as dictated by the corresponding flow rules without the traditional restrictions of IP-based networks. Quality of service enforcement can be distributed among network device based on network-wide view where the SDN-controller performs access control and traffic scheduling over the involved switches.

Network function virtualization (NFV) is recently attracting attention in carrier networks [16][18][19]. NFV aims at decoupling network functions, e.g. DNS, Caching, etc., from hardware-based architecture, and migrate those functions to optimized software architecture to accelerate service innovation and provisioning. NFV can enhance service delivery and reduce overall costs [26]. Migration of core nodes to the virtual environment enables dynamic and flexible allocation of EPC functionality and services using cloud computing features and virtualization to meet load demands and increase utilization based on user's location [30].

Network virtualization (NV) is also attracting attention in carrier networks [16][18][19]. Network Virtualization (NV) refers to creating logical networks decoupled from the actual network hardware. Network functionality in NV is hosted in a virtual environment as virtual instances where the hardware is a general, off-the-shelf platform. NV is used in multi-tenancy environments. Cellular networks can benefit from NV for providing isolation between different classes of traffic such as roaming subscribers from home subscribers' achieved using different logical instances for each traffic class [16]. SDN is inherently an enabler for virtualization and has been used in recent attempts such as the one reported in [27]. SDN-based cellular network can also actively share the infrastructure among other operators [12][27].

The network intelligence is evolved into software-driven and decoupled from hardware or vendor dependence [12][16]. This promotes evolution of the mobile network to new technologies due to decoupling of core functionality into virtual environment and enables new services.

SDN with a distributed EPC architecture could offer scalability and optimized routing features where the benefits of a centralized control plane with distributed data plane is achieved [10]. In addition, a distributed EPC architecture enables dynamic/ distributed mobility schemes [10][23], this distribution offers optimized performance relative to centralized conventional schemes where handover functions are handed by distributed nodes closer to users in the backhaul.

3.3 RELATED WORK

Various researchers have viewed applying SDN concept to the EPC architecture in different approaches. In this section we present few studies that summarize the wide view of SDN-based EPC network.

The work in [13] presents a qualitative analysis of virtualized EPC architecture based on SDN. The study suggests that transferring the complete functionality of EPC node to virtualized platforms would retain the conventional monolithic architecture of the EPC network. It provides classification of the main functionality of EPC nodes followed by a proposal of functions split to enable different levels of migration of these functions to the virtual environment. Because OpenFlow does not support GTP in its current release, they propose four frameworks to handle GTP matching: network application on top of OF-controller, dedicated middleboxes, customized hardware OF-enabled switches with GTP functionality, and software OF-enabled switches with GTP modules. The alternatives for EPC functions placement out of the EPC nodes and towards the virtual environment are suggested to be as: full functionality migration, control-plane migration, signaling control migration, and scenario-based migration.

In [18], the authors provide two strata for SDN in EPC architecture. First, an SDN-enabled mobile network accommodating EPC nodes and controlled by custom controller, called MobileFlow, supported by custom interfaces. Second, an OF-enabled transport network which is controlled by an OF-controller. Forwarding elements in mobile network stratum are composed of custom MobileFlow elements (MFFE) that meets carrier grade functionality. MFFEs could advertise a flat architecture, however, details about functionality split or distribution of SGW and PGW among MFFEs is not provided. The work is based on actual implementation and the study reports the success of essential EPC functions verification. Latency and mobility test cases were not provided; the work appears to be in progress.

The researchers in [22] present a semi-distributed mobility scheme based on OpenFlow. They argue for dynamically delegating a part of mobility management such as anchor points to the backhaul network, so that a large part of 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. An analytical evolution of latency time for initial attachment and handover is performed based on an abstract model with various configurations. The proposed architecture focuses on an SDN-based backhaul network elements and how they can participate with EPC core network in traffic management such as mobility procedures. The proposal does not necessarily affect the EPC architecture; however, it suggests the removal of GTP tunnels used in the data path and relies on OF operations. A proposal to dispose of the, well-established, GTP tunnel is non-3GPP compliant.

The study in [23], proposes a scalable SDN-based mobile architecture. Their design follows a data center architecture where the network consists of a fabric of simple core switches, effectively removing the existence of specialized nodes in the core network, i.e. SGW and PGW. To implement service policy enforcement and mobility anchors, a local agent is introduced at the access edge. The local agents perform fine grained packet classification based on OF-controllers commands and translate source IP addresses to location dependent IP addresses. Their simulation results show switch flow table size versus number and length of service policy clauses and network size. The proposed work is considered non-3GPP compliant and does not include any latency results or analysis with regards to EPC core functionality.

Reviewing related works it can be seen that SDN can be explored in two network areas: backhaul network and core network. In the backhaul network, its advantage can be summarized in advanced and flexible traffic engineering which can help in load balancing, performance gain and aiding in mobility procedure. However, an SDN-based backhaul is not necessarily limited to EPC architecture but rather an overall impact on the backbone transport network. Imposing SDN on core network would have more impact on the EPC architecture. As can be seen from previous works, the different configuration of migration of EPC functionality can be either as: complete or partial migration of EPC nodes functionality to SDN control realm.

Complete migration of functionality: In the complete migration, control-plane and data-plane functionality, besides packet forwarding, of MME, SGW and PGW nodes are implemented as network applications on top of OF-controller. The deployed core network elements would be normal OF-switches that rely on the controller to perform EPC functionality. This configuration could be viewed as sub-optimal as special services need the data-plane traffic to traverse several nodes for every service. This may introduce latency and challenges the scalability of the network.

Partial migration of functionality: In the partial migration, MME being a control entity is always implemented as a network application. As for SGW and PGW various configuration may be used. In one configuration the PGW is unchanged and a data-plane

SGW (SGW-D) is deployed where the control part of SGW (SGW-C) is implemented as a network application. A second configuration is to deploy SGW-D and PGW-D in the core network and implement SGW-C and PGW-C as network applications. The first and second configurations can leverage the distribution of SGW-D closer to the access network which can potentially reduce latency in mobility cases in addition to other benefits. The second configuration can leverage the use a more stripped down and cheaper gateways to function as a PGW-D instead of a full-fledged expensive PGW. A third configuration is to deploy SGW-D and a combined S/PGW-D. The combined S/PGW-D can be leveraged to have more distributed deployment of gateways and closer to the access network. Thus, the S/PGW-D gateway is assumed to handle less traffic and users service requests.

CHAPTER 4

CONVENTIONAL EPC AND OPENFLOW NETWORK AND PROPOSED ARCHITECTURE

In this chapter, a comparison between an OpenFlow based network and EPC network is investigated pointing out the similarities and differences between these two architectures. Based on this comparison a basis for the proposed new EPC architecture is developed and described in detail.

4.1 ANALYSIS AND COMPARISON OF CONVENTIONAL EPC NETWORK AND OF NETWORK

As explained in section 2.2, SDN is based on decoupling the control plane from the forwarding plane. Many approaches to exploit SDN are based on the OpenFlow architecture which transforms a conventional networking device to a remotely controlled entity by a central controller. Control plane communication between the OF controller and OF switches is carried over secure TCP connections in a one-to-one communication

EPC is a special type of networks. Looking deeply at the architecture, we can perceive the similarity between the SDN standard and the EPC architecture but with slight differences. The reason for this is that the principal concept of SDN, i.e. the decoupling of control plane from the forwarding plane, is inherently present in EPC architecture through the use of the GTP protocol over UDP and also the S1 Application Protocol (S1AP) over the SCTP protocol. Recalling the functionality of GTP protocol discussed in 2.1.2, it defines a control stack (GTP-C) and user stack. The GTP-C is responsible for control plane signaling amongst MME, SGW and PGW. The S1AP protocol handles control plane signaling between E-UTRAN (eNB nodes) and EPC's MME. MME is the main control entity in the EPC architecture that controls the establishment, maintenance and deletion of connections by signaling pertinent nodes (eNBs, SGW). Thus the MME has the same role as the SDN controller.

In an OpenFlow enabled network, when a packet arrives at the switch it searches for match in its flow table and signals the controller in case a match is not found. The controller responds with appropriate flow rules to be installed in that switch and possibly to other switches on the path of that packet to reduce latency. This operation is performed symmetrically in all OF switches, and hence data traffic is assumed to traverse from any switch to another. In contrast to the OF network, an EPC network has very well-defined configurations of traffic flows. Furthermore, EPC has a well-defined network edges that are either E-UTRAN (eNBs) or PGWs, where EPC traffic must arrive or exit. Connections management between these nodes for a subscriber are initiated at an eNB by signaling the MME which in turn signals the SGW to create what is called a bearer context. A bearer context is similar to flow rules in OpenFlow, however, with much more information used for policy and QoS functions at the SGW and PGW gateways. The SGW then signals PGW to create a bearer context as well. EPC differs in the control communication from OpenFlow is that control signaling of PGW is initiated from the SGW in response to MME control commands rather from MME communication directly. Figure **4-1** depicts control plane communication flow for OpenFlow architecture and EPC architecture.

Considering the protocol layers and looking closely at OpenFlow, it is perceivable that the protocol has control over Layer 2, 3 and 4 protocols, specifically the 12 fields used in the flow rules and tables. Control traffic is transported using TCP protocol between OF switches and the controller. While for the EPC network, being an overlay network operations are carried out entirely above Layer 4 protocols. Control plane and User plane traffic are primarily encapsulated in UDP packets except for control signaling between MME and eNBs which is encapsulated in SCTP packets.

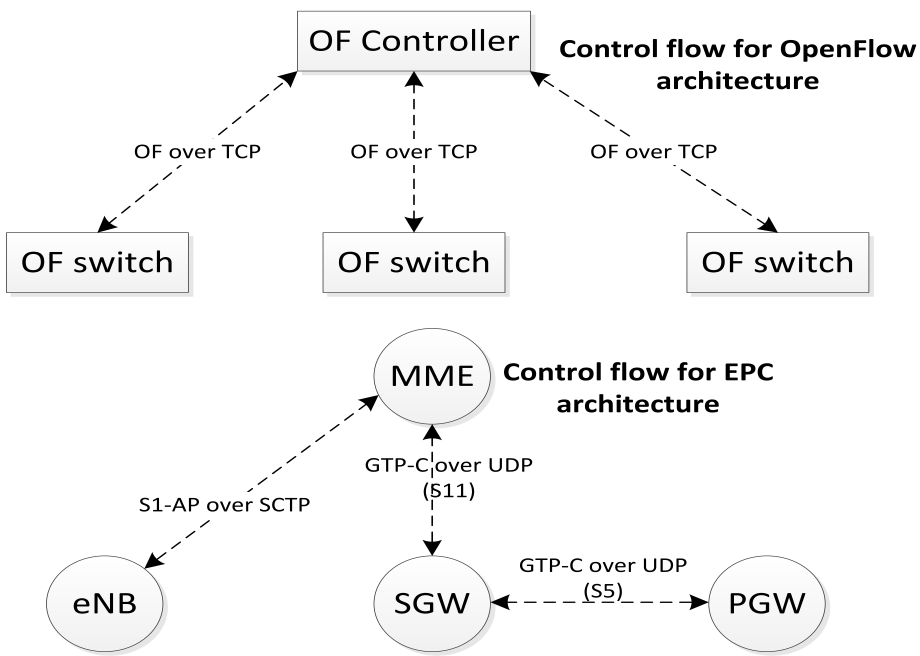


Figure 4-1: Control plane communication for OpenFlow architecture and EPC architecture

Various approaches to apply SDN in EPC network disregarded the fact that EPC is an overlay network. Previous works such as [19] treated the GTP protocol as an additional layer to OpenFlow. Therefore, new extensions to OpenFlow are proposed to accommodate EPC operations into OpenFlow domain where the OpenFlow controller is collocated with the MME. The new architecture suggests to enable OpenFlow switches to process GTP packets in both user plane and data plane, making them EPC-capable gateway nodes. However, GTP packet processing at high bandwidth is challenging for programmable OpenFlow switches. In this study perspective, this modification does not exploit the inherent features of EPC, that is already centrally controlled by the MME. The argument is that converting SGW and PGW into flow-based devices in the L2,L3, and L4, i.e. OpenFlow-like behavior, is not necessarily for the prime evolution of EPC into an SDN domain. This argument stems from two reasons. The first is that the number of EPC nodes in the data path is minimal which is three nodes at most: eNB, SGW and PGW. Thus the overall forwarding behavior of traffic is not largely affected by an OpenFlow-like behavior in comparison to OpenFlow-based backhaul network. The second reason is that EPC nodes have their own flow-like treatment of connections through bearer context defined in the GTP protocol. However, these approaches can be justified as an opportunity to develop EPC nodes gateways based on commodity hardware instead of customized and expensive hardware.

The researchers in [23] propose an integration solution on the control level between an OF-controller and the MME. OF-enabled devices are deployed in the backhaul network instead of conventional network switches. Information from the MME is exploited by the OF-controller to aid OF switches perform intelligent traffic forwarding and possibly assist mobility procedures. These solutions propose cooperation of backhaul network with EPC network to provide better traffic engineering solution in the backhaul, however, they don’t necessary affect inherent EPC architecture nor introduce SDN concept to the EPC architecture.

Other works such as [18], proposes to have an EPC system following the basic idea of SDN but not based on OpenFlow. Their proposal is to have custom nodes having GTP capabilities acting as EPC nodes and controlled by a central controller which can act as the MME. Their view is to have a pure SDN-based EPC architecture with GTP capabilities in the data plane, however, details on the needed control plane modifications or features are not provided.

4.2 Proposed SDN-based EPC Architecture

This study view of EPC architecture in the SDN domain is to have control signaling and control operations completely centralized in the central entity, the MME. This means to relocate control plane operations occurring between SGW and PGW to be controlled by the MME. The MME would be responsible for directly controlling PGW and SGW. Figure 4-2 and Figure 4-3 show an abstract view of the proposed control plane architecture for the SDN-based EPC network. The conventional control plane of EPC network depends on the MME-SGW interface, referred to as S11 interface, and the SGW-PGW interface, referred to as the S5 interface. The new proposed architecture extends fundamentally the S5 interface as control signaling would be initiated from the SDN controller (MME) rather than from the SGW. This consequently affects the GTP-C flow procedures of the S5 interface and shifts its functionality from the SGW into the SDN controller (MME). Therefore, a GTP-C interface extension should be created for the MME-PGW interface.

The SGW and PGW hold bearer contexts and state information about users' traffic, this entails that state information and bearer contexts, which was signaled from the SGW towards the PGW in the conventional architecture, is now initiated from the SDN controller (MME) towards PGW. The former requires a modification on the S11 interface as state information from the SGW may be required at the PGW in some cases to ensure consistent state information and compliant with the GTP protocol. The dependency between SGW and PGW must be resolved to ensure proper operation.

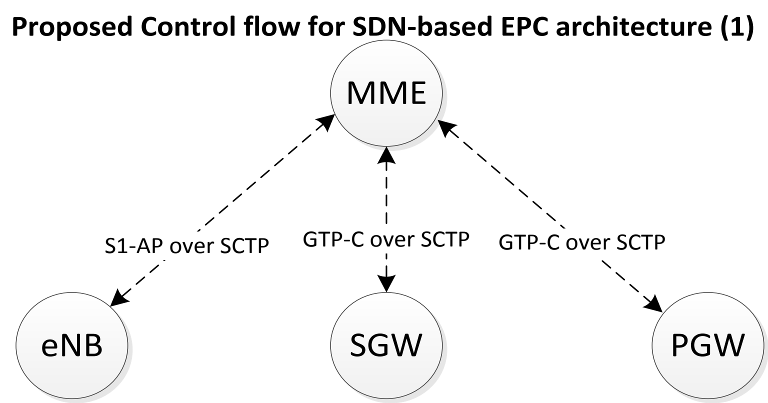


Figure 4-2: Proposed view of SDN-based EPC architecture (1)

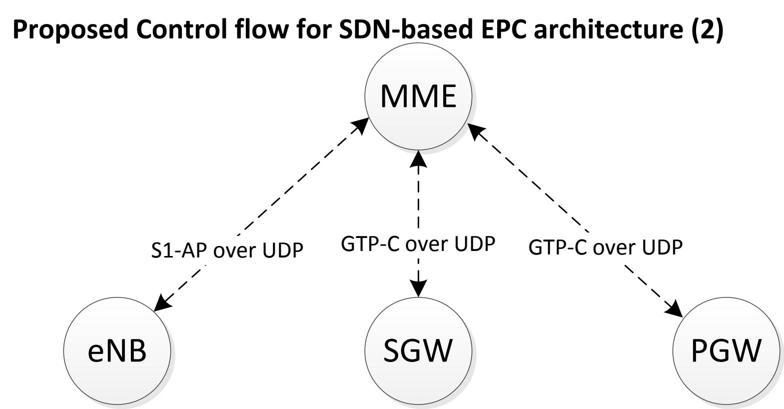


Figure 4-3: Proposed view of SDN-based EPC architecture (2)

As discussed previously in section 4.1 and shown in Figure **4-1**, the control plane communication in the conventional architecture is operating using GTP-C protocol over UDP and S1-AP protocol over SCTP. The S1-AP interface is kept unchanged as there is direct communication between eNBs and the MME and standardized interfaces are used. As for MME and gateways communications in the proposed architecture, two transport protocols are investigated as possible options for encapsulating control packets for the GTP-C interface. The first option is UDP protocol, as shown in Figure **4-3**. This choice would be in accordance with existing gateway's transport stack and does not add any overhead in terms of processing resources compared to conventional architecture. The second option is the SCTP protocol, as shown in Figure **4-2**. This choice is motivated by the requirement that an SDN-based EPC should promote distributed deployment of EPC nodes. With distributed deployment of nodes, the probability of lost packets is increased, thus we consider the SCTP protocol as the encapsulation layer for control packets to introduce reliability into packet delivery. Another advantage for the SCTP-based communication is the flow control capabilities of SCTP. For example, in case of signaling storms, SCTP can use throttling techniques to control the arrival rate of requests. The SCTP stack is not present in SGW and PGW in the conventional architecture, which means that additional firmware modifications are required at these nodes. Moreover, there is an overhead caused by acknowledgment packets and packet overhead introduced by SCTP headers, which are larger than that of UDP packets, and thus we expect slight variation in performance relative to the UDP option.

The effect of relocating control operation from SGW into MME leads to less required resource capacity at the SGW control plane and increases the resource requirement at the MME. This effect is desirable where less resources are deployed at the network edge which can be software based platforms rather than custom made, and more resources are required at the central location where higher processing capacity is located.

4.2.1 Modification required from conventional to proposed architecture in GTP protocol

By relocating control operations from gateways, i.e. SGW and PGW, into central control entity, i.e. MME, control information that are typically generated at the gateways will be now generated at the MME. The main part that this study focuses on are tunnel management information. It is also noted that other information may be relocated as well such as charging information; however this aspect is out of the scope of this study.

To modify GTP protocol to be centralized at the MME we identify control information flow characteristics which are as follows:

- Case-A: Information originating in eNB or MME and is destined to SGW and PGW
- Case-B: Information originating in eNB or MME and is destined to SGW only.
- Case-C: Information originating in eNB or MME and is destined to PGW only but transferred through SGW transparently.
- Case-D: Information originating in SGW and is destined to PGW.

The goal is to make these control operations generating these information located at the MME and the MME in return communicates with each gateway independently. That requirement means that the same GTP message sent in conventional architecture is now duplicated in the SDN architecture and sent to both gateways but with some modifications. These modifications, required in light of the control information flow characteristics stated above, will be reflected on the architecture in an abstract manner as follows:

- Case-A: same information elements are duplicated at the MME and sent in two different GTP message to SGW and PGW simultaneously. An example of these fields is UE information fields such as mobile subscriber identifications fields.

- Case-B: By default these information are not required to reach PGW, thus they will not be present in the duplicated GTP message.

- Case-C: Information elements that are previously transparently transferred through SGW is not required to be present in GTP message destined to SGW only. An example of this information field is the Protocol Configuration Options field which is required only at the PGW and originates from UE device.

- Case-D: These information are now generated at the MME and required to be added to both GTP message destined to SGW and PGW, hence both require them for operations. An example of these information fields are tunnel management information such as tunnel end identifiers required for establishing user plane tunnels between gateways for transporting user data traffic.

Imposing these changes should affect all operations of the control plane. In specific it will affect the sequence of message flow exchanged between EPC nodes and the information fields contained in these messages. This study focuses on registration procedure and S1-handover procedure, therefore the following sections present the message flow and message fields as per the conventional architecture and proposed SDN-EPC architecture. Other control operations will require an equivalent transformation; however we suffice this work for the registration and S1-handover procedure as representative of common control operations.

4.3 Control plane procedures under conventional and proposed architecture

Section 2.1.3, describes control plane procedures in EPC network. The main control procedures are session management and mobility management. The session management is comprised of the Initial attach procedure, user-initiated service procedure, network-initiated service procedure and others. The various session management procedures have the same sequence of message exchange, the difference among them lies with information being relayed among participating nodes, i.e. MME, SGW, and PGW. Thus, we choose to analyze only initial attach procedure, shown in Figure 4-4. Mobility management comprises of idle mobility and active mobility. In the idle mobility, most signaling interaction happens between the MME and the eNBs which is not the concern of this study. In active mobility, we tackle the S1-based handover procedure, shown in Figure 4-6. The S1-based handover is an intra-LTE handover and happens when the subscriber is in an active session and moves from his location. The S1-based handover procedure may relocate either the serving MME, the SGW or both (Intra-MME and Intra-SGW). In this study, we consider SGW relocation occurrence in the S1-based handover which is of particular interest in the case of many distributed SGWs.

4.3.1 Initial Attach (Registration) Procedure

When a UE first connects to access radio, it registers within EPC nodes and the connection information state is maintained continuously. The communication path to the PGW is established. The message flow for this procedure is shown in Figure 4-4. At the start of the procedure, the eNodeB where the UE is attached sends an ATTACH

REQUEST to the MME with all required identification of UE. The MME performs authentication of the UE by retrieving its credentials from the HSS. After authenticating UE identity and credentials, the MME sends CREATE SESSION REQUEST message to the closet SGW which contains all required information of UE. The SGW create a new entry in its bearer table and sends a CREATE SESSION REQUEST to PGW. The PGW also creates a new entry in its bearer table, allocates UE IP address, and generates the required context information such as charging identification and others. The PGW then sends a CREATE SESSION RESPONSE to SGW which in turn creates its similar response and send it toward the MME. When the MME receives the request response, it updates information about UE activated bearers in the HSS and responds back to E-UTRAN (eNB) with the operation result.

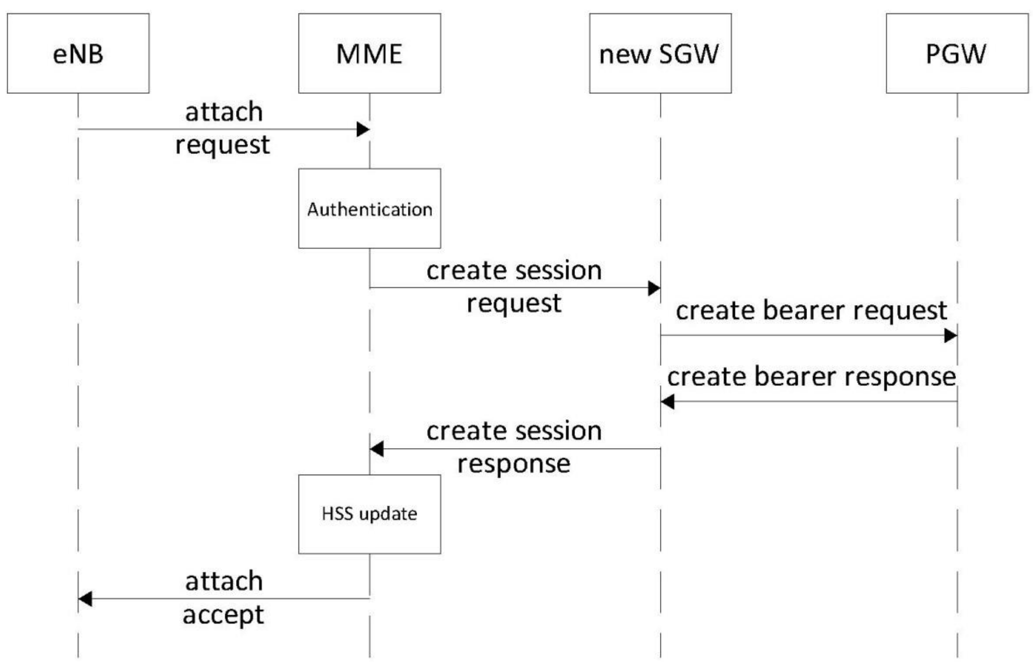


Figure 4-4: Registration procedure in GTP protocol [1]

The Initial Attachment procedure as described above involves control operation between SGW and PGW. To apply the SDN concept on this operation, we send CREATE SESSION REQUEST directly from MME towards PGW where all information required for PGW operation is decided at MME rather than SGW. Thus, PGW does not depend on SGW to create its bearer entries. Furthermore, SGW would depend on the MME in creating its bearer entries because MME is now responsible for bearer management rather than gateways themselves. The proposed message flow for SDN-based Initial Attach procedure is shown in Figure 4-5.

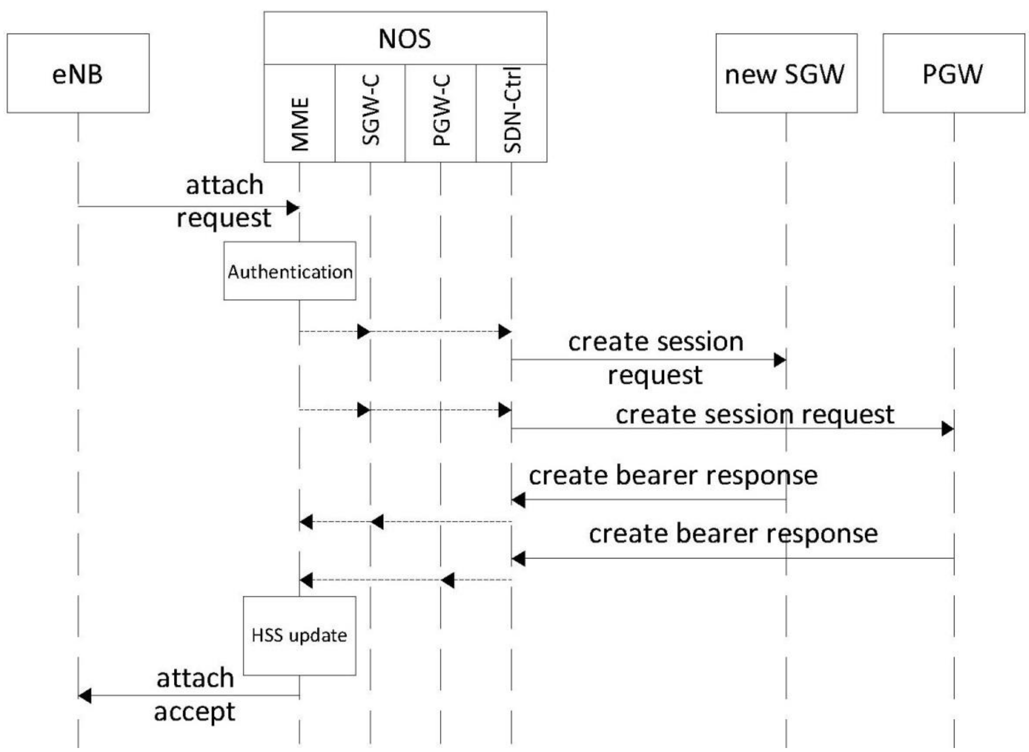


Figure 4-5: Proposed Registration procedure in GTP protocol for SDN architecture

The information fields in modified GTP messages required in both architectures with the proper modification is presented in Table 4-1 to Table 4-4. The table shows which fields that will be added or removed from modified GTP messages.

Table 4-1: Create Session Request GTP message

Message Type: Create Session Request		**Interface**			
		Conventional		**SDN-based**	
GTP-C Header	12 Bytes	√	√	√	√
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S5-Ext**
IMSI	11	√	√	√	√
MSISDN	12	√	√	√	√
ME Identity	12	√	√	√	√
User Location Information (ULI)	43	√	√	√	√
Serving Network	7	√	√	√	√
RAT Type	5	√	√	√	√
Indication Flags	0	√	√	√	√
Sender F-TEID	13	√	√	√	√
PGW S5/S8 Address	8	√		√	
APN	35	√	√	√	√
Selection Mode	5	√	√	√	√
PDN Type	5	√	√	√	√
PAA	9	√	√	√	√
APN Restriction	5	√	√	√	√
AMBR	12	√	√	√	√
EPS Bearer ID (EBI)	5	√	√	√	√
Bearer Context to be created	*	√	√	√	√
Bearer Context to be removed	9	√	√	√	√
Protocol Configuration Options	10	√	√		√
Trace Information	0	√	√	√	√
Recovery	0	√	√	√	√
MME-FQ-CSID	11	√	√	√	√
SGW-FQ-CSID	11		√	√	√
PGW-FQ-CSID	11			√	√
UE Time Zone	6	√	√	√	√
User CSG	12	√	√	√	√

Charging	6	√	√	√	√
Signaling Priority	5	√	√	√	√
Total Size		**323**	**313**	**356**	**345**
* note-1: values for Bearer Context to be created are in table below					

Table 4-2: Bearer Context Created within Create Session Request information elements

Message Type: Bearer Context Created within Create Session Request Message		**Interface**			
		Conventional		**SDN-based**	
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S5-Ext**
EPS Bearer ID (EBI)	5	√	√	√	√
Bearer TFT (Optional)	0	√	√	√	√
S1-U eNodeB F-TEID (S11)	13	√		√	
S5/S8-U PGW FTEID - S11	13	√		√	√
S5/S8-U SGW FTEID - S5	13		√	√	√
Bearer Level QoS	26	√	√	√	√
Charging Id (S5)	8			√	√
Total Size		**61**	**48**	**82**	**69**

Table 4-3: Create Session Response GTP message

Message Type: Create Session Response		**Interface**			
		Conventional		**SDN-based**	
GTP-C Header	12 Bytes	√	√	√	√
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S11-Ext**
Cause	6	√	√	√	√
Change Reporting Action	5	√	√	√	√

CSG Information Reporting Action	5	√	√	√	√
Sender F-TEID for CP (S11)	13	√		√	
PGW S5/S8/S2b FTEID	13		√		
PDN Address Allocation (PAA)	9	√	√	√	√
APN Restriction	5	√	√	√	√
APN-AMBR	12	√	√	√	√
Protocol Configuration Options	10	√	√		√
Bearer Contexts created	*	√	√	√	√
Bearer Contexts marked for removal	15	√	√	√	√
Recovery	5	√	√	√	√
Charging Gateway Address	8	√	√	√	√
PGW-FQ-CSID (S5)	11		√		
SGW-FQ-CSID (S11)	11	√			
Total Size (S11)		**187**	**182**	**175**	**177**
* note-1: values for Bearer Context to be created are in table below					

Table 4-4: Bearer Context Created within Create Session Request information elements

Message Type: Bearer Context Created within Create Session Response		**Interface**			
		Conventional		**SDN-based**	
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S5-Ext**
EPS Bearer ID (EBI)	5	√	√	√	√
Cause	6	√	√	√	√
S1-U SGW F-TEID (S11)	13	√			
S5/S8-U PGW FTEID (S11, S5)	13	√			
Bearer Level QoS	26	√	√	√	√
Charging Id (S5)	8		√		
Total Size (S11)		**67**	**62**	**41**	**41**

4.3.2 S1-based Handover Procedure with SGW Relocation

In general, a handover occurs when utilized radio channel needs to be exchanged with neighboring cell. S1-based handover is required for aiding seamless mobility when there is relocation of the SGW [1]. The 3GPP prescribed message flow of S1-based handover is shown in Figure 4-6. The information elements (IE) present in the GTP messages exchanged in this procedure are shown in APPENDIX A. After the decision for a handover is made at the source eNB, the eNB sends a HANDOVER REQUIRED message to the MME. The message would indicate which bearers are subject to data forwarding. The MME selects the appropriate new target SGW and sends a CREATE SESSION REQUEST that contains bearer context, PDN addresses, Tunnel End Identifiers (TEID) for GTP protocol, and other required information to this target SGW. The target SGW responds to the MME by a CREATE SESSION RESPONSE message.

Then MME sends a HANDOVER REQUEST message to the target eNodeB which creates UE contexts. The eNodeB sends a HANDOVER REQUEST ACKNOWLEDGEMENT to the MME with bearer setup list which contains addresses and TEIDs for downlink traffic and receiving forwarded data, if necessary. The MME sets up forwarding parameters at target SGW by sending CREATE INDIRECT DATA FORWARDING TUNNEL REQUEST to target SGW which responds a CREATE INDIRECT DATA FORWARDING TUNNEL RESPONSE. The MME sends HANDOVER COMMAND to source eNodeB with list of bearers subject to forwarding and bearers subject for release. The source eNodeB sends eNodeB STATUS TRANSFER to MME which sends this information to the target eNodeB via the STATUS TRANSFER message. After the UE has successfully synchronized with the target eNodeB, it sends a HANDOVER NOTIFY to the MME. The MME sends a MODIFY BEARER REQUEST to target SGW with target eNB address and allocated TEID for each PDN connection. The target SGW assigns addresses and TEIDs for downlink traffic from the PGW. It then sends a MODIFY BEARER REQUEST towards PGW for each PDN connection. The PGW updates its bearer contexts and responds with MODIFY BEARER RESPONSE. The PGW starts forwarding traffic towards SGW using the newly assigned address and TEID. The target SGW receives the message from PGW and responds to the MME with a MODIFY BEARER RESPONSE. After timer expiry the MME sends messages to source eNodeB for release of resources via UE CONTEXT RELEASE COMMAND and sends a DELETE SESSION REQUEST to the source SGW. After both source eNodeB and source SGW has responded, the MME initiates DELETE INDIRECT DATA FORWARDING TUNNEL REQUEST to target SGW to release temporary resources.

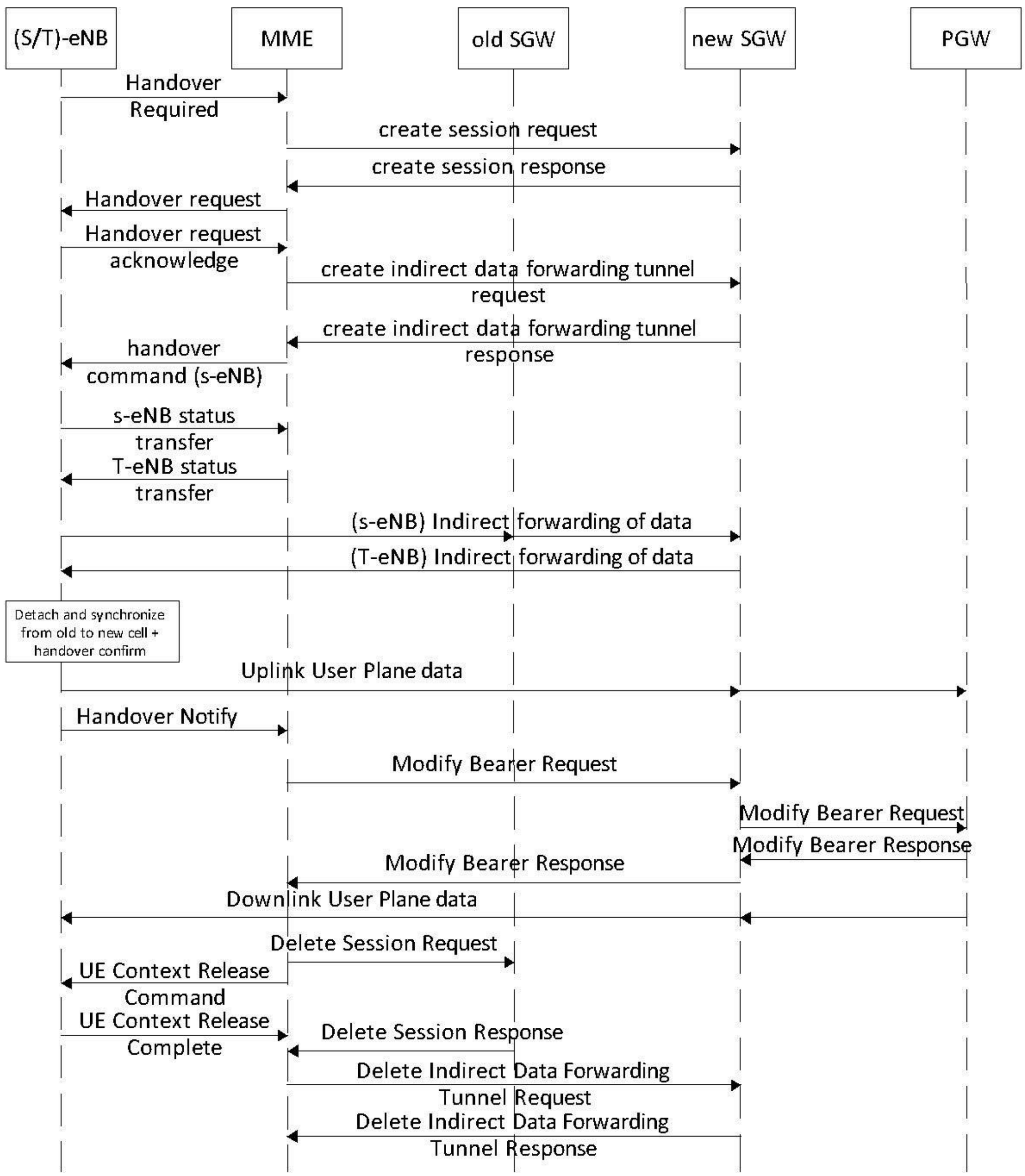


Figure 4-6: S1-based handover procedure in GTP protocol [1]

The procedure described earlier is executed in the conventional architecture. To apply the concept of SDN on this procedure, we identify the part where SGW-PGW control exchange occurs. This happens when the HANDOVER NOTIFY is received at the MME, the MME sends a MODIFY BEARER REQUEST to target SGW which in turn sends another MODIFY BEARER REQUEST to PGW. And the responses flow in the opposite direction. Applying the centralized concept of SDN, we identify the information required at the PGW and send it simultaneously as MODIFY BEARER REQUEST to PGW. The PGW in turn sends its response directly towards MME without SGW intervention. The proposed message flow procedure is shown in Figure 4-7.

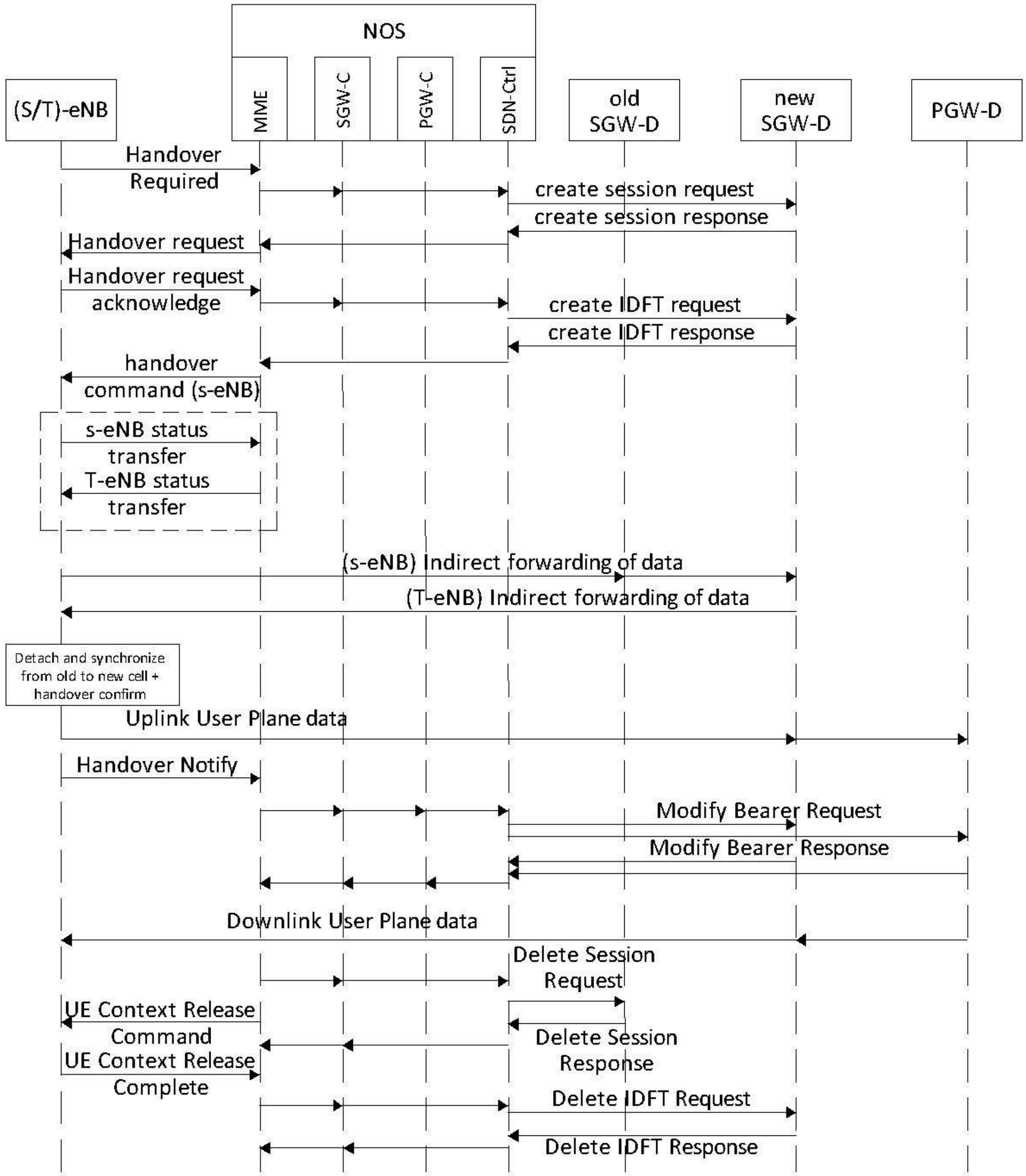


Figure 4-7: Proposed S1-based procedure in GTP protocol for SDN architecture

The information fields in the affected GTP messages required in both CONVENTIONAL and SDN-EPC architectures with the proper modification are presented in Table 4-5 to Table 4-8. The tables show which fields that will be added or removed from modified GTP messages.

Table 4-5: Modify Bearer Request GTP message

Message Type: Modify Bearer Request		**Interface**			
		Conventional		**SDN-based**	
GTP-C Header	12 Bytes	√	√	√	√
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S11-Ext**
MEI	12	√	√	√	√
ULI	43	√	√	√	√
Serving Network	7	√	√	√	√
RAT Type	5	√	√	√	√
Indication Flags	0	√	√	√	√
Sender F-TEID	13	√	√	√	√
AMBR	0	√	√	√	√
Delay Value (S11)	5	√		√	
Bearer Contexts to be modified	*	√	√	√	√
Bearer Contexts to be removed (S11)	9	√		√	
MME-FQ-CSID	11	√	√	√	√
SGW-FQ-CSID	11		√	√	√
PGW-FQ-CSID	11			√	√
UCI	12	√	√	√	√
Total Size (S11)		**155**	**152**	**177**	**163**
* note-1: values for Bearer Context to be created are in table below					

Table 4-6: Bearer Context Created within Modify Bearer Request information elements

Message Type: Bearer Context to be modified within	**Interface**

Modify Bearer Request		Conventional		SDN-based	
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S5-Ext**
EPS Bearer ID (EBI)	5	√	√	√	√
S1 eNodeB F-TEID (S11)	13	√		√	
S5/8-U SGW F-TEID (S5)	13		√		√
Total Size (S11)		**22**	**22**	**22**	**22**

Table 4-7: Modify Bearer Response GTP message

Message Type: Modify Bearer Response		**Interface**			
		Conventional		**SDN-based**	
GTP-C Header	12 Bytes	√	√	√	√
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S11-Ext**
Cause	6	√	√	√	√
MSISDN	12	√	√	√	√
EPS Bearer ID (EBI)	5	√	√	√	√
APN Restriction	5	√	√	√	√
Protocol Configuration Options	10	√	√		√
Bearer Contexts modified	*	√	√	√	√
Bearer Contexts marked for removal	15	√	√	√	√
CSG Information Reporting Action	5	√	√	√	√
SGW-FQ-CSID	11	√			
PGW-FQ-CSID	11	√	√		
Total Size (S11)		**124**	**108**	**87**	**97**
* note-1: values for Bearer Context to be created are in table below					

Table 4-8: Bearer Context Created within Modify Response Request information elements

Message Type: Bearer Context to be modified within Modify Bearer Response		**Interface**			
		Conventional		**SDN-based**	
Information Elements (IE)	**IE size (Bytes)**	**S11**	**S5**	**S11**	**S5-Ext**
EPS Bearer ID (EBI)	5	√	√	√	√
Cause	6	√	√	√	√
S1 SGW F-TEID (S11)	13	√			
Charging ID (S5)	8		√	√	√
Total Size (S11)		**28**	**23**	**23**	**23**

CHAPTER 5

SYSTEM DESIGN AND IMPLEMTATION

5.1 INTRODUCTION

The target of this study is to evaluate the performance metrics of the new proposed architecture in its both variants SCTP-based and UDP-based transport encapsulation relative to those for the conventional architecture. The evaluation is targeting the control plane operations of EPC network are the focus of the proposed and conventional architecture. The control plane procedures evaluated are the (1) Initial attachment procedure, and (2) S1-based handover procedure as described in section 4.3 for both architectures. The experiments are designed to reflect as closely as possible the real performance metrics characteristics in real deployed networks while sustaining controllable and adequate system simplicity, therefore some simplifications and engineering choices are made to generate this design. Not to mention that the combination of many network technologies and resources would lead to diverse results. However, the output of this study can serve as an indicator to realistic measurement metrics values, and most importantly, it may be used for comparison between architectures under study while running in the same network and under same conditions.

To perform this evaluation, we conduct a series of simulation experiments sets using the simulation package OMNeT++ [39]. The choice of simulation over a real testbed experiment lies in the fact that controlling various factors of the experiments is not possible using real platforms, not to mention the availability of such hardware. This study depends on varying many factors while setting others at specific levels at the same run. Consequently, the option of using simulation package is sought to be the optimum solution.

5.2 NETWORK SIMULATOR: OMNeT++

In this work, network simulation environment OMNeT++ is utilized to build the simulation test scenarios. OMNeT++ is one of the reliable and reputable network simulation packages. It is an object-oriented discrete event simulation system based on C++. It is designed to simulate various types of networks such as computer networks and wireless sensor networks. It is available under a free academic version for students use containing the main functionality required for building network simulations. OMNeT++ provides adequate packages to simulate various types of networks. In this study, the INET framework [33] is utilized within OMNeT++ package. The INET framework provides standard sets of APIs and standard protocols such as SCTP and UDP which will be used in building the simulation testbed. It enables creating application functionality flexibly through using the standard communication protocol APIs. OMNeT++ is capable of running large scale simulations with many repetitions, since it utilizes multithread technology and provides use-friendly graphical user interface. This allows us to simulate extreme scenarios under heavy loads and with many repetitions.

5.3 PERFORMANCE METRICS

This work attempts to evaluate the following measurement metrics related to network performance:

- **Operation End to End Delay (E2ED)**: the time required for control plane operation (Initial attachment and S1-based handover) to complete.
- **Resource utilization (CPU-Util)**: the amount of resources consumed by each network device, i.e. EPC nodes, in for performing all of its control plane operations.
- **Bandwidth utilization (BW-Util)**: the bandwidth consumed at MME links which consist of signaling load during its control plane operations.

5.3.1 PERFORMANCE BOTTLENECKS

It is important to understand performance bottlenecks to be able to explain variation in performance metrics. The following factors are the main contributors to the performance metrics:

- **Operation End to End Delay (E2ED)**:
  - Processing delay: depends on CPU, memory and load.
  - Transmission delay: depends on packets size and links data rate.
  - Propagation delay: depends on link distances and media.
  - Queuing delay: depends on factors such as background traffic and processing delay.
- **Resource utilization (CPU-Util):**
  - Processing delay: depends on processing capacity of computing entity in server and number of servers, e.g. multi-processor.
  - Rate of requests arrival: depends on control plane requests rate.
- **Bandwidth utilization (BW-Util)**:

- Number of control plane packets generated or received by MME: depends on control plane communication behavior.
- Size of control plane packets generated or received by MME: depends on application packet size (GTP-C) and transport layer encapsulation.

5.4 EXPERIMENT FACTORS

The experiment is based on varying multiple factors while setting others at constant values. The factors of this study are as follows:

- **Data plane background traffic (DBGT)**: this represents user traffic traversing the network from eNodeB towards its final destination, i.e. PGW. The DBGT shall be varied during simulation scenarios as percentage of the backhaul links.

- **EPC nodes control plane processing capacity**: each EPC node has finite processing capacity which is reflect in time required to complete requests arriving from UEs or other EPC nodes. In this study, the effect of finite processing capacity is reflected through predefined processing capacity. This factor concerns two entities: the MME and the gateways; each have separate configurations. The MME capacity factor shall be termed MMECAP, and the gateways, i.e. SGW and PGW, processing time shall be termed GWPT.

- **Control plane background traffic (CBGT)**: this represents control plane operations running besides control procedure under study during the test scenario. The nominal control plane background traffic is comprised mainly of session management traffic as reported in [6]. In this study, the CBGT is completely comprised of user initiated service request operations, which is a type of session management procedures similar to initial attach procedure. This factor is reported as requests per second and increased across different experiments.

- **Control operation request rate (CORR)**: this factor is the rate of control operation under study (mobility or registration) arriving at the MME.

- **Propagation delay of backhaul link (DEL)**: communication links connecting core and local mobile centers, and the radio access sites (eNBs) are configured to have propagation delays reflecting reasonable delay consistent with distance between the location of the entities and processing time of intermediate nodes.

The experiment includes six factors for which different settings of each factor could lead to variation in performance metrics. TABLE 5-1 summarizes factor levels in simulation experiments. DBGT is represented as percentage of backhaul link data rate; whereas all links are configured to 10Gbps bandwidth. CBGT is varied to reflect various loading effect on the MME processing resources, i.e. low, medium, and high load. GWPT which determines processing time for one request in EPC gateways, i.e. SGW and PGW, is configured to one of three values 10, 75, and 150 microseconds. 10 micro-second reflects the case of hardware-based platforms and 150 micro-second reflects the case of software-based platforms as reported in [36][37]. A Third option is added (75 microseconds), this option is assumed for an improved or hybrid software based platform that can perform more adequately for the operation of a mobile network. The MME-CAP is configured to 30Mbps which reflects an estimate of average capability for handling 5000 request per second based on information in [38] for MME dimensioning. This is considered the minimum MME processing capacity at 10Gpbs. CORR is configured to levels similar to deployed network requests distribution which is roughly around 5% to 15% of total control plane processing capacity[6]. Propagation delay of backhaul links is configured to reflect various distances of eNBs and local center from EPC core center location.

TABLE 5-1: FACTORS LEVELS

	Factor Type	**#levels**	**Parameter Levels**
1	Data plane back ground traffic (DBGT)	3	{ 20%; 50%; 80%} of backhaul link rate
2	EPC nodes CPU Process time (GWPT)	3	PGW/SGW Process time = 10, 75, 150 usec
3	MME processing capacity (MMECAP)	2	{30,60} MByte/sec
4	Control plane back ground traffic (CBGT)	3	{ 20%; 50%; 80%} of MME capacity
5	Control operation request rate (CORR)	3	{ 5%; 10%; 15%} of MME capacity
6	Backhaul link propagation delay (DEL)	3	0.1, 0.5, 1.0 msec

As shown in Table 5-2 and Table 5-3 that CBGT and CORR will be varied in percentages of MME capacity, this selection of load is based on the fact that the absolute number of requests arrival for both CBGT and CORR is not the effective approach to report this factor effect; it is the utilization of MME resources due to this factor that directly reflects on performance metrics. The tables show an experimental average estimation of CBGT and CORR requests rate of arrival on MME resource utilization for both EPC network types. An abbreviated term, e.g. CP1, is used hereinafter to refer a particular CBGT and CORR requests rate. It should be noted that MME utilization in conventional EPC and SDN-EPC have some variations, as discussed in previous chapter, due to shifting some operations from gateways to MME in the EPC architecture. This variation will be presented in the results section, but the table below shall be used as a common reference point.

Table 5-2: CONTROL OPERATIONS REQUEST RATE AT APPROXIMATED MME LOADING FOR S1-BASED HANDOVER PROCEDURE

CBGT LEVEL	CBGT REQUESTS RATE (per second) MME CAPACITY		CORR LEVEL	CORR REQUESTS RATE (per second) MME CAPACITY		CONTROL LOAD POINT ABBREVIATION	APPROXIMATE AVERAGE MME LOAD
	30 MB/sec	60 MB/sec		30 MB/sec	60 MB/sec		
CBGT-1	1300	2600	CORR-1	78	155	CL1	25%
CBGT-1	1300	2600	CORR-2	155	310	CL2	30%
CBGT-1	1300	2600	CORR-3	233	465	CL3	35%
CBGT-2	3200	6400	CORR-1	78	155	CL4	55%
CBGT-2	3200	6400	CORR-2	155	310	CL5	60%
CBGT-2	3200	6400	CORR-3	233	465	CL6	65%
CBGT-3	5200	10400	CORR-1	78	155	CL7	85%
CBGT-3	5200	10400	CORR-2	155	310	CL8	90%
CBGT-3	5200	10400	CORR-3	233	465	CL9	95%

Table 5-3: CONTROL OPERATIONS REQUEST RATE AT APPROXIMATED MME LOADING FOR REGISTRATION PROCEDURE

CBGT LEVEL	CBGT REQUESTS RATE (per second) MME CAPACITY		CORR LEVEL	CORR REQUESTS RATE (per second) MME CAPACITY		CONTROL LOAD POINT ABBREVIATION	APPROXIMATE AVERAGE MME LOAD
	30 MB/sec	60 MB/sec		30 MB/sec	60 MB/sec		
CBGT-1	1300	2600	CORR-1	188	375	CL1	25%
CBGT-1	1300	2600	CORR-2	375	750	CL2	30%
CBGT-1	1300	2600	CORR-3	563	1125	CL3	35%
CBGT-2	3200	6400	CORR-1	188	375	CL4	55%
CBGT-2	3200	6400	CORR-2	375	750	CL5	60%
CBGT-2	3200	6400	CORR-3	563	1125	CL6	65%
CBGT-3	5200	10400	CORR-1	188	375	CL7	85%
CBGT-3	5200	10400	CORR-2	375	750	CL8	90%
CBGT-3	5200	10400	CORR-3	563	1125	CL9	95%

5.5 FULL FACTORIAL VERSUS FRACTIONAL FACTORIAL EXPIEREMENT

5.5.1 FULL FACTORIAL EXPERIMENT DESIGN

A full factorial experiment takes in consideration all possible combinations of experiment factor levels in the set of factor configurations. Therefore, with reference to TABLE 5-1 which contains all factor levels, the number of unique combinations of factor levels sums up to 486 unique factor configuration where each is performed for each EPC network type and each simulation scenario. The result of a full factorial experiment is the quantification of contribution to response variable by: each single factor, second order interaction of factors, third order interaction of factors, and so forth. The advantage of this design is the availability of all information possible from an experiment and simplicity of approach. The disadvantage of this approach is that all unique combinations do not actually add extra information extracted from an experiment when the experiment contains non-interacting factors; a non-interacting factors are factors that do not depend on other factors level in their contribution and effect to the response variable, this indicates that some experiment factor combinations results can be numerically inferred from other combinations without the need to perform these experiments. Therefore it is neither an efficient use of resource nor the best engineering design to use full factorial experiment design in this case. In the following section, a fractional factorial design is explained and elaborated based on physical contribution of each factor and the interactions among them, which will provide adequate information for performance evaluation of architecture under test However, a full factorial design is performed in this study for a subset of simulation scenarios to support justifications and reasoning behind a fractional factorial experiment design and interactions among factors; the result for the full factorial design is reported in chapter 6.

5.5.2 FRACTIONAL FACTORIAL EXPERIMENT DESIGN

Interacting factors means factors that have a different effect on performance metrics based on other factors levels; and two factors are considered as none interacting factors when a factor have the same contribution to the response regardless of the other factor level. The set of combinations should detect most patterns of variations in metrics due to factors and their interactions. Using this information a fractional factorial design is sufficient when only interacting factors are tested using all of their respective combinations, and a full factorial design is not required to assess the effect of experiment factors interaction within different combinations. In addition, it is required to include one combination of factors that lead to worst case of performance metrics; this is represented by highest affecting level for each factor in this work. The design of this approach is explained in following paragraphs and the results for this design are reported in chapter 7.

We identified factors that affect performance metrics measured as shown in TABLE 5-1, and based on the physical basis of each factor we design the set of experiment combinations to discover interacting factors with logical justifications. The factors interactions are shown in Table 5-4, the logic in this analysis is as follows:

- Factor DBGT is physically independent from DEL, GWPT, and MMECAP factors, and the variation in DBGT does not in any logical way change the effect of these factors. Whereas factors CBGT and CORR might be affected by low and high DBGT levels, and

DBGT factor could create congestion points based on its level and CBGT and CORR factors.

- Factors CBGT and CORR control directly interacts with GWPT and MME-CAP factors, as the rate of requests arrival determines the load on each EPC entity based on processing time dictated by CAP factors. Furthermore CBGT and CORR rates are controlled by MME-CAP factor in these experiments.
- DEL factor contributes to E2ED metric only, it creates no effect on CPU-Util and BW-Util. DEL factor levels is physically independent from all other factors, that is cannot produce any interactions with other factors on E2ED metric beside its own direct effect.

Table 5-4: FACTORS INTERACTIONS IN EXPERIEMENT

Factors interactions	DBGT	CBGT	CORR	GWPT	MME-CAP	DEL
DBGT	-	Y	Y	-	-	-
CPBGT	Y	-	Y	Y	Y	-
CORR	Y	Y	-	Y	Y	-
GWPT	-	Y	Y	-	Y	-
MME-CAP	-	Y	Y	Y	-	-
DEL	-	-	-	-	-	-

Table 5-4 is used to derive sub-combinations that will be tested as a part of the fractional factorial experiment design, and these resultant sub-combinations will be used for detection of patterns of variations. Factors that are not part of combination factors are configured at their medium level, and in all combinations CBGT and CORR are varied to cover the spectrum of MME control load capacity. Table 5-5 shows these sub-combinations, for example, combination-2 and 22 is set of experiments where GWPT is varied between 10, 75, and 150 micro-seconds while other non-interacting factors, i.e.

DBGT and DEL factors are configured to their medium level only. Combination-0 is shown to indicate the average factor combination, while combination-4 is shown to indicate and measure the worst performing factor combination.

To simplify reporting experiments factor variation procedure, in each combination 3 factors shall be varied: CBGT, CORR and a third factor. The third factor shall be named the “fractional factor”, which is a term shall be used hereinafter for this factor being varied in its respective combination, besides CBGT and CORR. For example, combination 3 have DEL factor as the fraction factor, whereas combination 22 has GWPT as the fractional factor. The first set of combinations, combination 1, 2, and 3, have MME-CAP configured to 30 MB/sec, whereas the second combinations set, combinations 22 and 32, have MME-CAP configured to 30 MB/sec. The factors configuration in each combination is clearly defined in the table.

Table 5-5: EXPERIMENTS FACTORS COMBINATIONS

FACTOR CONFIGURATIONS	DBGT	DEL	MME-CAP	GWPT	FRACTINOAL FACTOR
COMBINATION-0	50%	0.5ms	30 Mbps	10usec	Average configuration
COMBINATION-1	**20%, 50%, 80%**	0.5ms	30 Mbps	10usec	DBGT
COMBINATION-2	50%	**0.5 msec**	30 Mbps	**10usec, 75usec, 150usec**	GW-CPU
COMBINATION-3	50%	**0.1ms, 0.5ms, 1.0ms**	30 Mbps	10usec	DEL
COMBINATION-22	50%	0.5ms	60 Mbps	**10usec, 75usec, 150usec**	GW-CPU

COMBINATION-32	50%	**0.1ms, 0.5ms, 1.0ms**	60 Mbps	10usec	DEL
COMBINATION-4	**80%**	**1.0ms**	**30 Mbps**	**75usec**	Worst configuration

5.6 SIMULATION NETWORK SETUP AND SCENARIOS CONFIGURATIONS

In order to further compare and simulate the effect of architectures under test, we assume two locations for EPC nodes, a core location and a remote, i.e. local, location. Figure 5-1 depicts the network deployment setup for the simulation scenarios. The core location is considered where all subscribers' information and policy entities for EPC network are located. The core location is considered as location for mobile operator services such as VOIP, SMS, MMS and others, whereas the local location is considered as a local offload datacenter which serves as an internet breakpoint, i.e. access point, for the operator that is closer to users. The MME is always located in core location, which is the widely adopted network deployment. The SGW and PGW are deployed in both locations core and local as shown in Figure 5-1. The various combinations of SGW and PGW locations for participating in the control operation under test are considered to showcase all possible actions; the scenario configurations for registration and S1-based handover mobility procedures are shown in Table 5-6 and Table 5-7. In the case of S1-based handover, the UE in the handover procedure changes from old SGW to new SGW which either could be local or core center located. As for the PGW, it could be as well local or core center located.

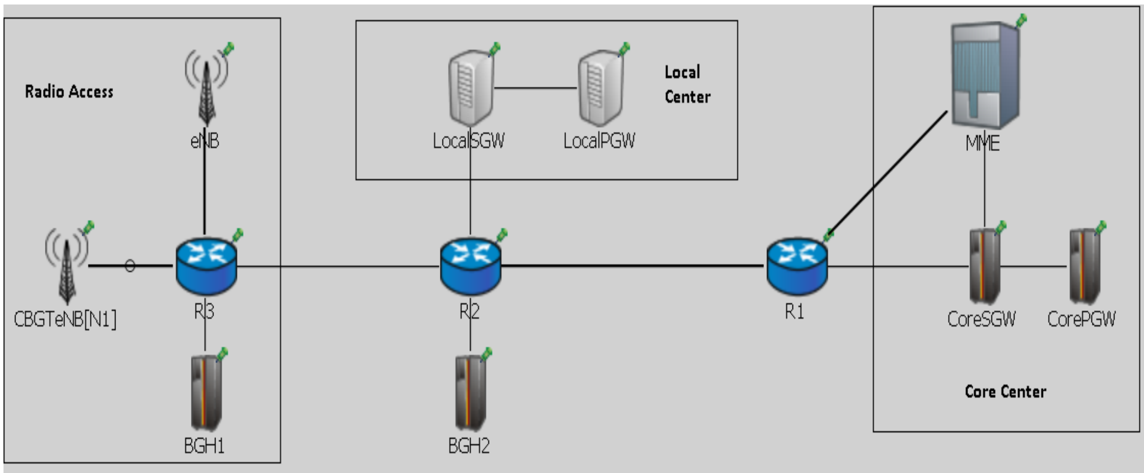


Figure 5-1: SIMULATED NETWORK SETUP

Table 5-6: 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 5-7: 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

Then by summarizing the complete set of tests, the collection of all simulation scenarios is: 3 EPC network architectures (Conventional, SCTP-based SDN, and UDP-based SDN), 2 control operations (initial attachment and S1-based handover), control operation scenario configurations (4 scenario each), and 10 replication per run, as shown in TABLE **5-8**.

TABLE 5-8: SET OF SIMULATION TESTS

Parameter Type	**# types**	**Parameter Levels**
Control plane operations	2	Initial attachment; S1-based handover
EPC Network Type	3	CONV-EPC; SDN-SCTP EPC; SDN-UDP EPC
Scenario configuration	4	**refer Table **5-6** and Table 5-7
Replication per run	10	Replicate runs

All links data rate is set at 10Gbps which is a reasonable carrier grade minimum data rate for an individual link. All propagation delay of links within location of EPC nodes are set to 0.1 micro second which is equal to 100 meter link physical length at 10Gbps data rate. The propagation delay of links connecting R3-R2 and R2-R1 is configured based on "DEL" parameter which reflects the distance of eNBs, local location, and core location from each other. The "DEL" parameter levels are chosen to be 0.1ms, 0.5ms, and 1.0ms, these values cover wide range of distance in the backhaul network. The sum of total propagation delays from core to eNB shall not exceed delay budget as reported in [35], which restricts backhaul delay between 1 milli-second and 15 milli-seconds. That will ensure our study parameters conformant to realistic deployment environments.

Node eNB is responsible for initiating control operation under test, i.e. initial attach and S1-based handover procedures. The rate of generating requests is controlled by the parameter CORR which is varied from 5% to 15% of total MME processing capacity; that reflects a reasonable and realistic level of operational conditions in deployed EPC networks.

Application servers (BGH1, BGH2) are used to generate data plane background traffic from backhaul towards local and core locations. Each server generate continuous UDP traffic packets with exponential inter-arrival time that leads to average throughput required for the configured DBGT level of link data rate, i.e. 20%,50%, and 80% of 10Gbps. BGH1 generates traffic towards the Core PGW, while BGH2 generates traffic towards the local PGW, as shown in Figure 5-2; this set up is made to ensure exact loading of all data links as per DBGT configurations. Each destination will echo back the same traffic to induce similar DBGT loading level on the reverse direction of data link; hence they are full duplex.

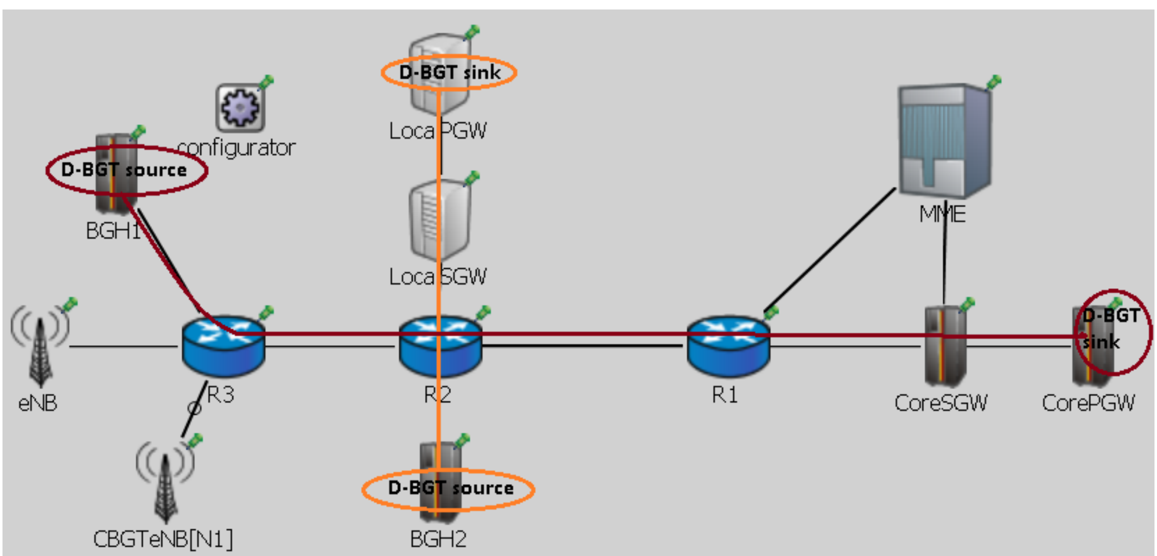


Figure 5-2: Data plane background traffic generation

Nodes named “CBGTeNB” are responsible for all control plane background traffic. The rate of control request arrival is controlled by CBGT parameter which is varied as percentage of MME processing capacity. The rate is configured on each simulation run to reflect the following levels of MME processing capacity: 20%, 50%, 80% load. However, as mentioned previously in section 5.4, CBGT operations are user initiated service request operations which are similar to registration operations and include two gateway nodes; i.e. a PGW and a SGW. All CBGT operations are configured to perform on gateways located in same center, i.e. either Core-SGW and Core-PGW or Local-SGW and Local-PGW; and CBGT operations are distributed evenly between core and local centers.

5.7 ASSUMPTIONS AND DESIGN DESCRIPTION

In this simulation model, the following engineering choices and assumptions are made:

- No erroneous packets occur. All transmitted packets are received correctly and no retransmissions are required.

- The control plane operations under test are as described in section 4.3.

- All control plane operations result in success. Each EPC node is considered to succeed in performing required control plane requests.

- Data plane processing capacity is of no concern to this study as control plane operations are the primary concern. Thus it is neglected.

- Each EPC node has finite processing capacity which is reflected in time required to complete serving requests arriving other EPC nodes. In this study, processing capacity of MME is based on a single queue-multi server model which is reasonably conformant with modern computing platforms and serving time of requests is linearly proportional to request application packet byte size based on the basic notion that time spent on each memory access and results computation is linearly proportional to output's size. The gateways processing time per request is a fixed processing time based on information from literature [36].

- Intermediate nodes, i.e. backhaul routers, are assumed to provide at least line rate processing capacity of forwarded packets, therefore processing delay in these nodes is negligible to accommodate large link capacity without imposing any bottlenecks due to extreme link loading.

- Each control request affects only one bearer context for UE. Although a UE might have several bearer contexts for different services, we limit the number of bearers to only one in all operations.

- SCTP and UDP protocol stack in all network devices are assumed to have no processing resource requirement, i.e. no processing delays due to SCTP and UDP protocol stacks. This assumption is made by the fact that benchmarking resource utilization of protocol stacks is not available and out of the scope of this study.

5.7.1 CONFIDENCE INTERVAL

To gain confidence in this study performance results, the experiments are replicated several times using different seeds for the random number generators. This would guarantee reliability in the collected results. The measured performance values are used to compute sample mean of the metric and a confidence interval for computed sample mean. The chosen number of replication is 10, which is clearly less than 30, thus the proper confidence interval computation would be using t-test (student test) [40]. A confidence interval of 90% is used for confidence interval computation which would result in interval less than 10% of the sample mean value.

The formula used for calculation of the confidence interval is:

$$\bar{X} \pm \ t_{\frac{\alpha}{2},N-1} \frac{\sigma}{\sqrt{N}}$$

where $\bar{X}$ is the computed mean, N is number of replications (equal to 10 in this study), σ is the standard deviation of replications, and $t_{\frac{\alpha}{2},N-1}$is the value of the $t-distribution$ with $\mathrm{N}-1$ degrees of freedom and α is the significance level.

5.8 CONTROL PLANE MESSAGE FORMAT AND SIZES

In this section, we present the message format of control plane messages used to complete operations under study. This information are obtained from related 3GPP standard documents in [2][3]. There are two message kinds used in control operations under test: GTP-C and S1-AP messages.

Messages exchanged amongst MME, SGW and PGW are defined in the GTP-C protocol. GTP-C messages consist of GTP-C header and followed by zero or more information elements (IE) as sown in Figure 5-3. The GTP-C message header consists of a 12 octet header. The information elements followed by the header are encapsulated as

Type-length-instance-value encoding. Each IE encoded as type-length-instance-value (TLIV) format and have a 4 octet header due to the encoding, and in some cases nested encoding of IE is possible. Each message contains a list of IEs based on their purpose.

The S1AP protocol provides signaling interface between the EPC and E-UTRAN. It is responsible for setting up, maintenance and release of radio access bearers at the E-TRAN. Similar to GTP-C, S1AP messages are formatted in IEs and each message contains a list IEs based on their purpose.

Table Figure 5-3 shows the complete byte size of GTP-C messages, which are calculated based on accumulation of all IEs sizes and GTP-C header; these are used in this work. The information about each IE and its size is detailed in appendix A and is extracted from 3GPP documents that standardize each message format [2][3].

OCTETS	8	7	6	5	4	3	2	1
1 to m	GTP-C HEADER							
m+1 to n	Zero or more Information Elements (IEs)							

Figure 5-3: GTP-C message format

Table 5-9: EPC control messages sizes [2][3]

Protocol Type	Message Type	Size (Bytes) CONV	Size (Bytes) SDN
GTP-C	CREATE SESSION REQUEST - S11	323	356
GTP-C	CREATE SESSION REQUEST – S5	313	345
GTP-C	CREATE SESSION RESPONSE - S11	187	175
GTP-C	CREATE SESSION RESPONSE - S11	182	177
GTP-C	MODIFY BEARER REQUEST - S11	155	177
GTP-C	MODIFY BEARER REQUEST – S5	152	163
GTP-C	MODIFY BEARER RESPONSE - S11	124	87
GTP-C	MODIFY BEARER RESPONSE – S5	108	97
GTP-C	DELETE SESSION REQUEST	102	102
GTP-C	DELETE SESSION RESPONSE	32	32
GTP-C	DELETE IDFT REQUEST -	16	16
GTP-C	DELETE IDFT RESPONSE	22	22
GTP-C	CREATE IDFT REQUEST	109	109
GTP-C	CREATE IDFT RESPONSE	98	98
S1-AP	HANDOVER REQUIRED	200	200
S1-AP	HANDOVER REQUEST	331	331
S1-AP	HANDOVER REQUEST ACKNOWLEDGE	175	175
S1-AP	HANDOVER NOTIFY	85	85
S1-AP	HANDOVER COMMAND	162	162
S1-AP	ENB STATUS TRANSFER	56	56

CHAPTER 6

FULL FACTORIAL DESIGN SIMULATION RESULTS

6.1 INTRODUCTION

In the experiment design in section 5.5, a full factorial and fractional factorial design are explained; a fractional factorial experiment design is justified to be sufficient for quantification of single factors and their interactions contribution to response variable, and we explain that full factorial experiment performs extra sets of experiments that do not necessarily add information that we cannot infer from the fractional factor design. The full factorial design is explained to include redundant experiment sets that comprise of factor combinations of non-interacting factors; the results for these combinations can be easily numerically inferred from the fractional factor design. In this chapter, a full factorial experiment design is performed and mathematical techniques are used to show and support the justification of a fractional factor design. Full factorial experiment is performed on subset of simulation scenarios for both SDN-EPC and CONV-EPC networks, and a mathematical formulation is derived using analysis of variation (ANOVA) technique, that will provide quantification of the contribution of each control factor in the response outcome and support the statement of interacting factors and their role in varying the response, it will also verify the derived interaction factors table shown in Table 5-4.

6.2 FULL FACTORIAL CONFIGURATION

In this section, a full factorial experiment is performed for simulation scenarios: Mob-D and Reg-2 as shown in TABLE 6-1; the set of experiment factors and setups for each scenario are shown in TABLE 6-2. The E2ED results for these experiments are then processed using analysis of variation (ANOVA) technique and results are used for explanation of fractional factorial design.

TABLE 6-1 shows experiment scenarios that undergo full factorial experiment; Mob-D and Reg-2 are chosen for full factorial experimentation in each EPC network type. Other simulation scenarios, Mob-A, B, C, Reg-1, etc., are not required for experimentation for a full factorial design, hence a simulation scenario represent different configurations of distances between EPC gateways, for example Mob-D differs from Mob-B is anchor PGW located in core center in Mob-D whereas Mob-B have it in local center, therefore the distance between anchor PGW and other EPC entities are varied, this leads to variation of effect of DEL factor at the same configuration. However, the factor configurations in full factorial design already include different levels of DEL factor and therefore the effect of DEL is already accounted for in the design. The inclusion of different simulation scenario determines the portion of contribution of this factor to the response variable but does not control the interaction among different factors, therefore it is considered irrelevant to the objective of the full factorial design in this chapter which is to quantify the contribution and effect of factor interactions that support the fractional factorial design.

TABLE 6-2 shows the full factorial experiment design factor combinations, it shows that for a single scenario a 486 unique combination are required. This number of combinations shows the large amount of different configurations used in a full factorial experiment, and also indicates that when using a fractional factorial experiment design large savings in time and resources are gained due to less number of combinations.

TABLE 6-1: FULL FACTORIAL EXPERIMENT UNDER TEST

Simulation Scenario	**# types**	**Parameter Levels**
Mob-D	1	Initial attachment
Reg-2	1	S1-based handover
EPC Network Type	2	CONV-EPC; SDN-SCTP EPC

TABLE 6-2: FULL FACTORIAL EXPERIMENT COMBINATIONS

	Factor Type	**#levels**	**Parameter Levels**
1	DEL	3	0.1, 0.5, 1.0 msec
2	DBGT	3	{ 20%; 50%; 80%} of backhaul link rate
3	MMECAP	2	{30,60} MByte/sec
4	GWPT	3	10, 75, 150 usec
5	CBGT & CORR	9	Refer Table 5-2 and Table 5-3
	Total combinations	486	Unique experiment factor combinations
	Replications	10	Replicates of same factor combination
	Total runs	4860	

6.2.1 ANALYSIS OF VARIARTION - ANOVA

In this section, a brief description of ANOVA study [41] is presented which provides insight about mathematical formulation used to quantify each experiment control factor and their contribution to response variable.

ANOVA provides a statistical test of whether the means of different factor combinations are equal or not equal. The ANOVA model computes: the mean response due to variation of all factor levels, this is termed grand mean (μ), the effect of each factor level variation, these are termed main effects coefficients, and the effect of factors interaction on response variable, these are termed factor interactions coefficients. From these coefficients, for main effects and factors interaction, the percentage of contribution of each of them is available. Equation (1) shows formulation of ANOVA model [41] for a two factor (A,B) experiment with many replications, where factor A has “a” levels, factor B has “b” levels, and “r” replications. The model computes the amount of change at each factor level change relative to grand mean, therefore it is available from these coefficients the exact contribution of each level separately. The model also provides quantification of the contribution of each effect using sums of squares of the factors coefficients as shown in equations (2) and (3). Although the model can be easily extended to 4 factors and more, we suffice by the explained model below.

The interpretation of ANOVA study results is that when factor levels have large contribution to change in response variable then the coefficient will show high value, and the contribution of the factor will show high value. When the coefficient are small that indicates the importance of the corresponding factor is limited. If it is extremely small that indicates that factor effect is irrelevant to response variable.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \text{ ... (1) where}$$

y_{ijk} is the response value at ith level of factor A, jth level of factor B, and kth replication

μ is the grand mean; average of all results obtained

α_i is the effect of factor A at ith level

β_j is the effect of factor B at jth level

γ_{ij} is the effect of interaction between factor A and B at A's ith level and B's jth level

e_{ijk} is the experimental errors that were not attributed to any factor effect

$$SSA = br\sum_j \alpha_j^2 \text{ ... } (2.a) \qquad SSA = r\sum_{ij} \gamma_{ij}^2 \text{ ... } (2.c)$$

$$SSB = ar\sum_i \beta_i^2 \text{ ... } (2.b) \qquad SSE = \sum_{ijk} e_{ijk}^2 \text{ ... } (2.d)$$

Factor A CONTRBUTION% = 100*SSA/SST…(3.a)

Factor B CONTRBUTION% = 100*SSB/SST…(3.b)

Factor A and B interaction CONTRBUTION% = 100*SSAB/SST…(3.c)

Error CONTRBUTION% = 100*SSE/SST …(3.d)

Applying this formulation to the full factorial experiment design, we have:

- The following four factors: DEL, DBGT, MMECAP, and GWPT.
  - DEL have 3 levels with "i = 1, 2, and 3"
  - DBGT have 3 levels with "j = 1, 2, and 3"
  - MMECAP have 2 levels with "k = 1, 2, and 3"
  - GWPT have 3 levels with "l = 1, 2, and 3"
- There are six first order interaction terms (interaction of two factors), these are: (1) DEL & DBGT, (2) DEL & MMECAP, (3) DEL & GWPT, (4) DBGT & MMECAP, (5) DBGT & GWPT, (6) MMECAP & GWPT.

- There are four second order interaction terms (interaction of three factors), these are: (1) DEL & DBGT & MMECAP, (2) DEL & DBGT & GWPT, (4) DEL & MMECAP & GWPT, (4) DBGT & MMECAP & GWPT.
- There is only one third order interaction term (four factors interaction), this is DEL & DBGT & MMECAP & GWPT.
- Number of replication is 10 with "r = 1, 2, …10"
- The response variable "y" is formulated using these factors and their interactions as follows:

$$
\begin{aligned}
y_{ijk} = {} & \mu + DEL_i + DBGT_j + MMECAP_k + GWCAP_l + + e_{ijkl} \\
& +(DEL \,\&\, DBGT)_{ij} \\
& +(DEL \,\&\, MMECAP)_{ik} \\
& +(DEL \,\&\, GWCAP)_{il} \\
& +(DBGT \,\&\, MMECAP)_{jk} \\
& +(DBGT \,\&\, GWCAP)_{jl} \\
& +(MMECAP \,\&\, GWCAP)_{kl} \\
& +(DEL \,\&\, DBGT \,\&\, MMECAP)_{ijk} \\
& +(DEL \,\&\, DBGT \,\&\, GWCAP)_{ijl} \\
& +(DEL \,\&\, MMECAP \,\&\, GWCAP)_{ikl} \\
& +(DBGT \,\&\, MMECAP \,\&\, GWCAP)_{jkl} \\
& +(DEL \,\&\, DBGT \,\&\, MMECAP \,\&\, GWCAP)_{ijkl}
\end{aligned}
$$

Each term will be used to measure the contribution of each factor and factors interactions according to formulation presented.

To determine if a factor has a significant effect on the response, statisticians compare its contribution to the variation with that of the errors. If unexplained variation due to errors is high the factor explaining a large fraction of the variation may turn out to be statistically insignificant. The statistical procedure to analyze the significance of various factors is called Analysis Of Variance (ANOVA). This is performed by computing an F-test result, which uses sum of squares for factor coefficients and sum of squares of the errors. The ratio of these two sums provides an F-computed value. This value is compared with an F-table value extracted from tabulated values that are part of F-distribution; these values depend on significance level, i.e. confidence interval, and number of degrees of freedom. If F-computed value is larger than that for F-table, then the factor effect on explaining variation is considered statistically significant. Otherwise, the factor effect is considered statistically insignificant relative to unexplained fraction of results.

6.2.2 FULL FACTORIAL EXPERIMENT RESULTS

The result of full factorial experiment of the selected scenarios will be used in this section to detect interaction between DEL & DBGT factors, DEL & MMECAP factors, DEL & GWPT factors, BGT & MMECAP factors, DBGT & GWPT factors, and MMECAP & GWPT factors by using ANOVA technique. If the contribution of these combination of factors are quantified to be 0 then that concludes there is interaction between them, and therefore conclude the viability and correctness of the fractional factorial experiment design introduced in this study.

6.2.2.1 S1-HANDOVER – Mob-D scenario

Table 6-3 and Table 6-4 show E2ED results for full factorial experiment for Mob-D in CONV-EPC and SDN-SCTP EPC respectively. The results are presented at each control load level from CL1 to CL9, since the fractional factorial experiment design introduced in chapter 5 includes control load in the full factorial experiments, the control load factor is exempted from the ANOVA study and the relationship between its effect and other factors will be tackled in subsequent sections. The results shown in these tables (raw results) show the following:

- Effect of main factors:
  - DEL factor is observed to cause major variation in E2ED; with each increase in DEL level, the overall E2ED is increased significantly.
  - DBGT factor variation has almost no perceptible impact on E2ED, that is indicated by extremely small variation on E2ED with DBGT level increase
  - MMECAP factor effect is observed to cause decrease in E2ED results, the decrease is observed to be constant at any other factor level which indicate the non-interaction of MMECAP factor from other factors effect; for example, the decrease in E2ED due to MMECAP level variation at DEL-1 and DEL-2 are the same.

GWPT factor effect is observed to be minimal relative to overall E2ED, the increase from GWPT-1 to GWPT-2 or GWPT-2 to GWPT-3 is observed to induce approximately 0.5 msec in E2ED. This amount of E2ED increase deems less important relative to overall E2ED when DEL factor level is high, i.e. DEL-3. However, it is observed for configuration of CL4 to CL9, MMECAP-2, and GWPT-3 in CONV-EPC network there is an abrupt increase in E2ED, this sudden increase, as shall be analyzed in subsequent section, is attributed to saturation of resources at the gateways and congestion occurrence. This shows the limited resource capacity of GWPT-3 compared to GWPT-1 and 2 for CONV-EPC, whereas SDN-EPC was able to accommodate same requests with no degradation of service as occurred in CONV-EPC network.

Table 6-3: E2ED results for Mob-D in CONV-EPC

			DEL-1			DEL2			DEL3		
			BGT-1	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**
CL1	**MME-1**	**GWPT-1**	7.39	7.40	7.41	17.70	17.71	17.73	30.59	30.60	30.61
		GWPT-2	7.79	7.81	7.82	18.11	18.12	18.13	31.00	31.01	31.02
		GWPT-3	8.32	8.33	8.35	18.70	18.71	18.73	31.52	31.54	31.54
	MME-2	**GWPT-1**	5.05	5.06	5.08	15.41	15.42	15.43	28.35	28.36	28.38
		GWPT-2	5.49	5.50	5.51	15.84	15.85	15.87	28.90	28.91	28.94
		GWPT-3	6.17	6.19	6.19	16.54	16.55	16.56	29.62	29.63	29.65
CL2	**MME-1**	**GWPT-1**	7.42	7.43	7.44	17.77	17.79	17.80	30.72	30.73	30.75
		GWPT-2	7.87	7.89	7.90	18.23	18.24	18.25	31.13	31.14	31.16
		GWPT-3	8.36	8.38	8.39	18.85	18.85	18.87	31.76	31.78	31.79
	MME-2	**GWPT-1**	5.06	5.07	5.09	15.44	15.45	15.46	28.41	28.42	28.44
		GWPT-2	5.54	5.55	5.56	15.93	15.95	15.96	28.96	28.98	29.00
		GWPT-3	6.29	6.31	6.32	16.72	16.73	16.73	29.74	29.75	29.76
CL3	**MME-1**	**GWPT-1**	7.43	7.44	7.46	17.80	17.81	17.83	30.76	30.78	30.79
		GWPT-2	7.84	7.85	7.87	18.21	18.23	18.24	31.18	31.20	31.22
		GWPT-3	8.45	8.46	8.47	18.80	18.83	18.84	31.72	31.74	31.75
	MME-2	**GWPT-1**	5.06	5.08	5.10	15.45	15.46	15.48	28.45	28.46	28.48
		GWPT-2	5.55	5.56	5.57	15.90	15.91	15.93	29.08	29.10	29.11
		GWPT-3	6.63	6.65	6.65	17.10	17.12	17.13	30.07	30.06	30.05
CL4	**MME-1**	**GWPT-1**	7.41	7.43	7.44	17.72	17.74	17.76	30.62	30.64	30.64
		GWPT-2	7.86	7.87	7.89	18.18	18.20	18.20	31.07	31.08	31.09
		GWPT-3	8.66	8.66	8.65	19.03	19.05	19.06	31.86	31.88	31.87
	MME-2	**GWPT-1**	5.06	5.08	5.09	15.42	15.43	15.45	28.39	28.40	28.41
		GWPT-2	5.65	5.66	5.68	16.02	16.03	16.05	29.09	29.09	29.11
		GWPT-3	258.58	266.28	266.72	263.92	286.21	239.80	293.76	280.62	267.96
CL5	**MME-1**	**GWPT-1**	7.46	7.48	7.49	17.81	17.83	17.84	30.76	30.77	30.78
		GWPT-2	7.94	7.96	7.97	18.30	18.32	18.34	31.22	31.24	31.25
		GWPT-3	8.74	8.78	8.78	19.26	19.27	19.25	32.17	32.16	32.18
	MME-2	**GWPT-1**	5.09	5.10	5.11	15.46	15.47	15.48	28.45	28.46	28.48
		GWPT-2	5.75	5.75	5.77	16.14	16.15	16.15	29.17	29.18	29.19
		GWPT-3	435.83	410.71	402.95	426.04	409.75	419.62	434.63	428.62	447.25
CL6	**MME-1**	**GWPT-1**	7.49	7.51	7.51	17.86	17.87	17.89	30.87	30.89	30.91
		GWPT-2	7.95	7.95	7.97	18.32	18.33	18.34	31.35	31.36	31.37
		GWPT-3	8.85	8.86	8.88	19.41	19.44	19.44	32.36	32.35	32.35
	MME-2	**GWPT-1**	5.10	5.12	5.13	15.51	15.53	15.54	28.66	28.67	28.69
		GWPT-2	5.78	5.80	5.81	16.21	16.24	16.25	29.31	29.32	29.34
		GWPT-3	517.40	518.82	526.60	533.64	533.44	509.10	513.06	521.26	551.73
CL7	**MME-1**	**GWPT-1**	7.88	7.90	7.90	18.23	18.24	18.28	31.14	31.15	31.19
		GWPT-2	8.37	8.36	8.40	18.71	18.75	18.73	31.59	31.63	31.65
		GWPT-3	10.60	10.50	10.52	20.98	20.97	21.06	33.89	33.90	33.80
	MME-2	**GWPT-1**	5.29	5.29	5.30	15.67	15.69	15.68	28.71	28.72	28.74
		GWPT-2	6.60	6.62	6.59	16.99	16.98	17.06	29.97	30.02	30.07
		GWPT-3	1271.35	1287.64	1217.25	1263.23	1160.64	1278.26	1147.21	1236.22	1238.43

CL8	**MME-1**	**GWPT-1**	8.13	8.17	8.14	18.58	18.55	18.56	31.51	31.55	31.47
		GWPT-2	8.68	8.64	8.67	19.07	19.06	19.12	31.93	31.96	31.99
		GWPT-3	11.71	11.80	11.69	22.51	22.60	22.29	35.24	35.20	35.07
	MME-2	**GWPT-1**	5.43	5.45	5.46	15.82	15.82	15.84	28.90	28.91	28.91
		GWPT-2	7.08	7.27	7.25	17.59	17.77	17.89	30.72	30.77	30.75
		GWPT-3	1273.34	1327.09	1213.97	1262.87	1191.18	1321.51	1244.09	1222.80	1274.98
CL9	**MME-1**	**GWPT-1**	8.51	8.62	8.57	18.91	19.07	19.02	32.19	32.25	32.19
		GWPT-2	9.04	9.03	9.05	19.52	19.54	19.58	32.69	32.74	32.80
		GWPT-3	16.43	15.79	15.15	26.47	26.39	25.46	38.66	42.18	39.96
	MME-2	**GWPT-1**	5.61	5.67	5.64	16.21	16.20	16.22	29.31	29.33	29.34
		GWPT-2	9.10	9.01	9.09	20.07	20.21	19.52	33.15	32.72	33.33
		GWPT-3	1221.18	1288.63	1306.94	1333	1351	1328	1225	1152	1330

Table 6-4: E2ED results for Mob-D in SDN-SCTP EPC

			DEL-1			**DEL2**			**DEL3**		
			BGT-1	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**
CL1	**MME-1**	**GWPT-1**	7.20	7.21	7.22	16.72	16.73	16.74	28.62	28.63	28.64
		GWPT-2	7.47	7.48	7.48	16.98	16.99	17.01	28.88	28.89	28.90
		GWPT-3	7.79	7.80	7.80	17.31	17.31	17.33	29.20	29.21	29.23
	MME-2	**GWPT-1**	4.84	4.85	4.86	14.40	14.41	14.42	26.35	26.36	26.37
		GWPT-2	5.12	5.13	5.14	14.68	14.69	14.70	26.64	26.65	26.66
		GWPT-3	5.47	5.49	5.49	15.04	15.05	15.05	27.06	27.07	27.09
CL2	**MME-1**	**GWPT-1**	7.23	7.24	7.25	16.79	16.80	16.81	28.74	28.75	28.76
		GWPT-2	7.50	7.51	7.52	17.10	17.11	17.12	29.06	29.08	29.08
		GWPT-3	7.84	7.86	7.87	17.46	17.47	17.48	29.46	29.48	29.49
	MME-2	**GWPT-1**	4.85	4.86	4.87	14.43	14.44	14.45	26.41	26.42	26.43
		GWPT-2	5.15	5.15	5.16	14.78	14.78	14.78	26.69	26.71	26.72
		GWPT-3	5.52	5.52	5.53	15.14	15.16	15.17	27.12	27.13	27.14
CL3	**MME-1**	**GWPT-1**	7.24	7.25	7.26	16.82	16.83	16.84	28.78	28.79	28.80
		GWPT-2	7.51	7.52	7.53	17.08	17.10	17.11	29.07	29.07	29.07
		GWPT-3	7.83	7.84	7.85	17.41	17.42	17.43	29.38	29.38	29.39
	MME-2	**GWPT-1**	4.86	4.87	4.88	14.44	14.46	14.47	26.46	26.47	26.48
		GWPT-2	5.17	5.19	5.20	14.72	14.73	14.74	26.85	26.86	26.87
		GWPT-3	5.53	5.54	5.55	15.10	15.11	15.12	27.28	27.28	27.29
CL4	**MME-1**	**GWPT-1**	7.30	7.30	7.32	16.84	16.86	16.86	28.71	28.74	28.74
		GWPT-2	7.56	7.57	7.58	17.11	17.11	17.12	28.99	28.99	29.01
		GWPT-3	7.91	7.92	7.94	17.44	17.46	17.47	29.33	29.35	29.36
	MME-2	**GWPT-1**	4.89	4.90	4.91	14.46	14.47	14.48	26.43	26.44	26.45
		GWPT-2	5.19	5.19	5.20	14.75	14.76	14.77	26.74	26.75	26.76
		GWPT-3	5.72	5.73	5.74	15.32	15.33	15.35	27.33	27.34	27.36
CL5	**MME-1**	**GWPT-1**	7.37	7.39	7.39	16.98	16.99	17.00	28.91	28.91	28.93
		GWPT-2	7.64	7.66	7.66	17.27	17.28	17.30	29.21	29.21	29.22
		GWPT-3	8.05	8.07	8.07	17.64	17.65	17.66	29.64	29.65	29.65
	MME-2	**GWPT-1**	4.92	4.94	4.95	14.52	14.54	14.55	26.50	26.51	26.52
		GWPT-2	5.24	5.24	5.25	14.86	14.87	14.88	26.81	26.82	26.83
		GWPT-3	5.87	5.89	5.89	15.59	15.60	15.60	27.42	27.43	27.44
CL6	**MME-1**	**GWPT-1**	7.41	7.43	7.44	17.07	17.08	17.08	29.02	29.03	29.03
		GWPT-2	7.69	7.71	7.71	17.33	17.35	17.34	29.28	29.29	29.30

		GWPT-3	8.06	8.07	8.08	17.67	17.69	17.69	29.62	29.64	29.64
	MME-2	**GWPT-1**	4.96	4.98	4.98	14.57	14.58	14.59	26.77	26.77	26.78
		GWPT-2	5.29	5.30	5.31	14.85	14.86	14.87	27.03	27.02	27.04
		GWPT-3	6.05	6.07	6.10	15.72	15.71	15.74	27.70	27.71	27.71
CL7	**MME-1**	**GWPT-1**	9.03	8.98	9.12	18.89	18.87	18.82	30.53	30.62	30.52
		GWPT-2	9.52	9.45	9.51	18.95	18.95	19.11	30.91	30.97	31.05
		GWPT-3	9.78	9.67	9.74	19.47	19.35	19.56	31.17	31.17	31.08
	MME-2	**GWPT-1**	5.73	5.87	5.82	15.45	15.43	15.43	27.47	27.45	27.47
		GWPT-2	6.07	6.04	6.07	15.75	15.81	15.78	27.80	27.78	27.75
		GWPT-3	7.67	7.57	7.51	17.07	17.12	17.09	29.38	29.21	29.34
CL8	**MME-1**	**GWPT-1**	10.94	10.75	10.77	20.92	20.61	20.66	33.23	32.85	32.56
		GWPT-2	11.60	10.95	12.48	21.10	21.01	20.71	32.88	32.85	32.90
		GWPT-3	11.69	11.23	11.63	21.49	21.28	20.90	33.22	33.45	33.69
	MME-2	**GWPT-1**	6.71	6.80	6.58	16.33	16.51	16.40	28.48	28.38	28.28
		GWPT-2	7.16	7.07	7.14	16.61	16.78	16.70	28.59	28.93	28.73
		GWPT-3	11.35	11.78	11.59	20.98	21.12	21.09	33.63	33.16	33.66
CL9	**MME-1**	**GWPT-1**	21.44	19.66	19.21	36.06	27.49	28.02	43.37	44.60	43.03
		GWPT-2	18.85	19.57	18.88	33.35	26.77	27.83	45.57	39.72	39.73
		GWPT-3	21.78	18.65	20.51	28.06	30.94	30.44	42.82	40.22	38.84
	MME-2	**GWPT-1**	10.70	11.26	10.58	22.29	21.92	20.31	33.68	31.90	34.78
		GWPT-2	10.96	12.34	10.81	21.94	21.70	20.79	33.63	33.29	33.56
		GWPT-3	141.63	142.31	134.60	136.63	131.13	146.81	167.97	157.65	162.96

Using results in tables above and ANOVA technique described earlier, we quantify the main factors effects and the interaction of factors along with their contribution to overall E2ED results at selected control load points CL2, CL5 and CL8 for CONV-EPC and SDN-SCTP EPC networks.

Table 6-5 and Table 6-6 show the ANOVA study results at CL2, the results assert the following:

- DEL factor have the major effect of variation of results, the coefficients of this factor is the highest of all other factors coefficients, the coefficients are shown to be -11.25 and -10.39 at DEL-1, and 12.11 and 11.18 at DEL-3 for CONV-EPC and SDN-SCTP respectively, these values are 10 times higher than other factors coefficients which indicate the dominance of this factor on other factors contribution. However, it is noticed that the variation in DEL levels leads to more increase in E2ED result for the case of CONV-EPC more than that for SDN-SCTP; this is indicated by higher factor coefficients for CONV-EPC than SDN-SCTP.

- The ANOVA study shows that in DEL contributes 98% of E2ED results which is a clear indication of factor dominance.
- DBGT coefficients show extremely small contribution to E2ED variation with almost 0% contribution to overall results.
- MMECAP effect is almost equal for CONV-EPC and SDN-SCTP with almost 1.1 msec variation of E2ED are attributed to MMECAP change. The percentage of MMECAP contribution to E2ED variation is around 1.5% in both network types.
- GWPT factor have least effect compared to DEL and MMECAP factors, it induces 0.54 msec and 0.35 msec variation in results for CONV-EPC and SDN-SCTP EPC respectively according to GWPT coefficients. Although CONV-EPC is observed to have higher impact due to GWPT level than SDN-SCTP EPC, it is extremely small relative to overall delay. The contribution of this factor is less than 0.24% to overall variations.
- Factors interactions: the ANOVA study shows that there is no interaction between any of the factors DEL, GWPT, MMECAP, and DBGT in any combination among them. The interaction between factors is indicated by values: DEL & BGT (SS), DEL & MMECAP (SS), DEL & GWPT (SS), BGT & MMECAP (SS), BGT & GWPT (SS), and MMECAP & GWPT (SS), these all have 0 value which indicates no interaction among any combination at this control load level. This result supports the fractional factorial experiment design explained in chapter 5, that no interaction

between these factors exist, and therefore a full factorial experiment is not required for these factors combinations. A fractional factor experiment that varies one of these factors and configures the others to average configuration would still capture all of the effects of these factors.

- The results show in F-test section that all F-computed values are larger than their respective F-table values. This indicates that the percentage of variation explained by each factor in the ANOVA model is statistically significant. These results support the robustness of this model.

ANOVA study for E2ED results for Mob-D

Table 6-5: CONV- EPC at CL2

Table 6-6: SDN-SCTP EPC at CL2

	MAIN FACTORS EFFECTS LEVEL (Table 6-5)			MAIN FACTORS EFFECTS LEVEL (Table 6-6)		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-11.25	-0.86	12.11	-10.39	-0.79	11.18
BGT	-0.01	0	0.01	-0.01	0	0.01
MMECAP	1.11	-1.11		1.17	-1.17	
GWPT	-0.54	-0.07	0.61	-0.33	-0.03	0.36

Component	**Square sum**	**Percent Variation**	**Degrees of freedom**	**Square Sum**	**Percent Variation**	**Degrees of freedom**
Response(SSY)	22555		54	19429		54
Mue (SS0)	17543		1	15147		1
y-Mue (SST)	5011	100	53	4281	100	53
Main effects	5011	100	7	4281.83	100	7
DEL (SS)	4932	98.42	2	4203.32	98.17	2
DBGT (SS)	0.01	0	2	0	0	2
MMECAP(SS)	66.86	1.33	1	74.29	1.73	1
GWPT (SS)	11.99	0.24	2	4.22	0.1	2
1st order	0.23	0	18	0.02	0	18
DEL &	0	0	4	0	0	4

DBGT (SS)							
DEL & MMECAP (SS)	0.02	0	2		0	0	2
DEL & GWPT (SS)	0.02	0	4		0.01	0	4
BGT & MMECAP (SS)	0	0	4		0	0	4
BGT & GWPT (SS)	0	0	2		0	0	2
MMECAP & GWPT (SS)	0.16	0	2		0	0	2
2nd order	0.01	0	20		0	0	20
DEL & MMECAP & GWPT (SS)	0	0	8		0	0	8
DEL & DBGT & MMECAP (SS)	0	0	4		0	0	4
DEL & DBGT & GWPT (SS)	0.01	0	4		0	0	4
DBGT & MMECAP & GWPT (SS)	0	0	4		0	0	4
3rd order							
All 4 factors	0	0	8		0	0	8
Totals	5011				4281.85		

	F-test study								
DEL	**F-test Computed**	19018941	**F-test table**	0.59		**F-test Computed**	1.03E+08	**F-test table**	0.59
DBGT		46243.35		0.59			103811.2		0.59
MMECAP		515627.2		0.66			3653765		0.66
GWPT		23.26		0.59			83.28		0.59

Figure 6-1 and Figure 6-2 show the quantile-quantile plot of normal quantile versus residual quantile. This plot is used to detect the distribution of errors in the ANOVA model. The Q-Q plot shows that errors are extremely small relative to the average response value. Most residual values are within -0.005 and 0.005 for both SDN-EPC and CONV-EPC networks. This indicates that the factor coefficients provide strong relation with response variable as computed by the ANOVA study. Although the residuals show a pattern in the Q-Q plot, their extreme small values relative to response variable values allows considering it unimportant.

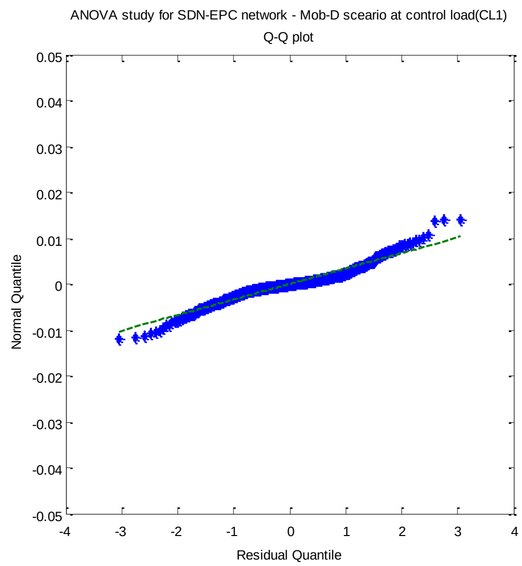


Figure 6-1: Q-Q plot of residuals for ANOVA study for SDN-EPC in Mob-D scenario at CL2

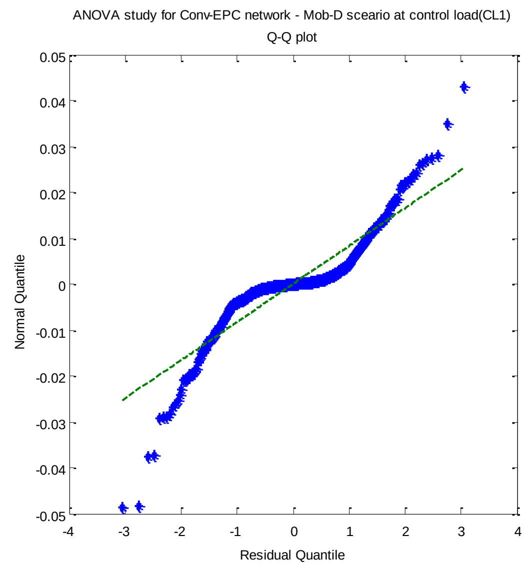


Figure 6-2: Q-Q plot of residuals for ANOVA study for CONV-EPC in Mob-D scenario at CL2

Table 6-7 and Table 6-8 show the ANOVA study results at CL5, the results assert the following:

- SDN-EPC have almost the same factor coefficients and contributions those at CL2
- CONV-EPC has major change of MMECAP and GWPT factors effects particularly at GWPT-3; it is observed that at GWPT-3 there is significant increase on GWPT coefficient, whereas GWPT-1 and 2 are almost identical. The contribution of GWPT and MMECAP has increased considerably to reach 40% and 20% respectively, and the interaction between them has reached almost 40%. However, relating these information with results shown in previous tables (raw results), we already identified that this behavior occurred due to extreme congestion at EPC gateways, therefore the

system is unstable and these results do not in fact represent interaction between GWPT and MMECAP, although the fractional factorial experiment design takes in consideration all GWPT and MMECAP factors combinations.

- The results show in F-test section that all F-computed values, except GWPT, are larger than their respective F-table values. This indicates that the percentage of variation explained by these factors in the ANOVA model is statistically significant. The exception for this is for GWPT factor in CONV-EPC. This indicates in the ANOVA model that the errors due to experiments have more contribution to variation than GWPT in CONV-EPC. Therefore, to test GWPT effect more elaborate experiment design that eliminates errors due to other factors and focuses on GWPT only. In other words, the effect of GWPT is masked by other factors errors.

The results also show that there is no interaction among DEL, DBGT and other factors, indicated by zero contribution of the corresponding interaction terms in the tables, which following results at CL2 supports the fractional factor experiment design.

ANOVA study for E2ED results for Mob-D

Table 6-7: CONV- EPC at CL5 | Table 6-8: SDN-SCTP EPC at CL5

	MAIN FACTORS EFFECTS LEVEL			MAIN FACTORS EFFECTS LEVEL		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-10.61	-1.62	12.24	-10.4	-0.77	11.17
BGT	1.36	-1.26	-0.1	-0.01	0	0.01
MMECAP	-66.56	66.56		1.17	-1.17	
GWPT	-68.35	-67.77	136.13	-0.38	-0.08	0.45

Component	Square sum	Percent Variation	Degrees of freedom	Square Sum	Percent Variation	Degrees of freedom

Response(SSY)	1638338		54	19744.9		54
Mue (SS0)	398200		1	15465.1		1
y-Mue (SST)	1240137	100	53	4279.79	100	53
Main effects	744414	60.03	7	4279.54	99.99	7
DEL (SS)	4769.7	0.38	2	4199.83	98.13	2
DBGT (SS)	62.25	0.01	2	0	0	2
MMECAP(SS)	239252	19.29	1	73.29	1.71	1
GWPT (SS)	500330	40.34	2	6.41	0.15	2
1st order	495136	39.93	18	0.23	0.01	18
DEL & DBGT (SS)	88.08	0.01	4	0	0	4
DEL & MMECAP (SS)	18.57	0	2	0	0	2
DEL & GWPT (SS)	35.1	0	4	0	0	4
BGT & MMECAP (SS)	125.93	0.01	4	0	0	4
BGT & GWPT (SS)	63	0.01	2	0	0	2
MMECAP & GWPT (SS)	494787.2	39.9	2	0.22	0.01	2
2nd order	427.6	0.03	20	0.02	0	20
DEL & MMECAP & GWPT (SS)	176.18	0.01	8	0	0	8
DEL & DBGT & MMECAP (SS)	88.37	0.01	4	0	0	4
DEL & DBGT & GWPT (SS)	37.33	0	4	0.02	0	4
DBGT & MMECAP & GWPT (SS)	125.72	0.01	4	0	0	4
3rd order						
All 4 factors	176.73	0.01	8	0	0	8
Totals	1240155			4279.79		

	F-test study								
DEL	**F-test Computed**	12.47	**F-test table**	0.59		**F-test Computed**	7870225	**F-test table**	0.59
DBGT		1308.44		0.59			12017.77		0.59
MMECAP		1251.36		0.66			274689.3		0.66
GWPT		0.16		0.59			6.14		0.59

Table 6-9 and Table 6-10 show the ANOVA study results at CL8, the results assert the findings observed in same raw results tables (Table 6-3 and Table 6-4) at CL8. In CONV-EPC GWPT and MMECAP factors are causing the means of the experiments to be large affected by their effect and they appear to be major contribution to E2ED, however it is already identified the cause of this matter, but it is observed their effect is more considerable than that at CL5.

In SDN-SCTP it is noticed that effect of GWPT and MMECAP is slightly increased than that at CL5 and CL8 but still minimal compared to DEL factor; DEL factor contribution is decreased by 4% to reach 94% contribution to E2ED overall results.

The results show in F-test section that not all factor effects pass the F-test. DBGT and MMECAP pass the significance test. GWPT also does not pass the F-test, for the same reason mentioned at CL5 that errors due other factors mask the effect of GWPT. This prevents the ANOVA model from providing confidence in GWPT effect. DEL factor pass the F-test in SDN-EPC, however, in CONV-EPC is fails the test due errors introduced from instability in the system. Therefore, even though DEL effect is confirmed at other control load points, the instability in the system introduces errors that reduce confidence in the results.

In terms of factors interactions it is still as before factors DEL and DBGT have no interactions with GWPT and MMECAP factors, and that supports the fractional factorial experiment design as other results have shown as well.

ANOVA study for E2ED results for Mob-D

Table 6-9: CONV- EPC at CL8 | Table 6-10: SDN-SCTP EPC at CL8

	MAIN FACTORS EFFECTS LEVEL (Table 6-9)			MAIN FACTORS EFFECTS LEVEL (Table 6-10)		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-7.36	-0.77	8.13	-10.45	-0.84	11.29
BGT	0.15	-1.99	1.84	0.03	-0.04	0.01
MMECAP	-205.33	205.33		1.44	-1.44	
GWPT	-208.07	-206.9	414.97	-1.03	-0.67	1.7

Component	**Square sum** (6-9)	**Percent Variation** (6-9)	**Degrees of freedom** (6-9)	**Square Sum** (6-10)	**Percent Variation** (6-10)	**Degrees of freedom** (6-10)
Response(SSY)	14306497		54	26884.3		54
Mue (SS0)	2761443		1	22363.5		1
y-Mue (SST)	11545054	100	53	4520.74	100	53
Main effects	6928300	60.01	7	4462.41	98.71	7
DEL (SS)	2174.88	0.02	2	4270.85	94.47	2
DBGT (SS)	132.71	0	2	0.06	0	2
MMECAP(SS)	2276554	19.72	1	112.07	2.48	1
GWPT (SS)	4649438	40.27	2	79.44	1.76	2
1st order	4602985	39.87	18	56.41	1.25	18
DEL & DBGT (SS)	2581.96	0.02	4	0.26	0.01	4
DEL & MMECAP (SS)	558.66	0	2	0	0	2
DEL & GWPT (SS)	1132.07	0.01	4	0.59	0.01	4
BGT & MMECAP (SS)	264.91	0	4	0.22	0	4
BGT & GWPT (SS)	136.83	0	2	0.21	0	2
MMECAP & GWPT (SS)	4597752	39.82	2	55.11	1.22	2
2nd order	9155.99	0.08	20	1.5	0.03	20
DEL & MMECAP & GWPT (SS)	5153.68	0.04	8	0.62	0.01	8
DEL & DBGT & MMECAP (SS)	2587.4	0.02	4	0.47	0.01	4
DEL & DBGT &	1141.87	0.01	4	0.29	0.01	4

GWPT (SS)							
DBGT & MMECAP & GWPT (SS)	273.03	0	4		0.13	0	4
3rd order							
All 4 factors	5171.39	0.04	8		0.42	0.01	8
Totals	11545612				4520.75		

	F-test study								
DEL	**F-test Computed**	0.21	**F-test table**	0.59		**F-test Computed**	1826.02	**F-test table**	0.59
DBGT		443.07		0.59			33.96		0.59
MMECAP		433.89		0.66			95.83		0.66
GWPT		0.01		0.59			0.02		0.59

6.2.2.2 REGISTRATION PROCEDURE – Reg-2 scenario

Table 6-11 and Table 6-12 show E2ED results for full factorial experiment for Reg-2 in CONV-EPC and SDN-SCTP EPC respectively. The results are presented at each control load level from CL1 to CL9, since the fractional factorial experiment design introduced in chapter 5 uses full factorial experiments with regard of control load, the control load factor is exempted from the ANOVA study and the relationship between its effect and other factors will be tackled in subsequent sections. The results shown in these tables show the following which are almost identical to that in S1-handover procedure:

- Effect of main factors:
    - DEL factor is observed to cause major variation in E2ED; with each increase in DEL level, the overall E2ED is increased significantly.
    - DBGT factor variation has almost no perceptible impact on E2ED, that is indicated by extremely small variation on E2ED with DBGT level increase

- MMECAP factor effect is observed to cause decrease in E2ED results, the decrease is observed to be constant at any other factor level which indicate the non-interaction of MMECAP factor from other factors effect; for example, the decrease in E2ED due to MMECAP level variation at DEL-1 and DEL-2 are the same.
- GWPT factor effect is observed to be minimal relative to overall E2ED, the increase from GWPT-1 to GWPT-2 or GWPT-2 to GWPT-3 is observed to induce approximately 0.2 and 0.35 msec in E2ED for CONV-EPC and SDN-SCTP EPC respectively. This amount of E2ED increase deems less important relative to overall E2ED when DEL factor level is high, i.e. DEL-3. However, it is observed for configuration of CL4 to CL9, MMECAP-2, and GWPT-3 in CONV-EPC network there is an abrupt increase in E2ED, this sudden increase, as shall be analyzed in subsequent section, is attributed to saturation of resources at the gateways and congestion occurrence. This shows the limited resource capacity of GWPT-3 compared to GWPT-1 and 2 for CONV-EPC, whereas SDN-EPC was able to accommodate same requests with no degradation of service as occurred in CONV-EPC network.

Table 6-11: E2ED results for Reg-2 in CONV-EPC at CL1

			DEL-1			DEL2			DEL3		
			BGT-1	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**
CL1	**MME-1**	**GWPT-1**	2.41	2.41	2.42	5.59	5.60	5.60	9.58	9.59	9.59
		GWPT-2	2.61	2.61	2.62	5.80	5.80	5.81	9.78	9.79	9.79
		GWPT-3	2.87	2.88	2.88	6.06	6.06	6.07	10.04	10.05	10.05
	MME-2	**GWPT-1**	1.63	1.63	1.64	4.82	4.83	4.83	8.81	8.82	8.82
		GWPT-2	1.84	1.85	1.85	5.04	5.04	5.05	9.03	9.03	9.04
		GWPT-3	2.17	2.17	2.18	5.36	5.37	5.37	9.39	9.39	9.41
CL2	**MME-1**	**GWPT-1**	2.41	2.42	2.42	5.60	5.61	5.61	9.60	9.60	9.61
		GWPT-2	2.61	2.62	2.62	5.81	5.81	5.82	9.80	9.81	9.81
		GWPT-3	2.88	2.88	2.89	6.07	6.07	6.08	10.07	10.07	10.08
	MME-2	**GWPT-1**	1.63	1.63	1.64	4.82	4.83	4.84	8.82	8.83	8.83

		GWPT-2	1.84	1.85	1.85	5.05	5.05	5.06	9.04	9.04	9.05
		GWPT-3	2.19	2.20	2.21	5.48	5.49	5.48	9.43	9.43	9.44
CL3	**MME-1**	**GWPT-1**	2.41	2.42	2.42	5.61	5.61	5.62	9.60	9.61	9.61
		GWPT-2	2.62	2.62	2.63	5.81	5.82	5.82	9.81	9.81	9.82
		GWPT-3	2.88	2.89	2.89	6.08	6.08	6.09	10.07	10.08	10.08
	MME-2	**GWPT-1**	1.63	1.63	1.64	4.83	4.83	4.84	8.82	8.83	8.83
		GWPT-2	1.84	1.85	1.86	5.04	5.05	5.05	9.04	9.05	9.05
		GWPT-3	2.27	2.27	2.29	5.47	5.48	5.49	9.50	9.52	9.51
CL4	**MME-1**	**GWPT-1**	2.41	2.42	2.42	5.60	5.61	5.61	9.59	9.60	9.60
		GWPT-2	2.63	2.64	2.64	5.82	5.83	5.83	9.81	9.82	9.82
		GWPT-3	3.01	3.01	3.01	6.18	6.19	6.19	10.17	10.18	10.19
	MME-2	**GWPT-1**	1.63	1.64	1.64	4.83	4.83	4.84	8.82	8.82	8.83
		GWPT-2	1.91	1.92	1.92	5.10	5.11	5.12	9.10	9.11	9.11
		GWPT-3	124.25	118.77	109.67	114.25	121.28	110.65	123.36	117.35	116.52
CL5	**MME-1**	**GWPT-1**	2.42	2.42	2.43	5.62	5.62	5.63	9.61	9.61	9.62
		GWPT-2	2.64	2.64	2.65	5.84	5.84	5.85	9.83	9.84	9.84
		GWPT-3	3.01	3.01	3.03	6.21	6.22	6.21	10.25	10.26	10.27
	MME-2	**GWPT-1**	1.63	1.64	1.64	4.83	4.84	4.84	8.82	8.83	8.84
		GWPT-2	1.92	1.92	1.93	5.14	5.15	5.16	9.12	9.12	9.13
		GWPT-3	211.91	206.62	204.39	218.56	204.36	214.20	218.53	213.32	225.81
CL6	**MME-1**	**GWPT-1**	2.42	2.43	2.43	5.62	5.62	5.63	9.62	9.62	9.63
		GWPT-2	2.65	2.65	2.66	5.84	5.85	5.85	9.84	9.84	9.85
		GWPT-3	3.04	3.05	3.05	6.25	6.27	6.28	10.25	10.25	10.26
	MME-2	**GWPT-1**	1.64	1.64	1.65	4.83	4.84	4.84	8.83	8.84	8.84
		GWPT-2	1.93	1.94	1.94	5.14	5.14	5.15	9.13	9.14	9.15
		GWPT-3	276.74	285.26	291.68	290.66	284.46	286.19	293.69	288.78	293.62
CL7	**MME-1**	**GWPT-1**	2.54	2.55	2.55	5.72	5.72	5.72	9.73	9.74	9.75
		GWPT-2	2.77	2.77	2.78	5.96	5.95	5.96	9.96	9.99	10.00
		GWPT-3	3.68	3.64	3.71	6.89	6.85	6.93	10.86	10.90	10.93
	MME-2	**GWPT-1**	1.69	1.70	1.70	4.90	4.91	4.91	8.87	8.88	8.88
		GWPT-2	2.25	2.22	2.27	5.47	5.50	5.49	9.48	9.50	9.47
		GWPT-3	595.87	597.93	604.40	599.44	601.23	587.75	599.35	595.19	610.26
CL8	**MME-1**	**GWPT-1**	2.56	2.57	2.57	5.81	5.80	5.81	9.78	9.79	9.79
		GWPT-2	2.80	2.81	2.81	6.02	6.04	6.02	10.02	10.02	10.04
		GWPT-3	4.17	4.03	4.18	7.43	7.41	7.27	11.42	11.35	11.45
	MME-2	**GWPT-1**	1.71	1.71	1.72	4.92	4.92	4.93	8.90	8.91	8.91
		GWPT-2	2.49	2.52	2.51	5.70	5.75	5.79	9.71	9.76	9.69
		GWPT-3	640.06	636.84	636.87	634.58	636.05	642.18	628.42	629.87	630.61
CL9	**MME-1**	**GWPT-1**	2.65	2.66	2.66	5.85	5.84	5.85	9.84	9.86	9.85
		GWPT-2	2.88	2.88	2.90	6.09	6.08	6.10	10.09	10.09	10.10
		GWPT-3	5.46	5.56	5.73	9.27	8.95	8.82	13.05	13.31	12.56
	MME-2	**GWPT-1**	1.75	1.76	1.77	4.95	4.95	4.96	8.94	8.95	8.95
		GWPT-2	3.27	3.17	3.50	6.45	6.44	6.38	10.86	10.40	10.32
		GWPT-3	666.84	658.79	661.72	670.08	661.94	667.13	659.43	674.12	666.94

Table 6-12: E2ED results for Reg-2 in SDN-SCTP EPC

			DEL-1			**DEL2**			**DEL3**		
			BGT-1	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**	**BGT-1**	**BGT-2**	**BGT-3**
CL1	**MME-1**	**GWPT-1**	2.24	2.24	2.24	4.63	4.63	4.63	7.62	7.62	7.62

		GWPT-2	2.30	2.30	2.31	4.69	4.70	4.70	7.68	7.69	7.69
		GWPT-3	2.38	2.38	2.39	4.77	4.78	4.78	7.76	7.77	7.77
	MME-2	**GWPT-1**	1.43	1.43	1.43	3.82	3.83	3.83	6.82	6.82	6.82
		GWPT-2	1.50	1.50	1.50	3.89	3.89	3.90	6.89	6.89	6.89
		GWPT-3	1.59	1.59	1.59	3.98	3.98	3.99	6.98	6.98	6.98
CL2	**MME-1**	**GWPT-1**	2.24	2.24	2.24	4.63	4.64	4.64	7.63	7.63	7.63
		GWPT-2	2.31	2.31	2.31	4.70	4.70	4.71	7.70	7.70	7.70
		GWPT-3	2.39	2.39	2.39	4.78	4.78	4.79	7.78	7.78	7.78
	MME-2	**GWPT-1**	1.43	1.43	1.43	3.83	3.83	3.83	6.82	6.83	6.83
		GWPT-2	1.50	1.50	1.50	3.90	3.90	3.90	6.89	6.90	6.90
		GWPT-3	1.59	1.59	1.59	3.98	3.99	3.99	6.98	6.98	6.99
CL3	**MME-1**	**GWPT-1**	2.24	2.24	2.25	4.64	4.64	4.64	7.63	7.64	7.64
		GWPT-2	2.31	2.31	2.31	4.70	4.71	4.71	7.70	7.70	7.71
		GWPT-3	2.39	2.39	2.39	4.79	4.79	4.79	7.78	7.78	7.79
	MME-2	**GWPT-1**	1.43	1.43	1.43	3.83	3.83	3.83	6.83	6.83	6.83
		GWPT-2	1.50	1.50	1.50	3.90	3.90	3.90	6.90	6.90	6.90
		GWPT-3	1.59	1.59	1.59	3.99	3.99	3.99	6.98	6.99	6.99
CL4	**MME-1**	**GWPT-1**	2.27	2.27	2.28	4.65	4.65	4.66	7.64	7.64	7.64
		GWPT-2	2.34	2.34	2.34	4.72	4.72	4.72	7.71	7.71	7.71
		GWPT-3	2.42	2.42	2.43	4.81	4.81	4.81	7.80	7.80	7.80
	MME-2	**GWPT-1**	1.44	1.45	1.45	3.83	3.84	3.84	6.84	6.84	6.84
		GWPT-2	1.51	1.52	1.52	3.91	3.91	3.91	6.91	6.91	6.91
		GWPT-3	1.65	1.66	1.66	4.04	4.04	4.05	7.03	7.04	7.04
CL5	**MME-1**	**GWPT-1**	2.28	2.28	2.28	4.66	4.66	4.67	7.67	7.67	7.68
		GWPT-2	2.34	2.34	2.35	4.73	4.74	4.74	7.74	7.74	7.74
		GWPT-3	2.43	2.43	2.43	4.83	4.83	4.83	7.83	7.82	7.83
	MME-2	**GWPT-1**	1.45	1.45	1.45	3.85	3.85	3.85	6.84	6.85	6.85
		GWPT-2	1.52	1.52	1.52	3.92	3.92	3.93	6.92	6.92	6.92
		GWPT-3	1.66	1.66	1.66	4.05	4.05	4.05	7.05	7.05	7.05
CL6	**MME-1**	**GWPT-1**	2.29	2.29	2.30	4.68	4.69	4.69	7.68	7.68	7.69
		GWPT-2	2.36	2.36	2.36	4.75	4.76	4.76	7.75	7.75	7.76
		GWPT-3	2.45	2.45	2.45	4.84	4.84	4.84	7.84	7.84	7.84
	MME-2	**GWPT-1**	1.45	1.46	1.46	3.85	3.85	3.86	6.85	6.85	6.86
		GWPT-2	1.53	1.53	1.53	3.93	3.93	3.93	6.92	6.93	6.93
		GWPT-3	1.68	1.68	1.68	4.06	4.07	4.07	7.06	7.06	7.07
CL7	**MME-1**	**GWPT-1**	2.79	2.80	2.87	5.17	5.19	5.16	8.11	8.13	8.13
		GWPT-2	2.91	2.96	2.93	5.26	5.26	5.22	8.21	8.22	8.20
		GWPT-3	3.01	3.01	3.02	5.29	5.33	5.34	8.29	8.30	8.30
	MME-2	**GWPT-1**	1.73	1.76	1.75	4.09	4.08	4.09	7.09	7.11	7.10
		GWPT-2	1.79	1.80	1.81	4.16	4.16	4.15	7.17	7.17	7.20
		GWPT-3	2.11	2.13	2.12	4.41	4.41	4.42	7.45	7.43	7.41
CL8	**MME-1**	**GWPT-1**	3.35	3.24	3.40	5.74	5.69	5.60	8.80	8.72	8.69
		GWPT-2	3.29	3.25	3.33	5.67	5.71	5.70	8.72	8.71	8.67
		GWPT-3	3.46	3.38	3.48	5.71	5.67	5.70	8.84	8.84	8.77
	MME-2	**GWPT-1**	1.96	1.94	1.92	4.37	4.34	4.34	7.34	7.35	7.30
		GWPT-2	2.05	2.02	2.09	4.47	4.41	4.40	7.38	7.38	7.38
		GWPT-3	2.48	2.60	2.53	4.77	4.81	4.83	7.72	7.69	7.78
CL9	**MME-1**	**GWPT-1**	4.91	5.28	4.49	7.47	6.87	7.72	10.14	10.16	9.73
		GWPT-2	5.14	4.76	4.92	7.53	7.31	7.54	10.29	10.11	10.50
		GWPT-3	4.73	4.89	4.56	7.53	7.47	7.06	10.37	10.85	

	MME-2	**GWPT-1**	2.80	2.77	2.85	5.29	5.05	5.49	8.32	8.13	8.26
		GWPT-2	2.82	2.74	2.86	5.31	5.41	5.20	8.21	8.15	8.03
		GWPT-3	3.74	3.85	4.00	5.87	6.36	6.06	8.64	8.98	8.93

Using results in tables above and ANOVA technique described earlier, we quantify the main factors effects and the interaction of factors along with their contribution to overall E2ED results at selected control load points CL2, CL5 and CL8 for CONV-EPC and SDN-SCTP EPC networks.

Table 6-13 and Table 6-14 show the ANOVA study results at CL2, the results assert the following:

- DEL factor have the major effect of variation of results, the coefficients of this factor is the highest of all other factors coefficients, the coefficients are shown to be -3.47 and -2.6 at DEL-1, and 3.73 and 2.8 at DEL-3 for CONV-EPC and SDN-SCTP respectively, these values are 5 times higher than other factors coefficients which indicate the dominance of this factor on other factors contribution. However, it is noticed that the variation in DEL levels leads to more increase in E2ED result for the case of CONV-EPC more than that for SDN-SCTP; this is indicated by higher factor coefficients for CONV-EPC than SDN-SCTP.
- The ANOVA study shows that in DEL contributes 97% of E2ED results which is a clear indication of factor dominance.
- DBGT coefficients show no contribution of DBGT factor to E2ED variation with 0% contribution to overall results.
- MMECAP effect is almost equal for CONV-EPC and SDN-SCTP with almost 0.36 msec variation of E2ED are attributed to MMECAP change. The percentage of

MMECAP contribution to E2ED variation is around 1.5% and 3.2% in both networks types respectively.

- GWPT factor have least effect than DEL and MMECAP, where it induces 0.25 msec and 0.1 msec variation in results for CONV-EPC and SDN-SCTP EPC respectively according to factor coefficients. Although CONV-EPC is observed to have higher impact due to GWPT level than SDN-SCTP EPC, it is extremely small relative to overall delay. The contribution of this factor is 0.55% and 0.08 to overall variations for both networks respectively.
- Factors interactions: the ANOVA study shows that there is no interaction between any of the factors DEL, GWPT, MMECAP, and DBGT in any combination among them. The interaction between factors is indicated by values: DEL & BGT (SS), DEL & MMECAP (SS), DEL & GWPT (SS), BGT & MMECAP (SS), and BGT & GWPT (SS), these all have 0 value which indicates no interaction among any combination at this control load level. MMECAP & GWPT (SS) has more than 0 value however extremely small, less than 0.01%, which can be neglected. This result supports the fractional factorial experiment design explained in chapter 5, that no interaction between these factors exist, and therefore a full factorial experiment is not required for these factors combinations. A fractional factor experiment that varies one of these factors and configures the others to average configuration would still capture all of the effects of these factors
- The results show in F-test section that all F-computed values are larger than their respective F-table values. This indicates that the percentage of variation explained by

each factor in the ANOVA model is statistically significant. These results support the robustness of this model.

ANOVA study for E2ED results for Reg-2

Table 6-13: CONV- EPC at CL2

Table 6-14: SDN-SCTP EPC at CL2

	MAIN FACTORS EFFECTS LEVEL (CONV)			MAIN FACTORS EFFECTS LEVEL (SDN-SCTP)		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-3.47	-0.26	3.73	-2.6	-0.2	2.8
BGT	-0.01	0	0.01	0	0	0
MMECAP	0.36	-0.36		0.4	-0.4	
GWPT	-0.25	-0.04	0.29	-0.07	-0.01	0.08

Component	**Square sum**	**Percent Variation**	**Degrees of freedom**	**Square Sum**	**Percent Variation**	**Degrees of freedom**
Response(SSY)	2254.18		54	1368.52		54
Mue (SS0)	1776.39		1	1096.76		1
y-Mue (SST)	477.8	100	53	271.77	100	53
Main effects	477.73	99.99	7	271.76	100	7
DEL (SS)	467.95	97.94	2	262.83	96.71	2
DBGT (SS)	0	0	2	0	0	2
MMECAP(SS)	7.14	1.49	1	8.72	3.21	1
GWPT (SS)	2.65	0.55	2	0.21	0.08	2
1st order	0.06	0.01	18	0	0	18
DEL & DBGT (SS)	0	0	4	0	0	4
DEL & MMECAP (SS)	0	0	2	0	0	2
DEL & GWPT (SS)	0	0	4	0	0	4
BGT & MMECAP (SS)	0	0	4	0	0	4
BGT & GWPT (SS)	0	0	2	0	0	2
MMECAP & GWPT (SS)	0.06	0.01	2	0	0	2
2nd order	0	0	20	0	0	20

DEL & MMECAP & GWPT (SS)	0	0	8	0	0	8
DEL & DBGT & MMECAP (SS)	0	0	4	0	0	4
DEL & DBGT & GWPT (SS)	0	0	4	0	0	4
DBGT & MMECAP & GWPT (SS)	0	0	4	0	0	4
3rd order						
All 4 factors	0	0	8	0	0	8
Totals	477.8			271.77		

	F-test study								
DEL	**F-test Computed**	13136518	**F-test table**	0.59		**F-test Computed**	219652717.5	**F-test table**	0.59
DBGT		74270.84		0.59			176152.43		0.59
MMECAP		400728.3		0.66			14579190.68		0.66
GWPT		29.36		0.59			185.61		0.59

Figure 6-3 and Figure 6-4 show the quantile-quantile plot of normal quantile versus residual quantile. This plot is used to detect the distribution of errors in the ANOVA model. The Q-Q plot shows that errors are extremely small relative to the average response value. Most residual values are within -0.002 and 0.002 for both SDN-EPC and CONV-EPC networks. This indicates that the factor coefficients provide strong relation with response variable as computed by the ANOVA study. Although the residuals show a pattern in the Q-Q plot, their extreme small values relative to response variable values allows considering it unimportant.

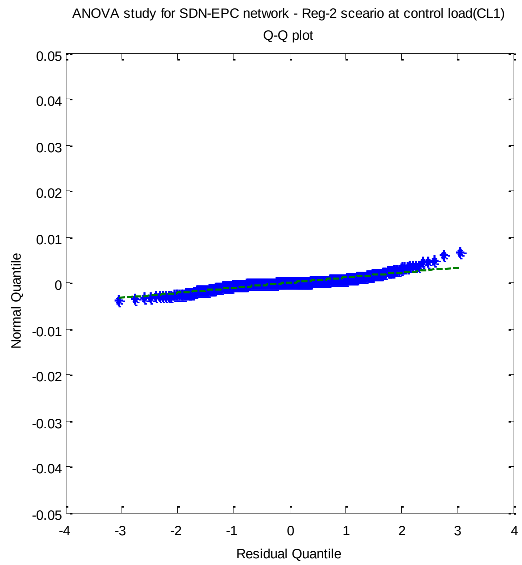


Figure 6-3: Q-Q plot of residuals for ANOVA study for SDN-EPC in Reg-2 scenario at CL2

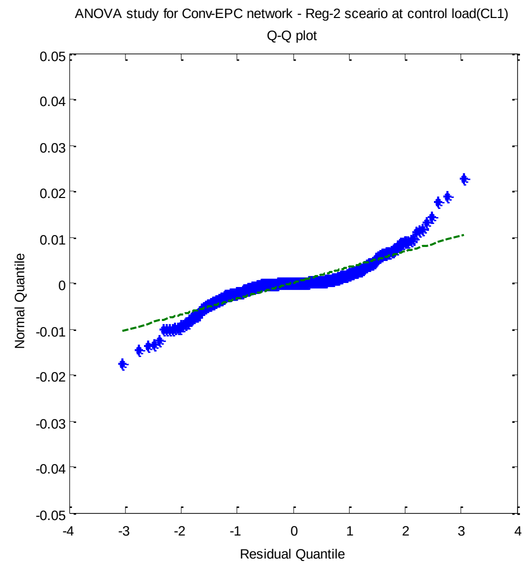


Figure 6-4: Q-Q plot of residuals for ANOVA study for CONV-EPC in Reg-2 scenario at CL2

Table 6-15 and Table 6-16 show the ANOVA study results at CL5, the results assert the following:

- SDN-EPC have almost the same factor coefficients and contributions those at CL2
- CONV-EPC has major change of MMECAP and GWPT factors effects particularly at GWPT-3; it is observed that at GWPT-3 there is significant increase on GWPT coefficient, whereas GWPT-1 and 2 are almost identical. The contribution of GWPT and MMECAP has increased considerably to reach 40% and 20% respectively, and the interaction between them has reached almost 40%. However, relating these information with results shown in previous tables, we already identified that this behavior occurred due to extreme congestion at EPC gateways, therefore the system is unstable and these results do not in fact represent interaction between GWPT and

MMECAP, although the fractional factorial experiment design takes in consideration all GWPT and MMECAP factors combinations.

The results also show that there is no interaction among DEL, DBGT and GWPT and MMECAP, indicated by zero contribution of the corresponding interaction terms in the tables, which following results at CL2 supports the fractional factor experiment design.

ANOVA study for E2ED results for Reg-2

Table 6-15: CONV- EPC at CL2 | Table 6-16: SDN-SCTP EPC at CL2

	MAIN FACTORS EFFECTS LEVEL (Table 6-15)			MAIN FACTORS EFFECTS LEVEL (Table 6-16)		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-3.47	-0.26	3.73	-2.6	-0.2	2.8
BGT	-0.01	0	0.01	0	0	0
MMECAP	0.36	-0.36		0.4	-0.4	
GWPT	-0.25	-0.04	0.29	-0.07	-0.01	0.08

Component	**Square sum**	**Percent Variation**	**Degrees of freedom**	**Square Sum**	**Percent Variation**	**Degrees of freedom**
Response(SSY)	410955.		54	1387.17		54
Mue (SS0)	87896.1		1	1115.08		1
y-Mue (SST)	323059.	100	53	272.09	100	53
Main effects	193883.7	60.01	7	272.09	100	7
DEL (SS)	569.58	0.18	2	263.02	96.67	2
DBGT (SS)	19.12	0.01	2	0	0	2
MMECAP(SS)	63091.3	19.53	1	8.76	3.22	1
GWPT (SS)	130203.	40.3	2	0.3	0.11	2
1st order	129049.5	39.95	18	0.01	0	18
DEL & DBGT (SS)	16.56	0.01	4	0	0	4
DEL & MMECAP (SS)	4.85	0	2	0	0	2
DEL & GWPT (SS)	10.15	0	4	0	0	4
BGT & MMECAP (SS)	38.36	0.01	4	0	0	4

BGT & GWPT (SS)	19.17	0.01	2		0	0	2
MMECAP & GWPT (SS)	128955.	39.92	2		0.01	0	2
2nd order	97.68	0.03	20		0	0	20
DEL & MMECAP & GWPT (SS)	33.1	0.01	8		0	0	8
DEL & DBGT & MMECAP (SS)	16.55	0.01	4		0	0	4
DEL & DBGT & GWPT (SS)	9.7	0	4		0	0	4
DBGT & MMECAP & GWPT (SS)	38.33	0.01	4		0	0	4
3rd order							
All 4 factors	33.14	0.01	8		0	0	8
Totals	323064				272.09		

	F-test study								
DEL	F-test Computed	4.6	F-test table	0.59		F-test Computed	9098786	F-test table	0.59
DBGT		1052.23		0.59			10301.71		0.59
MMECAP		1019.73		0.66			606376.5		0.66
GWPT		0.15		0.59			7.77		0.59

Table 6-17 and Table 6-18 show the ANOVA study results at CL8, , the results assert the findings observed in same raw results tables (Table 6-11 and Table 6-12) at CL8. In CONV-EPC GWPT and MMECAP factors are causing the means of the experiments to be large affected by their effect and they appear to be major contribution to E2ED, however it is already identified the cause of this matter, but it is observed their effect is more considerable than that at CL5.

In SDN-SCTP it is noticed that effect of GWPT and MMECAP is slightly increased than that at CL5 and CL8 but still minimal compared to DEL factor; DEL factor contribution is decreased by 4% to reach 92% contribution to E2ED overall results.

The results show in F-test section that not all factor effects pass the F-test. DBGT and MMECAP pass the significance test. GWPT also does not pass the F-test, for the same reason mentioned at CL5 that errors due other factors mask the effect of GWPT. This prevents the ANOVA model from providing confidence in GWPT effect. DEL factor pass the F-test in both SDN-EPC and CONV-EPC. However, in CONV-EPC the F-computed value is very close to the F-table which indicates that the confidence in the factor effect is not very strong due to errors from system instability.

In terms of factors interactions it is still as before factors DEL and DBGT have no interactions with GWPT and MMECAP factors, and that supports the fractional factorial experiment design as other results have shown as well.

ANOVA study for E2ED results for Reg-2

Table 6-17: CONV- EPC at CL8

Table 6-18: SDN-SCTP EPC at CL8

	MAIN FACTORS EFFECTS LEVEL (CONV- EPC)			MAIN FACTORS EFFECTS LEVEL (SDN-SCTP EPC)		
FACTOR	**1**	**2**	**3**	**1**	**2**	**3**
DEL	-2.43	0.21	2.21	-2.56	-0.22	2.79
BGT	-0.12	-0.14	0.25	0.01	-0.01	0
MMECAP	-104.38	104.38		0.6	-0.6	
GWPT	-105.42	-104.89	210.31	-0.1	-0.07	0.17

Component	**Square sum**	**Percent Variation**	**Degrees of freedom**	**Square Sum**	**Percent Variation**	**Degrees of freedom**
Response(SSY)	3631974		54	1813.55		54
Mue (SS0)	665728.		1	1533.78		1
y-Mue (SST)	2966245	100	53	279.76	100	53
Main effects	1782715	60.1	7	279.19	99.8	7
DEL (SS)	194.76	0.01	2	258.85	92.53	2
DBGT (SS)	1.73	0	2	0	0	2
MMECAP(SS)	588289	19.83	1	19.51	6.97	1

GWPT (SS)	1194229	40.26	2		0.82	0.29	2
1st order	1183434	39.9	18		0.55	0.2	18
DEL & DBGT (SS)	5.17	0	4		0.01	0	4
DEL & MMECAP (SS)	66.84	0	2		0.03	0.01	2
DEL & GWPT (SS)	130.76	0	4		0.04	0.01	4
BGT & MMECAP (SS)	3.35	0	4		0.01	0	4
BGT & GWPT (SS)	1.71	0	2		0	0	2
MMECAP & GWPT (SS)	1183159	39.89	2		0.42	0.15	2
2nd order	151.95	0.01	20		0.04	0.02	20
DEL & MMECAP & GWPT (SS)	10.17	0	8		0.01	0	8
DEL & DBGT & MMECAP (SS)	5.58	0	4		0.01	0	4
DEL & DBGT & GWPT (SS)	132.82	0	4		0.02	0.01	4
DBGT & MMECAP & GWPT (SS)	3.38	0	4		0.01	0	4
3rd order							
All 4 factors	10.87	0	8		0.01	0.01	8
Totals	2966312				279.79		

	F-test study								
DEL	**F-test Computed**	2.53	**F-test table**	0.59		**F-test Computed**	3971.9	**F-test table**	0.59
DBGT		15488.43		0.59			12.6		0.59
MMECAP		15259.52		0.66			598.77		0.66
GWPT		0.02		0.59			0.06		0.59

CHAPTER 7
FRACTIONAL FACTORIAL DESIGN SIMULATION RESULTS AND PERFROMANCE EVALUATION OF CONTROL OPERATIONS

7.1 INTRODUCTION

In this chapter, we present fractional factor simulation experiments results for control plane operations described in previous chapters. We present and analyze the results obtained; the simulation results for S1-based handover and registration procedures are presented in separate sections. Each section will evaluate the End-to-End delay (E2ED) metric for the procedures, bandwidth utilization (BW-Util) metric at the MME links and resource utilization of the MME and SGW involved in the control operation. A comparison is made among different configurations of the controlled factors and their impact on the metrics is measured. It should be noted that the simulation is based on assumptions and network configuration described 5.5.2.

A *simulation configuration* is the term used for the collection of *EPC network type*, *simulation scenario*, and *simulation parameters combination*. In simulation combinations the parameters are varied according to parameter levels described in previous chapters. The performance results for these experiments are reported in a systematic approach based on information desired from each parameter.

First, E2ED results are reported for each simulation combination, simulation scenario, and EPC network type in the following fashion:

- Effect of factors DEL, DBGT, and distribution of gateways through results for simulation scenarios and combinations: This mainly explores the interactions between DEL factor, DBGT factors and different locations of EPC gateways with isolation of control load factor effect. The combinations 1, 3, and 32 are the main concern for these factors; however for the sake of completeness we will show other combinations as well. The location change of the gateways is coupled with conditions that could vary in the operations, and these would be the propagation delay of connecting links involved in the operation and background traffic; as different location leads to different routes and consequently dependent on conditions of this route. The results in this section are reported for control load point CL5, i.e. CBGT 50% and CORR 10%, which is an average control load point as we are interested in effect of location change and link conditions on E2ED without the effect of varying control load, which is to be studied later on. The E2ED is reported at each fractional factor level for DBGT and DEL along with the relative E2ED performance of SDN-EPC to CONV-EPC is presented at each fractional factor level; the order of scenarios within same EPC network type and across all EPC network types is inferred.
- Effect of MME control load and EPC nodes processing capacity: This part presents exact E2ED **at** each control load point **at** a particular fractional factor level for each particular simulation configuration. This is the individual E2ED value recorded at each run, and shall be viewed as a function of control load levels. This part will confirm results presented in the previous method and show the interaction of results with control load. The relative variation of E2ED performance for SDN-EPC network

types from CONV-EPC network in each simulation configuration is deduced from these results.

Second, representation of MME links BW-Util results are reported as the ratio of increase in BW-Util for SDN-EPC compared to CONV-EPC. The reported value is averaged over all combinations as it is not expected to vary across simulation combinations.

Third, representation of MME resource utilization (MME-CPU) results for all EPC network types are reported as function of control load at each combinations and scenarios, in addition the relative variation of MME-CPU in SDN-EPC compared to CONV-EPC is presented through ratio of average MME-CPU UTILIZATOIN**.**

Lastly, representation of core SGW resource utilization (core SGW-CPU) results are presented similarly to representation of MME-CPU results; moreover, the effect of different GWPT levels is presented more clearly through ratio of average core SGW-CPU UTILIZATOIN**.**

7.2 S1-BASED HANDOVER SIMULATION RESULTS

Currently, there is no specific E2ED for S1-based handover procedure found in the literature; the only condition it has to meet is the quality of service required for the pertinent traffic flow in LTE system. On success of the S1-based handover procedure, data plane packets (user's traffic) will flow from the target (new) eNB through the new SGW towards the anchor PGW. The required time for completing this procedure introduces delay for the data plane packets. Table 7-1 shows the delay budget for different classes of service in LTE system. It can be seen that the most stringent budget allows for a 50 milli-second of E2ED per packet. The obtained results of average E2ED for S1-based handover procedure are below the most stringent delay budget that is 50 milli-seconds. In addition, the range of average E2ED for all combination under average control load is 4 to 32 milli-seconds. This is clearly below the delay budget for all QoS classes' values in Table 7-1. This gives a clear indication of the appropriateness of parameters configurations. It should be noted that delay budget of different QoS takes into consideration the transport delay budget which could range between 1 mill-second to 15 milli-seconds. However, in our study the chosen value of the transport delay is configured to values of 0.1, 0.5, and 1.0 milli-second for two reasons. The first is as an attempt to not overwhelm results of E2ED with the transport delay. The second is that the transport delay of over 1 milli-second is mainly contributed by intermediate nodes processing of data plane packets. The propagation delay of 100 Km links is well less than 1 milli-second, while in this situation the packets being transferred are control packets which do not require much processing hence their small size and higher priority in the transport network. For the reasons stated, the parameters configurations for this study are considered appropriate.

Table 7-1: Standardized Quality of Service Class Identifier (QCI) [42]

QCI	RESOURCE TYPE	PRIORITY	PACKET DELAY BUDGET (MS)	PACKET ERROR LOSS RATE	EXAMPLE SERVICES
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real-time gaming
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP-based (for example, WWW, e-mail), chat, FTP, p2p file sharing, progressive video and others
9	Non-GBR	9	300	10^{-6}	

7.2.1 E2ED PERFOMANCE METRIC

7.2.1.1 SIMULATION SCENARIOS

In this section, a detailed analysis of E2ED results for CONV-EPC and SDN-EPC networks is presented. S1-handover procedure E2ED results are presented at control load point CL5 at each fractional factor level for each simulation scenarios in the following pattern: (1) raw E2ED results; this is shown in Table 7-2. (2) Normalized E2ED results to the highest reported scenario of all scenarios in all EPC networks; this is shown in Table 7-3. For example, E2ED results for all scenarios in combination 1 at DBGT-1 are normalized to E2ED of CONV-EPC Mob-D hence it reports the highest E2ED at this fractional factor level across all scenarios and EPC networks. (3) Normalized E2ED results to the highest reported scenario of all scenarios in the same EPC network; this is shown in Table 7-4. For example, E2ED results for SDN-SCTP in combination 1 at DBGT-1 are normalized to E2ED of SDN-SCTP Mob-B hence it reports the highest E2ED at this fractional factor level across all scenarios in SDN-SCTP; the same is computed in SDN-UDP and CONV-EPC. (4) Normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC; this is shown in Table 7-5. For example, MOB-A in SDN-SCTP and SDN-UDP are both normalized to Mob-A in CONV-EPC, the same is computed for the other scenarios.

From these E2ED results tables, we deduce the following findings divided into sections: (1) scenarios E2ED performance order across all EPC network types at the same fractional factor level; based on Table 7-3. (2) Scenarios E2ED performance order within the same EPC network type; based on Table 7-4. (3) Comparison of E2ED results for SDN-SCTP and SDN-UDP EPC network types; based on Table 7-4. (4) E2ED performance of SDN-EPC compared to CONV-EPC network; based on Table 7-5. (5) E2ED performance at each fractional factor level and factors contribution to variation; based on Table 7-2. Finally, (6) the effect of increasing MME processing capacity is analyzed; based on Table 7-2.

Table 7-2: E2ED results (milli-seconds) at CL5 in each EPC network – mobility procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D
1	Comb-1 DBGT-1	13.91	16.93	12.87	17.89	14.01	17.08	12.98	17.05	13.94	17	12.89	17.05
2	Comb-1 DBGT-2	13.92	16.95	12.88	17.91	14.01	17.08	12.99	17.06	13.96	17	12.92	17.04
3	Comb-1 DBGT-3	13.93	16.95	12.89	17.91	14.02	17.09	13	17.07	13.96	17.02	12.91	17.07
4	Comb-2 GWPT-1	13.9	16.92	12.87	17.88	13.99	17.06	12.98	17.04	13.95	17.01	12.89	17.03
5	Comb-2 GWPT-2	14.37	17.41	13.33	18.37	14.28	17.35	13.27	17.34	14.22	17.29	13.18	17.34
6	Comb-2 GWPT-3	15.19	18.26	14.23	19.32	14.63	17.72	13.62	17.7	14.56	17.65	13.54	17.7
7	Comb-3 DEL-1	6.68	7.28	6.46	7.49	6.76	7.39	6.58	7.4	6.68	7.31	6.48	7.37
8	Comb-3 DEL-2	13.9	16.93	12.87	17.88	14	17.08	12.98	17.04	13.94	16.99	12.89	17.05
9	Comb-3 DEL-3	22.87	28.89	20.88	30.88	22.97	29.04	21.01	29.02	22.89	28.95	20.92	29.01
10	Comb-22 GWPT-1	11.48	14.49	10.48	15.48	11.52	14.55	10.53	14.55	11.48	14.51	10.5	14.54
11	Comb-22 GWPT-2	12.12	15.21	11.09	16.16	11.82	14.89	10.83	14.89	11.78	14.85	10.79	14.89
12	Comb-22 GWPT-3	427.39	429.44	429.86	438.82	12.48	15.61	11.49	15.61	12.45	15.56	11.45	15.6
13	Comb-32 DEL-1	4.29	4.88	4.08	5.09	4.33	4.93	4.14	4.93	4.3	4.89	4.1	4.92
14	Comb-32 DEL-2	11.48	14.49	10.48	15.48	11.52	14.55	10.53	14.55	11.48	14.51	10.5	14.54
15	Comb-32 DEL-3	20.5	26.5	18.48	28.5	20.54	26.56	18.53	26.55	20.52	26.52	18.49	26.54
16	Comb-4	23.57	29.7	21.56	31.65	23.37	29.5	21.4	29.49	23.31	29.41	21.32	29.48

Table 7-3: E2ED results normalized to highest E2ED at each fractional factor configuration in each EPC network in all scenario at CL5 – mobility procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D
1	Comb-1 DBGT-1	77.75	94.66	71.97	100	78.32	95.47	72.57	95.29	77.9	95.05	72.05	95.31
2	Comb-1 DBGT-2	77.72	94.64	71.96	100	78.25	95.4	72.57	95.31	77.98	94.97	72.14	95.18
3	Comb-1 DBGT-3	77.76	94.63	71.96	100	78.28	95.41	72.58	95.29	77.91	95.03	72.08	95.27
4	Comb-2 GWPT-1	77.74	94.63	71.99	100	78.27	95.4	72.58	95.32	78	95.11	72.08	95.26
5	Comb-2 GWPT-2	78.2	94.75	72.55	100	77.7	94.46	72.2	94.38	77.4	94.08	71.75	94.39
6	Comb-2 GWPT-3	78.63	94.5	73.64	100	75.72	91.72	70.51	91.63	75.38	91.34	70.06	91.63
7	Comb-3 DEL-1	89.16	97.22	86.31	100	90.24	98.74	87.8	98.77	89.24	97.55	86.51	98.35
8	Comb-3 DEL-2	77.74	94.69	71.99	100	78.34	95.54	72.63	95.34	78	95.06	72.08	95.39
9	Comb-3 DEL-3	74.07	93.58	67.62	100	74.39	94.04	68.03	93.99	74.14	93.77	67.75	93.95
10	Comb-22 GWPT-1	74.13	93.59	67.7	100	74.38	93.97	68.03	93.95	74.16	93.7	67.81	93.91
11	Comb-22 GWPT-2	74.99	94.11	68.62	100	73.13	92.15	67	92.15	72.91	91.91	66.74	92.12
12	Comb-22 GWPT-3	97.39	97.86	97.96	100	2.84	3.56	2.62	3.56	2.84	3.55	2.61	3.56
13	Comb-32 DEL-1	84.34	96.02	80.28	100	85.14	97	81.43	96.91	84.46	96.12	80.63	96.75
14	Comb-32 DEL-2	74.13	93.59	67.7	100	74.38	93.97	68.03	93.95	74.16	93.7	67.81	93.91
15	Comb-32 DEL-3	71.93	92.97	64.84	100	72.08	93.18	65.01	93.15	71.99	93.05	64.87	93.14
16	Comb-4	74.47	93.82	68.11	100	73.83	93.22	67.61	93.18	73.65	92.93	67.35	93.15

Table 7-4: E2ED results normalized to highest E2ED within same EPC network at each fractional factor configuration at CL5 – mobility procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D
1	Comb-1 DBGT-1	77.75	94.66	71.97	100	82.04	100	76.01	99.81	81.73	99.72	75.59	100
2	Comb-1 DBGT-2	77.72	94.64	71.96	100	82.02	100	76.07	99.9	81.92	99.78	75.79	100
3	Comb-1 DBGT-3	77.76	94.63	71.96	100	82.05	100	76.06	99.87	81.78	99.74	75.65	100
4	Comb-2 GWPT-1	77.74	94.63	71.99	100	82.04	100	76.09	99.92	81.88	99.85	75.67	100
5	Comb-2 GWPT-2	78.2	94.75	72.55	100	82.26	100	76.44	99.92	82.01	99.68	76.02	100
6	Comb-2 GWPT-3	78.63	94.5	73.64	100	82.56	100	76.88	99.91	82.26	99.69	76.46	100
7	Comb-3 DEL-1	89.16	97.22	86.31	100	91.36	99.97	88.89	100	90.73	99.19	87.96	100
8	Comb-3 DEL-2	77.74	94.69	71.99	100	82	100	76.02	99.79	81.77	99.65	75.56	100
9	Comb-3 DEL-3	74.07	93.58	67.62	100	79.11	100	72.34	99.95	78.92	99.81	72.11	100
10	Comb-22 GWPT-1	74.13	93.59	67.7	100	79.15	100	72.4	99.98	78.97	99.78	72.21	100
11	Comb-22 GWPT-2	74.99	94.11	68.62	100	79.36	100	72.7	100	79.15	99.77	72.45	100
12	Comb-22 GWPT-3	97.39	97.86	97.96	100	79.97	100	73.62	100	79.81	99.73	73.37	100
13	Comb-32 DEL-1	84.34	96.02	80.28	100	87.77	100	83.94	99.9	87.3	99.35	83.34	100
14	Comb-32 DEL-2	74.13	93.59	67.7	100	79.15	100	72.4	99.98	78.97	99.78	72.21	100
15	Comb-32 DEL-3	71.93	92.97	64.84	100	77.35	100	69.77	99.96	77.3	99.91	69.64	100
16	Comb-4	74.47	93.82	68.11	100	79.2	100	72.53	99.96	79.07	99.77	72.3	100

Table 7-5: E2ED results of SDN-SCTP and SDN-UDP normalized to E2ED results of CONV-EPC in the same respective scenario at CL5 – mobility procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D	Mob-A	Mob-B	Mob-C	Mob-D
1	Comb-1 DBGT-1	100	100	100	100	100.74	100.86	100.83	95.29	100.2	100.41	100.12	95.31
2	Comb-1 DBGT-2	100	100	100	100	100.68	100.8	100.85	95.31	100.33	100.35	100.24	95.18
3	Comb-1 DBGT-3	100	100	100	100	100.67	100.83	100.85	95.29	100.19	100.43	100.16	95.27
4	Comb-2 GWPT-1	100	100	100	100	100.68	100.81	100.83	95.32	100.34	100.5	100.12	95.26
5	Comb-2 GWPT-2	100	100	100	100	99.36	99.7	99.52	94.38	98.98	99.3	98.9	94.39
6	Comb-2 GWPT-3	100	100	100	100	96.3	97.06	95.75	91.63	95.86	96.66	95.14	91.63
7	Comb-3 DEL-1	100	100	100	100	101.21	101.56	101.73	98.77	100.09	100.34	100.24	98.35
8	Comb-3 DEL-2	100	100	100	100	100.77	100.9	100.88	95.34	100.34	100.39	100.12	95.39
9	Comb-3 DEL-3	100	100	100	100	100.43	100.49	100.6	93.99	100.09	100.2	100.19	93.95
10	Comb-22 GWPT-1	100	100	100	100	100.33	100.41	100.49	93.95	100.04	100.12	100.17	93.91
11	Comb-22 GWPT-2	100	100	100	100	97.51	97.92	97.64	92.15	97.22	97.66	97.26	92.12
12	Comb-22 GWPT-3	100	100	100	100	2.92	3.64	2.67	3.56	2.91	3.62	2.66	3.56
13	Comb-32 DEL-1	100	100	100	100	100.95	101.02	101.43	96.91	100.15	100.1	100.44	96.75
14	Comb-32 DEL-2	100	100	100	100	100.33	100.41	100.49	93.95	100.04	100.12	100.17	93.91
15	Comb-32 DEL-3	100	100	100	100	100.21	100.23	100.26	93.15	100.1	100.09	100.04	93.14
16	Comb-4	100	100	100	100	99.14	99.35	99.26	93.18	98.91	99.05	98.88	93.15

7.2.1.1.1 SCENARIOS PERFORMANCE ORDER ACROSS ALL EPC NETWORK TYPES

Table 7-3 shows normalized E2ED results to the highest reported scenario of all scenarios in all EPC networks. This is used to show E2ED performance order among all EPC networks at the same fractional factor level as indicated by the percentages. After further comparison between all simulations combinations, it is found that this order is the same across different fractional factor levels. This means that regardless of simulation combination and fractional factor level, the order does not change, and it is (excluding combination 22 with high GWPT level) as follows:

- CONV-EPC Mob-D as highest and the highest E2ED is recorded in combination 4 with 31.65 msec; refer Table 7-3 row 16.
- SDN-EPC Mob-D and Mob-B and CONV-EPC Mob-B as second with 93-98% ratio to highest across combinations; refer Table 7-3 rows 15 and 10.
- SDN-EPC and CONV-EPC Mob-A with 74-90% ratio to highest across combinations; refer Table 7-3 rows 15 and 10.
- And lastly scenario with the least E2ED is SDN-EPC and CONV-EPC Mob-C with 65-87% ratio to highest scenario across combinations; refer Table 7-3 rows 15 and 10. The least E2ED is recorded in combination 32 with 4.08 msec at 0.1 DEL level; refer Table 7-3 row 13.

7.2.1.1.2 SCENARIOS PERFORMANCE ORDER WITHIN SAME EPC NETWORK TYPES

Table 7-4 shows normalized E2ED results to the highest reported scenario of all scenarios in the same EPC network. This is used to deduce E2ED performance order within the same EPC networks at the same fractional factor. The E2ED relative performance order within any EPC network type has been found as: Mob-D having the highest E2ED, followed by Mob-B, Mob-A, and Mob-C with the lowest E2ED value, with the exception of combination 22 with high GWPT level. The relative percentage E2ED for each scenario relative to Mob-D scenario (E2ED at each scenario divided by E2ED in Mob-D in the same EPC network type) is:

- In combination 1: For SDN-EPC is approximately 99%, 82%, and 76%, and for CONV-EPC is approximately 94.6%, 77%, and 72%, for Mob-B, Mob-A, and Mob-C respectively regardless of DBGT factor levels; refer Table 7-3 rows 1, 2, and 3.
- In combination 2: For SDN-EPC is approximately 99%, 82%, and 76%, and for CONV-EPC is approximately 94.6%, 78%, and 72%, for Mob-B, Mob-A, and Mob-C respectively; refer Table 7-3 rows 4, 5, and 6.
- In combination 3: For SDN-EPC is approximately 99%, 79-82%, and 72-89%, and for CONV-EPC is approximately 93-97%, 74-89%, and 67-86%, for Mob-B, Mob-A, and Mob-C respectively, where the difference in E2ED between Mob-D and other scenarios increases with increasing DEL level; refer Table 7-3 rows 10, 11, and 12.
- In combination 22: For SDN-EPC is approximately 99%, 79%, and 76%, and for CONV-EPC is approximately 93.6%, 74%, and 76%, for Mob-B, Mob-A, and Mob-C respectively, excluding results for combination 22 at high GWPT level; refer Table 7-3 rows 7, 8, and 9.
- In combination 32: For SDN-EPC is approximately 99-100%, 77-87%, and 70-86%, and for CONV-EPC is approximately 92-96%, 72-84%, and 64-80%, for Mob-B, Mob-A, and Mob-C, respectively. The difference in E2ED between Mob-D and other scenarios increases with increasing DEL level; refer Table 7-3 rows 13, 14, and 15.
- In combination 4: For SDN-EPC is approximately 100%, 79%, and 72%, and for CONV-EPC is approximately 93%, 74.5%, and 68%, for Mob-B, Mob-A, and Mob-C, respectively; refer Table 7-3 row 16.

The results show that E2ED relative performance of simulation scenarios within any EPC network type is always the same; meaning the order of highest to lowest scenario is maintained regardless of EPC network type and parameters combination; however the relative performance of scenarios relative to highest within the same EPC network type is different for SDN-EPC from that for CONV-EPC. The similarity in E2ED relative performance of simulation scenarios among different EPC network types can be explained by the distance between new SGW and the MME, which leads to scenarios that have the same new SGW have comparable levels of E2ED results. Based on that for scenarios where distance is highest between new SGW and the MME, i.e. Mob-D and Mob-B where the new SGW is located in local center, E2ED values are the highest, and the opposite is true, when distance is lowest, E2ED values are lowest which is observed in cases Mob-A and Mob-C where the new SGW is located in core center and closer to the MME.

It is also observed that scenario Mob-A is always higher than Mob-C and Mob-D is always higher than Mob-B in the same EPC network, this variation can be attributed to distance between new SGW and anchor PGW involved in mobility operation; the farther the distance, which is the case in Mob-A and Mob-D, the higher the E2ED value. However, this analysis does not justify for the equality between Mob-B and Mob-D E2ED overall average for SDN-EPC, by inspecting SDN-EPC mobility procedure it can be justified by the fact that in SDN-EPC there is no trombone route between SGW and PGW, and E2ED is affected more by location of new SGW more than that of anchor PGW hence there is more MME communication with the former than the latter, thus Mob-B and Mob-D E2ED value in SDN-EPC is controlled by new SGW location which is in local center for these two scenarios.

7.2.1.1.3 PERFORMANCE OF DIFFERENT SDN-EPC NETWORK TYPES

Table 7-5 shows normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC. This is used to compare E2ED results between SDN-SCTP and SDN-UDP at each single same scenario and same fractional factor combinations. It is observed that no perceivable difference between SDN-SCTP EPC and SDN-UDP EPC E2ED results for the same simulation configuration. In addition, both types have almost the same E2ED in the same scenario, which indicates that the performance of both SDN-EPC types is comparable under same DBGT and DEL conditions.

The results showing no difference between SDN-SCTP and SDN-UDP is explained by the fact that abundant link capacity, i.e. 10Gbps, and small packets exchanged for control communication and as well acknowledgment packets does not have substantial contribution to E2ED compared to overall E2ED contributed by other factors such as DEL.

7.2.1.1.4 PERFORMANCE OF SDN-EPC NETWORK COMPARED TO CONV-EPC

Table 7-5 shows normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC at each fractional factor level as indicated by the percentages. Based on previous observation in 7.2.1.1.3 that SDN-SCTP and SDN-UDP have almost identical E2ED results, same or within 1% difference, in this section SDN-EPC is referred as the representative of both and the average E2ED of their results is reported. From this view we can deduce the difference in E2ED performance between EPC networks, and the following is observed:

- In scenarios Mob-A, Mob-B, and Mob-C, E2ED results of SDN-EPC are mainly identical to those of CONV-EPC with the same simulation configuration regardless of SDN-EPC type and at any fractional factor level. Note that difference of 1% or less is considered unimportant and deemed with equal performance. The exception for the above observation is at high GWPT level in combination 2 and 22 and average GWPT level in combination 22. The details for this exceptions are:
  - In combination 2 at high GWPT level, E2ED results are 95-97% of CONV-EPC E2ED, and the difference range is approximately 0.5-0.6 msec; refer Table 7-5 row 6.
  - In combination 22 at average GWPT level, E2ED results are 97.5% of CONV-EPC E2ED, and the difference range is approximately 0.3 msec; refer Table 7-5 row 8.
  - In combination 22 at high GWPT level, E2ED results show severe degradation in E2ED results in CONV-EPC where SDN-EPC are 2-4% of CONV-EPC E2ED results, and the difference is approximately 425 msec; refer Table 7-5 row 9.
- For scenario Mob-D:
  - In combination 1, E2ED results are 95% of CONV-EPC E2ED and the difference is approximately 0.84 msec at any DBGT level; refer Table 7-5 rows 1, 2, and 3.
  - In combination 2, E2ED results are 91.6-95.4% of CONV-EPC E2ED, where the difference in E2ED is increasing with increasing GWPT levels, i.e. higher GWPT level shows lower E2ED in SDN-EPC, and the difference range is approximately 0.84-1.62 msec; refer Table 7-5 rows 4, 5, and 6.

- In combination 22, E2ED results are 92-94% of CONV-EPC E2ED, where the difference in E2ED is increasing going from GWPT-1 to GWPT-2. The difference is approximately 0.93 and 1.29 msec; refer Table 7-5 rows 7 and 8, respectively. However, GWPT-3 leads to unstable network as will be explained later; refer Table 7-5 row 9.
- In combination 3, E2ED results are 94-99% of CONV-EPC E2ED, where the difference is increasing with increasing DEL levels and the difference range is approximately 0.1-1.86 msec; refer Table 7-5 rows 10, 11, and 12.
- In combination 32, E2ED results are 93-97% of CONV-EPC E2ED, where the difference is increasing with increasing DEL levels and the difference range is approximately 0.2-1.95 msec; refer Table 7-5 rows 13, 14, and 15.
- In combination 4, E2ED results are 93% of CONV-EPC E2ED and the difference is approximately 2.16 msec; refer Table 7-5 row 16.

Considering the pattern of the communications between MME and EPC gateways, E2ED relative variation among EPC network types is expected to behave in favor of SDN-EPC, i.e. less E2ED in SDN-EPC. The communication in CONV-EPC occurs in the pattern MME-SGW-PGW which translates to higher distance travelled, and thus larger E2ED. For the SDN-EPC the trombone route is not present. Therefore, this leads to favorable E2ED results for SDN-EPC. Based on this justification, in Mob-D case where distance travelled is larger in the trombone path for CONV-EPC compared SDN-EPC route, SDN-EPC is performing better than CONV-EPC, i.e. lesser E2ED. Whereas for Mob-A, Mob-B, and Mob-C almost all EPC network types perform the same since the trombone path is the same across these mobility scenarios for all EPC network types.

In the light of this analysis, it can be deduced that with increasing DEL levels the reduction in Mob-D E2ED in SDN-EPC from CONV-EPC is increased, and the results show that with increasing DEL levels in combinations 3 and 32 there is a 2-3% further decrease of E2ED in Mob-D at each DEL level increase. I.e. the benefit of SDN-EPC in terms of E2ED for this case is greater with higher propagation delays.

It is already explained in 4.2 and 4.2.1 that SGW in CONV-EPC requires more processing resources due to the fact that SGW is required to process responses from PGW, whereas in SDN-EPC SGW does not perform this role. The MME in the SDN-EPC performs control operations that were performed in the SGW-PGW interface in the CONV-EPC architecture, as indicated in Figure 4-2. Therefore, SDN-EPC requires less processing resources at the SGW and performs better than CONV-EPC, i.e. produces less delays. Increasing GWPT level contributes to higher delays at the EPC gateway system, which is independent of gateway location. GWPT levels and MME system capacity are tightly coupled, in combination 2 high GWPT increased E2ED for CONV-EPC. Whereas in combination 22, MME system capacity is doubled and henceforth increases number of requests at the gateways. In combination 22, average and high GWPT levels increased E2ED for CONV-EPC more than that for SDN-EPC. Therefore it is expected that with each increase in GWPT level, SDN-EPC would perform better with higher percentage than that in CONV-EPC. The results show in combination 22 at high GWPT level severe degradation in E2ED performance, which is an indication of congestion at the gateways, this will be tackled in detail later through E2ED results as a function of control load which will give a succinct view of E2ED behavior; and hence this E2ED results at high GWPT are excluded in this part of comparisons.

7.2.1.1.5 E2ED PERFORMANCE AT EACH FRACTIONAL FACTOR LEVEL AND FACTORS CONTRIBUTION TO VARIATION.

To identify effect of each fractional factor in the simulation combinations in each EPC network, Figure 7-1, Figure 7-2, and Figure 7-3 show the ratio of E2ED at low fractional factor level to E2ED at average fractional factor level, and the ratio of E2ED at high fractional factor level to E2ED at average fractional factor level for each EPC network type; that means for example E2ED at DEL equal to 0.1 msec divided by E2ED at DEL equal to 0.5 msec, and E2ED at DEL equal to 1.0 msec divided by E2ED at DEL equal to 0.5 msec.

The figures show this ratio for all scenarios as an interval. The variation of each fractional factor effect across different scenarios is found to be minimal; this is indicated by the small interval in the figures. This emphasizes the fact that the effect of fractional factor is more dominant than the effect of gateway location distribution.

- The effects of varying each fractional factor level based on results in the Figure 7-1, Figure 7-2, and Figure 7-3 are found to be:
  - DBGT have 0% change in E2ED across any of its level; no change was detected; refer to "Comb-1" in respective figures.
  - GWPT factor in combination 2 have:
    - 2% and 3% decrease in E2ED when GWPT changes from 75 usec to 10 usec for SDN-EPC and CONV-EPC respectively.
    - 3% and 5% increase in E2ED when GWPT changes from 75 usec to 150 usec for SDN-EPC and CONV-EPC respectively.
  - GWPT factor in combination 22 have:

        - 2.4% and 5% decrease in E2ED when GWPT changes from 75 usec to 10 usec for SDN-EPC and CONV-EPC respectively.
        - 5.36% increase in E2ED when GWPT changes from 75 usec to 150 usec for SDN-EPC, whereas E2ED suffered severe degradation when GWPT changes from 75 usec to 150 usec for CONV-EPC.
    - DEL factor in combination 3 have 54% decrease in E2ED at 0.1 msec DEL from E2ED at 0.5 msec DEL, and 67% increase in E2ED at 1.0 msec DEL from E2ED at 0.5 msec DEL, the same ratio is observed in all EPC network types
    - DEL factor in combination 32 have 64% decrease in E2ED at 0.1 msec DEL from E2ED at 0.5 msec DEL, and 80% increase in E2ED at 1.0 msec DEL from E2ED at 0.5 msec DEL, the same ratio is observed in all EPC network types

Factor DBGT had no perceivable impact at E2ED results in any EPC network type at any of its levels, i.e. DBGT-20%, 50%, and 80%; refer COMB-1 in Figure 7-1. This is opposite to the expected outcome for this factor effect. Further analysis of the simulation testbed indicates the reasons for this effect. The fact that physical links are configured to very high data rate (10Gbps) meant that even at very lengthy forward queues at connecting intermediate nodes, the queuing delay encountered by packets waiting for transmission is very limited. Packets will wait extremely short amount of time until the link is free for transmission due to the abundant link transmission rate. Furthermore, intermediate nodes, i.e. backhaul routers, are assumed to provide at least line rate processing capacity of forwarded packets that means processing delay in these nodes is negligible with comparison to accommodate large link capacity.

The effect of DEL levels on E2ED is clearly evident from large variation between its different levels, and there is a linear proportional increase in E2ED values for every increase in DEL level. The increase from 0.1 msec to 0.5 msec led to 1.7 to 2.15 and 2.5 to 3 times E2ED increase in combinations 3 and 32 respectively. The increase from 0.5 msec to 1.0 msec led to 1.65 and 2.05 times E2ED increase in combinations 3 and 32 respectively. This can serve as means for predicting E2ED performance using regression models for particular DEL values, especially when there are multiple distributed mobile centers hosting EPC nodes with variable distances.

The effect of GWPT in general is noticed to have minimal effect on E2ED; the increase of GWPT from 10 usec to 75 usec showed an increase of only 3% to 5% at an average control load which indicates that the contribution of this factor is limited. At high GWPT level two cases occurred, for SDN-EPC the difference is limited from the case of average GWPT, the increase in E2ED is only 5.6%, whereas for CONV-EPC E2ED suffered severe degradation, as explained before. Thus it is concluded that GWPT factor does not contribute substantially to E2ED under low and average GWPT in both architecture, but at high GWPT for CONV-EPC the E2ED suffers degradation due to limited resources at the gateways, whereas for SDN-EPC the control operations are minimally affected by the GWPT level.

7.2.1.1.6 EFFECT OF INCREASING MME PROCESSING CAPACITY

The portion of enhancement of E2ED results by increasing the MME capacity depends on the relative contribution of processing delays caused by the MME at different levels. For example at combination 32 with 60 Mbps capacity at high DEL level E2ED is 18.48 msec where in combination 3 with 30 Mbps capacity at the same DEL level E2ED is

20.88 msec, thus doubling the MME capacity led to 12% decrease in overall E2ED. Following this observation the enhancement (reduction) of E2ED results by doubling the MME capacity is computed from combinations 2, 22, 3, and 32 as follows:

- For low DEL level (0.1 msec): reduction is 33% - 37%, based on Table 7-2 rows 10 and 13.
- For average DEL level (0.5 msec): reduction is 14% - 19%, based on Table 7-2 rows 11 and 14.
- For high DEL level (1.0 msec): reduction is 8% - 12%, based on Table 7-2 rows 12 and 15.
- For low GWPT level (10 usec): reduction is 14.4% - 19%, based on Table 7-2 rows 4 and 7.
- For average GWPT level (75 usec): reduction is 12% - 19%, based on Table 7-2 rows 5 and 8.
- For high GWPT level (150 usec): reduction is 16% - 20% (for SDN-EPC only, excluding CONV-EPC results), based on Table 7-2 rows 6 and 9.

These results show that with increasing DEL values the gain of reducing E2ED by increasing the MME capacity is decreased. This is because DEL factor becomes more dominant in contribution to E2ED results with each DEL increase than MME resources capacity.

The reduction in E2ED at different GWPT levels is noticed to be similar to the reduction at an average DEL value (0.5 msec), which remind us that all GWPT experiments have DEL level configured to average DEL (0.5 msec). Thus it is concluded that the portion of enhancement in E2ED is mostly dependent on DEL value rather than GWPT level.

The results show that increasing MME capacity in combination 22 and using GWPT with high processing time requirement led to severe degradation of E2ED performance in

CONV-EPC. This E2ED degradation did not occur at GWPT-1 and GWPT-2 levels in combination 22. This indicates that EPC gateways at this level were not able to handle the control load and a congestion point has emerged at these gateways; this will be explained from resource utilization results shown in later sections.

7.2.1.1.7 PRIMARY FACTORS CONTRIBUTION ORDER

With respect to previous sections findings, the results conclude the primary factors that impact E2ED response, and these are DEL, and MME-CAP; whereas factor GWPT had minimal effect at a suitable control load level, and factor DBGT had no impact on E2ED values.

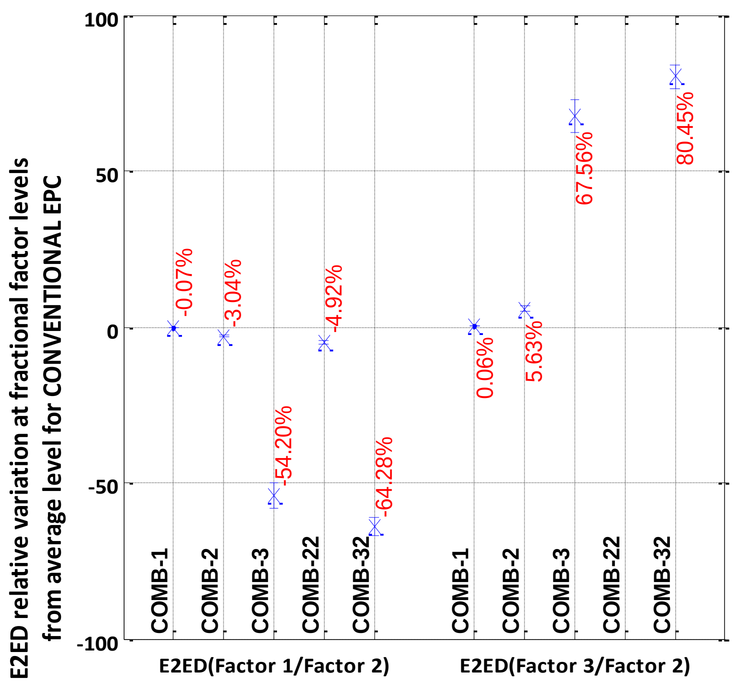


Figure 7-1:E2ED results relative variation from average fractional factor level for combinations (S1-handover experiment)

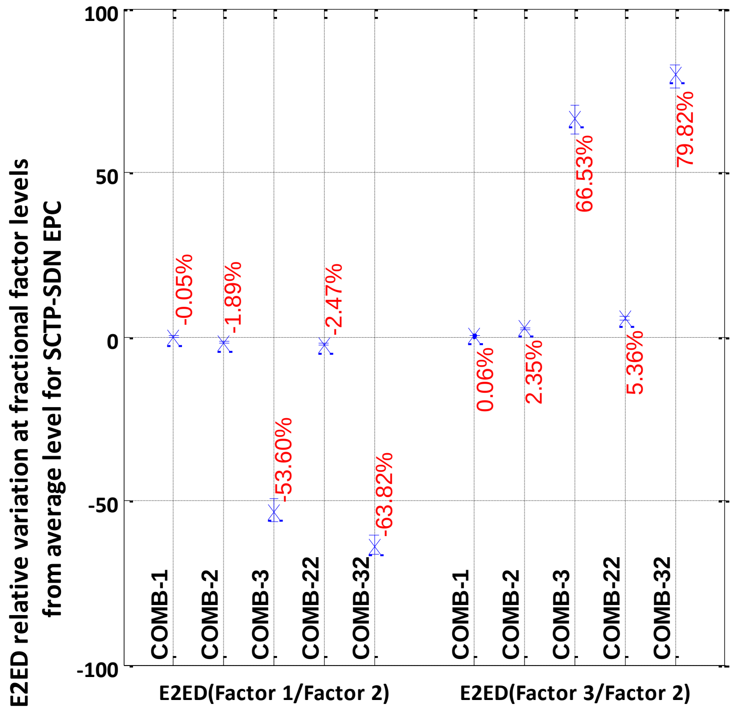


Figure 7-2: SCTP SDN-EPC E2ED results relative variation from average fractional factor level in combinations (S1-handover experiment)

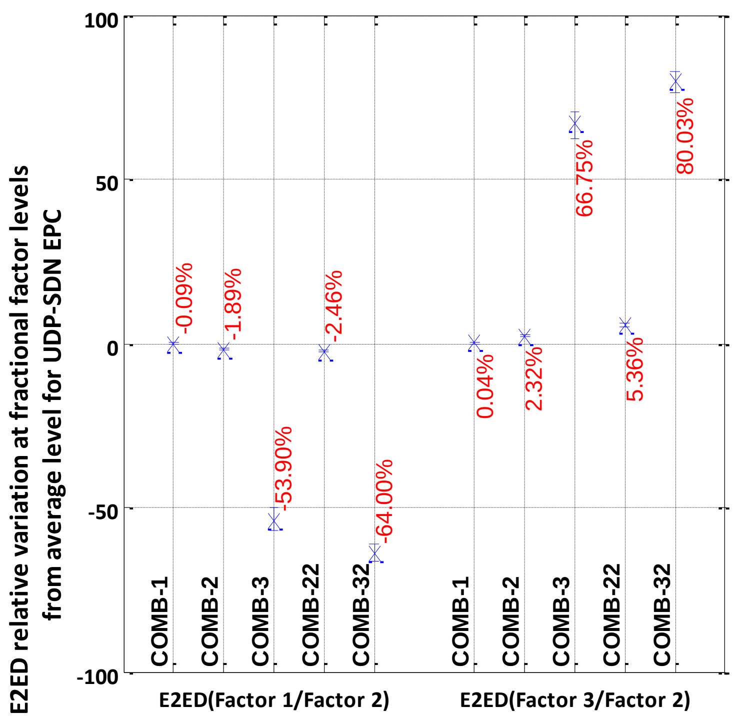


Figure 7-3: UDP SDN-EPC E2ED results relative variation from average fractional factor level in combinations

(S1-handover experiment)

7.2.1.2 E2ED VERSUS CONTROL PLANE LOAD

This subsection presents the E2ED results as a function of control load. The results viewed so far depicted the average performance of mobility operations, however, the exact performance at a particular control load is still required. Based on observations previously made, Mob-C and Mob-D are selected to view nominal E2ED values versus control load which both represents the cases of highest E2ED, i.e. Mob-D, and case of lowest E2ED, i.e. Mob-C. Mob-D also represents the case where there is variation between different EPC network types. Based on the fact that SDN-SCTP and SDN-UDP had little difference from each other, only SDN-SCTP results shall be presented for convenience and eliminating redundancy.

Figure 7-4 to **Figure 7-17** show the E2ED for mobility operation versus control plane load for scenario Mob-C and Mob-D in each combination. Each figure represents a particular scenario with all fractional factor levels, i.e. low, average, and high, for a combination along with combination 0, i.e. the average configuration. Each figure contains the E2ED results for both CONV-EPC and SDN-EPC. For example, **Figure 7-4** show combination 0 and combination 1 at DBGT 20%, 50%, and 80% for EPC networks.

- The E2ED curves show in general two patterns:
    - At low and average control load points, from CL1 to CL6, there is almost constant E2ED curves. The exception is for combination 22 the pattern is from CL1 to CL3, as shown in Figure 7-12 and Figure 7-13.

    - At high control load points, from CL7 to CL9, there is an increase in E2ED results from its E2ED at average control load points; the increase occurs at a rapid rate in SDN-EPC, whereas in CONV-EPC the increase is quite minimal.
- The E2ED results at average control load points at each fractional factor level are already reported in previous section which matches the values in the curves at CL5 in each simulation configuration, and are not repeated in this section.

- The increase in E2ED at highest control load (CL9) from average control load (CL5) is observed to be as follows:
    - In Combination 0: 8 msec and 3 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-4 and Figure 7-5.
    - In Combination 1: 8 msec and 3 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-4 and Figure 7-5.
    - In Combination 2: 8-10 msec and 1-5 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-6 and Figure 7-7.
    - In Combination 3: 8-10 msec and 1 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-8 and Figure 7-9.
    - In Combination 22:
        - At low GWPT level: 7 msec and 1 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-14 and Figure 7-15.
        - At average GWPT level: 7 msec and 3.5 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-14 and Figure 7-15.

    - At high GWPT level: 131 msec for SDN-EPC, whereas CONV-EPC have E2ED of 400 msec at CL5 and approximately 1400 msec with large standard deviation at CL9; refer Figure 7-12 and Figure 7-13.
- In Combination 32: 7-8 msec and 1 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-16 and Figure 7-17.
- In Combination 4: 9 msec and 1 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-10 and Figure 7-11.

These E2ED results show that a range of average increase of 7-10 and 1-5 msec when increasing control load from CL5 to CL9 for SDN-EPC and CONV-EPC respectively. From values above it is observed that the increase in E2ED at high control load (CL9) from E2ED at average control load (CL5) is independent from fractional factor level except for GWPT levels. However the percentage of increase in E2ED at CL9 compared to E2ED at CL5 is variable hence the CL5 E2ED level at different fractional factor level is different. The increase is observed to be slightly varying between different simulation scenarios but still close to each other within 1 msec or less.

The results show that SDN-EPC operations are more demanding for processing resources at the MME than CONV-EPC. This is indicated by the rapid rate of E2ED increase as a function of control load in SDN-EPC compared to CONV-EPC. It shows that for SDN-EPC there is a range of increase in E2ED results due to high utilization of the MME around 7-10 msec. However, for CONV-EPC the increase in E2ED is typically 1 msec when GWPT is configured to 10 usec, and increases to 3-5 msec when GWPT is configured to 75 usec. When GWPT is configured to 150 usec CONV-EPC can perform normally only with MMCAP is configured to 30 Mbps (MMECAP-1). However, when

MMECAP is configured to 60 Mbps (MMECAP-2) E2ED suffered severe degradation at control load point CL4 and above. These results show that CONV-EPC is more sensitive to gateway processing capacity than SDN-EPC, whereas SDN-EPC is more sensitive to MME resource utilization however with graceful degradation.

Relative performance of different EPC network types at average control load point (CL5) has shown in previous sections that Mob-D has less E2ED in SDN-EPC than CONV-EPC where the rest of scenarios Mob-A, Mob-B, and Mob-C are almost identical in different EPC networks. However, results in this section show that at high control load (CL9) and GWPT configured to 10 usec SDN-EPC always performs worse even in Mob-D. These findings show that any enhancement gained by eliminating trombone route can be negated if the delay produced by MME processing is higher than the savings in E2ED, if there is any.

In combination 1, E2ED curve had almost the exact values curve at any DBGT level, for any scenario type, which aligns with previous conclusions that DBGT factor in these experiments had no impact on the E2ED results.

The results extracted above of increase at high control load show that the increase in E2ED is independent from fractional factors levels except of GWPT, which is as expected there should be no interaction between MME processing and DBGT nor DEL factors.

GWPT and MMECAP factors are tightly coupled together; it is seen in combination 2 that different GWPT levels had minimal impact on E2ED, whereas at high MME capacity, combination 22, the effect of GWPT is stressed clearly. The reason is that expanding MMECAP, which actually reflects increased system capacity, led to increased request rate at high control load arriving at the gateways which leads to more resource utilization and eventually for high level of GWPT, i.e.150 micro-seconds, the gateway became completely congested as shall be seen in later results.

When GWPT is configured to 150 usec, in CONV-EPC E2ED has risen to almost 100 times the E2ED at GWPT 75 usec, whereas in SDN-EPC it has risen to almost 10 times the E2ED at GWPT 75 usec, as shown in Figure 7-12 to Figure 7-15. However, it is clear that GWPT of 10 and 75 usec have significantly lower E2ED values relative to GWPT of 150 usec, which indicates the interaction between MME capacity and GWPT factor negatively affects E2ED, where increasing MMECAP or GWPT to a certain level leads to deterioration in E2ED results.

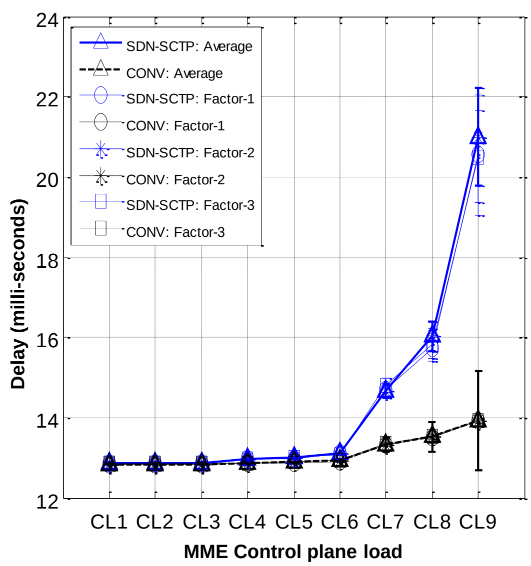


Figure 7-4: Mob-C in combination 0 and combination 1

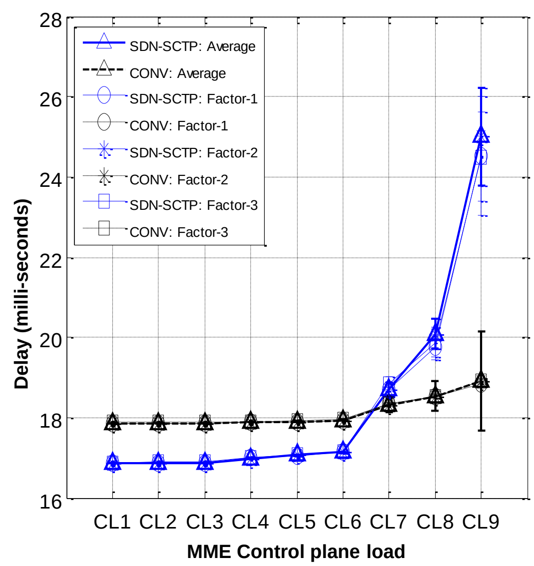


Figure 7-5: Mob-D in combination 0 and combination 1

(E2ED versus nominal control plane load at all fractional factor levels)

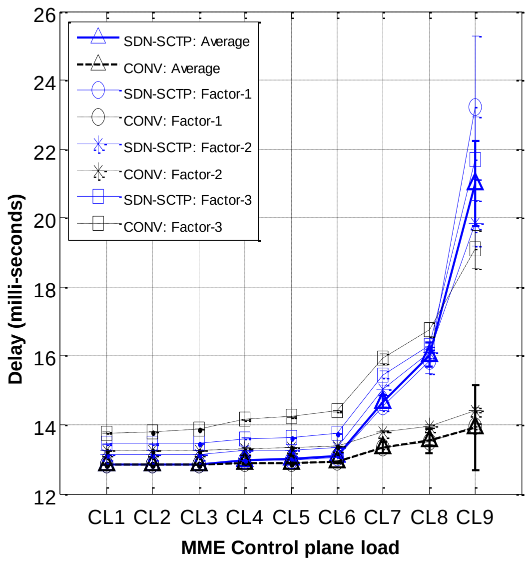


Figure 7-6: Mob-C in combination 0 and combination 2

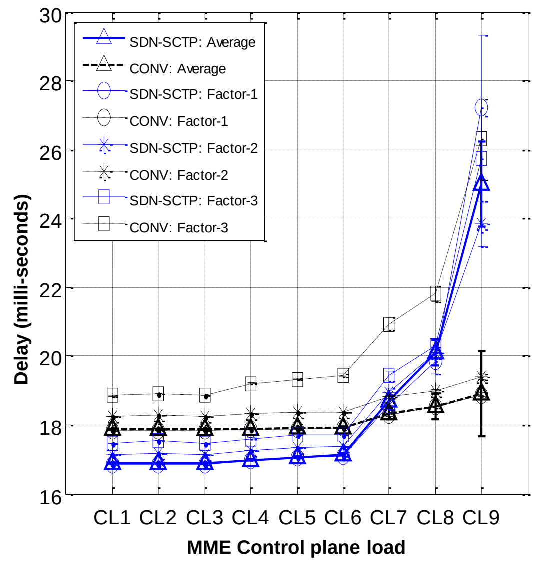


Figure 7-7 Mob-D in combination 0 and combination 2

(E2ED versus nominal control plane load at all fractional factor levels)

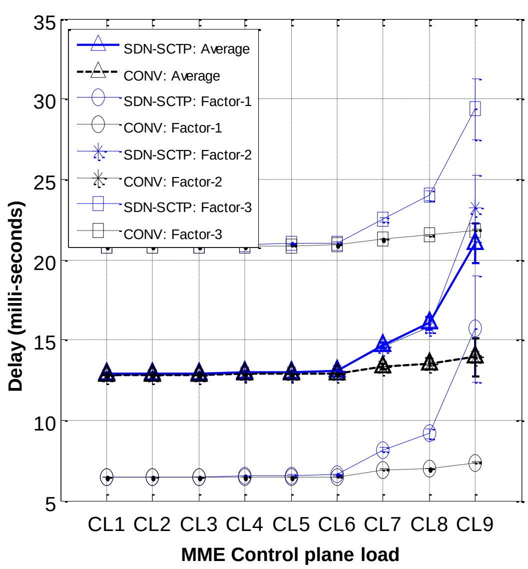


Figure 7-8: Mob-C in combination 0 and combination 3

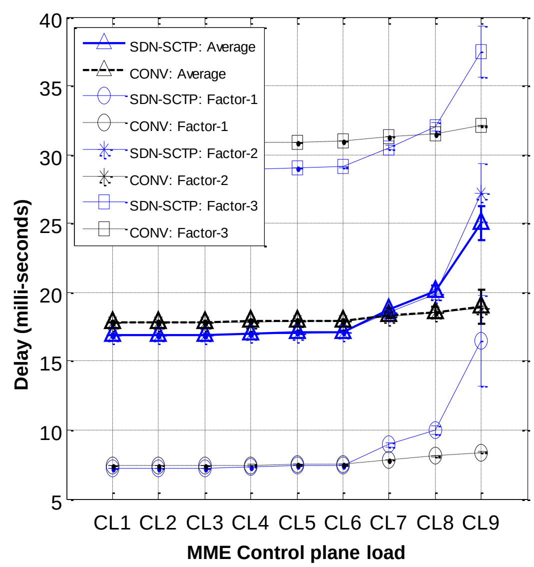
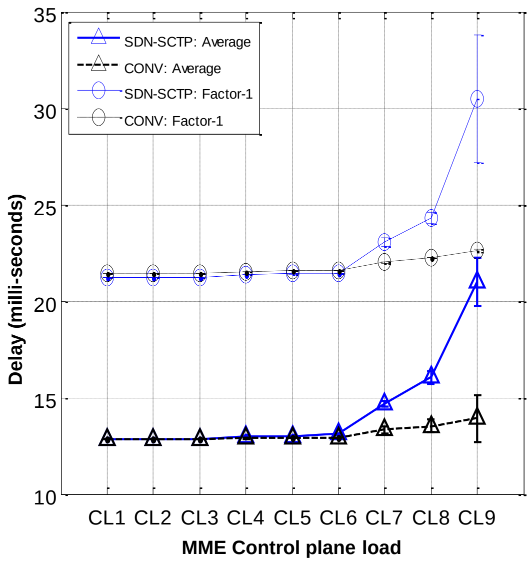


Figure 7-9: Mob-D in combination 0 and combination 3

(E2ED versus nominal control plane load at all fractional factor levels)



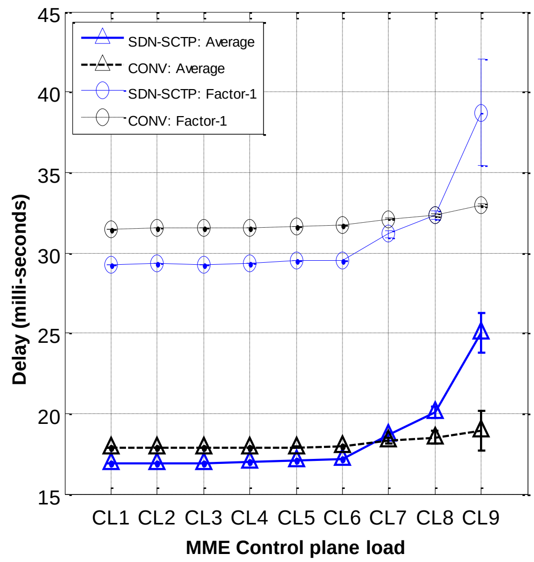
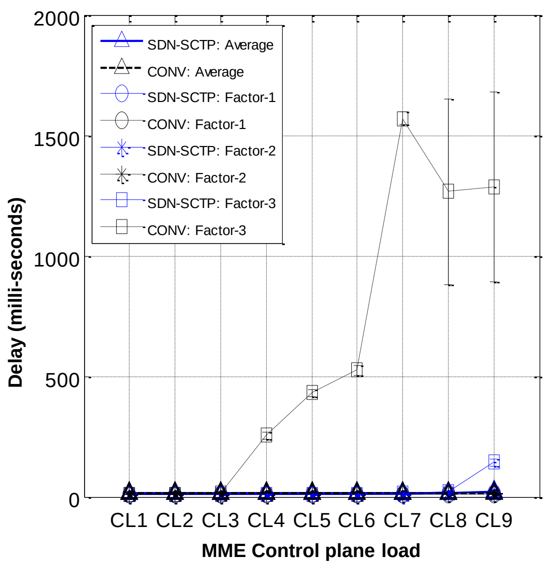


Figure 7-10: Mob-C in combination 0 and combination 4

Figure 7-11: Mob-D in combination 0 and combination 4

(E2ED versus nominal control plane load at all fractional factor levels)



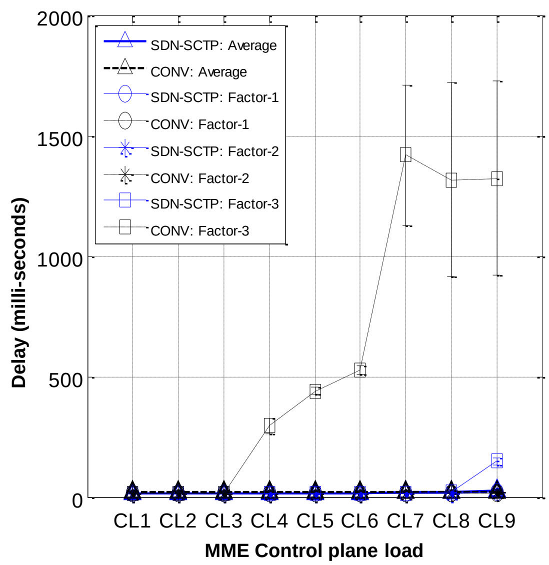
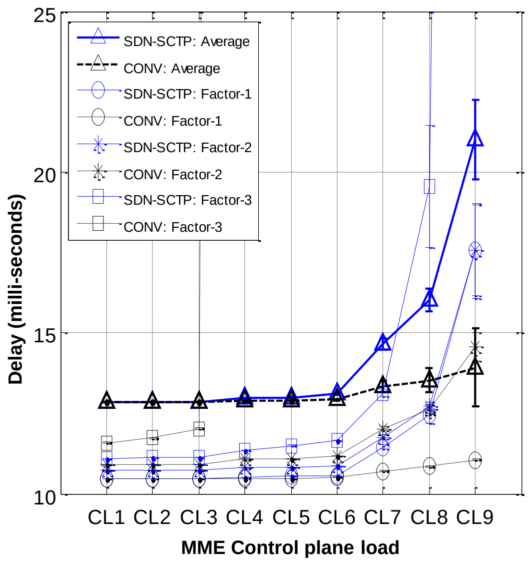


Figure 7-12: Mob-C in combination 0 and combination 22

Figure 7-13: Mob-D in combination 0 and combination 22

(E2ED versus nominal control plane load at all fractional factor levels)



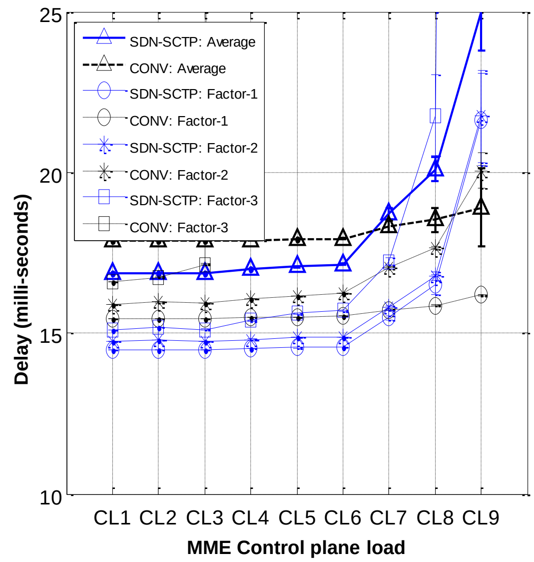
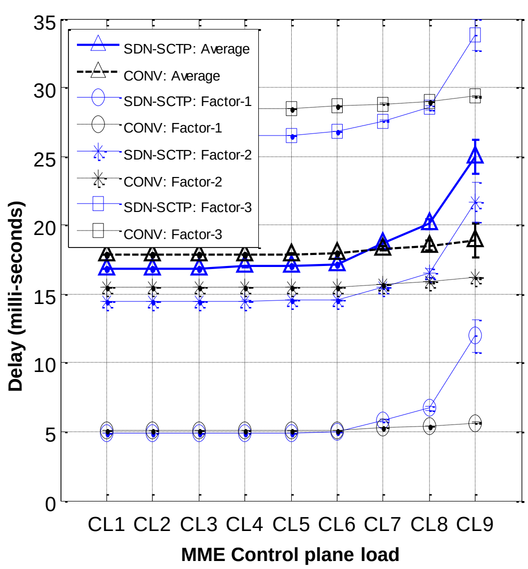


Figure 7-14: Mob-C in combination 0 and combination 22 (zoomed in)

Figure 7-15: Mob-D in combination 0 and combination (zoomed in)

(E2ED versus nominal control plane load at all fractional factor levels)



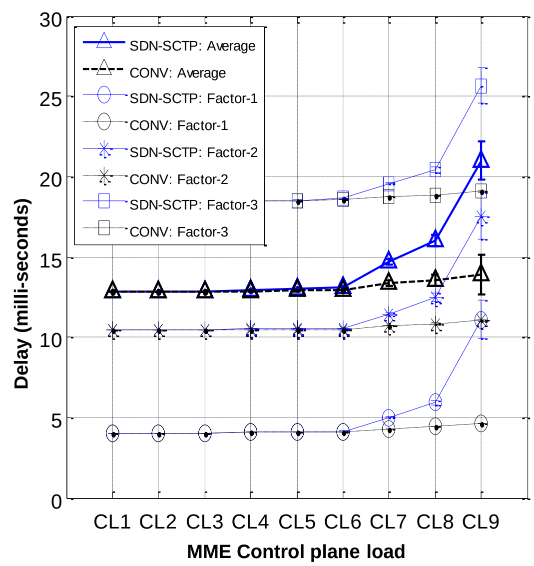


Figure 7-16: Mob-C in combination 0 and combination 32

Figure 7-17: Mob-D in combination 0 and combination 32

(E2ED versus nominal control plane load at all fractional factor levels)

7.2.2 MME LINKS BANDWIDTH UTILIZATION

This subsection refers to bandwidth utilization of MME physical links that are used in simulation setup, as shown in Figure 5-1. Both SDN-EPC variants have an effect on utilization of MME links where in SDN-EPC there is an increase in number of packets exchanged amongst EPC gateways. This effect is measured relative to CONV-EPC bandwidth utilization. Table 7-6 shows the increase of bandwidth utilization of MME-R1 link in all simulations scenarios, both SDN-EPC variants led to an increase on average of 52% and 33% for SDN-SCTP and SDN-UDP respectively relative to CONV-EPC.

Table 7-7 shows the increase of bandwidth utilization of MME-SGW link in all simulations scenarios, both SDN-EPC variants led to an increase on average of 158% and 75% for SDN-SCTP and SDN-UDP respectively relative to CONV-EPC. It is noted that the increase is much higher at the MME-SGW link, because most communication on this link is between MME and other core nodes, i.e. core PGW and core SGW, whereas MME-R1 is used for communication with local EPC nodes, i.e. local SGW and PGW, and as well eNBs. Traffic generated for communication with eNBs is not affected by SDN-EPC network relative to CONV-EPC thus the increase in MME-R1 is less than that of MME-SGW.

However, even with this high percentage of increase in bandwidth utilization, it is only relative to CONV-EPC values. The actual bandwidth consumed in all SDN-EPC variants reached at its maximum level 6 MBytes/seconds, which is greatly less than 1% utilization of the 10Gbps link. This concludes that even with downside of increased number of packets and packets overhead, the differentiation between SCTP and UDP variants for an SDN-EPC based on link utilization is negligible.

Table 7-6: MME-R1 LINK BANDWIDTH UTILIZATION CROSS SCENARIOS FOR ALL COMBINATIONS

SCENARIOS	**RELATIVE VARIATION FROM CONV-EPC**	
	SDN-SCTP EPC	**SDN-UDP EPC**
MOB-A	**+55%**	**+36%**
MOB-B	**+54%**	**+34%**
MOB-C	**+52%**	**+33%**
MOB-D	**+50%**	**+31%**

Table 7-7: MME-SGW LINK BANDWIDTH UTILIZATION CROSS SCENARIOS FOR ALL COMBINATIONS

SCENARIOS	**RELATIVE VARIATION FROM CONV-EPC**	
	SDN-SCTP EPC	**SDN-UDP EPC**
MOB-A	**+139%**	**+86%**
MOB-B	**+168%**	**+108%**
MOB-C	**+148%**	**+93%**
MOB-D	**+180%**	**+117%**

7.2.3 MME CPU UTILIZATION

As mentioned previously, both conventional and SDN-EPC will serve the exact same number of requests at each control operations settings but the effect on loading the

MME is different, as some control operations are now performed in MME for SDN-EPC instead of SGW as in CONV-EPC. It should be reminded that SDN-SCTP and SDN-UDP are not to have any difference in MME CPU utilization hence both process the same GTP messages in the application layer. However the difference in the process between them lies in the transport protocol stack, i.e. SCTP versus UDP protocol stack, which are not included in the processing resource modules of this study, as mentioned in the simulation setup assumption section 5.7.

MME CPU utilization is coupled with number of requests being served and the amount of resources required for each request. Thus it is expected that the MME CPU utilization profile is identical across combinations at the same control load configuration, except when request rate is affected which might happen if different configurations induces that. Therefore we choose combination 2 and 22 for viewing the ratio of the utilization for SDN-EPC compared to CONV-EPC at each control load points.

Figure 7-18 and Figure 7-19 show the ratio of MME CPU utilization for SDN-EPC compared to CONV-EPC in combinations 2 and 22; the figures indicate the following:

- MME resources requirement in SDN-EPC ranges between 10%-17% and 10%-15% more than that in CONV-EPC in combination 2 and 22 respectively across GWPT levels, except for high GWPT level in combination 22 which starts at 14% at CL1 and reaches 50% at CL9.
- The highest ratio of increase in required MME resource occurs when the CORR rate is lowest for the same CBGT rate, this occurs for points CL1, CL4, and CL7; the least

ratio of increase occurs when the CORR rate is highest for the same CBGT rate, this occurs for points CL3, CL6, and CL9.

- In general, the utilization ratio is almost identical for different GWPT levels, except in combination 22 at GWPT of 150 usec. There is a surge in this ratio when GWPT level is configured to 150 usec in combination 22. The increase reaches 50% of that for CONV-EPC at highest control load points. This behavior, as shall be seen in following results, is reflected due to the fact that there is a decrease in requests arrival at the MME at high GWPT which is caused by congestion at the new SGW. Therefore the ratio has increased since SDN-EPC is actually processing more requests than CONV-EPC.

These ratios indicate how much additional resources are required in SDN-EPC relative to CONV-EPC, and according to the first observation, at least 10% more resources are required for SDN-EPC MME operations and this ratio increases when CORR is at its lowest level at 5%, and decreases when CORR is at its highest level at 15%. This behavior occurs due to the fact that large part of the S1-handover mobility procedure, which is controlled by CORR rate, contains communication with new SGW only and eNBs only. This means no message exchange between new SGW and anchor PGW in that part, and GTP-C messages in both EPC networks for communication with new SGW and eNBs were largely identical because functions they support are independent of PGW operations. Thus this leads that increasing CORR rate increases number of operations that require same resources which lead to decreasing the difference between SDN-EPC and CONV-EPC resource utilization.

Figure 7-20 to Figure 7-22 show this actual MME CPU utilization versus control plane load. It is noticed that the gap between utilization of both architectures is increased with increasing load which is expected due to increased requests arrival. The difference between SDN-EPC and CONV-EPC under low control plane load is approximately 3% of MME-CPU capacity, and it increases under very high control loads to reach 15% of MME CPU capacity.

It is also noted that for combination 22 there is a decrease in MME utilization curve at high GWPT level relative to other GWPT levels, as shown in Figure 7-22. This occurs because of number of requests arriving at the MME from SGW and PGW gateways is decreased from the typical request rate, that indicates that MME capacity is NOT the reason for the performance bottleneck on the system occurring at this level as indicated by E2ED curves in previous sections, and this gives more indication that the bottleneck is occurring somewhere else in the system, as shall be presented in the next section is attributed to SGW resource saturation.

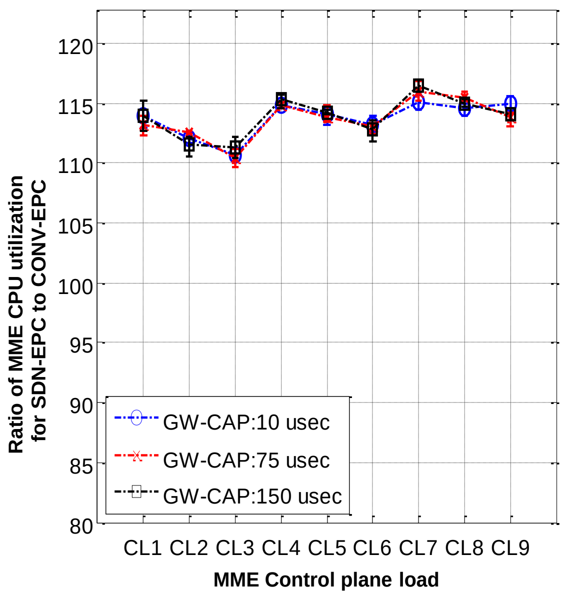


Figure 7-18: Ratio of MME CPU utilization for SDN-EPC to CONV-EPC in combination-2 as a function of control load

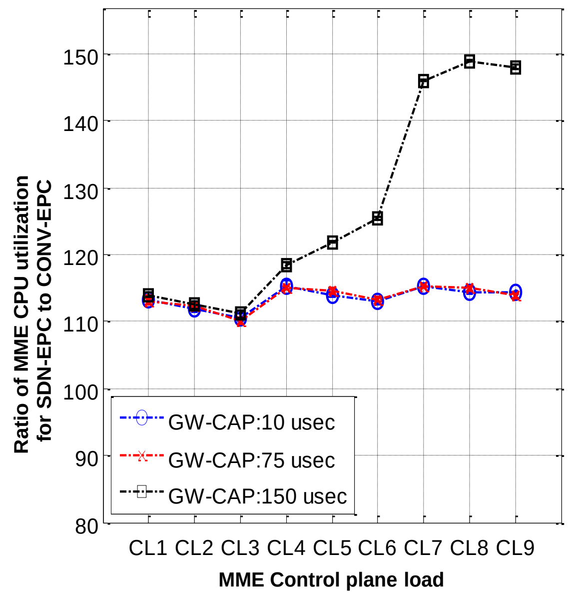


Figure 7-19: Ratio of MME CPU utilization for SDN-EPC to CONV-EPC in combination-22 as a function of control load

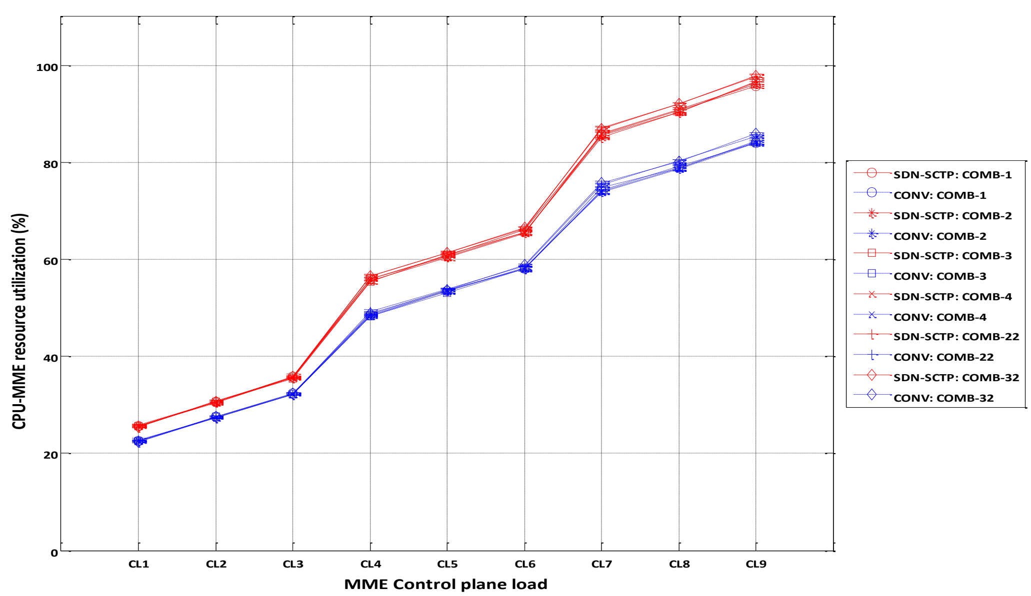


Figure 7-20: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios Mob-C at lowest fractional factors level

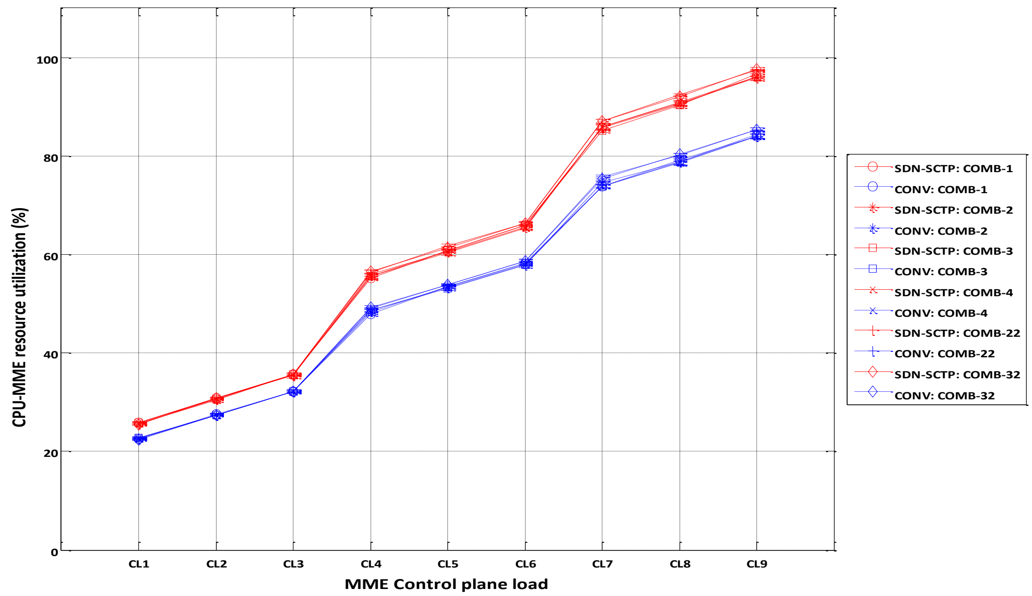


Figure 7-21: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios Mob-C at average fractional factors level

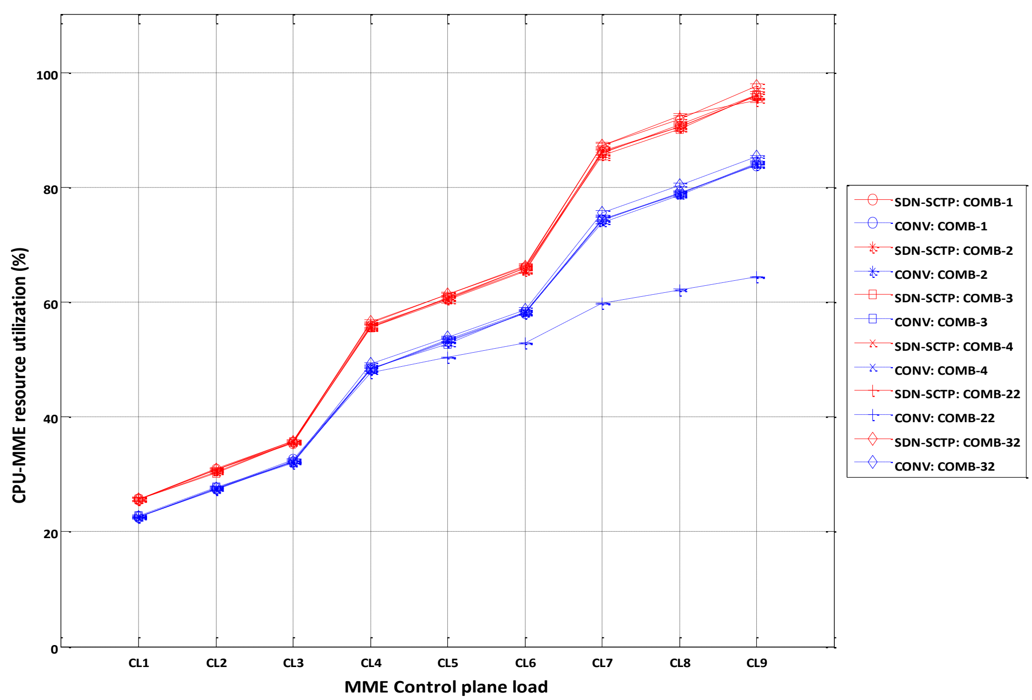


Figure 7-22: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios Mob-C at highest fractional factors level

7.2.4 CORE SGW CPU UTILIZATION

The effect of EPC gateways utilization can be observed in one of the SGWs deployed; either core SGW or local SGW. This refers to the network setup in Figure 5-1. It should be noted that both SDN-SCTP and SDN-UDP have the exact SGW CPU utilization in the results, as it is expected, and no distinction is made further between them.

Table 7-8 shows the average relative variation of Core SGW utilization from CONV-EPC. It is noticed that Mob-B and D have similar average utilization, whereas Mob-A and C have the same. This is attributed to the fact that in each of these pairs has the same role in mobility procedure scenarios which is either as an old SGW or a new SGW. There is a 44% decrease on average in SGW utilization in SDN-EPC compared to CONV-EPC. This reflects the amount of resources required in both architectures; these experiments show that SDN-EPC can reduce SGW gateway processing resources by 44% on average from resources required in CONV-EPC for control operations.

Table 7-8: AVERAGE RELATIVE VARIATION OF CORE SGW CPU UTILIZATION IN SDN-EPC ACROSS SCENARIOS

	Mob-A	**Mob-B**	**Mob-C**	**Mob-D**
RELATIVE VARIATION FROM CONV-EPC	**-44%**	**-47%**	**-44%**	**-47%**

SGW CPU utilization is coupled with number of requests being served and the amount of resources required for each request. Thus it is expected that the SGW CPU utilization profile is identical across combinations at the same control load configuration and same GWPT level. Therefore we choose combination 2 and 22 for viewing the ratio of the utilization for SDN-EPC compared to CONV-EPC at each control points. Figure 7-23 and Figure 7-24 show the ratio of core SGW CPU utilization for SDN-EPC compared to CONV-EPC in combinations 2 and 22. The figures indicate the following:

- Core SGW resources requirement in SDN-EPC ranges between 35%-50% less than that in CONV-EPC across GWPT levels. The exception for this is at high GWPT level in combination 22 which starts at 45% at CL1 and decreases to reach 8% at CL9.
- The resource utilization increases with increase in CORR level for the same CBGT level; this occurs at points CL3, CL6, and CL9. This behavior occurs due to the fact that large part of the S1-handover mobility procedure is for communication with new SGW only, and the same functions are performed in both SDN-EPC and CONV-EPC, therefore both consume the same resources and the reduction in required resources is decreased.
- In general, the utilization ratio is almost identical for different GWPT levels, except in combination 22 at GWPT of 150 usec. There is a surge in required resources for SDN-EPC when GWPT level is configured to 150 usec in combination 22. The ratio reaches 92% of that for CONV-EPC at highest control load points. This behavior is due to the fact that there is a decrease in requests arrival at the MME at high GWPT which is caused by the congestion at the new SGW. Therefore the ratio has increased since SDN-EPC is actually processing more requests than CONV-EPC.

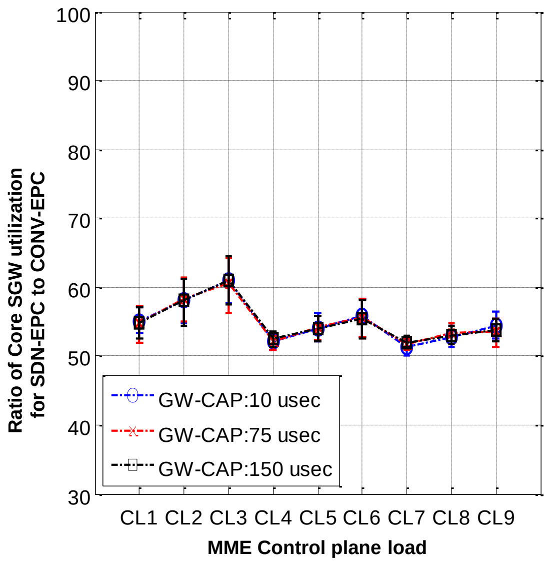


Figure 7-23: Ratio of core SGW CPU utilization for SDN-EPC to CONV-EPC in combination-2 as a function of control load

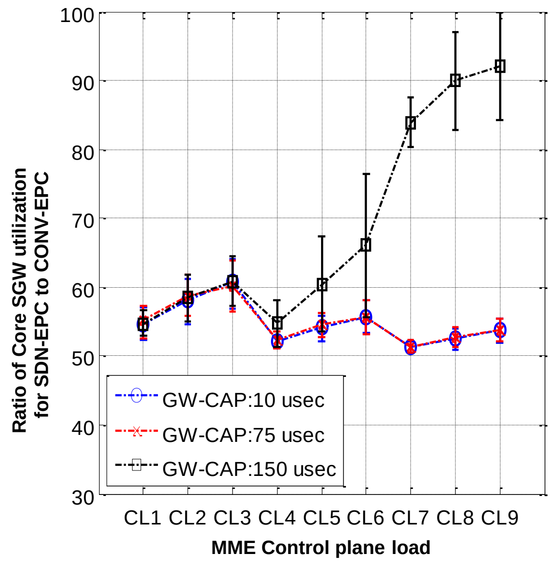


Figure 7-24: Ratio of core SGW CPU utilization for SDN-EPC to CONV-EPC in combination-22 as a function of control load

Figure 7-25 to Figure 7-29 show readings for core SGW CPU utilization for Mob-D at different fractional factors levels. It is clear that CPU utilization is increasing monotonically with control plane load as expected. The following can be deduced from the figures:

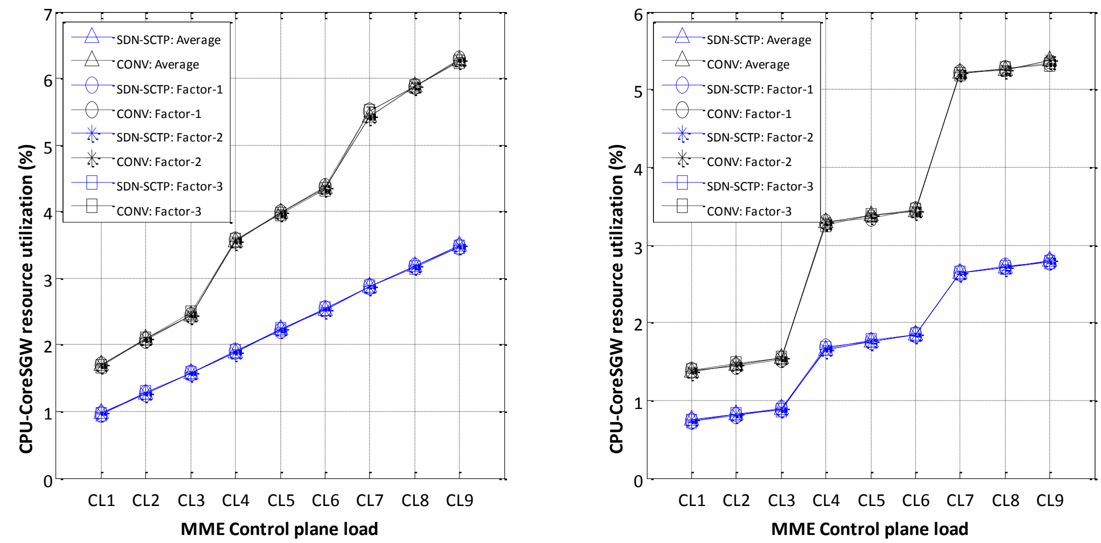
- The utilization at GWPT of 10 usec for all combinations and EPC network types never exceeded 10% of total resources; refer Figure 7-25, Figure 7-27, and Figure 7-29.
- The utilization at GWPT of 75 usec in SDN-EPC network reached at maximum control load to 50% and 96% of SGW CPU resources for combination 2 and 22 respectively, whereas CONV-EPC CPU utilization reached at maximum control load

to 50% and 95% of SGW CPU resources for combination 2 and 22 respectively; refer Figure 7-26 and Figure 7-28.

- The utilization at GWPT of 150 usec for combination 22 reached saturation, 100% utilization, under medium control load for CONV-EPC architecture. For SDN-EPC the utilization under highest control plane load reached 80% utilization for combination 22. For combination 2 the utilization for CONV-EPC reached 80% under high loads, whereas for SDN-EPC reached 40%; refer Figure 7-26 and Figure 7-28.

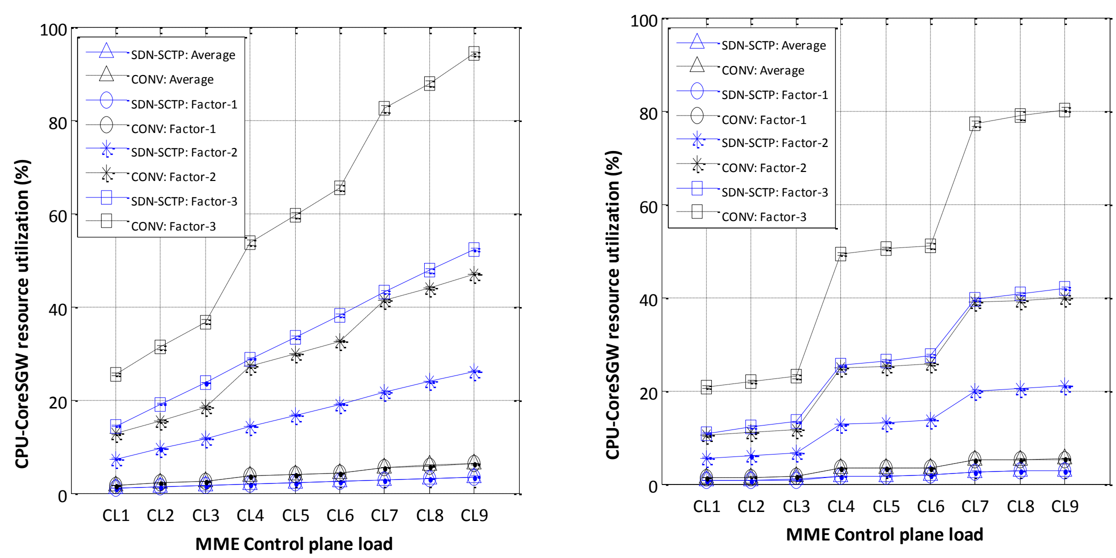
These results are consistent with previous results that SDN-EPC reduces utilization of EPC gateway resources, and thus SDN-EPC enables the usage of software-based gateways, whereas CONV-EPC were not able to handle large amount of requests using software-based gateways, although it does under medium control plane load. The use of software-based gateways (high GWPT level) in CONV-EPC leads to complete saturation of gateway resources under high control load (combination 22, CL4 and more) which, as observed before, contributed to degradation in performance of operation E2ED due to congestion happening at the gateways.

On the other hand, the improved software-based gateways (75 usec GWPT) were able to accommodate the high load of control plane, and it reached 50% and 80% utilization of the gateways capacity as such load for combination 2 and 22 respectively, whereas hardware-based gateways even under high loads do not exceed 10% utilization. This also indicates that further increase in control load might eventually lead to saturation, which stresses the constraints of using these gateway platforms and the importance of having careful engineering design of network dimensioning.



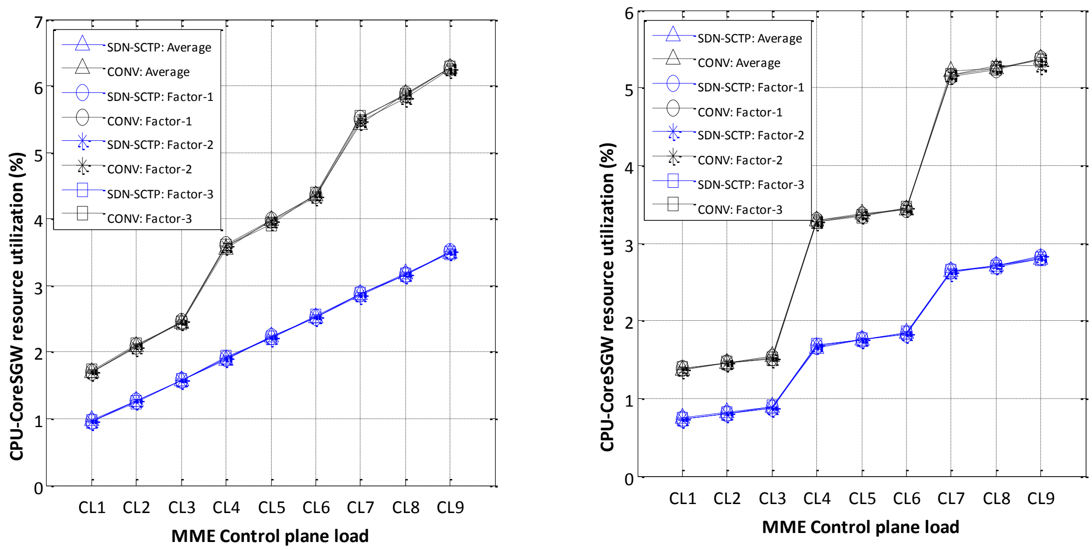
SCENARIO MOB-C SCENARIO MOB-D

Figure 7-25: CORE SGW CPU UTILIZATION versus nominal control plane load in combination 0 and combination 1 at all fractional factor levels



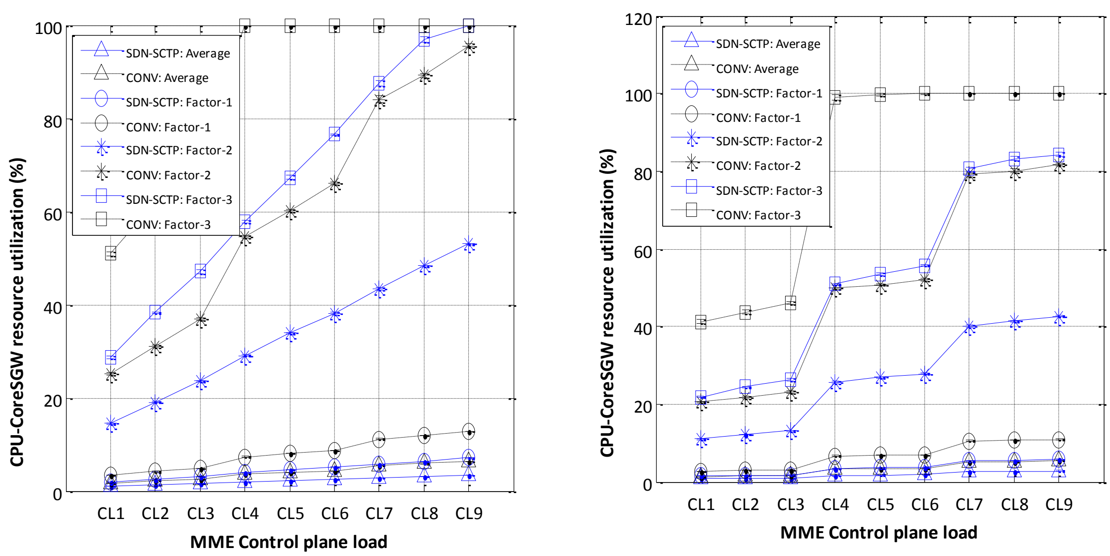
SCENARIO MOB-C SCENARIO MOB-D

Figure 7-26: CORE SGW CPU UTILIZATION versus nominal control plane load in combination 0 and combination 2 at all fractional factor levels



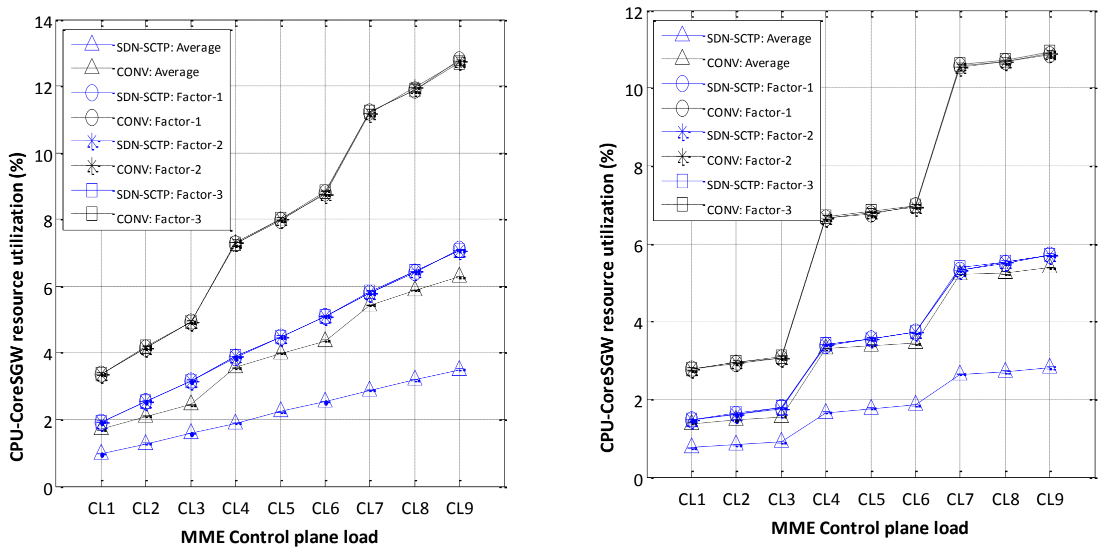
SCENARIO MOB-C SCENARIO MOB-D

Figure 7-27: CORE SGW CPU UTILIZATION versus nominal control plane load in combination 0 and combination 3 at all fractional factor levels



SCENARIO MOB-C SCENARIO MOB-D

Figure 7-28: CORE SGW CPU UTILIZATION versus nominal control plane load in combination 0 and combination 22 at all fractional factor levels



SCENARIO MOB-C SCENARIO MOB-D

Figure 7-29: CORE SGW CPU UTILIZATION versus nominal control plane load in combination 0 and combination 32 at all fractional factor levels

7.3 REGISTRATION OPERATION RESULTS

In [35], the recommended value for E2ED of registration procedure is 15 milli second for EPC processing delay that includes 10 milli-second for user information context retrieval from HSS, and 5 milli-seconds for message exchange between EPC nodes. In this recommendation study, a centralized deployment of EPC network is adopted. This means recommended delay values are subject to EPC network deployed in core centers rather than distributed in local centers. However, this simulation experiment does not include communication with HSS, which leave 5 milli-seconds of time for the complete procedure. Furthermore, this study investigates various EPC network deployment strategies between core and local centers that reflect multiple options a mobile network operator might adopt, represented in multiple registration procedure scenarios in the simulation experiment (Reg-1, Reg-2, etc.).

The average E2ED reported in this study is within range of 1 to 10 milli-seconds. Reg-1 scenario, which consists of core SGW and core PGW in the registration procedure scenario, recorded average E2ED within range of 1 to 6 milli-seconds under average control load. Whereas Reg-2, which constitutes of local SGW and core PGW in this registration procedure scenario, recorded average E2ED within range of 1 to 10 milli-seconds. E2ED results recorded for Reg-1 is mostly within range of [35] and in some cases slightly higher by 1 milli-second, Reg-1 also matches network deployment strategy used in [35]. Thus it is deemed that the factors configuration used in these experiments are appropriate enough for performance evaluation of registration procedure. Although Reg-2 results are clearly higher than those recommended values, this discrepancy is due to the fact of network deployment strategy used in this scenario that includes local centers for EPC nodes. Therefore, this represents the actual registration procedure performance under the same factor configuration used in Reg-1, which is appropriate for standardized evaluation.

In addition, It should be noted that recommended values in [35] are considered an ideal case which does not include factors such as background traffic and MME load condition, thus reported results in this study is considered a reflection of actual conditions.

7.3.1 E2ED PERFOMANCE METRIC

7.3.1.1 SIMULATION SCENARIOS

In this section, a detailed analysis of E2ED results for CONV-EPC and SDN-EPC networks is presented. Registration procedure E2ED results are presented at control load point CL5 at each fractional factor level for each simulation scenarios in the following pattern: (1) raw E2ED results; this is shown in Table 7-9. (2) Normalized E2ED results to the highest reported scenario of all scenarios in all EPC networks; this is shown in Table 7-10. For example, E2ED results for all scenarios in combination 1 at DBGT-1 are normalized to E2ED of CONV-EPC Reg-2 hence it reports the highest E2ED at this fractional factor level across all scenarios and EPC networks. (3) Normalized E2ED results to the highest reported scenario of all scenarios in the same EPC network; this is shown in Table 7-11. For example, E2ED results for SDN-SCTP in combination 1 at DBGT-1 are normalized to E2ED of SDN-SCTP Reg-4 hence it reports the highest E2ED at this fractional factor level across all scenarios in SDN-SCTP; the same is computed in SDN-UDP and CONV-EPC. (4) Normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC; this is shown in Table 7-12. For example, REg-1 in SDN-SCTP and SDN-UDP are both normalized to Reg-1 in CONV-EPC, the same is computed for the other scenarios.

From these E2ED results tables, we deduce the following findings divided into sections: (1) scenarios E2ED performance order across all EPC network types at the same fractional factor level; based on Table 7-10. (2) Scenarios E2ED performance order within the same EPC network type; based on Table 7-11. (3) Comparison of E2ED results for SDN-SCTP and SDN-UDP EPC network types; based on Table 7-10. (4) E2ED performance of SDN-EPC compared to CONV-EPC network; based on Table 7-12. (5) E2ED performance at each fractional factor level and factors contribution to variation; based on Table 7-9. Finally, (6) the effect of increasing MME processing capacity is analyzed; based on Table 7-9.

Table 7-9: E2ED results (milli-seconds) at CL5 in each EPC network – registration procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4
1	Comb-1 DBGT-1	3.61	5.63	4.62	4.62	3.68	4.67	4.68	4.69	3.61	4.6	4.61	4.61
2	Comb-1 DBGT-2	3.62	5.63	4.62	4.62	3.68	4.67	4.68	4.69	3.61	4.6	4.61	4.62
3	Comb-1 DBGT-3	3.62	5.64	4.63	4.62	3.69	4.68	4.69	4.7	3.62	4.6	4.61	4.62
4	Comb-2 GWPT-1	3.61	5.62	4.61	4.61	3.68	4.67	4.68	4.68	3.61	4.6	4.6	4.61
5	Comb-2 GWPT-2	3.83	5.84	4.84	4.83	3.75	4.74	4.75	4.76	3.68	4.67	4.67	4.68
6	Comb-2 GWPT-3	4.2	6.21	5.23	5.2	3.84	4.83	4.84	4.85	3.76	4.76	4.77	4.77
7	Comb-3 DEL-1	2.01	2.42	2.21	2.21	2.08	2.28	2.28	2.28	2.02	2.21	2.21	2.21
8	Comb-3 DEL-2	3.61	5.62	4.61	4.61	3.68	4.67	4.68	4.69	3.61	4.6	4.6	4.61
9	Comb-3 DEL-3	5.61	9.62	7.62	7.62	5.68	7.68	7.69	7.7	5.61	7.61	7.62	7.63
10	Comb-22 GWPT-1	2.82	4.83	3.83	3.83	2.85	3.85	3.85	3.86	2.81	3.81	3.82	3.83
11	Comb-22 GWPT-2	3.11	5.14	4.14	4.11	2.92	3.92	3.93	3.93	2.89	3.89	3.89	3.9
12	Comb-22 GWPT-3	159.2	169.52	151.88	165.6	3.05	4.05	4.05	4.06	3.02	4.02	4.02	4.02
13	Comb-32 DEL-1	1.22	1.63	1.43	1.43	1.25	1.44	1.45	1.45	1.21	1.41	1.41	1.42
14	Comb-32 DEL-2	2.82	4.83	3.83	3.83	2.85	3.85	3.85	3.86	2.81	3.81	3.82	3.83
15	Comb-32 DEL-3	4.82	8.83	6.83	6.83	4.85	6.85	6.85	6.86	4.81	6.81	6.82	6.83
16	Comb-4	5.61	9.62	7.62	7.62	5.68	7.68	7.69	7.7	5.61	7.61	7.62	7.63

Table 7-10: E2ED results normalized to highest E2ED at each fractional factor configuration in each EPC network in all scenario at CL5 – registration procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4
1	Comb-1 DBGT-1	64.21	100	82.04	82.01	65.45	83.06	83.1	83.35	64.23	81.71	81.89	82.01
2	Comb-1 DBGT-2	64.18	100	82.04	81.97	65.38	82.95	83.02	83.24	64.13	81.71	81.77	81.98
3	Comb-1 DBGT-3	64.17	100	82.09	81.99	65.43	82.98	83.11	83.3	64.21	81.66	81.82	82.01
4	Comb-2 GWPT-1	64.23	100	82.05	82.02	65.49	83.01	83.13	83.29	64.19	81.75	81.81	82.03
5	Comb-2 GWPT-2	65.53	100	82.85	82.73	64.19	81.13	81.25	81.41	63.02	79.87	79.97	80.16
6	Comb-2 GWPT-3	67.53	100	84.1	83.66	61.74	77.73	77.84	78.01	60.53	76.57	76.69	76.76
7	Comb-3 DEL-1	83.31	100	91.63	91.6	86.26	94.23	94.53	94.49	83.38	91.3	91.48	91.63
8	Comb-3 DEL-2	64.24	100	82.05	82.02	65.49	83.05	83.17	83.32	64.22	81.79	81.88	82.06
9	Comb-3 DEL-3	58.33	100	79.22	79.2	59.05	79.86	79.91	80.06	58.3	79.13	79.18	79.32
10	Comb-22 GWPT-1	58.44	100	79.28	79.25	58.99	79.7	79.78	79.93	58.25	78.96	79.02	79.18
11	Comb-22 GWPT-2	60.41	100	80.48	79.84	56.8	76.28	76.34	76.46	56.13	75.62	75.67	75.8
12	Comb-22 GWPT-3	93.91	100	89.59	97.69	1.8	2.39	2.39	2.39	1.78	2.37	2.37	2.37
13	Comb-32 DEL-1	75.16	100	87.61	87.54	76.74	88.72	88.9	89.11	74.61	86.54	86.71	86.96
14	Comb-32 DEL-2	58.44	100	79.28	79.25	58.99	79.7	79.78	79.93	58.25	78.96	79.02	79.18
15	Comb-32 DEL-3	54.64	100	77.36	77.34	54.94	77.58	77.62	77.75	54.53	77.16	77.2	77.32
16	Comb-4	58.33	100	79.22	79.2	59.05	79.86	79.91	80.06	58.3	79.13	79.18	79.32

Table 7-11: E2ED results normalized to highest E2ED within same EPC network at each fractional factor configuration at CL5 – registration procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4
1	Comb-1 DBGT-1	64.21	100	82.04	82.01	78.53	99.65	99.71	100	78.32	99.63	99.86	100
2	Comb-1 DBGT-2	64.18	100	82.04	81.97	78.54	99.65	99.74	100	78.23	99.67	99.75	100
3	Comb-1 DBGT-3	64.17	100	82.09	81.99	78.55	99.62	99.78	100	78.29	99.57	99.77	100
4	Comb-2 GWPT-1	64.23	100	82.05	82.02	78.63	99.67	99.81	100	78.25	99.65	99.73	100
5	Comb-2 GWPT-2	65.53	100	82.85	82.73	78.85	99.66	99.8	100	78.62	99.64	99.76	100
6	Comb-2 GWPT-3	67.53	100	84.1	83.66	79.15	99.64	99.78	100	78.86	99.75	99.91	100
7	Comb-3 DEL-1	83.31	100	91.63	91.6	91.24	99.68	100	99.95	90.99	99.64	99.84	100
8	Comb-3 DEL-2	64.24	100	82.05	82.02	78.6	99.67	99.82	100	78.27	99.68	99.78	100
9	Comb-3 DEL-3	58.33	100	79.22	79.2	73.76	99.75	99.81	100	73.5	99.76	99.82	100
10	Comb-22 GWPT-1	58.44	100	79.28	79.25	73.8	99.72	99.82	100	73.56	99.71	99.79	100
11	Comb-22 GWPT-2	60.41	100	80.48	79.84	74.28	99.76	99.84	100	74.04	99.76	99.83	100
12	Comb-22 GWPT-3	93.91	100	89.59	97.69	75.15	99.8	99.85	100	75.02	99.81	99.89	100
13	Comb-32 DEL-1	75.16	100	87.61	87.54	86.11	99.56	99.76	100	85.8	99.52	99.72	100
14	Comb-32 DEL-2	58.44	100	79.28	79.25	73.8	99.72	99.82	100	73.56	99.71	99.79	100
15	Comb-32 DEL-3	54.64	100	77.36	77.34	70.66	99.78	99.83	100	70.53	99.79	99.84	100
16	Comb-4	58.33	100	79.22	79.2	73.76	99.75	99.81	100	73.5	99.76	99.82	100

Table 7-12: E2ED results normalized to highest E2ED within same EPC network at each fractional factor configuration at CL5 – registration procedures

		CONV-EPC				SDN-SCTP				SDN-UDP			
		Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4	Reg-1	Reg-2	Reg-3	Reg-4
1	Comb-1 DBGT-1	100	100	100	100	101.93	83.06	101.29	101.62	100.04	81.71	99.82	99.99
2	Comb-1 DBGT-2	100	100	100	100	101.87	82.95	101.2	101.55	99.92	81.71	99.68	100.02
3	Comb-1 DBGT-3	100	100	100	100	101.96	82.98	101.24	101.6	100.06	81.66	99.67	100.03
4	Comb-2 GWPT-1	100	100	100	100	101.97	83.01	101.31	101.55	99.94	81.75	99.71	100.01
5	Comb-2 GWPT-2	100	100	100	100	97.95	81.13	98.06	98.41	96.17	79.87	96.52	96.89
6	Comb-2 GWPT-3	100	100	100	100	91.43	77.73	92.56	93.24	89.64	76.57	91.19	91.75
7	Comb-3 DEL-1	100	100	100	100	103.54	94.23	103.17	103.16	100.08	91.3	99.84	100.03
8	Comb-3 DEL-2	100	100	100	100	101.95	83.05	101.37	101.59	99.98	81.79	99.8	100.05
9	Comb-3 DEL-3	100	100	100	100	101.23	79.86	100.86	101.08	99.95	79.13	99.94	100.15
10	Comb-22 GWPT-1	100	100	100	100	100.93	79.7	100.63	100.86	99.66	78.96	99.67	99.92
11	Comb-22 GWPT-2	100	100	100	100	94.02	76.28	94.86	95.76	92.91	75.62	94.03	94.94
12	Comb-22 GWPT-3	100	100	100	100	1.92	2.39	2.67	2.45	1.9	2.37	2.65	2.43
13	Comb-32 DEL-1	100	100	100	100	102.1	88.72	101.46	101.79	99.27	86.54	98.97	99.33
14	Comb-32 DEL-2	100	100	100	100	100.93	79.7	100.63	100.86	99.66	78.96	99.67	99.92
15	Comb-32 DEL-3	100	100	100	100	100.54	77.58	100.33	100.53	99.81	77.16	99.79	99.97
16	Comb-4	100	100	100	100	101.23	79.86	100.86	101.08	99.95	79.13	99.94	100.15

7.3.1.1.1 SCENARIOS PERFORMANCE ORDER ACROSS ALL EPC NETWORK TYPES

Table 7-10 shows normalized E2ED results to the highest reported scenario of all scenarios in all EPC networks. This is used to show E2ED performance order among all EPC networks at the same fractional factor level as indicated by the percentages. After further comparison between all simulations combinations, it is found that this order is the same across different fractional factor levels, meaning regardless of simulation combination and fractional factor level the order does not change, and they are (excluding combination 22 with high GWPT level):

- CONV-EPC Reg-2 as highest and the highest E2ED is recorded in combination 4 with 9.62 msec; refer Table 7-10 row 16.
- SDN-EPC Reg-2 as second highest with 78-94% ratio to highest across combinations; refer Table 7-10 rows 15 and 10.
- SDN-EPC and CONV-EPC Reg-3 and Reg-4 with 77-94% ratio to highest across combinations; refer Table 7-10 rows 15 and 10.
- SDN-EPC and CONV-EPC Reg-1 with 55-86% ratio to highest across combinations; refer Table 7-10 rows 15 and 10. The least E2ED is recorded in combination 32 at 1.22 msec with low DEL level shown Table 7-10 row 13.

It is noted E2ED in CONV-EPC combination 22 had the same behavior as one encountered in mobility procedure, and for the same reason justified previously which is congestion at respective SGW as results shall present in following section.

7.3.1.1.2 SCENARIOS PERFORMANCE ORDER WITHIN SAME EPC NETWORK TYPES

Table 7-11 shows normalized E2ED results to the highest reported scenario of all scenarios in the same EPC network. This is used to deduce E2ED performance order within the same EPC networks at the same fractional factor. The E2ED relative performance order within the same EPC network type has been found in SDN-EPC as Reg-2, Reg-3, and Reg-4 equally as the highest average E2ED followed by Reg-1 as lowest average E2ED; in CONV-EPC Reg-2 with highest average E2ED followed by Reg-3 and Reg-4 having equal average E2ED, and as lowest average is E2ED Reg-1. The relative percentage E2ED for each registration procedure scenario relative to Reg-2 scenario (E2ED at each scenario divided by E2ED in Reg-2 in the same EPC network type) is:

- In combination 1: For SDN-EPC is approximately 100%, 100%, and 79%, and for CONV-EPC is approximately 82%, 82%, and 64%, for Reg-3, Reg-4, and Reg-1 respectively; refer Table 7-11 rows 1, 2, and 3.
- In combination 2: For SDN-EPC is approximately 100%, 100%, and 79%, and for CONV-EPC is approximately 82%, 82%, and 64-67%, for Reg-3, Reg-4, and Reg-1 respectively; refer Table 7-11 rows 4, 5, and 6.
- In combination 22: For SDN-EPC is approximately 100%, 100%, and 74%, and for CONV-EPC is approximately 80%, 80%, and 60%, for Reg-3, Reg-4, and Reg-1 respectively, excluding results for combination 22 at high GWPT level; refer Table 7-11 rows 7, 8, and 9.
- In combination 3: For SDN-EPC is approximately 100%, 100%, and 73-91%, and for CONV-EPC is approximately 80-91%, 80-91%, and 59-83%, for Reg-3, Reg-4, and Reg-1 respectively, where the difference in E2ED between Reg-2 and other scenarios increases with increasing DEL level; refer Table 7-11 rows 10, 11, and 12.
- In combination 32: For SDN-EPC is approximately 100%, 100%, and 70-85%, and for CONV-EPC is approximately 77-87%, 77-87%, and 55-75%, for Reg-3, Reg-4, and Reg-1 respectively, where the difference in E2ED between Reg-2 and other scenarios increases with increasing DEL level; refer Table 7-11 rows 13, 14, and 15.

- In combination 4: For SDN-EPC is approximately 100%, 100%, and 74%, and for CONV-EPC is approximately 79%, 79%, and 58%, for Reg-3, Reg-4, and Reg-1 respectively; refer Table 7-11 row 16.

The results show that E2ED relative performance of simulation scenario within the same EPC network type is always the same; meaning the order of highest to lowest scenario is maintained regardless of EPC network type and parameter combination; however the percentage of variation within EPC network types is different for SDN-EPC from that for CONV-EPC. The similarity in E2ED relative performance among registration procedure scenarios in EPC network types can be explained by the distance between the MME and EPC gateways, i.e. SGW and PGW, participating in the registration control operation and route of this message exchange for this procedure; for cases where route travelled is farthest between these EPC entities, i.e. Reg-2 in CONV-EPC, the E2ED average value is highest, and the opposite is true, when route travelled is least the overall average E2ED value is lowest which is observed in cases Reg-1 for all EPC types.

It is reminded that in Reg-2 the route for CONV-EPC follows the trombone route which is higher than that for SDN-EPC that does not have this trombone feature, thus Reg-2 E2ED for SDN-EPC is lower than that of CONV-EPC, whereas Reg-3 and Reg-4 have similar E2ED performance which is governed by location of farthest gateway, which is local PGW, even though in Reg-3 core SGW is involved, the fact that one of the gateways is farther than the other will determine the expected E2ED performance.

7.3.1.1.3 PERFORMANCE BETWEEN DIFFERENT SDN-EPC NETWORK TPYES

Table 7-12 shows normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC. This is used to compare E2ED results between SDN-SCTP and SDN-UDP at each single same scenario and same fractional factor combinations. The results show that difference in E2ED between SDN-SCTP and SDN-UDP is at most cases 1%-2% or less, which indicates no perceivable difference between SDN-SCTP EPC and SDN-UDP EPC E2ED results. In addition, both types have could be deemed with the same E2ED performance in the same scenario, and indicates that the performance of both SDN-EPC types is comparable under same DBGT and DEL conditions. This small difference is attributed for the same justification provided previously those communication links have abundant capacity whereas control packets have small sizes leading to no variation in the E2ED results.

7.3.1.1.4 PERFORMANCE OF SDN-EPC NETWORK COMPARED TO CONV-EPC

Table 7-12 shows normalized E2ED results of SDN-SCTP and SDN-UDP to the respective scenario in CONV-EPC at each fractional factor level as indicated by the percentages. Based on previous observation in 7.3.1.1.3 that SDN-SCTP and SDN-UDP have almost identical E2ED results, same or within 1% difference, in this section SDN-EPC is referred as the representative of both and the average E2ED of their results is reported. From this view we can deduce the difference in E2ED performance between EPC networks, and the following is observed:

- In scenarios Reg-1, Reg-3, and Reg-4, E2ED results of SDN-EPC are mainly 1% to 2% from those of CONV-EPC at the same scenario and fractional factor level regardless of SDN-EPC type. There are however small variations in combinations 2, 3, 22, and 32 as follows:

- In combination 3 at low DEL level there is an increase of 4% and it translates to only 0.08 msec which can be deemed as unimportant; refer Table 7-12 row 10.
- In combination 2 at average and high GWPT level and combination 22 at average GWPT level, there is a decrease in E2ED by 2-9% and it translates to 0.36 msec at maximum difference. Thus we deem this variation as important hence the small difference is not significant for the current configurations. However, this difference potentially increases with increasing MME capacity further than configured levels; refer Table 7-12 rows 6, 7, and 8.
- In combination 22 at high GWPT level, there is a severe degradation in E2ED results for reasons already justified in previous experiment; refer Table 7-12 row 12.
- SDN-EPC in both its types in scenarios Reg-1, Reg-3, and Reg-4 are similar in E2ED performance to that for CONV-EPC with GWPT configured to 10 usec; refer Table 7-12 rows 4 and 7.

- In scenario Reg-2:

- In combination 1, E2ED results are 83% of CONV-EPC E2ED and the difference is approximately 1 msec at any DBGT level; refer Table 7-12 rows 1, 2, and 3.
- In combination 2, E2ED results are 78-83% of CONV-EPC E2ED, where the difference it is increasing with increasing GWPT levels and the difference is approximately 1-1.4 msec; refer Table 7-12 rows 4, 5, and 6.

- In combination 22, E2ED results are 77-80% of CONV-EPC E2ED, where the difference it is increasing going from GWPT-1 to GWPT-2 with the difference as approximately 1-1.2 msec. GWPT-3 leads to unstable network as shown previously. refer Table 7-12 rows 7, 8, and 9.
- In combination 3, E2ED results are 80-94% of CONV-EPC E2ED, where the difference is increasing with increasing DEL levels and the difference is approximately 0.25-2 msec increasing with DEL level increase; refer Table 7-12 rows 10, 11, and 12.
- In combination 32, E2ED results are 77-89% of CONV-EPC E2ED, where the difference is increasing with increasing DEL levels and the difference is approximately 0.2-2 msec increasing with DEL level increase; refer Table 7-12 rows 13, 14, and 15.
- In combination 4, E2ED results are 80% of CONV-EPC E2ED, where it is decreasing with increasing DEL levels and difference is 2 msec; refer Table 7-12 row 16.

The reason E2ED performance in Reg-2 is less in SDN-EPC than CONV-EPC reverts to the fact of eliminating trombone route in this scenario, as the EPC gateways involved are core PGW and local SGW which form a larger route in the case of CONV-EPC network compared to SDN-EPC.

In most simulation configurations where DEL level is configured to 0.5 msec the difference is 1 msec, when DEL level increases to 1.0 msec the difference increases to 2 msec, and when DEL is decreased to 0.1 msec the difference is minimized to 0.2 msec. This concludes that similar to Mob-D scenario the reduction in SDN-EPC E2ED is dependent on DEL level and increasing proportionally with it. The results show for each DEL level increase there is approximately 6% decrease in SDN-EPC compared to CONV-EPC.

At different GWPT levels the E2ED for SDN-EPC is less than CONV-EPC by approximately 0.2 msec on average. Combination 22 at high GWPT level show severe degradation in E2ED performance, which is an indication of congestion at the gateways as the case detected in mobility procedure. This also will be tackled in detail in later sections through E2ED results as a function of control load.

7.3.1.1.5 E2ED PERFORMANCE AT EACH FRACTIONAL FACTOR LEVEL AND FACTORS CONTRIBUTION TO VARIATION

To identify effect of each fractional factor in the simulation combinations in each EPC network, Figure 7-30 shows the ratio of E2ED at low fractional factor level to E2ED at average fractional factor level, and the ratio of E2ED at high fractional factor level to E2ED at average fractional factor level. The figure shows this ratio for all scenarios as an interval. The variation of each fractional factor effect across different scenarios is found to be minimal; this is indicated by the small interval in the figures. This emphasizes the fact that the effect of fractional factor is more dominant than the effect of gateway location distribution, similar to finding in mobility procedure.

- The effects of varying each fractional factor level based on results in the Figure 7-30, Figure 7-31, and Figure 7-32 are found to be:
  - DBGT has 0% change in E2ED across any of its level; no change was detected; refer “Comb-1” in respective figures.

- GWPT factor effect is:
    - In SDN-EPC there is limited variation due to change in GWPT level in combination 2 and 22, most variation is recorded to be 2-3% change from average GWPT level.
    - In CONV-EPC, the effect is dependent on system capacity (MME capacity). In combination 2 there is a change of 8% at high GWPT level from value at average GWPT level. In combination 22 there is a severe degradation between high GWPT and average GWPT reaching 35 times increase.
- DEL in combination 3 have 50% decrease in E2ED when DEL changes from 0.5 msec to 0.1 msec. It also has a 62% increase in E2ED when DEL changes from 0.5 msec to 1.0 msec. The same ratio is observed in all EPC network types.
- DEL in combination 32 have 61% decrease in E2ED when DEL changes from 0.5 msec to 0.1 msec. It also has a 76% increase in E2ED when DEL changes from 0.5 msec to 1.0 msec. The same ratio is observed in all EPC network types.

Factor DBGT had no perceivable impact at E2ED results in any EPC network type at any of its levels; refer COMB-1 in Figure 7-30, Figure 7-31, and Figure 7-32. The reason for this is as explained previously in mobility procedure that the links have very high data rate while intermediate nodes are not imposing any processing capacity thus any delay caused in the transport network will be very minimal contributed by queuing delays at intermediate routers.

The effect of DEL levels on E2ED is clearly evident from large variation between its different levels, and there is monotonic increase in E2ED values for every increase in DEL level. The increase in DEL from 0.1 msec to 0.5 msec leads to 1.8 to 2.3 and 2.3 to

2.9 times E2ED increase in combinations 3 and 32, respectively. The increase in DEL from 0.5 msec to 1.0 msec leads to 1.55 and 1.75 times E2ED increase in combinations 3 and 32, respectively. The relative variation in E2ED performance is similar regardless of EPC network type which indicates the effect of this factor is independent from the network type itself. Furthermore, this can serve as means for predicting E2ED performance using regression models for particular DEL values.

The effect of GWPT is observed to have similar effect such that in mobility procedure. At low and average GWPT levels there is a limited increase for SDN-EPC and slightly higher for CONV-EPC reaching 5-7%. At high GWPT level and high control load (combination 22), SDN-EPC was not affected, the increase is limited to 4%. But for CONV-EPC, there is an increase of around 35 times that from the average GWPT level. Thus it is concluded with the same conclusion in mobility procedure that GWPT factor does not contribute to E2ED under low and average GWPT in both architecture, but at high GWPT for CONV-EPC the E2ED suffers degradation in E2ED due to limited resources at the gateways, whereas for SDN-EPC the control operations are minimally affected by the GWPT level.

7.3.1.1.6 EFFECT OF INCREASING MME PROCESSING CAPACITY

The portion of enhancement of E2ED results by increasing the MME capacity depends on the relative contribution of processing delays caused by the MME at different levels, for example at combination 32 with 60 Mbps capacity at high DEL level E2ED is 4.82 msec where in combination 3 with 30 Mbps capacity at the same DEL level E2ED is 5.61 msec, thus doubling the MME capacity led to 14% decrease in overall E2ED. Following this observation the enhancement (reduction) of E2ED results by doubling the MME capacity is computed from combinations 2, 22, 3 and 32 as follows:

- For low DEL level (0.1 msec): reduction is 33% - 40%; refer Table 7-9 rows 10 and 13.
- For average DEL level (0.5 msec): reduction is 14% - 22%; refer Table 7-9 rows 11 and 14.
- For high DEL level (1.0 msec): reduction is 8% - 14%; refer Table 7-9 rows 12 and 15.
- For low GWPT level (10 usec): reduction is 14% - 22%; refer Table 7-9 rows 4 and 7.
- For average GWPT level (75 usec): reduction is 12% - 19%; refer Table 7-9 rows 5 and 8
- For high GWPT level (150 usec): reduction is 16% - 20% (for SDN-EPC only and excluding CONV-EPC results); refer Table 7-9 rows 6 and 9.

At DEL 0.1 msec, the reduction of E2ED by increasing MME processing capacity is highest since processing delay have highest contribution to E2ED than at higher DEL levels. As DEL level increases the reduction is decreased since the total contribution of processing delay gets smaller relative to overall E2ED value, and that aligns with previous results in mobility procedure.

The reduction in E2ED at different GWPT levels is noticed to be similar to those for average DEL value, which remind us that all GWPT experiments have DEL level configured to average DEL (0.5 msec). Thus it is concluded that the amount of enhancement in E2ED is mostly dependent on DEL value rather than GWPT level.

In combination 22, the same outcome as in mobility procedure occurred for CONV-EPC. Increasing MME capacity in combination 22 with high GWPT level leads to severe degradation of E2ED performance, which did not occur at GWPT-1 and GWPT-2 in combination 22. This is attributed to high utilization of gateway resources at higher processing time in the gateways.

7.3.1.1.7 PRIMARY FACTORS CONTRIBUTION ORDER

The findings in registration operation regarding fractional factor aligns with findings from mobility procedure. These results show the primary factors that impact E2ED response, and these are in descending order: DEL, and MME-CAP; while factor GWPT had minimal effect in general and factor DBGT had no impact on E2ED values.

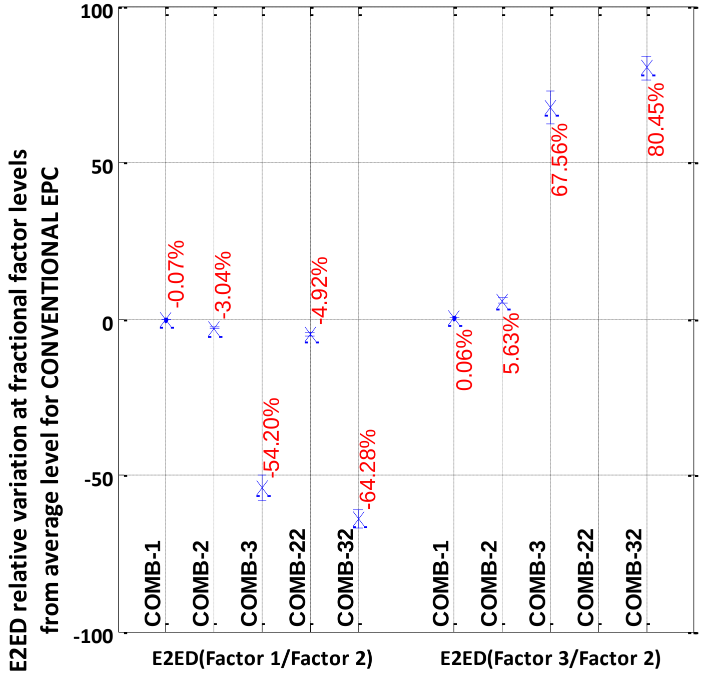


Figure 7-30: CONVENTIONAL EPC E2ED results relative variation from average fractional factor level in combinations

(registration experiment)

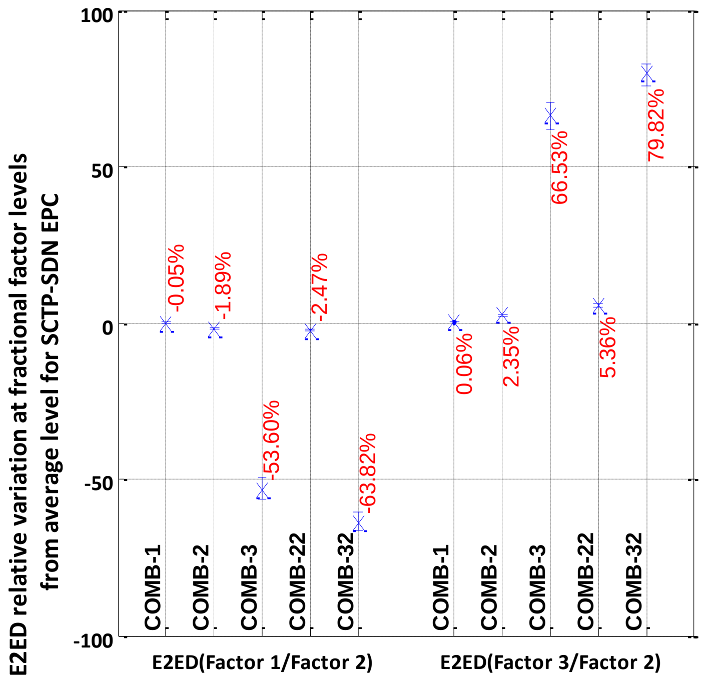


Figure 7-31: SCTP SDN-EPC E2ED results relative variation from average fractional factor level in combinations (registration experiment)

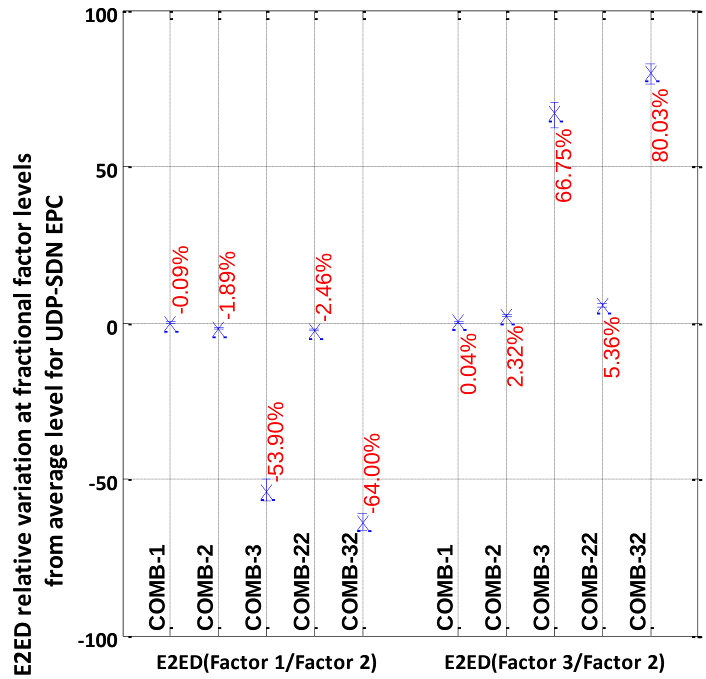


Figure 7-32: UDP SDN-EPC E2ED results relative variation from average fractional factor level in combinations (registration experiment)

7.3.1.2 E2ED VERSUS CONTROL PLANE LOAD

This subsection presents the E2ED results as a function of control load. The results viewed so far depicted the average performance of registration operations, however, the exact performance at a particular control load is still required. Based on observations previously made, Reg-1 and Reg-2 are selected to view nominal delay value versus overall control plane load which they represent the cases of highest E2ED, i.e. Reg-2, and lowest E2ED, i.e. Reg-1. Reg-2 also represents the case where there was variation between different EPC network types. Based on the fact that SDN-SCTP and SDN-UDP had little difference from each other, only SDN-SCTP results shall be presented for convenience and eliminating redundancy.

Figure 7-33 to **Figure 7-46** show the E2ED for registration operation versus control plane load for scenario Reg-1 and Reg-2 in each combination. Each figure represents a particular scenario with all fractional factor levels, i.e. low, average, and high, for a combination along with combination 0, i.e. the average configuration. Each figure contains the E2ED results for both CONV-EPC and SDN-EPC. For example, **Figure 7-33** show combination 0 and combination 1 at DBGT 20%, 50%, and 80% for EPC networks.

- The E2ED curves show in general two patterns, same as in mobility procedures:
  - At low and average control load points, from CL1 to CL6, there is almost constant E2ED curve. The exception is for combination 22 the pattern is from CL1 to CL3, as shown in Figure 7-41 to Figure 7-42.
  - At high control load points, from CL7 to CL9, there is an increase in E2ED results from its E2ED at average control load points. The increase occurs at a rapid rate in SDN-EPC, whereas in CONV-EPC the increase is quite minimal.
- The E2ED results at average control load points at each fractional factor level are already reported in previous section which matches the values in the curves at CL5 in each simulation configuration, and are not repeated in this section.
- The increase in E2ED at highest control load (CL9) from average control load (CL5) is extracted from figures to be as follows:
  - In Combination 0: 1.7 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-33 and Figure 7-34.

- In Combination 1: 1.5-2 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-33 and Figure 7-34.
- In Combination 2: 1.7-2.4 msec and (0.2 msec at low and average GWPT and 2 msec at high GWPT) increase from average point for SDN-EPC and CONV-EPC respectively; refer Figure 7-35 and Figure 7-36.
- In Combination 3: 2 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-37 and Figure 7-38.
- In Combination 22:
  - At low GWPT level: 1.5 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-43 and Figure 7-44.
  - At average GWPT level: 1.5 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-43 and Figure 7-44.
  - At high GWPT level: 2 msec for SDN-EPC, whereas CONV-EPC have E2ED of 354 msec at CL5 and approximately 515 msec with large standard deviation at CL9; refer Figure 7-41 and Figure 7-42.
- In Combination 32: 2-2.5 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-45 and Figure 7-46.
- In Combination 4: 2 msec and 0.2 msec increase from E2ED at CL5 for SDN-EPC and CONV-EPC respectively; refer Figure 7-39 and Figure 7-40.

The results show that CONV-EPC has almost no change with control load change in most configurations except combination 22. The recorded increase in E2ED at high control load CL9 is mostly 0.2 msec which is insignificant. For SDN-EPC, the results show that in most cases the increase in E2ED values is approximately 2 msec.

It is noticed that Reg-2 in SDN-EPC have lower E2ED curve on average load by approximately 1 to 2 msec and starts increasing to larger values from CONV-EPC at very high control load, i.e. CL9, whereas CONV-EPC have higher E2ED than SDN-EPC but almost does not react to increasing control load as the case with SDN-EPC. Thus the advantage of SDN-EPC is diminished by processing delays caused by the MME at high control load.

The similarity in E2ED due to high control load indicates that the increase in E2ED in most fractional factor levels is independent from fractional factor level except for high GWPT level. However, the percentage increase in E2ED at highest control load from average control load from average control load point is variable hence the average E2ED level at average fractional factor level is different. This is expected as there should be no interaction between MME processing on one side, and DBGT and DEL factors on the other side.

In combination 1 Figure 7-33, E2ED curve had almost the exact values curve at any DBGT level, for any scenario type, which aligns with previous conclusions that DBGT factor in these experiments had no impact on the E2ED results.

Different GWPT levels have very limited change in E2ED for SDN-EPC. The increase at high control load is similar at all GWPT levels. In CONV-EPC, the increase at low and average GWPT is minimal with increasing control load. At high GWPT level, there is slightly higher E2ED increase in combination 2. However, in combination 22 the performance suffered severe degradation for CONV-EPC where E2ED increased to 510 msec. For SDN-EPC, E2ED have consistent result with other GWPT levels at the same configuration with no degradation detected. The degradation in performance occurs at CL4 and upward for CONV-EPC, similar to that in S1-handover procedure, which again emphasizes the sensitivity of CONV-EPC to gateway processing capacity more than SDN-EPC.

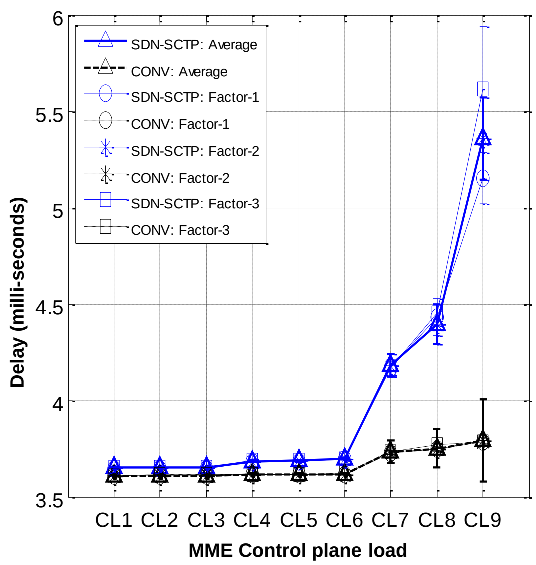


Figure 7-33: REG-1 in combination 0 and combination 1

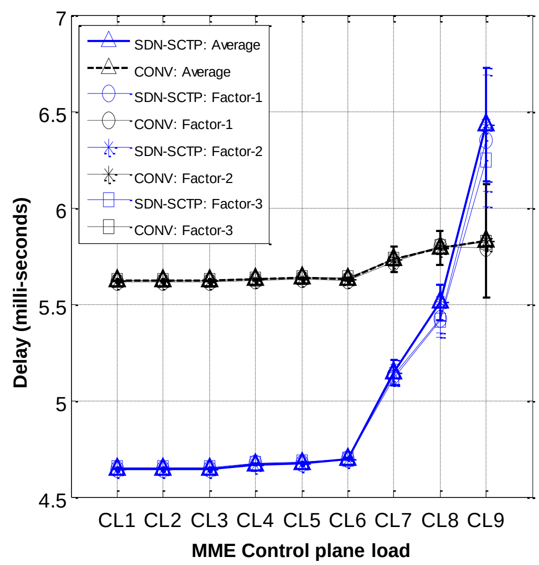


Figure 7-34: REG-2 in combination 0 and combination 1

(E2ED versus nominal control plane load at all fractional factor levels)

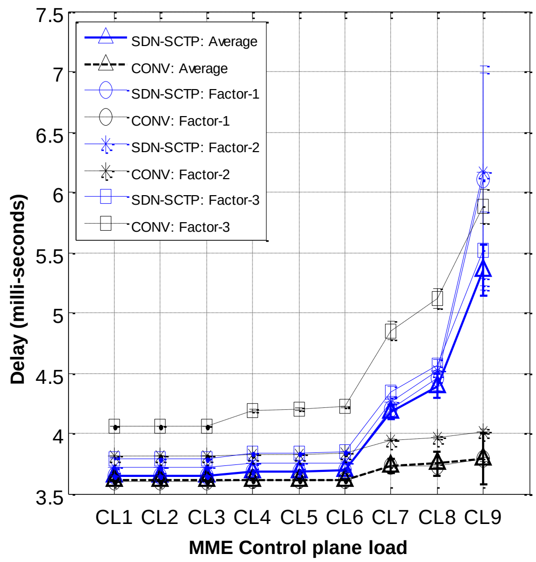


Figure 7-35: REG-1 in combination 0 and combination 2

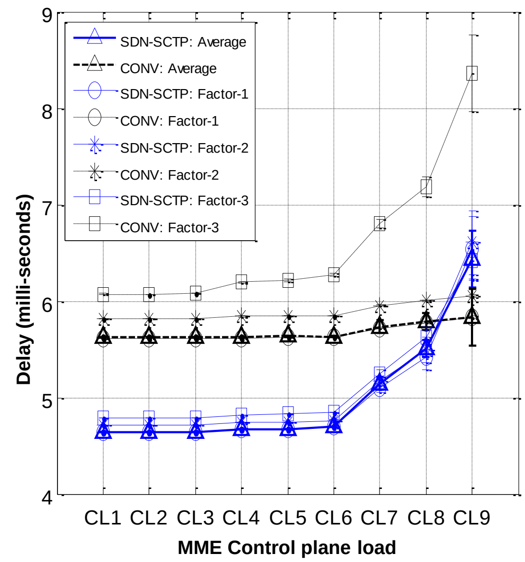


Figure 7-36: REG-2 in combination 0 and combination 2

(E2ED versus nominal control plane load at all fractional factor levels)

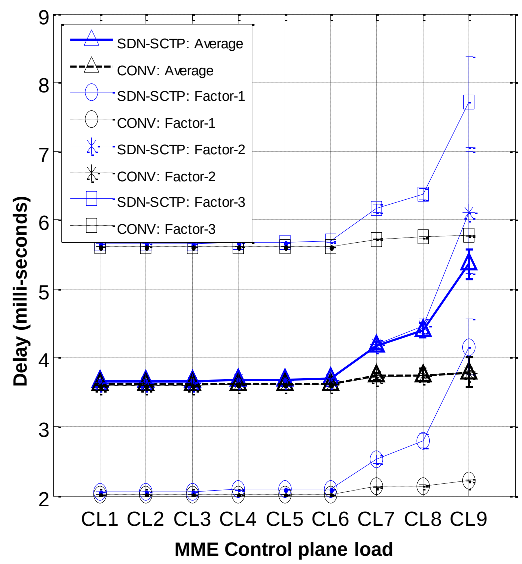


Figure 7-37: REG-1 in combination 0 and combination 3

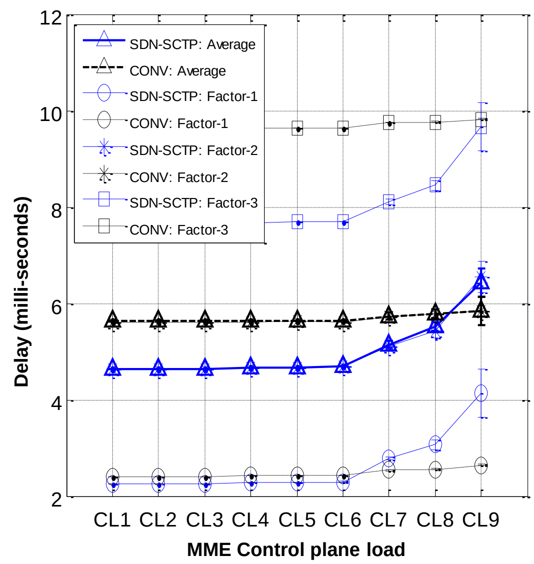
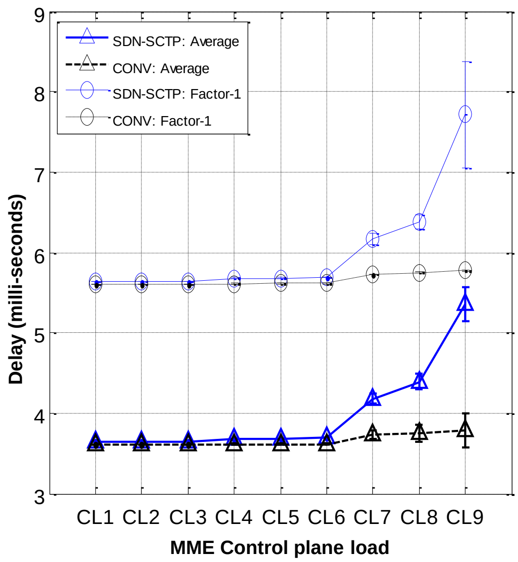


Figure 7-38: REG-2 in combination 0 and combination 3

(E2ED versus nominal control plane load at all fractional factor levels)



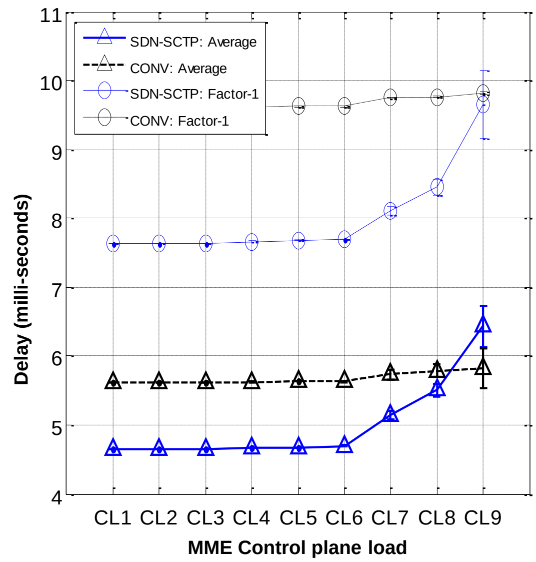
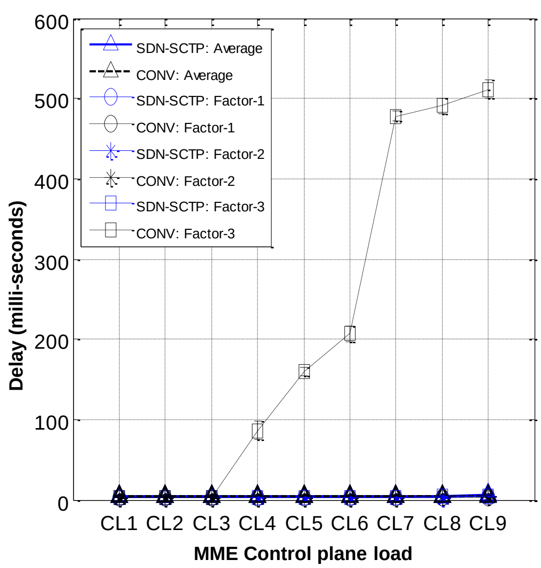


Figure 7-39: REG-1 in combination 0 and combination 4

Figure 7-40: REG-2 in combination 0 and combination 4

(E2ED versus nominal control plane load at all fractional factor levels)



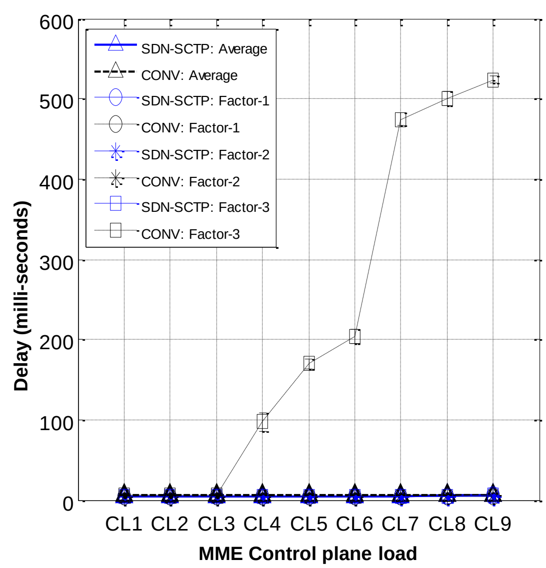
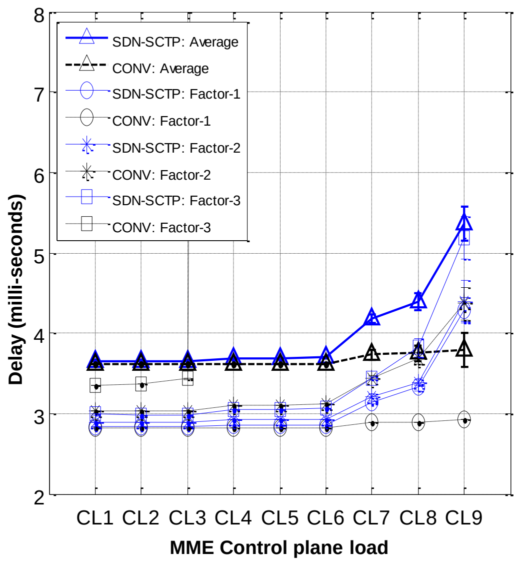


Figure 7-41: REG-1 in combination 0 and combination 22

Figure 7-42: REG-2 in combination 0 and combination 22

(E2ED versus nominal control plane load at all fractional factor levels)



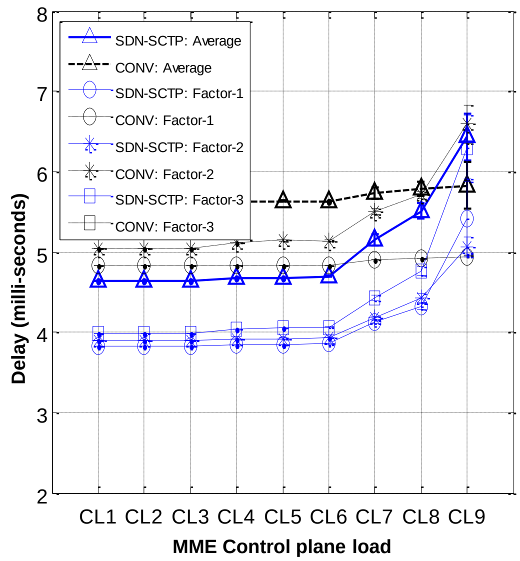
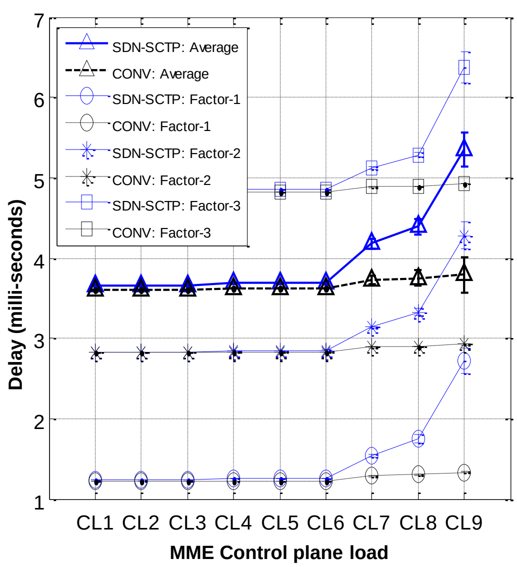


Figure 7-43: REG-1 in combination 0 and combination 22 (zoomed in)

Figure 7-44: REG-2 in combination 0 and combination 22 (zoomed in)

(E2ED versus nominal control plane load at all fractional factor levels)



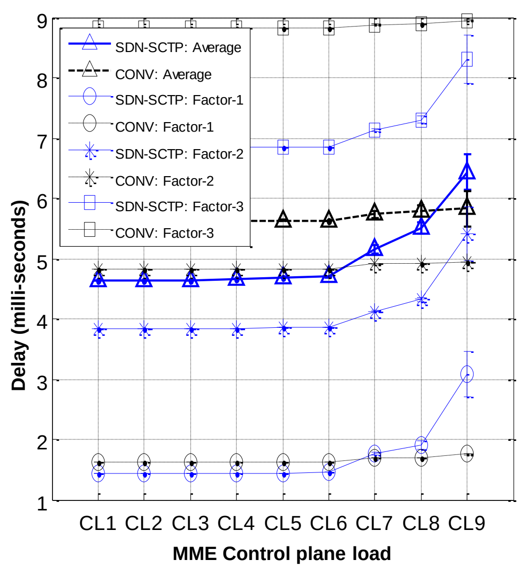


Figure 7-45: REG-1 in combination 0 and combination 32

Figure 7-46: REG-2 in combination 0 and combination 32

(E2ED versus nominal control plane load at all fractional factor levels)

7.3.2 MME LINKS BANDWIDTH UTILIZATION

This subsection refers to bandwidth utilization of MME physical links that are used in simulation setup, as shown in Figure 5-1. Table 7-13 shows the increase of bandwidth utilization of MME-R1 link and MME-SGW link in all simulations scenarios, both SDN-EPC variants led to an increase on average of 53% to 75% and 34% to 52% for SDN-SCTP and SDN-UDP respectively relative to CONV-EPC.

Table 7-14 shows the increase of bandwidth utilization of MME-SGW link in all simulations scenarios both SDN-EPC variants led to an increase on average of 130% to 230% and 83% to 160% for SDN-SCTP and SDN-UDP respectively relative to CONV-EPC. It is noted that the increase is much higher at the MME-SGW link, the same as noticed in mobility procedure for the same justification provided then due to communication with eNBs.

However, as well as mobility procedure the highest link bandwidth used in this experiment is 2 MB/sec, which is greatly less than 1% utilization of the 10Gbps link. This concludes that in registration operations even with downside of increased number of packets and packets overhead, the differentiation between SCTP and UDP variants for an SDN-EPC based on link utilization is negligible; Not to mention the utilization due to control operations itself is negligible.

Table 7-13: MME-R1 LINK BANDWIDTH UTILIZATION ACROSS SCENARIOS FOR ALL COMBINATIONS

SCENARIOS	RELATIVE VARIATION FROM CONV-EPC	
	SDN-SCTP EPC	SDN-UDP EPC
REG-1	+57%	+37%
REG-2	+53%	+34%
REG-3	+74%	+52%
REG-4	+67%	+47%

Table 7-14: MME-SGW LINK BANDWIDTH UTILIZATIONACROSS SCENARIOS FOR ALL COMBINATIONS

SCENARIOS	RELATIVE VARIATION FROM CONV-EPC	
	SDN-SCTP EPC	SDN-UDP EPC
REG-1	+168%	+116%
REG-2	+230%	+162%
REG-3	+130%	+83%
REG-4	+176%	+115%

7.3.3 MME CPU UTILIZATION

The results in this section are presented similar to that in mobility procedure section. MME CPU utilization in combination 2 and 22 are chosen for viewing the ratio of the utilization for SDN-EPC compared to CONV-EPC at each control points. Figure 7-47 and Figure 7-48 show the ratio of MME CPU utilization for SDN-EPC compared to CONV-EPC in combinations 2 and 22; the figures indicate the following:

- MME resources requirement in SDN-EPC ranges between 17%-22% and 17%-22% more than that in CONV-EPC in combination 2 and 22, respectively, at different

GWPT levels. The exception is for high GWPT level in combination 22 which starts at 18% at CL1 and reaches 50% at CL9.

- The highest ratio of increase in required MME resource occurs when the CORR rate is highest for the same CBGT rate, this occurs for points CL3, CL6, and CL9; the least ratio of increase occurs when the CORR rate is lowest for the same CBGT rate, this occurs for points CL1, CL4, and CL7. These findings are opposite to what was observed in S1-handover mobility procedure.
- In general, the ratio of utilizations is almost identical, with 1% to 2% difference, for different GWPT levels, except in combination 22 with GWPT of 150 usec. There is a surge in this ratio when GWPT level is configured to 150 usec in combination 22. The increase reaches 50% of that for CONV-EPC at highest control load points. This behavior is explained with the same reasoning that occurred in mobility procedure. The decrease in requests arrival at the MME at high GWPT leads to increasing this ratio, since SDN-EPC is actually processing more requests than CONV-EPC.

These ratios indicate how much additional resources are required in SDN-EPC relative to CONV-EPC. According to the first observation, at least 17% more resources are required for SDN-EPC MME operations and this ratio increases when CORR is at its highest level at 5%, and decreases when CORR at its highest level at 5%.

For each registration procedure request, the MME has to process 4 GTP-C messages in CONV-EPC and 5 GTP-C messages in SDN-EPC, in both EPC networks 2 GTP-C messages are for MME-eNB communication, that leaves 2 and 3 GTP-C messages for MME-SGW/PGW communications in CONV-EPC and SDN-EPC, respectively. The processing required for GTP-C message in registration is higher than that for bearer modification procedure (CBGT), hence GTP-C messages for registration procedure are larger in size than those for modification procedure, refer Table **5-9** control messages size. Therefore there is an increase in required resources in SDN-EPC compared to CONV-EPC for each increase in registration request rate.

Figure 7-49 to **Figure 7-51** presents MME-Util as a function of control load. It is noticed that the gap between both architectures is increased with increasing load which is expected due to increased requests arrival and difference in required resources in each architecture. The difference between CONV-EPC and SDN-EPC under low control plane load is approximately 5% of MME-CPU capacity, and it increases under very high control loads to reach 16% of MME CPU capacity. The reduction in MME-Util in combination 22 at highest fractional factor is, as in mobility procedure experiments, due to congestion at core SGW which delays control packet sent from SGW to MME and decreases the rate of incoming packets at the MME.

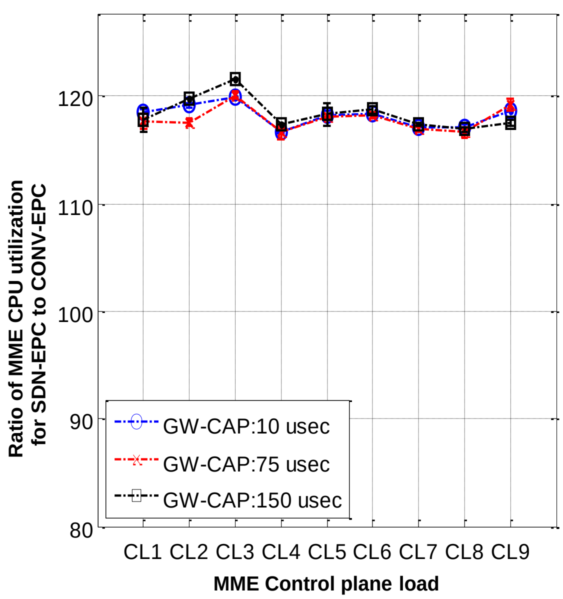


Figure 7-47: Ratio of MME CPU

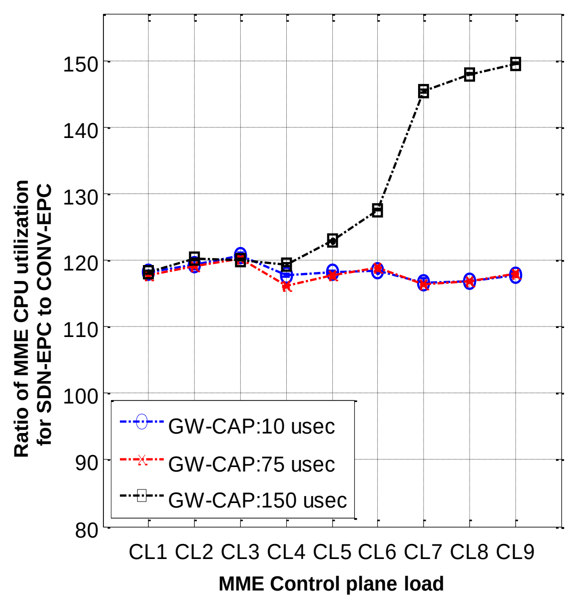


Figure 7-48: Ratio of MME CPU

utilization for SDN-EPC to CONV-EPC in combination-2 as a function of control load

utilization for SDN-EPC to CONV-EPC in combination-22 as a function of control load

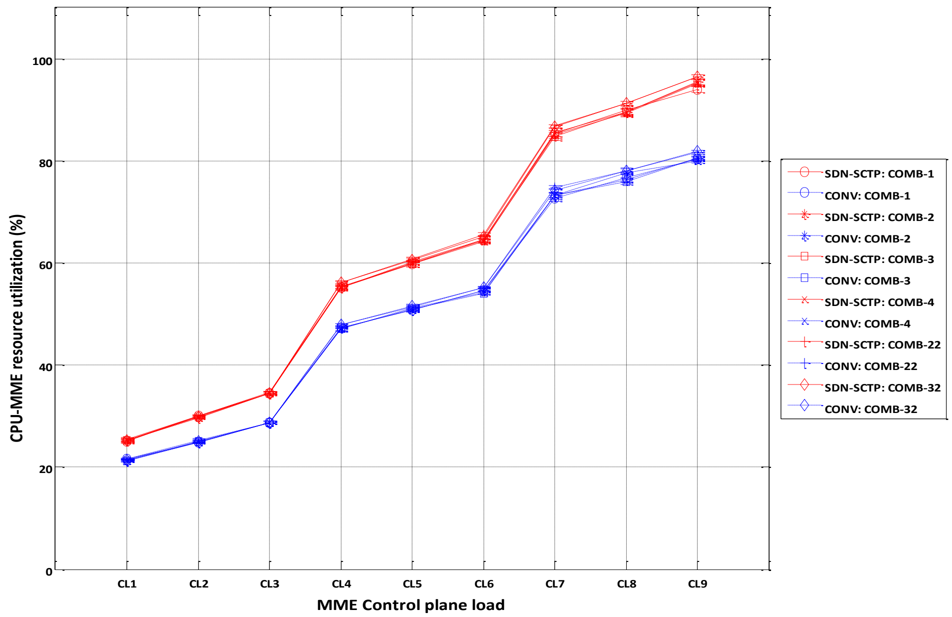


Figure 7-49: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios REG-1 at lowest fractional factors level

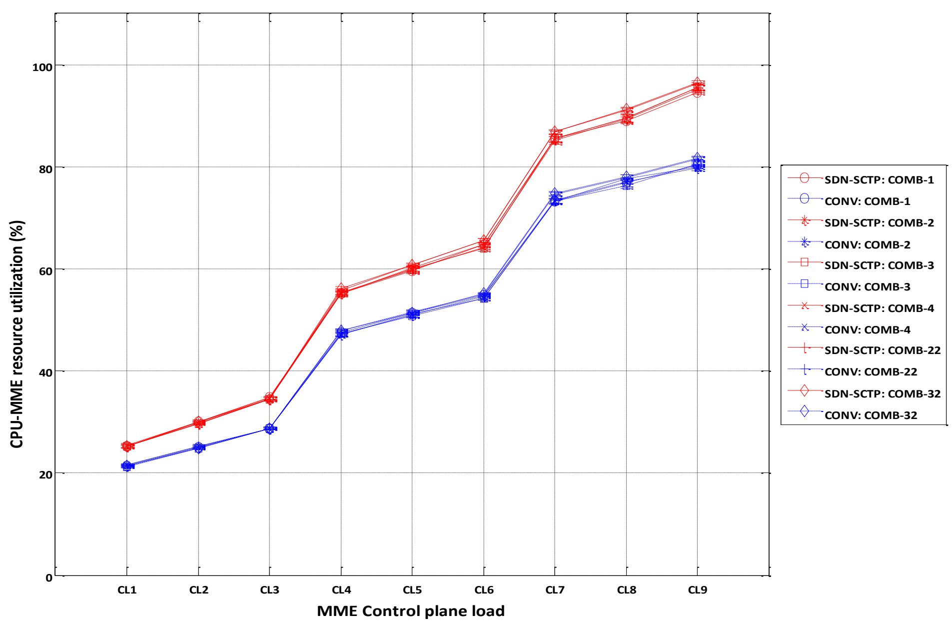


Figure 7-50: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios REG-1 at average fractional factors level

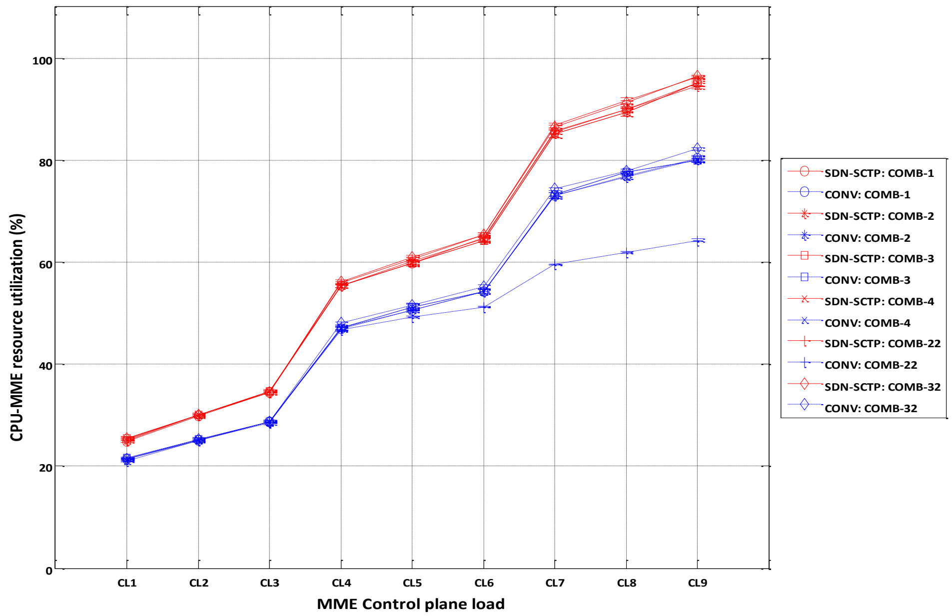


Figure 7-51: MME CPU UTILIZATION versus nominal control plane load in all combinations for scenarios REG-1 at highest fractional factors level

7.3.4 CORE SGW CPU UTILIZATION

The effect of EPC gateways utilization can be observed in one of the SGWs deployed; either core SGW or local SGW. This refers to the network setup in Figure 5-1. Both SDN-SCTP and SDN-UDP have the exact SGW CPU utilization in the results, as it is expected, and no distinction is made further between them. Table 7-15 shows the average relative variation of Core SGW utilization from CONV-EPC; all scenarios have the same average performance compared to CONV-EPC, which approximately on average 50% less. The table record scenario Reg-1 and 3 which are scenarios were Core SGW is involved in registration procedure, whereas Reg-2 and 4 have the same performance but for Local SGW. This similarity in performance is due to the fact that gateways role in registration procedure is symmetrical, meaning in Core SGW performed in Reg-1 and 3 are the same operations as Local SGW performed in Reg-2 and 4. There is at least 50% decrease in SGW utilization in SDN-EPC compared to CONV-EPC. This reflects the amount of resources required in both architectures. This experiments show that SDN-EPC can reduce SGW gateway processing resources by at least 50%.

Table 7-15: AVERAGE RELATIVE VARIATION OF CORE SGW CPU UTILIZATION IN SDN-EPC ACROSS SCENARIOS

	REG-1	**REG-3**
RELATIVE VARIATION FROM CONV-EPC	**-50%**	**-50%**

Figure 7-52 and Figure 7-53 show the ratio of core SGW CPU utilization for SDN-EPC compared to CONV-EPC in combinations 2 and 22. The figures indicate that Core SGW resources requirement in SDN-EPC is 50% less than that in CONV-EPC across GWPT levels. The exception is at high GWPT level in combination 22 which starts at 50% at CL1 and decreases to reach 5% at CL9.

The utilization ratio is almost identical for different GWPT levels, except in combination 22 at GWPT of 150 usec. There is a surge in required resources ratio for SDN-EPC when GWPT level is configured to 150 usec in combination 22. The ratio reaches 92% of that for CONV-EPC at highest control load points. This behavior is due to the fact that there is a decrease in requests arrival at the MME at high GWPT which is caused by congestion at the new SGW. Therefore the ratio has increased since SDN-EPC is actually processing more requests than CONV-EPC.

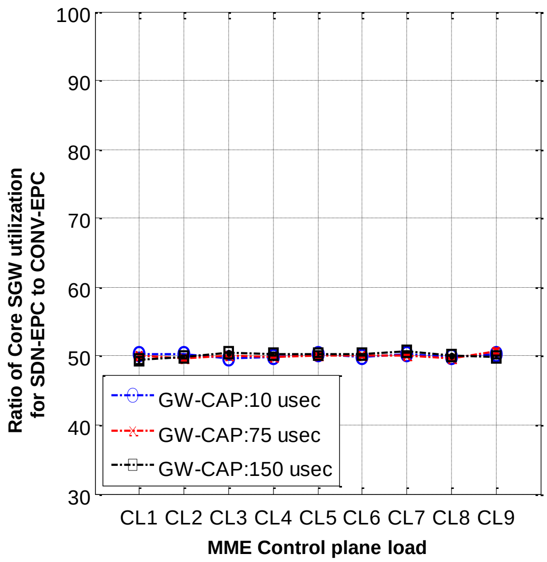


Figure 7-52: Ratio of core SGW CPU utilization for SDN-EPC to CONV-EPC in combination-2 as a function of control load

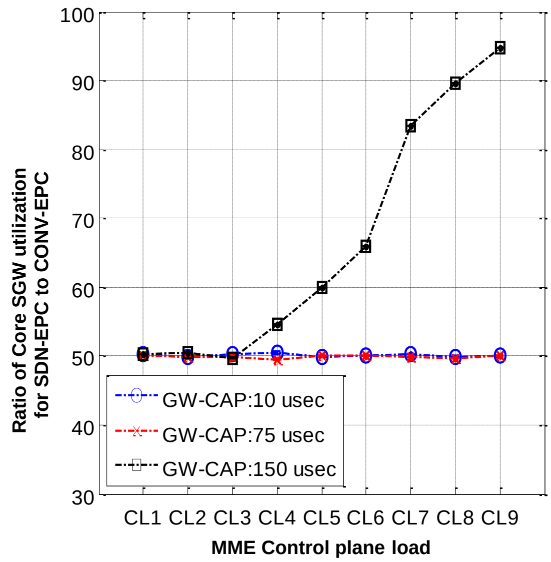


Figure 7-53: Ratio of core SGW CPU utilization for SDN-EPC to CONV-EPC in combination-22 as a function of control load

Figure 7-54 to **Figure 7-56** show the core SGW CPU utilization for Reg-1 at different fractional factors levels. It is clear that CPU utilization is increasing monotonically with control plane load as expected. The following can be deduced from the figures:

- The utilization at GWPT of 10 usec for all combinations and EPC network types never exceeded 6% of total CPU resources; refer Figure 7-54.
- The utilization at GWPT of 75 usec in SDN-EPC network reached at maximum control load to 22% and 44% of SGW CPU resources for combination 2 and 22 respectively, whereas CONV-EPC CPU utilization reached at maximum control load to 44% and 88% of SGW CPU resources for combination 2 and 22 respectively; refer Figure 7-55 and Figure 7-56.

- The utilization at GWPT of 150 usec in combination 22 in CONV-EPC architecture reached saturation (100% utilization) quickly under medium control load, whereas in SDN-EPC the utilization under highest control plane load reached 90% utilization. In combination 2 the utilization for CONV-EPC reached 88% under high loads, whereas for SDN-EPC reached 45%; refer Figure 7-55 and Figure 7-56.

These results are consistent with previous results for mobility procedure, where it emphasizes the fact that SDN-EPC has more capability to use software based platforms at high requests rate than that of CONV-EPC network. It is also observed that improved software-based gateways perform well for both network architecture, and it reaches at highest control load to 45% and 90% of their full capacity indicating that further increase in request rate a bottleneck is reached at CONV-EPC network. SDN-EPC also would reach that point but at a much higher request rate than CONV-EPC would.

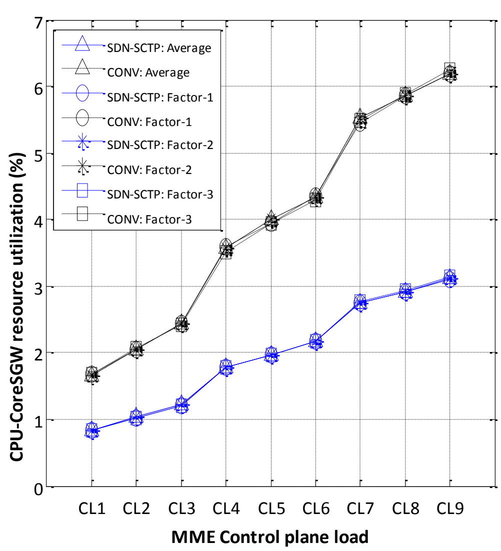


Figure 7-54: CPU-Util Core-SGW combination 0 and combination 1

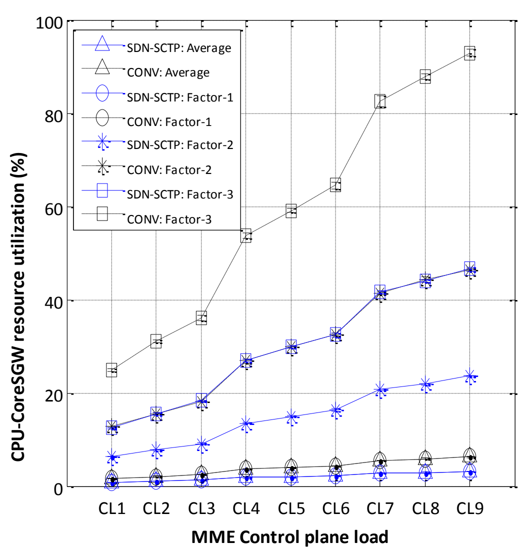


Figure 7-55: CPU-Util Core-SGW combination 0 and combination 2

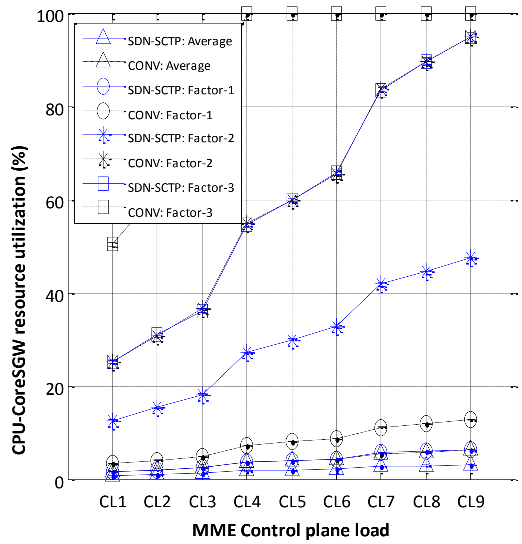


Figure 7-56: CPU-Util Core-SGW combination 0 and combination 22

CORE SGW CPU UTILIZATION versus nominal control plane load in Reg-1 scenario at all fractional factor levels

CHAPTER 8

CONCLUSION AND FUTURE WORK

Software defined networking 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-EPC should outweigh its disadvantages in order for manufacturers and operators to adopt this change in the control plane architecture.

In this study, we analyzed the characteristics of EPC architecture and EPC entities. We described different operational configuration and physical features which affect its performance. We provided qualitative analysis and comparison between conventional control plane and an SDN-based control plane in the EPC. An SDN-based EPC, in our view, is best suited to be an overlay architecture rather than controlling the lower layers as the case in OpenFlow architecture. We presented main control plane operations, particularly the registration procedure and S1-based handover, and provided the operational procedure for them. The modification of the operations from conventional architecture to SDN-based has been derived along with GTP-C message formats.

Lastly, simulation framework that reflects real deployment of an EPC network is built and justification for various configurations is presented. The simulation framework is built such that it enables exploring the performance of various configurations and different scenarios of network setup and EPC network deployment that would reflect the behavior of the control operations in diverse situations. The framework examines effect of various factors on performance metrics and explores their relative functions. The framework has the following controlled factors: EPC network type (CONV-EPC, SDN-SCTP, and SDN-UDP), distributed EPC gateways locations, data plane background traffic (DBGT), control plane background traffic (CBGT), control operation request rate (CORR), EPC gateways processing capacity (GWPT), MME processing capacity, and backhaul links propagation delay (DEL). In this framework, a fractional factorial experiment was designed to enable an effective and efficient exploration of performance metrics under sufficient and adequate factors combinations.

8.1 MAIN FINDINGS AND FUTURE RESEARCH DIRECTIONS

I. The proposed SDN-EPC exploits inherent centralized properties of EPC network and provides the characteristics of an SDN architecture without the need to make disruptive changes to GTP protocol. The proposed SDN-EPC is 3GPP standards compliant and paves for an easy migration guidelines from conventional EPC to SDN-based EPC.

II. Qualitative study of EPC functions and SDN concept justifies the use of overlay-based SDN-EPC rather than OpenFlow based architecture which preserves independence of EPC nodes functionality from underlying protocol layers functionality and in turn facilitates adoption of SDN-EPC architecture into existing EPC models.

Performance evaluation of SDN-EPC and CONV-EPC has shown the following in terms of end-to-end delay (E2ED) of registration and S1-handover procedures:

III. SDN-EPC have better E2ED performance compared to CONV-EPC, i.e. less, only when the MME is under low to average resource utilization and the delay caused by trombone route between the PGW and the SGW involved in control operation is maximal in the CONV-EPC; this latter condition occurred only when the PGW located in core center and the SGW located in local center are involved in the control operation.

III.1. The percentage of enhancement of E2ED between SDN-EPC and CONV-EPC in the former case depended on propagation delay of communication links between core center and local center. Larger propagation delay levels, i.e. more distance, lead to larger enhancement.

III.2. The reduction of SDN-EPC in E2ED in the former case recorded 1-7% and 6-23% in S1-handover and registration procedures, respectively, for backhaul links propagation delay configurations of 0.1-1.0 msec.

III.3. In terms of SGW processing time: the difference of E2ED between SDN-EPC and CONV-EPC is found to be minimal, less than 0.5 msec, as the contribution of processing time in SGW to the E2ED is minimal compared to the contribution of the other factors, i.e. propagation delay and MME processing delay.

IV. SDN-EPC has worse E2ED performance compared to CONV-EPC, i.e. higher, when the MME is under high resources utilization, i.e. 80% utilization or more. The SDN-EPC requires more resources at the MME which reflects more sensitivity to its utilization. The E2ED suffered 7-10 msec and 2 msec degradation in S1-handover and registration procedures, respectively, from its average E2ED value. The amount of degradation is irrelevant to the location of EPC gateways involved in control operation and as well of backhaul links propagation delay.

V. SDN-EPC have comparable E2ED performance, i.e. almost equal, when the MME is under low to average resource utilization and the delay caused by trombone route between PGW and SGW involved in the control operation is minimal in the CONV-EPC. This latter condition occurred when both the PGW and the SGW involved in the control operation are located in the center location or the PGW is in local center while SGW is in core center.

VI. SCTP-based SDN and UDP-based SDN have shown no significant difference in end to end delay performance. The difference remains within 0-2% of overall delay, and the difference in bandwidth utilization is considered unimportant compared to the overall link capacity using very high speed links, i.e. 10Gbps.

VII. Performance evaluation of SDN-EPC and CONV-EPC has shown the following in terms of EPC nodes resource utilization in registration and S1-handover procedures:

VII.1. SDN-EPC induces more resource utilization in the MME by 10%-17% and 17%-22% in S1-handover and registration procedures, respectively. This is caused by control operations relocated from EPC gateways into the MME control operations. The percentage of additional resources required increases with increasing registration requests arrival rate and decreases with increasing S1-handover requests arrival rate.

VII.2. SDN-EPC induces less resource utilization in the SGW gateways by 35%-50% and 50% in S1-handover and registration procedures, respectively. This is caused by control operations relocated from EPC gateways into the MME control operations, the reduction is observed to decrease with increasing S1-handover requests arrival rate.

VIII. Performance evaluation of SDN-EPC and CONV-EPC has shown the following in terms of bandwidth utilization of registration and S1-handover procedures:

VIII.1. SDN-EPC have produced more bandwidth utilization of link connecting the MME and ingress router by 53%-74% and 34%-52% for SCTP-based SDN and UDP-based SDN, respectively. In addition, an increase in bandwidth utilization of link connecting the MME and core SGW is observed to be 130%-230% and 83%-162% for SCTP-based SDN and UDP-based SDN, respectively. However, the overall bandwidth utilization of any of the links never exceeded 1% utilization for a 10 Gbps link, which renders this effect insignificant.

IX. The performance evaluation of control plane in EPC network have shown the order of factors contribution and determination of expected end to end delay of control operations are: propagation delay of backhaul links as dominant factor, location of EPC gateways participating in control operation, and MME processing capacity. The relative effect of each of these factors on E2ED variation is minimally affected by type of the EPC network. The variation of factor levels operates with same percentage on E2ED regardless of EPC network type.

X. Data plane background traffic have no effect on E2ED performance of control operations under the condition that intermediate nodes in the backhaul network do not impose any processing delays on forwarded packets. The transport links have very high capacity (10 Gbps) which would not reflect any sizeable queuing delay on forwarded packets compared to the overall E2ED of control operation.

XI. Propagation delay of backhaul links have a significant contribution to E2ED of any control procedure; E2ED is monotonically increasing with increased propagation delay values and it is possible to produce a prediction model from gathered data on an experiment to predict end to end delay of a particular control operation for a particular propagation delay. It is found that depending on location of gateway participating in control operation and the MME processing capacity the effect of increasing propagation delay on E2ED is as follows:

XI.1. Increasing propagation delay from 0.1 msec (30 km) to 0.5 msec (50 km) induces 1.8-2.9 times increase in E2ED results.

XI.2. Increasing propagation delay from 0.5 (50 km) msec to 1.0 msec (100 km) induces 1.55-1.75 times increase in E2ED results.

XII. Reduction of E2ED operation by increasing MME capacity is dependent on portion of contribution of the MME processing delay in overall control operation delay, which in turn depends on propagation delay of backhaul links, the larger the propagation delays the lesser portion of overall delay is attributed to the MME's processing delay.

XII.1. The reduction recorded by doubling the MME capacity in S1-handover procedure is 33%-37%, 14%-19%, and 8%-12% at 0.1, 0.5, and 1.0 msec propagation delay, respectively. For registration procedure it is 33%-40%, 14%-22%, and 8%-14% at 0.1, 0.5, and 1.0 msec propagation delay, respectively.

XII.2. This finding shows that increasing MME processing capacity can improve delay performance to a certain limited extent controlled by its contribution to the overall delay.

XIII. Hardware based EPC gateways provided the best performance and enhanced software based gateways provided second best with a minimal degradation at highest request rate regardless of EPC network type. However, software based gateways were not able to accommodate large number of request rate in CONV-EPC architecture, but in SDN-EPC, the E2ED performance was affected only at highest request rate and the degradation was much smaller than that for CONV-EPC.

XIV. SDN-EPC have shown more adaptability to software based gateways; whereas CONV-EPC had severe congestion at the SGW at high control request rate. Particularly for gateways with 150 usec processing time and single queue-single server model; in CONV-EPC at 6400 session modification procedure rate and more, E2ED is recorded with 100 times and 35 times increase in S1-handover and registration procedures, respectively. Whereas SDN-EPC is recorded only with 10 times increase at 10400 request rate using the same gateway model.

As a future research direction this work can be extend to include factors that were not present in the simulation. Some of these factors are SCTP protocol stack and UDP protocol stack processing effect. Another factor is the processing delay in intermediate nodes, e.g. routers; this might accumulate for large number of nodes which will show the effect of data plane background traffic. An important factor as well is a more accurate model of processing time in EPC gateways and the MME rather than the linear model used in the experiments; as it is known that software based platform does not perform linearly with load being processed, it can be an interesting area to look into.

APPENDIX A

GTP MESSAGE FORMATS

APPENDIX A contains GTP-C and S1-AP messages formats used in the simulation testbed. These formats are extracted from standardized GTP-C message formats listed in 3GPP TS 29.274 document and standardized S1-AP message formats listed in (3GPP TS 36.413. The list of messages reported here are:

- CREATE SESSION REQUEST
- CREATE SESSION RESPONSE - S11/S5
- MODIFY BEARER REQUEST - S11/S5
- MODIFY BEARER RESPONSE - S11/S5
- DELETE SESSION REQUEST
- DELETE SESSION RESPONSE
- DELETE IDFT REQUEST
- DELETE IDFT RESPONSE
- CREATE IDFT REQUEST
- CREATE IDFT RESPONSE
- HANDOVER REQUIRED
- HANDOVER REQUEST
- HANDOVER REQUEST ACKNOWLEDGE
- HANDOVER NOTIFY
- HANDOVER COMMAND
- ENB STATUS TRANSFER

The tables Table 8-1 to Table 8-16 present the detailed information elements in each GTP-C message and its standard byte size.

Table 8-1: Modify Bearer Request message information elements

Message Type: Modify Bearer Request	
Information Elements (IE)	**IE size (Bytes)**
ME Identity	12
User Location Information	43
Serving Network	7
RAT Type	5
Indication Flags	0
F-TEID	13
AMBR	0
Delay Value (S11)	5
Bearer Contexts to be modified (Handover)	22
Bearer Contexts to be removed/s11	9
MME-FQ-CSID (S11)	11
SGW-FQ-CSID (S5)	11
User CSG Information	12
Total Size (S11)	**155**
Total Size (S5)	**155**

Table 8-2: Create Session Request information elements

Message Type: Create Session Request	
Information Elements (IE)	**IE size (Bytes)**
IMSI	11
MSISDN	12
ME Identity	12
User Location Information (ULI)	43
Serving Network	7
RAT Type	5
Indication Flags	0
F-TEID	13
PGW S5/S8 Address	8
APN	35
Selection Mode	5
PDN Type	5
PAA	9
APN Restriction	5
AMBR	12

EPS Bearer ID (EBI)	5
Bearer Context	70
PCO	10
Trace Information	0
Recovery	0
MME-FQ-CSID	11
SGW-FQ-CSID	11
UE Time Zone	6
User CSG	12
Charging	6
Signaling Priority	5
Total Size (S11)	**334**
Total Size (S5)	**334**

Table 8-3: Create Session Response information elements

Message Type: Create Session Response	
Information Elements (IE)	**IE size (Bytes)**
Cause	6
Change Reporting Action	5
CSG Information Reporting Action	5
Sender F-TEID for CP (S11)	13
PGW S5/S8/S2b FTEID	13
PDN Address Allocation (PAA)	9
APN Restriction	5
APN-AMBR	12
PCO	10
Bearer Contexts created	62
Bearer Contexts marked for removal	15
Recovery	5
Charging Gateway Address	8
PGW-FQ-CSID (S5)	11
SGW-FQ-CSID (S11)	11
SGW LDN (S11)	24
PGW LDN (S5)	24
Total Size (S11)	**224**
Total Size (S5)	**206**

Table 8-4: Modify Bearer Response information elements

Message Type: Modify Bearer Response	
Information Elements (IE)	**IE size (Bytes)**
Cause	6
MSISDN	12
EPS Bearer ID (EBI)	5
APN Restriction (Handover)	5
PCO - HO	10
Bearer Contexts modified	36
Bearer Contexts marked for removal	15
CSG Information Reporting Action	5
FQ-CSID	11
Total Size (S11)	**121**
Total Size (S5)	**121**

Table 8-5: Delete Session Request information elements

Message Type: Delete Session Request	
Information Elements (IE)	**IE size (Bytes)**
Cause - S11	6
EPS Bearer ID (EBI)	5
ULI	43
PCO	10
Node Type - S11	5
F-TEID - S11	11
UE Time Zone	6
Total Size (S11)	**102**
Total Size (S5)	**102**

Table 8-6: Delete Session Response information elements

Message Type: Delete Session Response	
Information Elements (IE)	**IE size (Bytes)**
Cause	6
Recovery	0
PCO	10
Private Extension	0
Total Size (S11)	**32**

Total Size (S5)	**32**

Table 8-7: Delete IDFT Request information elements

Message Type: Delete IDFT Request	
Information Elements (IE)	**IE size (Bytes)**
Private Extension	0
Total Size (S11)	**16**

Table 8-8: Delete IDFT Response information elements

Message Type: Delete IDFT Response	
Information Elements (IE)	**IE size (Bytes)**
Cause	6
Private Extension	0
Total Size (S11)	**22**

Table 8-9: Create IDFT Request information elements

Message Type: Create IDFT Request	
Information Elements (IE)	**IE size (Bytes)**
IMSI	11
ME Identity	12
F-TEID	13
Bearer Contexts	57
Total Size (S11)	**109**

Table 8-10: Create IDFT Response information elements

Message Type: Create IDFT Response	
Information Elements (IE)	**IE size (Bytes)**
Cause	6
F-TEID	13
Bearer Context	63
Total Size (S11)	**98**

Table 8-11: HANDOVER REQUIRED information elements

Message Type: HANDOVER REQUIRED	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
eNB UE S1AP ID	5
Handover Type	4
Cause	4
Target ID	27
Direct Forwarding Path Availability	3
SRVCC HO Indication	3
Source to Target Transparent Container	65
Source to Target Transparent Container Secondary	65
CSG Id	6
Cell Access Mode	3
PS Service Not Available	3
Total Size	**200**

Table 8-12: HANDOVER REQUEST information elements

Message Type: HANDOVER REQUEST	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
Handover Type	4
Cause	4
UE Aggregate Maximum Bit Rate	10
E-RABs To Be Setup List	50
Source to Target Transparent Container	65
UE Security Capabilities	4
Handover Restriction List	47
Trace Activation	40
Request Type	4
SRVCC Operation Possible	3
Security Context	39
NAS Security Parameters to E-UTRAN	0
CSG Id	6
CSG Membership Status	3
GUMMEI	12

MME UE S1AP ID 2	6
Management Based MDT Allowed	3
Management Based MDT PLMN List	9
Masked IMEISV	10
Expected UE Behavior	**0**
Total Size	**331**

Table 8-13: HANDOVER REQUEST ACKNOWLEDGE information elements

Message Type: HANDOVER REQUEST ACKNOWLEDGE	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
eNB UE S1AP ID	5
E-RABs Admitted List	71
E-RABs Failed to Setup List	0
Target to Source Transparent Container	65
CSG Id	6
Criticality Diagnostics	13
Cell Access Mode	3
Total Size	**175**

Table 8-14: HANDOVER NOTIFY information elements

Message Type: HANDOVER NOTIFY	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
eNB UE S1AP ID	5
E-UTRAN CGI	13
TAI	11
Tunnel Information for BBF	10
LHN ID	34
Total Size	**85**

Table 8-15: HANDOVER COMMAND information elements

Message Type: HANDOVER COMMAND	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
eNB UE S1AP ID	5
Handover Type	4
E-RABs Subject to Forwarding List	63
E-RABs to Release List	0
Target to Source Transparent Container	65
Criticality Diagnostics	13
Total Size	**162**

Table 8-16: eNB STATUS TRANSFER information elements

Message Type: eNB STATUS TRANSFER	
Information Elements (IE)	**IE size (Bytes)**
Message Type	4
MME UE S1AP ID	6
eNB UE S1AP ID	5
eNB Status Transfer Transparent Container	39
Total Size	**56**

APPENDIX B

SIMULATOR GUIDE

APPENDIX B contains guidelines onto how to run the designed testbed. OMNeT++ is used in this work for running simulation experiments. OMNeT++ framework includes three types of files for implementation of network nodes functionality, these are: NED files, INI files, and regular C++ files. Any functionality first implemented through C++ files and the INET framework is used which includes network functionality for most network layers, e.g. physical to application layer functionality. The NED files are used for: network connections (network set up), network device type, e.g. router, host or any other custom device (EPC node), and assign some of network functionality such as links data rates. The INI files (initialization files) are used to configure network devices parameters through direct assignment.

In this work, modifications to some of the C++ files are present in various locations which are beyond explanation in this document, however, in short the modifications are implemented to perform control operations of EPC network as described in the thesis chapters. The C++ files concerning network node functionality are found in "src/application" folder in the following subfolders: UDPSDNMobility, UDPMobConv, UDPMobConv, SCTPBasic, SCTPMobUDP, SCTPSDN. These subfolders contain modified C++ files of UDP and SCTP applications derived from the INET framework. Moreover, a model for single queue-multi server is added, where the C++ files are found in "src/application/MMEPCU" subfolder. These folders contain C++ files that correspond to applications behaviors used in INI files.

There is only one NED file in the experiment setup; the NED file is named "defaultArch.ned". The file contains described setup in the thesis chapter 5 and assigns node devices to perspective roles, and assigns link data rate to 10Gbps. The file configuration remains unchanged while running simulation experiments.

The INI files are used for specific configuration of network nodes and functionality; there is one INI file for each EPC type, scenario, and control operation as shown in table below.

SCEN-ARIO	CONV	SDN-SCTP	SDN-UDP
Mob-A	M_UDP_4_2.ini	M_SCTP_4_2.ini	M_UDPSDN_4_2.ini
Mob-B	M_UDP_2_4.ini	M_SCTP_2_4.ini	M_UDPSDN_2_4.ini
Mob-C	M_UDP_3_1.ini	M_SCTP_3_1.ini	M_UDPSDN_3_1.ini
Mob-D	M_UDP_1_3.ini	M_SCTP_1_3.ini	M_UDPSDN_1_3.ini
Reg-1	UDP_Core_Core.ini	SCTP_Core_Core.ini	UDPSDN_Core_Core.ini
Reg-2	UDP_Local_Core.ini	SCTP_Local_Core.ini	UDPSDN_Local_Core.ini
Reg-3	UDP_Core_Local.ini	SCTP_Core_Local.ini	UDPSDN_Core_Local.ini
Reg-4	UDP_Local_Local.ini	SCTP_Local_Local.ini	UDPSDN_Local_Local.ini

Additional to INI files above there is the following files used for configuration: prefix.ini, BGHscript.ini, CP_BGT_UDP_local.ini, and CP_BGT_UDP_core.ini. The following is the procedure for changing parameter configurations throughout the experiments.

First, file "prefix.ini" contains general configuration of parameter which are:

```
result-dir = GRANDCOMB       # location of results files
repeat = 10                 # number of replication for same configuration
**.eth10GNew.delay = ${DEL= 0.1,0.5,1.0}ms
                    # PARAMTER used for propagation delay configuration
**.MME.cpu.capacity = ${MME= 60}
                    # Parameter used for MME CAP configuration
**.BGeNBNumber = ${CPV1= 26,65,104}
                # used for CBGT level configuration for core-core CBGT
**.BGeNBLocalNumber = ${CPV2= 26,65,104 ! CPV1}
                # used for CBGT level configuration for local-local CBGT, should match
CPV1 configurations

**.LocalSGW.cpu.hardProcTime = ${CAP1= 10,75,150 }
**.LocalPGW.cpu.hardProcTime = ${CAP2= 10,75,150 ! CAP1}
**.CoreSGW.cpu.hardProcTime =  ${CAP3= 10,75,150 ! CAP1}
**.CorePGW.cpu.hardProcTime =  ${CAP4= 10,75,150 ! CAP1}
                # These parameter are for GWPT configuration, all should match for
proper configuration
```

Second, in INI files of scenarios (the ones in table above) contain configuration of CORR parameter:

```
**.eNB[*].sctpApp[0].numRequestsPerSession = ${CPM2 = 233,465,698 ! CPM}
```

```
**.eNB[*].sctpApp[0].thinkTime = ${CPM=0.006451613,0.003225806,0.002150538}s
```

CPM2 contains number of request during simulation, computed from simulation time/CORR rate.

CPM contains CORR rate per second, computed from table described in thesis chapter 5 which specifies CORR rates

Third, in BGHscript.ini file contain configuration for DBGT parameter:

```
**.BGH1.udpApp[*].messageLength = 4250B
**.BGH2.udpApp[*].messageLength = 4250B
**.BGH1.udpApp[*].sendInterval=exponential(${BGT1=0.000017,0.0000068,0.00000425}s)
**.BGH2.udpApp[*].sendInterval=exponential(${BGT2=0.000017,0.0000068,0.00000425 ! BGT1}s)
```

These lines specify packet size of DBGT packets and sending rate. From these two values, you can compute the average link utilization through: (packet size in bits)/(sending rate*10^10 (10 Gbps))

Fourth, in CP_BGT_UDP_local.ini and CP_BGT_UDP_core.ini are the script for controlling CBGT operations. However, the script is designed to be controlled though commands in “prefix.ini” file:

```
**.BGeNBNumber = ${CPV1= 26,65,104}
**.BGeNBLocalNumber = ${CPV2= 26,65,104 ! CPV1}
```

Where `BGeNBNumber` and `BGeNBLocalNumber` control number of CBGT nodes to increase or decrease CBGT level. There are two configurations for CPV1 and CPV2, one for MMECAP-1 and the other for MMECAP-2. Due to limited features in simulator each have to be done separately from other.

MMECAP-1: *CPV1= 13,32,52*

MMECAP-1: *CPV1= 26,65,104*

The values are extracted from CBGT MME loading level and request rates described in table 5-2 and table 5-3 in chapter 5.

Running the simulation is done through “Run configuration manager” where we specify name of INI file and number of parallel runs, typically 10 parallel runs. Also “run number” field is assigned “*” to indicate the full combinations of assigned parameters, that is parameters: `DEL, MME, CPV1, CAP1, BGT1, and CPM2.` The run manager will take care of all possible combinations of these factors while running the simulation experiments. That means it will produce a loop of each factor with regard to other factor levels and produce the full combinations.

After running the experiments, results are stored in “result-dir” folder where it is in “.sca” and “.vec” files formats. For extraction of these results we require a tool that is used with OMNeT++ and is called “scavetool”. This tool is used in a bash script as follows:

```
declare -a DEL=("DEL=0.1" "DEL=0.5" "DEL=1.0")
declare -a GWPT=("CAP1=10" "CAP1=75" "CAP1=150")
declare -a BGT=("BGT1=0.000017" "BGT1=0.0000068" "BGT1=0.00000425")
….
for cpv in 0 1 2; do
  for cpm in 0 1 2; do
```

```
  for mme in 1 2; do
    for bgt in 0 1 2; do
      for del in 0 1 2; do
        for GWPT in 0 1 2; do
          for var in {0..9}; do # replications
….
scavetool vector -p "name(*End2End*) AND file(*${DEL[$del]}*${MME[$mme-
1]}*$CPVX*${GWPT[$GWPT]}*${BGT[$bgt]}*$CPMX*-$var.vec)"
-V -O "$Output" -F csv
*${DEL[$del]}*${MME[$mme-
1]}*$CPVX*${GWPT[$GWPT]}*${BGT[$bgt]}*$CPMX*-$var.vec
…
          done
        done
      done
    done
  done
 done
done
```

The “name” field is used to specify the name of variable used for storage of required value from results. In this example “End2End” is the name of the vector used for storing E2ED results. The “file” field is used to specify name of file with the required parameters values. For example, when “${DEL[$del]” is equal to “DEL=0.1” files with this specified parameter are extracted. In the “file” field it is specified each parameter value and that leads to one file that is processed in each time. Then the output is saved into specified name “$Output”.

After extraction of raw results from “.sca” and “.vec” files, the results are read using MATLAB code, which handles all remaining computation of average, confidence intervals, plotting of results, and any other method used in this work.

REFERENCES

[1] “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access” Release 10, 3GPP, TS 23.401, 2014.

[2] 3GPP TS 29.274. Evolved General Packet Radio Service (GPRS) Tunneling Protocol for Control plane (GTPv2-C); Stage 3 (3GPP TS 29.274 version 10.12.0 Release 10).

[3] 3GPP TS 36.413. Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP) (3GPP TS 36.413 version 12.3.0 Release 12)

[4] Karagiannis, Georgios, et al. "Mobile Cloud Networking: Virtualisation of cellular networks." Telecommunications (ICT), 2014 21st International Conference on. IEEE, 2014.

[5] Hayashi, Toshiki. "Evolved Packet Core (EPC) Network Equipment for Long Term Evolution (LTE)." FUJITSU Sci. Tech. J 48.1 (2012): 17-20.

[6] Alcatel-Lucent; The Impact of Small Cells on MME Signaling, METHODS TO REDUCE AND OPTIMIZE MME CORE SIGNALING CAUSED BY SMALL CELLS. APPLICATION NOTE, October 2013.

[7] Lara, Adrian, Anisha Kolasani, and Byrav Ramamurthy. "Network innovation using openflow: A survey." (2013): 1-20.

[8] Mendonca, Marc, et al. "A survey of software-defined networking: Past, present, and future of programmable networks." (2013).

[9] Ali-Yahiya, Tara, and Khaldoun Al Agha. Understanding LTE and its Performance. Springer, 2011.

[10] CELTIC/MEVICO D2.1; “Architectural EPC extensions for supporting heterogeneous mobility schemes”; 2013.

[11] OpenFlow Switch Specification, Version 1.1.0 Implemented (Wire Protocol 0x02). [Online]. Available: http://www.openflow.org/ documents/openflow-spec-v1.1.0.pdf.

[12] Naudts, Bram, et al. "Techno-economic analysis of software defined networking as architecture for the virtualization of a mobile network." Software Defined Networking (EWSDN), 2012 European Workshop on. IEEE, 2012.

[13] Basta, Arsany, et al. "A Virtual SDN-enabled LTE EPC Architecture: a case study for S-/P-Gateways functions." Future Networks and Services (SDN4FNS), 2013 IEEE SDN for. IEEE, 2013.

[14] Hampel, Georg, Moritz Steiner, and Tian Bu. "Applying software-defined networking to the telecom domain." Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on. IEEE, 2013.

[15] Yap, Kok-Kiong, et al. "Blueprint for introducing innovation into wireless mobile networks." Proceedings of the second ACM SIGCOMM workshop on Virtualized infrastructure systems and architectures. ACM, 2010.

[16] Li, Li Erran, Z. Morley Mao, and Jennifer Rexford. "CellSDN: Software-defined cellular networks." Computer Science, Princeton University, Princeton, NJ, USA, Tech. rep (2012).

[17] Katanekwa, Nicholas, and Neco Ventura. "Mobile content distribution and selective traffic offload in the 3GPP evolved packet system (EPS)." Information Networking (ICOIN), 2013 International Conference on. IEEE, 2013.

[18] Pentikousis, Kostas, Yan Wang, and Weihua Hu. "Mobileflow: Toward software-defined mobile networks." Communications Magazine, IEEE 51.7 (2013).

[19] Kempf, James, et al. "Moving the mobile evolved packet core to the cloud."Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on. IEEE, 2012.

[20] Said, Siwar Ben Hadj, et al. "New control plane in 3GPP LTE/EPC architecture for on-demand connectivity service." Cloud Networking (CloudNet), 2013 IEEE 2nd International Conference on. IEEE, 2013.

[21] Olsson, Magnus, et al. SAE and the Evolved Packet Core: Driving the mobile broadband revolution. Academic Press, 2009.

[22] Gurusanthosh, P., A. Rostami, and R. Manivasakan. "SDMA: A semi-distributed mobility anchoring in LTE networks." Mobile and Wireless Networking (MoWNeT), 2013 International Conference on Selected Topics in. IEEE, 2013.

[23] Jin, Xin, et al. "SoftCell: scalable and flexible cellular core network architecture."Proceedings of the ninth ACM conference on Emerging networking experiments and technologies. ACM, 2013.

[24] Liu, Binghan. "Software Defined Networking and Tunneling for Mobile Networks." (2013).

[25] Brief, ONF Solution. "OpenFlow-enabled SDN and Network Functions Virtualization." (2014).

[26] Sherwood, Rob, et al. "Flowvisor: A network virtualization layer." OpenFlow Switch Consortium, Tech. Rep (2009).

[27] Philip, Venmani Daniel, Yvon Gourhant, and Djamal Zeghlache. "OpenFlow as an Architecture for e-Node B Virtualization." e-Infrastructure and e-Services for Developing Countries. Springer Berlin Heidelberg, 2012. 49-63.

[28] Yousaf, Faqir Zarrar, et al. "SoftEPC—Dynamic instantiation of mobile core network entities for efficient resource utilization." Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013.

[29] Cerroni, Walter, Gaia Leli, and Carla Raffaelli. "Design and test of a software defined hybrid network architecture." Proceedings of the first edition workshop on High performance and programmable networking. ACM, 2013.

[30] Shimizu, Takashi, et al. "An experimental evaluation of dynamic virtualized networking resource control on an Evolved mobile core network: A new approach to reducing massive traffic congestion after a devastating disaster."Humanitarian Technology Conference (R10-HTC), 2013 IEEE Region 10. IEEE, 2013.

[31] Sankaran, C. B. "Data offloading techniques in 3GPP Rel-10 networks: A tutorial." Communications Magazine, IEEE 50.6 (2012): 46-53

[32] Wang, Meng, Michael Georgiades, and Rahim Tafazolli. "signalling cost evaluation of mobility management schemes for different core network architectural arrangements in 3GPP LTE/SAE." *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*. IEEE, 2008

[33] A. Varga. INET Framework for the OMNeT++ Discrete Event Simulator. http://github.com/inet-framework/inet, 2012.

[34] Alcatel.Lucent. The Impact of Small Cells on MME Signaling: METHODS TO REDUCE AND OPTIMIZE MME CORE SIGNALING CAUSED BY SMALL CELLS, APPLICATION NOTE. October 2013.

[35] 3GPP TR 25.912. Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (3GPP TR 25.912 version 12.0.0 Release 12).

[36] Jarschel, Michael, et al. "Modeling and performance evaluation of an OpenFlow architecture." Proceedings of the 23rd international teletraffic congress. International Teletraffic Congress, 2011.

[37] Basta, Arsany, et al. "Applying NFV and SDN to LTE mobile core gateways, the functions placement problem." Proceedings of the 4th workshop on All things cellular: operations, applications, & challenges. ACM, 2014.

[38] Gabriel Brown. LTE/SAE & the Evolved Packet Core: Technology Platforms & Implementation Choices." White Paper, Heavy Reading. April, 2009.

[39] OMNeT++, Discrete Event Simulator, available at http://www.omnetpp.org ,2014

[40] Natrella, Mary. "NIST/SEMATECH e-handbook of statistical methods." (2010).

[41] R. Jain, "The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling," Wiley-Interscience, New York, NY, April 1991, ISBN:0471503361.

[42] "Policy and charging control architecture" Release 10, 3GPP, TS 23.203, 2013.

VITAE

Name: Ahmad Azmi Abo Naser

Nationality: Jordan.

Date of Birth : 11/12/1987

Email : abonaser.12.aa@gmail.com

Address: King Fahd University of Petroleum and Minerals Dhahran, 31261, Saudi Arabia.

EDUCATIONAL QUALIFICATION:

MASTER OF SCEINCE, COMPUTER ENGINEERING

April, 2015

King Fahd University of Petroleum and Minerals

Dhahran, Saudi Arabia

Bachelor of Communication and Electronics Engineering, with First Honors

September 2008

Jordan University of Science and Technology

Irbid, Jordan

PROFESSIONAL EXPERIENCE

December 2009 – May 2012

Production Engineer, SilverTech ME FZCO

Dubai, UAE