Abstract— In this paper a novel mobility management scheme is proposed that is based on optical packet switching. Optical Packet based Mobility Management (OPMM) is proved to be more efficient for mobile traffic transport compared to conventional GPRS Tunneling Protocol (GTP) regarding packet overhead, transport latency and energy consumption. Meanwhile, OPMM reduces the complexity of the mobility management procedure. The system design is validated in a simulator and the performance is compared with the standard 3GPP implementation in realistic scenarios with a reference architecture.

Keywords—Mobile Network; Mobility Management; Packet Optical Network; GTP; MPLS

I. INTRODUCTION

Recent trends of mobile communications show that smartphones are continuously penetrating into our daily life and video based services are getting more and more popular. Together with the emergence of new radio access technologies such as Coordinated MultiPoint transmission (CoMP), Cloud RAN (C-RAN) and small cells, the mobile traffic will continue growing rapidly. Furthermore, new services can have much more stringent requirements for future mobile networks. E.g., cloud computing requires real-time communication and synchronization with multiple network computation entities, tactile Internet [1] requires extremely short latencies of 1ms. This pushes mobile network operators to seek for new mobile network transport solutions with much higher capacity, lower latency and at the same time lower cost and energy consumption.

Packet optical networks are good candidates for a solution due to the high capacity and low energy consumption feature inherited from optical transport technology. The study in [2][4] shows lower layer switching consumes less energy compared to higher layer switching in the Open Systems Interconnection (OSI) model. Based on this, we propose an Optical Packet based Mobility Management (OPMM) scheme, which shifts the user data transport from IP layer to the optical layer. In OPMM, mobility management is performed on optical packets instead of IP packets. This means, there is no need to convert the complete optical packet into electric form, process/route it in the IP layer and convert it back to the optical packet before inserting it into the optical fiber. In turn, the energy consumption and processing delay can be reduced.

There are numerous state of the art concepts on reducing the complexity of mobility management in the mobile networks. For instance, [6][7] propose to replace the conventional cellular mobility management protocol-GPRS Tunneling Protocol (GTP) [3] by MPLS [5] based mobility management. [8] attempts to include the mobility management information in the IPv6 header and use routing instead of GTP tunneling to support user mobility. However, we regard our work as the first attempt to treat the mobility management directly in the optical layer. This makes it possible to achieve higher transport efficiency and lower energy consumption compared to existing approaches.

We have shown in [10] a packet switched optical mobile network architecture based on our OPMM concept and made some performance estimation via numerical simulations. This paper further verifies the design with an actual implementation of the OPMM protocol stack in a simulator and a detailed performance comparison with GTP in both data plane and control plane. The simulations are performed based on a reference architecture with parameters from actual optical devices and they prove that our design can greatly increase the transport efficiency of mobile traffic and is able to satisfy the requirements of future mobile communications. Moreover, the deployment considerations for OPMM are discussed.

Section II describes the existing 3GPP mobility management protocol – GTP. Section III explains the detailed OPMM design and its QoS treatment in particular. Section IV describes the reference scenario and shows a numerical analysis of the gain of OPMM. The system design and performance is validated and evaluated in OMNeT simulation environment in Section V. Finally, Section VI concludes the paper.

II. GTP BASED MOBILITY MANAGEMENT

In addition to the general requirements of fixed networks, such as Operation Administration and Maintenance (OAM), network resilience, and QoS guarantees, mobile networks must
also support User Equipment (UE) mobility. When the UE moves from one base station to another, the user data flow needs to be switched to the new base station as well.

In the 3GPP standard, user data is carried by EPS bearers (data paths with a certain QoS guarantee) throughout the mobile network. The user data packets are mapped to EPS bearers at the UE for UpLink (UL) traffic and at the PDN Gateway (PGW) for DownLink (DL) traffic depending on the Traffic Flow Template (TFT). In the 3GPP standard, an EPS bearer is realized by GTP in the mobile backhaul. When a UE connects to one eNB (Evolved NodeB), GTP tunnels will be established between the eNB and SGW (S1 GTP tunnel), and between the SGW and PGW (S5 GTP tunnel). When the UE moves, new S1/S5 GTP tunnels will be established and the user data flow will be switched to the new GTP tunnels. The SGW checks the identifier of GTP tunnel (i.e., Tunnel Endpoint IDentifier (TEID)) carried by the GTP-U packet header to correctly interconnect the S1 and S5 GTP tunnels. Since a GTP packet is carried by UDP/IP where the IP destination address indicates the correspondent eNB, SGW or PGW node, the SGW always needs to process each packet electrically up to the GTP layer. This puts high processing load at the SGWs for user data transport. OPMM aims to reduce the efforts for such electrical packet processing and respective optical to electrical signal conversion.

III. OPTICAL PACKET BASED MOBILITY MANAGEMENT

A. Concept

As mentioned in the previous section, the transport cost, energy consumption and complexity decrease with the level of processing layers. The basic concept of OPMM is to replace GTP tunnels by optical tunnels.

In OPMM, the conventional mobile network node (i.e., eNB, S/PGW, etc.) works together with the optical transport layer underneath, e.g., Optical Packet Router (OPR), as one combined node. We call such nodes eNB_O, S/PGW_O and MME_O. The bearers are mapped into Optical Tunnels (OPTs) instead of GTP tunnels. Fig. 1 shows that the bearers are mapped into S1 OPTs at the eNB_O and S5 OPTs at the SGW_O for UL traffic. And for the DL traffic, the bearers are mapped into S5 OPT and at the PGW_O and S1 bearers at the SGW_O.

Each logical interface can have multiple OPTs defined for different QoS guarantees and network resiliencies. In the usual case, the user data flow goes through two OPTs in the mobile backhaul, i.e., S1 OPT and S5 OPT. For special cases such as the forwarding phase of handover, the user data flow may go through more OPTs, e.g., X2 OPT.

One OPT is shared by multiple bearers as far as the QoS requirements of each of these bearers can be satisfied. When UE moves, only the mapping tables of the bearers to the OPTs will be modified at the eNB_O, SGW_O and PGW_O nodes according to the instructions from the MME_O.

OPMM inherits the 3GPP bearer concept, and the bearer level QoS enforcement at the eNB_O and PGW_O stay the same as before.

Fig. 2 shows the definition of the optical packet format to realize the OPMM. The new optical control header consists of three fields: optical tunnel ID, QoS and Bearer IDentity (BID). Where the optical tunnel ID and QoS defines the treatment of the optical packet at each optical switches/routers, and the BID is only used at the eNB_O, S/PGW_O for bearer level treatment, such as bearer level policy control/enforcement and also mobility management. Here, we define the BID to be 32 bits long, where the most significant 28 bits identify the mobile user and the least significant 4 bits represent EPS bearer ID from 3GPP standard. The QoS value is derived from the QoS Class Identifier (QCI) parameter of the EPS bearer.

When a DL packet arrives at the PGW_O, it is mapped to a bearer according to the TFT as in the GTP case. This bearer is identified by the BID value which is added to the control header of this optical packet together. According to the BID, the PGW_O searches for the corresponding SGW_O address information and QoS information in its local data base which is provided by the MME_O. The PGW_O identifies the corresponding OPT to that SGW_O and adds the optical tunnel ID and QoS parameter into the control header of this optical packet. This optical packet is injected into the optical tunnel to the SGW_O. The SGW_O checks the BID of this optical packet and finds the corresponding eNB_O address and QoS parameter for this optical packet in its local data base. Similarly, SGW_O maps the eNB_O address and QoS parameter into a new OPT and replaces the old optical tunnel ID and QoS fields with the new one in the control header. The optical packet travels in the new OPT to the eNB_O and is finally mapped to the corresponding radio bearer according to the carried BID. Vice versa the procedure works for the UL packet.

B. Implementation

One possibility to realize OPTs is using MPLS tunnels. In this case, the Optical tunnel ID will be implemented as MPLS label. According to the network location of eNB_Os, S/PGW_Os, default MPLS tunnels are pre-configured using Label Distribution Protocol (LDP) or Resource reSerVation Protocol for Traffic Engineering (RSVP-TE) for each logical interface between these nodes. When a UE connects to the mobile network, the default MPLS tunnel will be used to transport the user data flow. eNB_O and S/PGW_O are
responsible to monitor the used bandwidth of the preconfigured MPLS tunnels. When the bandwidth usage exceeds certain limits (e.g., 90%), a new MPLS tunnel will be configured for the corresponding logical interface. For each optical packet that passes by, the SGW_O checks the BID in the optical packet header and adds a new MPLS label accordingly.

C. Benefits

In OPMM, the SGW_O only needs to check the control header of the optical packet to decide on the forwarding behavior. This avoids the unnecessary optical-to-electrical conversion of the payload of an optical packet. At the same time, since GTP is actually an IP-in-ip tunnel, OPMM will reduce the packet overhead of 28/48 bytes for IPv4/IPv6 (i.e., IP + UDP + GTP – 2 × MPLS header) by replacing it with lower layer tunnel. This overhead reduction also results in the decrease of the IP/UDP/GTP header processing load, latency, and energy consumption.

Besides data plane savings, the control plane complexity is also greatly reduced. As mentioned before, OPMM decouples OPT setup from individual user mobility. Instead of always establishing new GTP tunnels for individual UE mobility, a new OPT will be setup only in the case that the current allocated resource is not enough to accommodate the foreseen aggregated mobile traffic between the two mobile network nodes. Furthermore, GTP needs to exchange the TEID information between the two GTP tunnel end nodes. OPMM only needs to understand the mapping of BID to OPT, which is a node local issue and does not need to be communicated to other nodes. This further simplifies the tunnel setup signaling. Most important, the constraint for new tunnel setup time is relieved due to the decoupling to individual user mobility.

D. Deployment consideration

For a real deployment of OPMM, it is important to have an estimation of needed system capability, which is the switching capability of the intermediate packet optical routers/switchers and the combined packet optical routers/switchers, i.e., eNB_Os, and S/PGW_Os. Since OPMM is based on MPLS switching, the label length and mapping table size at the eNB_Os and S/PGW_Os are the key parameters to estimate the cost of the total system.

Using Japan as an example, we assume 32.9 Tbps total mobile traffic in 2020, 2 PGW_Os, 1000 SGW_Os, 1000000 eNB_Os distributed all over Japan and MPLS based OPTs. We considered a two-level hierarchical architecture as our reference mobile network architecture [10] (Fig. 3). A mesh structure is used in the core network level to interconnect the optical core network elements SGW_Os and PGW_Os (combined nodes of S/PGW and Hybrid Optoelectronic Router-HOPR [14]). And a ring structure is used in the access/metro network level to interconnect the eNB_Os (combined nodes of eNB and Packet-optical Add/Drop Multiplexer-POADM[11]).

1) Capacity and label space

All together 32.9 Gbps should be allocated for S5 MPLS tunnels and 3–6 Gbps should be offered by each S1 MPLS tunnel. The X2 MPLS tunnels should offer at least 0.58 Gbps capacity, which includes both the forwarding traffic during handover and the CoMP traffic for a 3 cell inter-site cooperation case.

The Next Generation Mobile Network (NGMN) envisions typically up to 6 operators and 16 S1 interfaces per operator per eNB, 16,000 S1 interfaces per operator per SGW, and 32 X2 interfaces per eNB [9]. Assuming 10 Gbps per MPLS tunnel, 100 SGW_Os per SGW_O pool, each SGW_O connects to 10 PGW_O for network redundancy, the rough number of needed preconfigured MPLS tunnels for one operator can be calculated:

- 1000×33=33000 at both PGW_Os
- 16000+10+99=16109 at each SGW_O
- 32×16=48 at each eNB_O.

Therefore the needed label length should be at least 16 bits at the PGW_O.

2) Size of the mapping table

The BID is mapped to an MPLS label at the eNB_O and S/PGW_O. The number of entries and the length of each entry in the mapping table decide on the needed memory as well as the processing complexity for table look up, which will in turn affect the energy consumption of the correspondent node. Assuming 60 million subscribers for one mobile operator, with up to 11 bearers each will result in 660 million entries at both PGW_Os and 660,000 entries at each SGW_O. An eNB is able to support 250 users per sector currently, which means 3000 entries for 6 sectors’ eNB. 30 bits is needed to indicate the BID for each entry in the mapping tables.

3) Performance estimation

The simulations in [10] show that such a system will provide at most 9.32 ms DL transport delay in 37 eNBs per ring case, which is able to meet the 2–15 ms transport delay specified by 3GPP[17]. The respective X2 delay will be 3.2 ms in the worst case which fits into the X2 interface latency requirements for both data forwarding (10–20 ms) and CoMP (1–5 ms) [10].

SGW_O based on optical transport router (e.g., HOPR[14]) can save up to 97% of power consumption compared to SGW based on electrical router (e.g., Cisco CRS-1 [15]) with equivalent performance.

IV. SYSTEM VALIDATION AND EVALUATION

The proposed optical mobility management protocol is implemented in an OMNest simulation environment and is compared with GTP based mobility management protocols. The implementation includes the reference architecture as shown in Fig. 3. One UE is first attached to the source eNB and then moves to the target eNB. The GTP-U/C packet is carried over UDP/IP and SIAP/X2AP is carried by SCTP/IP.

Fig. 3. Reference optical mobile network architecture
Both the U-plane performance and C-plane performance are compared. Below are the simulation parameters and assumptions based on the deployment scenario described in section III-D:

- SGW-MME distance: 0-205 km
- SGW-PGW distance: 0-1000 km
- Number of eNBs in the ring: 7, 19, 37
- Distance between neighboring eNBs: 0.8-4.6 km
- IP processing delay: 0.1ms
- POADM queuing delay: 2 ms
- POADM pro node processing delay: 100ns
- HOPR processing delay: 1.4 µs
- No queuing delay for the HOPR nodes
- Bandwidth of the data channel: 10Gbps
- Packet arrive interval: 0.5 ms
- RRC reconfiguration time during HO: 12.5 ms [16]

A. C-Plane saving

The savings in the control plane are coming from the fact that the TEIDs distribution becomes unnecessary in the OPMM, since the tunnels are identified via the BIDs and IP addresses.

For the GTPv2-C protocol, it means that the F-TEID Information Element (IE) can be reduced to carry IP address only, thus saving 4 bytes (length of the TEID IE) per message (see Fig. 4). Similarly, the TEIDs do not have to be sent by the S1AP [12], and X2AP [13], thus reducing the message by the size of the TEID IE. Generally, this size will be implementation specific, but for the simulation the estimated size of 4 bytes was used.

The comparison for the signaling overhead of both architectures for Initial Attachment (IA), X2 based HandOver (X2HO) and S1 based HandOver (S1HO) are shown in the Fig. 5. The size of the involved messages was estimated from the corresponding 3GPP technical specifications ([12][13][14]), though it may vary depending on the implementation. For all control plane procedures, there is a small signaling overhead reduction in OPMM case.

Another aspect of saving is specific for the Attach Procedure. Since the steps 11 and 13 (Fig. 4) are only used for the TEID distribution in the GTP-based MM, they can be omitted in case of OPMM. However, in order to let the SGW know the eNB, there should be an additional IE added to the Create-Session-Request as shown in Fig. 4.

![Fig. 4. Attach Procedure](image)

![Fig. 5. Comparison of signalling overhead](image)
reduction in the OPMM case, comes from the propagation delay between MME and SGW. In the best case scenario (B), the IP processing delay reduction becomes more important. For the intermediate case (I), the major difference between GTP and OPMM comes also from the propagation delay between MME_O and SGW_O. In all the cases, the optical delay difference of the GTP and OPMM-based approaches is almost imperceptible.

B. U-Plane saving

Since OPMM remove the necessity of the GTP tunnel, the GTP/UDP/IP overhead is reduced. Assuming the payload size is between 20 bytes (with VoIP G.711 codec) and 1396 bytes (largest payload of UDP packet in GTP case), the transport efficiency can be increased by 13% as shown in TABLE I. This will in turn reduce the energy consumption of the system for packet header processing and increase the network capacity.

<table>
<thead>
<tr>
<th>Payload size (bytes)</th>
<th>GTP IPv4</th>
<th>GTP IPv6</th>
<th>OPMM IPv4</th>
<th>OPMM IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>160 (VoIP G.71)</td>
<td>66.7%</td>
<td>57.1%</td>
<td>76.9%</td>
<td>70.1%</td>
</tr>
<tr>
<td>20 (VoIP G.72)</td>
<td>20.0%</td>
<td>14.3%</td>
<td>29.4%</td>
<td>22.7%</td>
</tr>
<tr>
<td>1396 (UDP)</td>
<td>95.4%</td>
<td>92.8%</td>
<td>97.5%</td>
<td>96.1%</td>
</tr>
</tbody>
</table>

Except for the network capacity saving, OPMM also reduces the U-Plane transport latency. DL/UL packets are processed at the SGWs in the optical layer instead of IP layer. This allows OPMM to save 2 times IP processing delay (one for incoming IP packet and the other for outgoing IP packet) for all the DL/UL packets. This effect is more obvious during the handover procedure. The procedure itself does not differ from the state-of-the-art workflow, except for the IEs change, as it was described above. However, due to the usage of OPTs, source eNB is able to bind the S1 OPT directly to the X2 OPT without IP/UDP/GTP processing and most important avoid the queuing from the IP layer to the optical layer. This will greatly reduce the forwarding delay of the X2 U-plane traffic.

Fig. 7 and Fig. 8 illustrate the DL and UL packet lifetime during the X2 based handover procedure. The packet life time is measured at the eNB, when the packet is forwarded to the radio layer for the DL case, and at the PGW, when the packet is extracted from the tunnel for the UL case. OPMM gives slightly smaller latency for both DL and UL case compared to GTP case (i.e., 0.2ms). The total U-plane DL transport latency is 5–7 ms. The UL transport latency is 2 ms more compared to the DL case due to the POADM queuing delay at the eNB.

As shown in Fig. 7, there are two types of forwarded packets. Type 1 are the packets which were stored in the buffer of the eNB when the handover command arrived at eNB. These packets were already sent to the UE, but were not acknowledged yet. Proceeding from the fact that the round-trip delay for the radio link is 10 ms ([16]), all the packets arrived 10 ms before the HO will belong to this category. The delays for these packets will only slightly differ in the OPMM case.

![Fig. 7. Packet lifetime during X2 based handover (DL)](image)

![Fig. 8. Packet lifetime during X2 based handover (UL)](image)
Type 2 are the packets, which arrived after the handover has started. In the OPMM case, those packets will be processed on the MPLS layer, skipping the IP/UDP/GTP processing, and, more important, avoiding the POADM queuing delay. The resulting difference will be significant, as it can be easily deduced from the Fig. 7. The reduced forwarding delay (by around 70% for X2 forwarding and by around 50% for S1 forwarding) results in lower end-to-end latency during handover procedure (by around 25%). At the same time, due to the fast packet forwarding path, the possibility of packets reordering during the handover procedure is greatly reduced (by 67% as shown in Fig. 7). All these facts help on improving the user experience during handover procedure (less jitter, shorter service interruption).

In general, S1 based handover affects the forwarded packets in the similar fashion as the X2-based handover. The packets forwarded from the source eNB to the target eNB are processed on the SGWs in additional to the eNBs. For the OPMM case, the SGWs only process the packets on the optical layer, and for the GTP case additional IP/UDP/GTP processing is required. Therefore, the absolute value for the delay reduction for the OPMM architecture will be larger compared to the X2 based handover.

V. CONCLUSION

This paper describes the concept and working principle of a novel mobility management protocol based on optical packet switching – OPMM. The system design is validated based on an OMNest simulation environment in a realistic scenario. The simulation results show that OPMM increases the transport efficiency of mobile traffic by up to 13%, and reduces the packet latency during handover procedure by around 2 ms compared to GTP. Moreover, OPMM greatly reduces the packet processing load and energy consumption in mobile networks thanks to optical routing/switching.

REFERENCES

[12] 3GPP TS 36.413 “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (SIAP)”