Double-NAT Based Mobility Management for Future LTE Networks

Morteza Karimzadeh[1], Luca Valtulina[1], Aiko Pras[1],
Hans van den Berg[1,2], Ricardo de O. Schmidt[1]
[1]University of Twente, [2]TNO, The Netherlands

Marco Liebsch
NEC Laboratories Europe,
Germany

Tarik Taleb
Aalto University,
Finland

Abstract—In this paper we discuss the major modifications required in the current LTE network to realize a decentralized LTE architecture and develop a novel IP mobility management solution for it. The proposed solution can handle traffic redirecting and IP address continuity above the distributed anchor points in a scalable and resource efficient manner. Our approach is based on the NAT (*Network Address Translation*) mechanism, which is a well-known and widely used procedure in the current Internet. We extend the NS3-LENA to implement a decentralized LTE network as well as the proposed scheme. The evaluation results show that the proposed solution efficiently fulfills the functionality and performance requirements (*e.g.*, latency and signaling load) related to the mobility management.

I. INTRODUCTION

Cisco forecasted that the worldwide mobile data traffic will be increased more than 8-fold between 2015 and 2020 [1]. Coping with such a demand in the current mobile networks is neither economically nor technically viable. The RAN (*Radio Access Network*) cannot be easily extended due to spectrum limitations. Furthermore, the core of mobile networks is highly centralized, introducing scalability and reliability issues.

Mobile network operators augment RAN capacity by improving spectrum utilization in several ways, *e.g.*, deployment of small cells and exploiting multi-carrier and multi-radio access approaches [2]. The major challenge related to the core networks (standardized by 3GPP, IETF) is mainly due to the fact that a few high level network elements, entitled *anchor points*, handle both the *Data plane* and the *Control plane*. Such centralization makes the network prone to several limitations, *e.g.*, sub-optimal routing, low scalability, signaling overhead, and lack of granularity on services [3, 4]. The straightforward and short-term solution to cope with the core networks issue may consist of operators investment to upgrade the resources. This approach is technically feasible. However, network operators always stand for the cost-effective and long-term solutions. Traffic offloading is an alternative approach to mitigate the traffic impact to limited resources in the core networks. That can be achieved by placing small-scale anchor points in the proximity of the access network to handle Mobile Nodes (MNs) connections and traffic locally [5]. This essentially leads to a decentralized network architecture. Relocation of the mobile devices' edge anchor points helps maintaining efficient routes for mobile users' connection. However, it demands additional mechanisms to maintain the mobile users' ongoing data sessions by enabling IP address continuity to the mobile devices and steering the data packets towards the new anchor points.

LTE (*Long Term Evolution*) network is expected to be the leading mobile networking technology in the coming 5-10 years. While it supports 47% of mobile traffic nowadays, it is estimated to handle about 72% of the worldwide mobile data

traffic by 2020 [1]. Therefore, in this paper we pay special attention to an architecture for realizing a decentralized LTE network for the 3GPP access and introduce a novel solution to support IP address mobility and to enable and maintain efficient routing paths in such an architecture.

Due to space limitation, we refer the readers to our previous research [6] as the related work. There we analysed and compared several key techniques as the enablers to manage IP address and traffic session continuity in a mobile network with distributed anchor points. Our main contributions in this paper can be summarized as follows:

- We discuss the modifications required in the current LTE network to realize a decentralized LTE architecture and to support the MNs' mobility in the new architecture.
- We develop a NAT-based mechanism to enable IP address and data traffic continuity in a decentralized LTE network, that can be easily implemented using the existing technology with low impact on the network implementation.
- We extend the NS3-LENA simulation environment to support the implementation of a decentralized LTE network as well as the proposed solution.
- Using the new implemented LTE architecture, we verify the function and performance requirements of the developed mechanism.

The remainder of this paper is organized as follows: § II introduces concisely the current LTE network architecture and its mobility management solutions. It also describes the main modifications required to realize a decentralized LTE network and handle MNs mobility in this architecture. § III presents our proposed solution and details its function and components. § IV describes the implementation of the LTE's new architecture and the developed solution in the NS3-LENA. § V defines the performance metrics and presents the obtained simulation results. Finally, the paper ends up with conclusion at § VI.

II. LTE NETWORK

This section gives a brief overview of the current LTE system (§ II-A) and its mobility management mechanisms for the 3GPP access (§ II-B). It also describes an approach to realize a decentralized LTE network deployment (§ II-C) and the required modifications to support the MNs' mobility in this architecture (§ II-D). These are essential for understanding the problem statement being addressed in this paper.

A. Current LTE Architecture

The existing LTE network architecture is hierarchical and defines the EPS (*Evolved Packet System*) consisting of E-UTRAN (*Evolved Universal Terrestrial RAN*) and EPC (*Evolved Packet Core*). The E-UTRAN consists of a network of radio base stations (eNodeB – *evolved Node B*), that

provides radio connectivity to the MNs. The EPC is a multi-access IP-based network that uses a common core network for the 3GPP and non-3GPP radio access, and fixed access.
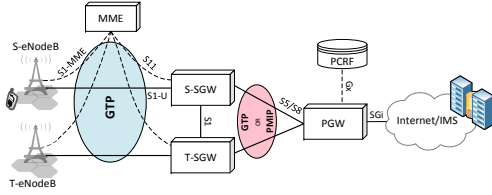

Fig. 1: Current LTE network architecture for the 3GPP access.

The EPC consists of four main elements (Fig. 1), that allow for the convergence of packet-based services [7]: The PGW (*Packet Data Network Gateway*) connects the EPC to external IP networks. It also carries out other functions, *e.g.,* IP address allocation, mobility management, and policy enforcement, *e.g.,* for QoS and charging. The SGW (*Serving Gateway*) provides data paths between the eNodeBs and PGW. It also handles the mobility of MNs between the local eNodeBs. The MME (*Mobility Management Entity*) controls the MNs in accessing to the LTE network. The PCRF (*Policy and Charging Rule Function*) that determines QoS policies and charging rules to the PGW (if GTP is used) and SGWs (if PMIP is used).

B. Mobility Management in the Current LTE Network

A mobility management mechanism supports a set of procedures to enable seamless IP address and data session continuity for moving MNs within the network. In the current LTE for the 3GPP access, mobility management is based on either GTP (*GPRS Tunneling Protocol*) or PMIP (*Proxy Mobile*) protocols, where the PGW acts as a central mobility anchor point. When a MN connects to an eNodeB, its traffic is encapsulated in a GTP tunnel between the eNodeB and SGW and in another GTP (or PMIP) tunnel between the SGW and PGW. When the MN performs a handover between two eNodeBs to keep the ongoing IP flow(s) active, the new S1-U and S5/S8 tunnels are established between the EPC and E-UTRAN entities (Fig. 1), depending on whether the target eNodeB is served or not by the same SGW. The procedure described above shows that the existing data plane and mobility management procedure are highly hierarchical, demanding management of several tunnels between the PGW and the MNs. In a large LTE network, the PGW needs to maintain a considerable number of (*e.g.,* a range of millions for a nationwide network) per-user tunneling data, which may cause scalability and performance issues.

LIPA (*Local IP Access*) and SIPTO (*Selected IP Traffic Offload*) have been introduced in the 3GPP Release 10 to alleviate data traffic load on the LTE's core network. However, they are limited in supporting the MNs' mobility only for the local nodes [8].

C. Decentralized LTE Architecture

The hierarchy of the data and control planes in the current LTE network can be eliminated by co-locating the SGW and PGW functions into a single entity, entitled as *S/PGW*. Accordingly, the S/PGWs are distributed closer to the edge network and can handle the MNs' connection functions, data traffic, and mobility locally. This approach basically leads to decentralized LTE architecture and effectively reduces the load on the core network entities (Fig. 2). In the current LTE, MNs

rarely change their attached PGW. If this happens (*e.g.,* during inter-operator roaming) a *PDN Disconnection* procedure will be triggered by the network for the IP flow(s) initiated at the previous PGW. Next, the new PGW anchors the MNs and serves the re-initiated traffic. This implies a disruption in the MN's ongoing traffic, since the MN's IP address is not maintained but a new one is assigned.
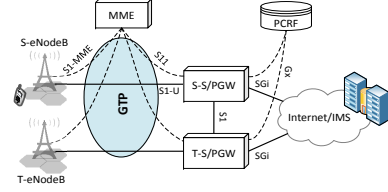

Fig. 2: Decentralization of the LTE core network architecture.

Following a decentralized architecture, relocation of the MNs' mobility anchors (S/PWG) will happen far more often. In this case, two layers of mobility management need to be handled to support IP traffic continuity for the MNs moving to a different S/PGW: (i) inside the EPC network (between the S/PGW and the MNs); and (ii) above the EPC network (between the S/PGW and the data networks), which hosts the MNs corresponding services. The following section describes the required modifications in the current EPC data and control planes for supporting the MNs' mobility inside the EPC network of a decentralized LTE architecture. The mobility support above the EPC network is based on our developed mechanism, presented in § III.

D. Mobility Management inside the EPC of a Decentralized LTE Architecture

IP address continuity is not supported in the current 3GPP's LTE standard, upon changing the attached PGW. This is due to the fact that there is neither a signaling nor data forwarding scheme available between two different PGWs. Following a decentralized architecture, the existing control messages and the traffic forwarding mechanism, used during a SGW relocation, can be revised to use the same IP address in different S/PGWs. This modification should enable the following functions: (*I*) the target S/PGW must be informed to implement a GTP bearer for the moving MN without requiring a new IP address allocation. This bearer is used to keep active the MN's flow(s) after handover; and (*II*) the source S/PGW must be informed when the IP address used by the moving MN can be released.

During a MN's attach procedure three types of bearers are set up to transport MN's traffic between the EPS entities (Fig. 3). The S1 and S5/S8 bearers use the GTP protocol to identify the individual connections between two nodes. A TEID[1] (*Tunnel Endpoint Identifier*) is assigned to each GTP bearer allowing the nodes to determine to which specific bearer a particular packet belongs. Each EPS Bearer is associated with one TFT (*Traffic Flow Template*), defining the filtering rules to differentiate data packets. In a decentralized EPC, the S5/S8 bearer is unnecessary and direct mapping from External bearer to S1 bearer (DL-TFT→S1-TEID) can be performed in the S/PGW. A procedure to adjust the S1-TEID in a X2-based handover with SGW relocation[2] is specified in § 5.5.1.1.3

---

[1]For the PMIP, a GRE (*Generic Routing Encapsulation*) key is used as a identifier.

[2]The S1-based handover can also be used to handle MN's mobility during a SGW relocation (§ 5.5.1.2.2 of [9]). As we used the X2-based for development of our solution, in this paper we ignore to present the required modifications for the S1-based approach.
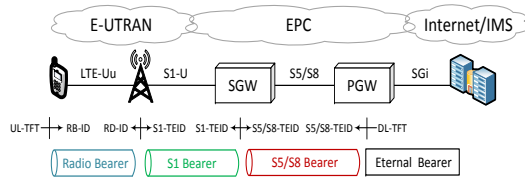
Fig. 3: EPS bearers in the current LTE network.

of [9]. We considered it as a baseline and defined a new handover mechanism to cope with the issues described in (***I***)[3]. We modified the *Create Session Request/Response* and *Modify Bearer Request/Response* messages between the MME and target S/PGW (which is replaced as PGW) to create a new DL-TFT in the target S/PGW. This changes handle the migration of an EPS bearer to a newly established data forwarding plane for the moving MN. To realize the function described in (***II***)[3], we modified the *PDN disconnection* procedure specified in § 5.10.3 of [9]. This procedure allows the MN to request for disconnection from the network. By receiving a *PDN Disconnection Request (LBI)* message, the MME exchanges the *Delete Session Request/Response* messages with the target S/PGW to inform the list of bearers to be released for a particular PDN connection. After step 10 (in § 5.10.3 of [9]), the MME exchanges other *Delete Session Request/Response* messages with the source S/PGW to inquire it to drop the corresponding bearer context from its list and to release the associated MN's IP address(es).

III. DOUBLE-NAT BASED MOBILITY MANAGEMENT

This section describes the functional approach, architecture, components, and control messages of our proposed solution to handle the MNs' mobility above the EPC network, during a S/PGW relocation in a decentralized LTE architecture.

A. Functional Approach

As discussed in [6], SDN (*Software Defined Networking*) paradigm and NAT mechanism are promising enablers to efficiently (re)direct the MNs' traffic upon anchor points relocation. Our proposed SDN-based solution to support the MNs' mobility discussed in [10], demands extra modification and management efforts in the network. In this paper, we develop a new approach relying on the NAT mechanism, that can be readily deployed with low impact on the existing network implementation. This solution adopts the *Identifier-Locator* split concept, previously discussed in the DMM (*Distributed Mobility Management*) working group Internet draft [11]. The use of per-host locator's IP address, allows translation of IP address above the EPC network to route the data traffic towards the MN's current S/PGW, while address translation is kept transparent to the endpoints. In this context, *Identifier* refers to the IP address of MN's flow(s), allocated by the first attached S/PGW (e.g., *A.X.X.1*) and kept active after a handover. The IP address assigned from the second attached S/PGW's address pool (e.g., *B.Y.Y.1*) is referred to the *Locator* address and is used to forward downlink traffic to the MN's new position.

---

[3]Due to space limitation, we only discuss the messages modified for our propose, and refer the readers to § 5.5.1.1.3 and 5.10.3 of [9] for more information.

B. Architecture

In a decentralized network architecture, MN's IP address is anchored at a distributed anchor point and may need to be (temporarily) maintained to keep the ongoing sessions active, when the anchor point is relocated. This may change the MN's IP address from a *routable* (topologically correct) into a *non-routable* (topologically incorrect) address at the new anchor point. In this condition, the transport network needs to steer the MN's downlink traffic to the new anchor point. Fig. 4 shows the architecture of proposed approach to handle the above mentioned aspects in a decentralized LTE network. The NAT mechanism is used at the edges of the transport network to redirect MN's downlink traffic to its new location. For this, two new entities, named *Ingress NAT* and *Egress NAT* routers, perform IP address translation from *Identifier → Locator* address and contrariwise, respectively. As the IP address translation is applied twice in the MN's downlink path, this approach is entitled as the *Double-NAT*.
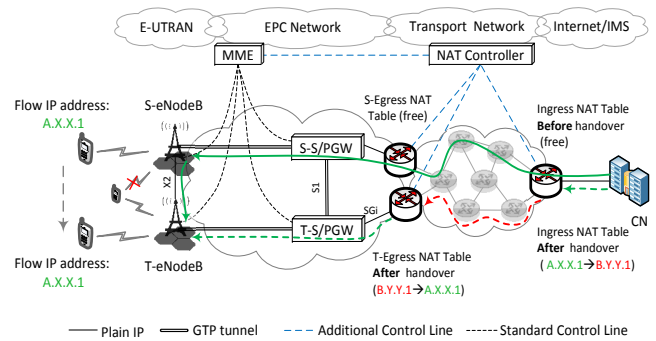

Fig. 4: Double-NAT based traffic redirection path and architecture.

During a handover procedure, the MN's traffic forwarding and IP address continuity between two eNodBs (inside the EPC) are handled by the mechanism explained in § II-D. After completing the handover, in the transport network (above the EPC network) the Double-NAT solution is used to redirect the flow(s), previously initiated by the MN, to the target S/PGW. For this, the target S/PGW allocates an IP address (e.g., *B.Y.Y.1*) from its address pool to the handed over MN, by receiving a specific signal from the MME (§ II-D). This IP is not advertised to the MN and is only used as the flow's *Locator* address within the transport network. Next, the Ingress and Egress NAT routers take care of the correct translation of the *Identifier* ⇌ *Locator* addresses to redirect the MN's flow(s) to the MN's new position. Upon arrival of the data packets at the target S/PGW, they are processed and encapsulated into a GTP tunnel and then forwarded to the target eNodeB to be delivered to the MN, via the LTE air interface.

Upon fulfilling a handover by the MN, a new NAT rule needs to be proactively defined in the Ingress and Egress NAT routers' tables. Furthermore, when the MN terminates the connection, its corresponding rules must be removed from the NAT tables. To accomplish these an entity entitled *NAT Controller* is introduced in the network. NAT Controller needs the following information to signal the NAT routers: (i) the flow's *Identifier* address; (ii) the flow's *Locator* address; and (iii) the IP addresses of the NAT routers. During a handover procedure the target S/PGW and its serving MME have the knowledge about this information. Being a control plane procedure, we chose the MME to signal the NAT Controller. Furthermore

one MME can serve multiple S/PGWs, reducing the amount of connections between the EPC and NAT Controller.

C. Control Messages

Two steps of signaling must be performed to create/remove a Double-NAT traffic path in the transport network:

1) Messages from MME → NAT Controller: MME uses two messages to signal the NAT Controller: (i) *Create Path* to setup a downlink path; and (ii) *Remove Path* to remove a previously established downlink path in the transport network. These messages carry the *Identifier* and *Locator* IP addresses. MME receives this information from the target S/PGW, via the *Modify Bearer Response* message (§ II-D).

2) Messages from NAT Controller → NAT Routers: NAT Controller uses two messages to signal the NAT routers: (i) *Update Rule* to add/modify a rule into the NAT tables; and (ii) *Remove Rule* to remove a rule from the NAT tables. These messages carry the *Identifier* and *Locator* IP addresses as well as the time that a rule is created/modified in the tables.

D. Control Tables

Three types of tables are used for signaling or traffic redirection procedures in the Double-NAT approach.

1) MME's Table: It contains information about the NAT Controller(s) present in the network. This table is used to find the appropriate entity for signaling.

2) NAT Controller's Table: NAT Controller has two tables containing the Ingress and Egress NAT routers information (*e.g.,* the IDs, IP addresses, TCP ports). This data is used by the NAT Controller to find the correct NAT routers to be signaled.

3) NAT Routers' Tables: The Ingress and Egress NAT routers keep per-host states in their tables to steer the MN's downlink traffic in the transport network, and towards the corresponding target S/PGW, respectively.

E. Flow of Messages

This section describes the flow of control messages between the EPC and Double-NAT components (Fig. 5).

1) During Handover Procedure: by (1) target S/PGW notifies the MME about the newly assigned IP address to the MN (§ II-D). (2) MME looks in the MN's context and retrieves the IP address of the flow(s) kept active by the MN after handover. Next, it sets the old and new IP addresses as the *Identifier* and *Locator* addresses in the *Create Path* message and sends it to the NAT Controller. (3) NAT Controller uses this information to construct the *Update Rule* message, sends it to every active NAT routers, and saves the IDs of the signaled routers. Upon receiving the *Update Rule* messages by the NAT routers, a new rule is added to the corresponding tables.

2) After Disconnection Procedure: by (1) MN sends a *PDN Disconnect Request* message including the LBI (*Linked EPS Bearer ID*) to the MME and asks to disconnect from the network (§ II-D). (2) MME extracts the current and previously assigned IP addresses of the MN from the MN's context and LBI. A *Remove Path* message is created by setting the old and new IP addresses as the *Identifier* and *Locator* addresses, and then it is sent to the NAT Controller. (3) The NAT Controller uses this data to create a *Remove Rule* message and then sends it to the signaled NAT routers. When the NAT routers receive this message, they look up at the related tables for an entry

whose data matches with the received information. If any entry is found, it is removed from the table.
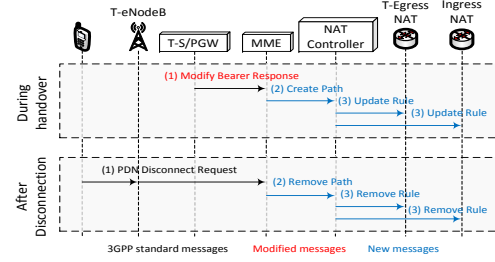

Fig. 5: Double-NAT control messages flow.

IV. SETUP OF SIMULATION STUDY

This section presents the simulation scenario, implementation of the components, and setting parameters in the simulation environment. We set up the simulation environment as realistic as possible to get reliable results, as follows.

A. Simulation Scenario

The logical network topology of the simulation scenario is the one shown in Fig. 4. During 20 second simulation, 120 MNs attach to both eNodeBs and generate the E-UTRAN traffic according to Table II. 30% of MNs from the source eNodeB, running VoIP application move to the target eNodeB at different times between 10.7 to 12.8 seconds. The X2-based (inside the EPC) and the Double-NAT solutions (above the EPS) forward the MNs' downlink traffic to the target position.

B. Implementation of Scenario in the NS3-LENA

To evaluate the proposed solution, we extended the NS3-LENA to implement the following components. The source codes to implement all of the modules can be found in [12].

1) Decentralized LTE Network: Several modifications are needed in the existing version of NS3-LENA to implement a decentralized LTE network: *(i) Instantiation of Multiple S/PGWs:* we enhanced the *EpcHelper* module of the NS3-LENA to implement two standalone EPS subsystems that use independent S/PGWs with different pool of IP addresses. The S/PGWs serve the separate eNodeBs, but use a shared MME; *(ii) Implementation of X2-based Handover Procedure with S/PGW Relocation:* having multiple S/PGWs in the simulation scenario, we modified the currently deployed X2-based handover procedure of the NS3-LENA to support relocation of S/PGW during the MNs handover and to realize the functions described in § II-D.

2) Transport Network: The transport network topology is according to a small part of the EBONE (one of the European ISPs) network, covering The Netherlands, north-east of Belgium and north-west of Germany. To implement it we used a map provided by the Rocketfuel project [13].

3) NAT Function: We used an external module developed in [14] to implement the NAT functions. Using the *Ipv4NatHelper* entity of it, IPv4 NAT capabilities can be added to a node defined in the NS3 environment (*e.g.,* NAT routers in Fig. 4).

4) Double-NAT Signaling: To implement the signaling procedures described in § III-E, the *UdpSockets* are set up between the MME, the NAT Controller, and the NAT routers entities.

5) MNs Movement: In NS3-LENA, the MNs' handover time is based on a pre-defined schedule, set up in the simulation. Therefore, no movement is needed to be performed by the MNs to trigger the handover procedures. The MNs are only placed in positions at the same distances from the eNodeBs and a distribution of dwell time is used to trigger the handovers. We used the *Fluid-Flow* mobility model [15] to derive the average dwell time of the MNs in each S/PGW. As a S/PGW relocation most likely happens for the MN that are on a highway, we used the *Free Speed Distributions* model [16] to compute the velocity of the MNs. We used a *Normal Distribution* with the *(Mean = 32.1 m/s)* and *(Standard Deviation = 4.33)* [16]. For the sake of simplicity we assumed that the MNs move in a straight road between the S/PGWs.

C. Simulation Parameters

This section presents the setting of parameters in the simulation environment.

1) E-UTRAN Setting: Table I summarizes values of the configured parameters in the RAN. The values are based on the LTE release-8 specifications, implemented in the NS3-LENA.

TABLE I: The E-UTRAN parameters.

Parameters	Value
Uplink and Downlink bandwidth	5 MHz
Source eNodeB uplink / downlink	2535 / 2655 MHz
Target eNodeB uplink / downlink	2540 / 2650 MHz
Transmission mode	MIMO 2x2
MN / eNodeB transmission power & noise figure	26 / 49 dBm & 5 dB
Cell radius / distance	2 / 4 Km

2) E-UTRAN Traffic: We used the *traffic mix* model (Table II) specified in [17] to generate the RAN traffic. VoIP is selected as the traffic generated by the moving MNs. Other types of traffic are used to generate the RAN background traffic by the fixed MNs attached to the both eNodeBs.

TABLE II: The E-UTRAN traffic model.

Application	Traffic	Percentage of MN
VoIP	Real-time	30%
FTP	Best effort	10%
HTTP	Interactive	20%
Video	Streaming	20%
Gaming	Interactive real-time	20%

3) EPC and Transport Networks Setting: The values of parameters set up for the EPC and transport networks are shown in Table III. The size of buffer for the entities in the both networks is set to $(\frac{C \times RTT}{10})$ [18]. Where C and RTT denote the link speed and the *Round Trip Time* $(= 250ms$ [18]) of the flows in the network, respectively.

TABLE III: The EPC and transport network parameters.

Parameters	Value
Transmission technology	Ethernet
MTU size	1.5 KB
Transport / EPC network links data rate	10 / 1 Gbps
Queue scheme	Drop-tail
Transport / EPC network nodes buffer size	31.250 / 3.125 MB
Radius of a tracking area (TA) of each S/PGW	50 KM

4) The EPC and Transport Networks Traffic: We used the PPBP (*Poisson Pareto Burst Process*) model [19] to generate a realistic Internet traffic in the wired networks. 80% is chosen as the maximum level of link utilization for the both networks. The rest of available capacity is used to transfer the VoIP traffic and also to keep as the safety capacity.

V. PERFORMANCE METRICS AND EVALUATION RESULTS

We evaluate the seamlessness of the proposed solution, using simulation experiments. That is, if no significant packet loss and delay are introduced during a S/PGW relocation. For this we define the following performance metrics and present the obtained results accordingly.

A. Average Latency of Data Packets Delivery Before and After Handover

The average latency of the data packets received by the moving MNs via the alternative downlink paths before and after handover is compared in Fig. 6a. The graphs clearly demonstrate that the proposed solution has no considerable impact on latency of the data packets redirected to the MNs after a handover. As the placement of NAT Controller in the transport network could be critical, we also studied the impact of it's position in the network topology (Fig. 4). Given that the MME uses the transport network to signal the NAT Controller, it is observed that no significant impact is appeared in the average latency after handover (≤ 3 ms), when distance between the NAT Controller and MME is changed from 1-7 hops. In the simulation scenarios, the Egress NAT routers is placed in a fixed position with one hop distance to the S/PGWs. Positioning of the Ingress NAT routers has no particular constrain and mainly depends on network design. The obtained results shows that after handover, the average latency of data packets delivery has slightly variation (≤ 8 ms), when distance between the Egress and Ingress NAT routers is changed from 1-7 hops. This accordingly could affect the throughput of the Double-NAT solution.

B. CDF of Latency of the First Downlink Data Packets Redirected Through the Double-NAT Approach, and Downlink Data Packets Forwarded Through the X2 Path

Fig. 6b presents CDF of latency of the first data packets delivered to the moving MNs after completion of the handover. We selected the first data packet (of the MNs), as it is the most delayed packet after the handover procedure. This is because its delivery time is directly influenced by the time required to establish the Double-NAT traffic forwarding path in the transport network. The obtained results show that the Double-NAT approach is fast enough to easily meet the maximum allowed one-way latency of 150 ms [20] as a threshold for the VoIP application. Fig. 6b also shows CDF of latency of the data packets forwarded through the X2 path during the MNs' handover. The graph shows that the Double-NAT approach significantly outperforms the X2 interface. Therefore, if a function to predict the mobility of MNs would be present in the network [21], Double-NAT traffic redirection path can be set up a priori to avoid the usage of X2 data forwarding.

C. Load of Messages in the Double-NAT Solution

In the proposed approach, to set up a traffic redirection path, several messages must be exchanged between the MME and the Double-NAT entities (Fig. 5). Assuming that only the messages of the NAT Controller towards the NAT routers are sent through the transport network, Fig. 6c shows load of the Double-NAT messages on the network, in terms of number of the MNs handed over to a new S/PGW. It also shows the load of messages of the GTP [7] and PMIP [9] protocols, during the MNs handover, given a SGW relocation, in the current

(a) Latency of the MNs' downlink data packets delivery before and after handover procedure.

(b) CDF of latency of the first downlink data packets redirected through the Double-NAT, and data packets forwarded via the X2 interface.

(c) Load of the MN's mobility support messages using different approaches in the centralized and decentralized LTE architectures.

(d) The MNs' downlink data packets loss ratio without using the X2-path traffic redirection.
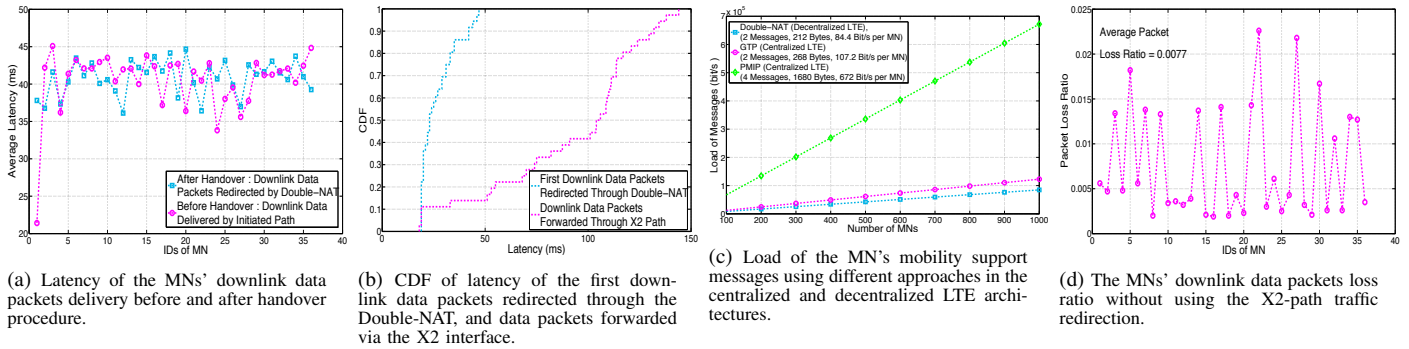
Fig. 6: Evaluation results of the Double-NAT traffic steering approach in a decentralized LTE network.

LTE network. It is clear that the Double-NAT outperforms the other mechanisms and has lower impact on the network load.

D. Packet Loss Ratio in the Downlink Data Traffic

The X2 path is used to forward the MNs' downlink data during the handover procedure, and also after when the Double-NAT traffic redirection path is not set up yet. By allowing the use of X2 path for 10 ms after the handover procedure (by setting *Delete Session Timer* = 10 ms in the MME [9]), no loss is presented in the MNs' data. It implies that 10 ms is enough to cope with the latency, needed to establish the Double-NAT traffic forwarding path. Mobile network operators may wish not to use the X2 path after the MNs' handover (setting *Delete Session Timer* = 0). In this case the moving MN's session is removed from the source S/PGW, upon the S1 bearer is initiated at the target S/PGW. This situation may results in some packet loss for the moving MNs (Fig. 6d).

VI. CONCLUSION

In this paper, we have developed a novel mechanism to support IP address continuity for the flow(s) kept active by the MNs performing handover with S/PGW relocation in a decentralized LTE architecture. Our solution is based on the NAT technique, which is a widely used procedure in the current Internet. Therefore, integration of the NAT functions into several network routers is easily feasible with a trivial overhead and complexity. Handling NAT tables on the NAT routers for steering the MNs' traffic above the EPC network can be performed using small size messages, with negligible impact on the network load. Besides this, deploying the NAT routers only at the edges of the transport network makes our solution to be deployable in large scale. Detailed simulation shows that the proposed solution is fast enough in setting up the traffic redirection path and readily meets the latency requirement, considering the maximum allowed delay threshold for real-time applications (*e.g.,* VoIP). Considering a scenario with the MME relocation as well as utilizing more than one NAT Controller within the network are other research areas of interest, as future work.

REFERENCES

[1] Cisco Visual Networking Index, Global Mobile Data Traffic Forecast Update, 2015−2020, February 2016.
[2] Dealing with Density: The Move to Small-Cell Architectures. In *RUCKUS Wireless White paper*, 2014.
[3] H. A. Chan, H. Yokota, J. Xie, P. Seite, and D. Liu. Distributed and Dynamic Mobility Management in Mobile Internet: Current Approaches and Issues. *JCM*, 2011.
[4] Z. F. L. Bokor and S. Imre. Flat architectures: Towards scalable future internet mobility. In *The Future Internet*. Springer, 2011.
[5] T. Taleb, K. Samdanis, and F. Filali. Towards supporting highly mobile nodes in decentralized mobile operator networks. In *IEEE ICC*, 2012.
[6] M. Karimzadeh, L. Valtulina, and G. Karagiannis. Applying SDN/OpenFlow in Virtualized LTE to support DMM. In *4th CLOSER*, 2014.
[7] 3GPP.TS.23.402. Universal Mobile Telecommunications System (UMTS); LTE. 2015.
[8] 3GPP SR 23.859. Technical Specification Group Services; LIPA Mobility and SIPTO at the Local Network, Release 11 and 12. 2012.
[9] 3GPP.TS.23.401. General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access. 2011.
[10] L. Valtulina, M. Karimzadeh, G. Karagiannis, G. Heijenk, and A. Pras. Performance evaluation of a SDN/OpenFlow-based DMM approach in virtualized LTE systems. In *IEEE CCSNA*, 2014.
[11] M. Liebsch. Per-Host Locators for Distributed Mobility Management. Internet draft, 2012.
[12] https://github.com/MortezaKarimzadeh/D-NAT_LTE.
[13] Rocketfuel: An ISP Topology Mapping Engine. http://research.cs.washington.edu/networking/rocketfuel/.
[14] Network Address and Port Translation (NAT) models. https://www.nsnam.org/wiki/GSOC2012Projects.
[15] Radhika Ranjan Roy. *Handbook of Mobile Ad Hoc Networks for Mobility Models*. Springer, 2010.
[16] S. P. Hoogendoorn. Vehicle-Type and LaneSpecific Free Speed Distributions on Motorways: A Novel Estimation Approach Using Censored Observations. In *Journal of the Transportation Research Board*, 2006.
[17] R. Irmer. Next Generation Mobile Networks Radio Access Performance Evaluation Methodology. In *NGMN Alliance White paper*, 2008.
[18] Y. Ganjali and N. McKeown. Update on Buffer Sizing in Internet Routers. *ACM SIGCOMM*, 2006.
[19] D. Ammar, T. Begin, and I.G. Lassous. A New Tool for Generating Realistic Internet Traffic in NS-3. In *ICST (SIMUTools)*, 2011.
[20] http://www.cisco.com/c/en/us/td/docs/ios/solutions_docs/qos_solutions/QoSVoIP/QoSVoIP.pdf.
[21] M. Karimzadeh, Z. Zhao, L. Hendriks, R. de O. Schmidt, S. la Fleur, H. van den Berg, A. Pras, and T. Braun. Mobility and Bandwidth Prediction as a Service in Virtualized LTE Systems. In *4th IEEE CloudNet*, 2015.