SDMA: A Semi-Distributed Mobility Anchoring in LTE Networks

P. Gurusanthosh*, A. Rostami†, R. Manivasakan*

*Optical Networks Lab, Indian Institute of Technology Madras, India. {gsanthosh.stu,rmani}@ee.iitm.ac.in
†Telecommunication Networks Group, Technical University of Berlin, Germany. rostami@tkn.tu-berlin.de

Abstract—We present a novel mobility management scheme for the Long Term Evolution (LTE) networks, where a part of mobility management tasks is transferred from the mobile core network (i.e., Evolved Packet Core (EPC)) to the backhaul network. In this approach, a part of inter-eNodeB handovers of user equipments (UEs) are managed by control nodes in the backhaul in contrast to the standard LTE, where such handovers are exclusively handled by the EPC. For this purpose, we utilize the concept of software-defined networking (SDN) and OpenFlow protocol to realize control nodes for the backhaul network, which facilitate implementation of sophisticated traffic management schemes. We present the architecture and associated protocols, and evaluate the performance gains and scalability of our proposed scheme. We find that the proposed scheme considerably reduces the mobility-related signaling load to EPC. Additionally, the use of SDN/OpenFlow eliminates the need for conventional mobility tunnels, which results in improving the efficiency of data transport in the network.

I. Introduction

Mobile communication networks play an ever-increasing role in providing broadband access to the Internet and data services. This has pushed over the last years for packet-oriented architecture in the cellular networks. One promising architecture in this direction is the Long Term Evolution (LTE)–which is a set of standard specifications for beyond 3G broadband mobile communications as the evolution of GSM/UMTS (Global System for Mobile Communications/Universal Mobile Telecommunications System) standards–developed by the 3rd. Generation Partnership Project [1]. In contrast to the previous standards, LTE enjoys a flat all-IP architecture, which greatly simplifies the network architecture and enables the integration of LTE with fixed broadband access technologies. Nevertheless, the support for user mobility in the LTE networks needs particular attention, since the IP network does not readily support the host mobilities. To address the user mobility in LTE networks, a tunneling approach based on the GPRS Tunneling Protocol (GTP) is adopted [1].

The GTP-based tunnels can mask the movement of endusers among different access points so that re-establishing an IP connectivity after every handover is avoided. This, however, comes at a cost of terminating and re-establishing the GTP tunnel between the LTE core network and corresponding access points (a.k.a. evolved node B or eNB) for

each handover. The disadvantages of such an approach include the signaling overhead to the core network associated with handovers, complexity of a central anchor point in the EPC, as well as the GTP-tunneling overhead for exchanged data packets. The situation becomes even worse taking into account that cellular network operators tend to utilize smaller cells for more efficient spectrum management, which in turn increases the frequency of handovers–particularly for users moving at higher speeds [2].

In order to address these issues we propose a semi-distributed mobility anchoring in LTE networks, where a part of mobility management tasks is delegated to the mobile backhaul network. Our motivation is based on the fact that in most cases when a user terminal is handed over to a new access point, a large part of the path that the traffic takes in the backhaul network would be the same before and after handover. Hence, it makes sense to transfer a part of handover management tasks from the core to the backhaul–closer to the mobile users–by dynamically placing the mobility anchor point at the nearest backhaul node to the UE, which is common in the route of traffic of the UE before and after handover. We intentionally make these transfers partial and selective, to accommodate those handovers which need specifically, the knowledge of network-wide traffic state. We term our scheme as Semi-Distributed Mobility Anchoring (SDMA) since the handover management functions are partially handled by distributed control nodes in the backhaul network (in contrast to standard LTE, where they are centralized).

To realize this approach, we need to address the following challenges. First, the control plane of the backhaul network should be aware of the traffic flows of the user terminals and their handovers. Second, once the backhaul network knows about a handover it has to identify an appropriate anchor point in the network and accordingly re-route the traffic of the user terminal to the right access point. To address these challenges, we propose to employ the concept of software defined networking (SDN) [3] and associated OpenFlow protocol [4] to realize control nodes for the backhaul network, which can implement sophisticated traffic management schemes. That is, making use of the SDN/OpenFlow, network operators would be able to manage traffic, to the extent of identification of individual traffic flows and implementing a per-flow scheduling, management and control of traffic in their networks. This, in turn, eliminates the need for GTP tunneling for mobility management.

The rest of this paper is structured as follows. In Section II,

we briefly review the mobility management procedure in LTE networks. In Section III, we introduce our proposed scheme for mobility management. Section IV presents an analytical modeling for signaling load on EPC in both SDMA and standard LTE. Section V presents numerical analysis of the proposed scheme and the comparison with the standard LTE. Finally, in Section VI we conclude the work by summarizing the achieved results.

II. Mobility Management in LTE Networks

An abstract model of the LTE network architecture is depicted at the top of Fig.1. The network includes the Evolved Packet Core (EPC) and Evolved UMTS Terrestrial Radio Access Network (E-UTRAN containing eNBs in the access side). Mobility Management Entity (MME), Serving GateWay (S-GW), Packet Data Network GateWay (PDN-GW) and Home Subscription Server (HSS) are among the main components of the EPC [1]. MME is a control plane entity that is responsible for, among other things, signaling, bearer establishment, gateway selection and tracking area management. The S-GW serves the User Equipments (UEs) via eNBs by forwarding and routing user data traffic. It also acts as an anchor point for handovers. The PDN-GW provides the interface to external packet data networks (e.g., Internet) and is also responsible for assigning IP addresses to UEs. The HSS keeps a database of users information such as identities and service profiles. The database is used for user authentication and authorization as well as for the mobility management. The E-UTRAN part of the architecture is only composed of eNBs. The traffic between the EPC and eNBs are transported through a backhaul network that provides the required capacity.

In order to access data services over the network, a user equipment has to register itself to the network so that a connection is established between the UE and EPC via an eNB. This process is called the Initial Attach (IA) procedure [1]. Upon successful completion of the IA, a virtual path is established between the UE and PDN-GW. This path is called the default bearer and once it is established, the IP connectivity is provided to the UE, i.e., an IP address is assigned to the UE by the PDN-GW. Each bearer is realized in the fixed part of the network in form of a GTP tunnel between eNB and S-GW. All data packets exchanged between a UE and the network are forwarded through the corresponding GTP tunnel.

When a UE moves from the coverage area of a source eNB to that of a target eNB, corresponding bearers of the UE need to be reconfigured. This reconfiguration must be carried out in a way that the IP connectivity of the UE is not affected, i.e., it should not require a change in the IP address assigned during IA procedure. This is achieved by tearing down the GTP tunnel between the source eNB and the S-GW and establishing a new tunnel between the S-GW and the target eNB. That is, the UE is not aware of the GTP tunnel and its reconfiguration and can keep its IP connectivity while it moves [1]. Although the use of the GTP tunnel helps in achieving a seamless mobility in the LTE network, it comes at a cost of excessive signaling load between E-UTRAN and the core network and the per packet tunneling overhead.

There are a few works on mobility management in LTE networks with the support of the backhaul network. One of the closest works to our approach in this area is the work presented in [6], where a layer-2-based backhaul network architecture is presented that can also support user mobility. The approach is similar to SDMA, in that they both eliminate the need for the GTP tunneling for the mobility management. Nevertheless, there are fundamental differences between the two works. In [6], TRILL (TRansparent Interconnection of Lots of Links) [8] technology is used together with Distributed Hash Tables (DHTs) in the access network, where as we use OpenFlow that provides more flexibility. Additionally in Trill-based approach, gratuitous ARP (address resolution protocol) packets has to be sent by eNB after a HO, whereas SDMA realizes a normal LTE procedure by just changing service VLAN identifiers (service VLAN id in carrier Ethernet). Also, broadcast messages flow around in access network in TRILL-based approach, whereas in SDMA they are controlled by OpenFlow.

III. Semi-Distributed Mobility Anchoring

In this section, we propose a novel mobility management scheme for LTE, which is based on using semi-distributed mobility anchors in the backhaul network. That is, in our scheme the backhaul network is partially aware of the bearers and the inter-eNB handovers. The basic idea of SDMA is as follows.

We assume that for a better management of mobility, the location addresses and identifiers of UEs are assigned separately, e.g., through the utilization of the location/identifier separation protocol (LISP) [9] in the network. Specifically, we assume that an IP address assigned to a UE during the IA procedure is merely regarded as the UE identifier, which remains fixed when the UE changes its point of attachment to the network. Additionally, the medium access control (MAC) address of the eNB, under which a UE exists, is assigned as the location address of that UE. The task of binding an identifier and a locator address to a UE is part of the mobile network. In this configuration, whenever a UE is handed over from a source eNB to a target eNB and the handover does not involve any change of the MME and S-GW, the main task of the handover management will be limited to changing the location address of the downstream packets of the UE from MAC address of the source eNB to that of the target eNB. In the SDMA approach, the SDN/OpenFlow is utilized in the network backhaul for an efficient realization of this process, i.e., dynamic identification of UE handovers and accordingly rewriting the location addresses in the downstream packets. While there are several possibilities for realizing the SDMA– e.g., in terms of the backhaul network technology and the signaling procedures–here we present a sample realization of the SDMA architecture which assumes the deployment of Carrier-Ethernet technology in the backhaul. Our proposed approach ensures an efficient exchange of mobility related signaling between the backhaul network and the EPC with minor modification to existing LTE mobility mechanisms[1].

---

[1]LTE specifications [1] are referred for initial attach, tracking area and handover procedures.

A. Network Architecture and Assumptions

We assume that the backhaul is a layer-2 network that is re-alized using the Carrier-Ethernet technology. The backhaul has two hierarchical areas, namely access and aggregation areas, where an aggregation area network provides the connectivity between the EPC and several access area networks. In the aggregation network, the IEEE802.1ah Ethernet technology [7] is used to provide scalability and quality of service (QoS) differentiation. Each access network connects a group of eNBs to the aggregation network, and is realized using the IEEE802.1ad Ethernet technology [7]. Each access domain network has a centralized controller, which controls the operation of all the Ethernet switches in that part of the network using the OpenFlow protocol (OFP) [5]. The controller has a knowledge of the topology and resources available in the controlling network and makes decisions about how packets of different traffic flows are forwarded and routed, and accordingly instructs corresponding switches. To hierarchically separate traffic flows in access networks we use service VLAN (S-VLAN) and customer network VLAN (C-VLAN) ids of 802.1ad. Specifically, S-VLAN ids are used to separate traffic of the users under different OpenFlow controllers (OFC), i.e., traffic associated with each access network has a unique S-VLAN id. Also, C-VLAN id is used to separate users traffic under each OFC. The assumption is that the assignment of VLAN ids are known to EPC. A sample realization of the network architecture is depicted in Fig. 1. Additionally, we make the following assumptions.

- MME and S-GW maintain a mapping between tunnel endpoint identifier (TEID) and VLAN pairs (C-VLAN, S-VLAN) to map the packets accordingly.
- A special S-VLAN called "Signaling S-VLAN" is reserved to signal the OFC whenever mobility related signaling is exchanged between UE and EPC and vice-versa.
- Each OpenFlow-controlled switch (OFS) has two pre-installed flow entries: one for forwarding the packets that are destined to EPC and another one for those packets that should be forwarded to OFC (the packets marked with the signaling S-VLAN).
- In order to track the location of UEs and also to facilitate the mobility management, each OFC is equipped with a mobility management module. The module keeps a table of all UEs, which are associated with the corresponding access network. Each entry of the mobility management table (MMT) has five elements: UE IP address, corresponding eNB MAC address, C-VLAN id, S-VLAN id and S-GW MAC address. UE IP address and C-VLAN id are assigned by EPC after Initial attach, and S-GW MAC address is the destination for the user plane packets from all UEs. Additionally, it is assumed that the mobility management module can process some of the LTE signaling messages (e.g., path switch) and react appropriately (e.g., generating end marker message).

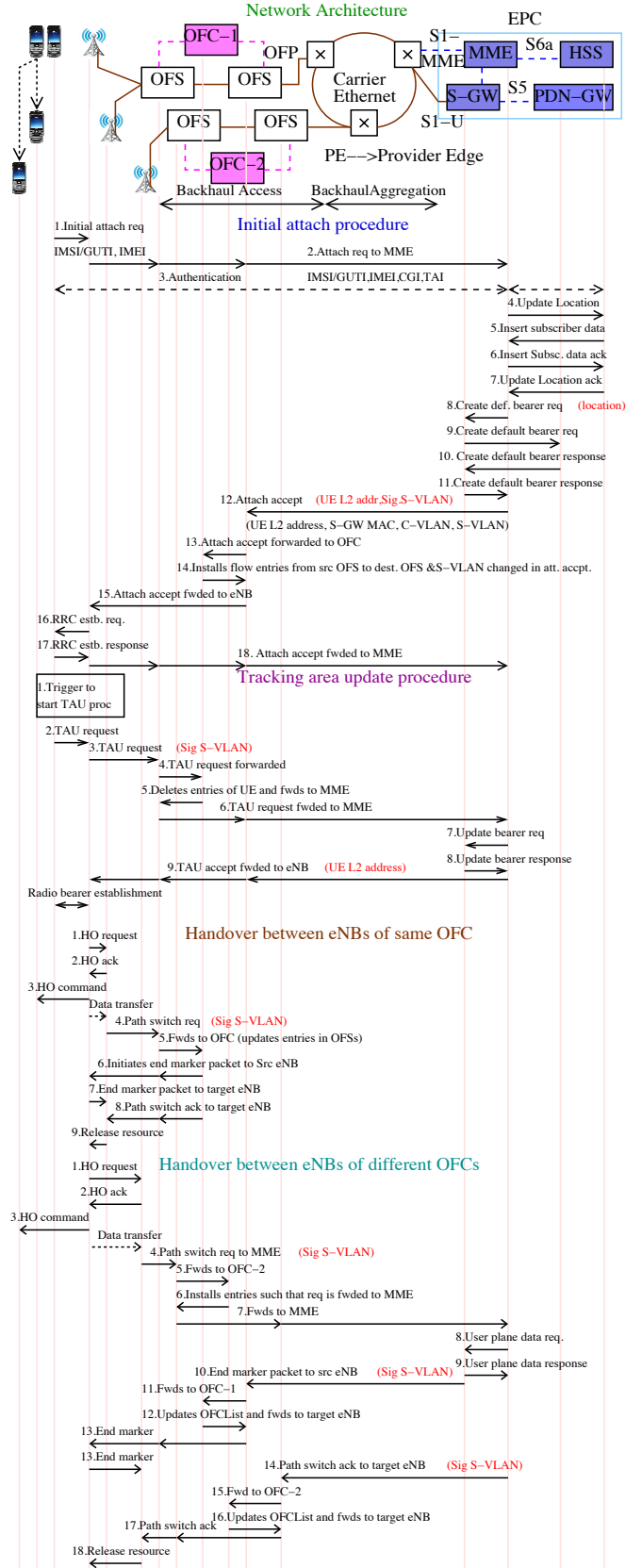| Mobility Management Table (MMT)				
UE IP	eNB MAC	C-VLAN	S-VLAN	S-GW MAC

Fig. 1. Initial Attach, Tracking Area Update and Handover procedures.

In the following sections, we describe how different mechanisms of the LTE network work. In our description of the handover management, we focus only on the X2-based inter-eNB handovers without change of MME and S-GW, and it is assumed that the other types of handovers–e.g. handovers with change of S-GW and MME, or handovers to a non-3GPP networks–are carried out following the standard procedures. Nevertheless, our scheme can be extended to the cases where there are more MME and gateways in the network and a mobility results in changing the gateway and MME.

B. Initial attach (IA) procedure

Once the UE is switched on, it initiates an IA request to an eNB, which forwards the request to the right MME (steps 1-2 in Fig. 1). In the next step, MME authenticates the UE and fetches subscriber data from the HSS (steps 3-7). Then, MME initiates a default bearer request to S-GW by forwarding the corresponding eNB MAC address (steps 8-11). S-GW forwards the request to PDN-GW, which assigns a unique IP address and C-VLAN id to the UE. Now S-GW forms UE address by binding the UE IP address and eNB MAC address and forwards the address to the MME. Then, MME initiates an IA accept to the eNB and changes S-VLAN of the IA accept message to the signaling S-VLAN, because this is to be interpreted by the corresponding OFC. Accordingly, on reaching the first OFS in its path, the packet is forwarded to the controlling OFC. The OFC updates its MMT and installs flow entries from source OFS to target OFS. Also, OFC installs an extra entry in the OFS, from which the IA accept message is sent to OFC, such that signaling S-VLAN of IA accept is changed to original S-VLAN (steps 12-15). Finally, IA accept and acknowledgement (ack) messages are embedded in radio resource control (RRC) bearer setup messages and IA accept ack is forwarded to MME (steps 16-18).

Note that in this scheme the transport layer address in initial context setup is S-GW MAC instead of its IP address, and GTP identifiers are not used here anymore. Additionally, two new fields are used in the initial context setup, i.e., UE L2 address and C-VLAN id.

C. Tracking Area Update (TAU) procedure

A TAU is initiated by an idle UE when it detects that it has entered a new TA or when a periodic timer expires. In our scheme, only the TA updates due to detecting a new TA are sent to OFC, and other TA updates are directly sent to the MME so that OFC is not overloaded. The modification to S1-AP protocol messages are similar to those in the IA procedure.

As shown in Fig. 1, when a TAU is initiated by an UE (steps 1-2), the corresponding eNB forwards the request to the MME by changing the S-VLAN id of the message to signaling S-VLAN and therefore the update is first sent to OFC. OFC observes it as TAU and removes the entries of UE in the edge OFS. Also S-VLAN is changed to the original one (steps 3-5). Then, TAU is forwarded to the MME, which initiates an update bearer request to the S-GW. Finally, the MME initiates a TAU accept message, which is sent together with the UE L2 address to the eNB (steps 6-9).

D. Handover (HO) procedures

There are two kinds of handovers, namely, handovers where both source and target eNBs are under the control of the same OFC, and handovers where the target eNB is controlled by a different OFC than that of source eNB. Below, we consider these two cases separately.

1) HO under the same OFC: Once a decision is made for handover, the source eNB initiates and sends a *HO-request* message including the UE L2 address and C-VLAN id to the target eNB (over the X2 interface). After receiving handover acknowledgement at the source eNB, a handover command is issued to the UE (steps 1-3). Then the target eNB initiates path switch request to MME using the signaling S-VLAN id since it is mobility related signaling (step 4-5). The OFC mobility management module receives this *path-switch* message and observes that both eNBs are under its control. Therefore, an *end-marker* message is generated and forwarded to the source eNB. Also, the OFC accordingly updates the flow tables in the corresponding OFSes and then generates and sends a *path-switch-ack* message to the target eNB (steps 6-8). To complete the mobility management task, the OFC further needs to install a flow entry in the access edge OFS (the OFC that connects the access network to the aggregation network) to redirect the traffic of the mobile UE to the target eNB. The matching filter of this flow entry includes UE IP address, source eNB MAC address and the C-VLAN id, and the action of the entry consists in changing the destination MAC address of incoming frames to the target eNB MAC address and forwarding it.

It can be seen that this kind of HO is completely delegated to the mobility management modules in OFCs and no signaling message is sent to the EPC. In fact, the access edge OFS takes the role of mobility anchoring in the network.

2) HO under different OFCs: The initial steps (1-3) are the same as the case with the same OFC. However, here the source OFC observes that the target eNB is not under its control, so it forwards the *path-switch* message to the MME. Then, the MME informs the S-GW to update its user plane and initiates and forwards an *end-marker* message to the source eNB and a *path-switch-ack* message to the target eNB using the signaling S-VLAN id. The *end-marker* message is processed by the source OFC, which then removes the entries of the UE from its MMT and the corresponding OFSes. Also, the *path-switch-ack* message is processed by the target OFC, which installs new entries in the path from a source OFS to a destination OFS as in IA procedure.

Note that, in this scheme, GTP identifiers are not used in HO request message and two new fields are used in *HO-request* message (UE L2 address & C-VLAN). Also, in step 4 of the both HO cases, target eNB needs to set the S-VLAN id of the initiated *path-switch* message to the signaling S-VLAN id.

IV. ANALYTICAL MODELING OF LOAD ON EPC

In this section we present a mathematical model for calculating the signaling load to the EPC under both standard mobility management as well as our proposed SDMA mobility management scheme. For this purpose we assume that the shape of cell sites is approximated with a circle, and that the speed and direction of each UE at every instant is modeled as

Gauss-Markov [10], since it imitates the real human mobility. Also, without loss of generality we assume that the time is slotted. The procedure to calculate the signaling load is as follows.

Let N be the number of UEs in the network and k be the average number of time slots over which the UE stays in the same cell site.

Average load (L_{LTE} in packets/sec) on EPC for the standard LTE can be calculated as the sum of packets sent to EPC due to Initial Attach (L_{IA}) and Handover (L_{HO}) events per second. Therefore, $L_{LTE} = L_{IA} + L_{HO}$.

The time interval between IA and detach events and vice-versa for each UE is modeled as an exponential random variable with parameter λ. An update is sent to EPC for every IA and detach events. Therefore, average number of packets per second sent to EPC for each UE due to IA events is obtained from alternating renewal process as λ. In an average for every k time slots, a HO update is sent to EPC for each UE. Since UE exists either in attach or detach states, average number of packets per second sent to EPC for each UE due to HO events is $1/2k$. So, average load on EPC for N users is given by,

$$L_{LTE} = L_{IA} + L_{HO} = \left(\lambda + \frac{1}{2k}\right) N \qquad (1)$$

Now let us calculate the average duration of UE in a cell site as characterized by k. From Gauss-Markov mobility, speed and direction of each UE at every instant n is modeled as [11]

$$S_n = \alpha S_{n-1} + (1-\alpha)\mu + \sqrt{1-\alpha^2} S_{x_{n-1}} \qquad (2)$$

$$\Theta_n = \alpha \Theta_{n-1} + (1-\alpha)\theta + \sqrt{1-\alpha^2} \Theta_{x_{n-1}} \qquad (3)$$

where α is degree of memory ($0 \le \alpha \le 1$) indicating the randomness ($\alpha = 0$ gives brownian motion and $\alpha = 1$ gives linear motion), S_n and Θ_n are speed and direction of UE at instant n; μ and θ are mean speed and mean direction; $S_{x_{n-1}}$ and $\Theta_{x_{n-1}}$ are independent and identically distributed (i.i.d.) gaussian random variables with means (μ and θ respectively) and standard deviations (σ_μ and σ_θ respectively).

The random variable S_n in terms of S_0 (initial speed distribution) is given by [10]:

$$S_n = \alpha^n S_0 + (1-\alpha^n)\mu + \sqrt{1-\alpha^2} \sum_{i=0}^{n-1} S_{x_{n-1}} \alpha^{n-i-1} \qquad (4)$$

In equation 4, S_0 is a uniform random variable [U($\mu - \sigma_{\mu_0}, \mu + \sigma_{\mu_0}$)] and is approximated by gaussian with mean μ and variance $\frac{\sigma_{\mu_0}^2}{3}$ [N($\mu, \frac{\sigma_{\mu_0}^2}{3}$)]. Similarly Θ_0 is G($\theta, \frac{\sigma_{\theta_0}^2}{3}$) Therefore,

$$S_n \sim N\left[\mu + \sqrt{\frac{1-\alpha}{1+\alpha}}(1-\alpha^n)\mu, \frac{\alpha^{2n}\sigma_{\mu_0}^2}{3} + (1-\alpha^{2n})\sigma_\mu^2\right] \qquad (5)$$

$$\Theta_n \sim G\left[\theta + \sqrt{\frac{1-\alpha}{1+\alpha}}(1-\alpha^n)\theta, \frac{\alpha^{2n}\sigma_{\theta_0}^2}{3} + (1-\alpha^{2n})\sigma_\theta^2\right] \qquad (6)$$

Given that a UE is inside a cell site (the shape of cell site is approximated as circle with radius r) at this time slot, n, the probability that it stays in the same cell site in the next time slot, $n+1$ is denoted by p. Therefore, the number of

time slots it stays in the same cell site can be modeled as a Geometric random variable with parameter $1 - p$.

$$p = \frac{4}{\Pi r^2} \int_0^r \int_0^{\sqrt{r^2 - x^2}} \int_{-\Pi}^{\Pi} P(s_n \le d) f_{\Theta_n}(\phi) d\phi dy dx \qquad (7)$$

where d is the distance from any point in the circle say (x, y) to a point on the circle that is making an angle ϕ with the x-axis; and given by

$$d = x\cos(\phi) + y\sin(\phi) + \sqrt{r^2 - (x\sin(\phi) - y\cos(\phi))^2} \quad (8)$$

Accordingly, k can be calculated iteratively as follows:
1) Assume n=1.
2) Calculate p from equation 7 by substituting n.
3) Calculate n_{temp} using $n_{temp} = 1/(1-p)$.
4) Increment n value and goto step-2 until $n_{temp} \approx n$.
5) Set $k \leftarrow n_{temp}$

The obtained k value is substituted in 1 to get average signaling load on EPC for LTE.

Similarly, the average load to EPC under the SDMA mobility scheme can be calculated as,

$$L_{SDMA} = \left(\lambda + \frac{f}{2k}\right) N \qquad (9)$$

where f indicates the conditional probability that a UE takes HO between eNBs of different OFCs; given that a HO took place. f is a function of number of OFCs, number of OFSes under each OFC and the way in which eNBs are assigned to each OFS. f is calculated as the ratio between number of possible HOs between eNBs of different OFCs and total possible HOs in the network.

Also, note that for the case where there is an extra controller for managing the backhaul aggregation section under SDMA, there will be no signaling load to the EPC due to the HO in the considered scenario. Accordingly, in this case the average load on EPC is limited to the IA, which is given by

$$L_{SDMA} = \lambda N. \qquad (10)$$

V. NUMERICAL ANALYSIS

In this section, we numerically analyze the performance of the SDMA and compare it with that of the mobility management scheme based on the LTE standard specifications. The objective is to get the first insight into the achievable gains of implementing SDMA as well as incurred costs. As discussed in the last sections, we expect the SDMA to reduce both the handover latency and the signaling load to the EPC. In addition, since we do not use the GTP tunneling, the encapsulation overhead per data packet is obviously reduced in SDMA as compared to standard LTE. For the current analysis we only focus on the reduction in the handover latency and signaling load to EPC. The HO latency is defined as the difference between the time at which a source eNB initiates the handover process and the time at which the same eNB receives the handover acknowledgement message.

On the other hand, since a part of mobility management process is delegated to the elements of the backhaul network, it is important to evaluate the additional (processing and

signaling) load incurred to these elements. For this purpose, we consider two metrics, namely, the signaling load to OFCs and the number of flow entries in OFSes in the backhaul network. These metrics can further help in understanding the scalability of SDMA. Additionally, we consider initial attach latency as another cost metric, since we expect it to increase due to the additional stage of processing in the backhaul network. The IA latency is defined as the difference between the time at which UE entered into the network and the time at which UE is assigned its unique identifier. To evaluate the performance metrics, we have implemented an abstract model of the network in software using C++.

A. Parameters and Scenarios

The following assumptions are made in developing the simulation model. The network has 120 hexagonal cells (eNBs) each of side 0.4 km, which are arranged in a spiral fashion (The unique numbering to eNB and finding the unique id for a given co-ordinates are done according to [12]). The access area networks have 12 OFSes in total, where each OFS serves 10 eNBs. The nodes in each access network are connected according to a ring topology. The number of OFCs controlling the access networks is a varying parameter in the set $\{2, 3, 4, 6\}$, i.e., the size of the access networks is varying from 6 OFSes to 2 OFSes. The number of users in the network (i.e., UEs) is varying from 1000 to 10,000. Also, for simulating the user mobility the Gauss-Markov model [11] has been used, and the mean and standard deviation of speed of UEs is set to 5 m/s and 4 m/s, respectively. We have evaluated and compared the performance metrics for four different scenarios. In first scenario (denoted below by LTE) we consider the standard LTE without using OpenFlow switches. Second scenario (denoted by SDMA-I) considers the case, where we have OFCs and OFSes only in the access area network and the aggregation area network is realized using standard 802.1ah. In the third scenario (denoted by SDMA-II), the aggregation network is also OpenFlow-controlled and has a dedicated OFC. Finally, in the last scenario (denoted by SDMA-III) the whole access and aggregation networks are controlled by a single OFC. Average processing times (PT) of different networking elements and average propagation times (GT) in different parts of the network are given below.

	average time (msec)
PT of UE	0.8
PT of eNB	4
PT of access OFS	0.02
PT of access non-OFS	0.03
PT of aggregation network	8
PT of EPC for an IA request	30
PT of EPC for handover	20
GT between OFS and OFC in access	0.01
GT between OFS and OFC in aggregation	0.6

B. Numerical Results and Discussion

First, let us consider the IA and HO latencies, which are numerically calculated. The results for all four scenarios are shown in Table I for two different values of OFC processing time. For the each of the scenarios SDMA-I and SDMA-II,

two values are presented for handover. These are for the cases, where handover is under the same OFC (denoted in the table as HOS) and under different OFCs (denoted as HOD).

From the results, we observe that SDMA can lead to a great improvement in the HO latency. This reduction in the HO latency can be up to around 86%, which is achieved for HOs under the same OFC and with OFC processing time set to 5 ms. We also see a similar reduction in the latency of HOs with a single OFC in network (SDMA-III). The least amount of reduction is achieved when the HO is performed under different access OFCs. In fact, for SDMA-I this latency can even be slightly larger than that in standard LTE, because in this case there is one extra OFC processing. Nevertheless, even in this case (SDMA-I), the average net effect of using SDMA is positive, because the average number of inter-OFC HOs is less that that of intra-OFC handovers. Another important point here is the impact of OFC processing time on the HO latency. In fact, the HO latency increases almost linearly with the OFC PT. In Table I, we also have the IA latency for different scenarios. As expected, the IA latency slightly increases under SDMA in comparison with the standard LTE. The worst case occurs for SDMA-II and with OFC PT set to 10 ms. In this case, the IA latency increases by around 35%. Nevertheless, the absolute value of IA latency for all cases is within the acceptable standard range.

TABLE I
INITIAL ATTACH AND HANDOVER LATENCIES (MSEC)

OFC	LTE		SDMA − I			SDMA − II			SDMA − III	
PT	IA	HO	IA	HOS	HOD	IA	HOS	HOD	IA	HO
5	56.1	36.1	61.1	5.1	41.2	66.1	5.1	16.2	62.1	6.8
10	56.1	36.1	66.1	10.1	46.2	76.1	10.1	31.2	67.1	11.8

Now let us consider other performance metrics as defined earlier in this Section. Fig. 2-(a) depicts the average load (in packets per second) on different elements of the network as a function of number of users. The results are shown only for the representative case of having two access networks (i.e., two OFCs), and the other results are omitted due to space limit. The results show the linear increase of the load on all elements of the network. Also, we observe a considerable reduction in the load to EPC when SDMA is used in comparison with standard LTE. The gain is particularly larger at the higher number of users, where most of the HOs are managed by the OFCs in the backhaul network. As a result, the amount of gain is also larger for the case where both access and aggregation area networks are equipped with OpenFlow-based mobility management entities. The figures also demonstrate that, as expected, the reduction in the signaling load to EPC is reflected in the increase of OFC loads. The results for the cases with different number of access networks, though not shown here, demonstrate similar results and conclusion.

In Section IV, we developed analytical models for the average load on EPC both for standard mobility and the SDMA. For cross-validation of our analytical and numerical models, here we consider a comparison of the results achieved under both models. Fig. 2-(b) depicts the comparison
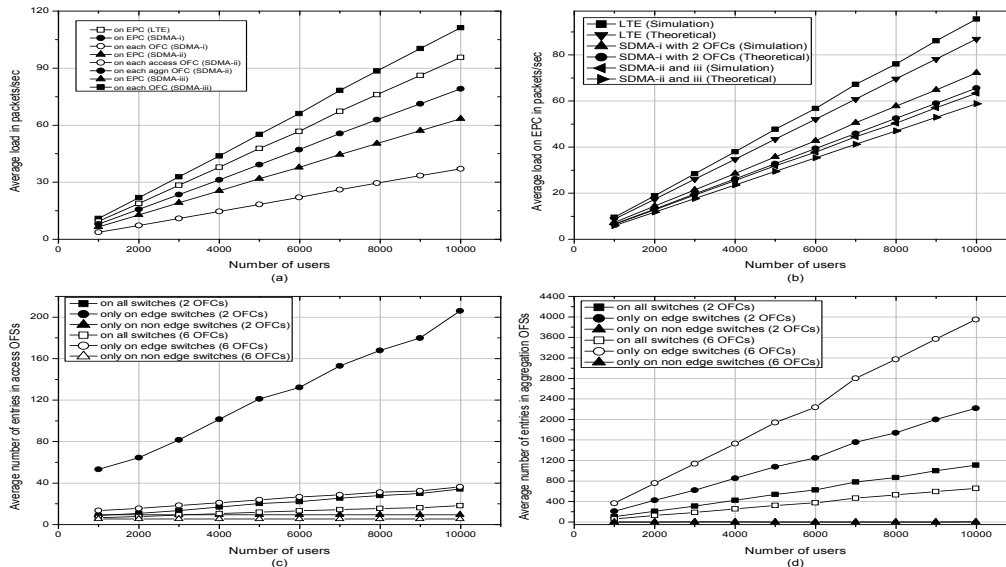
Fig. 2. Performance metrics vs. number of users: (a) Average load on the network's elements, (b) Comparison between analytical and simulation models for the average load on EPC, (c–d) Average number of flow entries in OFSes.

between the analytical model (denoted by Theoretical) and the numerical results (denoted by Simulation) for both standard LTE mobility scheme and the SDMA. We observe that the analytical models closely approximate the load on EPC for all the cases. Specifically, the average difference between the analytical and numerical model is around 8.7%, which is due to overestimating the shape of hexagons as circles with the radius of 400 meters.

Now, let us consider the scalability of our approach by looking at the number of flow entries in the switches. The average number of flow entries in OFSes in access and aggregation area networks for the case with 2 and 6 access networks are presented in Figs. 2-(c and d). It is observed that the average number of flow entries in access OFSes in all cases is tractable. The results also indicate that the edge OFSes (the switches connecting an access network to the aggregation network) require a much larger space for flow entries. This is because, the edge switches are also responsible for changing the eNB MAC addresses of flows belonging to all UEs in that access that have handed over to a new eNB. For the non-edge switches the required space for flow entries due to mobility management is below 15 in all the evaluated cases. As for the number of flow entries in aggregation OFSes we observe similar relation between the edge and non-edge switches. Nevertheless, here the total number of flow entries is almost an order of magnitude larger than that in access OFSes. In fact, as the number of controllers increases (the number of access networks), load on access OFCs decreases and load on aggregation OFCs increases since the number of intra-OFCs HO decreases and the HOs need to be managed by the OFC in the aggregation network. In the worst case, an aggregation edge OFS would need space for 4,000 flow entries to manage the HOs in a network with 6 access networks and 10,000 active UEs. In summary, the results demonstrate the scalability of the SDMA approach.

VI. CONCLUSION

We presented a novel mobility management scheme called SDMA for LTE networks, where SDN/OpenFlow is used to distribute the mobility anchor points in the backhaul network. We demonstrated though the numerical analysis that the SDMA is a scalable approach that can reduce the HO-related signaling load to the EPC by up to 33%, and the HO latency by up to 86%. Additionally, SDMA eliminates the need for GTP tunneling, which in turn improves the efficiency of data transport. All these improvements come at the cost of a marginal increase in the initial IA latency and an increase in the processing and signaling load to the switches and controllers in the backhaul network. Nevertheless, through the examples we showed that the amount of increased load on the network elements in the backhaul can be easily handled by existing commercial off the shelf (COTS) equipments.

REFERENCES

[1] 3GPP Technical Specifications, "3GPP TS 23.401 V8.3.0," 2009. http://www.3gpp.org/
[2] Cisco White Paper, "The future of mobile networks," Cisco, 2013. http://www.cisco.com/web/about/ac79/docs/sp/Future-of-Mobile-Networks.pdf/
[3] Open Networking Foundation White Paper, "Software-defined networking: The new norm for networks," ONF, 2012.
[4] N. McKeown *et.al.*, "OpenFlow: Enabling innovation in campus networks," in *ACM SIGCOMM*, 2008.
[5] OpenFlow specifications v1.1.0, 2011. http://openflow.org/
[6] N. Varis, J. Manner and J. Heinonen, "A layer-2 approach for mobility and transport in the mobile backhaul," in *ITS Telecommunications*, 2011.
[7] Metro Ethernet Forum. http://metroethernetforum.org
[8] J. Touch, and R. Perlman, "Transparent interconnection of lots of links (TRILL): Problem and applicability statement," RFC 5556, IETF, 2009.
[9] Locator Identifer Split Protocol. http://www.lisp4.net
[10] B. Liang and Z.J. Haas, "Predictive distance-based mobility management for PCS networks," in *INFOCOMM*,1999.
[11] T. Camp, J. Boleng and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing (WCMC)*, 2002.
[12] H. Cao, G. Wang and Y. Xie, "A two-dimensional logical co-ordinate system for hexagonal grids in manets," in *International conference on Communications and Mobile Computing*, 2009.