

Taming Mobility Management Functions in 5G: Handover Functionality as a Service (FaaS)

Albert Codes Morales[†], Adnan Aijaz[‡], Toktam Mahmoodi[‡]

[†] Polytechnic University of Catalonia, Spain

[‡] King's College London, UK

Contact author: toktam.mahmoodi@kcl.ac.uk

Abstract—With the shift towards decoupling of data and control plane and the softwarization in mobile networks, delivery of mobile network control functionalities can be foreseen as over the top services. We envisage such move within the paradigm of Functionality as a Service (FaaS); a platform that can also deliver requirements of the 5th generation mobile, i.e. 5G. In this paper, we briefly explain architectural elements of mobile networks that can materialize the FaaS platform, and discuss the mobility management within the FaaS platform. Relying on these architectural elements, we propose modelling the dynamics of mobile network topology, including fixed base stations and mobile users, as a self-regulatory network. Therefore, users' mobility is provisioned and traffic flow is moved to the next point of attachment, i.e. the next base station, with no handover request from the mobile device. Clearly, there is a trade-off between keeping the abstract model up-to-date and the increased overhead of the updates. Hence, we further examine the effect of such out-of-date and insufficient information in the abstraction. The investigations are carried out in network simulator (ns-3) and it can be seen while reducing frequency of updates results in the increased packet loss, an optimal balance between overhead and packet loss can be achieved.

Index Terms—5G mobile; mobility management; Software-defined Networking; Functionality as a Service (FaaS)

I. INTRODUCTION

Mobility management at the IP level has evolved significantly over the last decade and with the current accelerated pace towards “flatten” core networks and heterogeneity on the access, the quest for supporting mobility becomes more than ever important. With the advances in decoupling data and control plane, and softwarization of control, as well as the emerging distributed mobility management, there is a great opportunity for mobility management to be offered as a service to the mobile users.

Mobile data traffic also continues to increase. According to Cisco annual statistics, global mobile data grew 69 percent in 2014 and traffic reached 2.5 exabytes per month at the end of 2014, up from 1.5 exabytes per month at the end of 2013. On the other hand, signaling in mobile network is accountable for a larger part of growth in traffic¹. The tsunami of data traffic over mobile networks can be tackled through various *mobile data offloading* techniques [1]. However, such solutions are not applicable for signaling traffic. With the

development of small cells, mobility-related signaling is particularly increasing and therefore, a fresh approach to mobility management is the need of the hour. We have previously studied the effect of integrating Openflow switches in the core network of 4G/LTE (long term evolution) on reducing the handover signaling by half [2]. Bringing Software-defined networking (SDN) in the LTE core network has been subject of various studies including [2] and [3].

In this paper, we take a more radical approach and use network abstraction to deliver mobility management as an over the top (OTT) service in mobile networks. We model the dynamics of mobile network topology, i.e. users' mobility and changes in their point of attachments, as a self-regulatory network. Therefore, traffic flow is moved to the next point of attachment, i.e. the next base station, with no handover request from the mobile device. We simulate this setting in the ns-3 LTE model [4], while interface between the network abstract and the mobile core gateway is OpenFlow. Therefore, the decision to move certain traffic flow from one base station to another is translated to an update in the flow table entry of the gateway at the LTE core network. Clearly, there is a trade-off between keeping the abstract model up-to-date and the increased overhead of the updates. Hence, we further examine the effect of such out-of-date and insufficient information in the network abstraction. The out-of-date information could be compensated by estimating the mobility patterns and predicting the next state of the network graph.

The remainder of this paper is organised as follows. In Section II, an overview of mobility management in the LTE network is presented. Section III details abstraction of mobile network topology and its evolution. Simulation model and results are presented in Section V. Finally, some highlights and the road ahead in further development of this research are discussed in Section VI.

II. MOBILITY MANAGEMENT IN 4G-LTE

In this section, we briefly discuss the key mobility management function of handover in state-of-the-art 4G LTE systems. Mobility management in LTE is handled by the Mobility Management Entity (MME), which is located in the Evolved Packet Core (EPC) network. It is responsible for initiating paging, managing tracking area updates, and authentication of the UEs. MME is also involved in bearer activation and its deactivation procedures. MME connects to the eNodeB

¹“Signaling is growing 50% faster than data traffic”, NSN White Paper, 2012.

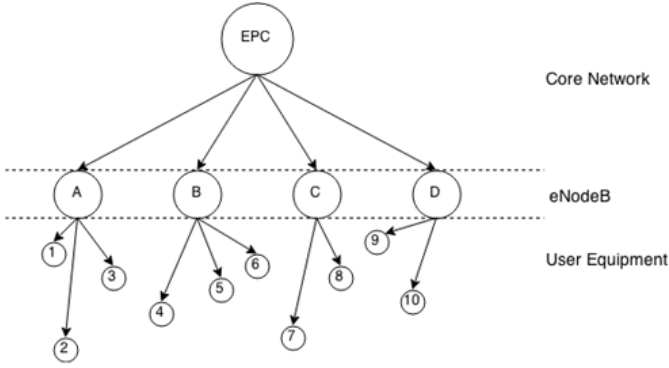


Fig. 1. Mobile Network Topology as a Weighted Graph

through the S1-MME interface. One of the most complex MME functions is seamless interworking in inter-eNodeB, inter-system (2G/3G), and roaming scenarios. In this paper our focus is strictly on the handover function supported by the MME.

In LTE systems, the handover process is UE-assisted and network controlled such that the UE periodically reports the Received Signal Strength Indicator (RSSI) from the connected base station. The handover is initiated by the UE when the signal strength falls below a certain threshold. The key steps involved in this process are discussed as follows ².

- Based on the reported signal strength from the UE, the source eNodeB sends a handover request over the x2 interface to the target eNodeB. The target eNodeB sends a handover acknowledgement message to the source eNodeB with some RRC (radio resource control) reconfiguration parameters for the UE.
- The UE is notified by the source eNodeB and detaches from the source cell. After that the UE joins the target cell through a contention-free random access procedure.
- The UE indicates the completion of handover for itself to the target eNodeB through a RRC reconfiguration complete message.
- The target eNodeB informs the MME of the new location of the UE. After that the MME sends a message to the serving gateway (S-GW) to update the path of user plane data for the UE.
- the S-GW confirms the update of path for the UE to the MME, after which the target eNodeB informs the source eNodeB of successful handover.
- During the handover process, the source eNodeB buffers all the packets destined to the UE and transfers to the target eNodeB.

III. HANDOVER FAAS IN 5G: DYNAMICS OF MOBILE NETWORK TOPOLOGY

In our proposed mobility management framework, we model the topology of the mobile network as a graph assuming each UE is connected to only one eNodeB at any time instant

which implies that an edge between the UE and the eNodeB exists in the graph. Owing to the dynamic nature of mobile networks, this graphical representation is evolving with time and therefore, it is obtained by iteratively applying some given operator on the initial graph.

Let, $X = \{E, V\}$ and Γ denote the initial graph and the operator, respectively. The dynamic nature of graph can be expressed as (X, Γ) , while $\Gamma : X \rightarrow X$, recursively. A given graph X will remain invariant under some given operator Γ , if X and $\Gamma(X)$ are isomorphic. Further, each edge of the graph is assigned a weight $\omega : E \rightarrow \mathbb{R}$. Let, $\mathcal{L} \triangleq \{1, \dots, l, \dots, L\}$ and $\mathcal{K} \triangleq \{1, \dots, k, \dots, K\}$ denote the set of eNodeBs and users, respectively. Each user will connect to the strongest eNodeB instantaneously i.e., the eNodeB providing highest received power in the downlink. The received power for the k^{th} user connected to the l^{th} eNode B is given by

$$\mathcal{P}_k^l = P_l h_k^l |x_k^l|^{-\alpha}, \quad (1)$$

where P_l is the transmit power of the l^{th} eNodeB, h_k^l denotes the channel fading coefficient between the k^{th} user and the l^{th} eNodeB, x_k^l represents the distance between the k^{th} user and the l^{th} eNode B, and α is the path loss exponent.

The eNodeBs are connected to the Serving Gateway (S-GW) which is further connected to an SDN-controller. The S-GW captures all packets from the eNodeBs and forwards them to the remote hosts through the Packet Gateway (P-GW). The forwarding rules followed by the S-GW are dictated by the SDN-controller. Details of this architecture design can be seen in Figure 2 and functionality of the SDN-controller is further explained in Section V-B.

The SDN-controller also stores the abstract view of the network, i.e., it is aware of the current network topology and its future states. Figure 1 shows an example of the network topology abstract view at the SDN-controller. The larger nodes represent the parent nodes and have fixed locations while the smaller nodes represent the UEs which are mobile. The arrows indicate the existence of a wireless link between eNodeBs and the UEs. In our proposed framework, the weights to individual links are assigned according to distances of the UEs from the connected eNodeBs. Thus, the SDN-controller's decision on reconstructing the topology graph depends on the weight of the eNodeB-UE links'.

Let, r_{max}^l denote the maximum coverage radius of the l^{th} eNodeB. Similarly, let r_{min}^l denote the minimum possible distance from the l^{th} eNodeB at which a UE will be connected. We defined two different thresholds on the distance based on which SDN-controller executes a new action such that

$$\begin{cases} r_{min}^l < x_k^l < T_1 & \rightarrow A_0 \\ T_1 < x_k^l < T_2 & \rightarrow A_1 \\ T_2 < x_k^l < r_{max}^l & \rightarrow A_2 \end{cases} \quad (2)$$

In (2), A_0 denotes no action since the UE is in close proximity to the eNodeB and therefore, no information is requested by the SDN-controller. In case of A_1 , the SDN-controller obtains signal quality information from the eNodeB

²3Gpp release 12, available on <http://www.3gpp.org/>

and might decide to reconstruct the topology i.e., trigger an handover. Finally, in case of A_2 , the SDN-controller triggers an handover automatically as the UE might move out of the coverage of the eNodeB in a short while. After executing the handover the graph is updated by inserting the link between the UE and new eNodeB.

IV. SDN-CONTROLLER DESIGN

The network architecture foreseen in this paper is based on the previously discussed design in [2]. The SDN-controller can implement multiple applications such as mobility management, pseudowire creation and assignment, allocation of radio resources and many others, while also interworking with other programmable elements in the protocol stack. In a nutshell, the design of control plane in SDN-controller comprises of: (1) Separating the control plane functionalities in all network entities starting from eNodeB up to the PGW; (2) Implementing the control plane functions as SDN-controller applications; (3) Encapsulating the management plane functions within individual virtual machines (VM) that interact with the SDN-controller. The SDN controller can potentially have control over the number, placement and workload-assignment of instances of each function instantiated as VMs.

As mentioned earlier, for the Handover FaaS platform considered in this paper, the SDN-controller stores the abstract view of the network, and is aware of the current network topology and its future states. Based on this topology then, pushes the packet forwarding rules to the mobile gateway.

The SDN-controller implementation can be envisioned based on OpenFlow [5], with two classes of modules (viz., platform essentials, value-added services) as shown in Figure 3. The platform essentials are involved in state dissemination, L3 routing, tunnel setup, and interfacing with management plane. The value-added services handle intelligence for automatically managing resources, mobility, DiffServ/QoS and network selection for UE. Hence, the SDN-controller will push rules to all devices (including PGW) in a proactive manner to handle the large scale of subscribers.

V. PERFORMANCE STUDY

A. Simulation Model

We use ns-3 for our performance study ³since a complete model of LTE network is available in this network simulator [4]. The LTE architecture model in ns3 is depicted in Figure 4. We added the SDN-controller on top of this model and an openflow interface to the SGW/PGW. The dynamic topology of mobile nodes is modelled and abstracted at the SDN-controller as detailed in Section III. Our simulated area is a squared area with four eNB placed in a way that there is no “black hole”, where no eNodeB has coverage. Ten UEs have been deployed, with a constant speed mobility model.

In this simulation, we show the effect of updating frequency of mobility information on the network performance. Apart from mobility management, the controller also has to correctly

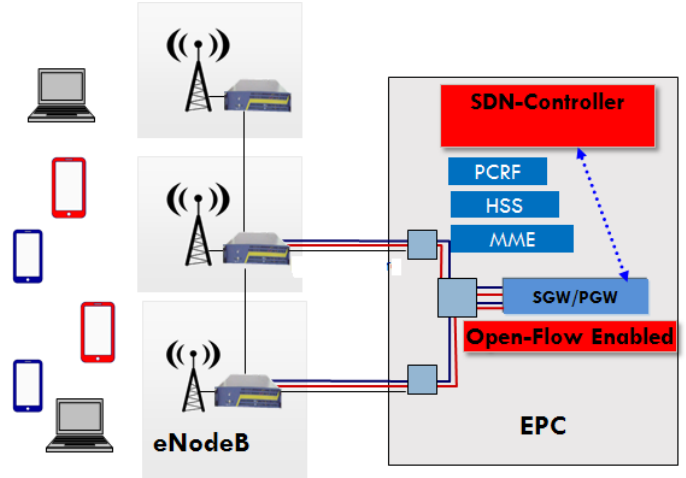


Fig. 2. Network Architecture.

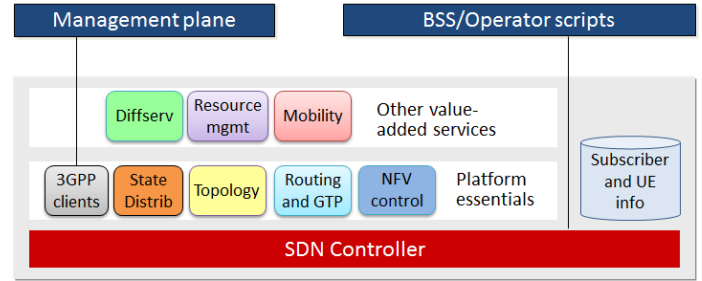


Fig. 3. Design of the SDN controller for the LTE network [2].

update all flow routes in the core of the cellular network. If a handover has been triggered, the controller has to alert the MME that it has to reroute all traffic to the convenient final eNodeB. Depending on the frequency of updates, some packets will be lost in the lapse between the handover is made and the switch has updated its flow table. Therefore, the performance indicator that is measured here is the probability of packet loss.

Since both initial placement and direction are decided using random parameters, we used Monte Carlo method and replicate the same environment for different values of these variables. Here, we ran twenty independent simulations. After twenty independent simulations, an average of all users packet loss ratio for different update times of the flow table in the EPC of the network is calculated.

B. Simulation Results

Figure 5 demonstrate packet loss on the right axis, and the communication overhead between the SDN-Controller and radio access on the left axis, versus time between two consecutive updates in the network topology. It is clear that by increasing time between each two updates, packet loss increases while communication overhead would be decreased. However, from this figure it can be seen that increase in the packet loss ratio follow a non-linear pattern, and there is an area where increasing the update time has a more detrimental

³Version 3.21 of the simulator has been used to run the simulations.

effect on the QoS parameter than in others. On the other hand, the communication overhead has close to exponential decrease by increasing the time between two updates. An interesting optimal point can be extracted for the updating frequency, that is where the two plots intersect, unless the corresponding packet loss does not meet the QoS requirements.

Another consideration for deciding on the frequency of update could be the applications type and its corresponding QoS requirements, i.e. mobility and location of a user with stronger QoS in terms of packet loss can be reported more frequently to the network. If the application can hold a reasonable packet loss ratio, like video on demand, then a larger update time can be used without affecting the overall performance of this applications and at the same time having very little overhead. Instead, for applications that are interactive, like gaming, packet loss ratio is a very crucial QoS parameter, and a very low update time would be required.

Observing from Figure 5, the update time should be between ten and thirty seconds, because the gain obtained from reducing the update rate lower than ten seconds is not significant while the packet error rate increase dramatically for larger than thirty seconds update time with no equivalent decrease in the communication overhead. The exact value of update time can be decided depending on the users' application.

VI. REMARKS

In this paper, we explain architectural elements of mobile networks that can materialize the Functionality as a Service (FaaS) platform, as a platform that delivers decoupled and softwarezied control plane. We examine Handover FaaS by modelling the mobile network topology as an evolving graph. While the volume of signalling traffic is dramatically increasing in mobile network, executing handover as an OTT service with reduced signalling, can affect the performance of mobile networks significantly. Through some preliminary investigations, we observe that frequency of updating the mobile network topology at the FaaS provider (or the so-called SDN-controller in the architecture design of this paper) can be designed such that users' experience can remain unchanged. Clearly, an out-of-date view of the network topology at the SDN-controller results in packets being forwarded to an eNodeB that is no longer in the range for the UE, hence resulting packet loss. But, on one hand prediction of UE's mobility pattern, for updating the topology, and on the other hand deciding the frequency of UE's location update at the controller depending on the tolerance of the running application to packet loss, can leave the users' experience unchanged.

VII. FUTURE WORKS

The future work will focus on developing a complete analytical model of the proposed mobility management framework wherein the eNodeBs and UEs would be distributed according to a Poisson Point Process (PPP). Further, the impact of having small cells in the form of another network tier would be investigated. In addition, the impact of different

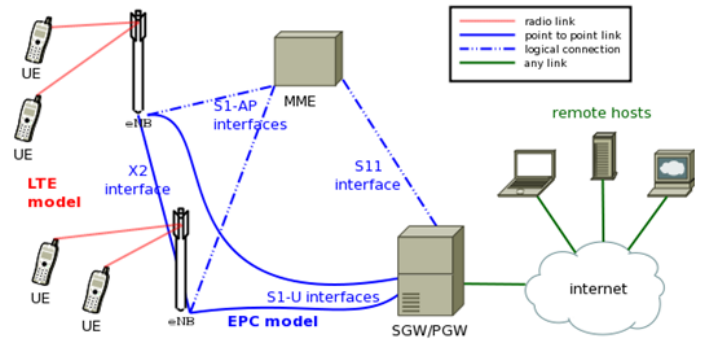


Fig. 4. ns3 LTE network architecture [4].

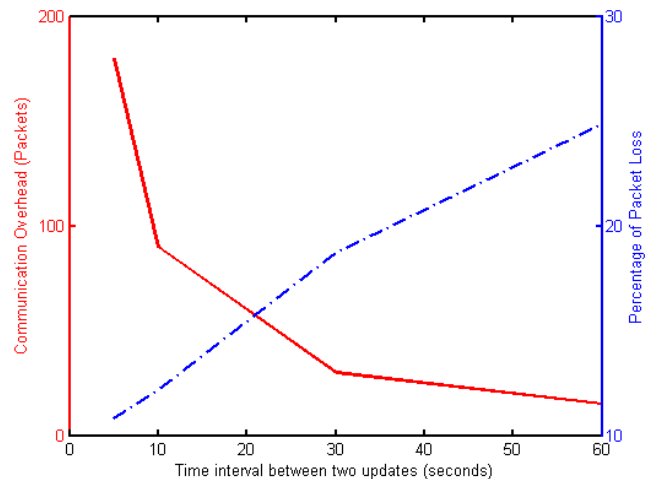


Fig. 5. Communication Overhead and Packet Loss Vs frequency of updates

base station association policies on the signalling load would be investigated. Considering users' mobility prediction at the controller is another avenue for our future research

VIII. ACKNOWLEDGEMENT

This work has been supported in part by the 5GPP NORMA (NOvel Radio Multiservice adaptive network Architecture) Project.

REFERENCES

- [1] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data off-loading: technical and business perspectives," *Wireless Communications, IEEE*, vol. 20, pp. 104–112, April 2013.
- [2] T. Mahmoodi and S. Seetharaman, "Traffic jam: Handling the increasing volume of mobile data traffic," *Vehicular Technology Magazine, IEEE*, vol. 9, pp. 56–62, September 2014.
- [3] K. Nagaraj and S. Katti, "Procel: Smart traffic handling for a scalable software epc," in *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking, HotSDN '14*, pp. 43–48, ACM, 2014.
- [4] G. Piro, N. Baldo, and M. Miozzo, "An lte module for the ns-3 network simulator," in *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques (SIMUTools'11)*, pp. 415–422, 2011.
- [5] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: enabling innovation in campus networks," *ACM SIGCOMM CCR*, vol. 38, no. 2, 2008.