Aggregating LTE and Wi-Fi: Towards Intra-cell Fairness and High TCP Performance

Boram Jin, Segi Kim, Donggyu Yun, Hojin Lee, Wooseong Kim, and Yung Yi, *Member, IEEE*

Abstract—The data explosion and resource scarcity of mobile cellular networks require new paradigms to effectively integrate heterogeneous radio resources. Of many candidate approaches, smart aggregation of LTE and Wi-Fi radios is a promising solution that bonds heterogeneous links to meet a mobile terminal's bandwidth need. Motivated by the existence of a significant number of carrier-operated Wi-Fi APs, we propose a mechanism, called *LTE-W,* which efficiently utilizes LTE and Wi-Fi links only with the minimum change of eNodeBs, LTE backhaul networks, and mobile terminals, thus easily deployable. LTE-W, which is a link-level aggregation mechanism, has the following two key components: (i) mode selection and (ii) bearer-split scheduling. First, in the mode selection LTE-W internally decides who should be served by either of LTE-only or LTE-Wi-Fi aggregation considering intra-cell fairness rather than just following users' intention of aggregation. For the users' preference to be offered the aggregation service, we choose a bearer (roughly defined in LTE as a set of flows with a similar QoS) as a basic unit of aggregation and propose a smart intra-bearer scheduling algorithm that splits a bearer's traffic into LTE and Wi-Fi links, considering the performance of TCP flows that take two heterogeneous wireless links. We evaluate our mechanism using the NS-3 with LENA, under various configurations including nodes with mobility and HTTP traffic, and compare it to a transport-level aggregation mechanism, MPTCP, demonstrating that LTE-W significantly improves MPTCP, e.g., up to 75% in terms of Jain's fairness index.

Index Terms—5G, LTE-WiFi aggregation, fairness, TCP performance, scheduling;

I. INTRODUCTION

Cisco forecasts that the amount of mobile data has grown exponentially up to eight-fold within 5 years (2015 - 2020), and in 2020, portion of the large file sharing services (e.g. Internet video) will be over 75 percents of all mobile data traffic [1], [2]. Thus, for individual mobile users, Mobile Network Operators (MNOs) try to find methods to solve scarcity of resource problem, and integrating LTE/Wi-Fi could be one of the solutions. The idea is that when there does not exist a single radio access technology (RAT) that offers sufficient bandwidth to meet an application's requirement (e.g., users at the cell edge), two or more RATs are merged so that the application is able to experience a scaled-up capacity [3]–[5]. In fact, it is expected to be a key ingredient of the next-generation 5G wireless to efficiently use heterogeneous wireless networks in an integrated manner [6]. Two candidate RATs to aggregate are LTE and Wi-Fi, due to their popularity in the state-of-the-art mobile devices, which are also the focus of this paper.

There exists an extensive array of research and development efforts both in the academic and industrial domains, where a variety of approaches are taken at different layers (see the related work in Section II) and their unique pros and cons exist. A lot of research efforts on opportunistically using Wi-Fi APs mostly for delay-tolerant applications have been made in the name of Wi-Fi offloading, e.g., [7], [8]. Other notable examples to aggregate LTE and Wi-Fi, rather than opportunistically using one of them, include MPTCP (Multi Path TCP) at the transport layer [9]–[11]. This paper is motivated by the fact that a significant portion of Wi-Fi APs are being deployed by MNOs[1], referred to as MNO-operated Wi-Fi, enabling many mobile users to be under the coverage of both LTE and Wi-Fi. This may provide more opportunities to MNOs to tightly optimize the aggregation of LTE and Wi-Fi links at the link-level for higher efficiency, but without much change of the current LTE architecture.

We summarize our contributions in what follows:

(a) Fairness, split-scheduling, and their impact on TCP performance: Unconditionally providing the aggregation service to all aggregation-requested users may lead to serious intra-cell unfairness, hurting the system-level QoS and suboptimally utilizing the system resource (see a motivating example in Section III-C). Also, in splitting the packets inside a bearer (which is the basic unit of aggregation, as discussed later) served by two highly heterogeneous links (LTE and Wi-Fi), we need to consider the compatibility of such splitting mechanism with TCP, since TCP's protocol features are highly sensitive to out-of-order packets which could be affected by the other incoming packet stream patterns. In this paper, we take a time-scale separation approach that we first decide a *mode* of a bearer (i.e., either of LTE-only or LTE-W mode), whenever a new bearer is created, using the solution of an utility maximization problem that formulates the (intra-

[1]For example, KT, which is one of the biggest three MNOs in South Korea, have already about 200 million APs [12].

cell) fairness. We prove that our proposed mode selection algorithm outputs the optimal solution in polynomial time. This intra-cell fairness provisioning is the major gain coming from a link-level aggregation, forming the key difference from MPTCP. In MPTCP, LTE and Wi-Fi subflows would achieve the throughputs [2] depending only on available path-bandwidths (a LTE path and another Wi-Fi path), often ignoring intra-cell fairness (see Section IV). Once the modes of all bearers are decided, eNodeB then performs split-scheduling that strips the incoming packets inside a LTE-W bearer into two links, so that TCP flows inside the bearer experience good throughput at the receiver. This "infrequent" decision of a bearer mode allows our design to separately focus on intra-cell fairness and flow-level TCP performance, thereby leading to a simple, yet efficient design of a link-level aggregation service.

(b) Architecture design: We propose an architectural design, called *LTE-W,* that aims at achieving the key features mentioned above, as depicted in Fig. 1. The key direction of our design is to achieve our goal with the minimum change of the current LTE implementation, e.g., a simple software upgrade. To this end, we use a bearer (a group of flows with a similar QoS defined in LTE [13]) as a basic aggregation unit in order to minimize the eNodeB modification [14] as well as avoid high complexity per-flow based processing. In LTE, the Packet Data Convergence Protocol (PDCP) layer is responsible for handling bearers, and thus our bearer split-scheduling is also proposed to be implemented at the PDCP layer. We also propose a modification of Wi-Fi AP MAC architecture (that are MNO-operated) to employ per-bearer queueing (i.e., separate queues for LTE-W bearers), which is also implementable by a simple software upgrade, so that the bearer-splitting function at an eNodeB is operated in a more predictable manner. This per-bearer queueing enables the system to provide the predictable Wi-Fi throughput, which is importantly used in our mode selection decision module, as well as to sustain more stable behavior of TCP flows than that with only FIFO queueing.

(c) NS-3 LENA implementation and evaluation: We implement our LTE-W design by extending NS-3 LENA [15], [16], and evaluate LTE-W under various scenarios including mobile users with HTTP traffic in order to generate realistic situations. We compare LTE-W with MPTCP and demonstrate that TCP flows inside a LTE-W bearer achieve stable throughputs, and more importantly, LTE-W outperforms MPTCP in terms of the system-wide Jain's fairness index by 75%, and verify that our proposed bearer split mechanism achieves high link utilization.

II. Related Work

One way of smartly using LTE and Wi-Fi links is to adaptively select either of those, being categorized into network-driven [17]–[20] and user-driven [3], [4] approaches. For example, the issue of deciding who should use which radio access technology is formulated by an NP-hard optimization problem, where a greedy-like algorithm [19] or a low-complexity distributed algorithm [20] are proposed. As an

array of closely related work to this paper, there exist various proposals on aggregating LTE and Wi-Fi links. In the International Telecommunication Union's (ITU) Plenipotentiary Conference 2014 in Busan, Samsung demonstrated Download Booster on Galaxy S5 as a bandwidth aggregation in the application layer and SKT and KT show demonstrations of the commercial deployment of Multipath TCP (MPTCP) [21], [22]. In academy, various MPTCP methods have been researched to support high-quality video streaming [23]–[26], where [25], [26] proposed protocols to efficiently utilize mobile energy conservation and bandwidth, respectively.

MPTCP [9]–[11] utilizes multiple physical *paths* simultaneously to improve throughput and resilience, but lacks in providing the intra-cell fairness (see Section IV for details). In 2015 Mobile World Congress at Barcelona, KT demonstrates the LTE-H (LTE-HetNet) using Samsung's LTE base station and Wi-Fi AP, and Qualcomm's modem chip for mobile devices based on LTE PDCP layer bandwidth aggregation. Recently, the third-Generation Partnership Project (3GPP) defines LTE-WLAN Aggregation (LWA) [27]–[29] and the specific LWA bearer in Release 13 [27] (formally completed in March 2016).

In 3GPP, as a different way of utilizing unlicensed spectrum, LTE-U or LAA (Licensed-Assisted Access), which applies LTE carrier aggregation to unlicensed spectrum, has been proposed and standardized in 3GPP Releases 13 and 14. To compare LTE-U/LAA with LTE-W of this paper or LWA, there exist pros and cons for both approaches. LTE-U/LAA is expected to provide more natural aggregation between licensed and unlicensed spectrums at the physical layer due to the homogeneity of access technology, but its coexistence at unlicensed spectrum with other access technologies, e.g., Wi-Fi, still remains to be solved. A couple of approaches such as Listen Before Talk (LBT) or Carrier Sense Adaptive Transmission (CSAT) [30], [31] in Qualcomm [32], [33] and Ericsson [34] have been proposed, but it still remains to study what is the best mechanism to fairly share unlicensed spectrum among competing access technologies. LTE-W or LWA is designed with minimum changes of the current LTE and Wi-Fi systems, but integrating different access technologies may entail the performance loss incurred by merging two heterogeneous access technologies. This paper aims at proposing novel mechanisms in terms of intra-cell fairness and split scheduling for graceful aggregation of LTE and Wi-Fi, so upper-layer protocols enjoy the aggregation gain as much as possible.

The general goal and practical consideration of LWA are similar to LTE-W of this paper, e.g., RAN aggregation in the downlink connections and per-packet scheduling based on the feedback and measurements from LTE and WLAN systems on PDCP layer. To the best our knowledge, LWA only specifies how and where LTE and Wi-Fi are integrated, but does not specify the core algorithmic components such as what we do in this paper: how to provide intra-fairness and bearer-splitting mechanism. We can find an analogous case in *user scheduling* algorithms in 3G/4G cellular systems. There are many proposals on user scheduling such as Proportional-fair scheduler, MAX-SNR scheduler, etc, which are not specified in the 3G/4G standard, but rather vendor-specific. Similarly,

---

[2]We abuse a term "throughput" to denote the throughput value through this paper.
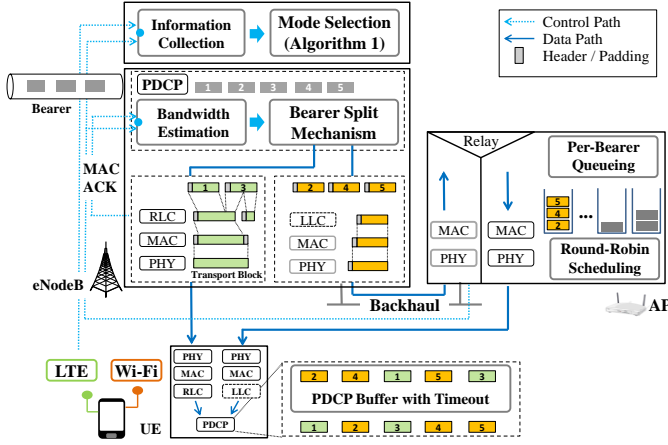
Fig. 1: LTE-W architecture overview.

LWA does not specify the algorithms on how to split bearers and under what criterion a user should be associated to both LTE and WiFi. LWA only provides where this aggregation occurs, and specifies the required field and frame format changes.

III. LTE-W DESIGN

A. LTE Background

Bearer. A bearer, which is a unit of traffic management, is a tunnel connection between User Equipment (UE) and Packet-data-network GateWay (PGW), where multiple flows are grouped in a bearer based on the level of QoS provided by the Policy and Charging Rules Function (PCRF) [13]. Throughout the initial certification processes, the default bearer is established. When a UE requests a service which requires higher Quality of Service (QoS), then a dedicated bearer is established on demand. Depending on the QoS of the requested service, each dedicated bearer is classified as a Guaranteed Bit Rate (GBR) bearer and a Non-GBR bearer. LTE-W does not support the Guaranteed Bit Rate (GBR) bearer, which may be served through LTE system, because the tight resource scheduling in LTE guarantees more than just a long-term throughput. However, as shown in Table 6.1.7 of [13], Constant Bit-Rate (CBR) applications such as live streaming or buffered streaming services are categorized as Non-GBR, so that our proposed scheme could be applied to CBR applications.

LTE user plane protocol stack. User plane protocol stack in LTE has four commonly related layers in eNodeB and UE: (a) Packet Data Convergence Protocol (PDCP), (b) Radio Link Control (RLC), (c) Medium Access Control (MAC), and (d) PHY. PDCP is responsible mainly for IP header compression and ciphering, and supports lossless mobility in case of inter-eNodeB handovers and provides integrity protection to higher layer control protocols. RLC supports data segmentation and concatenation to fit the size required by the MAC (mostly transport block size), and performs Automatic Repeat reQuest (ARQ). Indeed, RLC has three modes for data transmission: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). The main functions of MAC are Hybrid-ARQ (HARQ) and reporting of scheduling information. We refer the readers to [29], [35] for more details.

B. LTE-W operation overview

Modes. We consider a scenario where a UE that is under the coverage of both LTE and Wi-Fi, has three options to choose in each device as follows: **M1.** Wi-Fi only, **M2.** LTE-only, and **M3.** LTE-W. Both **M1** and **M2** are the modes that are available at the state-of-the-art smart devices in the market. The mode **M3** is assumed to be added due to our LTE-W service. From the perspective of a LTE-W service provider, the major target to be considered is **M2** and **M3**, where the UEs choosing **M1** are treated as exogenous ones. Note that we consider the case where, even if a UE chooses the LTE-W mode, the UE is *not* guaranteed to be served by both LTE and Wi-Fi, which is determined by the operator, probably the operator's intention of achieving some goals such as intra-cell fairness. The operator decides this based on the result coming from its mode selection module (see Section III-C).

Operating procedures. We now describe a summary of LTE-W operation procedures, see the architecture depicted in Fig. 1. As an initial procedure, an eNodeB manages the basic information on the UEs and the MNO-operated APs under its cell coverage (e.g., MAC/IP addresses and SSID, etc).

(a) *UE's choice of mode:* Whenever each UE expresses and often changes its choice out of two modes, **M2** and **M3**, it is reported to the eNodeB, and the RRC layer of the eNodeB records the *bearer-information,* such as the mode preference, bearer ID, LTE and Wi-Fi throughputs that are estimated when each UE will be connected (see Section III-C for more details).

(b) *Mode selection and bearer split-scheduling:* Using the recorded bearer-information, the eNodeB performs mode selection in the RRC layer and bear split-scheduling in the PDCP layer. Note that there exist a variety of ways of where to perform mode selection, e.g., Mobility Management Entity (MME) or some centralized server in the managed LTE infrastructure [6]. The role of bearer split-scheduling is to strip the incoming PDCP Packet Data Units (PDUs) in a LTE-W bearer directly into LTE, or into Wi-Fi links by forwarding the PDUs to the AP that serves the corresponding UE.

In the rest of this section, we describe two key components of LTE-W, (i) mode selection in Section III-C and (ii) bearer split-scheduling in Section III-D. We assume that eNodeBs are connected to MNO-operated APs by a high-speed backhaul link, which holds in practice in the most of MNOs' infrastructure in the present market.

C. Mode Selection

Motivating example. To intuitively understand why mode selection is necessary, consider a simple fairness concept having equal throughput value among users in a simple network with one eNodeB and one Wi-Fi AP, where the Wi-Fi network has capacity 8 Mbps. We assume that two users exist and their individual LTE throughputs are (9 Mbps, 1 Mbps), respectively, and their Wi-Fi throughputs are the same, i.e., 4 Mbps per user. These different LTE thorughputs are mainly due to their location/channel conditions and the employed user scheduler in the eNodeB, e.g., proportional-fair scheduling [36]. If all

TABLE I: Used Notations.

For OPT and Algorithm 1	
M, N, J	Set of Wi-Fi APs, set of UEs, and index set of bearers, respectively.
(i, j)	j-th bearer of UE i.
b_{ij}	1 if bearer (i, j) is opened, and 0 otherwise.
p_i	1 if UE i chooses LTE-W, and 0 otherwise.
$B_0(B_1)$	Set of opened bearers excluded from (eligible to be served by) LTE-W service.
$\boldsymbol{x} = (x_i^k)$	Physical association vector of UE $i \in N$ to AP $k \in M$.
λ_{ij}	1 if bearer (i, j) is decided for the LTE-W service, 0 otherwise.
γ_{ij}	Actual throughput of bearer (i, j).
L_{ij}(resp. W_{ij})	LTE (resp. Wi-Fi) throughput for bearer (i, j).
C^k	Capacity of AP $k \in M$.

For Algorithm 2 ($i \in \{$L (LTE), W (Wi-Fi)$\}$)	
B_i	Per-bearer available bandwidth over link i.
S_i	Total bytes of unAcked PDUs over link i.
D_i	Total delay of link i.
C_i	Processing and propagation delays of link i.
size(p)	Size (byte) of packet p.
T(p)	Scheduled time when packet p departed from PDCP layer of eNB.
RTT(p)	Round-trip-time of packet p.

users request the aggregation service, and the system agrees in an unrestricted manner, then the total throughputs of both users become (13 Mbps, 5 Mbps), respectively. Now, consider the case where only the user with 1 Mbps LTE throughput is provided the aggregation service. Then, the total throughputs of both users are (9 Mbps, 9 Mbps), which seems more fair than when both users receive the aggregation service.

Model and formulation. We now describe our mode selection algorithm, called **ModeSel**, which runs whenever a new bearer is created and at periodic intervals determined by the characters of users in a cell (e.g., mobility pattern). For ease of exposition, we consider a system as presented next when a new bearer is created. The notations used in this paper are summarized in Table I.

We consider a single cell, where there exist a set M of Wi-Fi APs and one eNodeB, and a set N of UEs. Let J be an index set of bearers, where $|J|$ is the maximum number of bearers that a UE can open[3], and we denote by (i, j) the j-th bearer of UE i. Let $\boldsymbol{b} = (b_{ij} : i \in N, j \in J)$ be the 0-1 matrix that represents which bearers are opened, i.e., $b_{ij} = 1$ when the bearer j is opened at UE i, and 0 otherwise. We also use $\boldsymbol{p} = (p_i : i \in N)$ to denote the user preference for LTE-W service, i.e., $p_i = 1$ (resp. $p_i = 0$) when a UE chooses LTE-W (resp. LTE-only). Then, for a given $\boldsymbol{p}$, the entire bearers can be expressed as a union of three disjoint sets $B_0 = B_0(\boldsymbol{b}, \boldsymbol{p}) = \{(i, j) \mid p_i = 0, b_{ij} = 1\}$, $B_1 = B_1(\boldsymbol{b}, \boldsymbol{p}) = \{(i, j) \mid p_i = 1, b_{ij} = 1\}$, and $B_2 = B_2(\boldsymbol{b}, \boldsymbol{p}) = \{(i, j) \mid b_{ij} = 0\}$. In other

[3]In LTE-A, $|J| = 8$ [37].

words, B_0 is the set of all bearers that are excluded from the LTE-W aggregation service, B_1 is the set of all opened bearers that are *eligible* for both LTE and Wi-Fi, and B_2 is the set of bearers that are not opened. Let $\boldsymbol{x} := (x_i^k : i \in N, k \in M)$ be the physical association vector of UEs for both LTE and Wi-Fi, where $x_i^k = 1$ when UE i is associated with Wi-Fi AP $k \in M$. Denote by B_1^k the set of all opened bearers of UEs who has LTE-W preference and physical association with AP k. For notational convenience, we use $k(i)$ to refer to the UE i's associated Wi-Fi AP, and $k(i) = 0$ when it is only associated with LTE. We let $(C^k, k \in M)$ be the link-level capacity of AP k. We specially use the superscript 0 to refer to eNodeB, so $x_i^0 = 1$ when associated with LTE. We assume that each UE is always associated with the eNodeB, but may not have Wi-Fi connection. Thus, we have $1 \le \sum_{k=0}^{m} x_i^k \le 2$ for all $i \in N$.

We consider the following optimization problem **OPT** that dertermines which bearers will be served LTE-W:

$$\textbf{OPT:} \quad \max \sum_{i,j} U(\gamma_{ij}),$$

$$\text{subject to } \lambda_{ij} = 0 \quad \text{if } (i, j) \in B_0 \cup B_2, \quad (1)$$

$$\gamma_{ij} = \begin{cases} L_{ij} & \text{if } (i, j) \in B_0, \\ L_{ij} + \lambda_{ij} \cdot W_{ij}(\boldsymbol{\Lambda}) & \text{if } (i, j) \in B_1, \\ 0 & \text{if } (i, j) \in B_2, \end{cases} \quad (2)$$

$$\text{constants } \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{x}, (L_{ij} : i \in N, j \in J), (C^k : k \in M),$$

where variables $\boldsymbol{\Lambda} = (\lambda_{ij} \in \{0, 1\} : (i, j) \in B_0(\boldsymbol{b}, \boldsymbol{p}) \cup B_1(\boldsymbol{b}, \boldsymbol{p}) \cup B_2(\boldsymbol{b}, \boldsymbol{p}))$. We add interpretations on **OPT** in the following:

○ $U(\cdot)$ is a utility function that satisfies the standard conditions in literature, i.e., concavity, differentiability, and monotonicity. One can consider the famous $\alpha-$fair utility function [38]: $U(\gamma_{ij}) = \gamma_{ij}^{1-\alpha}/(1 - \alpha)$ for $\alpha \neq 1, 0 \le \alpha$, and $U(\gamma_{ij}) = \log(\gamma_{ij})$ for $\alpha = 1$. The $\alpha-$fair utility function is related to the well-known fairness, such as proportional fairness ($\alpha = 1$), and max-min fairness ($\alpha \to \infty$).

○ The variable $\lambda_{ij} \in \{0, 1\}$ for $(i, j) \in B_1$ represents whether the bearer (i, j) is served by both LTE and Wi-Fi ($\lambda_{ij} = 1$) or not ($\lambda_{ij} = 0$). We call $\boldsymbol{\Lambda} = (\lambda_{ij})$ *LTE-W service vector.* The constraint (1) means that $\lambda_{ij} = 0$ for each uninterested user i in LTE-W and unopened bearers in B_2.

○ The constants $(L_{ij} : i \in N, j \in J)$ correspond to the given throughputs of LTE for a given set of opened bearers, which are determined by the underlying resource allocation scheme of a LTE enodeB, e.g., user scheduling.

○ $W_{ij}(\boldsymbol{\Lambda})$ is the Wi-Fi throughput of LTE-W bearer (i, j) if it is decided to be served by Wi-Fi. Note that $W_{ij}(\boldsymbol{\Lambda})$ is the function of the LTE-W service vector due to its dependence of how many UEs are associated with the corresponding AP. It also depends on a LTE-W designer who may or may not provide differentiated service from other normal Wi-Fi users to LTE-W service users. Despite such dependence, it seems reasonable to assume that $W_{ij}(\boldsymbol{\Lambda})$ is determined by the AP capacity $C^{k(i)}$ and the number of (actually) associated bearers, say n, i.e., $W_{ij}(\boldsymbol{\Lambda}) = W_{ij}(C^{k(i)}, n)$.

Algorithm 1: ModeSel

1: **INPUT**
   $\boldsymbol{x}$: Physical association vector of UEs
      and $\boldsymbol{p} = (p_i : i \in N)$ LTE-W preference vector,
   $(C^k : k \in M)$: Link layer capacity of AP k,
   $(L_{ij} : i \in N, j \in J)$: LTE throughput of bearer (i,j).
2: **OUTPUT** $\boldsymbol{\Lambda} = (\lambda_{ij} : i \in N, j \in J)$

3: **Initialization:** $\boldsymbol{\Lambda} = \boldsymbol{0}$.
4:   **for** AP $k \in M$ **do**
5:     Sort the bearers $(i,j) \in B_1^k$ in ascending order of L_{ij}.
6:     **for** $n \in [1 : |B_1^k|]$ **do**
7:       Select n bearers in B_1^k at the front of the sorted set.
8:       $U_n = \sum_{\text{selected } n \text{ bearers}} U(L_{ij} + W_{ij}(C^k, n))$
9:     **end for**
10:     $n^\star = \arg\max_n U_n$
11:     Set $\lambda_{ij} = 1$ for the bearers (i,j) that achieves $n^\star$.
12: **end for**

○ Thus, in the constraint (2), γ_{ij} corresponds to the actual throughput of the bearer (i,j) after $\boldsymbol{\Lambda}$ is decided. Also, the constraint covers for the trivial case of unopened bearers in B_2 as $\gamma_{ij} = 0$.

ModeSel algorithm and optimality. We now describe our algorithm, which we call **ModeSel**, in **Algorithm 1**, followed by the explanation of how it operates. In **ModeSel**, we first decompose **OPT** from the perspective of each Wi-Fi AP k as follows:

$$\max \sum_{i,j} U(\gamma_{ij}) = \max \left[\sum_{(i,j) \in B_0} U(\gamma_{ij}) + \sum_{(i,j) \in B_1} U(\gamma_{ij}) \right] \tag{3}$$

$$\leq \sum_{(i,j) \in B_0} U(\gamma_{ij}) + \sum_{k \in M} \left[\max \sum_{(i,j) \in B_1^k} U(\gamma_{ij}) \right].$$

From this decomposition, to solve **OPT**, it suffices that each AP k finds the optimal λ_{ij} for its associating UEs. Using this decomposition in (3), for each AP k, we select which UE should be served by LTE-W. In Lines 4-5, for each AP k, the LTE throughputs of UEs associated with AP k are sorted in ascending order. For each n with $1 \leq n \leq |B_1^k|$, we select n bearers that have smallest LTE throughputs (Line 7), and calculate the total utility U_n of the sum of LTE throughputs and Wi-Fi throughputs, denoted by $W_{ij}(C^k, n)$ in (4), if the selected n bearers are also served over Wi-Fi (Line 8), we choose $n^\star$ that maximizes U_n. Finally, we decide to serve the bearers (i,j) that outputs $n^\star$ (Line 10). **ModeSel** has the worst-case $O(|M||J|^2|N|^2)$ time complexity, because $|B_1^k| \leq |N||J|$, where recall that $|J|$ is the maximum number of bearers that a UE can open, thus $|J||N| \log |J||N|$ is the complexity for sorting.

Per-bearer queueing. The Wi-Fi throughput $W_{ij}(C^k, n)$ in (4) depends on the designer of LTE-W, in particular, in relation to how to treat LTE-W bearers in Wi-Fi AP. In our design, we propose the rule, called *per-bearer queueing with LTE-W*

priority that each bearer with the LTE-W service is assigned a separate queue and all other normal flows are served in a FIFO queue. Under this policy,

$$W_{ij}(C^k, n) = \frac{C^k}{n+1}, \tag{4}$$

where n refers to the LTE-W service bearers and '1' for all other normal Wi-FI flows. We believe that this queueing policy is plausible, because Wi-Fi APs in this paper are operated by MNOs, and they try to maximize their revenue to provide a better QoS to the LTE-subscribing users who make monthly payment. It is typical that Wi-Fi services provided by a MNO are additional, often used just to obtain more market share in the competition with other MNOs. This per-bearer queueing is also beneficial in the bearer split-scheduling in Section III-D, contributing to predictable throughput estimation, as demonstrated in Section IV.

LTE throughput L_{ij}: The algorithm operates assuming L_{ij} and C^k are given. In practice, it is reported that over 70% of bearers contain only one TCP flow, and almost 50% of flows are shorter than 5.0 sec [39]. As we consider the patterns of mobile phone users, a flow is persistently transmitted in bearer for long time including only one flow. Thus, in our mechanism, we estimate the LTE throughput in a fixed time less than 5.0 sec (in simulation we consider 3.0 sec). These measured LTE throughputs for each bearer are used as L_{ij} in **ModeSel**. This L_{ij} measurement is practically valuable, because it renders **ModeSel** independent of the underlying resource allocation in LTE. For example, user scheduling in LTE is highly vendor-specific.

Optimality analysis. **OPT** is an integer program, which is in many cases NP-hard. However, as stated in Theorem 1, **OPT** outputs an optimal solution in polynomial time.

Theorem 1. *Under the per-bearer queueing policy at APs, **ModeSel** outputs an optimal solution of **OPT**.*

Proof. Since $(L_{ij} : i \in N, j \in J)$ is given and APs are not coupled (due to our assumption that each UE is associated with one AP), it is sufficient to solve the following per-AP optimization:

$$\textbf{Per-AP OPT:} \qquad \max \sum_{(i,j) \in B_1^k} U(\gamma_{ij}).$$

In this proof, we assume that we focus on an arbitrary AP k. We will prove that U_{n^*} is larger than or equal to the total utility for any other combinations of UEs in B_1^k, for which for any given $n \leq |B_1^k|$, we will prove that the total utility for the n smallest LTE throughput bearers exceeds that for any n selection of bearers.

For a given n, let $G_n = \{g_1, g_2, \ldots, g_n\}$ be the set of LTE throughputs of n-smallest bearers, and consider an arbitrary set $S_n = \{s_1, s_2, \ldots, s_n\}$ of n LTE throughputs of any n bearers. Without loss of generality, $g_i \leq g_j$ and $s_i \leq s_j$ for all $i < j \leq n$. Let $A = G_n \cap S_n$, and $|A| = a$, where let

$$G_n' = G_n \setminus A = \{g_1', g_2', \cdots, g_{n-a}'\},$$
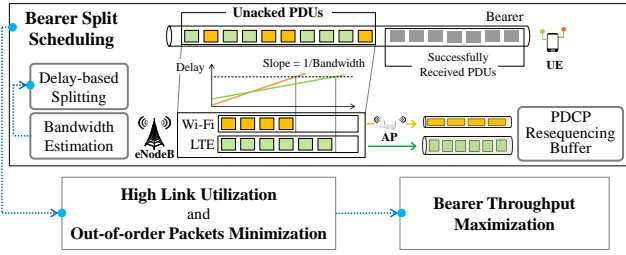$$S_n' = S_n \setminus A = \{s_1', s_2', \cdots, s_{n-a}'\},$$

Fig. 2: Overview of bearer split-scheduling.

Due to the concavity of $U(\cdot)$, it is easy to check that for any $1 \leq j \leq n - a$,

$$\left(U(g_j' + \frac{C^k}{n+1}) - U(g_j')\right) - \left(U(s_j' + \frac{C^k}{n+1}) - U(s_j')\right) \geq 0,$$

where $\frac{C^k}{n+1}$ is Wi-Fi throughput defined in (4). Then, the total utilities of G_n and S_n when served by LTE-W service are compared by:

$$T(A, n) + T(S', n) + \sum_{i=1}^{n-a} U(g_i')$$

$$\leq T(A, n) + T(S', n) + \sum_{i=1}^{n-a} U(g_i')$$

$$+ \sum_{i=1}^{n-a} \left[(U(g_i' + \frac{C^k}{n+1}) - U(g_i')) - (U(s_i' + \frac{C^k}{n+1}) - U(s_i')) \right]$$

$$= T(A, n) + T(G', n) + \sum_{i=1}^{n-a} U(s_i'),$$

where $T(S', n) = \sum_{i=1}^{n-a} U(s_i' + \frac{C^k}{n+1})$, and similarly $T(G', n)$. Therefore, the total utility of G_n is larger than that of S_n for a fixed number of queues $1 \leq n \leq |B_1|$. This completes the proof. □

D. Bearer Split-Scheduling

Key issues and challenges. We first discuss the key issues and challenges when we design a scheduling algorithm that splits inside the packets of a LTE-W bearer into LTE and Wi-Fi links. Note that there can exist multiple TCP flows inside one bearer.

○ *TCP throughput:* Suppose that a scheduler knows the per-bearer throughputs for LTE and Wi-Fi links a priori, denoted with symbols x_L and x_W, respectively. For a reception-rate oriented transport protocol such as UDP, it would be enough to split and schedule the packets in a bearer as a ratio of x_L and x_W. The challenge comes from the fact that TCP is a dominant transport mechanism and has very complex protocol behaviors such as Ack-clocked, loss-based Congestion Window (CWND) control, being highly sensitive to out-of-order packets. A vanilla approach to completely remove out-of-order packets is to adopt a flow-based split-scheduling [40], i.e., directing a TCP flow inside a bearer only to a single link, LTE or Wi-Fi. However, this flow-based scheduling is undesirable, due to its low link utilization as well as the cost of per-flow processing in eNodeB. Thus, it is necessary to devise a split-scheduling algorithm that (i) minimizes out-of-order packets as well as (ii) maximizes

---

Algorithm 2: BSplit

1:    **INPUT** For $i \in \{$L (LTE), W (Wi-Fi)$\}$,
      B_i : Per-bearer available bandwidth over link i,
      S_i : Total bytes of unAcked PDUs over link i,
      C_i : Processing and propagation delays of link i,

2:    **Split-scheduling algorithm:** For incoming PDCP PDU p,
3:       Compute D_i for each link i as follows:
         $D_i = (S_i + \text{size}(\text{p}))/B_i + C_i$.
4:       $i^\star = \text{argmin}_{i \in \{L, W\}} D_i$.
5:       $S_{i^\star} = S_{i^\star} + \text{size}(\text{p})$.
6:       Schedule p to link $i^\star$.
7:       Record following two p-specific values for link $i^\star$:
         $S_{i^*}(\text{p}) = S_{i^\star}$.
         $T(\text{p}) = \text{getCurrentTime}()$.

8:    **Bandwidth estimation:**
      For Ack of PDCP PDU p over link i^*,
9:       $S_{i^*} = S_{i^*} - \text{size}(\text{p})$.
10:      $RTT(\text{p}) = \text{getCurrentTime}() - T(\text{p})$.
11:      Estimate B_{i^*} in (6) based on $S_{i^*}(\text{p})$ and $RTT(\text{p})$.

---

link utilization, probably using various statistics from on-line measurement and off-line computations.

○ *Tracking network variations and modularization:* Another important issue is that network conditions are time-varying due to dynamic arrivals and departures of users/flows as well as mobility, which affect the results of on-line measurements, e.g., x_L and x_W values and thus the rule of splitting the per-bearer packets. Also, MAC-layer implementations of LTE are highly vendor-specific. For example, it is known that different LTE eNodeBs employ different user scheduling algorithms, e.g., proportional-fair scheduling [36] and MAX SNR scheduling [41], which often changes as MNO's policy. It is necessary to develop a split-scheduling mechanism that works without knowledge of such specific MAC-layer implementations.

Our design. To maximize link utilization and minimize the number of out-of-order TCP packets, we employ a scheduling algorithm, called **BSplit**, that splits the packets between LTE and Wi-Fi links, based on the measured *delay* and *bandwidth* of the target bearer over two access technologies (see Fig. 2). We also install a resequencing buffer at the UE PDCP layer to compensate for the possible imperfect handling with respect to out-of-order packets of our delay-based split-scheduling. In the following, we elaborate three key components of **BSplit**: (1) split scheduling algorithm, (2) bandwidth estimation, and (3) PDCP resequencing buffer.

(1) Split-scheduling algorithm: Delay-based. The basic idea of our delay-based split-scheduling is as follows: for an incoming PDCP PDU, we schedule it to be transmitted at the link having a smaller delay, where "delay" corresponds to the time from when the eNodeB transmits a packet at PDCP layer to when a UE receives it at the PDCP layer.

BSplit is described in **Algorithm 2**. In **BSplit**, for each incoming PDCP PDU p, we compute the value of D_i for

each link $i \in \{$L (LTE) or W (Wi-Fi)$\}$, which corresponds to the inferred delay of link i, computed by:

$$D_i = (S_i + \text{size(p)})/B_i + C_i, \tag{5}$$

where C_i is the sum of processing and propagation delays, and B_i is the estimated available bandwidth, i.e., the average bandwidth that is expected to be provided to the bearer including PDU p (we will discuss how to obtain the value of B_i shortly). Thus, S_i/B_i and $\text{size(p)}/B_i$ correspond to approximate queueing and transmission delays, where "approximate" is due to the fact that the B_i comes from the measurement (thus may not be perfect) and S_i also includes the bytes of PDUs on-the-fly, rather than counting the exact total bytes of PDUs in the eNodeB RLC queue or AP MAC queue. In particular, our method of simply counting the unAcked PDUs at the PDCP layer[4] is for implementation convenience by avoiding the information gathering from the underlying MAC layer. Then, the PDCP PDU p is finally scheduled at a link i^* having smaller delay. Note that when scheduling each incoming PDU p over the link, we record two p-specific values: $S_{i^*}(\text{p})$ and $T(\text{p})$ which correspond to the total bytes of unAcked PDUs and the scheduled time when p is scheduled (Line 7). For each Ack for PDU p sent over link i^*, we decrement S_{i^*} (Line 9), compute the (last hop link-level) round-trip-time $RTT(\text{p})$ (Line 10), and estimate the value of B_{i^*} in (6) (Line 11: we will present this shortly) using the earlier-recorded $S_{i^*}(\text{p})$ and $T(\text{p})$ when p was scheduled.

Rationale. We now present the rationale on why **BSplit** helps in achieving high efficiency in terms of both link utilization and packet ordering. First, in terms of packet ordering, it is clear that choosing the link with lower delay guarantees in-order packet delivery, as long as B_i and S_i are accurate, so as for the delay in (5) to be the exact delay that the target PDU p would experience. As mentioned earlier, we use a measurement-based scheme for B_i and an approximate value S_i of the actual bytes of queued packets, whose impact will be validated through extensive simulations. Second, in terms of link utilization, **BSplit** is effective as explained in what follows: Note that the reason why link utilization may be negatively affected is because our effort of sustaining the right ordering may render our split-scheduler non-work conserving, i.e., some link is kept idle to make the right sequence. To show that **BSplit** achieves high link utilizataion, it is important to check whether the ratio of using both links is the same as that of the offered available per-bearer bandwidths B_L and B_W, because in that case the scheduler becomes work-conserving. To this end, we first believe that the sum of propagation and processing delays C_i becomes negligible compared to D_i, because the last single hop Wi-Fi and LTE propagation delay is highly small and eNodeBs and Wi-Fi APs have enough computation powers to process the PDUs quickly. Under this condition, it is easy to see that from (5),

$$\triangle S_L : \triangle S_W = B_L : B_W,$$

where $\triangle S_i$ denotes the increment of pumped-in unAcked PDUs over link i over a fixed time interval.

[4]The LTE standard requires an Ack for each PDCP PDU.

The rationale in designing **BSplit** is demonstrated in a microbenchmark shown in Fig. 3, where the per-bearer available bandwidths for LTE and Wi-Fi links are set to be 7Mbps and 1.5Mbps, respectively. We consider three kinds of scheduling, **BSplit**, Round-robin, and Ratio-based in Fig. 3, where Ratio-based simply scheduled so that the transmission ratio is equal to the given available bandwidth ratio 7:1.5 by using e.g., weighted round-robin. **BSplit** achieves link utilization as almost 1, and sending rates for two links with the same ratio of available throughputs even over short time intervals. Ratio-based scheduling tracks the available bandwidths of two links, but we see that their instantaneous throughput is highly fluctuating. This is because relatively late PDUs from Wi-Fi incur a lot of out-of-ordering, so as to hurt TCP performance. It is interesting to observe that Ratio-based scheduling is even worse than the simple Round-robin scheduling, which, however, performs worse than **BSplit** because of the throughput over LTE is restricted by that over Wi-Fi.
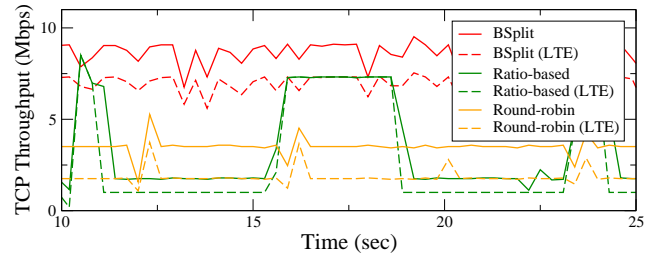

Fig. 3: Instantaneous TCP throughput of total LTE-W and LTE link with **BSplit** and Ratio-based, and Round-robin schedulings.

(2) Measurement-based per-bearer bandwidth estimation. Computing the value of B_i, which is crucial in **BSplit**, is non-trivial due to (i) sensitivity to network configuration changes and (ii) significant dependence to the underlying resource allocation in the MAC layer. In (i), by network configuration, we mainly mean the locations and the number of UEs in the network. In (ii), eNodeBs employ a vendor-specific user scheduling mechanism which in turn changes how much bandwidth is provided to a single bearer. Motivated by these challenges, we propose a measurement based B_i estimation mechanism which "conjectures" the value of B_i from the past PDU transmissions and their Ack reception status, as stated next.

It remains to explain which earlier PDU q is used to obtain the sample. LTE RLC provides concatenation and segmentation for the PDCP PDUs and RLC Acks can be generated in a cumulative manner for efficiency of LTE resource [35], similarly to the concept of delayed ack in TCP. However, Wi-Fi MAC Ack is immediately transmitted whenever there exists a successful data transmission as demonstrated in Fig. 4. Thus, in choosing q, in LTE, we use the last Ack of simultaneously received RLC Acks, including relatively short time in a queue, but in Wi-Fi, we simply use two MAC Acks for two consecutive packets p and q.

(3) Further tuning: PDCP resequencing buffer. Despite our design of **Bsplit** that aims at minimizing the number of out-of-order packets, random features of wireless channels, highly

Per-bearer bandwidth B_i estimation

S1. Whenever eNodeB PDCP layer receives Ack for PDCP PDU p sent over link i, using the earlier already-Acked q, compute the sampled available bandwidth $\tilde{B}_i$ by:

$$\tilde{B}_i = \frac{S_i(\mathrm{p}) - S_i(\mathrm{q})}{RTT(\mathrm{p}) - RTT(\mathrm{q})}. \tag{6}$$

S2. B_i is updated by its exponential moving average with the weighting constant α for the current sample $\tilde{B}_i$, where $\alpha = 1/512$.
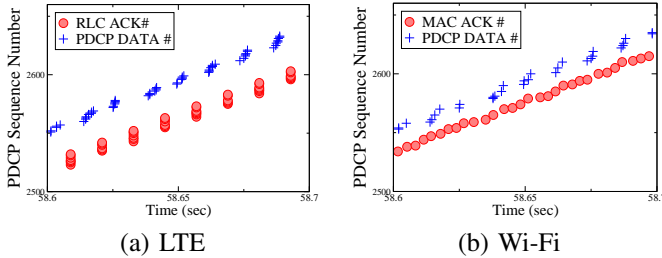

(a) LTE   (b) Wi-Fi

Fig. 4: Different Ack transmission patterns: LTE and Wi-Fi.

dynamic external flows in Wi-Fi, and possible imperfectness in estimating the available bandwidths B_L and B_W for two links, may not be perfect in fully obtaining the gains of aggregating two links. For this reason, we install additional safety device, which is a resequencing buffer at each UE PDCP layer to correct the order of the incoming PDCP PDUs. One of the crucial parameters in the design of the resequencing buffer is the time length T, which determines how long out-of-order packets stay in the resequencing buffer before they are pushed up to the upper layer. As T grows, we can increase the chance of correctly ordering out-of-order packets, which, however, also increases RTT, resulting in the decrease of TCP throughput. In our design, we choose $T = 100$ msec obtained from our various simulation experiments. Note that in Release 13, reordering functions for split bearers (e.g., LWA) are defined [28], [29].

IV. Performance Evaluation

A. Setup

Simulation environments. We show our simulation results to evaluate LTE-W with focus on the mode selection and bearer split-scheduling, where we extend NS-3 LENA [15], [16], an NS3 LTE and EPC implementation. Fig. 5 summarizes our simulation setup and scenarios, where we consider a heterogeneous network of one LTE and Wi-Fi cell that an eNodeB is connected to a MNO-operated AP by a high speed link, and ten UEs intend to use LTE-W, placed in two different geographical places, formed as two groups: `Group A` and `Group B`. In all scenarios, `Group A` is static, but `Group B` changes its location depending on the scenarios (which will be presented shortly). We describe other network operating parameters in Table II reflecting the state-of-the-art enviroment and specifications. In all of our simulation results, we have
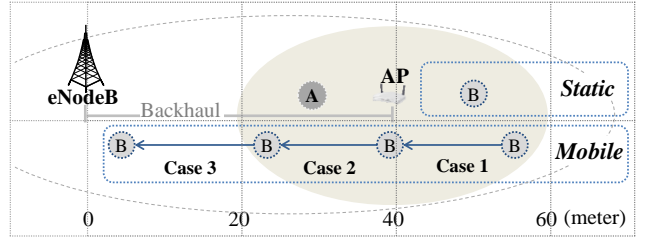

Fig. 5: Simulation node setup: 10 LTE-W UEs, `Group A` (5 UEs) and `Group B` (5 UEs).

used a collision model in Wi-Fi, which is a default setup in NS-3, meaning that whenever two transmissions are within their interference range, they can collide, which are in many cases avoided due to CSMA/CA. We let all the TCP Acks to be transmitted only through LTE link.

TABLE II: Simulation Parameters.

Parameter	Value
LTE configuation	20 MHz TDD
Wi-Fi PHY rate	48 Mbps (W/O frame aggregation)
Mode of RLC layer	Acknowledge Mode
LTE user scheduler	Proportional fair
Wi-Fi user scheduler	Round-robin
LTE pathloss model	Okumura-Hata
Wi-Fi pathloss model	Log-distance
PDCP resequencing buffer timeout	100 msec
Resequencing TCP version	New Reno

Scenarios. We considered two scenarios: *Static* and *Mobile*.

○ *Static:* Each UE in two groups is statically located where UEs in `Group A` are at 30 meters and UEs in `Group B` are at 50 meters from the eNodeB, and all of them have the same distance from the AP, 10 meters.

○ *Mobile:* UEs in `Group B` walk on the way to eNodeB with consistent walking speed 3.1 km/h for 60 seconds, from the end of the Wi-Fi coverage as shown in Fig. 5, whereas UEs in `Group A` is statically located at 30 meters from the eNodeB. After 40 seconds, UEs in `Group B` turn out to leave from the Wi-Fi coverage, thus mode of each UE in `Group B` is changed to LTE-only at that time. We observe the performance of UEs in `Group A` and `Group B` at 20, 40, and 60 seconds (called Cases 1, 2, and 3).

In all simulations, we consider the case when each UE of `Group A` and `Group B` has one dedicated bearer consisting of one flow[5], unless explicitly mentioned, and tries to download a content with an infinite size from a remote server outside of a LTE network. We also test LTE-W for HTTP/1.1 traffic and a bearer consisting of multiple TCP flows in some simulations to investigate more practical impact of LTE-W.

Comparison. We compare LTE-W with MPTCP [9]–[11] which is a well-known transport-level bandwidth aggregation.

---

[5]In [39], it is reported that there exists only one flow in more than 70% bearers.

We are interested in both total throughput and fairness, where total throughput is measured simply by the aggregate throughput of all UEs. Among the several MPTCP versions, we use the Linked Increases Algorithm (LIA) [5], available as a public open-source code [42]. We modify this code so as to be compatible to the LTE-W system.

B. Results: Static

MPTCP and LTE-W. Table III shows the average per-group and total throughputs of MPTCP and LTE-W UEs for 100 seconds. We observe that LTE-W exceeds MPTCP by 20%, where LTE-W achieves the total throughput 42.5 Mbps, whereas 35.3 Mbps in MPTCP. We see that LTE-W offers more bandwidths to the UEs of `Group B` (located 50 meters from the eNodeB), whereas MPTCP allocates too much bandwidth to `Group A` nearer to the eNodeB. This unfairness of MPTCP comes from the fact that LIA avoids aggressive behavior in order not to harm others, and thus carefully increases the congestion window size considering two links' environments, and decreases window size as regular TCP does [5].

TABLE III: Average throughput/UE in LTE-W and MPTCP: *Static*

(Mbps)		Group A	Group B	All
LTE-W	LTE	3.8522	1.7777	2.8150
($\alpha = 1$)	Wi-Fi	0.0	2.8590	1.4295
	Total	3.8522	4.6367	**4.2445**
MPTCP	LTE	3.0724	0.3730	1.7227
	Wi-Fi	2.8592	0.7643	1.8117
	Total	5.9316	1.1373	**3.5345**

Fairness of MPTCP and LTE-W. We first investigate the per-bearer fairness of both protocols, MPTCP and LTE-W, whose results are summarized in Table IV. We vary the fairness parameter α in the α-fair utility functions, where we observe that the utilities for all tested α of LTE-W outperform those of MPTCP. When we see the fourth column who gets LTE-W?, in our LTE-W, only a part of the entire UEs are decided to get both LTE and Wi-Fi, because our mode selection leads to assigning priority to UEs having smaller LTE throughput in order to maximize the increment of the network utility (e.g., UEs located at the cell edge).

TABLE IV: Utilities of MPTCP and LTE-W: *Static*

α	MPTCP	LTE-W	Who gets both LTE and Wi-Fi?
0.01	35.0981	42.2563	Group B (5)
0.1	33.1660	40.7972	Group B (5)
1	0.9457	14.4131	Group B (5)
10	-4.66E+12	-4.20E-6	Groups A (1), B (5)
100	-3.60E+147	-1.16E-55	Groups A (3), B (5)

This trend is slightly different across different values of α in that more users in `Group A` are selected for the LTE-W mode with larger α, since as α increases, the utility gain and the slope of utility function decrease, so that the mode selection algorithm chooses a larger number of UEs for both LTE and Wi-Fi. Figs. 6(a) and 6(b) show the instantaneous
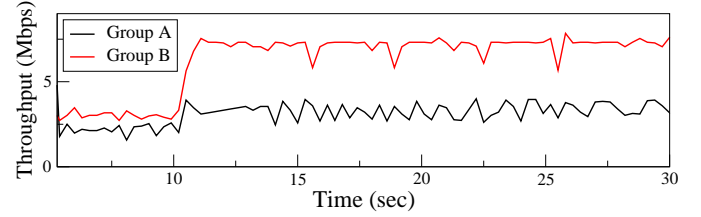
(a) Instantaneous TCP throughput of each LTE-only UE `Group A` and LTE-W UE `Group B` ($\alpha = 1$)
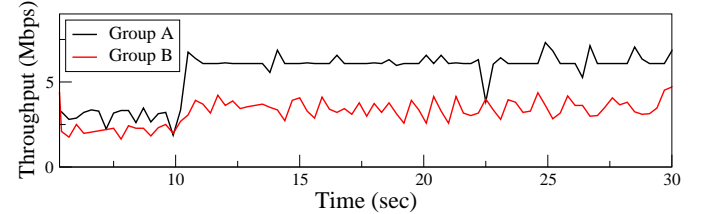
(b) Instantaneous TCP throughput of each LTE-W UE in `Group A` and `Group B` ($\alpha = 10$)

Fig. 6: Instantaneous TCP throughputs of UEs: *Static*

TCP throughputs for two cases, $\alpha = 1$ and $\alpha = 10$. The instantaneous TCP throughputs of LTE-W UEs in `Group B` increase due to the addition of Wi-Fi, while the throughputs of UEs in `Group A` do not change. However, for $\alpha = 10$, instantaneous TCP throughputs of LTE-W UEs in both groups increase.

○ *Jain's fairness index:* Fig. 7(a) shows the Jain's fairness index of LTE-only, MPTCP and LTE-W UEs, where by definition the index ranges over the interval [0,1]. LTE-W outperforms LTE-only by 15%, and MPTCP by 75%. This implies that unconditional providing of the aggregation service to all UEs, as in MPTCP, causes serious unfairness.

○ *GAT of* `Group A` *and* `Group B`*:* We confirm this by investigating what happens in each group, where we plot the geometric average of UE throughputs (GAT[6]) for $\alpha = 1$, shown in Fig. 7(b). Due to our distance setting from the eNodeB, the GAT of LTE-only UEs in `Group A` are two times larger than that of `Group B`, and LTE-W UEs achieves higher GAT than LTE-only UEs for all cases. However, GAT of MPTCP in `Group B` is even lower than that of LTE-only UE, as also seen Fig. 7(c), i.e., bandwidth concentration on `Group A`. However, LTE-W UEs in `Group B` receive higher GAT than those in `Group A`, thanks to our mode selection. Moreover, LTE-W UEs in `Group B` achieve high utilization of LTE/Wi-Fi links, since GAT from LTE link is almost the same as the GAT of LTE-only UEs in `Group B`.

C. Results: Mobile

To observe the dynamics in *Mobile*, we consider that **ModeSel** is periodically performed for each 10 seconds with $\alpha = 1$, where the period is determined based on the mobility pattern of users in a cell. Therefore, mode of each UE in `Group A` is changed to LTE-W at 20 seconds, when UEs in `Group B` locate nearby `Group A`. And mode of each UE in

---

[6]For a set of numbers $\{x_i\}_{i=1}^N$, GAT = $(\prod_{i=1}^N x_i)^{1/N}$

(a) Jain's fairness index of LTE-only, MPTCP and LTE-W UEs

(b) GAT of LTE-only, MPTCP and LTE-W for each `Group A` and `Group B`
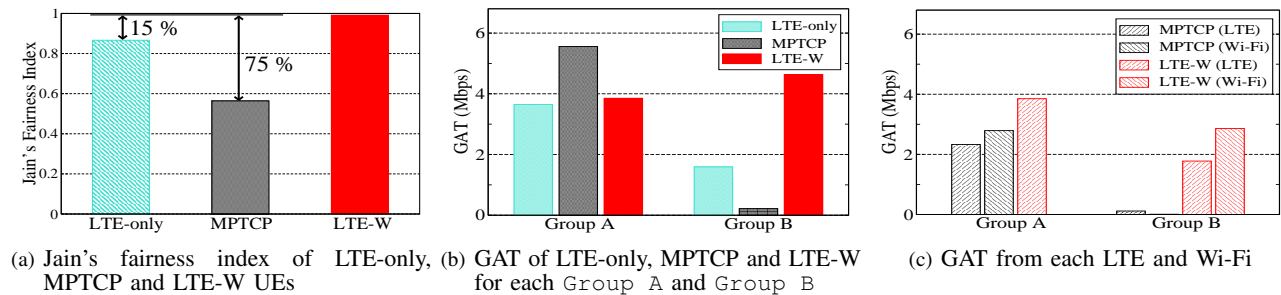
(c) GAT from each LTE and Wi-Fi

Fig. 7: Fairness and GAT comparisons: LTE-W and MPTCP: *Static*

`Group B` is changed to LTE-only near 40 seconds, after UEs in `Group B` leave the Wi-Fi coverage. Thus, we separate 60 seconds by 20 seconds, and evaluate the average performance of MPTCP and LTE-W UEs for each case, 20 second interval, as shown in Table V.

Fairness and GAT of MPTCP and LTE-W. For three cases, utility, Jain's fairness index and GAT of MPTCP are similar to or less than those of LTE-W, as shown in Table V, Figs 8(a) and 8(b). Since we consider MPTCP in current system utilizing a FIFO queue, the Wi-Fi resource is concentrated to the UEs having higher LTE throughput, thus MPTCP has low Jain's fairness index due to the unbalanced resource utilization between `Group A` and `Group B`, whereas LTE-W having per-bearer queueing maintains Jain's fairness Index as 1.

TABLE V: Utility comparison: MPTCP and LTE-W: *Mobile*

|	MPTCP	LTE-W	Who gets both LTE and Wi-Fi?
Case 1	5.7158	14.0136	`Group B`
Case 2	11.7589	17.8250	`Group A`, `Group B`
Case 3	16.2021	16.8360	`Group A`
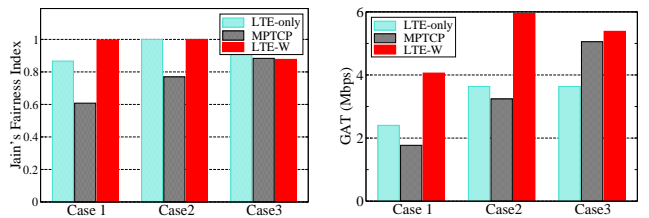
(a) Jain's fairness Index of LTE-only, MPTCP and LTE-W UEs

(b) GAT of LTE-only, MPTCP and LTE-W UEs

Fig. 8: Jain's fairness index and GAT of LTE-only, MPTCP and LTE-W UEs for Cases 1, 2, and 3: *Mobile*

Dynamics of MPTCP and LTE-W. According to the results of **ModeSel** in Table V, instantaneous TCP throughputs of UEs in `Group A` increase as depicted in Fig. 9(a), and the LTE throughputs slightly decrease in the Case 3, because UEs in `Group B` achieve a higher channel quality on the way to eNodeB. On the other hand, as UEs in `Group B` move closer to the eNodeB, instantaneous LTE throughputs of UEs increase in the Case 1, and they consistently utilize a constant amount of Wi-Fi resources guaranteed by per-bearer queueing in AP. After 40 seconds, UEs in `Group B` move out of the Wi-Fi coverage, but they could seamlessly receive remaining data

through the LTE link as shown in Fig. 9(b) without degrading QoE (Quality-of-Experience). Because split-scheduling is performed on the eNodeB, UE status report is completed in a short time, and eNodeB adaptively changes the mode of each UE in `Group B` from LTE-W to LTE-only to minimize the stacked packets in Wi-Fi queue causing TCP retransmission or timeout events.
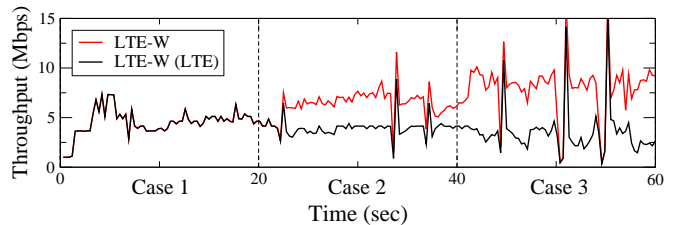
(a) Instantaneous TCP throughput of each LTE-W UE in `Group A`
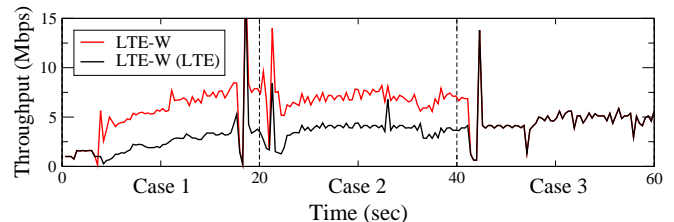
(b) Instantaneous TCP throughput of each LTE-W UE in `Group B`

Fig. 9: Instantaneous TCP throughputs: *Mobile*

### D. Results: HTTP and Multi-flow Traffic Bearer

In this subsection, we demonstrate the performance of LTE-W, when UEs in `Group B` are located nearby `Group A`, thus ten UEs in both groups can utilize LTE-W, as shown in Case 2 of scenario *Mobile*.

LTE-W performance in HTTP traffic. Most LTE traffic is HTTP traffic and 90% of flows carry less than 36 KB downlink payload [39], and large size video streaming traffic is transmitted as several blocks having size less than few MBytes [43], [44]. Thus, we test the performance of LTE-W for HTTP/1.1 traffic by modifying open-source code [45], [46]. Each UE in `Group A` and `Group B` downloads HTTP traffic for 100 seconds, where HTTP traffic is generated based on the statistical model of main object size (mean: 31,561 Byte), inline object size (mean: 23,915 Byte), number of inline objects (mean: 31.39), and user reading time (mean: 39.7s) [47], [48] obtained by the parameters from the real
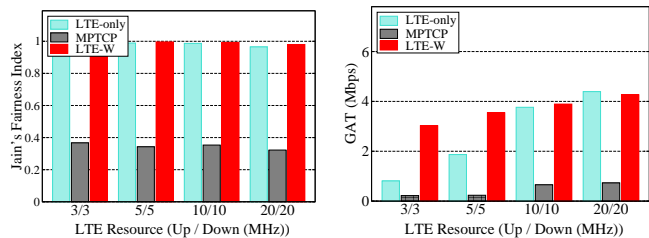
(a) Jain's fairness Index of LTE-only, MPTCP and LTE-W UEs

(b) GAT of LTE-only, MPTCP and LTE-W UEs

Fig. 10: Jain's fairness Index and GAT of LTE-only, MPTCP, LTE-W UEs for HTTP traffics.
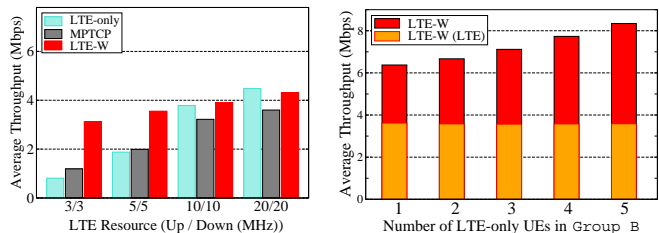
(a) Average throughput of LTE-W and LTE-only for HTTP traffic

(b) Average throughput of LTE-W and LTE link of UEs in Group A

Fig. 11: Average throughput comparison.

web page transactions. To verify the performance for diverse environments, we perform the additional simulations for constant size of object (50 KByte and 200 KByte) with the same parameter setting for number of inline objects and reading time. We consider four kinds of LTE bandwidth 3, 5, 10 and 20 MHz [27] with the fixed Wi-Fi PHY rate 48 Mbps which could indirectly show various network environments of users, and compare LTE-W to LTE-only and MPTCP in the same situation.

As shown in Figs. 10(a), 10(b), and 11(a) the average throughput, GAT and Jain's fairness Index of LTE-W UEs are similar to or bigger than those of LTE-only UE except the case of 20 MHz. Since the total amount of transmitted data from eNodeB is smaller than LTE resource capacity of 20 MHz, average throughput of LTE-only is varied depended on amount of data and number of scheduled users in a moment, thus LTE-only has higher average throughput rather than MPTCP and LTE-W as depicted in Figs. 11(a) and 12(a), different from the result in Fig. 13(a). These trends are related to the HTTP's feature and MPTCP LIA's behavior carefully increasing congestion window size [5] where several evaluation results of MPTCP have already reported that we could not achieve significant gain from MPTCP for smaller file sizes such as less than 50 KByte [49], [50], since each subflow has not enough time to probe path's capacity for efficient utilizing. Since the size of each main object or inline object is less than 30 KB on average, and reading time (mean 39.7 seconds) is comparably longer than object download time (2.067 seconds, downloading 775 KB through 3 Mbps (775 KB is total size of one 32 KB main object and thirty-one 24 KB inline objects)), the available bandwidths of two links are dynamically changed due to asynchronous download pattern of UEs. Thus, the estimated bandwidths of LTE-W in

the past could be totally different from the current available bandwidths, and this phenomenon might be amplified for the case of having large bandwidth (quick download time).

Therefore, we can conclude that LTE-W provides similar or better performance than LTE-only for HTTP traffic, and especially LTE-W UEs in scarcity of LTE resource (e.g., 3 MHz, located cell edge) could obtain significant improvement. Although the average throughput of LTE-only UE is slightly higher than that of LTE-W UE, the gap of average total delays is negligible (e.g., 0.033 seconds to download 775 KB). Although MPTCP achieves similar average throughput as LTE-W when the object size is 200 KByte, MPTCP has lower GAT (Figs. 10(b), 12(b) and 13(b)) and Jain's fairness index (Figs. 10(a), 12(c) and 13(c)) rather than LTE-W, since short-term bursty HTTP traffics are transmitted through Wi-Fi FIFO queue with concentrated scheduling to a UE, so that some of the MPTCP UEs achieve high throughput, but the other UEs receive insufficient allocated resource.

High link utilization and TCP performance. As illustrated in Fig. 11(b), LTE-W UEs achieve the high link utilization when UEs in Group B utilize LTE-only instead of LTE-W. As the number of LTE-only UEs in Group B increases from one to five, the throughput of LTE link is consistently almost 3.9 Mbps, and that from Wi-Fi link increases, where the total sum of the throughput from Wi-Fi link of all LTE-W UEs is maintained as a constant. This implies that our **BSplit** works well for TCP. Note that bearer split-scheduling operates on a bearer basis that contains a multiple of TCP flows. Thus, one may wonder about the individual TCP performance, and the fairness among them. Figs. 14(a) and 14(b) shows the traces of CWNDs for the cases when a bearer contains 2 and 4 flows, respectively (e.g. simultaneously transmitting two files or web browsing while downloading a file.). We observe that they stably follow the standard behavior of TCP's CWND, and there is no starvation for a specific flow.

V. CONCLUSION

Adaptive bandwidth aggregation is a promising solution to cope with scarcity of mobile network capacity and high bandwidth-hungry applications. In this paper, we proposed a link-level LTE/Wi-Fi bandwidth aggregation, called LTE-W. As two key modules, we proposed mode selection and bearer split-scheduling that smartly consider per-bearer fairness and efficiently merge LTE and Wi-Fi links to achieve high TCP performance and link utilization. LTE-W was implemented at the NS-3 LENA platform, and evaluated in terms of fairness and TCP performance with comparison to a transport-level bandwidth aggregation, MPTCP, under various scenarios including mobile users and HTTP traffic. We demonstrated that LTE-W outperforms over MPTCP by up to 75% with regards to Jain's fairness index, and our proposed bearer split mechanism achieves high link utilization.

REFERENCES

[1] Cisco, "Cisco visual networking index: global mobile data traffic forecast update, 2015-2020," Cisco Systems Inc., Tech. Rep., 2016.

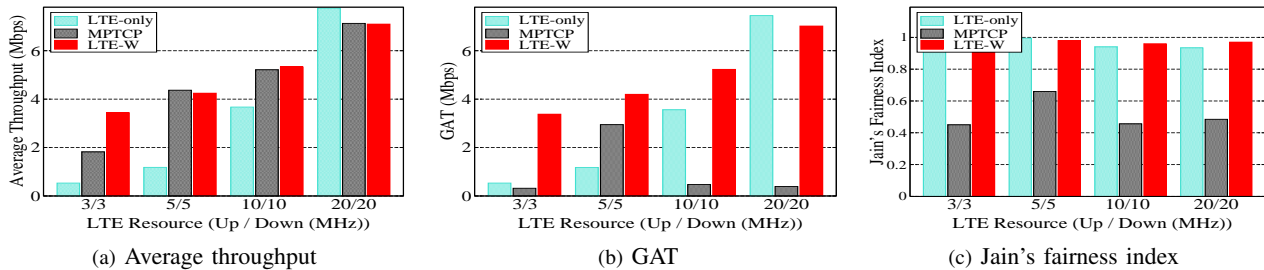[2] ——, "Cisco visual networking index: forecast and methodology, 2015-2020," Cisco Systems Inc., Tech. Rep., 2016.

(a) Average throughput

(b) GAT

(c) Jain's fairness index

Fig. 12: Average throughput, GAT and Jain's fairness index of LTE-only, MPTCP, LTE-W UEs for HTTP traffics having object size 50 KByte.

(a) Average throughput

(b) GAT

(c) Jain's fairness index
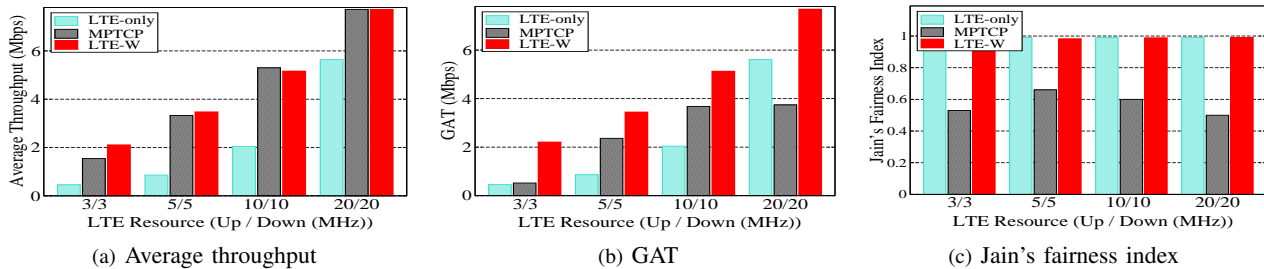
Fig. 13: Average throughput, GAT and Jain's fairness index of LTE-only, MPTCP, LTE-W UEs for HTTP traffics having object size 200 KByte.

(a) CWND of LTE-W for a two-flow bearer

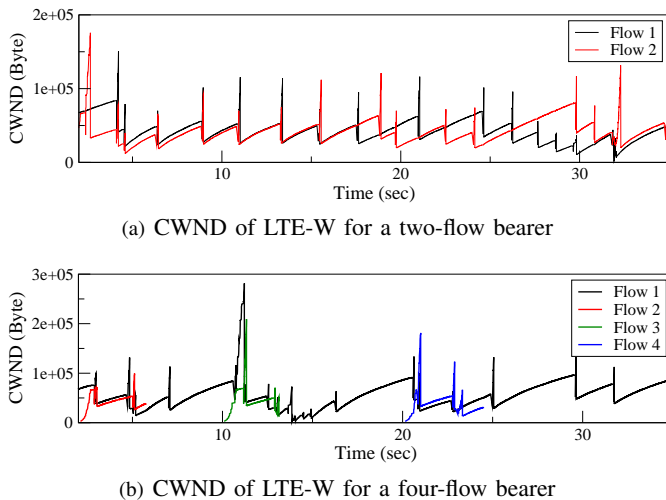
(b) CWND of LTE-W for a four-flow bearer

Fig. 14: CWND traces for a bearer containing multiple TCP flows.

[3] K. Zhu, D. Niyato, and P. Wang, "Network selection in heterogeneous wireless networks: evolution with incomplete information," in *Proceedings of IEEE WCNC*, 2010.

[4] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "Rat selection games in hetnets," in *Proceedings of IEEE INFOCOM*, 2013.

[5] C. Raiciu, M. Handley, and D. Wischik, "Coupled congestion control for multipath transport protocols," IETF, Informational RFC 6365, 2011.

[6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[7] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 580–603, 2014.

[8] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.

[9] D. Wischik, C. Raiciu, A. Greenhalgh, and M. Handley, "Design, implementation and evaluation of congestion control for Multipath TCP," in *Proceedings of NSDI*, 2011.

[10] C. Raiciu, C. Paasch, S. Barre, A. Ford, M. Honda, F. Duchene,

O. Bonaventure, and M. Handley, "How hard can it be? designing and implementing a deployable Multipath TCP," in *Proceedings of NSDI*, 2012.

[11] A. Ford, C. Raiciu, M. Handley, S. Barr, and J. R. Iyengar, "Architectural Guidelines for Multipath TCP Development," IETF, Informational RFC 6182, Mar. 2011.

[12] "Kt wi-fi service." [Online]. Available: http://zone.wifi.olleh.com/

[13] 3GPP, "Policy and charging control architecture," TS 23.203 V14.2.0, www.3gpp.org.

[14] ——, "Study on small cell enhancements for e-utra and e-utran; higher layer aspects," TR 36.842 V12.0.0, www.3gpp.org.

[15] "NS-3 LTE-EPC Network simulAtor (LENA)." [Online]. Available: http://networks.cttc.es/mobile-networks/software-tools/lena/

[16] "Network Simulator-3." [Online]. Available: http://www.nsnam.org

[17] W. Shen and Q.-A. Zeng, "Cost-function-based network selection strategy in integrated wireless and mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3778–3788, 2008.

[18] L. Wang and G.-S. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networksa tutorial," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 271–292, 2013.

[19] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated lte-wifi networks," in *Proceedings of ACM MobiCom*, 2014.

[20] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[21] "KT's GiGA LTE." [Online]. Available: https://www.ietf.org/proceedings/93/slides/slides-93-mptcp-3.pdf

[22] "SKT develops GB wireless Internet technology." [Online]. Available: http://www.businesskorea.co.kr/english/news/ict/10492-85-seconds-skt-develops-gb-wireless-internet-technology

[23] J. Wu, B. Cheng, M. Wang, and J. Chen, "Delivering high-frame-rate video to mobile devices in heterogeneous wireless networks," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4800–4816, 2016.

[24] J. Wu, C. Yuen, B. Cheng, M. Wang, and J. Chen, "Streaming high-quality mobile video with multipath tcp in heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2345–2361, 2016.

[25] ——, "Energy-minimized multipath video transport to mobile devices in heterogeneous wireless networks," *IEEE JSAC*, vol. 34, no. 5, pp. 1160–1178, 2016.

[26] J. Wu, C. Yuen, B. Cheng, Y. Yang, M. Wang, and J. Chen, "Bandwidth-efficient multipath transport protocol for quality-guaranteed real-time video over heterogeneous wireless networks," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2477–2493, 2016.

[27] 3GPP, "Evolved universal terrestrial radio access (e-utra) and evolved universal terrestrial radio access network (e-utran); overall description," TS 36.300 V14.1.0, www.3gpp.org.

[28] ——, "Evolved universal terrestrial radio access (e-utra); Radio Resource Control (RRC); protocol specification," TS 36.331 V14.1.0, www.3gpp.org.

[29] ——, "Evolved universal terrestrial radio access (e-utra); Packet Data Convergence Protocol (PDCP) specification," TS 36.323 V14.1.0, www.3gpp.org.

[30] Qualcomm, "Qualcomm research lte in unlicensed spectrum: Harmonious coexistence with wi-fi," 2014, www.qualcomm.com.

[31] 3GPP, "Feasibility study on licensed-assisted access to unlicensed spectrum," TR 36.889 V13.0.0, www.3gpp.org.

[32] Qualcomm, "Progress on LAA and its relationship to LTE-U and multefire," 2016.

[33] ——, "Best use of unlicensed spectrum," 2016.

[34] Ericsson, "LTE license assisted access," 2015.

[35] 3GPP, "Evolved universal terrestrial radio access (e-utra); Radio Link Control (RLC) protocol specification," TS 36.322 V13.2.0, www.3gpp.org.

[36] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in lte," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 461–464, 2009.

[37] Huawei, "eLTE broadband access solution QoS technical guide," 2014. [Online]. Available: http://www.huawei.com/ilink/enenterprise/download/HW_372670

[38] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 200.

[39] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of LTE: effect of network protocol and application behavior on performance," in *Proceedings of ACM SIGCOMM*, 2013.

[40] S. Kandula, D. Katabi, S. Sinha, and A. Berger, "Dynamic load balancing without packet reordering," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 2, pp. 51–62, 2007.

[41] S. Schwarz, C. Mehlführer, and M. Rupp, "Low complexity approximate maximum throughput scheduling for LTE," in *Proceedings of IEEE ASILOMAR*, 2010.

[42] "Multi-Path TCP (MPTCP)." [Online]. Available: http://code.google.com/p/multipath-tcp/

[43] P. Ameigeiras, J. J. Ramos-Munoz, J. Navarro-Ortiz, and J. M. Lopez-Soler, "Analysis and modelling of youtube traffic," *Transactions on Emerging Telecommunications Technologies*, vol. 23, no. 4, pp. 360–377, 2012.

[44] A. Rao, A. Legout, Y.-s. Lim, D. Towsley, C. Barakat, and W. Dabbous, "Network characteristics of video streaming traffic," in *Proceedings of ACM CoNEXT*, 2011.

[45] "HTTP traffic generator for NS-3." [Online]. Available: https://codereview.appspot.com/4940041

[46] "HTTP traffic generator for NS-3." [Online]. Available: http://www.saulodamata.com/codes/85-http-traffic-generator-for-ns-3.html

[47] R. Pries, Z. Magyari, and P. Tran-Gia, "An HTTP web traffic model based on the top one million visited web pages," in *Proceedings of IEEE NGI*, 2012.

[48] J. J. Lee and M. Gupta, "A new traffic model for current user web browsing behavior," Intel Coperation, Tech. Rep., 2007.

[49] O. Bonaventure, C. Paasch, and G. Detal, "Use cases and operational experience with multipath TCP," IETF, Tech. Rep., Oct. 2016.

[50] Y.-C. Chen, Y.-s. Lim, R. J. Gibbens, E. M. Nahum, R. Khalili, and D. Towsley, "A measurement-based study of multipath tcp performance over wireless networks," in *Proceedings of ACM IMC*, 2013.

Segi Kim received his B.S degree in the School of Electrical and Computer Engineering from the University of Seoul in 2014, and his M.S degree in the Department of Electrical Engineering from KAIST, South Korea, in 2016. He is currently a research engineer in Korea Telecom. His research interests include software defined networking, network function virtualization, and wireless networks.

Donggyu Yun received his B.S. and M.S. and Ph.D degree in the electrical engineering from the KAIST, South Korea, in 2009 and 2011 and 2016. He is currently with Naver Search Division, Naver, South Korea. His research interests include machine learning, data analysis, and wireless networks.

Hojin Lee is a senior software engineer in Cloud Platform Group, Samsung Electronics, where he is responsible to develop software for firmware update over the air. Before joining Samsung Electronics, he was with KAIST, South Korea as a postdoctoral researcher from 2012 to 2014. He was a visiting researcher at University of Minnesota in 2009 and at Simon Fraser University in 2005. He received his B.S. and Ph.D. degrees in computer science and engineering from Seoul National University in 2003 and 2012, respectively. His recent research interest includes neighbor discovery in Internet-of-Things (IoT) and realistic interference generation in wireless networks.

Wooseong Kim is currently an assistant professor of computer engineering department, Gachon university, South Korea. He received his Ph.D. degree from computer science, UCLA. He used to work as a researcher of Samsung electronics, Hyundai motor, LG electronics, SK Hynix semiconductor, etc. He had standardization activity in several SDOs like 3GPP, ITU, TTA and ETSI. He is interested in multi-hop adhoc networks, LTE and 5G wireless telecommunication systems, wireless LAN,SDN/NFV, IoT protocols, etc.

Boram Jin received her B.S. and M.S. degrees in the College of Science from Kyunghee University, South Korea, in 2009 and 2011, respectively. From 2010 to 2012, she was a researcher in the National Institute for Mathematical Sciences, and received Ph.D. degree in the Department of Electrical Engineering at KAIST in 2017. Her current research interests include mathematical modeling and analysis for wireless systems such as LTE and Wi-Fi in 5G and learning approaches in large scale analysis in cache networks (e.g. Information centric network).

Yung Yi received his B.S. and the M.S. in the School of Computer Science and Engineering from Seoul National University, South Korea in 1997 and 1999, respectively, and his Ph.D. in the Department of Electrical and Computer Engineering at the University of Texas at Austin in 2006. From 2006 to 2008, he was a post-doctoral research associate in the Department of Electrical Engineering at Princeton University. Now, he is an associate professor at the Department of Electrical Engineering at KAIST, South Korea. His current research interests include the design and analysis of computer networking and wireless communication systems, especially congestion control, scheduling, and interference management, with applications in wireless ad hoc networks, broadband access networks, economic aspects of communication networks, and green networking systems. He received the best paper awards at IEEE SECON 2013, ACM MOBIHOC 2013, and IEEE William R. Bennett Award 2016.