2017-09

Assay of IP mobility management in SDN based mobile network architecture

Dawood, M

http://hdl.handle.net/10026.1/13874

Assay of IP Mobility Management in SDN based Mobile Network Architecture

Dawood M[1, 2], Fuhrmann W[1, 2], Ghita BV[2]

[1]Faculty of Computer Science, University of Applied Sciences Darmstadt, Germany
woldemar.fuhrmann@h-da.de
[2]Centre for Security, Communications and Network Research (CSCAN),
Plymouth University, Plymouth, PL4 8AA, United Kingdom
e-mail: {muhammad.dawood, bogdan.ghita}@plymouth.ac.uk

Abstract—The evolution towards mobile networks flat architecture entreat a key-role for IP mobility management in providing the ubiquitous always-on network access services. This paper provides prospects for efficient mobility management in SDN plus mobile network architecture and describe important call control flows in inter system handover scenario. Distribution of gateway function approach has been followed and evolved with SDN technology. Key improvements with proposed architecture are to support seamless mobility in heterogeneous access environments, remove the chains of IP preservation and optimal data path management according to application needs. The paper assays the proposed evolution in terms of numbers of signaling messages processed by control entities for an inter system handover procedure relative to current mobile network architecture.

Keywords— Software Defined Mobile Network; Intersystem Inter networking; SDN based DMM; IP mobility management;

I. INTRODUCTION

The rapid growth of mobile data traffic has been widely recognized and reported and the mobile communications industry is preparing to cope with a 1000x increase of traffic by 2020 over 2010 [1]. More and more people see their handheld devices as an annex of their workspace while on move and the continuous improvement to the mobile network architecture is becoming increasingly important to support the performance requirements for the ubiquitous wireless connections. Specifically to support different levels of Mobile Unit (MU) mobility, future wireless access infrastructure is required to a) Support integration of heterogeneous Radio Access Technologies (RATs), the 3GPP Long Term Evolution (LTE) [2] and trusted/untrusted non-3GPP access types (for example Wireless LAN) efficiently combining multiple simultaneous connections of MU via multiple access nodes with unified mobility management b) provide seamless IP mobility for session as well as service continuity as per application needs. c) maintaining optimal service level quality for services that have different latency requirements between the MU and the Packet Data Network (PDN); and d) Support optimized mechanisms to control signaling overhead i.e. Minimize the number of messages required to enable the traffic exchange between the MU and the PDN.

Current and upcoming wireless communication systems exploit many techniques to meet these requirements. However, the current mobile network architecture is highly hierarchical and centralized, several limitations have been identified when compared to the always-on network access service requirements of seamless mobility. With centralized network architecture, since the MU uses the single address anchored at central IP anchor, user traffic will always need to go first to the home network and then to the correspondent service node even it is not connected via home access network, leading to paths that are in general longer than the direct one between the MU and its correspondent service node. This poses excessive traffic concentration on a single gateway element and possibly un-optimized routing adding in turn unnecessary delay and wasting resources. [3]. To cope with problems of centralized mobility management: the concept of distributed mobility management (DMM) has been proposed in contrast to centralized approach in a hierarchical model. The basic concept of DMM is the distribution of mobility and IP anchors and select and re-locate when necessary gateways that are topologically/geographically close to the MU. DMM architecture solves some of the problems of centralized mobility management, however when a MU moves to a new set of anchors, due to IP relocation, either tunneling need be used between initial router and new router or the active flows have to maintained until the flow is finished. Consequently, it may lead to a complex process and a high signaling cost.

During the recent years Software Defined Networking (SDN) [4] have gained a lot of attention from the research community and standardization organizations. SDN provides flexible ways to monitor and manage network efficiently with separation of data plane and control plane. In line with these trends we set out to take up the current mobile network architecture what we refer as SDN Plus Mobile Network (S+ MN), combined with configurable SDN technologies to solve the many of issues described above. To limit the potentially huge design space and the envisioned research towards distribution of gateway functions we focus mainly on seamless mobility in considered scenario, which include maintaining QoS, smooth handover, session continuity and sharing across multiple wireless technologies. Although there have been several proposals in this field as surveyed in [5-8] but these studies have

only proposed that SDN can provide much-needed benefits but have not shown the details for the evolved architecture. This paper not only proposed the S+ MN Architecture but also provide realization of core network functions with details of procedure commonly occur in current mobile networks elaborated with the separation of data and control plane. To summarize, the main contributions of this paper can be stated as follows:

- Improvements for SDN based mobile network with details of functional elements.

- Description of the procedures of current mobile networks to SDN context using the proposed S Plus controllers; Initial attachment, service and tunnel setup, intersystem handover and the data plane management in a distributed manner, etc.

- By analysing current research of IP mobility management in SDN based mobile network architecture; the proposed architecture implied Multipath TCP (MPTCP) functions to remove the chains of IP address preservation for session continuity during IP anchor relocation without the use of bi-directional tunnels between the initial and new router and avoiding a large number of signalling messages.

- Development of a functional validation setup to show the working of S+ MN architecture, application oriented forwarding of IP traffic through the use of distributed IP anchors.

The rest of this paper is organized as follows. The paper first gives an overview of related concepts and initiatives within Open Networking Foundation (ONF) [9] and 3GPP that are complement to S+ MN. Then introduces the S+ MN architecture, key features, interpretation of the internal functions with evolution of procedures of current mobile networks elaborated in SDN based mobile networks, analysis of current research towards IP mobility management and proposed improvements. In the next section it provides the functional validation and evaluation of proposed S+ MN architecture relative to a traditional architecture. The last section concludes the paper and discuss research directions.

I.   RELATED WORK

A.   Software Defined Networking (SDN):

SDN features various ways of using software to manage and manipulate network devices in which the control of gateway functions are lifted up to a logically centralized entity called SDN controller. The key concepts of SDN include the separation of the control and data plane, flexibility, high rate of innovation, and network programmability that speeds the deployment of new services. Figure 1 depicts a typical architecture of SDN.

Data plane comprises a set of forwarding devices, (e.g., virtual switches or physical switches). The control plane consists of SDN controllers and the application plane consists of one or more applications, such as routing, and monitoring applications. The SDN applications communicate their network requirements towards the controllers via northbound interfaces. The controllers communicate with each other using east/westbound
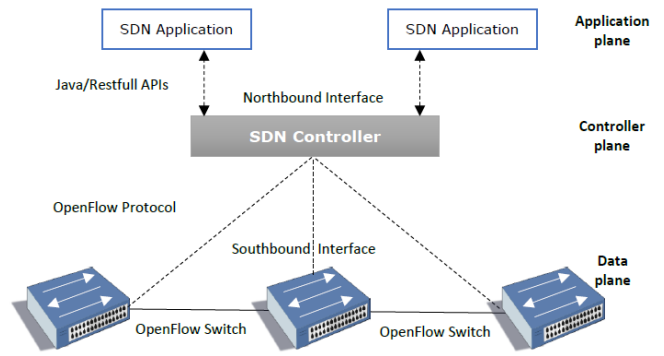

Figure 1: Typical Open SDN Architecture

interfaces. An open protocol is used to configure, monitor and manage forwarding devices through southbound interfaces. OpenFlow [10] protocol is the most extended SDN interface maintained by the Open Networking Foundation (ONF). The ONF has established the Wireless and Mobile Working Group to explore how the OpenFlow can adapt for use in mobile networks. In a solution brief [9] they illustrate the benefits of SDN/OpenFlow with use cases of inter-cell interference management, Mobile traffic management and envisioned that the SDN/OpenFlow benefits can also be realized throughout the mobile networks from access to backhaul and to the EPC.

B.   Multipath TCP (MPTCP)

MPTCP is an ongoing effort within Internet Engineering Task Force (IETF) to support multipath operation, a set of extensions to enables a regular TCP connection to use multiple different IP addresses and interfaces [11].

In the mobility context, when MU moves from one point of attachment (in 3GPP terminology PGW) to another i.e., it receives or configures a new IP address through new network attachment. MPTCP enable multiple IP addresses by adding subflows. Each sub-flow behaves as a separate regular TCP connection inside the network. Subflows can be added and removed at any point of time, in any MPTCP ongoing communication, with the help of ADD_ADDR option and REMOVE_ADDR option for any interface [11]. To maintain the
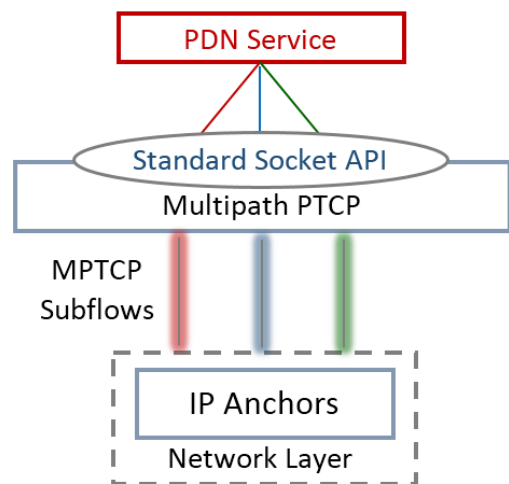

Figure 2: Multipath TCP multipath operations

ongoing communication MPTCP support "make before break" method and uses MP-PRIO option to specify any subflow as backup mode. In Fig. 2,, in the mobility scenario, With the MU have multiple IP addresses so in this case MPTCP can create multiple subflows and the MU is connected to Packet Data Network (PDN) service. Defined by MPPRIO option MPTCP support different flow modes, in the single-path mode only one TCP sub-flow is used at any time or using all subflows simultaneously between two communication nodes or uses only a subset of subflows for transmission of data packets. this paper explore the implied key role of transport layer protocols for IP mobility management by simultaneous exchange of IP traffic flows through the use of multiple distributed IP anchors.

II. S+ MOBILE NETWORK ARCHITECTURE

Using the concepts of separation of control and data plane, we design an evolution of current mobile network architecture what we refer as SDN Plus Mobile Network (S+ MN). Figure 3 shows the conceptual view of S+ MN architecture, data plane enable the user traffic to be forwarded through the RAN to core network gateways which terminate the traffic to external network. SDN Plus Radio access network Controller (S+ RC) and SDN Plus core Network Controller (S+ NC) are proposed, the S+ controllers (S+ RC and S+ NC) can control the related nodes through programmable interfaces and have the ability to interact both access as well as forwarding nodes by consolidating redundant Access Stratum (AS) and Non-access Stratum (NAS) signaling and management functions.
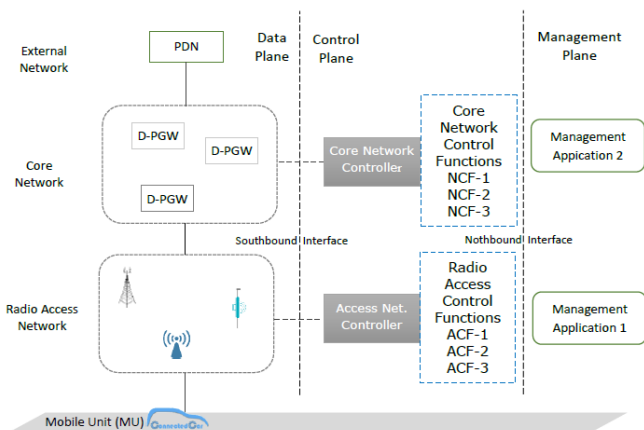

Figure 3: Conceptual view of S+ Mobile Network Architecture

The RAN comprises the heterogeneous access environment that includes multiple different RATs links such as LTE (3GPP RAN), Wi-Fi (non 3GPP trusted RAN) and any new RAN. New communication entity S+ Access Nodes can be presented that includes the collection of logical components of multiple different RATs nodes. The evolved access nodes directly interacts with the physical radio resources and should enabled by the attributes of SDN (simple OpenFlow switch shall be added to these devices in order to support Open SDN protocols). The S+ RC shall acts as a network administrator, coordination, policing and configuration of the infrastructure. The network control information is delivered from S+ RC to access nodes through the standardized southbound interface, i.e. OpenFlow. S+ RC shall apply access network-wide functions such admission control, radio resource scheduling, backhaul links

management and QoS etc. In considered context to provide mobility support with Intersystem inter networking be applied the MUs can obtain connectivity either from 3GPP or non-3GPP access node and move between them. Being the research focused on provision of IP session continuity during GW relocation and due to the paper length limitations, we cannot delve into details of S+ RAN and mainly describe the realization of S+ core network.

The Core network of S+ MN is composed of simple network gateways in contrast to legacy EPC, explicitly separate the data plane from control plane. Furthermore, it applied the distribution of gateway function. S+ NC manage the forwarding paths, calculates, installs rules into the user forwarding elements, path management, tunnel management, etc. In the following we provide details of S+ MN core network with evolution of procedure of current mobile networks elaborated with the separation of data and control plane.

III. REALIZATION OF S+ MN CORE NETWORK

The realization of S+ MN core network architecture is depicted in Figure 4, the S+ NC is the brain of the operations. All EPC control functions are realized as set of control plane applications (MME, SGW-C, PGW-C) running on top of the S+ NC. The application uses the northbound interface to access the substrate from the S+ NC. Depending on the implementation of control plane and requirements of supported services these applications can have different functionalities. The MME control function, can interact with subscription and content server named HSS+ to create a local copy for user authentication and authorization information. For mobility management MME tracks and updates the user's location information as well as their requirements. It delegates various event to S+ access node such as initial context setup request and response for user attachment. MME can also interact with SGW-C directly and PGW-C indirectly to setup user session. Additionally, we propose coordination between MME and ANDSF to periodically collect and analyse reports about the network state in order to take the decision about handovers and trigger the handover by sending a Handover Request message to the target access node.

The SGW-C performs the control plane functionalities of mobility anchor (SGW) including installing/uninstalling OpenFlow rules into the SGW-U switch. It also acts as an
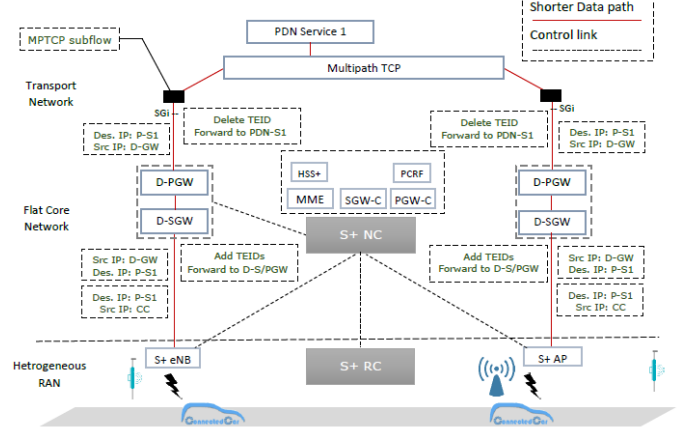

Figure 4: Realization of the S+ MN core network architecture

interface between MME and PGW-C sends a paging message to each access node belonging to the Tracking Area (TA) where the MU is registered and assign appropriate forwarding gateway. The PGW-C inherits the control functions performed by the PGW such as IP assignments, bearer setup, Allocation of the tunnel endpoint identifiers (TEIDs), establishing GTP tunnels and QoS provisions etc. The PGW-C needs to interact with the PCRF server to get the corresponding policies in the case where the session requires a specific QoS policies.

The SGW-C and PGW-C control applications can be combined as a single application. In this case the combined control function of SGW and PGW allocates the tunnel IDs for user flows, acting as a mobility anchor for inter system communication. Managing the forwarding elements, calculates and installs rules into the user forwarding elements, path management, tunnel management and processing GTP packets. The SGW-C and PGW-C function will use a northbound interface (e.g., REST API) with the S+ NC which translates northbound messages into OpenFlow messages. When the flow rules are associated for subsequent packets that belong to the same section and installs them for the data plane. Furthermore in order to support fully flexible IP flow mobility during GW relocation SGW-C and PGW-C applications can have additional functionalities. Such as, provide distinct ID address for each flow/bearer. The default bearer, if required, is anchored in the core at initial GW with a fixed IP address of the user so that the user is always reachable for incoming requests. Other IP flows/bearers may be anchored with temporary IP addresses at local GWs supporting session continuity when needed and decided upon by the user. For application oriented forwarding setup, change the forwarding rules of the data plane switches for a specific flows or for all flows of the handled MU.

Data plane is composed of simple network gateways named Data Plane Gateways (D-GWs) in contrast to legacy EPC, explicitly separate the data plane from control plane. Distribution of gateway function approach has been followed and evolved with SDN technology. Using 3GPP terminology each D-GW implements Data plane Serving Gateway (D-SGW) or Data plane PDN Gateway (D-PGW) functions and communicates with an S+ NC through OpenFlow protocol that implements the southbound interface. D-GWs use the flow rules received from the S+ NC and have standard mobile network tunnel processing capabilities with a GTP-U extension (e.g., encapsulation/decapsulation facilities) and forward packets to the transport network. Collocated D-SGW and D-PGW is adopted to avoid unnecessary longer data path. Data plane may also include the key functionality of a wireless access node, such as an interface to manage radio bearers. During an inter GW handover or gateway relocation, the S+ NC needs to send OpenFlow Packet Out messages to the source as well as target GW in order to modify their flow tables at the same time.

IV. EVALUATION METHODOLOGY

We discuss a scenario used to explain the behaviour is the inter system handover of a MU from a 3GPP access network towards a non 3GPP access network and describe the important call control flows being exchanged across the various network elements of the proposed architecture to show how it can be systematically exploited to gain benefits.

A. Initial attachment procedure:

The Figure 5 shows the signalling diagram of initial attachment, illustrating an expression of proposed architecture. The 3GPP/LTE compliant evolved access node S+ eNB receives an attachment request message (1) from MU. This message is embedded in an OpenFlow Initial User message (2) and is sent to the S+ NC, where the proposed MME application initiates MU authentication and fetches users subscription profile from HSS+.
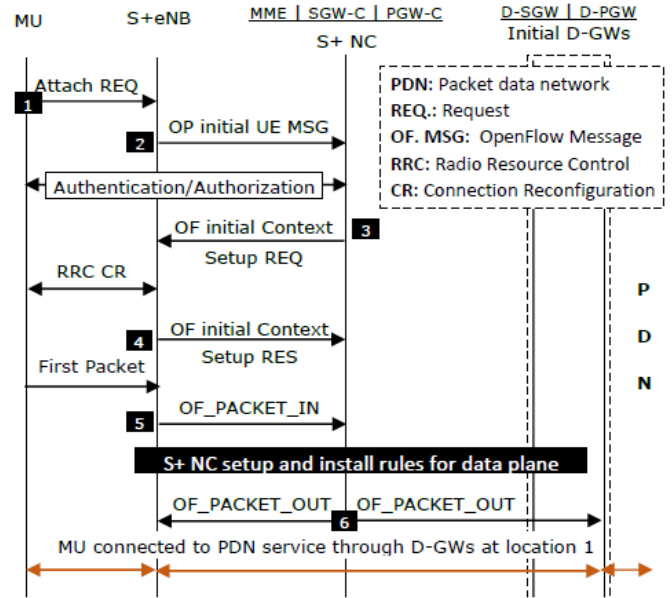

Figure 5: Initial attachment operation

After the successful authentication and authorization the OpenFlow Initial Context Setup Request message (3) is sent from the S+ NC to the S+ eNB. The S+ eNB reconfigures the radio connection, and forward the attach accept message to the MU. The acknowledgment of attach complete from MU is included by S+ eNB as context setup response message (4) to the controller. The state of MU transitions from deregistered to registered and transmits the first session packet to the S+ eNB over the radio connection. Since, there is no matching flow entry in the flow table of the S+ eNB. The S+ eNB triggers an OpenFlow packet-in message (5) to the S+ NC. This message includes some information that is necessary to establish the data plane. The S+ NC analyse the packet header to obtain the session information such as the source and destination nodes IP addresses and collects the tunnel information. S+ NC assign the D-GWs by interacting with the SGW-C and the PGW-C applications. In turn the MU obtain an IP address from PGW-C, this also includes as conferred in [12] creating Binding Cache Entry (BCE) to keep track the MU's location as well as the related information. Then, the S+ NC creates flow rules for subsequent packets that belong to the same section and send them as OpenFlow packet-out message (6) to installs them for the S+ eNB and D-GWs. In the case where the session requires a specific QoS policy, the S+ NC needs to interact with the PCRF server through PGW-C functions to get the corresponding policies before downloading the flow rules to the data plane. When the flow rules are associated with a QoS

parameter, the S+ NC will install them at the D-GWs. Furthermore considering the expected handover with MU mobility, The MME in cooperation with the S/PGW control functions notify the needed support for GW relocation and keep the established PDN-Connection context. At the end of this procedure, the MU state is transitioned from Idle to Connected and a data forwarding path is established from the MU to the correspondent PDN service

B. Inter System Handover IP mobility management

The As shown in Figure 6, the MU is initially connected to 3GPP access network, when the MU moves away from source S+ eNB and determine the presence of appropriate access node (in this case a non-3gPP AP), S+ eNB sends a Handover required message packaged in an OpenFlow message to the S+ NC (1).

The MME function in turn trigger the handover by sending a Handover Request message (2) to the target non-3GPP compliment access node (S+ AP). If targeted S+ AP response with handover ACK message (3) to the S+ NC, the handover command (4) is send toward the MU by S+ NC and then MU trigger the attachment procedure (Access technology specific procedure for the interworking of the 3GPP EPC and the non-3GPP access networks) [14]. After attachment S+ NC receives packet-in message (5) from S+ AP and it must connect a set of appropriate D-GWs that are close to the MU. Together with this procedure MU gets a new IP address to be used in the new flow. S+ NC allocates new IP address to MU and have to establish a data forwarding path according to application needs.
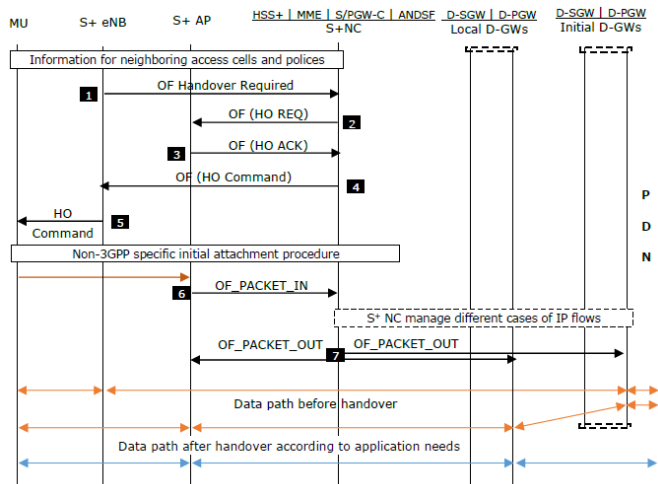

Figure 6: S+ MN based inter system handover

C. IP mobility management

There have been several approaches for IP mobility management in SDN based mobile networks, as surveyed in [12, 13] , In, [12] the author applied the SDN concept to DMM architecture for routing optimization with a DMM service within SDN controller. In proposed architecture, when a user attach initiate router, the SDN controller stored user information in Binding cache entry (BCE). If the user moves to the new router, the controller which receives packet-in message will check BCE. Mobility management is supported to user by binding update with DMM service. The new data path is set by SDN controller and it sends Flow Modify message to previous and new router. On receiving flow modify messages, the routers will update their flow tables. That is, user can be supported optimized path by flow table, without tunneling. But It is noted that after handover, first, the packets have to follow the route from the previous to the new router. The packets are then redirected between the new router and fInally to the user. Consequently, it may lead to a complex process and a high signaling cost.

In [13], the author considered two schemes for mobility management in DMM scenario called Tunneling mode and Optimal mode DMM, in first scheme different prefixes are allocated to user at old and each new attachment nodes. For the ongoing flows which require mobility support, the list of old and new prefixes are included in packet-out message named router advertisement message. The controller updates the forwarding rules at both old and new OpenFlow switches (OFSs). Using tunneling mechanism the flows are being redirected between distributed OFSs. In second scheme, the controller calculates the new route for the on-going flows and populates the new forwarding rules to all the intermediate OFSs along the new route between the new OFS and the Correspondent Node (CN). In other words, an explicit path between the user node and CN is defined and established by the controller. In this way, tunneling mechanism can be avoided. However, it may lead to a complex task giving the large number of switches/flows to handle. Besides, the simplicity of the first scheme can come at the cost of tunneling overhead and sub-optimal routing. the authors indicated that the optimal mode is likely more suitable for latency sensitive services while the tunnel mode seems to be better for the packet loss-and interruption-sensitive services.

To fully appreciate these particular challenges associated with IP mobility, the proposed architecture implied MPTCP at transport layer. Since MPTCP does not change the socket API, it can be used tranparently by any PDN service and change of IP address of a MU does not force the connection to be restarted. A combination of SDN capabilities and MPTCP options could be used to remove the chains of IP preservation of current mobility management solutions and disburden the process of flow forwarding during GW relocation. With implied MPTCP connection of MPTCP capable MU and PDN-Service and will be able to synchronize the MU traffic using different IP addresses. As shown in figure 7, the S+ NC can manage different cases of IP flows, the data packets from service requiring static IP anchoring are directed to initial flow, so the MU is enabled to

SDN for optimal path managment, and MPTCP to remove the chains of IP perservation or tunneling
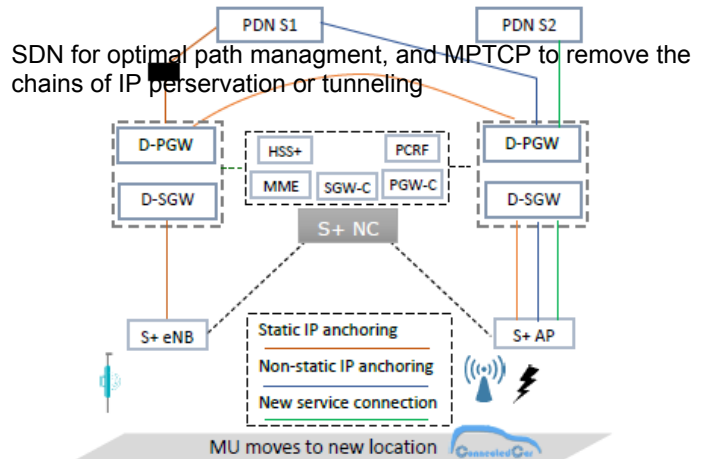

Figure 7: Flow setup to application needs

keep the initially assigned IP address despite its location changes and where no static anchoring is required, the MU uses the new flow for active communication, while maintaining the reachability for the IP address that is still in use. In this way the need for the maintaining a tunnel between source and target anchors and flow redirection is not required to link different flows, leading in turn to seamlessly session continuity during IP anchor relocation avoiding a large number of signalling messages. The S+ NC monitor the MU activity, as no traffic is carried in the initial subflow (during the timer interval) it starts the releases procedure for the removal of initial IP address from the PDN-S1 IP list. The release cause tag along the session management and perform the GW binding information update.

V. FUNCTIONAL VALIDATION

To evaluate the working and key benefits of the proposed S+ Mobile Network architecture, we developed a functional validation test setup which primarily consists of Floodlight controller (for S+ NC), MySQL database (for HSS+), Open vSwitch (OVS), S+ Access Nodes and PDN Service modules. Using the northbound APIs provided by the Floodlight controller, the S+ NC module implements the LTE/EPC control functions (e.g. MME, SGW-C, and PGW-C). OVS instances acts as D-GWs using the OpenFlow protocol and apply forwarding rules provided by S+ NC in distributed manner. The source code of implemented modules is available via [15]. The simulation starts with a MU initially attached to the PDN service 1 via 3GPP compliant access network, then after a few seconds the MU starts moving out of S+ eNB coverage area. The MU discovers and switches to non- 3GPP compliant access connection. This vertical handover from source to target access network undergoes D-GW relocation. We measured, i) The time required by the MU to change from S+ eNB to S+ AP. ii) The total time consumed by MU in dissociation from the S+ eNB and becoming ready to route network traffic through S+ AP, that is, the interval between the last Ping packet received or sent by the MU before the handover and the first Ping packet received or sent after the handover. However, it does not consider the duplicate address detection process that should be performed after configuring an IPv6 address on an interface. For each handover case, we performed the measurement of network traffic at MU's S+ AP interface for more then 100 handovers. Figure 8 depicts detailed handover distribution pattern.

VI. CONCLUSIONS

This paper presented evolution of the mobile network architecture in SDN context and provided improvements for DMM to enable seamless handover while roaming across heterogenous access networks, novel perspective to maintain session continuity during the movement of MU that tag along with GW relocation. The MPTCP protocol is implied to remove the chains of IP preservation of current mobility management solutions, leading in turn to seamlessly change the IP address during a connection without breaking and re-establishing the connection. A functional validation setup have been devolved
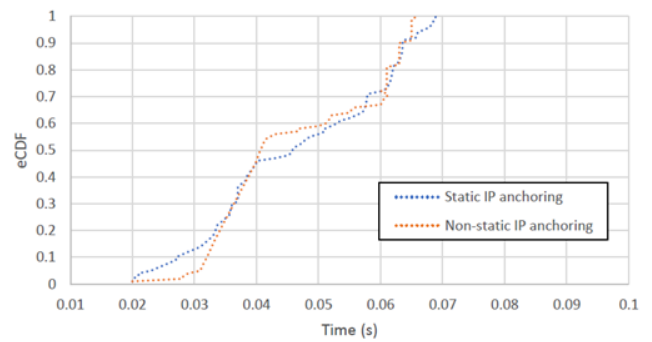

Figure 8: Empirical CDF of handover measurements for ping traffic

to show the working of proposed architecture. The pictorial call control flows thus created has managed different cases of IP flows and shown a significant decrease in signaling and delay compared to tunnel and routing based approaches of IP mobility management. MPTCP provides an overall umbrella and a longer evolution perspective of using simultaneously multiple network path, In future work we intend to perform more detailed experiment with MPTCP to illustrate further system realization of the architecture presented.

REFERENCES

[1] C1. Index, C., Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018. 2015.

[2] 3GPP, TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN).

[3] H. Chan, D.L., P. Seite, H. Yokota, J. Korhonen, , IETF RFC 7333 "Requirements for Distributed Mobility Management. 2014.

[4] Nguyen, V.-G., T.-X. Do, and Y. Kim, SDN and Virtualization-Based LTE Mobile Network Architectures: A Comprehensive Survey. Wireless Personal Communications, 2016. 86(3): p. 1401-1438.

[5] Karimzadeh, M., L. Valtulina, and G. Karagiannis, Applying sdn/openflow in virtualized lte to support distributed mobility management (dmm), in In The 4th international conference on cloud computing and services science (CLOSER) 2014. p. 1–6.

[6] Sama, M.R., et al. Enabling network programmability in LTE/EPC architecture using OpenFlow. in Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on. 2014.

[7] Yaz, V., U.C. Kozat, and M.O. Sunay, A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management. IEEE Comm. Magazine, 2014. 52(11): p. 76-85.

[8] Sama, M.R., et al., Software-defined control of the virtualized mobile packet core. IEEE Communications Magazine, 2015. 53(2): p. 107-115.

[9] ONF, OpenFlow Enabled Mobile and Wireless Networks, ONF Solution Brief. 2013.

[10] ONF, TS-020: OpenFlow switch specification version 1.5.0.

[11] Ford, A., et al., TCP Extensions for multipath operation with multiple addresses, in RFC 6824, IETF. 2013.

[12] Hyunsik Yang , K.S., Younghan Kim, IETF RFC: Routing Optimization with SDN.

[13] Nguyen, T.T., C. Bonnet, and J. Harri. SDN-based distributed mobility management for 5G networks. in 2016 IEEE Wireless Communications and Networking Conference. 2016.

[14] 3GPP, TS 23.402 Architecture enhancements for non-3GPP accesses. 2016.

[15] Dawood, M. and M.S. Shafi. S plus mobile network archtecture implementation. 2017; Available from: https://github.com/mshary/S-plus-Mobile-Network.