Towards Location Management in SDN-based MCN

Maja Sulovic, Clarissa Cassales Marquezan, Artur Hecker
European Research Center, Huawei Technologies, Munich, Germany
Email: \{maja.sulovic, clarissa.marquezan, artur.hecker\}@huawei.com

Abstract—One of the key functionalities in EPC (Evolved Packet Core) is Mobility Management (MM), which provides procedures for service continuity as the mobile devices (UE) move across the network. In a quest for more flexibility of MM, recent studies suggested to map MM procedures to SDN applications in an SDN controller. So far, these proposals investigated and showed how path switching during handover, one of the procedures of MM, is achieved in such an SDN-based Mobile Core Network (MCN). However, Location Management (LM) procedures, such as paging, are equally important. In this paper, we address this gap. We notably design and implement SDN-based UE state management and paging procedure for MCN. Our design only uses the data from the SDN controller, such as UE connectivity and flow information, and defines a new set of UE states: deregistered, idle, and active. We introduce new, dynamically configurable UE inactivity timers that regulate the transitions among these states. With this, we implement a purely SDN-based paging procedure capable of bringing a UE into a connected state on request, e.g. to deliver the downlink traffic. The experiments with our implementation on Floodlight SDN controller under different network loads and configuration settings for UE inactivity timers in Mininet demonstrate the practical feasibility of an SDN-based LM. Further, we show how different UE inactivity timer values can flexibly regulate the amount of paging. All in all, our results suggest that, leveraging the programmability provided by SDN, operators can flexibly install and use LM SDN apps with different settings to tailor the amount of paging according to the specific UE or user application activity patterns and its needs.

I. INTRODUCTION

In the ongoing design of the 5G mobile networks, flexibility is believed to be the key feature to support the expected variety of use cases and massive amounts of mobile devices. Achieving such flexibility while keeping the complexity low is a challenge for the next generation Mobile Core Networks (MCN) [1], [2]. The state of the art suggests Software Defined Networking (SDN) and Network Function Virtualization (NFV) as key enablers for 5G MCN [3], [4]. Current approaches can be divided in two groups. The first encloses proposals that maintain almost all EPC control signaling and entities [5], [6], [7]. They use virtualized EPC entities as VNFs (Virtual Network Functions), and the SDN controller to flexibly manage the interconnections of the latter. The other group includes solutions changing the signaling and the entities of the control plane. In this case, the services of the EPC are maintained, but the implementation uses SDN Applications as basic service elements [8], [9], [10]. While both approaches can achieve the flexibility, the latter has an important advantage of leveraging the SDN infrastructure for a broader set of functions, hence removing the need for an additional NFV infrastructure or for non-trivial coordination between the SDN and the NFV subsystems. Our view on the evolved MCN in this paper is aligned with the solutions in the second group.

The native MCN functionality is mobility management (MM), including the prominent handover management for UEs with active sessions, i.e. in “connected” state. In the current LTE EPC (Evolved Packet Core), the Mobility Management Entity (MME) and the basestations (eNB) are responsible for handling UE mobility. To spare battery resources, UEs regularly become idle. The transition from connected to idle state in LTE depends on a static UE inactivity timer preconfigured in each eNB in the network [11]. With this, whenever an eNB detects inactivity of a UE, it signals this event to the MCN. On this event, the MME orders a partial release of data plane resources for that UE and changes its state from connected to idle.

Hence, another set of crucial MM procedures is location management (LM), employing tracking area (TA) updates and paging to track the attachment point of idle UEs. Here, the TA update allows the MCN to track the approximate attachment point of idle UEs. Now, to deliver downlink data to idle UEs, paging must be performed within the latest tracking area to switch the UE state to connected and to retrieve its precise attachment point. LTE defines one state transition mechanism to fit all UE and application types in the network, which may not be optimal as discussed in [12]. Yet, paging represents an important load: while LM in the overall LTE signaling traffic accounts for up to one third of LTE signaling [13], TA updates account for 4.9% and paging represents more than 28% of the overall MME load. Still, previous research did not investigate, how such procedure can be provided in an SDN-based MCN.

To address this gap, we design and implement an SDN-based state management and paging procedure for MCN. Our idea is to use the data available at the SDN controller to design a new set of UE states. Then, we introduce new, dynamically configurable UE inactivity timers that regulate the transitions among such states. Overall, our approach achieves the flexibility by enabling the configuration of these timer values per UEs and applications; what is more, it supports dynamic changes of the configured values in runtime, e.g. in response to network events and conditions.

We implemented our ideas on the Floodlight SDN controller and used Mininet to run tests under different network loads. When all UEs have the same inactivity timer value, our experiments show that: i) the higher the UE activity, which we control with the inter-arrival time of session requests, the lower is the number of page messages that need to be triggered; ii) a variation of 1s in a scenario of high UE activity (e.g., changing from 1s to 2s the inter-arrival time for session requests) has a large effect in the increase of paging to be triggered (e.g., it shows an increase of 40% of paging triggered), while when this variation occurs in low UE activity, the increase in the amount of paging to be triggered increases by only 5%. As our goal was not to improve the paging procedure per se, these results,
which prove a comprehensible system response, underline the practical feasibility of our approach. What is more, they show that different UE inactivity timer values effectively regulate the amount of paging. Combined with the flexibility of the SDN paradigm, where instances of various SDN apps could be used for different groups of UEs, distinct user application classes, etc., our main contribution is a system that allows an operator to precisely tailor the overall paging in the system to the needs and policies. We believe that such fine tuning is key for 5G, where much more devices of different types are expected.

This paper is organized as follows. Next session discusses related work. Section III presents the technical backgrounds. Section IV introduces the design of our SDN-based location management solution. Section V presents emulation results of the solution. Finally, in Section VI, we conclude.

II. RELATED WORK

Several approaches addressed optimization of location management signaling in wireless networks. Bagaa et al. [14] proposed the implementation of a framework for efficient tracking area list management (ETAM). ETAM tunes the trade-off between tracking area update (TAU) and paging signaling messages. It finds the optimal distribution of Tracking Areas (TAs) in the form of Tracking Area Lists (TALs) and assigns TALs to the UEs, according to their mobility patterns or geographical distribution. The authors have shown that optimization of TALs in the network achieves better balance between the amount of triggered location updates and paging procedures. As it is dedicated to TALs, we regard this work as complementary to ours. Arouk et al. [15] acknowledged the need to provide different mechanisms of paging in 5G networks. They argue that the increase in data traffic volumes in 5G could cause congestion in both AN and MCN, which would further introduce intolerable delays, packet loss, or even service unavailability. They proposed an optimization of current 3GPP Group Paging (GP), which reduces the amount of signaling overhead in RAN. Similar ideas can be combined with our approach. Most of previous proposals for MCN on SDN are focused on handover management and path switching solutions. For instance, Marquezan et al. [8] proposed an SDN-based MCN for path switching using reactive and proactive design choices for handling the MCN signaling when a handover occurs. Ali-Ahmad et al. [16] also addressed handover management and proposed introduction of novel mechanisms that must be both distributed, in order to avoid bottlenecks, and offered dynamically, to reduce the signaling load and improve the overall performance. They propose a distributed MM to manage two types of handover events: when a UE moves within one local area, the handover event can be managed solely by the local controller, and when a UE attaches to the access point in another local area, the handover event must be managed by two local and one regional controllers. While similar SDN designs could also host our LM SDN apps, the authors did not consider LM in their work. Sama et al. [12] is one of the few proposals to consider LM in SDN. They define a new set of messages for current LTE procedures such as attachment and session setup. They also consider the flow expiry mechanism of OpenFlow switches in order to determine, whether a UE is in idle of connected states. They provide analytical means to calculate the signaling, and one of their formulae indicates the signaling load for session setup, for the precise case when the RAN bearer expires but not the MCN bearer. Therefore, this work is dedicated to the RAN idle modes and does not specify how to page a UE when the MCN bearers expire as well. While this approach fits the LTE “always on” principle, 5G is expected to support other types of services and devices.

III. BACKGROUND

In this session, we first discuss how UE state management and paging procedure work in LTE. Then, we introduce the basic concepts of the Mobility Management SDN Application (MMA) [8], which we used as a basis for the contribution in this paper.

A. Location Management in EPC

In LTE, the MME must know the exact attachment point of the UE in order to provide mobile service. The UE is not required to constantly stay connected. At the same time, the MME will not page the UE in the entire network on downlink data delivery. Instead, 3GPP partitions the network in so-called paging or tracking areas (TA). The UE can effectively move within the TA without sending anything, while the MME only has to flood the known TA when paging a UE [11].

The procedures defined within EPS Mobility Management (EMM) protocol and EPS Connectivity Management (ECM) protocol define different states for a UE upheld by the MME, as illustrated in Figure 2. The EMM states result from the procedures like Attach and Tracking Area Update. UE can be in one of the two EMM states, EMM-DEREGISTERED or EMM-REGISTERED. EMM-DEREGISTERED implies that the UE is not attached to the network; MME does not have routing or location information. After a successful attachment, the state of the UE from MMEs perspective changes to EMM-REGISTERED and the MME knows the location of the UE at least to the accuracy level of the TA (depending on the ECM status). The ECM states describe the signaling connectivity between the UE and the EPC. UE can be in one of two ECM states, ECM-IDLE and ECM-CONNECTED. When UE attaches to the network, MME determines which Serving and PDN Gateways (SGW and PGW) will serve the UE and creates an always-on GTP tunnel between them, dedicated to that UE (S5 bearer). MME also establishes the tunnel between the eNodeB and SGW (S1 bearer) and the default radio bearer between the UE and the eNB. With the established S1, S5 and radio bearers, the UE is in ECM-CONNECTED state, and MME knows its exact location to the accuracy of the serving eNB. After a certain period of UE’s inactivity (detected by the expiration of the RRC inactivity timer at the eNB, usually in range of few seconds to few tens of seconds), the eNB sends a “UE Context Release” message to the MME. The MME then requests the SGW and the eNB to release the S1 bearer and radio resources dedicated to that UE, respectively. The MME switches the state of the UE to ECM-IDLE and considers its location to be known to the accuracy level of TA. The tunnel between the SGW and PGW remains established, when the UE’s state is switched to ECM-IDLE (hence the name “always on”).

If a downlink data needs to be delivered to a UE in ECM-IDLE state, the SGW buffers the data from PGW and
B. SDN Based Mobility Management Application

In [8], the authors proposed a proactive (MMP) mobility management design. Their solution is solely based on SDN to handle changes at the MCN level. This assumes that eNBs support the OpenFlow protocol. When a UE needs to be handed over to another eNB, the eNB will send a message to the controller (using a PACKET_IN) to signal the controller and MMP that a handover is happening. The MMP pre-installs flow rules, which match UEs data traffic in the neighboring eNBs, proactively, i.e. before the handover event occurs. However, the MMP does not consider location management and how specific UE states can influence the downlink data delivery. We adopted the MMP as the baseline architecture for the UE LM, state management and paging procedure proposed in this paper.

IV. PROPOSAL OF LOCATION MANAGEMENT IN 5G MCN

Our design leverages three mechanisms. First, we rely on the Flow Information Base (FIB) from the previous work [8]. Originally, this FIB on the SDN controller stored information about all UE flows, deployed flow matches and the current UE attachment point. We extended it to store flow expiration timers that control session activity. Second, we only use the native capabilities of spec-conform OpenFlow switches. OpenFlow defines the IDLE_TIMEOUT field in a flow entry as the period of time in seconds without a single packet matching that flow entry; after that time the flow entry expires and must be removed from the flow table. The IDLE_TIMEOUT field of each flow entry is set to the value of previously mentioned flow expiration timer from FIB. A flow entry also has a flag field (i.e., OFPFF_SEND_FLOW_REM), which indicates to the OF switches to send flow removal notifications towards the SDN controller, using the message OFP_FLOW_EXPIRED. Third, we used the SDN controller data about network devices. We extended this data to store information about the UE tracking area and UE state switching timers, which are essential for the paging procedure, as discussed in Section IV-A. We combine the timers that control the UE states with flow (session) expiry timers and flow entry removal notifications to indicate, when a UE needs to change its state.

Our architecture is depicted in Figure 3. We defined two new SDN applications: UE State App and SDN Paging App. The former manages the state transitions for each UE, while the latter performs the paging procedure when necessary. The other three modules, FIB, Mobile Service State Information Manager (MSSIM), and MMP are from [8], extended in this work. The MSSIM interprets a message from the OF Switch (i.e., a PACKET_IN) against the data in the FIB to decide, which event is associated with the message. We extended this module to identify the need to trigger paging, when the destination UE is in idle state. The MMP was extended to set up active flows using the flow expiry information from the FIB. One of the key features of our solution is the possibility to program both the timers that regulate flow expiration and the timers that regulate the UE state switching. Hence, our solution goes beyond the current design in LTE, in which one model fits all types of devices and service needs.

In our design, we assume a correctly dimensioned SDN layer. For approaches how to build this, see the comprehensive SDN survey [17]. Hence, in our design, as per [18], we make sure that the proposed SDN apps do not introduce global locks and states. Indeed, in our design, any added state is strictly per UE, therefore allowing to add new application instances for (groups of) UEs as necessary, which solves scalability.
A. UE State Management

The proposed UE state management enables three states as illustrated in Figure 4: ACTIVE, IDLE or DEREGISTERED. UE with at least one active flow installed in the network is in state ACTIVE. Its access point is known to the accuracy of access switch and the port number, hence state ACTIVE corresponds to the state ECM-CONNECTED from LTE. The UE State App registers to receive flow entry expiration notifications sent by the OF switches. Every time a flow expiry notification arrives, the UE State App will mark the expired flow in FIB as inactive and check the remaining active flows of the source and destination UE. If there are no more active flows originating or terminating in any of these UEs (or in one of them), the UE State App starts the timer T_IDLE for that UE. If new PACKET_IN message arrives, indicating new originating or terminating session on the UE, whose timer T_IDLE is counting, the UE State App will stop the timer, reset its value and forward the PACKET_IN to the MMP to establish flow path. Otherwise, if the timer T_IDLE expires, the UE transitions from ACTIVE to IDLE. The attachment point of a UE in IDLE state is considered to be known to the accuracy of the 5G TA (TA in LTE), which corresponds to the state ECM-CONNECTED in LTE. If a new PACKET_IN requires establishment of a new flow path towards a UE in IDLE state, the MSSIM will interact with the Paging App to trigger paging and bring UE to ACTIVE state. Finally, a UE transits from IDLE state to DEREGISTERED, if its inactivity period exceeds the predefined period T_DEREGISTER (⇒T_IDLE), also controlled by the UE State App. Such UE is considered to be detached from the network and will not be paged. State DEREGISTERED corresponds to the state EMM_DEREGISTERED in LTE. A UE has to register to the network to change its state from DEREGISTERED to ACTIVE.

B. Paging procedure in SDN-based MCN

The overall steps for performing paging are illustrated in Figure 5. The trigger for paging (step 1) is PACKET_IN analyzed by the MSSIM application. Then, the Paging App starts the paging procedure by sending PACKET_OUT messages (step 2) to all access switches that belong to the 5G TA, where the UE in IDLE state is registered. Each PACKET_OUT message encapsulates the paging message that targets the UE in IDLE state. To achieve this, we set the destination port in all PACKET_OUT messages to ALL. Each switch in 5G TA de-capsulates the paging message and floods all ports with it, except the ingress port. The paging message reaches all UEs registered in 5G TA (step 3.a), and only the one that is actually being paged sends paging response (step 3.b). The access switch will, upon reception of the paging response message, issue PACKET_IN towards the controller (step 4). The controller internally passes the message to MSSIM that recognizes this PACKET_IN as paging response and forwards the message to UE State App, which updates UE’s current attachment point and last seen timestamp. At this point, the UE State App changes the UE state back to ACTIVE and resets its T_IDLE timer. Once active, the SDN Apps can establish flow paths towards such UE.

C. Implementation

We implemented our solution, addressing SDN-based state management and paging procedure, using the open source Floodlight SDN controller (version 1.0). We emulated our solution on a Mininet platform extended to support mobility. This platform has limitations related to the emulation of the radio signaling. For this reason, we could not implement the paging at the radio side. Instead, we created paging request and response messages that could be transmitted from the APs (which in our platform are OF Switches). The paging request message is a UDP packet targeting a certain closed port at the UE in IDLE state. When the UE in IDLE state receives this UDP packet, it sends back a response message “ICMP UDP Port Unreachable”. This message is the paging response in our solution. When AP (OF Switch) receives paging response, it sends a PACKET_IN to the SDN controller. When this PACKET_IN arrives at the controller, the Device Manager will update its internal database with UE’s attachment point. The PACKET_IN message is further passed to the UE State App to update the UE state to ACTIVE and reset the T_IDLE timer.

V. EVALUATION

In this section, our goal is to demonstrate the practical feasibility of the SDN-based LM approach. Note that we are not aiming to evaluate the LM procedures per se (i.e., paging). As our design is SDN-based, it is relatively easy to implement other paging logics, such as discussed in Section II.

Instead, we pursue two goals: first, we want to know whether our SDN-based LM works as expected. Second, and more importantly, we want to show that the parameters of our proposal effectively regulate the amount of paging in the system. To do this, we study the impact of two parameters on the system paging: the frequency of PACKET_IN messages and the T_IDLE timer value. The evaluated metric is the percentage of downlink deliveries that triggered paging. The
testbed for the evaluation is a two-layered (access and core) Mininet topology. Parameters used in the emulation are shown in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of access switches</td>
<td>10</td>
</tr>
<tr>
<td>Number of core switches</td>
<td>500</td>
</tr>
<tr>
<td>Number of sessions</td>
<td>250</td>
</tr>
<tr>
<td>Gamma distribution shape</td>
<td>2</td>
</tr>
<tr>
<td>Session duration</td>
<td>10s</td>
</tr>
<tr>
<td>Session expiry timer IDLE_TIMEOUT</td>
<td>5s</td>
</tr>
</tbody>
</table>

In our emulation, each host in the network starts communication with other randomly chosen host, whereby the time between two subsequent communication requests from one host is random and follows exponential distribution. Each new session request results with PACKET_IN on the SDN controller, so their arrival follows the same distribution. Since all hosts in the network communicate simultaneously and independently, the inter-arrival time of all PACKET_IN messages on the SDN controller (triggered by all subscribers) will be a sum of independent exponentially distributed random variables and, as such, it will follow a two-parameter gamma distribution (PDF and mean value as in Eq. 1 and Eq. 2, respectively). By varying the parameters of the gamma distribution, shape \(k\) and scale \(\theta\), it is possible to simulate more or less frequent arrival of PACKET_IN messages on the SDN controller. These inter-arrival times correspond to the different overall UE activity in the network, where less and more frequent arrival implies lower and higher overall activity of UEs, respectively.

\[
f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1}e^{-\frac{x}{\theta}} \quad x \geq 0 \text{ and } k, \theta > 0 \quad (1)
\]

\[
E[x] = k\theta \quad (2)
\]

The goal of the first test is to determine, how the overall UE activity in the network impacts the amount of paging requests. In scenarios with increased UE activity (higher frequency of PACKET_IN arrivals), the probability that a randomly chosen host at a randomly chosen time has at least one active flow (i.e., it is in the ACTIVE state) will also be increased. Hence, the probability that such chosen UE will be paged before setting up the flow path for a received PACKET_IN is accordingly lower. Figure 6 shows the results of the experiments for this scenario. We varied the load of the network and observed the percentage of paging triggered. The x-axis represents the average PACKET_IN inter-arrival time in the network and y-axis represents the percentage of PACKET_INs that triggered paging before the establishment of the flow path between the two communicating UEs. The first observation is that the higher the UE activity in the network, i.e., the lower the inter-arrival time, the lower the amount of triggered paging. This is aligned with our expectations. For instance, when average inter-arrival time of PACKET_IN messages is set to 1s (shape 0.5 and scale 2), the percentage of paging is around 32%. If the mean inter-arrival time is reduced and set to 2s, the percentage of PACKET_INs that triggered paging almost double. The increase in the amount of paging remains, as we increase the inter-arrival time (which in its turn, reduces the UE activity in the network); nevertheless, this increase in the percentage has a lower trend of growth. The results show non-linear dependency on the overall UE activity, where the small difference, when
there is high UE activity can significantly increase the amount of triggered paging requests, but the same variations with low UE activity does not generate the same increase in triggered paging requests. We can conclude that the proposed solution can implement UE state aware paging procedure as in LTE today. However, the observed non-linear behavior confirms that an approach, where the timers can be programmed is able to optimize the amount of signaling produced by the paging procedure.

The second test investigates the impact of timer $T_{\text{IDLE}}$ value on the amount of paging requests. The tests have been carried out using three scenarios with different UE activity the inter-arrival time between two subsequent PACKET_IN messages in the network was modeled using gamma(0.5, 2), gamma(0.75, 2) and gamma(1, 2), which corresponds to average inter-arrival time equal to $1s$, $1.5s$ and $2s$, respectively. Values used for the timer $T_{\text{IDLE}}$ are 2, 5, 10, 20 and 30 seconds. The results are depicted in Figure 7. As expected, the increase of the inactivity timer reduces the amount of triggered paging requests. Moreover, there exists approximately linear relationship between $T_{\text{IDLE}}$ timeout value and the percentage of PACKET_INs that triggered paging. When the $T_{\text{IDLE}}$ value is increased of 5 seconds, we observe an average reduction of 4% in the amount of triggered paging requests in all tested scenarios. This experiment shows that dynamic programmability of $T_{\text{IDLE}}$ in the proposed approach unfolds the possibility to tailor paging curve according to the needs of each type of device, application, or any other MCN operator requirement.

![Fig. 7: Paging percentage depending on the value of $T_{\text{IDLE}}$](image)

VI. CONCLUSION AND FUTURE WORK

This paper shows that it is possible to design and implement an SDN-based UE state management and paging procedures for MCN. Furthermore, it also shows that the programmability offered by SDN can improve the performance of paging procedures, reducing signaling load by enabling the capability of tailoring the need to trigger paging to UE types, applications or other MCN operator requirements. As a future work, we plan to investigate solutions for dynamically defining UE inactivity timer values based on different network performance metrics.

REFERENCES


