IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 11, 2016; accepted December 7, 2016, date of publication January 9, 2017, date of current version January 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2643690

Long Term Evolution-Wireless Local Area Network Aggregation Flow Control

DAVID LÓPEZ-PÉREZ[1], DANIELA LASELVA[1], EUGEN WALLMEIER[2], PÄIVI PUROVESI[2], PETTERI LUNDÉN[1], ELENA VIRTEJ[1], PIOTR LECHOWICZ[2], ESA MALKAMAKI[1], AND MING DING[3]
[1]Nokia Bell Laboratories
[2]Nokia Mobile Networks
[3]Data 61, Sydney, Australia

D. López-Pérez (david.lopez-perez@nokia-bell-labs.com)

ABSTRACT Long term evolution-wireless local area network (LTE-WLAN) aggregation (LWA) has recently emerged as a promising third generation partnership project (3GPP) Release 13 technology to efficiently aggregate LTE and WLAN at the packet data convergence protocol layer, allowing uplink traffic to be carried on LTE and downlink on both LTE and WLAN. This removes all the contention asymmetry problems of WLAN and allows an optimum usage of both licensed and unlicensed band for downlink. In this paper, we present a new feature of LWA, its flow control scheme, which controls how to aggregate downlink traffic in licensed and unlicensed bands. This aggregation technique exploits user equipment-based flow control feedback in the form of LWA status reports, and can be expanded to work with any number of frequency bands and radio technologies. The same concepts apply to 5G networks, although the performance evaluation provided here is in the context of LTE-Advanced Pro. Simulation results in a typical enterprise scenario show that LWA can enhance user performance up to 8 times over LTE-only, and 3.7 times over WLAN only networks, respectively. The impact of the file size and LWA status report frequency on network performance is also investigated.

INDEX TERMS Cellular networks, long term evolution (LTE), wireless local area network (WLAN), LTE WLAN aggregation (LWA), flow control, aggregation.

I. INTRODUCTION

Long term evolution (LTE) networks carry a continuously increasing amount of data driven by the growing number of worldwide LTE subscribers, which reached nearly 1.5 billion in 2016. In parallel, radio capabilities are also rapidly evolving with the development of LTE-Advanced (LTE-A) and LTE-A Pro, enabling peak data rates of 450 Mbps with carrier aggregation. This fast uptake of LTE in different regions of the world shows how the demand for mobile broadband continues to increase, and that LTE and its evolutions are a successful platform to meet such demand [1], [2].

In order to keep up with customers needs, cellular operators around the world are looking with interest at unlicensed spectrum as a complementary tool to expand the capacity of their networks and augment their service offering [3]. With a tight integration to operator's licensed spectrum, unlicensed spectrum represents a valuable set of new carrier frequencies available for deployment in small cells [4]. The simultaneous usage of licensed and unlicensed spectrum offers the end-users access to a larger bandwidth and a better performance, *e.g.*, higher (peak) data rates. Moreover, this heterogeneous spectrum access means that the licensed spectrum can take over the unlicensed one to provide quality of service (QoS), if the latter becomes unusable for any reason, *e.g.*, reduced coverage, interference from another system or avoidance of, for example, a radar operating in the band. Therefore, compared to currently available loose interworking solutions defined by the Third Generation Partnership Project (3GPP), such as access network discovery and selection function (ANDSF) [5] and radio access network (RAN)-assisted/controlled LTE-wireless local area network (WLAN) radio interworking (RALWI/RCLWI) [6], which entirely switch data bearers between LTE and WLAN, the tight integration of licensed and unlicensed spectrum can provide higher data rates and better QoS guarantees. Tight integration also makes the access to unlicensed spectrum transparent to the operator's evolved packet core, which simplifies the overall network maintenance by avoiding multiple solutions for network management, security and authentication.

In order to realise this tight integration, the 3GPP has recently standardised a set of new features as part of the

Release 13 of the evolved UTRA (EUTRA) specifications (a.k.a. LTE-A Pro), *i.e.*, licensed assisted access (LAA) [7],[1] LTE-WLAN radio level integration with IPsec tunnel (LWIP) [6] and LTE-WLAN aggregation (LWA) [6].

The main difference between LTE-U/LAA and LWIP/LWA is that LAA uses LTE radio access technology to operate the unlicensed spectrum, while LWIP and LWA access it through IEEE 802.11 standards, thus allowing to leverage incumbent wireless local area network (WLAN) deployments. It is important to note that the use of unlicensed spectrum carries some regulatory requirements to allow co-existence [8], such as being able to detect if a radar system is using the band or co-exist with other nodes using the band. The latter is often referred to as listen-before-talk, and implies that it is not possible to transmit immediately if the intended channel is occupied. Due to regulations, the allowed transmit power also varies depending on the region and the part of the band used. With this in mind and in contrast to LAA, regulatory requirements are not a concern for LWIP and LWA, since they access the unlicensed spectrum through already certified WLAN standards.

With regard to LWIP and LWA, the main difference between them both in Release 13 is that LWIP supports downlink switching of internet protocol (IP) flows at the IP layer (*i.e.*, packets of an IP data flow are transmitted in the LTE or WLAN link but never in both), while LWA is able of aggregating packet data convergence protocol (PDCP) flows at the use PDCP layer (*i.e.*, packets of a PDCP data flow can be simultaneously transmitted in both the LTE or WLAN link). The use of a different integration layer, IP versus PDCP, has important deployment and performance implications. While the former provides a better performance more universal solution since it can work with any WLAN node, the latter may not be as universal but can provide a larger performance due to its aggregation capabilities. In this paper, we focus on LWA and its aggregation capabilities. Please refer to [9] for further details on LWIP and its current switching capabilities.

Techniques for single/multi-RAT traffic aggregation at the radio layer have been considered in the literature. The work in [10] proposes a flow control scheme for LTE dual connectivity (DC) neglecting TCP impacts. An elegant closed form this solution for LWA traffic splitting is proposed in [11]. However, this solution relies on the knowledge at the LTE eNodeB of the UE instantaneous transmission, which also requires rate in WLAN, which is exchanged over the backhaul, in order to run a joint resource allocation on LTE and WLAN, which also requires performing a computationally a new traffic aggregation-based LWA flow control algorithm and expensive sorting of the active UEs.

In this paper, we present a new traffic aggregation-based LWA flow control algorithm and user equipment (UE)-based flow control feedback that *i)* allows an efficient downlink traffic aggregation of both licensed and unlicensed spectrum, *ii)* leverages the existing feedback to estimate and minimise the per-packet delay, thus optimising TCP performance, *iii)* can cope with any number of radio links, spectrum bands, and backhaul latency, and *iv)* can be applied to other technologies such as DC or LWIP.

[1] LTE-unlicensed (LTE-U) is an industry standard that allows the operation of LTE-like technology in the unlicensed band [4] in geographical areas where listen before talk is not mandatory, e.g., United States.

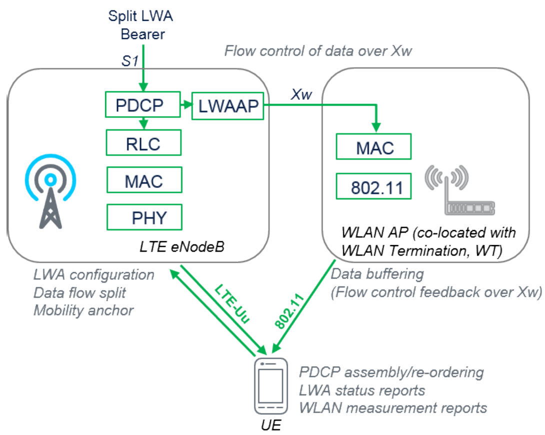

FIGURE 1. **Release 13 LWA architecture and main functions.**

The rest of the paper is organised as follows: In Section II and Section III, the architecture and the UE-based feedback framework of LWA Release 13 are introduced, respectively. In Section IV, the new flow control devised for LWA is presented. In Section V, simulation results, which show the performance gain of LWA and the proposed flow control with respect to LTE only and WLAN only networks, are discussed. Finally, in Section VI, the conclusions are drawn.

II. LTE-WLAN AGGREGATION (LWA) RELEASE 13

Leveraging the LTE DC bearer-split architecture [12], LWA supports downlink aggregation at the PDCP layer and re-uses the PDCP based reordering mechanism introduced for split bearers. In more detail, PDCP protocol data units (PDUs) of the same IP flow can be independently routed by the LTE eNodeB through the LTE and WLAN links, while the PDCP layer re-ordering mechanism at the UE ensures in-sequence delivery to the upper layers based on the sequence numbering of each PDU. Differently from the existing aggregation schemes in the industry, which mostly occur at the application layer, *e.g.*, multi-path transmission control protocol (MPTCP), per-PDCP PDU split in LWA permits to exploit a faster adaptation to radio and traffic fluctuations in both LTE and WLAN downlinks, as LWA works at a radio protocol layer, and benefits from the knowledge of further radio link statistics. In Release 13, uplink transmissions do not benefit from aggregation and are only supported on the LTE network. Release 14, instead, extends the aggregation flexibility in the uplink direction.

Fig. 1 illustrates the overall architecture of LWA. LWA supports both co-located and non co-located scenarios. For the latter, it features a new direct interface *Xw*, as shown in Fig. 1, defined between LTE and WLAN, which has similar capacity and latency requirements as the *X2* interface. The new *Xw* interface is terminated at the WLAN termination (WT), which is a newly defined 3GPP logical node that may be in control of one or more WLAN access points (APs), and can be used to carry not only PDCP PDUs (user plane) from LTE to WLAN, but also control plane information from WLAN to LTE and flow control feedback.

LWA, however, may not always be capable of leveraging sophisticated *Xw*-based flow control mechanisms in order to assist PDCP PDU routing decisions, as legacy WLAN APs may not be able to provide such feedback. To address this issue and avoid any WLAN to LTE feedback dependencies, a new UE-based flow control feedback framework for the PDCP layer has been standardised in Release 13, which allows UEs to directly sent information related the WLAN link performance to the serving LTE eNodeB via the new LWA status reports. This new feedback framework is discussed in more detail in Section III.

It is also important to note that LWA adds a new LWA-specific EtherType to each PDCP PDU routed over WLAN, as part of the Ethernet frame, to allow the UE to differentiate LWA traffic from other WLAN traffic. Therefore, in order for LWA to work, the involved WLAN APs need not to discard frames containing the LWA-specific EtherType value. Simple upgrades of WLAN APs already deployed may be needed to overcome this issue and enable the LWA operation.

LWA also requires new UE chipset support. A dual-radio terminal capable of LWA needs to perform and feed back WLAN radio measurements, including WLAN connection failure detection to support the configuration, pausing and release of the LWA operation by the LTE eNodeB. If the WLAN connection fails, the mobility anchor on the LTE eNodeB ensures that the UE connection is maintained.

Moreover, a novel LTE eNodeB-assisted WLAN mobility procedure has been defined which re-uses existing WLAN mobility mechanisms and minimises WLAN related signalling. The LTE eNodeB controls the WLAN APs involved in the LWA operation providing the WLAN mobility set. Then UE performs mobility procedures between WLAN APs belonging to the provisioned set without informing the LTE eNodeB. In contrast, any mobility event, which occurs towards a WLAN AP outside the WLAN mobility set, is controlled by the LTE eNodeB through measurement reporting. Such events may trigger LWA reconfiguration, as in case of inter-WT mobility.

III. LWA UE-BASED FLOW CONTROL FEEDBACK

Having a proper estimation at the eNodeB of the number of bits successfully received at the UE through the WLAN link since the previous LWA status report is key for a proper PDCP level scheduling and an efficient spectrum aggregation. In order to allow such estimation when *Xw*-based feedback is not available, a new UE-based feedback framework in the form of LWA status reports has been standardised in Release 13, which contains the following information:

- First missing sequence number (FMS): The first missing PDCP sequence number (SN) in sequence of received sequence numbers.
- Highest received SN on WLAN (HRW): The highest successfully received PDCP sequence number on the WLAN link, or FMS if no PDCP PDUs have been received on the WLAN link.
- Number of missing PDUs (NMP): The number of missing PDCP PDUs with PDCP sequence numbers below HRW starting from and including FMS.

The report periodicity can be flexibly set as fast as 5 ms and as slow as 50000 ms, and it is important to note that a more frequent LWA status report should result in a better aggregation performance due to more accurate and up-to-date channel statistics, e.g., more frequent LWA status reports deal better with the randomness of the wireless channel. However, these gains come at the cost of uplink overhead, which should be kept moderate. Moreover, it is important to note that QoS is taken care independently in the two systems, LTE and WLAN, at the medium access control (MAC) layer according to regular mechanisms, i.e., irrespective of LWA. In addition, fairness across LWA and non-LWA UEs could be taken care at the MAC layer, if desired.

In an example, if a UE has received PDCP PDUs 1, 2, 4 through LTE and 5, 7, 9, 11 through WLAN, its LWA status report will be as follows FMS=3, HRW=11, NMP= 4. It is important to note that there are 4 PDCP PDUs indicated as lost, but actually only PDCP PDU=3 is surely lost,[2] since PDCP PDUs 6, 8 and 10 may still be under transmission via LTE. Note that since LWA status reports work on the basis of PDCP PDUs, it is transparent to lower layers operations, e.g., hybrid automatic repeat request (HARQ) operations.

It is obvious that from the above LWA status report, the eNodeB cannot directly infer the number of bits successfully received at the UE, denoted ACK_{WLAN}. However, it can rather accurately perform an estimation of it using the provided information. The estimation procedure is described in the following.

Since the eNodeB knows that WLAN AP delivers in order (however possibly with gaps), ACK_{WLAN} can be calculated as

$$ACK_{WLAN} = D_{WLAN} - L_{WLAN}, \tag{1}$$

where D_{WLAN} (departure) are the sum bits of all PDCP PDUs passed onto the WLAN link, whose PDCP sequence number is greater than the HRW indicated in the last message and smaller or equal than the HRW reported in the current message, and L_{WLAN} (lost) are the sum bits in all PDCP PDUs lost in the WLAN link, which in turn can be estimated as,

$$L_{WLAN} = (\text{NMP} - \alpha) \cdot \text{PDU}_{size}^{avg}, \tag{2}$$

[2]With regard to the lost PDU=3, it is important to note that it is up to the eNodeB whether to re-transmit or not this PDU. This eNodeB's freedom, however, could make the UE re-ordering wait for a PDU that could never arrive.

where α is the number of PDCP PDUs passed onto the lower layer through the LTE link, whose PDCP sequence number is greater or equal than FSM and less than HRW, and PDU_{size}^{avg} is the average PDCP PDU size through the WLAN link.

With regard to D_{WLAN}, the eNodeB can derive this value because it knows the sequence numbers of the PDCP PDUs that it scheduled in the WLAN link, and has information on HRW of the last and current LWA status report.

With regard to L_{WLAN}, NMP should include any PDU missing over the WLAN and LTE links. However, due to HARQ and radio link control (RLC) acknowledge mode properties (automatic repeat request (ARQ) retransmissions and in-sequence delivery to upper layer), the likelihood of missing a PDCP PDU over the LTE link is negligible and therefore neglected in the calculation, i.e., a PDCP PDU can only be missed over the WLAN link. Thus, the proposed computation of L_{WLAN} is a good estimation of the sum bits in all PDCP PDUs lost in the WLAN link, since the LTE eNodeB is aware of the PDCP PDU under retransmission.

IV. LWA FLOW CONTROL ALGORITHM

In this section, we present our proposed LWA flow control algorithm that relies on the UE-based flow control feedback presented in the previous section, and that is targeted at providing an efficient downlink traffic aggregation leveraging both licensed and unlicensed spectrum. When deploying it in a traditional architecture, the algorithm could sit at the PDCP layer of the eNodeB. Alternatively, when using a cloud RAN architecture, it could reside in a central cloud server, which handles the PDCP layer operations of multiple cells.

The *objective* of the algorithm is to optimally split the traffic flow between the LTE and WLAN links, which may have quickly fluctuating capacity and latency conditions. In order to perform such split, each PDCP PDU is routed through the link that provides the shortest packet delay. This chosen metric results in a minimised PDCP re-ordering delay and a faster in-order delivery to the upper layer of the UE. This plays a key role in improving the performance of transmission control protocol (TCP)-based applications, thanks to the reduced TCP round trip times (RTTs) and the avoidance of the potential TCP congestion control mechanisms. The algorithm is also targeted at avoiding buffer underflow and overflow (the former causes link starvation when there is available capacity, while the latter results in packet losses and triggers TCP congestion mechanisms) and in general avoiding too long queueing times, which may result in a suboptimal splitting as the link capacity may vary rapidly.

In this paper, we focus on a system comprised of an LTE and a WLAN link, but it is important to note that the algorithm can cope with any number of radio links and spectrum bands, and can be applied to other technologies such as DC (aggregating licensed bands) or LWIP (aggregating licensed and unlicensed bands at the IP layer), if the appropriate feedback is in place.

In order to realise the mentioned objectives, we propose an algorithm that is comprised of three phases, which act

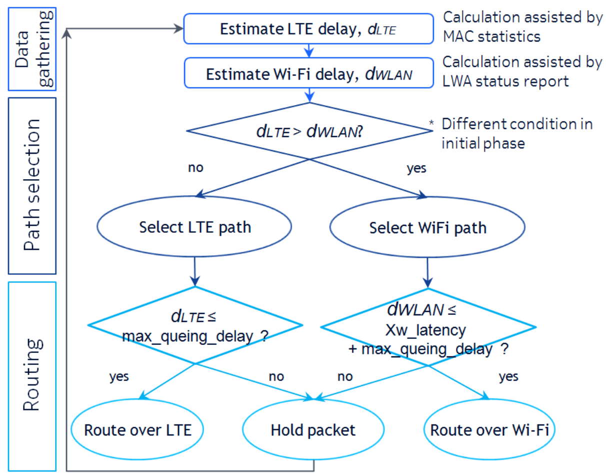

FIGURE 2. Flow control diagram.

sequentially: the data gathering phase, the path selection phase and the routing decision phase. Fig. 2 illustrates the proposed flow control algorithm.

In essence, if the LWA capability is activated for a given data flow, our LWA flow control algorithm decides whether an arriving downlink service data unit (SDU) to the PDCP layer of a split bearer should be:

- forwarded to the LTE link or through the *Xw* interface to the WLAN link (this decision is done during the path selection phase), or
- hold at the PDCP layer, and its routing decision postponed until congestion conditions are favourable in one of the links (this decision is done during the routing phase).

According to the selected metric, the path selection decisions are based on delay estimates, obtained during the prior data gathering phase. Then, each PDCP PDU is forwarded to the link that provides the shortest delay. The purpose of the posterior routing phase is to avoid aggravating congestion conditions, by limiting the number of PDCP PDUs that are in-flight between the splitting function at the eNodeB and the reordering function at the UE upon congestion detection.

In the following, the three different phases of the algorithm are described in more detail.

A. DATA GATHERING PHASE

The PDCP layer at the eNodeB stores downlink SDUs of a split bearer in a bearer specific first-in first-out (FIFO) queue, and when the PDCP layer processes an SDU (ciphering, adding a PDCP sequence number, etc.), the proposed algorithm *i)* estimates, for each link, LTE or WLAN, how long the PDCP PDU would need to travel from the eNodeB splitting function to the UE (a.k.a. link delay), and then *ii)* forwards it through the link that provides the shortest delay. The performance of the algorithm is sensitive to the delay estimation, and in this case, we calculate the time needed for a PDCP PDU to reach the UE through a given link using Little's law [13]. This time depends on the capacity and latency of the link, including the amount of queuing along the link as well as external factors such as the *Xw* interface latency.

According to Little's law, the mean delay δ of a packet in an arbitrary stable queuing system can be computed as $\delta = \frac{B}{R}$, where B is the mean number of bits in the system, and R is the mean rate of departure or the throughput of the system. In our Little's law queueing system, the bits of a PDCP PDU enter the queueing system when the eNodeB puts them onto the lower layer towards, e.g., the LTE or WLAN link, and leave the queueing system when they have been received by the UE.

With this in mind, let us define some notation. Let t denote the time at which a scheduling decision for a PDCP PDU is made, let PDU_{size} denote the size of the PDCP PDU, let $Xw_{latency}$ denote the Xw interface latency, and let $l = [LTE, WLAN]$ denote the type of link.

Moreover, let B_l denote the number of *bits-in-flight* in link l, which are the sum bits of the PDCP PDUs passed onto the lower layer, but not acknowledged yet. In the WLAN link, the bits-in-flight are acknowledged by the LWA status report, thus when an LWA status report arrives, the B_{WLAN} is updated as follows

$$B_{WLAN}(t) = B_{WLAN}(t\prime) - ACK_{WLAN}(t), \tag{3}$$

where $t\prime$ is the time when B_{WLAN} was updated the last time before t.

In addition, let R_l denote the *expected rate* in link l, which is an estimate of the throughput supported in each link. In the WLAN link, the expected rate is the ratio between the PDCP PDU bits acknowledged by an LWA status report, ACK_{WLAN}, and the time ΔT in between this LWA status report and the previous one, *i.e.*

$$R_{WLAN}(t) = \frac{ACK_{WLAN}(t)}{\Delta T}. \tag{4}$$

When link statistics are not available, *e.g.*, no LWA status report has arrived and thus there is no ACK_{WLAN} information, or $B_l = 0$, then $R_{WLAN} = R^{int}_{WLAN}$, an initialisation value for the WLAN link.

Similarly, the bits-in-flight in LTE can be acknowledged by the RLC status reports, and when link statistics are not yet available $R_{LTE} = R^{int}_{LTE}$, the initialisation value for the LTE link.

In the following, we explain how the eNodeB can estimate the time the PDCP PDU would need to reach the UE when either link is used.

1) LTE LINK DELAY

Let t_{LTE} denote the point in time when the PDCP PDU reaches the UE, if the LTE link is selected. Then, d_{LTE} is an estimate of the delay $t_{LTE} - t$ for the PDCP PDU at time t, where

$$d_{LTE} = \frac{PDU_{size}}{R_{LTE}} + \frac{B_{LTE}}{R_{LTE}}, \tag{5}$$

2) WLAN LINK DELAY

Let t_{WLAN} denote the point in time when the PDCP PDU reaches the UE, if the WLAN link is selected. Then, d_{WLAN} is an estimate of the delay $t_{WLAN} - t$ for the PDCP PDU at time t, where

$$d_{WLAN} = \frac{PDU_{size}}{R_{WLAN}} + \max(Xw_{latency}, \frac{B_{WtoUE}}{R_{WLAN}}), \tag{6}$$

where in turn B_{WtoUE} is the number of bits in the considered Little's queueing system. B_{WtoUE} is not known at the eNodeB, but it can be estimated from the bits-in-flight B_{WLAN} by subtracting an estimate of the number of bits that have been received by the UE but not acknowledged at the eNodeB yet, i.e.,

$$B_{WtoUE} = B_{WLAN} - \min(B_{WLAN}, (t - t_{upd} + Uu_{latency}) \cdot R_{WLAN}), \tag{7}$$

where t is the current time, t_{upd} is the time when the last UE-based feedback in the form of LWA status report was received and $Uu_{latency}$ is the time that it takes the LWA status report to travel from the PDCP layer of the UE to the PDCP layer of the eNodeB.[3] This estimate accounts for non-ideal backhaul (Xw latency) and potential backhaul congestion, which would result in increasing delays.

B. PATH SELECTION PHASE

Using the link delay estimates d_{LTE} and d_{WLAN}, the algorithm can now define a criterion for selecting the shortest path. A PDCP PDU should be routed via WLAN if and only if

$$d_{WLAN} - \beta Xw_{latency} < d_{LTE}; \tag{8}$$

Otherwise, it should be routed through LTE.

Note that β is a parameter used to control the fairness between LWA and non-LWA UEs. If $\beta = 0$, the fastest of the two links is chosen. This minimises the PDU's RTT and optimises TCP performance for split bearers. If $\beta = 1$, the algorithm equalises the queueing delays along both paths, *i.e.*, if there is no bottleneck in the Xw interface, the algorithm equalised the load at the air interface of the LTE and WLAN links.

Let denote by l^* the selected link and by d^*_l its delay.

C. ROUTING DECISION PHASE

It is important to note that the proposed algorithm should forward PDCP PDUs to one of the available links as late as possible in order to keep the amount of PDCP PDU between the splitting point in the eNodeB and the reordering function in the UE small. This helps to avoid the following problems:

1) If the available capacity of one of the links suddenly decreases, *e.g.*, due to changing radio conditions, high reordering delays will occur at the UE.
2) If the queues overflow, packet discarding will occur and the reordering function in the UE will stop forwarding the received PDCP PDUs to the next layer

[3]Since according to our definition, the bits of a PDCP PDU leave the queueing system when they have been received by the UE and not when they are acknowledged at the eNodeB, $Uu_{latency}$ needs to be subtracted from the WLAN link delay.

until the reordering timer expires. This interruption may take several hundreds of milliseconds. For TCP applications, this will trigger TCP congestion control mechanisms and may require PDCP or TCP level retransmissions to recover the discarded packets.

In contrast, there is also the risk that the algorithm forwards PDCP PDUs too late, leading to link starvation. The algorithm should avoid cases in which SDUs are queued in the eNodeB RLC queue, while the WLAN queue runs empty and its available capacity is not used.

To address these issues, the proposed algorithm will hold a PDCP PDU at the PDCP layer, and delay its routing decision to a later point in time, if the expected link delay d_l^* of the selected link l^* is larger than a maximum queuing delay limit d_{max}, i.e., $d_l^* > d_{max}$; Otherwise, it is routed through link l^*. The maximum queuing delay limit d_{max} has to be defined properly considering the LWA status report frequency, ΔT, the maximum PDCP PDU size, PDU_{size}^{max}, the expected rate on the link l*, R_{l*}, and the Xw latency if $l* = $ WLAN. For example, if the LWA status report frequency is kept fix, when the link throughput is large, more buffering should be allowed, while when the link throughput is small, less buffering should be permitted. The maximum amount of data of a split bearer that can be buffer in the Xw interface is

$$(Xw_{latency} + d_{max}) \cdot R_{WLAN}. \quad (9)$$

V. SIMULATION RESULTS

In this section, simulation results are presented to validate the performance of the presented LWA flow control in terms of downlink capacity and UE throughput performance. The performance evaluation is conducted over an enterprise scenario of 50 m×120 m, where there is an LTE small cell eNodeB located at the centre of it, and two WLAN APs within it (see Fig. 3). Most simulation assumptions in terms of eNodeB, AP and UE deployment as well as antenna gain, path loss, shadowing and multi-path fading modelling follow the 3GPP recommendations in [7]. 100 simulation drops are performed, and in each drop 10 seconds are simulated. Please refer to [14] for a more complete description of the simulator.

a: LTE eNodeB deployment

The cell, located at the centre of the enterprise, has a transmit power of 24 dBm and deploys 10 MHz in the 1.9 GHz band. No inter-cell interference is assumed. Two omnidirectional antennas with a 5 dBi gain are considered.

b: WLAN AP deployment

2 WLAN channels of 20 MHz in the 5 GHz band are considered, and 2 AP are deployed in the enterprise where the inter-AP distance is 60 m. Each AP has a transmit power of 24 dBm, and selects upon deployment the channel in which the least load and interference is observed. Two omnidirectional antennas with a 5 dBi gain are considered.

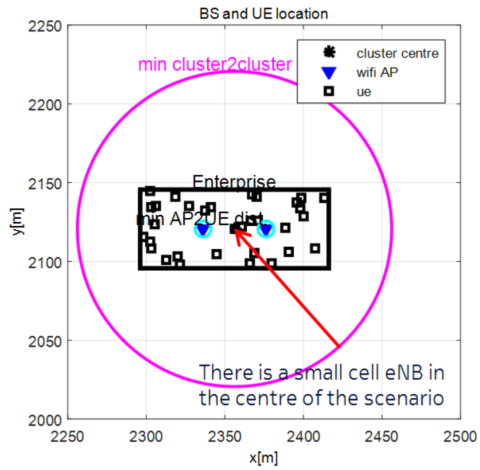

FIGURE 3. **Enterprise scenario with one LTE small cell eNodeB and two WLAN APs.**

c: UE deployment

1, 4, 20 or 32 stationary UE are uniformly deployed within the enterprise, where the minimum AP-to-UEs distance is 3 m. Each UE has a transmit power of 18 dBm, and associates to the eNodeB and the AP with the strongest pilot signal, provided that the AP pilot is detected at or above −82 dBm in the 20 MHz channel. Two omnidirectional antennas with a 0 dBi gain are considered, thus allowing 2×2 multiple input multiple output (MIMO) transmissions. Fast fading channel gains are assumed with a UE speed of 3km/h.

d: Services

All UEs use a bidirectional file transfer protocol (FTP) service (3GPP FTP traffic model 2). The FTP file size is 0.5 Mbytes (or 2 Mbytes) in the downlink and half of it in the uplink, while the mean reading time is 0.1 s. Note that TCP acknowledgments (ACKs) are generated in response to FTP traffic, where 1 TCP ACK is sent for every 3 TCP data packets.

Other relevant WLAN parameters are set as follows: DIFS= 34 μs, SIFS= 16 μs, time slot = 9 μs, TXOP=3 ms.

A. BENCHMARKED TECHNOLOGIES

Three system configurations are considered:

1) *LTE only*: All downlink and uplink traffic is carried by the LTE small cell eNodeB in the licensed band.
2) *WLAN only*: All downlink and uplink traffic is carried by the WLAN APs in the unlicensed band.
3) *LWA*: Downlink FTP traffic and downlink TCP ACKs are routed over LTE and WLAN (aggregation mode), while uplink FTP traffic and uplink TCP ACKs are routed over LTE according to Release-13 LWA. WLAN MAC ACKs remain in the WLAN network. This follows the so-called WLAN Boost configuration principle presented in [14] and [15], where WLAN DL performance are boosted by offloading WLAN UL traffic to LTE. The LWA flow control presented in Section IV is adopted.

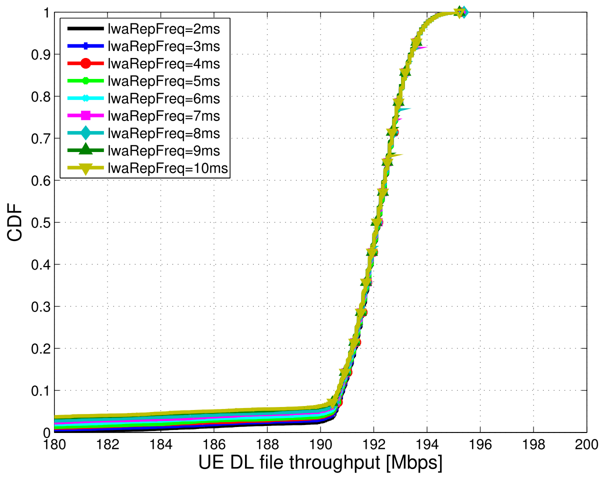

FIGURE 4. UE throughput distribution for the case of 1 user per enterprise with 0.5 MB file size.

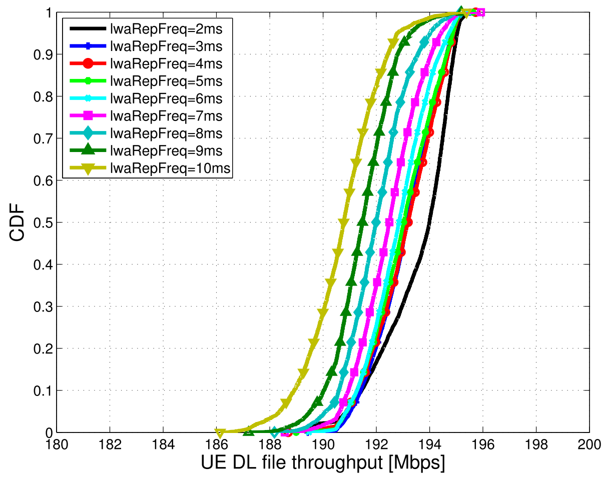

FIGURE 5. UE throughput distribution for the case of 1 user per enterprise with 2 MB file size.

Note that the LWA flow control is configured with the following parameters: $Xw_{latency} = 0$ ms (enterprise scenario), $\beta = 0$ (all UEs have LWA capabilities), $\Delta T = 5$ ms (unless otherwise stated) and $d_{max} = 30$ ms.

Note that R_l^{init} is equal to the ratio between the peak throughput of the cell divided by the number of connected users. More sophisticated methods to compute R_l^{init} that account for congestion estimation are left to further study.

B. PERFORMANCE COMPARISON

In order to access the performance of the LWA aggregation algorithm, let us first focus on the single UE case, where there is only one UE in the enterprise. For reference purposes and according to our simulations, let us note that the median peak UE throughput for this single UE case was estimated to be 63 Mbps for LTE only and 140 Mbps for WLAN only. Thus, since there is only one UE in the scenario, the resulting R_l^{init} are $R_{LTE}^{int} = 63$ Mbps and $R_{WLAN}^{int} = 140$ Mbps. Moreover, it is important to note that the maximum queuing delay limit d_{max} was set to 30 ms.

Fig. 4 and Fig. 5 show the cumulative distribution function (CDF) of the UE throughput when the downlink file size is 0.5 Mbytes and 2 Mbytes, respectively. Different LWA status report frequencies from 2 ms to 10 ms are considered.[4] When the downlink file size is 0.5 Mbytes (Fig. 4), one can observe that the LWA status report frequency does not have any impact in the performance. This is because a buffer, when $d_{max} = 30$ ms, can absorb, in one go, 0.5 Mbytes with a throughput estimation of $R_{LTE}^{int} = 63$ Mbps and $R_{WLAN}^{int} = 140$ Mbps. Therefore, a proper tuning of R^{int} is key in this case to avoid starvation. In contrast, when the downlink file size is 2 Mbytes (Fig. 5), one can see that the LWA status report frequency has an impact in the performance. This is because a buffer, when $d_{max} = 30$ ms, cannot absorb, in one go, 2 Mbytes with a throughput estimation of $R_{LTE}^{int} = 63$ Mbps and $R_{WLAN}^{int} = 140$ Mbps. In this case, the proper tuning of R^{int} is less important, and the LWA status report becomes more relevant. The more frequent the LWA status reports, the better performance of the WLAN link due to the more accurate statistics, *e.g.*, the initial phase uncertainty is avoided earlier, better adaptability to the changing channel conditions and selected modulation selections. One can tune d_{max} to adapt the buffer size to the traffic load and system setup, e.g., to avoid single shot decisions, which are risky if R_l^{init} is not properly set.

In the following, LWA status report frequency will be set to either 5 or 10 ms in line with the 3GPP allowed range which does not allow faster reporting than 5 ms in order to avoid extremely large signalling overhead. Fig. 6 shows the UE throughput distribution for the case where there are 4 UEs in the enterprise. The LTE only case provides a median UE throughput of 21.9 Mbps, while the WLAN only case provides a larger median UE throughput of 60.9 Mbps (2.7× gain). This is because the WLAN only case benefits from the following favourable deployment choices: More cells (2 instead of 1), more system bandwidth (2 × 20MHz instead of 1 × 10MHz carrier) and modulation scheme support with higher bits/symbol (256QAM instead of 64QAM). Results also show that LWA have a substantial gain over the LTE and WLAN only cases of around 6.7× and 2.4×, respectively. This is due to *i)* the offloading of UL traffic from the unlicensed to the licensed band and the resulting collision-free usage of the unlicensed spectrum for downlink (the so-called Boost effect), and *ii)* the LWA aggregation capabilities. Again, the more frequent the LWA status reports, the better the performance of the aggregation algorithm due to the better statistics, as well as the avoidance of WLAN link starvation (since the WLAN provides a large UE throughput due to the low system load, the WLAN queue for a UE may run out of packets before the next LWA status report arrives to take new scheduling decisions, creating WLAN link starvation).

[4]Note that 5 ms and 10 ms are standardised LWA status report frequencies, while all the others used here, e.g., 3 ms, are not and are investigated for research purposes.

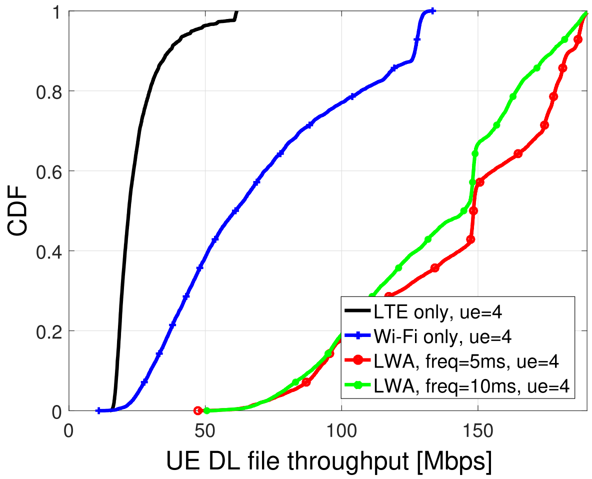

FIGURE 6. UE throughput distribution for the case of 4 users per enterprise with 0.5 MB file size.

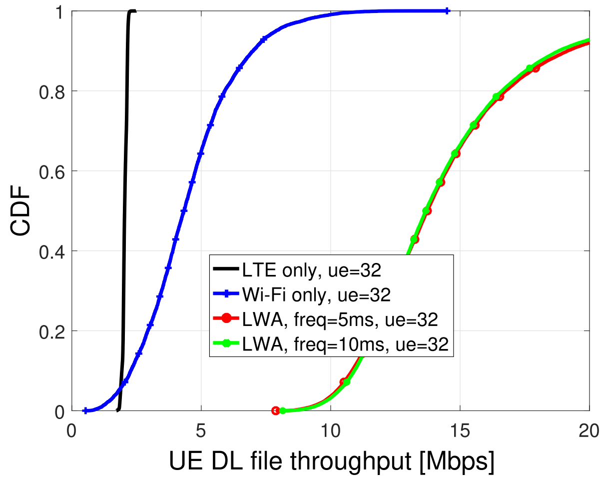

FIGURE 8. UE throughput distribution for the case of 32 users per enterprise with 0.5 MB file size.

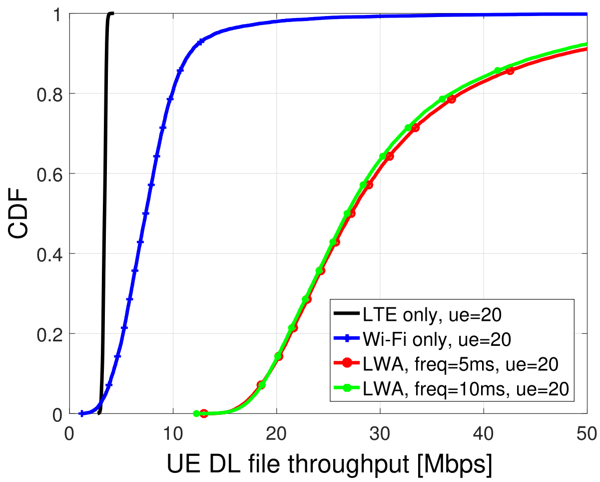

FIGURE 7. UE throughput distribution for the case of 20 users per enterprise with 0.5 MB file size.

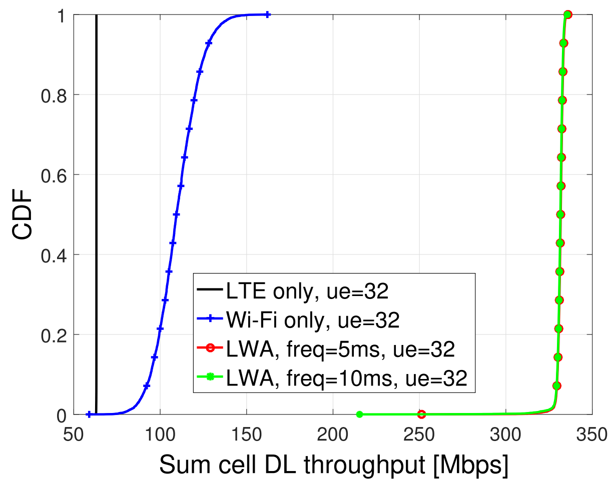

FIGURE 9. System throughput (sum of all cells) distribution for the case of 32 users per enterprise with 0.5 MB file size.

Fig. 7 shows the UE throughput distribution for the case where there are 20 UEs in the enterprise. Now, the LTE only case provides a median throughput of 3.37 Mbps/UE, while the WLAN only case provides a larger median throughput of 7.35 Mbps/UE (2.18× gain). The performance gain of WLAN over LTE in this case is smaller in comparison to the 4 UE case due to the more inefficient sharing of resources between nodes and higher contention/collision level in the former. As in the previous scenario, LWA significantly outperforms the LTE and WLAN only cases by around 8.07× and 3.70×, respectively. The gain is larger than before because the larger load and contention significantly degrades the performance of the benchmark with respect to the previous case, leaving more space for gains due to the offloading of UL traffic from the unlicensed to the licensed band. However, since UEs have a much lower throughput than before, statistics become less predictable and starvation of the WLAN link does not occur, thus the more frequent LWA status reports do not provide much gains.

Fig. 8 shows the UE throughput distribution for the case where there are 32 UEs in the enterprise. Due to the larger load, and the resulting larger contention and congestion, the gap between the performance of the LTE only and WLAN only cases reduces further (2.12× gain). This shows how the CSMA/CA mechanism becomes more and more inefficient as the traffic load increases. LWA now outperforms the LTE and WLAN only cases by around 6.73× and 1.55×, respectively. The gain is smaller than in the previous case since the larger load and contention does not degrade the performance of the benchmark at the same pace as before,[5] and because the room for aggregation vanishes as the spectrum becomes more and more loaded. Due to this larger traffic load, the more frequent the LWA status reports do not provide any gain. This is inline with the results in the previous case.

Fig. 9 shows the sum cell throughput distribution for the case where there are 32 UEs in the enterprise. The sum

[5]In this line with Bianchi's model, which shows how the rate at which WLAN network throughput decreases slows down with the increase of access attempts when this number is large [16].

cell throughput is the sum of throughput of all cells in the scenario. Since there is only one LTE eNodeB in the scenario, the LTE only case provides a median cell sum throughput of 63 Mbps, around its peak throughput thanks to good SINR conditions allowing any UE to support the highest modulation and coding scheme. Instead, the WLAN only case provides a median cell sum throughput of 110 Mbps. Downlink and uplink contention within the WLAN cell yields to decreased medium usage and therefore reduced throughputs. For the LWA case, since such contention disappears due to the uplink traffic offloading from the unlicensed to the licensed band, the system almost reaches the peak aggregated cells throughput, i.e., 63 Mbps $+ 2\times 140$ Mbps $=$ 343 Mbps. This shows how LWA can help to get the most of a combined LTE and WLAN infrastructure. As discussed before, due to the large offered load, the LWA reporting frequency has less of an impact.

VI. CONCLUSION

In this paper, we have presented the architecture of 3GPP Release 13 LWA technology and its UE-based flow control feedback. Moreover, we have also proposed a LWA flow control algorithm at the PDCP layer with downlink aggregation capabilities, which optimally splits the traffic to get the most of the licensed and unlicensed spectrum. In essence, the algorithm selects for each PDCP PDU the link with the shortest delay, taking into account the backhaul latency as well, in order to minimise the RTT of data split between LTE and WLAN. This optimises TCP performance. The capability to hold PDCP PDUs at the PDCP layer is also proposed to avoid aggravating congestion conditions. Same concepts apply to 5G networks, although the performance evaluation provided here is in the context of LTE-A Pro. Simulation results in a typical enterprise scenario show that LWA can greatly enhance user throughput performance up to 8.0× over LTE only, and 3.7× over WLAN only networks. Results also show the impact of the file and buffer sizes as well as LWA status report frequency on network performance, where a proper set up is required to avoid queue starvation as well as long waiting times in the queue.

ACKNOWLEDGEMENT

The authors would like to acknowledge Niko Kollehmainen for his contributions to this work.

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update (2013-2018)*, CISCO, San Jose, CA, USA, Feb. 2014.

[2] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078–2101, 4th Quart. 2015.

[3] M. Bennis, M. Simsek, W. Saad, S. Valentin, and M. Debbah, "When cellular meets Wi-Fi in wireless small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 44–50, Jun. 2013.

[4] "LTE-U technical report coexistence study for LTE-U SDL V1.0," Alcatel-Lucent, Ericsson, Qualcomm, Samsung, and Verizon, LTE-U Forum, Tech. Rep. V1.0 (2015-02), Feb. 2015.

[5] 3GPP, *Access Network Discovery and Selection Function (ANDSF) Management Object (MO)*, document TS 24.312 v.12.10.0, 3GPP, Sep. 2015.

[6] *Evolved Universal Terrestial Radio Access (E-UTRA) and Evolved Universal Terrestial Radio Access Network (E-UTRAN); Overall Description; Stage 2, V.13.3.0, Section 22A.1*, document TS 36.300, 3GPP, Apr. 2016.

[7] *Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum, V.13.0.0*, document TR 36.889, 3GPP, 2015.

[8] *Broadband Radio Access Networks (BRAN); 5 GHz High Performance RLAN*, Standard ETSI EN 301 893 V1.7.1, Harmonized European Standard, Jun. 2012.

[9] D. López-Pérez, J. Ling, B. Kim, V. Subramanian, S. Kanugovi, and M. Ding, "LWIP and Wi-Fi boost flow control," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, San Francisco, CA, USA, Apr. 2017.

[10] H. Wang, C. Rosa, and K. I. Pedersen, "Inter-eNB flow control for heterogeneous networks with dual connectivity," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2015, pp. 1–5.

[11] S. Singh, M. Geraseminko, S. P. Yeh, N. Himayat, and S. Talwar, "Proportional fair traffic splitting and aggregation in heterogeneous wireless networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1010–1013, May 2016.

[12] "Evolved universal terrestial radio access network (E-UTRAN); extension of dual connectivity in E-UTRAN," document TR 36.875 v.13.1.0, 3GPP, Sep. 2015.

[13] J. D. Little and S. C. Graves, *Building Intuition: Insights from Basic Operations Management Models and Principles*. Springer, 2008, ch. Little's Law, pp. 81–100.

[14] D. López-Pérez, J. Ling, B. Kim, V. Subramanian, S. Kanugovi, and M. Ding, "Boosted WiFi through LTE Small Cells: The Solution for an All-Wireless Enterprise," in *Proc. IEEE Pers. Indoor Mobile Radio Conf. (PIMRC)*, Valencia, Spain, May 2016, pp. 1–6.

[15] J. Ling, S. Kanugovi, S. Vasudevan, and A. K. Pramod, "Enhanced capacity & coverage by Wi-Fi LTE integration," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 165–171, Mar. 2014.

[16] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

DAVID LÓPEZ-PÉREZ (M'12) received the B.Sc. and M.Sc. degrees in telecommunication from Miguel Hernández University, Spain, in 2003 and 2006, respectively, and the Ph.D. degree in wireless networking from the University of Bedfordshire, U.K., in 2011. He is currently a member of Technical Staff with Nokia Bell Laboratories. He was a RF Engineer with Vodafone, Spain, from 2005 to 2006, and a Research Associate with King's College London, U.K., from 2010 to 2011. He has authored the book titled *Heterogeneous Cellular Networks: Theory, Simulation and Deployment* (Cambridge University Press, 2012), and over 90 book chapters, journal, and conference papers, all in recognised venues. He also holds over 30 patents applications. He received the Ph.D. Marie-Curie Fellowship in 2007 and the IEEE ComSoc Best Young Professional Industry Award in 2016. He was also a finalist for the Scientist of the Year prize in The Irish Laboratory Awards in 2013 and 2015. He is an Editor of the IEEE Transaction on Wireless Communications since 2016 and was awarded as an Exemplary Reviewer of the IEEE Communications Letters in 2011. He is or has also been a Guest Editor of the number of journals, such as the IEEE Journal on Selected Areas in Communications, and the IEEE Communication Magazine.

DANIELA LASELVA received the M.Sc. degree in electrical engineering from the Polytechnic University of Bari, Italy, in 2002. From 2002 to 2006 she was active in COST273 and EU FP6 Project WINNER on MIMO Radio Channel Modeling with Elektrobit, Finland, and worked as a Senior Design Engineer with Nokia, Finland. Since 2006, she has been with Nokia Aalborg, Denmark, currently Nokia Bell Laboratories, where she is a Senior Researcher. Her current work is mainly related to concept research and standardization of LTE-Advanced Pro and 5G systems, including IoT wireless systems and end-to-end performance optimization. She has (co-)authored a few tens of scientific publications, has contributed to a couple of book chapters, and is the inventor of several patents. Her research interests and previous work embrace radio aggregation and interworking with WLAN, cellular operations in unlicensed spectrum, self-organizing networks (SON), topic developed within the EU FP7 project SEMAFOUR, traffic steering in heterogeneous networks, radio resource management, QoS provisioning, control channel optimization, and performance evaluation.

EUGEN WALLMEIER received the Diploma degree and the Ph.D. degree in mathematics from the University of Münster, Germany, in 1980 and 1983, respectively. In 1985, he was with the Public Communication Networks Group of Siemens AG, Munich, where he was engaged in SW development and performance analysis for public switching systems. In 1992, he was a Supervisor of a group, where he was involved in ATM traffic and performance issues and on the design of traffic control functions for ATM switches. In 2000, he was with the Mobile Radio Division of Siemens, where he has been involved in designing new controllers for 2G and 3G mobile networks. Since 2007, he has been the Head of various systems engineering teams with the Mobile Radio Business Lines of Nokia Siemens Networks and Nokia, respectively. His current man working area is requirements engineering for transport features in LTE networks. In 1980, he became a member of the Scientific Staff with Institute for Mathematical Statistics, University of Münster.

PÄIVI PUROVESI received the M.Sc. degree in computer science from the University of Helsinki, in 1990. She was with Nokia Research Center in 1989, and has worked at Nokia since then. She was also with the University of Helsinki in several positions before graduation for five years. She is currently a Senior Specialist with Nokia Mobile Networks. She contributed many research and product projects related to cellular networks of different technologies, for example GSM, GPRS/EGPRS, LTE, and 5G. Her main focus has been in user plane protocols and real-time embedded systems. She is co-author of two granted patents.

PETTERI LUNDÉN received the M.Sc. degree in computer and information science from the Helsinki University of Technology, Finland, in 2004. He was with Nokia Research Center, in 2002 and has been with Nokia since then he was involved in design, standardization, and performance analysis of wireless communication systems. He is currently a Research Specialist, Radio Research with the Nokia Bell Laboratories, Espoo, Finland. He has (co-)authored about 30 book chapters, journal, and conference papers and is the (co-)inventor of over 70 patent applications. His research interests include mobility and radio resource management solutions in LTE, MulteFire, and 5G.

ELENA VIRTEJ received the Dipl.Ing. from the Politehnica University of Bucharest, Romania, in 2004. She was with Nokia Research Center, Finland, in 2003 and ever since she has been with Nokia. She is currently with the Nokia Bell Laboratories, Espoo, Finland, as a Senior Specialist, Radio Research. She has expertise in system design and system performance evaluation, especially in areas of mobility and radio resource management for LTE licensed assisted access, dual connectivity, and heterogeneous networks. She has (co-)authored several conference papers and over 50 patent applications in the domain. Her current research interests include LTE, MulteFire, and 5G.

PIOTR LECHOWICZ received the M.S. degree from the Faculty of Automatic Control, Electronics and Computer Science, Politechnika Śląska. Since early nineties has been with the telecom industry in Poland, Lebanon, and Belgium. He is currently accountable for Nokia eNB specifications in the PDCP and the RLC areas, and has two patent applications pending.

ESA MALKAMAKI (M'97) received the M.Sc. (Tech), Lic.Sc. (Tech), and D.Sc. (Tech) degrees from the Helsinki University of Technology, in 1989, 1992, and 1998, respectively, all in electrical and communications engineering. He was with Communications Laboratory, Helsinki University of Technology, participating in RACE Mobile Project, since 1988. He was also with Nokia Research Center in 1992, where he was involved in RACE ATDMA Project. Since 1998, he has been with standardization research, first with ETSI (EGPRS) and since 1999 with 3GPP (WCDMA, HSPA, LTE, LTE-A, 5G), working first with physical layer and recently with higher layer protocols (MAC, RLC, PDCP, RRC) with carrier aggregation, dual connectivity, license assisted access (LAA), latency reductions, and 5G user plane. He is currently a Senior Specialist, Radio Research with Nokia Bell Laboratories, Espoo, and participates 3GPP RAN2 meetings as a Standards Delegate. He has published over 30 scientific papers and has been granted several patents over 50 patent families.

MING DING (M'12) received the B.S. and M.S. degrees (Hons.) in electronics engineering and the Ph.D. degree in signal and information processing from Shanghai Jiao Tong University, Shanghai, China, in 2004, 2007, and 2011, respectively. He is currently a Senior Research Scientist with Data61, CSIRO, Australia. He has authored about 40 papers in the IEEE journals and conferences, all in recognized venues, and about 20 3GPP standardization contributions, and a book titled *Multi-Point Cooperative Communication Systems: Theory and Applications* (Springer). Also, as the first inventor, he holds 15 CN, 7 JP, 3 US, 2 KR patents and co-authored another 100+ patent applications on 4G/5G technologies. He is or has been Guest Editor/Co-Chair/Co-Tutor/TPC Member of several IEEE Top-Tier Journals/Conferences, e.g., the IEEE Journal on Selected Areas in Communications, the *IEEE Communications Magazine*, and the IEEE Globecom Workshops. For his inventions and publications, he was a recipient of the President's Award of Sharp Laboratories of China in 2012, and served as one of the key members in the 4G/5G Standardization Team when it was awarded in 2014 as Sharp Company Best Team: LTE 2014 Standardization Patent Portfolio.