Optical Slot Switching Latency in Mobile Backhaul Networks

N. Benzaoui, Y. Pointurier, T. Bonald, Q. Wei and M. Lott

***Abstract*—Next generation mobile networks such as LTE-Advanced will offer high data rates to users utilizing new radio access technologies. In addition, new applications as well as devices such as smartphones and tablets will result in a rapid mobile traffic growth. Consequently, a rapid increase of the load in the mobile backhaul segment is expected, resulting in turn into a substantial increase of the energy consumption. Moreover, by allowing the possibility of direct communications between base-stations for instance to support interference cancellation mechanisms such as "CoMP" and expedite handovers, the traffic pattern turns into any-to-any node communication in the next generation mobile networks. In previous papers we proposed an optical mobile backhaul network to transport the mobile user data in an energy-efficient fashion, while supporting any-to-any traffic exchanges. In this paper we evaluate, analytically and by simulation, the latency in such next generation optical mobile backhaul networks for realistic scenarios. We also study the impact of the support of several classes of service.**

***Index Terms*—Optical packet switching, network performance, queue modelling, CoS, LTE-A mobile backhaul**

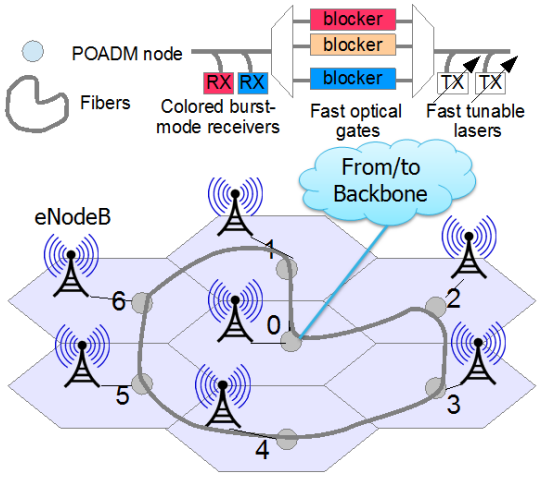


Fig. 1. POADM-based mobile backhaul ring.

I. Introduction

With the evolution of the optical transport technologies, energy-efficient solutions for the transport in the radio access part of cellular backhaul networks have become available. In previous work the use of a fine granularity optical transport solution, optical slot switching (OSS) was proposed as an energy-efficient switching technology [1] to interconnect the Evolved Node Bs (eNodeBs) in a Long Term Evolution Advanced (LTE-A) mobile backhaul network. LTE-A leverages direct node-to-node communication [2] through X2 interfaces [3] to transport Coordinated MultiPoint (CoMP) traffic [4], which improves coverage, cell-edge throughput, and system efficiency.

Packet Optical Add/Drop Multiplexer (POADM [5]; see top of Fig. 1) is an optical slot switching technology that natively supports any-to-any communication and is hence particularly suitable to the LTE-A mobile backhaul application. In this paper, optical slot switching is applied for metropolitan mobile backhaul networks with a few (up to a few dozens) of eNodeBs; Fig. 1 shows LTE eNodeBs interconnected by a mobile backhaul network [2].

POADM is a Wavelength-Division Multiplexed (WDM) time-slotted ring that provides sub-wavelength switching granularity. A POADM ring consists of several (e.g., 40) data channels, and of one additional "control channel" which carries all slot headers and is (electronically) processed at each node. Each node is connected to one or several clients (e.g., a Passive Optical Network (PON), an Ethernet switch, etc.). Data coming from/going to the clients is encapsulated/decapsulated within fixed-size packets called slots (of few µs) that can transit transparently, i.e. they are not converted to the electronic domain at the intermediate nodes [5]. Each node decides whether to drop or let transit each slot on each wavelength according to information carried by the control channel. This optical transparency for the transit traffic is a key feature that helps to minimize the power consumption. Indeed [6] shows that POADM rings can be up to 5 times more energy-efficient than Ethernet rings and 30% than Reconfigurable-OADM rings.

In [3], the authors show that the latency constraints of LTE - A can be very stringent in order to support sophisticated CoMP scheme (ideally in the scale of 1-5 ms). Future mobile networks even need to satisfy 1 ms end-to-end delay requirement for "tactile" type of services. All of these requirements have to be matched for the backhaul network. Many of today's backhaul networks are still implemented using legacy Time Division Multiplexing (TDM) technology based on electronic circuit switching. Due to the capacity restrictions of such systems and to the surge of packet-based mobile Internet traffic at the expense of voice traffic, new electronic, packet (typically, Ethernet-based) networks were proposed around 2010 [7]. In order to make such Ethernet networks QoS-aware and provide latency guarantees, the Metro Ethernet Forum defined "services" in the context of Carrier Ethernet, an extension to Ethernet. Although Carrier Ethernet does address the latency

N. Benzaoui is with Alcatel-Lucent Bell Labs, Nozay, France, Telecom ParisTech, Paris, France, and LINCS, Paris, France.
Y. Pointurier are with Alcatel-Lucent Bell Labs, Nozay, France.
T. Bonald is with Telecom ParisTech and LINCS, Paris, France.
Q. Wei and M. Lott are with DOCOMO Euro-Labs, Munich, Germany.

TABLE I
TRAFFIC CHARACTERISTICS [11], [12]

Type	Demand	Pkt size (Bytes)	Pattern	CoS
Data	1.5 Gb/s up, 3 Gb/s down	1500	Centralized	RT, NRT
Voice over IP	8.5 Mb/s up and down	162	Centralized	RT
CoMP data	1 Gb/s	1500	Peer to peer	NRT
CoMP signaling	774 Kb/s	92	Peer to peer	RT
Handover inter-ring	0.1 Gb/s up and down	1500	Centralized	NRT
Handover intra-ring	0.1 Gb/s	1500	Peer to peer	NRT

constraint in mobile backhaul networks, it does not address the high energy consumption issue that is associated with such electronic packet switching-based networks. More recently, Passive Optical Network (PON)-based mobile networks were proposed to enable the re-use of existing, high-capacity fixed access network infrastructure and hence decrease the cost of deploying next-generation mobile backhaul networks [8]. In [9] it is shown that latency in mobile backhaul PONs can be reduced down to 1 ms, making it a relevant LTE-A mobile backhaul network candidate. However, PON systems are based on a tree structure where extremely short latency for X2 communication (between the optical network units) is difficult to achieve. As explained above POADM is an energy-efficient network concept that enables direct basestation-to-basestation X2 communication. For this reason, we study POADM as a potential candidate for next-generation mobile backhaul networks. Compared with Ethernet and PON, POADM is a relatively novel concept relying on advanced optical components, with the possible associated cost premium. This paper deals with performance only.

The latency of POADM-based networks was studied in [10] for general-purpose, single class-of-service networks. In this work we study the performance (latency) of a POADM ring for the specific case of the LTE-A mobile backhauling application. We propose for the first time an analytical model to quantify the latency of an OSS-based LTE-A backhaul network using real traffic demand with variable client packet sizes and handling different classes of service (CoS). We investigate the impact on end-to-end latency of a CoS management policy affecting the slot formation and insertion processes. We compare analytical results with simulations for small scale networks and we extrapolate the analytical model to larger scale networks to avoid time-consuming simulations. In particular, we show that the overall latency is compatible with the LTE-A requirements.

This paper is organized as follows. In Section II we explain the different policies for class of services management. We propose in Section III an analytical model to evaluate the per-class of service latency in an OSS network. In Section IV we evaluate the performance of a POADM ring for an LTE-A mobile backhauling application, and using the same scenario we compare analytical and simulation results. We give in Section V the most important conclusions of the performance studies.

II. SUPPORT OF DIFFERENTIATED CLASSES OF SERVICE

Typical traffic in an LTE-A mobile backhaul network is described in Tab. I. In this work we consider 2 classes of service: real-time (RT) and non-real-time (NRT), however the analytical study can be generalized to any number of CoS. RT traffic is latency-sensitive and consists of a part of the data traffic, Voice over IP (VoIP), and signaling for CoMP, while NRT (or best effort) traffic consists of the remainder of the data traffic, data for the CoMP mechanism, and handover (HO) data and signaling. The strictest latency requirement comes from CoMP signaling, which should experience less than 1 ms end-to-end latency [12]. Given that such traffic may traverse a small distance over the backbone network to ensure communication between two eNodeBs in two neighbor metropolitan mobile backhaul networks (here, rings), the per-ring maximum latency for RT traffic is thus at most 500 µs. Observe that, thanks to transparency, the end-to-end latency can easily be calculated from the queuing delay by adding the propagation delay, which is deterministic for a given physical topology, and negligible in metropolitan networks. For this reason, we focus on the packet queuing delay analysis (Section II-A). This paper also quantifies the impact on traffic latency of CoS management during slot formation and slot insertion on channel (Section II-B).

A. Slot formation without CoS management

We first describe how slots are formed in a non-CoS managed network. Without CoS management, all client packets may be mixed in the same optical slot, as described in Fig. 2. At a POADM (eNodeB) node, packets arriving from clients (e.g. mobile terminals) are placed in a temporary packet queue identified by the destination (eNodeB) node. Once enough packets to fill a slot have arrived in the temporary packet queue, this queue is emptied and the optical slot is formed and placed in the optical slot queue, waiting for its insertion on the channel. To cap the optical slot formation time, we use a timer, which is triggered by the arrival of the first client packet, and which expiration causes the completion of the formation of the optical slot. Without CoS management, since there is no differentiation between the different classes of service, a unique timer value is used during slots formation. This global timer value is selected so as to respect the strictest traffic latency constraint, here, the CoMP latency constraint.

B. Slot formation with CoS management

As described in Fig. 3, temporary queues are now identified not only by destination node but also per CoS, such that each optical slot contains packets with the same CoS. The utilization of per-CoS queues enables sending RT traffic before NRT traffic. CoS differentiation is performed only during slot assembly (through the utilization of separate queues for the various classes of service) and channel insertion (through priotization of the various classes) and never at intermediate nodes; with transparency, all slots are treated equivalently by the transport layer, irrelevant of their CoS. With CoS management it is possible to differentiate between client packets of both classes

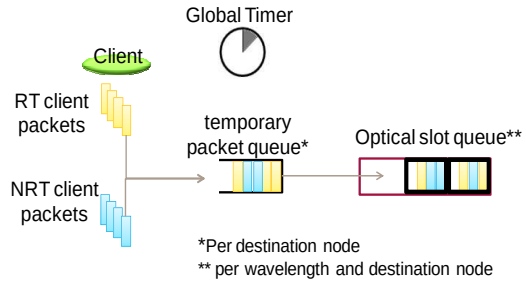


Fig. 2. Slot formation without CoS management.

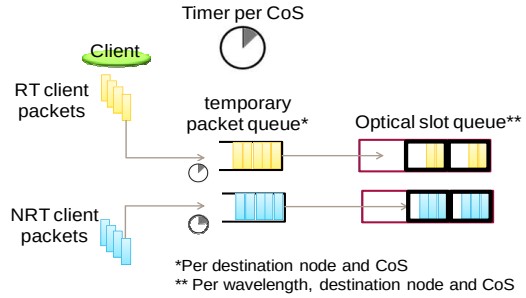


Fig. 3. Slot formation with CoS management.

of service at the slot assembly step thus we apply one timer per CoS: a short timer for RT traffic (to respect latency restriction) and a larger timer for the NRT traffic (since it is less sensitive to latency). At the channel insertion, optical slots containing RT packets are served first.

Fig. 4 represents the electronic architecture of a POADM node with the different steps followed by the client packets during the transmission, transit and reception process. For clarity, we mapped on Fig. 4 the temporary and optical slot queues mentioned in previous sections II-A and II-B.

III. Analysis

In this section we propose an analytical model to evaluate the per-class of service latency in an OSS network. The packet queuing delay is the waiting time between the sending of the packet by the source client and the insertion of the slot containing this packet on the channel. The queuing delay results from the optical slot formation time *OSFT*, time due to the slot assembly process happening in the temporary packet queue, and the buffering time *OBT*, time spent in the optical slot queues, as shown in Figs. 2 and 3. The mean queuing delay *QD* experienced by a client packet is thus:

$$QD = OSFD + OBD, \tag{1}$$

where *OSFD* is the mean delay experienced by a client packet during the optical slot formation, and *OBD* is the mean delay experienced by a client packet during the optical slot buffering.

A flow corresponds to the traffic sent from a given source node to another given destination node, and to the type of this traffic as defined in Tab. I. Hence, all packets with the same triplet (source node, destination node, traffic type) are

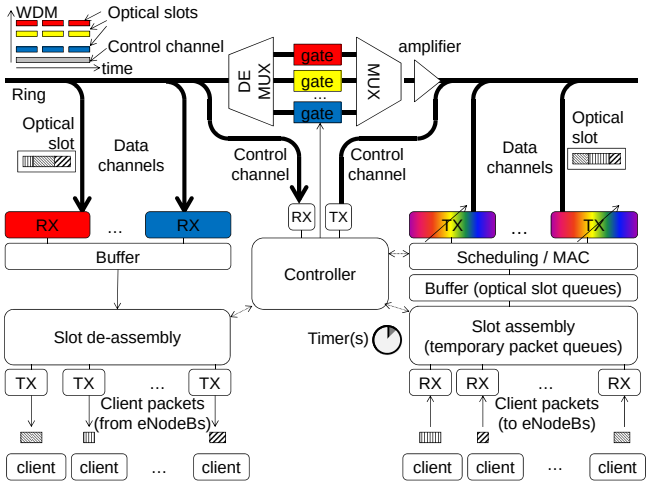


Fig. 4. Electronic POADM node architecture.

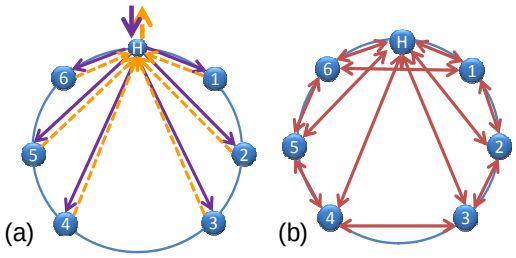


Fig. 5. Traffic patterns: (a) centralized; (b) peer to peer.

considered to be from the same flow, irrespective of the end-user application. Our analytical model is based on client-to-client traffic with Poisson packet arrivals, where the mean packet arrival rate for flow i is λ_i. The optical slots service rate is μ, and is deterministic and equal to the inverse of the optical slot duration. A slot is formed and inserted on the medium either because (a) it is (almost) filled or (b) the timer has expired. We now formalize (a). Since client packets of a same flow may have different sizes, optical slots may not be fully filled even with a very long timer. An optimal slot formation condition must take in account the trade-off between the slot filling ratio and the latency[1]. Thus, we propose to define the slot formation condition as the time when, after the arrival of a client packet, the remaining free room is smaller than the more frequent client packet size. This of course requires a certain knowledge of the distribution of the client packet sizes, which is shown for instance in Tab. I for the LTE-A application; there, the most frequent client packets are 1500 bytes long, which is also the maximum packet size. Hence, we formulate the condition for a slot formation as:

$$S - s_{\max} < V_t \leq S, \tag{2}$$

[1]A naive approach would be to define the slot formation condition as the moment when there is no room left for any incoming client packet. This solution results in a higher slot filling ratio than the one proposed in this paper but also in additional latency. Indeed a slot may wait for a small client packet to improve its filling ratio; if instead a client packet that is too large to fit in the slot arrives, the slot now departs with increased latency and no improved filling ratio.

where S is the optical slot size (in bit), $s_{\max}$ is the maximum client packet size, and V_t is the volume of data accumulated in the slot during a time interval t:

$$V_t = \sum_{i=1}^{N} x_i \cdot s_i, \tag{3}$$

where N is the number of flows departing a considered node and filling a same slot, i is a flow index, s_i is the size of a packet for flow i; and x_i is the number of packets of flow i contained in an optical slot after a time interval t. In the following we show how to compute *QD* by first deriving *OSFD* and then *OBD*.

A. Optical Slot Filling Delay derivation

As explained above the optical slot formation time *OSFT* will be either equal to the time t of the slot formation, under the condition (2), or to the timer value T. This implies that the *OSFT* depends on the combination of the types of client packets within the slot. In case (a) (i.e. before timer expiration), before the arrival of the last packet that triggers the sending of the slot, V_{t^-} (t^- the time before the last packet arrival) should be below or equal to $S - s_{\max}$. The probability of the slot formation depends then on V_{t^-} and the type (size) of the last arriving packet. A slot will be formed after the arrival of a client packet of a type j if and only if $S - s_{\max} - s_j < V_{t^-} \leq S - s_{\max}$. We take this into account by introducing the arrival probability of a client packet during a short time interval $\mathrm{d}t$: $\lambda_j \,\mathrm{d}t$. Hence *OSFT* is expressed as:

$$\begin{aligned} OSFT \;=\; & \int_0^T \sum_{j=1}^{N} P(S - s_{\max} - s_j < V_{t^-} \leq S - s_{\max}) \lambda_j t \,\mathrm{d}t \\ & + TP(0 \leq V_T \leq S - s_{\max}), \end{aligned} \tag{4}$$

where, for $\mathbf{x} = (x_1, \cdots, x_N)$,

$$P(S - s_{\max} - s_j < V_{t^-} \leq S - s_{\max}) = \sum_{\mathbf{x} \in A_j} P_t(\mathbf{x}), \tag{5}$$

$$P(0 \leq V_T \leq S - s_{\max}) = \sum_{\mathbf{x} \in B} P_T(\mathbf{x}), \tag{6}$$

where $P_t(\mathbf{x})$ and $P_T(\mathbf{x})$ are the probability of $\mathbf{x}$ packet arrivals, corresponding to flows $\{1, \ldots, N\}$, during a time interval t or during a time interval T, respectively. $\mathbf{A_j}$ is a set of $\mathbf{x}$ such that:

$$S - s_{\max} - s_j < V_{t^-} \leq S - s_{\max}, \tag{7}$$

and $\mathbf{B}$ is a set of $\mathbf{x}$ such that:

$$0 \leq V_T \leq S - s_{\max}. \tag{8}$$

Note that the slot formation process, and timer, are triggered by the arrival of the first packet, hence $t = 0$ at the first packet arrival. Consequently, we are looking for the probability of having $K - 1$ arrivals during t. We recall that, for a Poisson packet arrivals, the probability of having $K - 1$ packet arrivals during a time interval t is:

$$P_t(X = K - 1) = (\lambda t)^{(K-1)} \frac{e^{-\lambda t}}{(K-1)!} = \frac{K}{\lambda t} P_t(X = K). \tag{9}$$

Using (5), (6) and (9) in (4), *OSFT* is:

$$\begin{aligned} OSFT = & \int_0^T \sum_{j=1}^{N} \sum_{\mathbf{x} \in \mathbf{A_j}} \frac{\sum_{i=1}^{N} x_i}{\lambda t} \lambda_j P_t(\mathbf{x}) t \,\mathrm{d}t \\ & + T \sum_{\mathbf{x} \in \mathbf{B}} \frac{\sum_{i=1}^{N} x_i}{\lambda T} P_T(\mathbf{x}). \end{aligned} \tag{10}$$

We recall that:

$$P_t(\mathbf{x}) = \prod_{i=1}^{N} (\lambda_i t)^{x_i} \cdot \frac{e^{-\lambda_i t}}{x_i!}, \tag{11}$$

hence, injecting (11) in (10):

$$\begin{aligned} OSFT = & \int_0^T \sum_{j=1}^{N} \sum_{\mathbf{x} \in \mathbf{A_j}} \frac{\sum_{i=1}^{N} x_i}{\lambda t} \prod_{i=1}^{N} \left((\lambda_i t)^{x_i} \cdot \frac{e^{-\lambda_i t}}{x_i!} \right) \lambda_j t \,\mathrm{d}t \\ & + T \sum_{\mathbf{x} \in \mathbf{B}} \frac{\sum_{i=1}^{N} x_i}{\lambda T} \prod_{i=1}^{N} \left((\lambda_i T)^{x_i} \cdot \frac{e^{-\lambda_i T}}{x_i!} \right). \end{aligned} \tag{12}$$

We define $\lambda = \sum_{i=1}^{N} \lambda_i$, $X = \sum_{i=1}^{N} x_i$ and $Y = \prod_{i=1}^{N} x_i!$. Integrating by parts *OSFT* is developed into:

$$\begin{aligned} OSFT = & \sum_{j=1}^{N} \sum_{\mathbf{x} \in \mathbf{A_j}} \frac{\lambda_j X}{\lambda} \frac{X!}{Y} \left(\prod_{i=1}^{N} \lambda_i^{x_i} \right) \left(\frac{I_0}{\lambda^X} + \ldots \right. \\ & \left. \sum_{n=1}^{X} \frac{a_n}{n! \lambda^{X-n}} \right) + \sum_{\mathbf{x} \in \mathbf{B}} \frac{X}{\lambda} e^{-\lambda T} T^X \frac{\prod_{i=1}^{N} \lambda_i^{x_i}}{Y}, \end{aligned} \tag{13}$$

where $a_n = -\frac{T^n}{\lambda}$ and $I_0 = -\frac{e^{-\lambda T} + 1}{\lambda}$.

Knowing that during a time interval t the client packet arrivals are uniformly distributed [13], we consider the mean inter-arrival time equal to t/K (the interval time divided by the number of packet arrivals). Consequently, the first client packet that has arrived waits during time $(t/K) \cdot K$, the second packet during $(t/K) \cdot (K - 1)$, and so on. The total waiting time T_w of the client packets in a slot is therefore:

$$T_w = \frac{t}{K}(K + (K-1) + \ldots + 1) = t \cdot \frac{K+1}{2}. \tag{14}$$

Since *OSFD* is the average time waited by the client packets during the optical slot filling process, it depends on the client packet composition in the slot:

$$\begin{aligned} OSFD = & \sum_{j=1}^{N} \sum_{\mathbf{x} \in \mathbf{A_j}} \frac{\lambda_j X}{\lambda} \left(\prod_{i=1}^{N} \lambda_i^{x_i} \right) \frac{X!}{Y} \sum_{n=1}^{X} \frac{a_n}{n! \lambda^{X-n}} \frac{(X+1)}{2\overline{X}} \\ & + \sum_{j=1}^{N} \sum_{\mathbf{x} \in \mathbf{A_j}} \frac{\lambda_j X}{\lambda} \left(\prod_{i=1}^{N} \lambda_i^{x_i} \right) \frac{X!}{Y} \frac{I_0}{\lambda^X} \frac{(X+1)}{2\overline{X}} \\ & + \sum_{\mathbf{x} \in \mathbf{B}} \frac{X}{\lambda} e^{-\lambda T} T^X \frac{\prod_{i=1}^{N} \lambda_i^{x_i}}{Y} \frac{(X+1)}{2\overline{X}}, \end{aligned} \tag{15}$$

where $\overline{X}$ is the mean number of client packets per slot. To compute $\overline{X}$ we follow the same reasoning as for the calculation of the Mean Optical Slot Filling Time, but since all slots may not contain the same number of packets, we weigh $\overline{X}$ by the number of packets contained by the slot; which is X+1 in the case where the timer does not expire, "+1" corresponding to

the last packet triggering the slot sending. For the case where it is the timer that triggers the slot sending we do not add "1" since the formation slot is not conditioned on the arrival of a specific (last) packet; thus:

$$\begin{aligned}\overline{X} =& \int_0^T \sum_{j=1}^{N} \sum_{\mathbf{x}\in\mathbf{A_j}} \frac{X}{\lambda t} P_t(\mathbf{x}) \lambda_j (X+1)\,\mathrm{d}t \\ &+ \sum_{\mathbf{x}\in\mathbf{B}} \frac{X}{\lambda T} P_T(\mathbf{x}) X, \\ =& \sum_{j=1}^{N} \sum_{\mathbf{x}\in\mathbf{A_j}} \frac{\lambda_j X(X+1)}{\lambda} \left(\prod_{i=1}^{N} \lambda_i^{x_i}\right) \frac{(X-1)!}{Y} \sum_{n=1}^{X-1} \frac{a_n}{n!\lambda^{X-1-n}} \\ &+ \sum_{j=1}^{N} \sum_{\mathbf{x}\in\mathbf{A_j}} \frac{\lambda_j X(X+1)}{\lambda} \left(\prod_{i=1}^{N} \lambda_i^{x_i}\right) \frac{(X-1)!}{Y} \frac{I_0}{\lambda^{X-1}} \\ &+ \sum_{\mathbf{x}\in\mathbf{B}} \frac{X^2}{\lambda} e^{-\lambda T} T^{X-1} \frac{\prod_{i=1}^{N} \lambda_i^{x_i}}{Y}. \end{aligned} \tag{16}$$

Observe that *OSFD* only depends on the slot duration, the client packet sizes distribution and traffic demand distribution, which are given by the traffic matrix. *OSFD* is totally independent of the network state. The impact of CoS on the *OSFD* is through the intensities λ_i (and hence λ) of the flows participating on the formation of a same slot.

B. Optical Buffering Delay derivation

We now compute the *OBD* part of (1). *OBD* depends on two variables, per-CoS network load ρ (the traffic demand normalized by the maximum demand carried by the ring), and service time ν (the mean time to serve a slot). Those depend on the CoS management policy, as will be seen in Section III-C. Let τ be the optical slot filling ratio (amount of data in the slot divided by slot size). For the derivation of τ we follow the same reasoning as for the calculation of *OSFT*, shown in Section III-A. Since τ express a size and not a number of packets, instead of reasoning in term of number of arrivals $\sum_{i=1}^{N} x_i$ as in (10), here we consider the filling ratio of the slot, which is the size of the data in the slot ($s_j + V_{t^-}$) divided by the slot size (S). Hence, τ is:

$$\begin{aligned}\tau =& \int_0^T \sum_{j=1}^{N} \sum_{\mathbf{x}\in\mathbf{A_j}} \frac{s_j + V_{t^-}}{S\lambda t} P_t(\mathbf{x}) \lambda_j X\,\mathrm{d}t \\ &+ \sum_{\mathbf{x}\in\mathbf{B}} \frac{V_T}{S\lambda T} P_T(\mathbf{x}) X. \end{aligned} \tag{17}$$

Note that for case (a) (i.e., before timer expiration), we weigh τ by X instead of $X+1$ since the last packet that triggers the slot sending is already taken in account by adding its size s_j to the optical slot filling ratio. In [14], we derived in detail *OBD*:

$$OBD = \frac{1}{2} \cdot \frac{1}{\overline{X}.\nu} \cdot \frac{\rho}{\tau - \rho} + \frac{1}{2\mu}. \tag{18}$$

Unlike *OSFD*, (18) shows that *OBD* depends not only on the slot duration, client packet sizes distribution and traffic demand distribution, but also on the network dimensioning through ρ and ν. CoS management has an impact on *OBD* during slot formation through the terms $\overline{X}$ and τ, and during buffering and slot insertion process through the terms ρ and ν. In the following we derive ρ and ν according to the CoS management policy.

C. Impact of CoS management policy on QD

Without CoS management, all flows are served without differentiation; hence the load ρ_{transit} observed at a node is the same for all flows. ρ_{transit} is defined as the load transiting on the channel on which the considered slot is to be inserted, while load ρ_{in} is inserted by the source node; both ρ_{transit} and ρ_{in} are computed from the traffic demand:

$$\rho = \rho_{\text{transit}} + \rho_{\text{in}}. \tag{19}$$

The service rate is also the same for all flows and given by:

$$\nu = \mu \cdot (1 - \rho_{\text{transit}}), \tag{20}$$

The average queuing delay is obtained by plugging ρ and ν into (18), and plugging (15) and (18) into (1).

With CoS management, flows are treated differently depending on their CoS, hence the mean queuing time is computed separately for each CoS $k \in \{\text{RT, NRT}\}$ and (1) is replaced with:

$$QD_k = OSFD_k + OBD_k. \tag{21}$$

$OSFD_k$ and OBD_k are computed using the flows i for the relevant CoS in (15) and (18). Note that in (18) ρ and ν now depend on the CoS. Since RT traffic has priority over NRT traffic, loads ρ_{RT} and ρ_{NRT} are given by:

$$\rho_{\text{RT}} = \rho_{\text{transit}} + \rho_{\text{in\_RT}}, \tag{22}$$

$$\rho_{\text{NRT}} = \rho_{\text{transit}} + \rho_{\text{in}}, \tag{23}$$

where the ρ_{transit}, ρ_{in}, $\rho_{\text{in\_RT}}$ terms are inputs that are computed from the traffic matrix. Correspondingly, the service rates are:

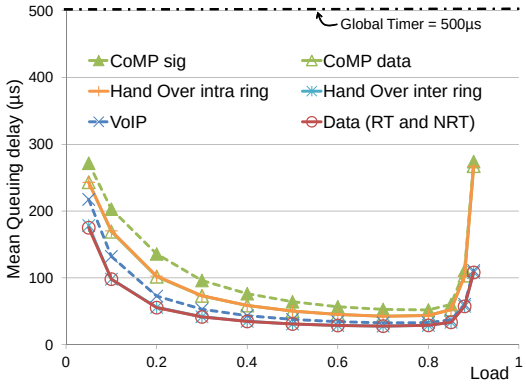
$$\nu_{\text{RT}} = \mu \cdot (1 - \rho_{\text{transit}}), \tag{24}$$

$$\nu_{\text{NRT}} = \mu \cdot (1 - \rho_{\text{transit}} - \rho_{\text{in\_RT}}). \tag{25}$$

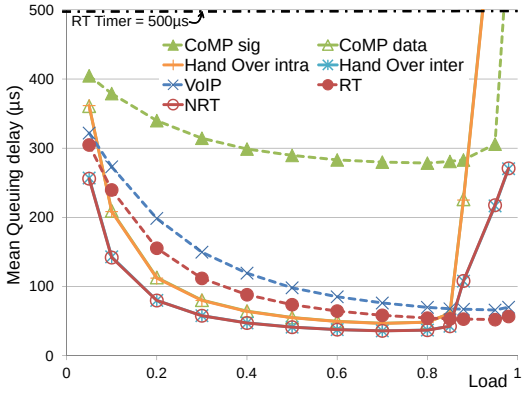
ρ_{RT}, ρ_{NRT}, ν_{RT}, ν_{NRT} are plugged in OBD_{RT} and OBD_{NRT} using (18); QD_{RT} and QD_{NRT} are obtained with (21). Note that (21) can be generalized for more than two classes of service as follows: consider k classes of service; each CoS has a different priority, if p is a given priority, and $p+m$ $(m = 0, \cdots, k-p)$ are higher priorities, then we obtain (21) for any CoS with a priority p using ρ_p and ν_p, defined as:

$$\rho_p = \rho_{\text{transit}} + \sum_{m=0}^{k-p} \rho_{\text{in\_p+m}}, \tag{26}$$

$$\nu_p = \mu \cdot (1 - \rho_{\text{transit}} - (\sum_{m=0}^{k-p} \rho_{\text{in\_p+m}}). \tag{27}$$



(a) without CoS management



(b) with CoS management

Fig. 6. Mean queuing delay per traffic type.

TABLE II
RING DIMENSIONING

node	H	1	2	3	4	5	6
number of TRX	3	1	1	1	1	1	1
RX colors	w_1, w_2, w_3	w_2	w_3	w_4	w_5	w_1	w_6

IV. PERFORMANCE EVALUATION

In this section we evaluate the performance of a 7-node POADM ring for an LTE-A mobile backhauling application. Using the same scenario we compare analytical and simulation results to validate the analytical model. Once the analytical model is validated we use it to evaluate the network performance for larger scenarios.

A. Scenario description

We consider a POADM ring connecting 7 eNodeBs; each eNodeB is connected to a POADM node. The traffic sent between eNodeBs, belonging to the same ring, remains inside the ring. The traffic exchanged with eNodeBs on other rings or with the Internet (via the backbone) is transmitted via a selected OSS node, which we call hub and denote by H. As mentioned earlier we consider 6 types of traffic. Each type of traffic has a specific pattern: centralized traffic is between any node and the backbone (i.e., it transits via H) while peer-to-peer traffic is between any two eNodeBs. The pattern and per-flow demands for each type are summarized in Fig. 5. The dimensioning of the POADM ring that ensures the support of such traffic demand is computed using algorithm [15] and results in a 6-wavelength ring with 3 transponders (TRX) on H and 1 transponder on every other node. Tab. II shows how the wavelengths $w_1, \ldots, w_6$ are assigned, and the receivers (RX) colored, for each node.

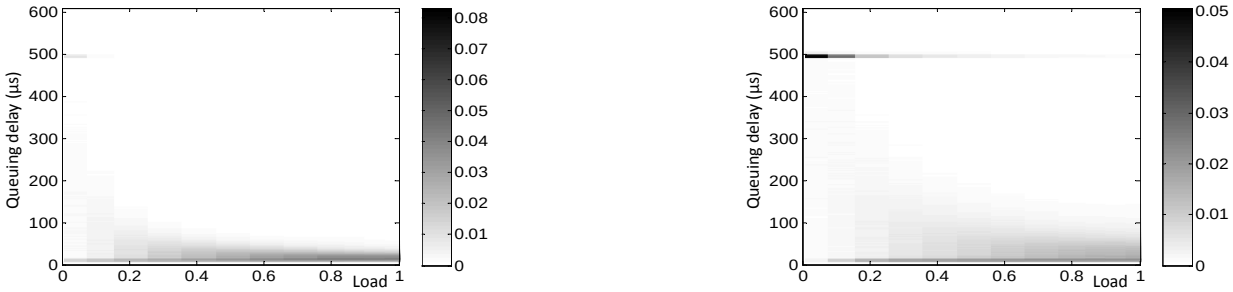
B. Queuing delay and CoS impact evaluation

First we evaluate the performance of the POADM solution with no CoS management in the 7-node mobile backhaul network. The slot duration is set to 8 µs and client traffic is assumed to be Poisson. In the following we consider that a load of 1 corresponds to the demand of Tab. I and other loads are generated by scaling this reference load. Propagation (a few µs in a 7-node mobile backhaul ring) and insertion/extraction on the medium (1 slot duration each) are small compared with the target real-time traffic latency (500 µs) such that queuing delays, which are reported here, are representative of the latency experienced by client data within the ring. We first set the global timer to 500 µs; Fig. 6a shows *QD* for each traffic type. Seemingly counter-intuitively, the queuing delay decreases as load increases for low loads; this is because queuing delay is then dominated by the time to actually fill the optical slots [16]. At higher loads the queuing delay mainly depends on channel congestion and queuing delay increases, as expected[2].

Fig. 6a shows that the mean queuing delay of all traffic classes is well below the timer limit for load up to 0.88. The approximately 10% wasted load is due to the impossibility to fully fill slots with client data. This problem can be solved by allowing segmentation of client packets in several slots; this is out of the scope of this work. Note that, using simulations results not shown here, there is no loss up to load 0.85 and we measured a client frame loss below 1e-3 up to load 0.9.

In Fig. 6b we show the impact of CoS management. We consider that half of the data traffic is RT and the other half is NRT. We keep a timer of 500 µs for RT traffic and set a timer of 1 ms for NRT traffic. Both RT and NRT queuing delays are higher than in the scenario without CoS management. The mean queuing delays remain below the respective timer values. However, as we can see in Fig. 6b, the CoS management policy gives an advantage to RT traffic, which can achieve a load close to 1 for a packet loss probability below 1e-3. This is at the expense of a lower maximum amount of NRT traffic that can be carried in the network for a given client frame loss rate. Indeed in Fig. 6b, NRT traffic can still achieve a

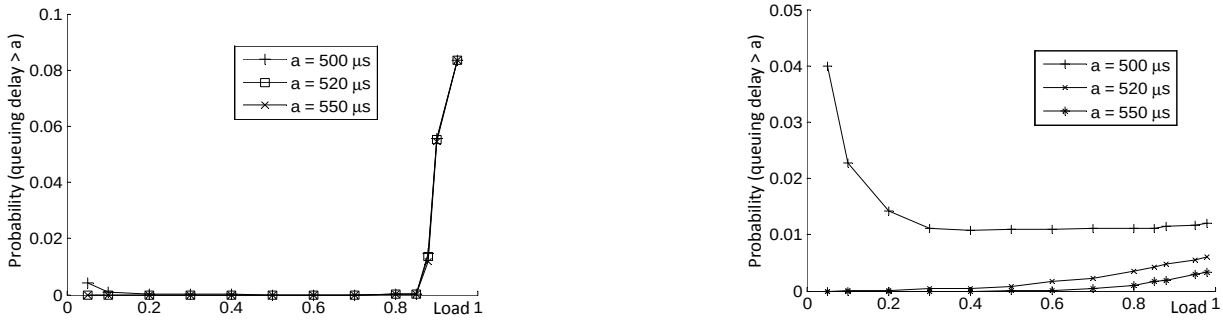
[2]The small differences across the different plots in Fig. 6a are due to the traffic matrix pattern, which depends on the type of flow. As explained in Section III, the mean queuing delay depends on λ, which is the sum of the arrival rates for all aggregated flows participating in the filling of a slot. Since λ differs according to the destination node (without CoS management, slots aggregate packets with same destination node) and since destination node is related to traffic type (some traffic types follow centralized traffic pattern, and others follow a peer-to-peer traffic pattern), the optical slots formed from different flows experience different queuing delays.



(a) All traffic types, without CoS management

(b) RT traffic types, with CoS management

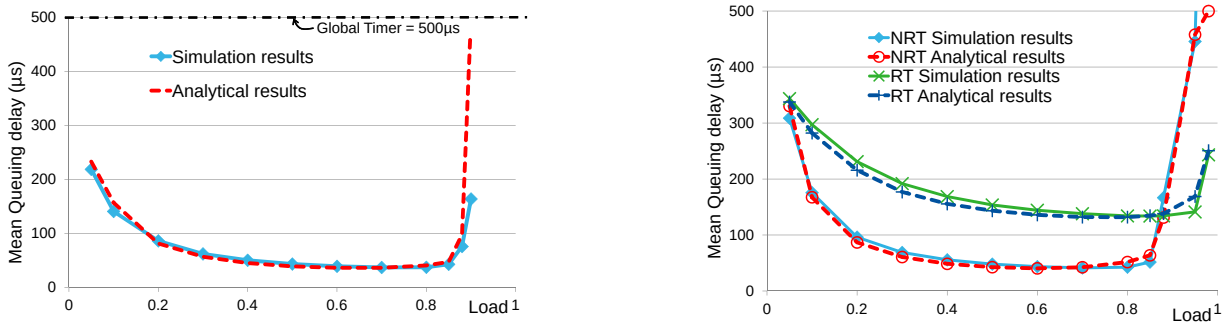
Fig. 7. Probability Mass Function for optical client packet queuing delays (bin width for each histogram: 1 µs; histograms cropped at 600 µs.).



(a) All traffic types, without CoS management

(b) RT traffic types, with CoS management

Fig. 8. Probability that a client packet experiences a delay larger than {500, 520, 550} µs.



(a) without CoS management

(b) with CoS management

Fig. 9. Mean queuing delay, simulation vs. analytical results for a 7-node ring.

load of 0.85 but beyond this value the queuing delay is higher with than without CoS management: RT traffic experiences almost no loss while NRT traffic experiences high losses for load greater than 0.85. Fig. 7 depicts, for several load values between 0.05 and 0.98, the probability mass function (PMF) for the queuing delay. When QoS management is not used (Fig. 7a), the PMF for all traffic, both real-time and non real-time, is shown, since they are mixed in the same optical slots. When QoS management is used (Fig. 7b), we depict the PMF for real-time traffic only, since non real-time traffic is latency-insensitive. In addition, Fig 8 shows the probability that the queuing delay experienced by a client packet is higher than 500, 520 or 550 µs. The horizontal darker bars around 500 µs in Figs. 7a and 7b correspond to traffic encapsulated in near-empty slots, which are inserted only upon timer expiration; note that the upper bound on queuing delay is slightly above the timer value of 500 µs because a slot may have to wait for one or two timeslots before being inserted on the channel. We observe that, even for high loads (between 0.8 and 0.95), very few packets (< 2 %) experience a queuing delay higher than 550 µs. If CoS management is used, only less than 2 % of the RT traffic experiences a delay higher than 550 µs, and less than 1 % of the RT traffic experiences a delay higher than 500 µs for a load of 0.98. Hence, whichever CoS management policy is used, we can claim that the proposed optical slot switching ring fulfills the strict latency constraints of LTE-A mobile backhaul network.

For the evaluation of the analytical model we compare

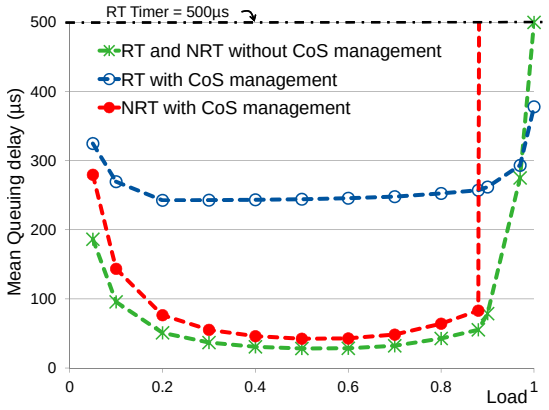


Fig. 10. Queuing delay analysis; with and without CoS management for a 19-node ring.

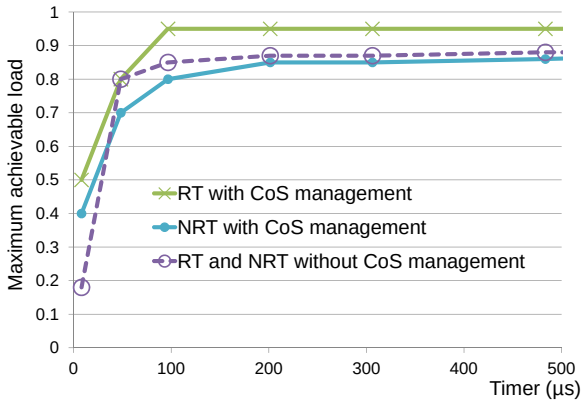


Fig. 11. Maximum achievable load for packet loss rate below 1e-3, using analysis for a 7-nodes ring.

the mean queuing delay for both traffic classes (RT and NRT) with and without CoS management for a 7-node ring. Figs. 9a and 9b show the mean queuing delay, for all traffic types, analytically and by simulation. Analytical results closely match simulation results, thereby validating our analytical model. We can then use the analytical model to predict the performance of a 19-node ring. Fig. 10 shows that the trends for the queuing delay for a 19-node ring are similar to those of a 7-node ring. This result proves that the optical slot switching ring is a scalable solution and that its performance remains the same with little dependence on the network size.

C. Impact of timer on network performance

In order to further reduce the network queuing delay, and hence the latency, it is possible to decrease the timer(s) value(s) (global timer or per-CoS timers). However setting timer(s) to very short durations would force the sending of near-empty optical slots, and hence waste a part of the offered channel capacity. Therefore in this section we show the trade-off between the latency reduction and the maximum achievable load, defined as the maximum load for which the client packet loss rate is below 1e-3. Fig. 11 shows the impact of the timer (without CoS management: global timer; with CoS management: RT timer, the NRT timer is set to 1 ms) on the maximum achievable load for each traffic class (RT and NRT). Without CoS management we can reach a load of 0.8 while ensuring a queuing delay for any traffic of 50 µs. Using CoS management, for RT traffic, we can still achieve a load close to 1 with a timer of 100 µs and load of 0.8 using a timer of 50 µs. For very short timers (below 200 µs) the maximum load achieved for NRT traffic is impacted by the prioritization of RT traffic but remains greater than 0.7 for a timer of 50 µs, and than 0.8 for a timer of 100 µs.

V. Conclusion

We showed that an optical slot switching network fulfills the latency constraints of a next generation LTE-A switching mobile backhaul network, which can reach as low as a few hundreds of microseconds in the considered segment. Even without differentiated CoS handling, real-time traffic can be supported with load up of to 88%. When CoS handling is used while forming slots it is possible to support even more real-time traffic, at the expense of a reduced non-real-time maximum carried load. Without CoS management, we showed that we can ensure a queuing delay of 50 µs while reaching a load of 80% for both RT and NRT traffic; with CoS management we can ensure a queuing delay of 100 µs while reaching a load of nearly 100% for RT and 80% for NRT traffic. We also proposed an analytical model with realistic assumptions and showed that the model closely matches the simulation results. Using the analytical model we showed that the optical slot switching ring is a scalable solution since its performance in terms of latency does not degrade even with a growing number of nodes per ring and the associated traffic demand.

Acknowledgment

This work was partly supported by DGE (France) and BMBF (Germany) through CELTIC + SASER-SAVENET project.

References

[1] Q. Wei, J. Bazzi, K. Kazuyuki, and M. Lott., "Packet switched optical mobile network - architecture design and performance analysis," in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, Mar. 2013.

[2] Q. Wei, J. Bazzi, M. Lott, and Y. Pointurier, "Multicast in mobile backhaul with optical packet ring," in *Selected Topics on Wireless and Mobile Computing, Networking and Communications (STWiMob)*, Oct. 2013, pp. 181–188.

[3] Cambridge Broadband Networks Limited, "Backhauling x2," white paper, 2010.

[4] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in lte-advanced: deployment scenarios and operational challenges," *Communications Magazine, IEEE*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[5] D. Chiaroni, G. Buforn Santamaria, C. Simonneau, S. Etienne, J.-C. Antona, S. Bigo, and J. Simsarian, "Packet OADMs for the next generation of ring networks," *Bell Labs Technical Journal*, vol. 14, no. 4, pp. 265–283, Winter 2010.

[6] Y. Pointurier, B. Uščumlić, I. Cerutti, A. Gravey, and J.-C. Antona, "Dimensioning and energy efficiency of multi-rate metro networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 22, pp. 3552–3564, Nov. 2012.

[7] P. Briggs, R. Chundury, and J. Olsson, "Carrier ethernet for mobile backhaul," *IEEE Communications Magazine*, vol. 48, pp. 94–100, Oct. 2010.
[8] T. Orphanoudakis, E. Kosmatos, J. Angelopoulos, and A. Stavdas, "Exploiting PONs for mobile backhaul," *IEEE Communications Magazine*, vol. 51, pp. S27–S34, Feb. 2013.
[9] N. Cvijetic, A. Tanaka, M. Cvijetic, Y.-K. Huang, E. Ip, Y. Shao, and T. Wang, "Novel optical access and digital processing architectures for future mobile backhaul," *IEEE/OSA Journal of Lightwave Technology*, vol. 31, pp. 621–627, Feb. 2013.
[10] R. M. Indre, "Performance analysis in subwavelength switching optical networks," Ph.D. Thesis, Telecom ParisTech, 2012.
[11] NGMN alliance, "Guidelines for lte backhaul traffic estimation," white paper, 2011.
[12] S. Brueck, L. Zhao, J. Giese, and M. Amin, "Centralized scheduling for joint transmission coordinated multi-point in lte-advanced," in *Smart Antennas (WSA), 2010 International ITG Workshop on*, Feb. 2010, pp. 177–184.
[13] T. Bonald and M. Feuillet, *Network Performance Analysis*. Wiley-ISTE, Oct. 2011.
[14] N. Benzaoui, Y. Pointurier, T. Bonald, and J.-C. Antona, "Impact of the electronic architecture of optical slot switching nodes on latency in ring networks," *Optical Communications and Networking*, vol. 6, no. 8, pp. 718–729, Aug. 2014.
[15] B. Uščumlić, I. Cerutti, A. Gravey, P. Gravey, D. Barth, M. Morvan, and P. Castoldi, "Optimal dimensioning of the WDM unidirectional ECOFRAME optical packet ring," *Springer Photonic Network Communications*, vol. 22, no. 3, pp. 254–265, Jul. 2011.
[16] N. Benzaoui, Y. Pointurier, and J.-C. Antona, "Electronic architectures of optical slot switching nodes," in *Proc. ONDM*, 2013, pp. 94–99.

Nihel Benzaoui received the engineering degrees from Institut National des Télécommunications et des Technologies de l'Information et de la Communication, Oran, Algeria in 2010, a Master Spécialisé en Conception et Architecture Réseau from Telecom ParisTech, Paris, France in 2012. She then joined Alcatel-Lucent Bell Labs, France, as a Ph.D. student working on multi-layer mechanisms for optical slot switching networks.

Yvan Pointurier (S'02–M'06–SM'12) received a Diplôme d'Ingénieur from Ecole Centrale de Lille, France in 2002, a M.S. in Computer Science in 2002, and a Ph.D. in Electrical Engineering in 2006, both from the University of Virginia, USA. He spent two years at McGill University in Montreal, Canada and then one year at Athens Information Technology, Greece, as a Postdoctoral Fellow. In 2009 he joined Alcatel-Lucent Bell Labs, France, as a research engineer. His research interests span design, optimization and monitoring of networks in general, and optical networks in particular. Dr. Pointurier is the author or co-author of 16 European patents (including 4 active patents) and 70 technical articles. He is a co-recipient of the Best Paper Award at the IEEE ICC 2006 Symposium on Optical Systems and Networks.

Thomas Bonald is Professor in the department of Computer Science and Networking at Telecom ParisTech, France. He graduated as an Engineer from Ecole polytechnique in 1994 and from Telecom ParisTech in 1996. He received the Ph.D. degree in Applied Mathematics from Ecole polytechnique in 1999. From 1999 to 2009 he was a Research Engineer at Orange Labs, working on traffic modeling for wireline and wireless networks. His main research interests are in queueing theory and stochastic models of communication networks. He is especially interested in applications to the design and performance evaluation of resource allocation schemes like packet scheduling and medium access control algorithms. Thomas Bonald is the author of 70 scientific papers and 10 patents. He received with Alexandre Proutire the Best Paper Award of ACM SIGMETRICS / IFIP Performance 2004. He is with Mathieu Feuillet the author of the book Network Performance Analysis published by Wiley in 2011. He is a member of the Board of Directors of ACM SIGMETRICS and an Associate Editor of the IEEE/ACM Transactions on Networking.

Wei Qing received the B.Sc. in Electrical Engineering, from Shanghai Jiao Tong University, P.R.China in 1997 and the M.Sc. in Communications Engineering from Technical University of Munich, Germany in 2001 with honor. Between 1997 and 1999 she worked in Philips Lighting Electronics (Shanghai) Co. Ltd as a research engineer. In 2002, she joined DOCOMO Euro-Labs in Munich and has been worked on context awareness, active networks, wireless sensor networks, network coding, 802.11 MAC, mobility management, MPLS, optical networking, Cloud RAN, small cells and 5G mobile network architecture. She holds over 10 IPRs and 30 publications in these areas. She is also involved in several EU projects and contributes to various standardization/industry forums.

Matthias Lott received the diploma degree in electrical engineering from University of Braunschweig and the Ph.D. (Dr.-Ing.) degree from RWTH Aachen, Germany, in 1996 and 2001, respectively. In 2000 he joined Siemens AG, and in 2007 he became an employee of Nokia Siemens Networks. During this time he was involved in European and German research projects, investigated future core network strategies, developed prototypes, and finally, he was responsible for research in the areas of service control and identity management. Since 2012 he was a director at DOCOMO Euro-Labs in Munich, focusing on future radio access and network virtualization, quality of experience and optical mobile networking. He has published over 100 papers, wrote book chapters, and holds more than 30 patents. He received the ITG best-paper award in 1999, the Borcher's Medal from RWTH Aachen in 2002, the European Wireless best-paper award in 2005, he was Siemens Inventor of the year 2005, received the Nokia Siemens Networks Quality Award in 2010, and is a member of VDE.