Chapter 5

Radio Access Network Backhauling Using Power Line Communications

Francesco Marcuzzi and Andrea M. Tonello

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.73390

Abstract

Nowadays, radio access networks (RANs) are moving toward a small cell-based paradigm, where the macro-cell antennas are aided by additional ones with lower coverage capabilities. This paradigm shift brings a new problem into the equation: a backhaul link is needed to carry the traffic from the small cell base stations to the network gateway. Currently, both wired and wireless solutions exist, but none is universally considered optimal. Power line communications (PLC) can be considered as a broadband access solution for this backhaul branch. Recent developments helped to push PLC performances to a point where state-of-the-art solutions can achieve very high-speed data transfers. Aided by traffic generation-based simulations, we will show how the PLC technology can be assessed for the described application. The reader will be guided through the process by discussing small cell networks, the power line infrastructure and basics of traffic generation modeling. The chapter will then discuss a quality-of-service (QoS)-driven analysis and use numerical results to show how requirements can be defined for the backhauling technology. Overall, this chapter will address how PLC and small cell network technologies can be brought together in a unified model to foster future small cell technology.

Keywords: power line communications, small cell networks, backhauling, optimization, broadband access, random channel modeling, stochastic geometry, transmission line theory, traffic generation

1. Introduction

Nowadays, Internet connectivity is a consumer good, and its usage is growing faster and faster. A recent report forecasts that the worldwide mobile data traffic will clock in at 71

exabytes per month in 2022, while in 2016 this figure equalled about 9 EB/month [1]. It also predicts that by 2022 five out of nine billion of mobile subscriptions will be LTE (in 2016 this value was 1.5 out of 7.5), and that video services will generate 75% of the amount of mobile data traffic (in 2016 it was 50%). This growth in the amount of traffic is causing a shift in the way we design networks and the technologies used for their deployment. In fact, other sources predict that new-generation network technologies will see a manifold increase in the strictness of the requirements on capacity, latency, energy consumption, cost and mobility [2, 3].

In the world of mobile communications, connectivity is mainly deployed through what are technically called radio access networks (RANs). A RAN is a part of a system of telecommunications: it is the conceptual intermediary between the mobile end-user, which is the person requesting service through a user device, and the core network, which enables the exchange of data from all the endpoints of the network. A RAN is often implemented through a system of high-powered antennas which partition the geographical territory in areas of a few kilometers of diameter each. These are referenced to as macro-cells, because of their capacity of coverage. RANs are widely spread to enable provision of communication standards as GSM, EDGE, UMTS and LTE. On the other hand, there are also other solutions to provide mobile connectivity.

Small cells are low-power, high-capacity radio access nodes that improve directly the spatial fragmentation and frequency reuse of the network and indirectly local capacity and global coverage. They also bring the network closer to the user, thus helping to save energy and stretching end-device battery life. Small cells can be characterized based on their capacity of coverage, both in terms of distance and number of simultaneous users, into the following broad families: microcells, picocells and femtocells. The latter, which are the smallest and less powerful of the bunch, are used mostly for in-home, private purposes, because of their low cost and high capacity [4], and are the kind of cells that will be considered throughout this work. Small cells are strategically deployed in environments where the network coverage generated by the macro-cell is weak, in order to bring the said network closer to the final user.

This technology introduces the necessity for a backhaul segment to bring data from the small cells to the RAN access points (network gateways) and back further to the core network. Solutions for this link can be wired or not; popular choices are optic fibers, ADSL on twisted pair or radio connections. Furthermore, this additional element in the network's hierarchy needs to be designed properly, taking into account the end-user requirements, and granting easy deployment and no use of end-user spectrum are all necessary action points in network planning [5].

Power line communications (PLC) is a technology that uses the same medium both for power supply and for data connectivity. An immediate advantage of this technology is the fact that there is no need for a new infrastructure, leading to savings in deployment time and costs. Despite the power cables not being designed for data transmission, they are currently employed both for low data rate (e.g., meter reading, alarm signals, emergency communications) and for data-hungry (e.g. in-home HD video streaming) applications [6, 7]. New advancements in the PLC research boosted the communication capacity, even beyond gigabit per second in the in-home environment. While PLC promises easy deployment, its performance for backhauling applications still needs to be thoroughly assessed.

To the best of our knowledge, never-before PLC was regarded as a broadband solution for RAN backhauling. Considering how femtocells are nowadays used inside buildings and how power is fed to this premises through the power line infrastructure, this could be an opportunity to better integrate PLC and provide general connectivity in the small cell environment. To analyze the problem, the contribution in this chapter is organized as follows.

In Section 2 the rationale of the wireless small cell networks, the power line infrastructure paradigm and the traffic generation methods used for the first part of the assessment will be explained.

In Section 3 the overall network structure will be described. A bottom-up transmission line model will be used to simulate the physical PLC channel between network nodes.

Section 4 will explain how, on top of the previous results, an optimization problem can be formulated to perform resource allocation. This also enables smart routing for improved connectivity of far cells.

Section 5 will be dedicated to showing and commenting results from the simulations carried out for several system setups.

Eventually, conclusions will be drawn in Section 6, along with comments on possible future applications and research paths.

2. Setting up the network: the physical layer

The network to investigate comprises small cells and power lines in a joint infrastructure. This section specifically aims at finding out the requirements that a PLC network needs to satisfy to be able to support a small network of private small cells. To do so, the following main aspects have to be addressed: (a) the small cell network deployment on the geographical territory, (b) the power line infrastructure and how it topologically connects the radio base stations (c) the traffic generated by the cells that needs to be backhauled.

Most of the material presented in this section is a synthesis of [8, 9].

2.1. Small cells

Small cells are low-power, high-capacity radio nodes, controlled either by an operator or a private customer. As mentioned before, small cells can be categorized into the following classes:

- Microcells: their coverage can reach up to a few kilometers and is used in large premises such airports or transportation hubs. They enable macro-cell base station controllers to manage the power in the network while optimizing spectrum usage.
- Picocells: these can cover up to a few hundred meters. While being employed for medium-sized venues (malls, train stations and the like), they also enable smoother handoffs of end-users between different macro-cells.

- Femtocells: lowest coverage of all the classes, it clocks in at a few tens of meters. They provide advantages both to operators and end-users, by bringing the network closer to the devices and improving battery life and quality-of-service (QoS). Femtocells can usually support 4, 8 or 16 simultaneous users.

While the first two categories are generally deployed and managed by operators, femtocells are directly controlled by the private user. It is legit to assume that, in a residential neighborhood, connection to the core network is provided in each building through one of these femtocells, meaning that there is a one-to-one relationship between cells and buildings in our model.

The next generation of cellular technology envisions a very different paradigm for networks, where small cells will play a fundamental role in assisting the macro-cells in providing the service in blind spots or where the QoS needs to be improved; in fact, a great increase in the density of small cell is forecasted for the coming years. Our model inquires the possibility of providing the connection to the small cells through simulations where the density of such elements is adjustable, thus allowing the model to relax or tighten the capacity requirement on the PLC technology.

In order to simulate the envisioned network, the first thing needed is a deployment of small cells. To do this, we employ the stochastic geometry tool [10], as to create a random dispersion of small cells, based on their individual coverage and global density, which is indirectly controlled through the portion of geographically territory to be actually covered. Stochastic geometry is a tool that simulates spatial averages and has been recognized as valid and powerful in the wireless communication world; herein, it is applied to generate and study the spatial distribution of PLC nodes (residing in the radio base stations) yielding, when wiring is considered, a random topology model whose idea was already presented in [11].

In our model, base stations are deployed through the following random processes:

- The number of base stations to be deployed is determined through random Poisson's process, whose average depends on the area that each cell covers (it is supposedly constant for every base station) and the quantity of geographical territory to be covered.
- A two-dimensional coordinate plane simulates our geographical territory; each cell's position is determined as a pair of uniform random variables.

A certain degree of overlapping is also allowed between different cells in order to account for neighboring houses or even cells in the same building but located on different floors. **Figure 1** shows a random deployment of cells. The small red circles represent the connected area reached by each base station located in their centers. The served area is the same for all the cells. To implement the specific deployment in **Figure 1**, a 20% value of covered territory was set.

2.2. Power line infrastructure for the small cell network

Once the small cells are deployed, the supporting power line infrastructure is created. Some assumptions are made to develop the model, which is based on the European paradigm

http://dx.doi.org/10.5772/intechopen.73390

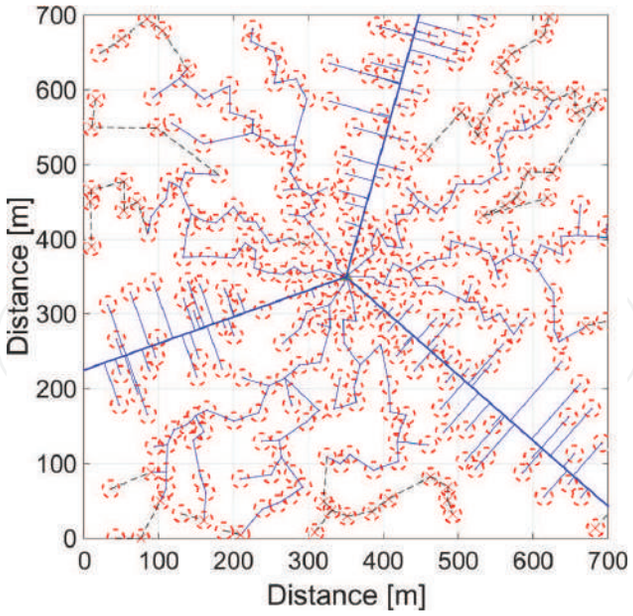

Figure 1. A random network deployment is shown, as produced by the implemented simulation software. The dashed circles represent small cells, while the continuous lines represent power cables from the supply infrastructure.

presented in [6]. This basically means the following things: the network is assumingly structured in a radial architecture, whose branches are all departing from a central concentrator that acts as a front-haul gateway (FHG), which may be located in a medium-voltage (MV) to low-voltage (LV) transformer station. In **Figure 1** the FHG is shown as a triangle in the center of the area. This is where high-speed data links are assumingly realized towards the core network, so the LV network of power lines might enable the backhaul portion of the network, between the private femtocells and the FHG.

On average, from each MV/LV transformer station, a maximum of ten branches are deployed to feed the end-users. Each one of these branches supplies a maximum of 30 to 35 buildings and extends to a maximum distance of 1 km from the FHG, measured in power line cable length. The LV network is the most geometrically complex part of the supply network: the model described in the following stems from observation and real-life cases described in literature [6].

In our model, different topologies for the power line infrastructure are realized to depict different geometrical properties of real networks. The bus topology drives a central backbone through the territory and connects the nearby cells following a minimum policy regarding the global amount of cable used; then, it draws branches to connect the cells to the backbone. The tree and chain topologies connect directly the cells to each other, preferring, respectively, a branching or serial connection policy. The only condition in the creation of these two structures is to avoid intersections of the power line medium between different segments.

While this completes the description of the physical layout of the network, the mobile traffic that is generated needs to be addressed and evaluated to define a requirement on the PLC backhaul technology.

2.3. Traffic generation models

We assume that small cells enable two different types of traffic: voice and data. Based on [4], we assume the probability of each type of call to be 3% for voice and 97% for data. We consider Poisson's distribution for the number of connection requests performed by the population of end-users; this, however, requires to know beforehand the average number of calls per time unit. This parameter is tuned in the model to account for different traffic loads in the considered networks.

Voice traffic is analytically treated as a phone call: 64 kbps are allocated to the user for the whole duration of each call, which follows an exponential distribution. This makes the whole statistic of voice traffic requests Poissonian. On the other hand, data traffic is trickier to model, as amount of data and duration of each communication are variable and co-dependent. This can be modeled through self-similar processes, such as the Pareto and Weibull distributions, which implement the well-known 80/20 rule [12]. Data to model the related probability distribution functions were retrieved from the same paper.

The developed network manages the traffic and resources on the PLC branch under analysis. It is implicitly assumed that resource management of the radio users inside the femtocells is carried out by the cell computational unit itself, and the same goes for the front-haul gateway (FHG) for the traffic to be routed towards the core network.

Results regarding the requirements on the PLC technology are reported in Section 5.1.

3. Transmission line-based bottom-up statistical PLC model

Transmission line theory allows to describe the behavior of a cable used as a telecommunication channel by modeling its medium as a circuit and including effects due to the propagation of short-length waves [6]. This method takes into consideration the nature of the transmission medium by means of:

- Propagation parameters, which allow to describe how signals are attenuated and distorted along the power lines based on the cable type
- Geometry of the cables, meaning how they are actually deployed on the territory based on length and direction
- Topology of the network, how cells are connected to each other and how their presence introduces interference through signal reflection on the channel

Translating this theory into a mathematical model enables bottom-up analysis of the network, as it starts from its lowest layer (physical) and builds models upon it to retrieve typical performances in a deterministic fashion. On the other hand, top-down approaches use real data measurements to infer the possible behavior of the network. As this application for broadband outdoor PLC, to our best knowledge, is not yet described in literature through statistical data, the former type of analysis was preferred.

While geometry and topology are defined in the first part of the random network generation, propagation parameters are generated afterwards based on the type of cable that is chosen to represent the network.

The method herein employed is the same described in [11]. Basically, the medium whose channel behavior is to be calculated is split into different elementary sections, and for each one of these sections, a transfer function is calculated. Sections of the medium are delimited either by loads (private premises) or discontinuities (change of cable type). Since the power line channel is not symmetrical, it is important to identify the direction of the signal that is being transmitted and whose attenuation profile we want to obtain. This is due to the fact that the transfer function of a certain section depends on all the previous sections passed by the signal. In this phase of the experiment, we consider the cells as transmitting towards the FHG. We consider the MV/LV transformer station to completely separate the channels of different sectors; thus, each one of the main branches departing from the station accounts for its own set of channel responses.

The channel under analysis sweeps the frequency domain between 1 MHz and 30 MHz, as this is the range of current PLC applications, although performances at higher frequencies are being explored [13]. Each private premise is modeled as an impedance whose value is statistically chosen in a uniform interval between 5 and 200 ohms, as the frequencies of the broadband channel impedances tend to drop to very low magnitudes (order of a few ohms). Further information on impedance modeling for broadband PLC channels can be found in [14]. The channel is also assumed to be time invariant for simplicity; in reality the channel in PLC is LPTV (linearly periodic time variant), although mostly at low frequency and only in the presence of strongly time variant loads connected closed to the transmitter or receiver (**Figure 2**).

The channel is evaluated for each cell of the network and each one of the 100 KHz subchannels. The evaluation of the channel is a computation of the chain of transfer functions between the receiving node and the transmitting one, which returns a channel gain as a ratio between the magnitudes of voltages at the ports of the two nodes. When combining this data with the transmitting power of nodes and the power spectral density of the noise, it is possible to retrieve the upper bound of capacity of the physical link through the Shannon formula. By using this method, it is implicitly assumed that inputs and noise behavior are Gaussian, which

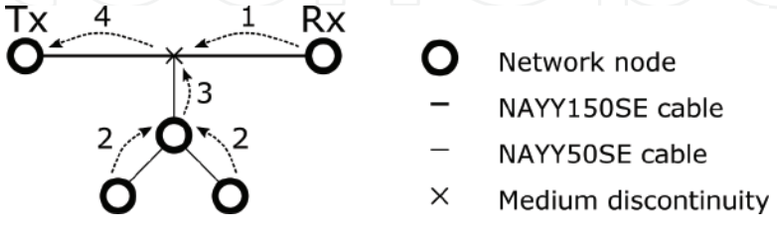

Figure 2. The channel transfer function is computed with the voltage ratio approach method. Starting from the receiving node, each section voltage ratio is used to retrieve the transfer function of that portion of the channel. Whenever a branch is found on the backbone connecting the receiving and the transmitting nodes, it is necessary to compute the transfer function related to this part of the network external to the communication path in order to model the reflections and attenuation it introduces.

is not actually the case of PLC channels, but the results can still be used as a frame of reference to a very well-documented case in literature.

$$C_{TOT} = \sum_{i=1}^{N} B_i \log_2(1 + SNR_i) \tag{1}$$

3.1. Remarks

As mentioned, this transmission line analysis allows to characterize the transfer function of the channel for each cell based on the power line distance of the cell from the FHG, the local density of loads connected to the power line infrastructure and the power spectral density of the noise on the channel.

This transfer function represents the channel gain of the signals transmitted between two points as a chain of voltage ratios, which can be used to compute the Shannon link capacity. The average channel gain (ACG) is the value of channel gain average over the whole spectrum considered for data transmission [15] and is thus strongly correlated to the value of this capacity upper bound calculated with Shannon's formula. By creating a large database of simulations through this model, it is possible to let a node infer the capacity of the medium simply by measuring the intensity of the noise on the cables, thus reducing the computational overhead load needed to map the channel and create an adequate scheduling. This would allow a node (the FHG in a downlink situation, a private cell in the uplink) to infer the capacity between itself and a receiver without sounding the medium.

4. Resource allocation and optimization

Each link between the FHG and a generic cell is characterized by a capacity value. This allows the server of the network, located in the FHG, to allocate the time resource to all the users that in each instant require a certain amount of throughput to connect to the core network.

The MAC protocol is assumed to be a TDMA as the channel is assumed time invariant; nevertheless, the time resource is divided into time frames as the PLC channel displays an LPTV (linear periodically time variant) behavior to simplify the integration of frequency variability in a more advanced stage of the study. The time frame naturally is meant to represent the duration of the mains cycle in an AC power supply network. As a frame of reference, in the European paradigm, this duration equals 20 ms.

Resource allocation schemes are applied to traffic requests that are generated over the network according to the models and metrics discussed in Section 2.3. Whenever one of the small cells generates a connection request, a throughput requirement is created and added for the whole duration of the call to the throughput already required during the said time range. As mentioned before, connection requests can be of two kinds, specifically voice or data. The former kind requires a continuous connection and a fixed throughput of 64 kbps for the whole

duration of the call, while the latter generates a bigger quantity of traffic not necessarily requiring continuity.

Initially, the network's performance is evaluated by operating a very simple TDMA protocol: during each time frame, the time frame is equally shared between the pools of users requesting access to the medium. If the allocated resource completely covers the required throughput for one or more users, then a smaller chunk is assigned to these users, and the remainder of the resource is shared accordingly.

The TDMA method described above is ideal as it allows a continuous allocation of the time resource, while real communication standards divide the time frame into slots that can be distributed to users. This is due to the fact that the time resource must be split into a finite number of units to respect the packet structure of data transmission. HPAV [16] envisions a duration of the elementary OFDM symbol in the tens of microseconds, which is quantitatively adapted here to divide each frame into 500 elementary symbols. For the optimization problem presented in the following, the time frame is divided into time slots in order to satisfy a system where the number of users and the magnitude of their throughput requests in each time frame are not fixed. Each time slot consists of an integer number of elementary symbols such that the total number of symbols is eventually covered by an integer number of time slots.

The aim of the optimization problem is structured in the following way: first, all the users requesting resource to satisfy calls of the voice kind are assigned enough resource to cover this throughput. When voice calls are satisfied, the rest of the throughput requests are considered, and the rest of the resource is allocated according to the following:

$$\begin{aligned} &\max_{\alpha, N_s} \sum_{k}^{N_U} \sum_{j}^{N_S} \alpha^{(k,j)} C_k(N_S), s.t. \\ &(i) \sum_{k}^{N_U} \alpha^{(k,j)} = 1, j = 1, 2, \ldots, N_S \\ &(ii) \max_p \sum_{j}^{N_S} \alpha^{(k,j)} C_k(N_S) \geq \frac{p}{100} C_k, k = 1, 2, \ldots, N_U \end{aligned} \tag{2}$$

where N_S and N_U equal, respectively, the number of slots in which the time frame is divided and the number of base stations operating an access request during the current time frame, $\alpha^{(k,j)}$ is the logic operator that equals 1 when the jth time slot is assigned kth base station and C_k is the theoretical capacity of the link between the FHG and the kth base station in case where the whole resource is assigned to the said base station, while $C_k(N_S)$ is the capacity that the kth base station gets from a time slot when the time frame is split into N_S slots. The p parameter allows the equalization of the resource, and it is maximized in order to allow each base station to exploit at least p (%) of the capacity of the link towards the FHG. Condition (i) makes sure that each time slot is assigned to only one user. Condition (ii) on the other hand assures that each user is assigned enough resource to exploit a minimum percentage (p/100) of their actual capacity. The aim of the optimization process is to maximize the aggregated throughput in

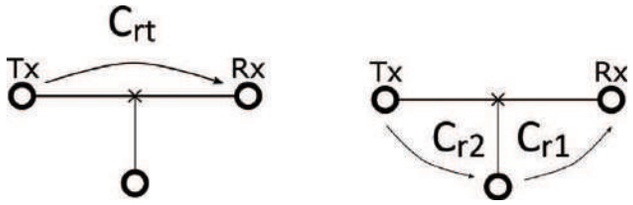

Figure 3. Routing enables a better coverage of the cells far from the FHG by decoding or amplifying the message received from the transmitting node before forwarding it to the receiver. In the network depicted on the left, the central cell is only creating interference and reflections on the PLC medium, while on the right is tentatively considered as a repeater for routing purposes.

each sector under a condition of resource scarcity by changing the overall number of time slots and the number of slots assigned to each user.

4.1. Enabling routing

The optimization problem can also be updated to include routing in it. A node can be chosen to act as a repeater based on the capacity of the channel between the FHG, the repeater and the cells that would be reached by the extended coverage. This also requires to distinguish between uplink and downlink cases because of the asymmetry of the PLC channel.

Simple routing can be enabled by assigning the role of repeater to one of the cells of the network, depending on the typical throughput required by a generic cell and the capacity of the channel between the end-user cell and the FHG with or without the repeater.

Figure 3 shows a case where only one cell is considered for routing-enabled extended coverage. This technique can actually be used to reach multiple cells in the region far from the FHG. Firstly, for each cell to be tentatively covered with routing, it is necessary to consider if the capacities created by a hopped communication give a better performance than the original one:

$$\begin{aligned} C_{DT} &= C_{rt} \\ C_{DF} &= \min\{\tau C_{r2}, \tau C_{rt} + (1-\tau)C_{r1}\} \end{aligned} \tag{3}$$

where C_{DT} and C_{DF} represent, respectively, the power line link capacity in the direct transmission case and the one where Detect-and-Forward strategy is used in the relaying node, as per **Figure 3**. Also, τ represents the fraction of the time frame that is dedicated to a specific connection. It is possible to adapt the problem to a case where the repeater needs to cover multiple cells. Basically, optimization is operated by tuning the aforementioned parameters so that the resource is used at its best. More details about this can be found in [17].

5. Results

In this section, numerical results from the simulations are presented. All the simulated networks consider small cells with a round coverage area whose surface is constant across

the whole territory, while the simulation area (visible also in **Figure 1**) is square shaped. The length of the side equals the supposed average distance between neighboring FHGs. Cells that are further than a kilometer away (considering power line distance) from the FHG are considered as not connected neither to the communication nor to the supply network. This set of geographical parameters enabled the representation of both sparsely populated areas (as rural environments) and strongly occupied ones (as urban settings) through the tuning of the density parameter. By varying gradually this parameter, it is possible to see how a change in population affects the requirements imposed on the PLC technology and its overall performance. In Section 3 it was mentioned that the type of cable affects the behavior of the network: in our case, two types of cable were used to describe the network, specifically NAYY150SE for the LV lines that depart from the transformer station and NAYY50SE to connect this main bus to the point of connection of private premises. Furthermore, for the resource allocation problem, only cells with a nominal capacity over a certain threshold are considered as covered by service. It is also supposed that the FHG is transmitting with a power spectral density of −50 dBm/Hz and that the channel is characterized by a Gaussian noise with a power spectral density having a floor of −140 dBm/Hz. To evaluate the performance of the network, the grade of service (GoS) will be used. This is defined as the ratio between the throughput supported (ST) by the technology and the one required (RT) by the aggregated users across all the base stations:

$$GoS = \begin{cases} 1, RT \geq ST \\ \dfrac{ST}{RT}, otherwise \end{cases} \tag{4}$$

With the model described in Section 2, it was possible to assess preliminarily how the cells generate traffic and how they define a requirement for the PLC technology to enable broadband access. It was found that, when generated traffic resembled models of real cellular networks, the magnitude of this traffic was compatible with the highest capacity found on the PLC medium in outdoor applications. For a more detailed report, please refer to [18].

As mentioned before, transmission line theory enables bottom-up analysis of the network performance. In **Figures 4** and **5**, it is possible to see how the channel varies its behavior based on frequency and the power line distance between the communicating nodes; this behavior changes for different densities of cells deployed on the geographical territory.

Figure 4 shows how the channel gain behaves at different frequencies and power line lengths of analyzed links for a fixed density. In this specific case, the density was tuned in order to obtain an average of 35 cells for each one of the main branches departing from the FHG: this is reportedly [6] the maximum number of private premises fed by one main LV bus in the European paradigm. As in the Japanese/American paradigm, this number is much lower; it made sense to consider the European one as it sets the strictest conditions on the capacity of the channel. The underlying scatter plot shows the actual real channel gain of cells at different frequencies and distances; from the clear pattern these data create, a second-order interpolation was extracted in order to easily describe the behavior.

Figure 5, on the other hand, shows how the capacity of a generic link between a cell and the FHG can be correlated to the ACG of the same link. As seen in **Figure 4**, it is possible to infer the ACG of a link by knowing the distance of a cell from the FHG and the density of cells in the territory. Furthermore, it is easily possible to see how the ACG strongly depends on the length

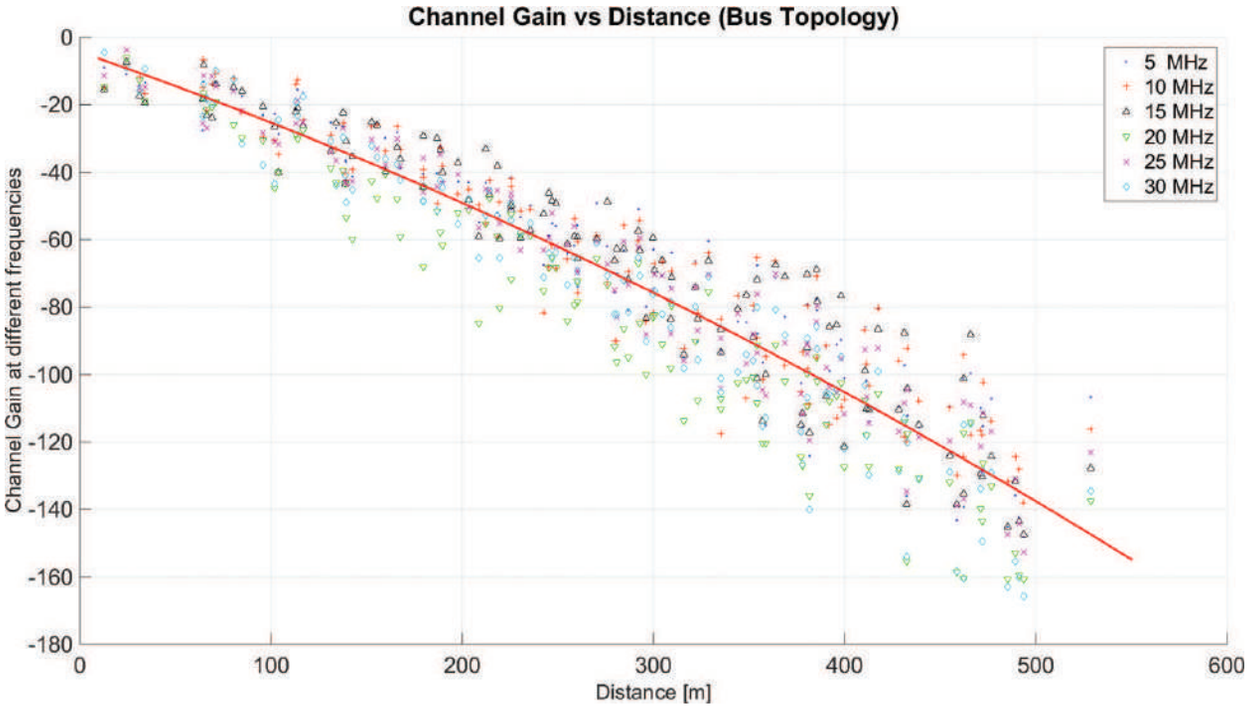

Figure 4. The scatter plot shows the channel gain computed with Shannon's formula and transmission line-based techniques in relation to the power line distance of the considered cells. A pattern is very clear; thus, a second-order interpolation was driven to retrieve the central red curve, which represents the ACG. This depends on the density of cells on the territory.

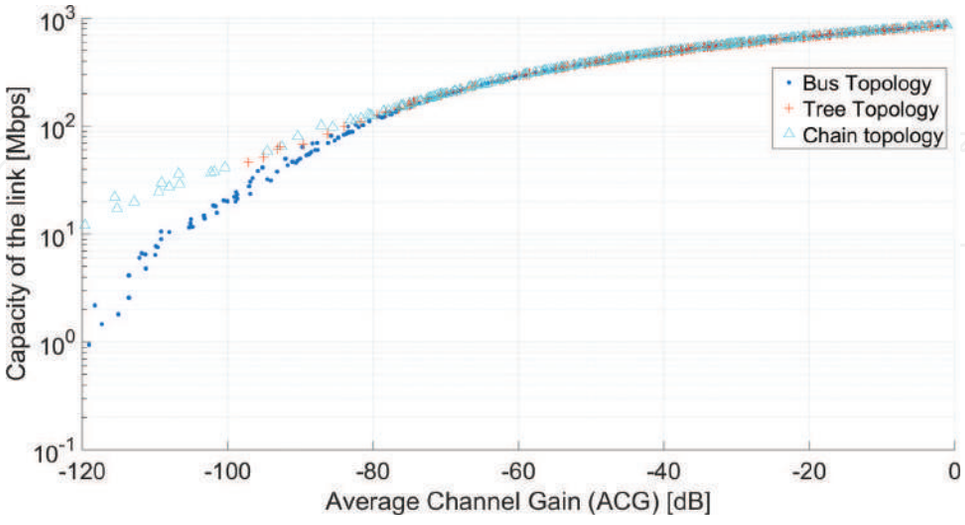

Figure 5. This semilogarithmic scatter plot shows the relation between the capacity of a link and the ACG attributed to the same. This relation does not depend on the distance; thus, it is possible to use it to infer a cell's capacity towards the FHG by measuring the ACG of the link.

of medium a signal must go through to enable communication. In **Figure 6**, the FHG (represented with a triangle in the center of the network) is always considered the transmitting node, thus enabling a downlink kind of communication. As mentioned in Section 3, it is possible to correlate this value with the capacity of the link between the FHG and the generic cell, when a few parameters of the network are known, such as distance of the communicating cell, density of the cells and power spectral density of the noise on the medium. Analysis from further simulations revealed that capacity correlated to the ACG neither depend on distance nor density of cells in the network. With reference also to **Figures 4** and **5**, it is possible to see that for a high-density network, coverage can reach almost 400 meters of power line length. In situations of lower cell density, cells that are further from the FHG can be covered by the PLC connection service. It follows that PLC coverage depends on distance also in relation to the density of base stations in the territory and their average coverage, i.e., the number of loads connected to the same shared medium.

Figures 7 and **8** report the grade of service (GoS) and the average aggregated throughput on the main buses departing from the FHG for the evaluation of the time resource assignment. The first one represents the network's performance when a simple TDMA protocol is employed, namely, each transmitting user gets a time slot whose duration is inversely proportional to the total number of transmitting users, while the second one shows the same values when the optimization process is operated. The average aggregated throughput is calculated as the sum of the effective throughput carried on the medium. Each user contributes to this figure either with the minimum value chosen between the throughput assigned by the FHG and the throughput required to satisfy the users in the cell. On the other hand, the grade of service is how much the throughput required by the users in the cell is actually satisfied: 1

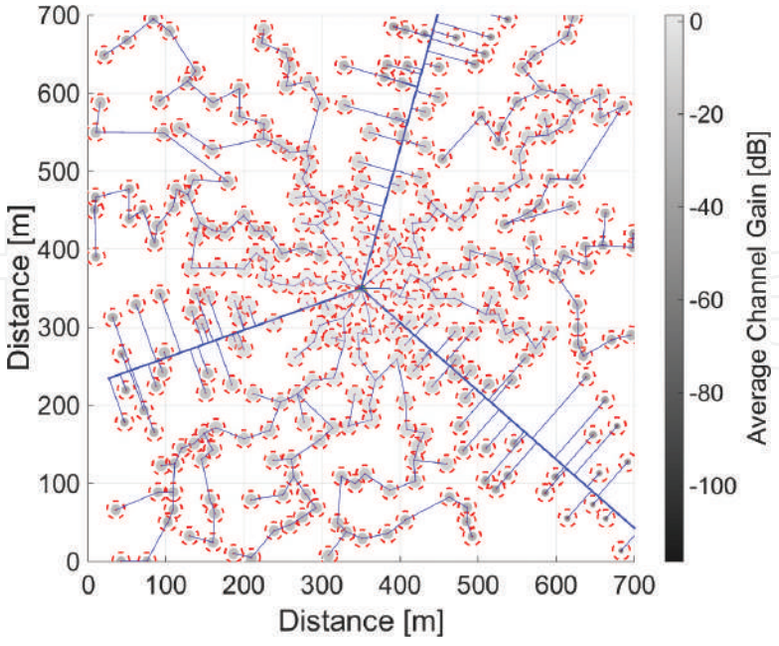

Figure 6. In this randomly generated network, a map of colors is used to show how the ACG fares in comparison with topology and distance of the cells from the FHG. As mentioned in section 5, cells whose ACG is circa −120 dB or higher can achieve a capacity of about 1 mbps or more.

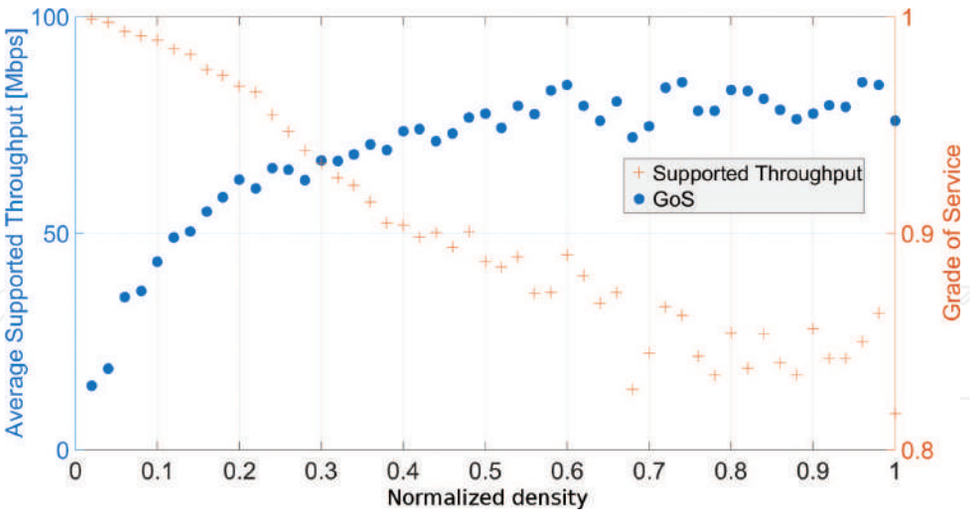

Figure 7. This graph shows the average supported throughput on each main bus departing from the FHG and the grade of service in relation to the normalized density of cells in a network where a simple TDMA protocol is applied.

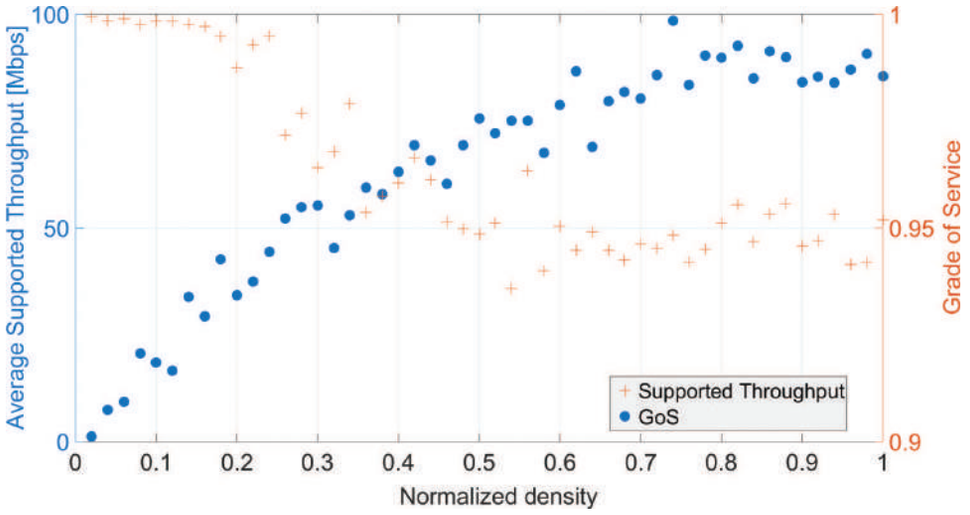

Figure 8. This graph shows the average supported throughput on each main bus departing from the FHG and the grade of service in relation to the normalized density of cells in a network where the optimization of the resource is carried out.

accounts for a network where all the users are satisfied, whereas 0 represents the opposite condition.

Also, for each density step considered for the simulations, the p factor mentioned in the definition of the optimization problem is reported to show how much resource is on average assigned to the requesting users to satisfy the generated calls. Its trend can be observed in **Figure 9**. This value is higher for low-density values because of the low number of users that request connections to the FHG.

From **Figures 7** and **8**, it is possible to see that the GoS is behaving better for the cases where the optimization is carried out. On the other hand, for lower densities, optimization seems to decrease the average supported throughput, probably due to the granular nature of the time

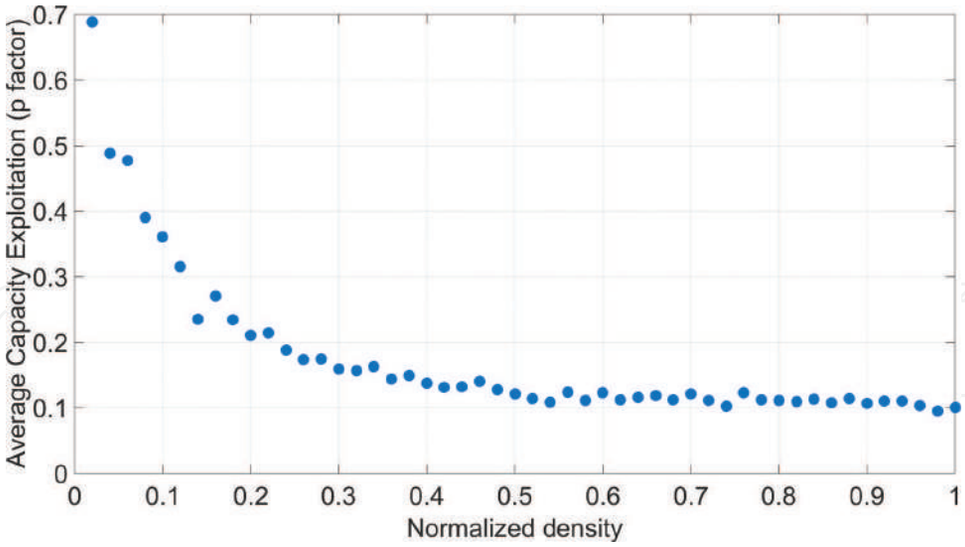

Figure 9. When the optimization is carried out on the time resource in a network, the p factor actualizes how much of the nominal capacity of the link towards the FHG each cell is able to exploit, based on how much time resource is assigned to each user.

resource in the optimization case versus the continuous one assumed in the simple TDMA case used for the former graph.

6. Conclusions

PLC is regarded nowadays as an enabling technology for IoT, smart grid and smart home applications. This study explores the possibility of using PLC as a broadband access solution for radio access networks by exploiting traffic generation models, transmission line theory and an optimization process to show how a TDMA protocol can be improved upon to let a central coordinator element schedule the time resource in an appropriate manner.

The experiments carried out on the simulation level showed how it is possible to employ PLC as a technology for broadband access in radio access networks. Studies carried out in the past showed which requirements were put on the shared medium, whereas here a general model for PLC performance was developed to infer which performance the technology is fit to provide. Specifically, herein it was shown that the technology is able to support connectivity for a small network of private femtocells covered by a LV power supply network, ranking as an adequate last mile solution. It was also shown how this performance can be enhanced for high-density networks by employing an optimization method, aiming at maximizing the supported throughput.

Moreover, it is mentioned in additional remarks (Sections 3.1 and 4.1) how the techniques used in the study will be exploited more in the future to gain a deeper insight into how hybrid broadband PLC-radio access networks can be designed to offer a better connection service to the end-users that populate them, namely:

- The transmission line theory model will be employed in conjunction with the topological and geometrical information to develop a system to infer the capacity of the link based on these information, which do not require a sounding of the channel, thus decreasing overhead and enhancing data speed.
- By using the method above, it will also be possible to consider different methods of routing inside the network to extend the coverage provided by the FHG. This will involve considering the combined capacities of the links that are created by partitioning the network. This will also depend on the type of link enabled, be it uplink or downlink.

A more realistic model could be developed by including different cell coverage sizes while also considering a more realistic behavior of users in the base stations via a traffic generator based on inferred statistics. Future endeavors will be devoted to experimentation in the field.

Author details

Francesco Marcuzzi* and Andrea M. Tonello

*Address all correspondence to: francesco.marcuzzi@aau.at

Networked and Embedded Systems Institute, Alpen-Adria-Universität Klagenfurt, Austria

References

[1] Cerwall P, Jonsson P, Carson S, editors. Ericsson Mobility Report (June 2017). 1st ed. Heuveldop N; 2017. p. 36

[2] KMS H, Rodriguez J, editors. Backhauling/Fronthauling for Future Wireless Systems. 1st ed., Chichester, United Kingdom: Wiley; 2016. p. 232

[3] Osseiran A, Boccardi F, Braun V, Kusume K, Marsch P, Maternia M, et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. IEEE Communication Magazine. 2014;**52**(5):26-35. DOI: 10.1109/MCOM.2014.6815890

[4] Quek TS, de la Roche G, Güvenç İ, Kountouris M, editors. Small Cell Networks: Deployment, PHY Techniques, and Resource Management. 1st ed., Cambridge, United Kingdom: Cambridge University Press; 2013. DOI: 10.1017/CBO9781139061421

[5] Small Cell Forum, editor. Backhaul Technologies for Small Cells: Use Cases, Requirements and Solutions. Small Cell Forum; 2013. p. 78

[6] Lampe L, Tonello AM, Swart T, editors. Power Line Communications: Principles, Standards and Applications from Multimedia to Smart Grid. 2nd ed., Chichester, United Kingdom: Wiley; 2016. p. 618. DOI: 10.1002/9781118676684

[7] Cano C, Pittolo A, Malone D, Lampe L, Tonello AM, Dabak AG. State of the art in power line communications: From the applications to the medium. IEEE Journal on Selected Areas in Communications. 2016;**34**(7):1935-1952. DOI: 10.1109/JSAC.2016.2566018

[8] Tonello AM, Marcuzzi F. Statistical assessment of PLC networking for front-hauling in small radio cells. In: WSPLC 2016; Oct 10–11 2016; Paris, France. IEEE; 2016

[9] Marcuzzi F, Tonello AM. Stochastic geometry for the analysis of small radio cells and PLC back-hauling. In: SCC 2017; 11th International ITG Conference on Systems, Communications and Coding; Proceedings of; 6–9 Feb. 2017; Hamburg, Germany. VDE; 2017

[10] ElSawy H, Hossain E, Haenggi M. Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey. IEEE Communications Surveys & Tutorials. 2013;**15**(3):996-1019. DOI: 10.1109/SURV.2013.052213.00000

[11] Tonello AM, Versolatto F. Bottom-up statistical plc channel modelling - part I: Random topology model and efficient transfer function computation. IEEE Transactions on Power Delivery. 2011;**26**(2):891-898. DOI: 10.1109/TPWRD.2010.2096518

[12] Han Y, Seo S, Hwang C, Yoo J, Won-Ki Hong J. Flow-level traffic matrix generation for various data center networks. In: Network Operations and Management Symposium (NOMS), 2014 IEEE; 5–9 May 2014; Krakow, Poland. IEEE; 2014. DOI: 10.1109/NOMS.2014.6838394

[13] Versolatto F, Tonello AM. PLC channel characterization up to 300 MHz: Frequency response and line impedance. In: GLOBECOM 2012; Global Communication Conference; Proceedings of; 3–7 Dec 2012; Anaheim, CA, USA

[14] Pittolo A, Tonello AM. A synthetic MIMO PLC channel model. In: ISPLC 2016; 20th International Symposium on Power Line Communications and its Applications; Proceedings of; 20–23 Mar 2016; Bottrop, Germany

[15] Tonello AM, Versolatto F. Bottom-up statistical PLC channel Modelling - part I: Inferring the statistics. IEEE Transactions on Power Delivery. 2012;**25**(4):2356-2363. DOI: 10.1109/TPWRD.2010.2053561

[16] Tonello AM, Cortes JA, D'Alessandro S. Optimal time slot design in an OFDM-TDMA system over power-line time-variant channels. In: ISPLC 2009; 13th International Symposium on Power Line Communications and its Applications; Proceedings of; 29 Mar-1 Apr 2009; Dresden, Germany

[17] D'Alessandro S, Tonello AM. On rate improvements and power saving with opportunistic relaying in home power line networks. EURASIP Journal on Advances in Signal Processing. 2012;**2012**:194. DOI: 10.1186/1687-6180-2012-194

[18] Marcuzzi F, Tonello AM. Stochastic geometry for the analysis of small radio cells and PLC back-hauling. In: SCC 2017; 11th International ITG Conference on Systems, Communications and Coding; Proceedings of; 6–9 Feb 2017; Hamburg, Germany

IntechOpen

Chapter 6

Co-Channel Interference Cancellation for 5G Cellular Networks Deploying Radio-over-Fiber and Massive MIMO Beamforming

Sheng Xu

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72727

Abstract

In future fifth-generation (5G) wireless cellular networks, distributed massive multiple-input multiple-output (MIMO) techniques will be applied worldwide. Recently, much more challenges on efficient resource allocation to large numbers of user equipment (UE) are raised in order to support their high mobility among different micro-/pico-cells. In this chapter, we propose a framework to enable an optical back-haul cooperation among different optical network units (ONUs) with distributed MIMO techniques in wireless front-haul for next-generation optical-wireless cellular networks. Specifically, our proposal is featured by a downlink resource multi-cell sharing scheme for OFDMA-based passive optical network (PON) supporting radio-over-fiber (RoF). We first consider system architecture with the investigation of related works, and then we propose a co-channel interference mitigation and delay-aware sharing scheme for real-time services allowing each subcarrier to be multi-cell shared by different active ONUs corresponding to different micro-/pico-cells. Furthermore, a heuristic algorithm to mitigate co-channel interference, maximize sharing capacity, and minimize network latency is given by employing the graph theory to solve such sharing problems for future 5G. Finally, simulations are performed to evaluate our proposal.

Keywords: co-channel interference, 5G, distributed MIMO, passive optical networks, OFDMA, radio-over-fiber

1. Introduction

The increasing demand of real-time services (e.g., VoIP, the video telephony, and streaming) poses high requirements on communication quality (e.g., interference mitigation and delay constraint) and bandwidth increase in the network for the future era of big data [1]. Nowadays, the OFDMA-based passive optical network (PON) has been applied to provide such a large-capacity and also high-flexibility solution for wireless cellular networks with radio-over-fiber (RoF) technology [2, 3].

Figure 1 describes a RoF-based optical-wireless system adopting OFDMA-PON, while one of prominent challenges in such networks for future 5G communications is the algorithm for effective resource allocation. In the related works, a dynamic bandwidth allocation (DBA) in OFDMA-PON has been implemented in [4] with fixed subcarriers for data scheduling, which adopts a traditional grant/report polling scheme. Moreover, dedicated resource allocation (DRA) and shared resource allocation (SRA) as two DBA methods were proposed in [5]. The DBA protocol in OFDM-PON therefore has been proposed in [6], where protocols are summarized in two schemes: the fixed burst transmission (FBT) and the dynamic circuit transmission (DCT). FBT employs a round-robin, IPACT algorithm [6] while DCT adopts bandwidth estimation. Furthermore, a power-efficient DBA scheme of OFDM-PON has also been given in [7] for the purpose of minimizing the optical network units (ONUs) transmitting power. In addition, a lot of works such as for attaining the low power consumption with OFDM-PON have been finished on a system hardware level. Specifically, a 36.86-Gb/s optical wavelength conveying six 100-MHz-bandwidth LTE-A signals has been proposed in [8]. The system supports 5-carrier aggregation, 2 × 2 MIMO, and three sectors, over a 40-km SSMF front-haul adopting a single 1550-nm directly modulated laser. In addition, the system [2] adopts a fixed RF channel on subcarriers; however, it becomes inflexible to satisfy DBA when high mobility of large number of user equipment (UE) occurs in the wireless front-haul. The structure [2, 3] deploys an optical distribution network (ODN), which is different with [4, 5], but the DBA problem is still the same.

However, considering a very high density level of UEs and their high mobility in future 5G cellular networks, a prominent problem waiting to be solved is the co-channel interference jointly employing radio-over-fiber, massive multiple-input and multiple-output (MIMO), and beam-forming [9] technologies. For example, in optical-wireless networks, when the same wireless frequency resources carried by different optical wavelengths overlap in the same beam direction

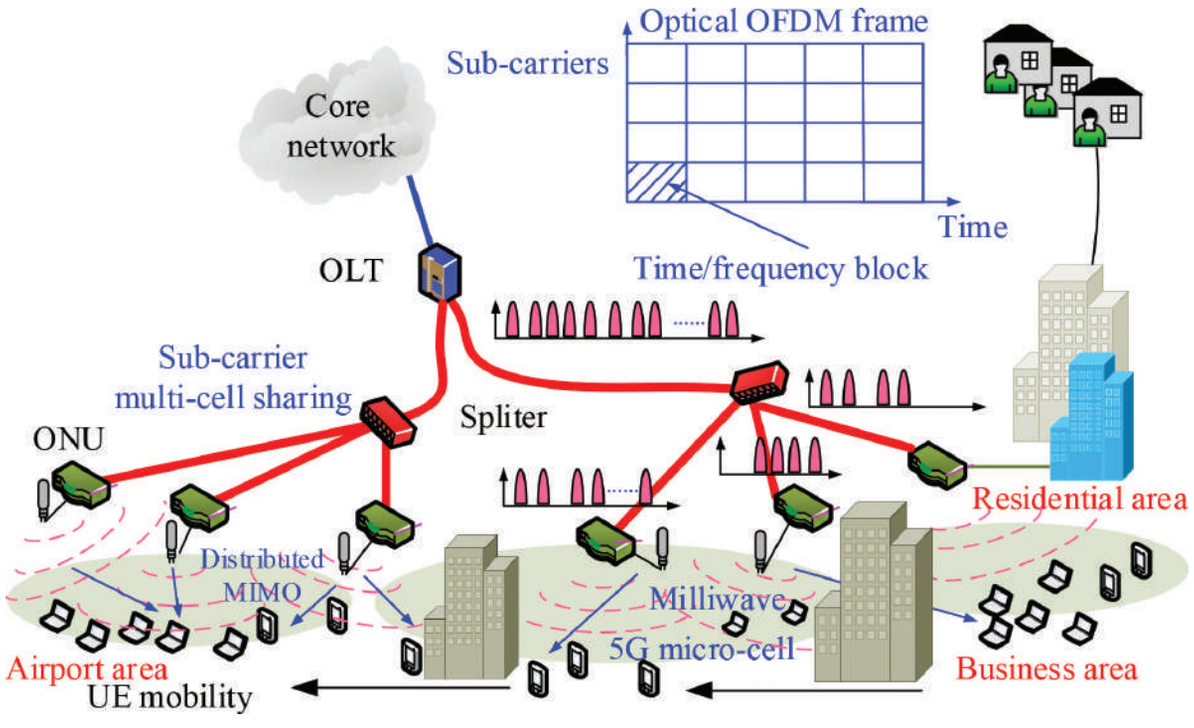

Figure 1. Network architecture of RoF-OFDM-PON based on 5G [9].

and are received by different UEs, these UEs which emerge in this beam direction under their mobility will suffer high interference. Typically, in the case of distributed massive MIMO, numbers of UEs move frequently among a few of micro-/pico-cells in any time; wireless data carried over the wavelength are shared by all served UEs in the pico-cell and adjacent cells. The minimizing of co-channel interference as mentioned will become much more imperative. Hence, it is expected to seek a scheduling optimization of wireless resources to each micro-cell mitigating the interference due to high UE mobility. Furthermore, we consider a given limited number of optical subcarriers, when an ONU needs additional resources, and in order to support the bandwidth demand for the rest of ONUs, optical subcarriers will not be reallocated in congestion cases, the problem herein is also to find ways to share optical subcarriers among local different cells by ONUs. However, it brings the additional delay problem and also configuration and control problem for selecting ONUs because of resource sharing and transmission. To achieve these targets, we propose and observe an interference mitigation and delay-aware sharing scheme for real-time services allowing that each subcarrier of RoF-OFDM-PON [2] can be multi-cell shared by UEs accessed from different micro-cells. Namely, each UE is arranged to receive multiple data streams demodulated from different ONUs simultaneously.

In this chapter, we address the aforementioned problems in the system, which have not been studied in other works before. The proposed method in this chapter can be employed by a future 5G operator to run radio-over-fiber based optical OFDM (OOFDM) [3] networks with multiple micro-/pico-cells as shown in **Figure 1**; it could be used as a method on network design to reduce resource waste and improve the performance of network.

The rest of this chapter is organized as follows. Section 2 presents our system architecture and resource allocation model. Section 3 introduces our resource sharing proposal. A heuristic algorithm guaranteeing minimum co-channel interference, maximum sharing capacity, and minimum delay time is presented in Section 4. Section 5 provides evaluation results with simulations. Finally, the chapter is concluded in Section 6.

2. RoF-OFDM-PON system

2.1. Link architecture in RoF-OFDM-PON networks

Figure 2 illustrates an experimental link architecture for signal processing in RoF-OFDM-PON [10] to support our proposal and to be employed as a physical infrastructure for one data stream in system. From the transmitter side, this implementation firstly modulates experimental data through an OFDM processing with performing of the PRBS, NRZ pulse, and QAM sequence generation (e.g., 4-QAM) in advance [10]. After that, a RF-IQ mixer is used to deal with the OFDM signal to analog RF with a proper quadrature modulation. The output signal then experiences an optical OFDM (OOFDM) modulation with a 193.1 THz CW laser by LiNbO3 mach-zehnder modulator (MZM) [10, 11] and then is sent into fiber through EDFA to amplify the signal.

On each receiver of ONU side, signals from fiber are received by photo-detector (PD) [10, 11] and are executed with a RF de-multiplexing and OFDM demodulation followed by QAM sequence generator and NRZ pulse generator in order to recovery the experimental data [10]. It is important to note that one set of optical OFDM subcarrier on fiber could be modulated to

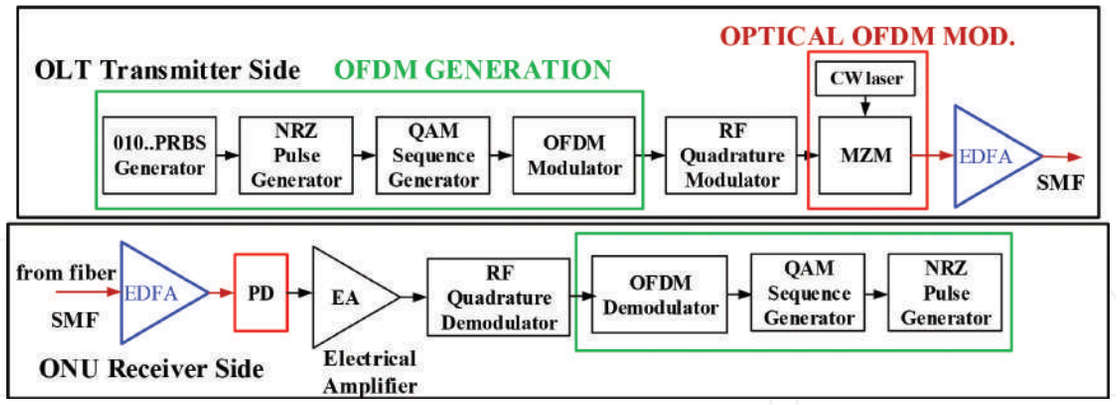

Figure 2. The downlink signal processing of RoF-OFDM-PON.

accommodate different UEs belonging to two or more receivers/ONUs in cellular networks, and the wireless data allocated to UEs belonging to any micro-cell could be transmitted by the broadcasting of multiple sharing streams from other ONUs with antennas in other micro-cells nearby (e.g., by the distributed massive MIMO [9]). In this chapter, our work thus mainly consider these resource allocation problems, while detailed physical discussions on the control and configuration issues (e.g., protocol specification) for dynamic transmission from multiple ONUs for resource sharing are out of the scope of this chapter.

2.2. Optical subcarrier allocation model

Employing the downlink signal processing mentioned in Section 2, the current OFDM-PON access networks flexibly allocate the time/frequency blocks in OFDM frame logically as shown in **Figure 3** under a mixed access rate. **Figure 3** describes an example about resource allocation of optical time/frequency block distributed to three different ONUs under different time slots and optical subcarriers. In this case, multiple wireless UE data are modulated onto each time/frequency block, and each block could be allocated to a single ONU, while each ONU could receive several such time/frequency blocks in the same time slot [12, 13]. Each optical subcarrier of time/frequency block in **Figure 3** could be addressed by a RoF modulation with radio frequency in the same or different wireless radio frequency spectra (e.g., a LTE radio frame frequency spectrum from 2110 to 2170 MHz) [14]. Moreover, according to the bandwidth capacity of single optical carrier, multiple radio frame could be conveyed on a single carrier (e.g., as reported in [8], six 100 MHz LTE-A signals are conveyed on a 36.86 Gb/s optical carrier).

Adopting this subcarrier allocation method, different UEs are fed by its ONU within its cell, and the subcarrier number allocating to each ONU could be appended according to the increase of traffic in this cell. However, the very high UE mobility in future 5G pico-cells [9] results that a few idle resource appears on a signal optical time/frequency block so that much more wasted resource is produced during resource allocation. In order to rationally allocate these idle resources (e.g., the remnant resource in **Figure 4** during slot t_2), it should be noted that the physical optical modulation process in **Figure 2** can be easily controlled to make UE data belonging to different ONUs be modulated onto the same optical time/frequency block with radio-over-fiber, as shown in **Figure 4**. For instance, in **Figure 4**, ONU 3 has idle resources in time slot t_1; however, its data requirement exceeds the allocated amount from time slot t_3 on.

http://dx.doi.org/10.5772/intechopen.72727

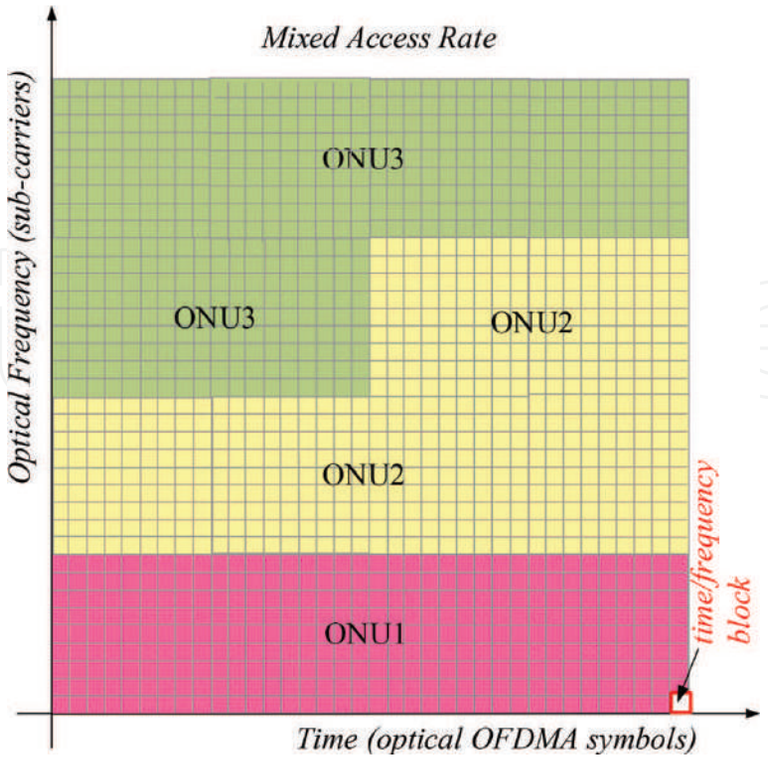

Figure 3. Optical OFDMA frame with time/frequency block allocation to different ONUs in RoF-OFDM-PON systems [12].

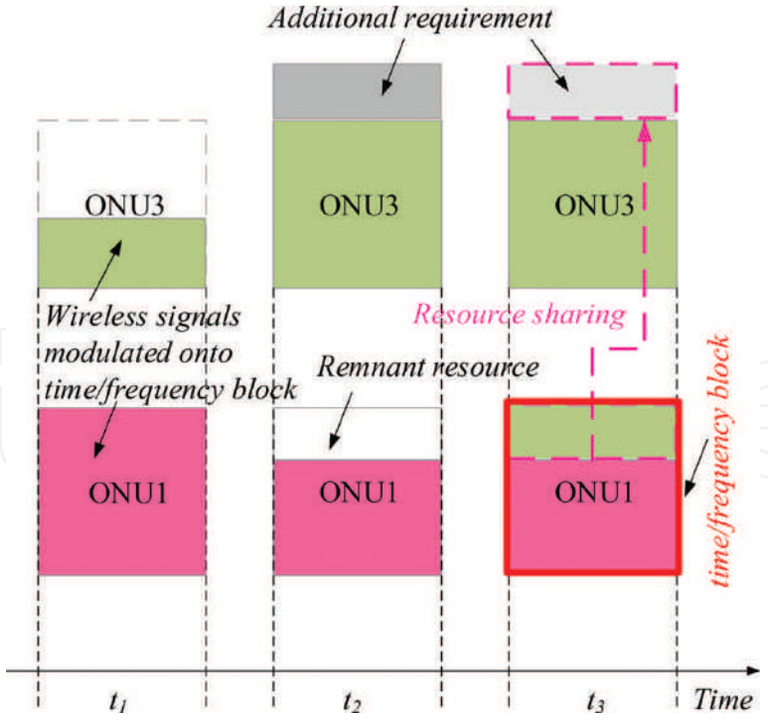

Figure 4. Multi-cell sharing of wireless resources on optical time/frequency blocks allocated to different ONUs corresponding to different wireless cells.

With our proposal, it will receive remnant resources of ONU 1 in time slot t_3 by real-time resource sharing for its additional requirement.

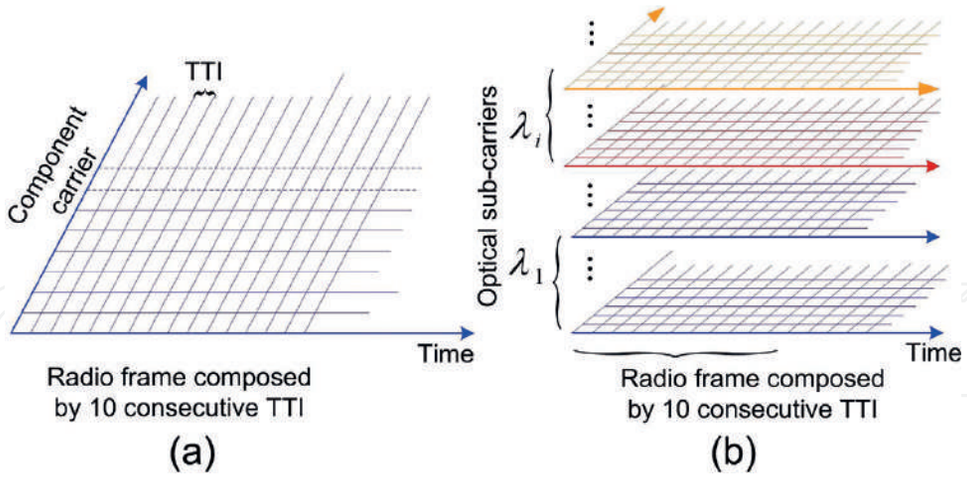

Figure 5. Wireless resource sharing logically on optical time/frequency blocks by different ONUs.

2.3. Radio frame model on optical subcarriers

The wireless spectrum resource can be illustrated by **Figure 5(a)**. In contrast to the single-layer radio frame which is illustrated in **Figure 5(a)**, by allocating more subcarriers, multiple radio frames can be delivered on fiber to each ONU, forming the multi-layer radio frames which are shown in **Figure 5(b)** for each ONU. Note that **Figure 5(b)** describes multiple wireless frames carried by a single optical subcarrier λ_i. Therefore, the time slot in **Figure 5** is different from that in **Figures 3** and **4**, for example, optical scheduling time slot t_1 in **Figure 4** contains several consecutive time slots in **Figure 5** which is named as transmission time interval (TTI) [15].

Note that the smallest optical resource unit in **Figures 3** and **4** is named as time/frequency block, while the smallest radio resource unit in **Figure 5** is named as resource block (RB). They are different concepts in this chapter. Each component carrier (CC) contains several RBs [14, 15]. One UE can receive several CCs in a certain time slot simultaneously.

3. Mathematical optimization

3.1. Assumptions of the model

Assumption 1: There are total $N^i_{\text{sub-c}}$ optical subcarriers allocated to ONU i, and l is the indicator of optical subcarrier. For each l-th optical subcarrier, it contains C_l layers of frames, as shown in **Figure 6(b)**. p is the indicator of frame on each optical subcarrier. The concept of multi-layer frames will be adopted in the following problem description and resource sharing algorithm.

Assumption 2: Denote $R_{k,t}$ as the minimum capacity requirement for user k in slot t. The UE set $\{1, 2,\ldots, \tilde{k},\ldots, \tilde{K}\}$ is served by ONU i, while UE set $\{1, 2,\ldots, k,\ldots, K\}$ is served by ONU j. Denote $N_{\text{sub-c}}$ as the consecutive subcarrier number on frequency of each RB and N_{sym} as OFDM symbol [14] number on time domain of each RB. In addition, denote $N_{\text{sc}}{}^{(\text{d})}{}_s$ as the subcarrier number for date transmission in the s-th OFDM symbol, and $N_{\text{sc}}{}^{(\text{d})}{}_s < N_{\text{sub-c}}$, because of the existence of subcarriers used for control signals in each RB.

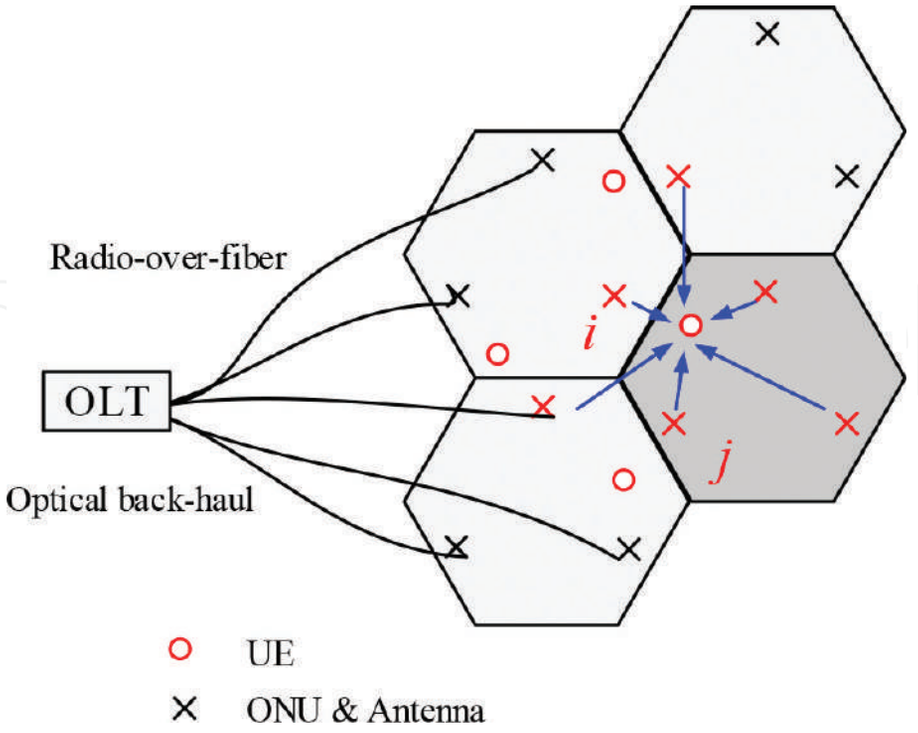

Figure 6. The illustration of multi-cell RoF-OFDM-PON scenarios with distributed massive MIMO deployment. (Each cross represents an ONU/antenna location, while the red circle represents the UE. Six red crosses in different cells highlight that one UE receives multiple data streams in the subcarrier sharing scenario).

Assumption 3: There are J numbers of modulation and coding scheme (MCS) for each RB to choose, and $R_o^{(c)}$ is code rate in a dedicated MCS, $o \in \{1, 2, \ldots, J\}$. M_o describes the constellation size of MCS [14]. In each TTI of scheduling, the capacity $r^{(o)}{}_{RB}$ for one RB under MCS o is then given by Eq. (1):

$$r_{RB}^{(o)} = R_o^{(c)} \log_2(M_o) \sum_{s=1}^{N_{sym}} N_{sc}^{(d)}{}_s \tag{1}$$

Assumption 4: Suppose that $g_{k,l,p,n}$ indicates the wireless channel quality indicator (CQI) [14] of the n-th RB in each CC carried on the p-th optical subcarrier dedicated to UE k. The CQI of RB in each CC carried on the p-th frame on l-th optical subcarrier can be given by $g_{k,l,p} = [g_{k,l,p,1}, g_{k,l,p,2}, \ldots, g_{k,l,p,N_{RB}}]^T$. Each UE can employ at most z CCs to receive data in each slot. Each UE can only adopt one MCS for its assigned RB of CCs.

Assumption 5: We consider beam-forming [7] for wireless signal propagation. In terms of beam direction of the i-th ONU/antenna pair, two categories of UEs' conflict relationship according to any two UEs' locations (from the view point of the i-th ONU/antenna pair) are (1) same angle UEs and (2) different angle UEs. Define a matrix $\chi_i = [\chi_i(1, 2), \ldots, \chi_i(k, k'), \ldots, \chi_i(K\text{-}1, K)]$. The value of $\chi_i(k, k')$ is then defined in the Eq. (2).

$$\chi_i(k, k') = \begin{cases} 0; \text{different angle UEs} \\ 1; \text{the same angle UEs} \end{cases} \tag{2}$$

Naturally, different beam directions can mitigate interference. For any two UEs, from a view point of the ONU/antenna pair, the first category (i.e., $c_i(k, k') = 1$) is that two UEs locate at the same angle of beam direction. The second category (i.e., $c_i(k, k') = 0$) is that two UEs locate at different beam directions. Hence, the co-channel interference of second category UEs will be mitigated, even if these UEs employ the same RB of CCs on different radio frames which are carried by different optical subcarriers. However, for the first category UEs, interference still occurs if the UEs employ the same RB of CCs on different radio frames.

Therefore, $\gamma^{(n,y,t)}_{k,\ k'}$ is also defined as a binary variable. As shown in Eq. (3), $\gamma^{(n,y,t)}_{k,\ k'} = 1$ indicates that the same RB n of the y-th CC is allocated to UE k and k' in slot t at different frames. The same RB of CCs here means the RB on component carriers in the same frequency and also the same time slot carried by different frames:

$$\gamma^{(n,y,t)}_{k,k'} = \begin{cases} 1; & \text{UE } k \text{ and } k' \text{ are allocated with the same RB of CC} \\ 0; & \text{otherwise} \end{cases} \tag{3}$$

Definition 1: The UE set {1, 2,…, $\tilde{k}$,…, $\tilde{K}$} is located outside the cell ξ and served by ONU i, while UE set {1, 2,…, k,…, K} is in the cell ξ and served by ONU j.

Definition 2: For any two UEs, from the viewpoint of the MIMO antenna, we define that the same angle UEs in Eq. (2) are two UEs located at the same angle of beam direction. The angle space depends on the coverage of a beam released by antennas (e.g., 30° or the case of narrow beam in 5G). Otherwise, they are different angle UEs which locate at different beam directions (base station MIMO antenna arrays in the cell are in the same place and treated as one point).

3.2. Modeling of resource sharing proposal

The optimization model we proposed is more applicable for the deployment of small cell coverage with a high UE mobility scenario, so that the sharing capacity can be maximized and the delay time from the OLT to each UE could be minimized by the model. In the system, we suppose a remnant resource of bandwidth of each ONU after its inter-cell allocation can be delivered to the UEs in different cells for resource sharing by the broadcasting of distributed antennas. It is assumed that the antenna transmission for wireless signals in each cell could well reach the UEs in several adjacent cells. We also suppose that each ONU is attached by one antenna element in its location by default. In this chapter, for simplicity, we directly denote i or j as an ONU/antenna pair, that is, the ONU i means the ONU in i-th ONU/antenna pair, and the ONU j means the ONU in j-th ONU/antenna pair. Especially, in terms of UE k which located in cell ξ, we define ONU i for the ONU placed outside cell ξ and ONU j for the ONU placed in cell ξ. The UE set {1, 2,…, $\tilde{k}$,…, $\tilde{K}$} is located outside the cell ξ and served by ONU i, while UE set {1, 2,…, k,…, K} is in the cell ξ and served by ONU j. The classification of different UE sets and different ONU/antenna pairs is to clearly describe the optimization problem of subcarrier multi-cell sharing.

Consider the single UE k which is accommodated by ONU j, and UE k receives data from ONU j and a shared ONU i. For the data stream from ONU i to UE k, we define $d^{k}_{i,t}$ as its delay of sharing data for UE k in time slot t by ONU i from optical back-haul in OLT to the UE.

We consider the case that multiple ONUs share their data for a single UE *k*. As the system model depicted in **Figure 6**, we define a set $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ which represents the set of ONUs outside the cell where UE *k* is located for bandwidth sharing to UE *k*.

On the other hand, we define $\mathcal{P} = \{j \mid j = 1, 2, \ldots, b, \ldots, n\}$ representing the set of ONUs inside the cell where UE *k* belongs and $\mathcal{N} = \{i \mid i = 1, 2, \ldots, b, \ldots, n\}$ representing the set of total ONUs in a local network, respectively. Here, *m* is less than *n* and $\mathcal{M} \cup \mathcal{P} \subseteq \mathcal{N}$. For the parameter *b*, note that the delay time $d^k_{b,t}$ is the maximum delay in sharing links among all the links through the selected ONUs to the UE *k* satisfying:

$$b = \arg\max_{i \in \mathcal{M}} d^k_{i,t} \tag{4}$$

We could enrich our model to the real-time scenario for multiple UEs. The joint objective to (*i*) maximize sharing capacity with minimum delay and (*ii*) to minimize co-channel interference in a time duration *T* can be formulated by Eq. (5) in detail.

Objective:

$$\max\left\{\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\sum_{i=1,\, i\neq j}^{m} w^k_{i,t}\cdot c^k_{i,t} - \sum_{i=1,\, i\neq j}^{m} q^k_{i,t}\cdot d^k_{i,t}\right) - \sum_{i=1,\, i\neq j}^{m}\sum_{t=1}^{T}\sum_{y=1}^{N_{CC}}\sum_{n=1}^{N_{RB}}\left(\sum_{C_i(k,k')=1}\gamma^{(n,y,t)}_{k,k'} + \sum_{C_i\left(k,\tilde{k}\right)=1}\gamma^{(n,y,t)}_{k,\tilde{k}}\right)\right\} \tag{5}$$

where $w^k_{i,t}$ and $q^k_{i,t}$ can be further described in Eqs. (6) and (7), respectively:

$$w^k_{i,t} = \beta^k_{i,t}\cdot G^k_{i,t}\cdot \min\left\{h^k_{i,t-1}, h^k_{i,t-2}, \ldots, h^k_{i,1}\right\} \tag{6}$$

$$q^k_{i,t} = \beta^k_{i,t}\cdot D^k_{i,t}\cdot \max\left\{U^k_{i,t-1}, U^k_{i,t-2}, \ldots, U^k_{i,1}\right\} \tag{7}$$

Here, $\beta^k_{j,t}$ represents a binary indicator that UE *k* is served or not by ONU *i* in slot *t*. Different from $c^k_{i,t}$ which is a current capacity that could be provided to UE *k* in slot *t*, while $G^k_{i,t}$ is a current capacity which is obtained by UE *k* finally in slot *t*. Moreover, $h^k_{i,t\text{-}1}$ is a historical capacity obtained by UE *k* in slot *t*-1. It should be noted that $c^k_{i,t}$ is the shared capacity available for UE *k* from ONU *i*. Meanwhile, in Eq. (7), $D^k_{i,t}$ and $U^k_{i,t\text{-}1}$ are current delay constraint of UE *k* in slot *t* and historical delay record of UE *k* in slot *t*-1, respectively.

The optimization objective in Eq. (5) may be seemed indeed as an interference mitigation problem of finding $c^k_{i,t}$ subjected to the delay requirement from a set of $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ for UE *k* severed by ONU *j*. This will be solved in more details in our heuristic algorithms later. Firstly, we discuss all the constraints of objective as follows;

1) Capacity constraints for UE *k*:

$$\sum_{i\in\mathcal{M},\, i\neq j} c^k_{i,t} \geq A^k_t - \sum_{j\in\mathcal{P}} F^k_{j,t} \tag{8}$$

Equation (8) describes the total sharing capacity should not be less than the capacity requirement for each UE *k* in each time slot *t*. $A^k{}_t$ is total data capacity demand of UE *k* and $F^k{}_{j,t}$ is data capacity provided by ONU *j* to UE *k*.

2) Delay constraints for UE *k*:

$$d^{k}_{b,t} \le D^{k}_{i,t} \tag{9}$$

The delay constraint in Eq. (9) in each slot *t* means that the delay time spent on the path from the source of OLT to the destination of UE should not exceed the maximum tolerable transmission delay time (TDT) of UE *k* in a real-time service.

3) Capacity constraint for ONU *i*

Considering the 5G communication with carrier aggregation from [14, 16], we may further discuss the constraint of $c^{k}_{i,t}$:

$$\sum_{k=1}^{K} c^{k}_{i,t} \le E_{i,t} \tag{10}$$

Denote $E_{i,t}$ as the total remaining capacity of ONU *i* after the allocation for its UEs accommodated. Equation (10) describes that the total amount of sharing capacity of UEs should be less than $E_{i,t}$.

With respect to our resource allocation model for optical time/frequency blocks with RoF and downlink signal processing in Section 2, we formulate $E_{i,t}$ approximately with the aforementioned assumptions which are detailed in the aspect on resource allocation.

The remaining capacity $E_{i,t}$ of ONU *i* in time slot *t* is then given as Eq. (11) approximately:

$$E_{i,t} = \sum_{l=1}^{N^{i}_{\text{sub-c}}} C_l \left(\frac{1}{Q} N_{\text{cc}} N_{\text{RB}} \sum_{o=1}^{Q} r^{(o)}_{\text{RB}} \right) - \sum_{l=1}^{N^{i}_{\text{sub-c}}} \sum_{p=1}^{C_l} \sum_{\tilde{k}=1}^{\tilde{K}} \sum_{y=1}^{N_{\text{cc}}} \sum_{n=1}^{N_{\text{RB}}} \sigma^{(n,y,t)}_{\tilde{k},l,p} \sum_{o=1}^{Q} \mu_{\tilde{k},o} \cdot r^{(o)}_{\text{RB}} \tag{11}$$

Especially, $\sigma^{(n,y,t)}{}_{k,l,p}$ is a binary variable to define whether or not the *n*-th RB of the *y*-th CC is assigned to the *k*-th UE on the *p*-th frame on the *l*-th optical sub-carrier in slot *t*, and $\sigma^{(n,y,t)}{}_{k,l,p} = 1$ expresses that allocating the *n*-th RB of the *y*-th CC to the *k*-th UE in slot *t*. Here, we define a binary variable $\mu_{k,o} = 1$ to express that the *k*-th UE employs the *o*-th MCS. *Q* in Eq. (13) describes the highest MCS employed by the *k*-th UE corresponding to a CQI of RB "$max(g_{k,l,p,\delta^*})$" in each CC of the *p*-th frame on the *l*-th optical subcarrier. Here:

$$\delta^{*} = \underset{n \in \{1,2,\ldots,N_{\text{RB}}\}}{\arg\max} \left(g_{k,l,p,n} \right) \tag{12}$$

$$Q_{k,l,p,\max\left(g_{k,p,\delta^*}\right)} = \underset{j \in \{1,2,\ldots,J\}}{\operatorname{argmax}} \left(R^{(c)}_{j} \log_2\left(M_j\right) \middle| g_{k,l,p,\delta^*} \right) \tag{13}$$

In the Eq. (11), $N^{i}_{\text{sub-c}}$, N_{cc}, and N_{RB} are the number of optical subcarriers allocated to ONU *i* by OFDM-PON, the total number of wireless component carriers (CC) [15, 16] modulated onto a single optical subcarrier, and total number of wireless resource blocks (RB) carried by a single wireless component carrier (CC), respectively. The second term of polynomial in Eq. (11)

should be larger than the summation of total UE minimum capacity requirement. Hence, a constraint of $c^{k}_{i,t}$ can be further formulated as in Eq. (14),

$$\sum_{k=1}^{K} c^{k}_{i,t} \le E_{i,t} \le \sum_{l=1}^{N^{i}_{\text{sub-c}}} C_l \left(\frac{1}{Q} N_{\text{cc}} N_{\text{RB}} \sum_{o=1}^{Q} r^{(o)}_{\text{RB}} \right) - \sum_{\tilde{k}=1}^{\tilde{K}} R_{\tilde{k}} \tag{14}$$

In the following allocation algorithm, we will satisfy all the aforementioned constraints to find the sharing capacity $c^{k}_{i,t}$ for UE k from ONU i.

4) Interference constraint for UEs

In terms of the i-th ONU/antenna pair, the constraint of $\gamma^{(n,y,t)}_{k,\tilde{k}}$ is described in Eq. (15). It means that the number of same RB of CC allocated to different UEs on all the frames carried by optical subcarriers has an upper limitation. Since $\gamma^{(n,y,t)}_{k,\tilde{k}}$ is a binary indicator, Eq. (15) could be treated as two cases. First, when $\gamma^{(n,y,t)}_{k,\tilde{k}} = 1$, the same RB n of the y-th CC is allocated to UE k and $\tilde{k}$ in slot t at different frames, the product of all the numbers of these RBs allocated to UE k and all the number of same RBs allocated to UE $\tilde{k}$ must not be larger than their arithmetic mean square, while the upper limitation of their arithmetic mean equals half of the number of total frames. Second, when $\gamma^{(n,y,t)}_{k,\tilde{k}} = 0$, any RB n of the y-th CC is not allocated to both UE k and $\tilde{k}$ in slot t at different frames; therefore, for all the number of RBs allocated to UE k and $\tilde{k}$, their product must equal to 0 (i.e., without RB overlapping on the same time/frequency domain) as described in Eq. (15):

$$\sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} \cdot \sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{\tilde{k},l,p} \le \gamma^{(n,y,t)}_{k,\tilde{k}} \left[\frac{1}{2} \left(\sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} + \sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{\tilde{k},l,p} \right) \right]^2 \le \gamma^{(n,y,t)}_{k,\tilde{k}} \left[\frac{1}{2} \left(\sum_{l=1}^{N^{i}_{\text{Sub-c}}} C_l \right) \right]^2, \tag{15}$$

$\forall k, \tilde{k}, \forall y, \forall n, \forall t$

Similarly, we hereby obtain the following constraint of $\gamma^{(n,y,t)}_{k,k'}$ as described in Eq. (16):

$$\sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} \cdot \sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k',l,p} \le \gamma^{(n,y,t)}_{k,k'} \left[\frac{1}{2} \left(\sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} + \sum_{l=1}^{N^{i}_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k',l,p} \right) \right]^2 \le \gamma^{(n,y,t)}_{k,k'} \left[\frac{1}{2} \left(\sum_{l=1}^{N^{i}_{\text{Sub-c}}} C_l \right) \right]^2, \tag{16}$$

$\forall k, k', \forall y, \forall n, \forall t$

4. Proposed resource sharing algorithm

In this section, we propose a heuristic algorithm for obtaining sub-optimal solutions because solving the objective in Section 3 is highly complex. A natural and simple approach to address the joint objectives of Eq. (5) is to treat it as a maximum flow and minimum cost problem about resource allocation (e.g., RB of CC) by assigning $\sigma^{(n,y,t)}_{k,l,p}$ we defined for UE in each slot, approximately. We record and observe some historical information (e.g., $h^{k}_{j,t\text{-}1}$ and $U^{k}_{j,t\text{-}1}$) as timely references and evaluations for finding a maximum flow ($c^{k}_{i,t}$) subjected to the delay requirement ($d^{k}_{i,t}$) with minimum delay time from a set of $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ for UE

k severed by ONU j. In the following sections, the sharing path assignment and the corresponding resource allocation will be detailed in algorithm descriptions by transferring the problem into the optimization with the aid of graph theory.

4.1. Problem statement

- *Given parameters*:
  - $\boldsymbol{G}(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is UE k and set of all ONUs and $\mathcal{E}$ is the set of resource sharing path through multiple ONUs to UE k
  - Matrix $C_{k,t} = [c^k_{1,t},\ldots, c^k_{i,t},\ldots, c^k_{N,t}]$, $\forall\, k$ in $\mathcal{K}$, $c^k_{i,t} > 0$
  - Matrix $\mathcal{D}_{k,t} = [d^k_{1,t},\ldots, d^k_{i,t},\ldots, d^k_{N,t}]$, $\forall\, i$ in $\mathcal{N}$, $\forall k$ in $\mathcal{K}$
  - Matrix $\mathcal{X}_i = [\chi_i(1,2),\ldots, \chi_i(k,k'),\ldots, \chi_i(K\text{-}1,K)]$
  - Matrix $\mathcal{Y}_i = [Y_i(k,1),\ldots, Y_i(k,\tilde{k}),\ldots, Y_i(k,\tilde{K})]$
  - Set of UE in the cell: $\mathcal{K} = \{k \mid k = 1, 2,\ldots, K\}$
  - Set of UE outside the cell: $\tilde{\mathcal{K}} = \{\tilde{k} \mid \tilde{k} = 1, 2,\ldots, \tilde{K}\}$
  - Set of ONU: $\mathcal{N} = \{i \mid i = 1, 2,\ldots, b,\ldots, N\}$
  - $R_{k,t}$: Minimum data capacity requirement for UE k
  - $B_{k,t}$: Allocated data capacity to UE k

Note that $\mathcal{X}_i$ and $\mathcal{Y}_i$ are two matrices which store UE conflict relationships.

- *Objective*:
  - Minimize the co-channel interference which is generated by sharing data received for the same angle UEs in their located cell and also in their adjacent cells.
  - Maximize the sharing capacity in terms of UEs.
  - Minimize the average delay of sharing data transmission by ONUs to satisfy UE requirements.

4.2. Algorithm description

The sharing algorithm tries to search the idle RBs over each optical subcarrier delivering to each cell and shares them to the UEs in the adjacent cells. From the sharing paths with minimum delay time, the algorithm selects the paths with maximum number of idle RBs for each UE so that it could maximize the sharing capacity for each UE. Consequently, the algorithm as a solution of optimization problem for our resource allocation model is suggested to be executed on the OLT side of optical back-haul. Considering the single UE k ($k = 1, 2,\ldots, K$) which is accommodated by multiple sharing paths from different ONUs, UE k can receive the

data from each sharing ONU i ($i = 1, 2,\ldots, N$). The algorithm is divided into several steps as shown in **Tables 1**, **2**, and **3**.

In the first step, a dynamic sharing graph is generated, for instance, **Figure 7(a)** illustrates an example of a sharing tree that five ONUs share bandwidth resources to UE k. The rooted vertex represents UE k, and any leaf vertex i represents ONU i, respectively. Each edge represents a sharing path. In the second step, the weights of vertex and edge are assigned. For the data of UE k from ONU i, $d^k_{i,t}$ denotes the overall delay on sharing path through ONU i to UE k in time slot t. $c^k_{i,t}$ denotes the available sharing data capacity for UE k from ONU i in time slot t.

ALGORITHM 1 Real-time Sharing Algorithm (RTSA)
<Note>:**Algorithm 1** contains **FUNCTION 1, 2** and **3**

Input: Matrix $C_{k,t} = \{c^k_{i,t} \mid c^k_{1,t},\ldots, c^k_{i,t},\ldots, c^k_{N,t}, \forall\, c^k_{i,t} > 0\}$
Matrix $\mathcal{D}_{k,t} = \{d^k_{1,t},\ldots, d^k_{i,t},\ldots, d^k_{N,t}\}$
Set of UE: $\mathcal{K} = \{k \mid k = 1, 2,\ldots, K\}$
Set of ONU: $\mathcal{N} = \{i \mid i = 1, 2,\ldots, b,\ldots, N\}$
Initialization: $\boldsymbol{G_k} = \emptyset$ for all $k = 1, 2,\ldots, K$; $\boldsymbol{G_k}$: total RB set to UE k
While ($B_{k,t} < R_{k,t}$) do
STEP 1: Make a sharing graph $\boldsymbol{G}$ ($\mathcal{V}$, $\mathcal{E}$); $\mathcal{V} = \{1,\ldots, N\}\cup\{k\}$, $\mathcal{E} = \{(i, k) \mid d^k_{i,t} \le D_k\}$; D_k is a maximum tolerable delay(MTD).
STEP 2: Assign weights to $\mathcal{E}$ by matrix $\mathcal{D}_{k,t}$ and sort the edges in $\mathcal{E}$ according to its weight in graph $\boldsymbol{G}$ ($\mathcal{V}$, $\mathcal{E}$) in an ascending order.
STEP 3: Find an edge (i, k) with minimum $d^k_{i,t}$ in the ascending order of $\mathcal{E}$ as a current link (i, k) for resource sharing.
STEP 4: $B_{k,t} \leftarrow c^k_{i,t}$ by allocating $\sigma^{(n,y,t)}_{k,l,p}$ until $B_{k,t} \ge R_{k,t}$, otherwise, go to **STEP 3**
STEP 5: Traverse all UEs in set $\mathcal{K}$ to allocate resource satisfying, $i = \text{argmin}\, d^k_{i,t}$, $B_{k,t} \leftarrow c^k_{i,t}$, find $c^k_{i,t}$ employing
FUNCTION 1: form $\boldsymbol{G_k}$
End while
STEP 6: Traverse all ONUs in set $\mathcal{N}$ to allocate resources, repeat **STEPS 1–5**.
STEP 7: Record historical information (e.g., $h^k_{j,t}$ and $U^k_{j,t}$) by **FUNCTION 2**
STEP 8: Update matrix $\boldsymbol{C_{k,t}}$, and matrix $\boldsymbol{\mathcal{D}_{k,t}}$.
End
Out put: $\boldsymbol{G_k} = \{\boldsymbol{C_k}(1), \boldsymbol{C_k}(2), \boldsymbol{C_k}(i),\ldots\ldots, \boldsymbol{C_k}(N)\}$
End
FUNCTION 1: form $\boldsymbol{G_k}$, for any k; $\boldsymbol{C_k}(i)$: a RB set from ONU i to UE k
Initialize $\boldsymbol{G_k} = \{\boldsymbol{C_k}(1), \boldsymbol{C_k}(2), \boldsymbol{C_k}(i),\ldots\ldots, \boldsymbol{C_k}(N)\} = \emptyset$, set $B_{k,t} \leftarrow 0$.
1: Find ONU i for current link(i, k), where $i = \text{argmin}\, d^k_{i,t}$, set $p \leftarrow 1$.
2: Set $l \leftarrow 1$; l is layer (radio frame) indicator.
3: Find the idle RB for any UE by **FUNCTION 3**, where $\sigma^{(n,y,t)}_{k,l,p} = 0, \forall k$.
4: Allocate a corresponding idle RB to UE k, put RB of $\sigma^{(n,y,t)}_{k,l,p} = 0$ into set $\boldsymbol{C_k}(i)$, then for the UE k, $\sigma^{(n,y,t)}_{k,l,p} \leftarrow 1$. Increase capacity $B_{k,t}$, where, $B_{k,t} = B_{k,t} + r^{(n,y,t)}_{\text{RB}k,l,p}$.
5: If $B_{k,t} \ge R_{k,t}$, **break**; $r^{(n,y,t)}_{\text{RB}\,k,l,p}$ is the capacity of RB.
Else if $\forall k, \forall \sigma^{(n,y,t)}_{k,l,p} = 1$, i.e., the set of RBs on current layer have been occupied and cannot be scheduled to UE k for sharing, and **if** $l \le L$ add l, go to **STEP 3**.
Else if $p \le P_{\text{max-sub}}$, add p, go to **STEP 2**.
Else go to **STEP 1**.
End IF
6: Output $\boldsymbol{G_k} = \{\boldsymbol{C_k}(1), \boldsymbol{C_k}(2), \boldsymbol{C_k}(i),\ldots\ldots, \boldsymbol{C_k}(N)\}$

Table 1. Real-time sharing algorithm.

FUNCTION 2 Historical State Recording

1: Set $h^{k}_{i,t} = \sum_{p=1}^{N^{i}_{\text{sub-c}}} \sum_{y=1}^{N_{\text{cc}}} \sum_{n=1}^{N_{\text{RB}}} \sigma^{(n,y,t)}_{k,p} \sum_{o=1}^{Q} \mu_{k,o} \cdot r^{(o)}_{\text{RB}}$
2: Set $U^{k}_{i,t} = d^{k}_{i,t}$
3: For t from 1 to T, $\forall k,i$
4: If $h^{k}_{i,t} < h^{(\min)}_{k,t}$, then $h^{(\min)}_{k,t} = h^{k}_{i,t}$
5: If $U^{k}_{i,t} > U^{(\max)}_{k,t}$, then $U^{(\max)}_{k,t} = U^{k}_{i,t}$
End

Table 2. Historical state recording function.

FUNCTION 3 Interference Mitigation
<Note>:This function addresses RB allocation according to UE conflict relationships

1: If $\chi_i(k, k') = 0$, $\forall k, k'$ and $Y_i(k, \tilde{k}) = 0$, $\forall k, \tilde{k}$
Find any RB where $\sigma^{(n,y,t)}_{k,l,p} = 0$, $\sigma^{(n,y,t)}_{k',l,p} = 0$, and $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 0$ to allocate, for UE k
2: Else if $Y_i(k, \tilde{k}) = 1$, find RB where $\sigma^{(n,y,t)}_{k',l,p} = 0$ and $\sigma^{(n,y,t)}_{k,l,p} = 0$ to allocate,
but refrain RBs where $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 1, \forall l,p$
3: **Else if** $\chi_i(k, k') = 1$, find RB where $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 0$ and $\sigma^{(n,y,t)}_{k,l,p} = 0$ to allocate,
but refrain RBs where $\sigma^{(n,y,t)}_{k',l,p} = 1, \forall l,p$
End

Table 3. Interference mitigation function.

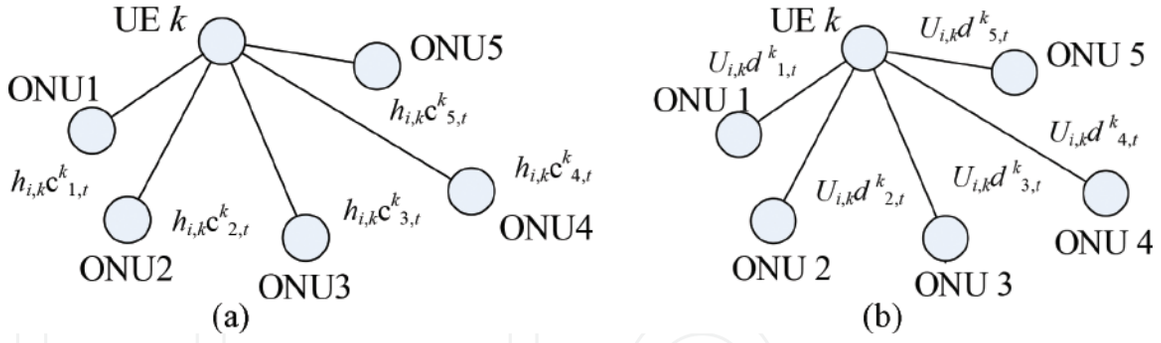

Figure 7. Wireless resource sharing logically on optical time/frequency blocks by different ONUs.

For each time slot t, the weight of each leaf is $h_{i,k}c^{k}_{i,t}$. The weight of edge between the rooted vertex and any leaf vertex i is $U_{i,k}d^{k}_{i,t}$, as shown in **Figure 7(b)**. Specifically, for $t = 1, 2,\ldots, T$, we define $h_{i,k} = min\{h^{i}_{k,t}, h^{k}_{j,t-1},\ldots, h^{k}_{j,1}\}$ and $U_{i,k} = min\{U^{i}_{k,t}, U^{k}_{j,t-1},\ldots, U^{k}_{j,1}\}$, respectively.

In the third step, the sharing path selection is performed by finding minimum delay time on each ONU. A sharing ONU combination could be found for UE k aiming to a minimum delay under its data rate demand.

In the fourth step, we allocate RB of CC to UEs in each ONU. Resource sharing for each UE (e.g., RB of CC) is executed by assigning $\sigma^{(n,y,t)}_{k,l,p}$. Here, $\sigma^{(n,y,t)}_{k,l,p}$ is a binary variable to define whether or not the n-th RB of the y-th CC on the l-th radio frame on p-th optical subcarrier is assigned to the k-th UE in slot t for finding a proper $c^{k}_{i,t}$ subjected to the minimum delay $d^{k}_{i,t}$ from a set of ONUs.

In the fifth step, the algorithm loops to another UE allocating RBs to it until all the UEs of current set have been finished. In the sixth step, the algorithm traverses all served ONUs to finish the RB allocation. In the seventh step, we compute the capacity obtained by UE k from ONU i in slot t and its delay time. Then we compare them to the previous historical values and update the historical peak value in the case that the current one exceeds it. After executing these steps, the algorithm outputs the RB set allocated to UE k classifying them into each subset of RBs obtained by each sharing ONU individually.

To meet all constrains of mathematical descriptions in Section 3, we formulate three different functions for the heuristic algorithm herein as practical approaches to achieve the target of optimization. Function 1 is one of solutions to search idle RBs for resource sharing satisfying minimum optical wavelength cost. Function 2 addresses historical information recording and their updating. Function 3 finds idle RBs and allocates them according to different UE conflict relationships in order to mitigate co-channel interference, which will be intensively evaluated and discussed in the next section of this chapter.

5. Simulations and numerical results

5.1. Simulation parameters

In this section, we provide a deep observation for the proposed resource sharing approach on the performance of wireless UEs in the OFDM-PON system. The simulation and analytic evaluation by large-scale C++ programming mainly focus on the interference mitigation of mobile UEs in the cell under different mobility and times.

In intensive large-scale C++ simulations, a RoF-OFDM-PON covering up to 256 cells assuming random UE mobility is deployed to evaluate our proposal as shown in **Figures 8** and **9**. Optical subcarriers with per λ_i 10-Gb/s digital-equivalent data rate are adopted. LTE-like wireless resources carried on optical subcarriers are assigned to UEs corresponding to the scheduling solution in the well-known network simulator 3 (ns-3) [17], supporting maximum five-carrier aggregation, simultaneously. MCS is assigned to UEs corresponding to Eqs. (12) and (13) by the scheduling in the ns-3. The main simulation parameters are described in **Table 4**.

In the C++ simulations, according to LTE-EPC model [18] in ns-3 simulator and with respect to its resource allocation models, we modify the scheduler significantly based on our proposed real-time sharing algorithm (RTSA). We evaluate the throughput performance of UEs by comparing RTSA with maximum throughput (MT) and proportional fair (PF) schemes [6, 19]. Note that for a fair comparison, we also modify the scheduler to serve multiple wavelength scheduling (i.e., multiple radio frames carried on one optical wavelength) since MT and PF themselves have no such functionality.

From the entire network perspective, the total UE number is set to 36,000 in simulations under different mobility ratios (a = number of mobile UEs/number of total UEs). We set a position for each UE with position allocator by model library of NS3 [17] (e.g., random waypoint). Meanwhile, we aim to simulate a difference specifically on UE mobility, for example, changing the residential position of UEs (e.g., migrate and recall UEs) within the scope of all cells regularly

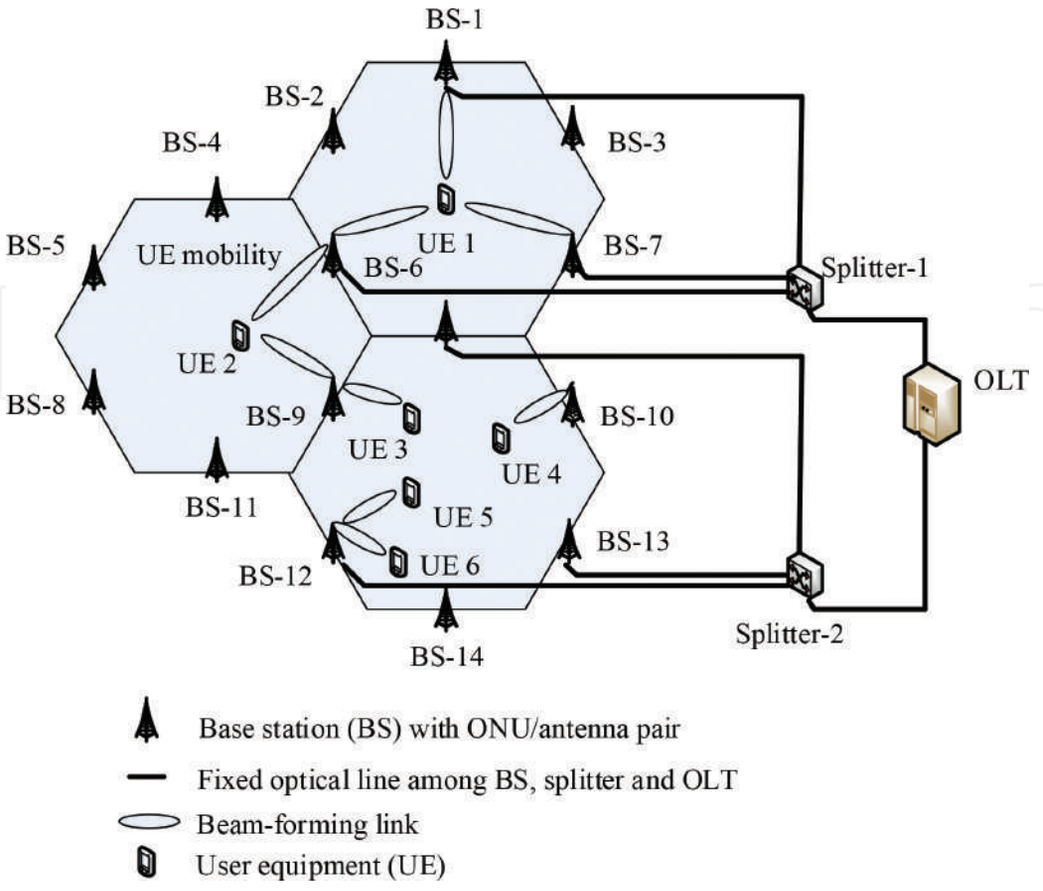

Figure 8. Scenarios of UEs and distributed antenna allocation in simulations with beamforming in the cluster of wireless cells.

in different times as shown in **Figure 9(a)** and **(b)** respectively. For instance, we define different mobility ratios of UEs equaling to 0.2 and 0.8, respectively.

5.2. Results and analysis on interference mitigation

Next, we observe the effect on interference mitigation under different UE mobility rates within different time slots by generating massive number of same angle UEs (the conflict UEs) in each beam direction. As an example, in **Figure 10** we typically illustrate four windows of UE distribution states under random mobility in four different time slots, respectively. With the irregular movement of UEs, new UE conflict relationships will be generated randomly in terms of different antennas. For instance, in the time slot 1, UE k_4 and k_5 are located in different directions in terms of antenna 1. Simultaneously, UE k_1 and k_7 which have a conflict relationship with each other are located in the same direction for antenna 1. However, in the time slot 2, a new UE conflict relationship is generated between UE k_4 and k_5, while UE k_1 and k_7 are located in different directions, and their conflict relationship disappears. Similarly, in a continuous time scope (e.g., 10 minutes) containing many more time slots, we then observe our proposed scheme in the aspect of interference cancellation. We therefore evaluate the block error rate (BLER) of RTSA, PF, and MT under Gauss interference model in ns-3 model library so as to compare our proposed scheme with conventional schemes in terms of their effectiveness on interference mitigation. In simulations, BLER is observed under fair channel condition, for instance, the same level of signal power which is set by signal-to-noise ratio (SNR), and the

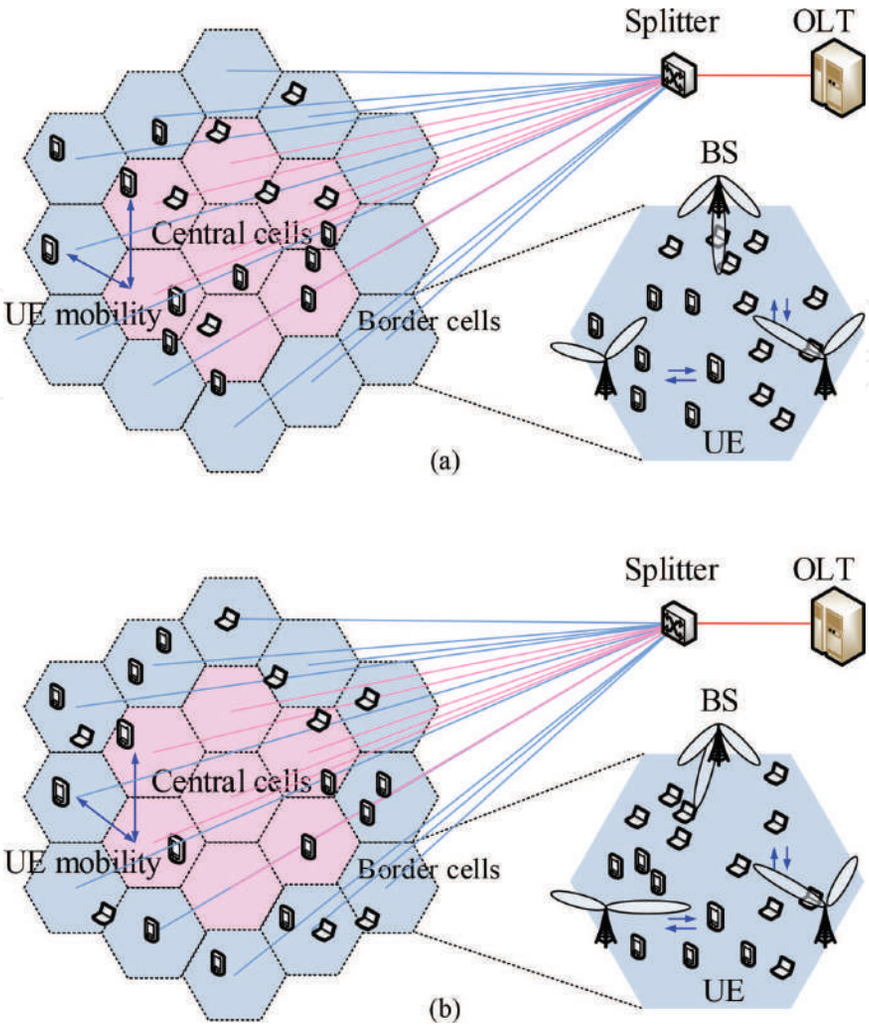

Figure 9. Scenarios of UE long-distance migration with position distribution for each UE by random model library (e.g., random waypoint) in simulations. (a) The aggregation of UEs at central cells. (b) The spreading of UEs to border cells.

Parameter	Value	Parameter	Value
LTE subcarrier	15 kHz	Frame duration	10 ms
Resource block	180 kHz	TTI	1 ms
RB carriers ($N_{sub\text{-}c}$)	12	UE date rate$_{min}$ (R_k)	200 Mbps
RB OFDM symbols	7	MCS (J)	29
UE received CC$_{max}$ (z)	5	Bandwidth of CC	20 MHz
Single CC length (m)	100 RBs	Testing MIMO per cell	4 × 4
BS TX power	30 dBm	Number of cell	256
Noise spectral density	−174 dBm/Hz	Cell radius	500 m
Path loss (distance R), in dB	128.1 + 37.6lg$R$	SMF fiber distance	20 km

Table 4. Simulation parameters.

SNR is obtained according to the parameters in **Table 4**. In addition, we set random UE location and irregular mobility with the increase of time.

The change of BLER is observed under low mobility (a = 0.2) and high mobility (a = 0.8) cases, respectively. As shown in **Figure 11**, RTSA has the lowest level of BLER than MT and PF, which

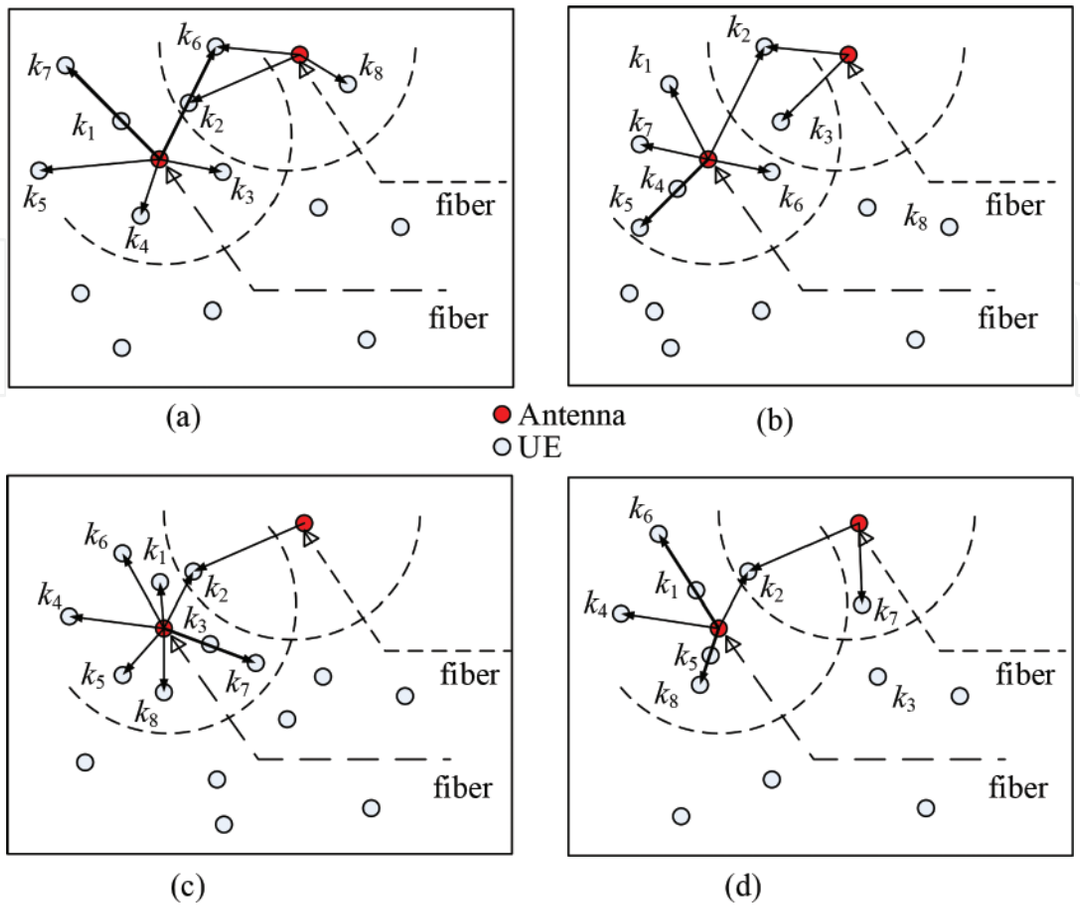

Figure 10. A group of observations about interference with UE/antenna distribution under different time slots and UE migration when employing the proposed scheme. In (a)–(d), the dashed line represents the propagation scope of each antenna. The solid arrow line represents a link with one beam direction (the thick arrow line represents a link which has conflict UEs with the potential interference).

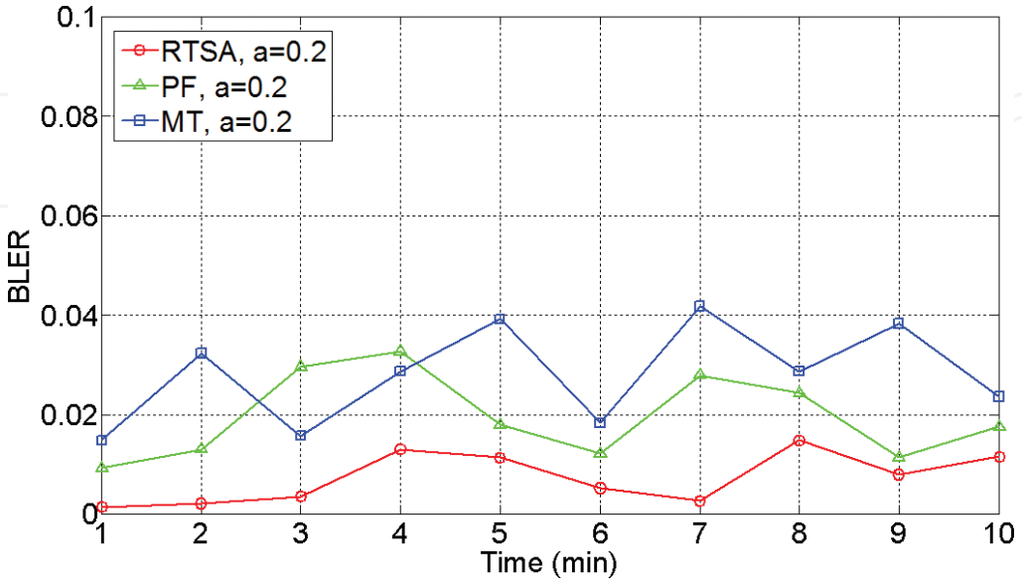

Figure 11. Comparisons of BLER under different time durations (mobility ratio 0.2).

http://dx.doi.org/10.5772/intechopen.72727

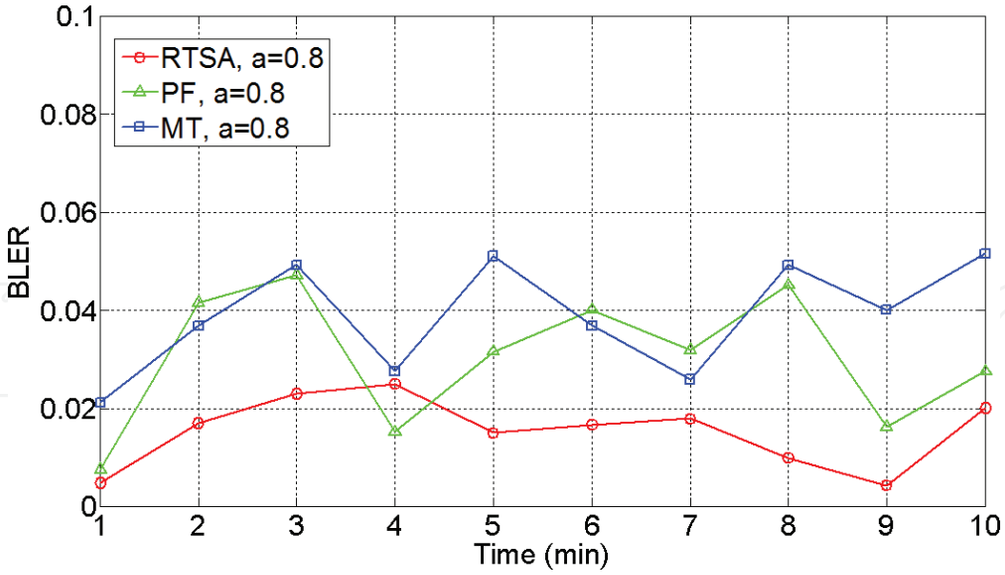

Figure 12. Comparisons of BLER under different time durations (mobility ratio 0.8).

also can be found in **Figure 12**. This feature reflects that the proposed RTSA has a benefit on BLER improvement. Compared with the results in **Figures 11** and **12**, we could further observe that the level of BLER for a = 0.8 is larger than that in a = 0.2 and the fluctuation of BLER in different time is also obvious in terms of the case where a = 0.2. The reason for a higher level of BLER is that a higher mobility brings much more interference with the conflict UEs generated in each beam direction. Nevertheless, our proposed RTSA will still has the lowest BLER level even in the case of higher UE mobility.

6. Conclusions

This chapter investigates the resource sharing problems for future 5G cellular networks [20], which jointly employ distributed massive MIMO, beamforming, and OFDMA-based passive optical network supporting radio-over-fiber (RoF). We have surveyed the system and its physical transmission features to explore reasonable solutions. With the assumptions based on physical features of system given in this chapter, we describe the latest hot problem with mathematical optimization for minimizing co-channel interference, etc. Since it is highly complex to get optimal results, then we heuristically formulate a real-time sharing algorithm as a practical solution. Simulation results also reveal that the proposed scheme is the most efficient one at the interference mitigation compared to conventional schemes.

Author details

Sheng Xu

Address all correspondence to: xusheng@akane.waseda.jp

Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan

References

[1] Bi S, Zhang R, Ding Z, Cui S. Wireless communications in the era of big data. *IEEE Communications Magazine*. 2015;**53**(10):190-199

[2] Lin YM, Tien PL. Next-generation OFDMA-based passive optical network architecture supporting radio-over-fiber. *IEEE Journal on Selected Areas in Communications*. 2010;**28**(6): 791-799

[3] Tien PL, Lin YM, Yuang MC. A novel OFDMA-PON architecture toward seamless broadband and wireless integration. In: *Conference on Optical Fiber Communication*; San Diego, CA, USA; 22-26 March 2009; IEEE; pp. 1, 3, 22-26

[4] Zhang J, Wang T, Ansari N. An efficient MAC protocol for asynchronous ONUs in OFDMA PONs. In: *Optical Fiber Communication Conference and Exposition (OFC/NFOEC 2011) and The National Fiber Optic Engineers Conference*; Los Angeles, CA, USA; 6-10 March 2011; IEEE; pp. 1-3

[5] Wei W, Hu J, Qian D, Ji PN, Wang T, Liu X, Qiao C. PONIARD: A programmable optical networking infrastructure for advanced research & development of future internet. *Journal of Lightwave Technology*. 18 February 2009;**27**(3):233-242

[6] Wei W, Wang T, Qian D, Hu J. MAC protocols for optical orthogonal frequency division multiple access (OFDMA)-based passive optical networks. In: *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC 2008)*; San Diego, CA, USA; 24-28 February 2008; IEEE; pp. 1-3

[7] You W, Yi L, Huang S, Chen J, Hu W. Power efficient dynamic bandwidth allocation algorithm in OFDMA-PONs. *IEEE/OSA Journal of Optical Communications and Networking*. 2013;**5**(12):1353-1360

[8] Liu X, Effenberger F, Chand N, Zhou L, Lin H. Efficient mobile fronthaul transmission of multiple LTE-A signals with 36.86-Gb/s CPRI-equivalent data rate using a directly-modulated laser and fiber dispersion mitigation. In: *Asia Communications and Photonics Conference*; Shanghai, China; 11-14 November 2014; Optical Society of America (OSA); paper AF4B.5

[9] Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong ACK, Zhang JC. What will 5G be? *IEEE Journal on Selected Areas in Communications*. 2014;**32**(6):1065-1085

[10] Almasoudi F, Alatawi K, Matin MA. Study of OFDM technique on RoF passive optical network. *Optics and Photonics Journal*. 2013;**3**:217-224

[11] Chow CW, Yeh CH, Wang CH, Wu CL, Chi S, Lin C. Studies of OFDM signal for broadband optical access networks. *IEEE Journal on Selected Areas in Communications*. 2010;**28**(6):800-807

[12] Kanonakis K, Tomkos I, Pfeiffer T, Prat J, Kourtessis P. ACCORDANCE: A novel OFDMA-PON paradigm for ultrahigh capacity converged wireline-wireless access networks. In: *12th International Conference on Transparent Optical Networks (ICTON2010)*; Munich, Germany; 27 June-1 July 2010; IEEE; pp. 1-4

[13] Qian D, Cvijetic N, Hu J, Wang T. 108 Gb/s OFDMAPON with polarization multiplexing and direct detection. *Journal of Lightwave Technology*. 2010;**28**(4):484-493

[14] Liao HS, Chen PY, Chen WT. An efficient downlink radio resource allocation with carrier aggregation in LTE-Advanced networks. *IEEE Transactions on Mobile Computing*. 2014;**13**(10): 2229-2239

[15] Cox C. An Introduction to LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications. West Sussex: Wiley; 2014

[16] Akyildiz IF, Gutierrez-Estevez DM, Reyes EC. The evolution to 4G cellular systems: LTE-advanced. *Physical Communication*. 2010;**3**:217-224

[17] NS-3 consortium. "NS-3 network simulator," http://www.nsnam.org/, INRIA and the University of Washington [Accessed September 2017]

[18] CTTC. "LTE-EPC network simulator (LENA)," CTTC, http://iptechwiki.cttc.es [Accessed October 2017]

[19] Guan N, Zhou Y, Tian L, Sun G, Shi J. Qos guaranteed resource block allocation algorithm for LTE systems. In: *IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2011)*; Wuhan, China; 10-12 October 2011; IEEE; pp. 307-312

[20] Xu S, Xu SG, Tanaka Y. Sub-carrier sharing in OFDM-PON for 5G mobile networks supporting radio-over-fibre. In: *21st OptoElectronics and Communications Conference (OECC 2016) held jointly with International Conference on Photonics in Switching (PS)*; Niigata, Japan; 3-7 July 2016; IEEE; pp. 1-3

IntechOpen

Chapter 7

Transport Protocol Performance and Impact on QoS while on the Move in Current and Future Low Latency Deployments

Eneko Atxutegi, Jose Oscar Fajardo and Fidel Liberal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71779

Abstract

Transport protocols and mobile networks have evolved independently leading to a lack of adaptability and quality of service (QoS) degradation while running under the variability circumstances present in cellular access. This chapter evaluates the performance of state-of-the-art transmission control protocol (TCP) implementations in challenging mobility scenarios under 4G latencies and low delays that model the proximity service provisioning of forthcoming 5G networks. The evaluation is focused on selecting the most appropriate TCP flavor for each scenario taking into account two metrics: (1) the goodput-based performance and (2) a balanced performance metric that includes parameters based on goodput, delay and retransmitted packets. The results show that mobility scenarios under 4G latencies require more aggressive TCP solutions in order to overcome the high variability in comparison with low latency conditions. Bottleneck Bandwidth and Round-Trip Time-RTT (BBR) provides better scalability than others and Illinois is more capable of sustaining the goodput with big variability between consecutive samples. Besides, CUBIC performs better in lower available capacity scenarios and regarding the balanced metric. In reduced end-to-end latencies, the most suitable congestion control algorithms (CCAs) to maximize the goodput are NewReno (low available capacity) and CUBIC (high available capacity) when moving with continuous capacity increases. Additionally, BBR shows a balanced and controlled behavior in most of the scenarios.

Keywords: transport protocols, performance, mobility, 4G, low latency, 5G

1. Introduction

Mobile broadband (MBB) usage has risen significantly in the last years and so has done the customers' awareness regarding the quality of service (QoS). Thus, the measurement of QoS

in MBB networks has become a key issue. In this regard, the proper performance of the transport layer constitutes a critical feature in order to fulfill the QoS requirements of the clients (user equipments—UE) [1, 2]. In MBB, due to the multiple sources of variability (related to the client or self-inflicted effects such as channel quality reporting, propagation and fading pattern alterations due to mobility—and related to the intrinsic features of MBB such as bandwidth sharing, modulation and so forth), the network conditions become more volatile than in fixed networks and therefore the accuracy of transport protocols to adapt their sending rate as close to the available capacity as possible is reduced, impacting on the final performance [1, 2]. Therefore, there is an urgent need to study the relation between the transport protocol performance in MBB and its impact on the actual QoS results.

We consider that the mobility is one of the biggest features that differentiates cellular access from other connectivity schemes. Besides, the wide range of moving possibilities makes every use-case distinct and independent, creating different network conditions for transport protocols. In this regard, it is important to focus the performance-based analysis of current state-of-the-art transport protocol solutions over different mobile network mobility circumstances and understand the implications of the movement in the interaction between the transport protocol and the MBB variability.

Regarding mobile networks, this work analyses different schemes in order to give a wider view of the impact that the performance of transport protocols have in the QoS: 4G scenarios and low latency scenarios targeting assisted by mobile edge computing (MEC-assisted) future 5G deployments [3]. Future 5G networks aim at allowing improved capabilities in terms of achievable capacities, modulation and end-to-end latency among other features. The reduction in the transmission latency is one of the main beneficial evolutions for the suitable performance of transport protocols. It shortens the feedback time between consecutive management decisions in the server, increasing the responsiveness of the transport layer to the fluctuations of the radio side. Therefore, we focus our second scheme in the proximity service of 5G deployments. To that end, we mimic with low latency 4G scenarios a 5G-alike service provisioning. All in all, our evaluation covers the performance of transport protocols in current and future MBB.

Considering that TCP is the predominant transport protocol on the Internet, we focus our study in the evaluation of TCP over distinct mobility circumstances over 4G latencies and low latency deployments. TCP is not a single entity but a family of different congestion control algorithms (CCAs) that manage the outstanding data of the server (clamped by the congestion window—CWND) in a different way based on pre-defined features such as throughput maximize algorithms with loss-based mechanisms, delay-aware implementations or hybrid developments. So far, despite many CCAs being available, none of them have demonstrated to both be easily deployable and appropriately face the variability of MBB fluctuation. This work selects and evaluates the performance of distinct CCAs that count on different features and implementations that in the end result in a different performance outcome in each precise network conditions.

The great success of TCP and user datagram protocol (UDP) have led to the widespread utilization of both of them, either as the selected transport solution or as a substrate to enable

http://dx.doi.org/10.5772/intechopen.71779

the so-called transport services [4–6]. These transport services are ad-hoc layers that work between the transport layer and the application layer, taking advantage of the substrate transport protocol (mainly TCP and UDP) and gaining some freedom due to its development in the user space of the operating system (OS) and additional functionalities (i.e. congestion control of QUIC over UDP). However, the utilization of TCP and UDP forces the system to stick with the infrastructural characteristics of the selected substrate transport protocol from the beginning of the transmission until it is closed. This limitation has been named "ossification" and it has three main effects: (1) it closes the opportunity to select and modify transport layer protocols at the beginning of a certain transmissions; (2) it leaves little room for transport protocol innovation; (3) it provides limited or non-existent flexibility of the application programming interface (API) [5]. This API serves as the connection point between the application layer and the transport layer. The existence of constraints in the communication between these two layers is directly translated to a standard behavior of the transport protocol with no consideration of the requirements from the application layer. In current MBB, the transport protocols misbehave due to the incapability of adapting to the actual network circumstances and the impossibility of adapting its features to the requirements from upper layers.

In a close future with the implementation of evolved transport services, the API would not only select the best transport protocol based on application requirements, but it would also consider the network status. Even though, this mechanism would require further signaling and interaction, recent advances [7] are evolving in this sense and could provide with a more complex and complete API. Taking into account that each CCA could be more suitable for certain network circumstances or application requirements than others, our work evaluates the best CCA candidate for each combination of conditions. In this regard, out of all possibilities, our analysis covers the study of TCP CUBIC, NewReno, Illinois, Westwood+ and the recently released BBR. Thus, being capable of providing hints in the complex process of improving the behavior in the transport layer and therefore in the resultant enhanced QoS.

The main findings of the chapter are the following. The mobility scenarios under 4G latencies require more aggressive TCP solutions to overcome the high variability in comparison with low latency conditions. Merely focusing on goodput: (1) although BBR provides with the best scalability, it also induces greater mean delays and lost packets; (2) in scenarios that evolve with continuous capacity reductions, Illinois shows the best adaptability to the variable conditions when the achievable capacities are high, whereas CUBIC demonstrates the same for lower bandwidth assignments. Considering the performance with a combination of goodput, delay and retransmissions, CUBIC presents a more balanced behavior with average achieved rates but greater awareness of the self-inflicted delay and retransmissions. With reduced end-to-end latencies, the most suitable CCAs are: (1) NewReno for low available capacity circumstances that moves with continuous capacity increases; (2) BBR as the most balanced CCA that allows both high bandwidth achievement and low delay and retransmissions; (3) CUBIC when scalability is required in presence of big changes between consecutive samples of assigned radio capacity and (4) similar goodput-based performance of all CCAs while moving to worse quality positions.

The chapter is structured as follows: Section 2 introduces the related work in the analysis of TCP in MBB in general and with mobility circumstances in particular. Section 3 shortly describes the utilized CCAs in the analysis. Section 4 covers the methodology with the description of the testbed and the utilized measurement and evaluation process. The analysis and results are explained in Section 5 divided by the 4G latency schemes and the low latency scenarios that mimic 5G deployments. Finally, Section 6 gathers the most important conclusions and proposes future lines.

2. Related work

When analyzing how the mobile networks' features have an effect on transport protocol behavior and therefore impact the final QoS, there are several characteristics that have to be mentioned.

2.1. Delay

It is clear that comparing mobile networks and fixed networks, the former has more variable channel conditions that could lead to achieve a degraded throughput [1, 2]. However, there are effects that from a macroscopic point of view are shared among the distinct networks. For instance, it has been proved [8, 9] that even mobile networks suffer due to the excessive buffering in intermediate queues leading to an increase in the end-to-end delays and dropped packages that severely impact the performance of TCP (bufferbloat effect). Measurements over both 3G and 4G cellular networks of four U.S. providers and Swedish networks have concluded that bufferbloat represent a problem in MBB too. Our study not only considers the achieved capacities but also the induced delay due to the possible impact that may well have as a cross-traffic.

2.2. Impact of variability

It has been demonstrated that there are differences among distinct mobile networks. A comparative work of 3.5G and 4G [10] showed that 4G networks are worse in regards to the TCP efficiency due to the superior throughput and variability. This is the higher variability, the worse scenario for TCP due to the lack of rapid adaptability. Garcia et al. [11] carried out measurements in the cellular networks of different Swedish operators so as to analyze the variation of TCP throughput and delay throughout the day. Sudden increases in traffic load leads to bandwidth variability and latency increment [2]. Therefore, TCP happens to drastically reduce its throughput. Additionally, TCP experiences timeouts many times. The timeout events are especially harmful because the CWND is reduced to one segment. In another study, Alfredsson et al. [12] proved that the variable modulation on the 4G link layer is contributing to retransmissions' increment and therefore higher delay and less throughput. Huang et al. [13] carried out a comprehensive study related to TCP throughput and latency estimation over a live long term evolution (LTE) network. In their measurements, they found out similar timeout events. Our work precisely evaluates how different variable mobility circumstances

http://dx.doi.org/10.5772/intechopen.71779

affect the performance of different CCAs in order to better understand the implications of fluctuations in the channel quality.

2.3. Uplink impact and bursty behavior

Regarding the evaluation of TCP throughput, it has been proved [14] that the uplink performance tends to degrade its performance due to scheduling policies, severely impacting on Acknowledge packet (ACK) arrival and therefore downlink injection ability. Those TCP flavors that merely depend its CWND management upon the reception of ACKs are drastically affected by these ACKs reception, also called ACK-compression. Related to this issue, it has been demonstrated [15] that modern cellular networks' traffic has a tendency to become bursty. For this reason, there can be a large variation in the actual throughput during a short period of time (varying by up to two orders of magnitude within a 10-min interval). This variability could be even harder due to the fact that mobile providers often maintain a large and individual downlink buffer for each UE, provoking high latency instability. Our analysis aims at detecting the most suitable CCA for different MBB mobility circumstances that also implicate distinct bursty conditions over the network.

2.4. Impact of speed

If mobile networks themselves suffer high variability, the channel conditions could be even more variable and challenging for TCP due to the movement of UEs. The movement leads to have distinct propagations and fading patterns over time that at the same time impact on the assigned modulation to the UE, provoking "jumps" between consecutive channel quality reports. Merz et al. [16] studied the performance of TCP in live LTE networks in mobility scenarios with speeds up to 200 km/h. They mainly evaluated the spectral efficiency depending on the modulation and the bandwidth share among the attached users to the eNodeB, together with the ability of those users to make the most of the assigned capacities. Li et al. [17] compared the performance of TCP in static positions with moving scenarios, resulting in harmful RTT spikes, massive dropped packets and eventual disconnections while on the move. Our work complements the mentioned studies by adding the evaluation of multiple state-of-the-art CCAs as well as the inclusion of different MBB schemes with 4G latencies and low latencies.

Taking into account the research and standardization momentum regarding 5G in which most of the work is yet to be fulfilled, there are few works that have considered the performance of TCP in the future 5G MBB. Pedersen et al. [18] demonstrated the potential of using different transmission-time intervals (TTI) in the eNodeB of 5G deployments depending on the metadata related to a certain channel. They showed that shorter TTIs were capable of allowing higher throughputs for short communications, whereas longer TTIs could overall benefit the performance of large transmissions. Sarret et al. [19] study the forthcoming benefit of using full duplex at the radio link layer (RLC) in comparison with the current half-duplex implementation for an improved throughput and delay. Besides, they covered the possible configurations in ultra-dense 5G deployments that could limit the envisioned rates.

Even though several research studies and proposals have reported their concerns and findings regarding the effects between mobile networks and TCP under challenging conditions, none of them have considered the analysis of distinct state-of-the-art CCAs of TCP under different mobility patterns and circumstances over 4G latencies and low latencies targeting proximity MEC-assisted provisioning in 5G.

3. Selected CCAs

The studied TCP variants fall into five categories with regard to their employed CCAs: loss-based, combined loss- and delay-based (with or without bandwidth estimation), and delay-based. As examples of loss-based CCAs, we study both NewReno [20] and CUBIC [21]. NewReno was selected due to its prevalence in research and its large implementation base, and CUBIC by the fact that it is the default CCA in Linux. The Westwood+ [22] congestion control was taken as a CCA example of a combined loss- and delay-based with bandwidth estimation technique. In many ways, TCP Westwood and its successor TCP Westwood+ laid the foundation for the work on designing a CCA that is able to distinguish between congestion and non-congestion related packet losses in wireless networks without any support from the wireless MAC layer. Also, Illinois [23] was selected as an example of a combined loss- and delay-based CCA. In contrast to Westwood+, Illinois primarily targets high-speed and long-delay networks. Finally, TCP BBR (Bottleneck Bandwidth and RTT) [24, 25] constitutes a model-based CCA that drives the congestion avoidance management based on two parameters: measured baseline RTT (delay-based) and the timing and rate of ACK packets (bandwidth estimation). A brief overview of the five studied TCP variants is given below.

TCP NewReno [20] is the basic TCP implementation that drives its CWND based on the additive increase and multiplicative decrease (AIMD) principle. In this regard, NewReno increases its CWND by one packet for each ACK reception during the Slow Start phase. Instead, during the congestion avoidance phase, it increases the CWND by one segment for each RTT. The increment is performed until a timeout period is consumed or a notification of a loss packet is received (with a triple duplicate ACK—3DUPACK). Depending on the event, NewReno would back-off differently, halving the CWND in case of 3DUPACK and establishing the CWND in one segment when a timeout is detected.

TCP CUBIC [21] uses a cubic equation during the congestion avoidance phase to manage the CWND. The closer the CWND is to the previous congestion point in terms of outstanding packets, the slower increment is applied. This function leads to a zero increment while the previous congestion point is achieved. If CUBIC does not detect congestion at that point, it increases the ramp-up pace of the CWND with a convex shape until a new loss event happens. One of the main features of CUBIC in comparison with NewReno is that the CWND is not ACK-clocked and therefore, depend less significantly in the RTT. CUBIC also introduces a modification in the Slow Start phase in order to avoid massive packet

http://dx.doi.org/10.5772/intechopen.71779

losses at the end of the ramp-up. The modification is called Hybrid Slow Start and tries to transfer the management of the CWND to the congestion avoidance phase prior to overfeed the network. To that end, two exit conditions are added: (1) if a delay increase over a predefined threshold is detected and (2) if the ACK train is lengthen. If any of the conditions is met, the Slow Start phase is left and the congestion avoidance phase would follow driving the CWND.

TCP Westwood+ [22] is a sender-side modification of TCP to allow estimating the available bandwidth by assessing the incoming ACK packets. The measured capacity serves to adjust the CWND during back-off phases after a loss event occurs. The selected CWND tries to establish the potential outstanding packets as close to the maximum capacity as possible but avoid building-up the queue of the bottleneck. To that end, the bandwidth-delay product (BDP) is calculated with the estimated bandwidth and the minimum assessed RTT.

TCP Illinois [23] is a loss-based AIMD mechanism that drives the CWND with certain knowledge of the queuing delay and buffer size of the bottleneck. This delay awareness is taken from the RTT measurements and consequently it is updated upon ACK arrival. If no excessive queuing delay is detected, the CWND would increase faster than in conditions of high induced latency. The maximum increment is established in 10 segments per RTT, while the minimum is set to 0.3. When the RTT is close to the maximum, the loss is considered as buffer overflow, whereas in low RTT the loss counts as packet corruption.

TCP BBR [24, 25] is the recently developed TCP implementation that bases its CWND management in a model of the bottleneck's BDP. It considers the estimated bottleneck bandwidth and the measured RTT in every update of the model. The estimated bottleneck bandwidth is measured by calculating the timing and rate of receiving ACKs in the sender. The calculated model determines whether the packet injection rate is below or over the capacity of the bottleneck, being able to appropriately adjust to the network requirements. Such an adjustment of the injection rate is carried out following the principle of *pacing*, either by using Fair Queue packet scheduling or the native and fall-back implementation of *pacing* developed in the transport layer. Besides the main behavioral features, BBR is handled with a four stages workflow:

- In the Startup stage BBR ramps-up as the Standard Slow Start until it detects that the obtained throughput gain is below the 25% throughout three consecutive RTTs.
- In the draining stage BBR tries to get rid of all the excessive packets in the bottleneck queue.
- In the probing bandwidth stage BBR uses an eight state cycle to cruise at different pacing rates. Throughout six states, BBR injects at the measured bottleneck's BDP rate if no change of the available capacity is detected. The other two states are a bandwidth probing phase with a 25% of the injection rate increment and a draining phase with the ability to drain the excess packets injected in the previous phase if the bottleneck does not tolerate greater throughputs.

- In the probing RTT stage, BBR re-measures the baseline RTT for a proper modeling of the network path. To this end, BBR reduces the CWND to four segments for at least 200 ms and then, re-established the previous CWND.

4. Methodology

This section presents the measurement testbed with the equipment involved in the assessment process and describes the measurement and evaluation procedure that have led to the results presented in the following analytical section.

4.1. Testbed

The LTE deployment uses a digital radio testing emulator or a LTE-in-a-box as the main equipment responsible for the LTE side. The main radio configuration parameters are the utilization of the seventh band of LTE due to its widespread usage and the availability of 100 physical resource blocks (PRB) and 20 MHz channels in order to be capable of performing at the full potential of the cell. This emulator plays among other attributes (full EUTRAN/EPC testbed) the role of the eNodeB, creating the LTE signaling to support the attachment and registration of any LTE device through a radiofrequency (RF) cable.

Figure 1 shows the experimental testbed and how the LTE-in-a-box is placed and connected to other equipment in the deployment. Apart from the LTE emulator, the testbed is formed by different parts:

A LTE UE is included with the capability of connecting to the network through the RF cable. Such connection is directly done to avoid undesired effects of the environment in the transmission.

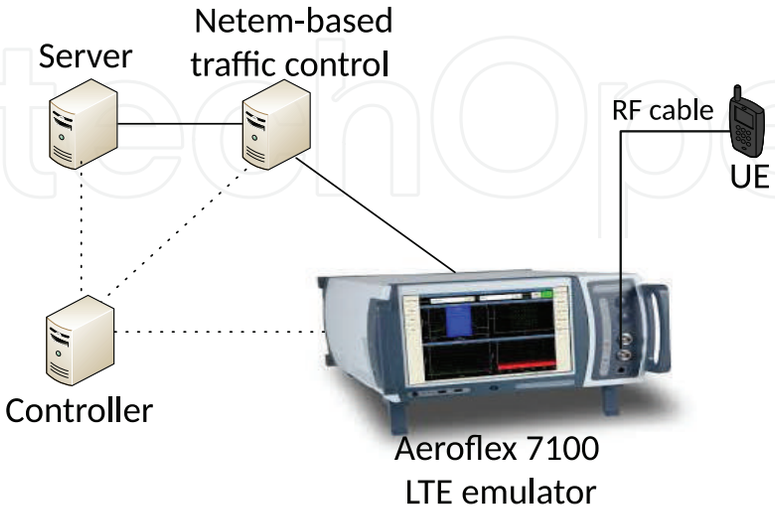

Figure 1. LTE testbed with Aeroflex 7100 LTE-in-a-box.

A Linux server is placed with a 4.10 kernel that contains the recent advances in the transport layer as well as TCP BBR. Besides, the server contains the required files that are used during the experimental phase. The files are devoted to ensure long experiments or greedy sources so as to appropriately detect differences in the behavior and performance of distinct CCAs. In addition, the server is responsible for gathering the logging files created by the *ss* (socket information) and *tcpdump* (whole pcap file of each transmission) programs that later on will be used for the analysis.

One traffic-control bottleneck server is located between the end-server and the LTE-in-a-box to manage the end-to-end delay with *netem*.

A controller is used to automatize and synchronize the rest of the equipment throughout the experiments. The controller is also responsible for commanding the precise configuration that each part of the testbed should apply. For instance, the controller selects which CCA is used in the end-server, the delay of the bottleneck, the baseline signal-to-noise-plus-interference (SINR) of the channel between the LTE-in-a-box and the UE (either as a fixed value or with variations that could allow emulated mobility) or the fading pattern to be utilized in the mentioned channel.

The emulated testbed enables representative reporting and signaling with real UEs, gives the opportunity to configure the LTE-in-a-box including fading patterns and it is capable of collecting logging traces. All this features make the selected testbed realist enough to be a representation of certain real-world network circumstances with the additional control and parameterization of the measurements outcome that a close testbed provides.

4.2. Measurement and evaluation procedure

In order to experiment with 4G latency MBB scenarios as well as low latency schemes that model the proximity service of MEC-assisted 5G future deployments, we have used different delays in the network path. Due to the privacy and non-disclosure information of operators, there is little data regarding the latencies present in 4G and the ones expected in real-world deployments for 5G. Therefore, our study is based on a report [26] that shows the delay results of four operators being between 68 and 85 ms on average for 4G. Thus, we configure our 4G latency scenarios with a minimum latency of 68 ms and the low latency scheme with the lowest possible value in our testbed, 18 ms.

The emulated effect of movement is obtained by the application of two parameters: (1) a selected fading pattern in the LTE-in-a-box that would affect the channel between the eNodeB and the UE and (2) the external SINR traces that periodically command the baseline SINR. These messages are sent by the controller and applied by the emulator in order to modify the baseline SINR of the channel towards the UE.

Regarding fading, since different maximum achievable rates and variability lead to different challenges for TCP, we have applied two distinct fading patterns in regards to modeling common fading effect under mobility in real deployments. The patterns are the following:

- A mobility scenario with Extended Vehicular A model 60 (EVA60) fading model. This fading and variability tries to mimic the vehicular scenarios at 60 km/h, which is a common limitation in rural roads.
- A mobility scenario with High Speed Train (HST) 300 (HST300) fading pattern. The selected fading models the signal fluctuation of current highspeed trains at 300 km/h.

Figure 2 shows the distribution of CQIs while applying the combination of a baseline SINR of 20 dB and the selected fadings. It is clear that both patterns depict a variable behavior even under static baseline SINR circumstances. However, both fading patterns result in a completely different environment for TCP. EVA60 demonstrates a variable behavior in high CQI values (high capacities), whereas HST300 shows even more fluctuation in lower CQI values (lower achievable rates).

Apart from the variability that the fading traces create in the radio link, in order to run under different mobility patterns, the Controller in the testbed would periodically command the baseline SINR of the channel to the emulator. Our experiments evaluate the responsiveness and suitability of the selected CCAs in two simplified mobility conditions:

- Forward movement or out-cell: this mobility pattern evaluates the performance of different flavors of TCP with a UE moving from the eNodeB towards a constantly worsened radio channel conditions. Therefore, the applied mobility traces would start from the baseline SINR of 20 dB and would go progressively worsening until 5 dBs are reached. The speed of such transition is determined by the scenario under study and thus, the 60 km/h mobility scenario spends five times the time is needed to perform the same at 300 km/h. The main idea is to evaluate TCP under variable conditions that progress with an average continuous capacity reduction.
- Backward movement or in-cell: this mobility pattern evaluates the performance of the selected CCAs with a UE moving in the other way around. The mobility traces would start by applying a baseline SINR of 5 dBs to the channel and the quality would progress increasing in accordance to the selected scenario and its modeled speed. The idea is to assess the performance of TCP in variable channel conditions when the quality of the channel tends to continuously increase its available capacity.

We understand that the simplification of movement into backward movement and forward movement does not directly represent the mobility circumstances in the real-world. However, we stand that many movement patterns can be divided or split into the aforementioned cases.

In regards to the utilized traffic, greedy sources are employed with no cross-traffic. The decision of performing with isolated flows is based on the better understanding of the impact that different mobility patterns and fading traces have, together with the better detection of the CCAs' adaptability.

In order to analyze NewReno, CUBIC, Westwood+, Illinois and BBR in the selected mobility scenarios under 4G latencies and low latencies that model the potential MEC-assisted service

http://dx.doi.org/10.5772/intechopen.71779

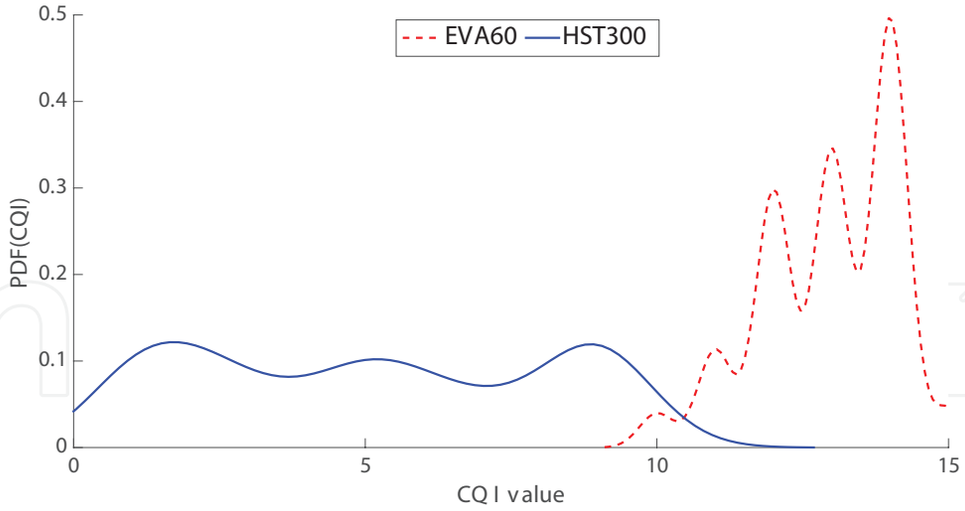

Figure 2. CQI distribution while combining a baseline SINR of 20 dBs and the selected fading patterns.

provisioning delay in 5G, we consider different metrics in the evaluation process. Instead of merely measuring the performance only based on the goodput performance, two different approaches are followed:

1. Criterion 1—pure goodput performance in order to assess the adaptability of CCAs and which are the rates capable of achieving.
2. Criterion 2—a performance metric that includes not only the goodput samples but the delay as well as the retransmission events. This way, a single value is able to include several important fields of the performance of a certain CCA. Thus, the evaluation could consider a more balanced parameter that does not rely on a single performance side, avoiding the cyclic dependency of TCP (i.e. in many situations the achievement of greater goodput—positive fact, also means the injection of greater delay—negative impact). In order to build a metric that considers all abovementioned parameters, we decided to create the following one:

$$A = \frac{K}{Kt} \tag{1}$$

$$B = \frac{Dmin}{D} \tag{2}$$

$$C = \frac{BDP - MSS * R}{BDP} \tag{3}$$

$$\alpha = 1; \beta = 1; \gamma = 1 \tag{4}$$

$$Pm = \frac{(\alpha * A) + (\beta * B) + (\gamma * C)}{\alpha + \beta + \gamma} \quad (5)$$

The first parameter (Eq. (1)) measures how much out of the available capacity (Kt) is reached in a precise sample (K). The second one (Eq. (2)) indicates the growth of current delay (D) considering the baseline delay of the transmission ($Dmin$). The third one (Eq. (3)) takes into account out of the current BDP (BDP), how many bytes are wasted in retransmissions (number of retransmissions—R times the maximum segment size—MSS). α, β and γ (Eq. ((4)) are parameters to weight the importance of the three principal. These parameters could be configured for the precise requirements of a certain application (critical with losses or critical with delays among other options). In our evaluation (Eq. (5)), our performance metric (Pm) considers every main parameter equally important to get a balanced performance.

The evaluation process aims at deciding the most appropriate options among selected CCAs depending on the selected mobility use-case and application requirement (evaluation criterion). The analysis and evaluation is linked with the current and state-of-the-art possibilities that are being opened in the transport layer so as to select the most suitable protocol at the beginning of every transmission [7]. Our work contributes to the better selection of CCAs under 4G latency and low latency MBB mobility scenarios.

5. Analysis

This section covers the overall performance and most suitable selection of TCP under mobility in 4G latencies and under low latency scenarios. In order to better explain the analysis, the section is divided into two subsections: (1) Section 5.1 covers the analysis regarding the performance of TCP under 4G latencies in mobility scenarios, together with the evaluation and selection of most TCP candidate considering goodput requirements and performance metric requirements; (2) Section 5.2 performs the same analysis and evaluation in low latency mobility scenarios. Each subsection covers three main steps in the explanation, it presents the overall performance, the goodput-based evaluation and the assessment based on the performance metric.

Since throughout the analysis we will use a performance metric that also considers the injected delay and number of retransmissions, the representation of the overall performance will take into account three parameters: mean goodput, mean delay and mean retransmissions per time slots of 100 ms. In order to avoid complicated graphs that may well be misinterpreted, the overall performance is depicted in spider plots of three axis, one per performance parameter. Each figure comprises four subplots: the first two subplots on the top depict the results for forward movement pattern, whereas, the bottom line shows another two subplots for backward movement pattern. The two subplots per line are representative of the different speed and fading conditions: from the left to the right, the scenarios at 60 and 300 km/h.

http://dx.doi.org/10.5772/intechopen.71779

Supported by the spider plot of the overall performance and adding a table that gathers the mean values and the average confidence intervals of goodput in each mobility scenario, the most appropriate CCA is selected. In this sense, the selection may well be utilized by requirement of the applications to pick among all CCAs the option that best maximizes the achieved rates in a precise mobility scheme.

Once selected the best candidates for mobility scenarios merely based on goodput, it is important to carry out a similar task but evaluating the performance of distinct TCP flavor from a point of view that gathers more information about the outcome and possible side-effects. To that end, a figure will depict the empirical cumulative distribution function (ECDF) results of the performance metric in all scenarios. The distribution of subplots is the same as before with forward movement scenarios in the first row and backward movement ones in the bottom line. Each column represents a different speed and fading combination, being from the left to the right the scenarios at 60 and 300 km/h, respectively.

5.1. TCP performance with mobility under 4G latencies

This section covers the analysis under 4G latencies of the selected movement patterns of five CCAs. **Figure 3** shows the spider plot results in mobility scenarios for CUBIC (straight line), NewReno (dotted line), Westwood+ (straight thick line), Illinois (dash-dotted line) and BBR (dashed line).

Out of all the results in **Figure 3**, the most important ones are the following:

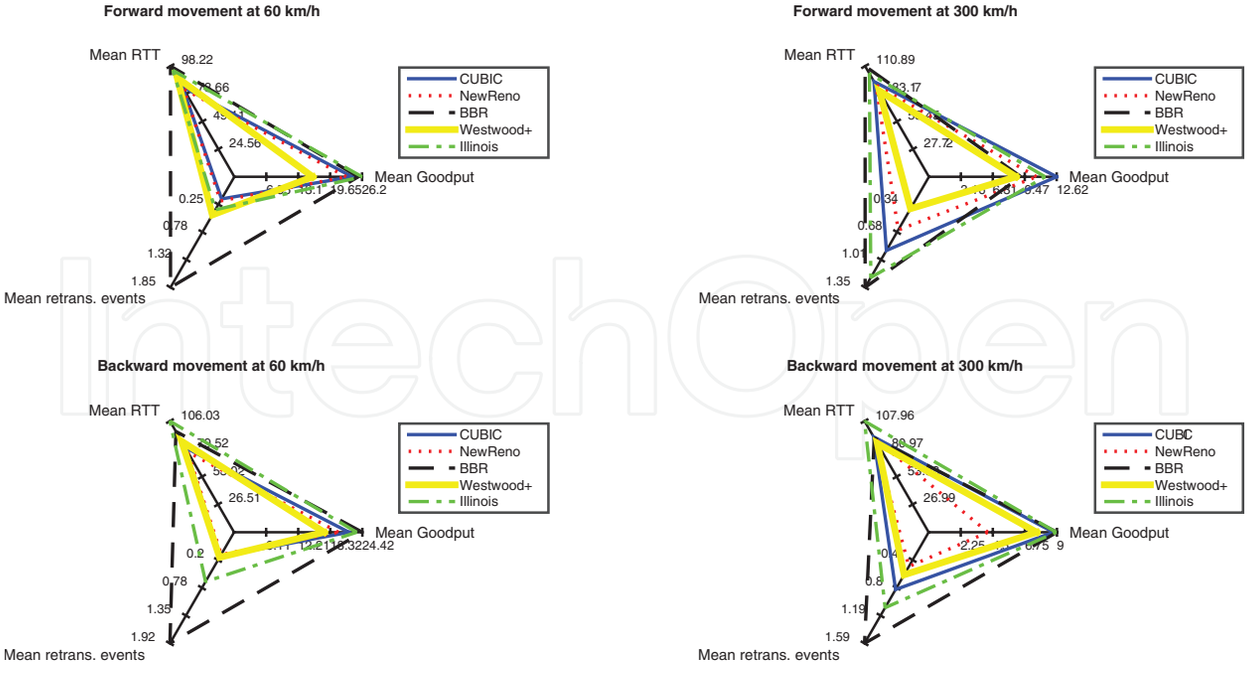

Figure 3. Performance spider plot of five CCAs while moving at different speeds under 4G latencies: forward movement on top; backward movement on bottom line.

- In forward movement pattern: We find two different cases: (1) when the variability happens at high capacities (scenario at 60 km/h) since the movement itself regardless the CCA under use allows having almost throughout the whole experiments packets in-flight, the mean achieved goodput is very similar in most of CCA cases. The best candidates are Illinois and BBR in terms of goodput but they also induce more delay and retransmissions (more significant with BBR) than NewReno or CUBIC; (2) at 300 km/h where the mean fluctuation is harder due to the superior speed, the self-inflicted effects of delay and retransmissions severely impact the goodput performance, dropping the goodput performance of BBR and Illinois and being clearly surpassed by CUBIC; (3) as detected in Ref. [27] Westwood+ suffers due to the excessive reduction in the back-off application after Slow Start, leading to an underuse of the radio capacity.
- In backward movement pattern: (1) both Illinois and BBR has shown better scalability than CUBIC in all scenarios achieving greater goodput rates, but as a drawback, inducing more delay in the network and suffering more retransmissions; (2) the scenario itself due to its low available rates at the beginning of the transmission helps appropriately perform weak CCAs such as Westwood+. In this sense, the deficiency of Westwood+ is minimized, leading to better results in terms of goodput while the delay is kept low; (3) considering goodput, apart from NewReno that demonstrates to underperform in variability circumstances when scalability is required under 4G latencies, the rest of the candidates achieve similar (not equal) results. The scenario at 60 km/h shows bigger differences among the CCA candidates due to the greater capacities present in the scenario, allowing more aggressive CCAs adapt better to the available bandwidth.

We have detected cases in which a similar goodput is achieved but significantly more delay and retransmissions are suffered. These examples, that suppose a difficult performance trade-off to analyze, are the foundation for the evaluation of the protocols based on different points of view in the performance in order to appropriately select the best candidate for each network circumstances but also considering the application requirements.

The overall performance has depicted the goodput performance as one of the parameters in the spider plot. Now instead, **Table 1** covers the performance of the goodput showing the actual average values of the transmission throughout each mobility scenario, together with the average confidence interval as a representation of the differences between independent tests.

Table 1 shows that depending on the scenario a proper CCA selection could allow the achievement of greater capacities. The best practises regarding the goodput-based evaluation are the following under 4G latencies:

- In backward movement pattern (right side of the table), BBR shows great variability in the behavior. Even with that, it is able to scale better than any other candidate. This ability could be of a great value in other MBB scenarios in which scalability is required, either as a necessary feature or as a response to available bandwidth increments.

http://dx.doi.org/10.5772/intechopen.71779

Context	Forward movement		Backward movement	
	Mean	Mean CI	Mean	Mean CI
60 km/h with EVA60 under 4G latencies				
CUBIC	24.18	6.59	21.87	6.89
NewReno	22.95	7.95	19.87	7.16
BBR	25.43	18.7	24.42	17.54
Westwood+	16.08	8.01	17.16	7.68
Illinois	26.2	10.7	23.35	14.84
300 km/h with HST300 under 4G latencies				
CUBIC	12.62	7.81	8.76	5.51
NewReno	10.63	7.58	4.15	2.89
BBR	9.09	10.21	9	6.76
Westwood+	8.61	4.13	7.58	4.28
Illinois	11.48	6.8	8.95	5.84

Table 1. Goodput-based evaluation of the selected CCAs in mobility scenarios under 4G latencies.

- In forward movement pattern (left side of the table), the two different scenarios require a distinct treatment and therefore CCA candidate. The best practises in forward movement pattern under 4G latencies are the following:

  1. Illinois is selected in scenarios with big bandwidth jumps between consecutive RTTs due to its aggressiveness to handle such variability but also due to its delay awareness. The combination of features provides the best and most consistent (less variation in the behavior than the second candidate according to the results—BBR) goodput performance.
  2. CUBIC is picked in variable scenarios with smaller changes between consecutive available bandwidth samples. The aggressive features of CUBIC are suitable to handle great fluctuation in smaller capacities but do not make the most when it comes to bigger bandwidth jumps.

Table 1 reveals that in forward movement, an appropriate CCA selection could provide with a great gain in terms of goodput even using a movement pattern that makes easier the task of TCP due to the avoidance of starvation events in the bottleneck buffer.

Once selected the best candidates for mobility scenarios under 4G latencies merely based on goodput, it is important to carry out a similar task but evaluating the performance of distinct TCP flavor from a point of view that gathers more information about the TCP performance. To that end, **Figure 4** depicts the ECDF results of the performance metric in all scenarios. The

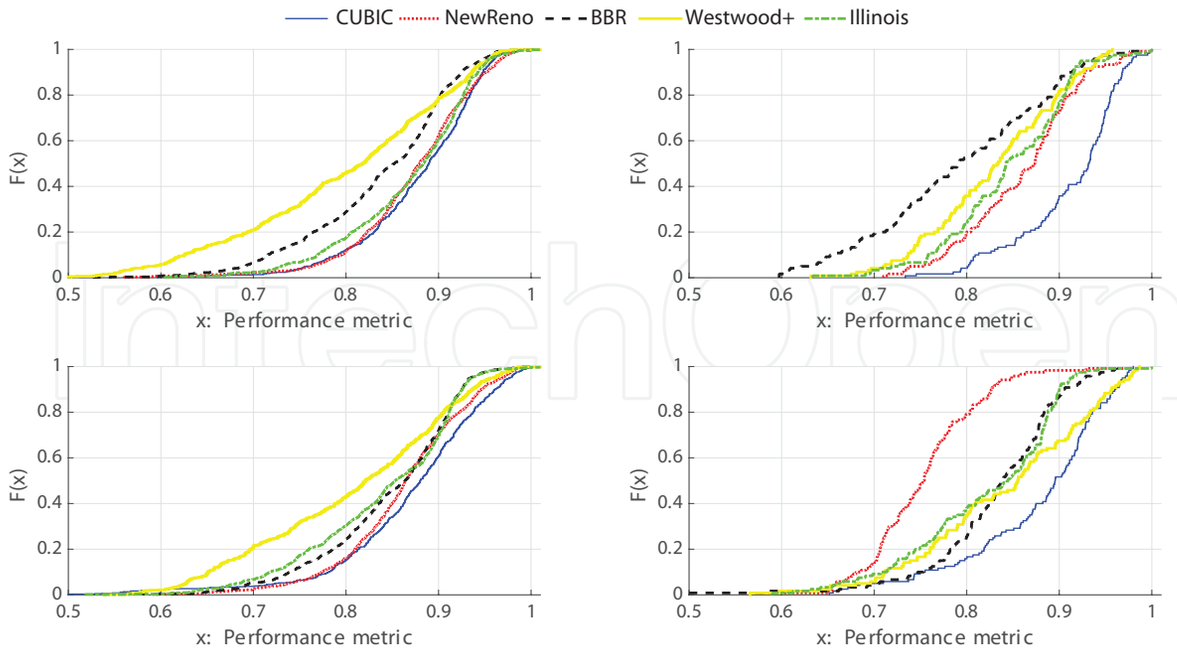

Figure 4. Comparison of five CCAs performance metric while movement at different speeds under realistic 4G latencies: forward movement on top; backward movement on bottom line.

distribution of subplots is the same as before with forward movement scenarios in the first row and backward movement ones in the bottom line.

Figure 4 contrasts with previous goodput-based evaluation in many ways. It shows that sometimes the TCP performance depends on the proportion between parameters, suffering a performance cycle that impacts negatively certain performance fields when others are improved and vice-versa. In general in forward movement, **Figure 4** demonstrates that the performance enhancements are not that clear in one CCA over the other. Maybe one CCA is able to perform better in terms of goodput but at the same time suffering from delay and retransmitted packets. Therefore, taking into account the three selected performance factors, it is not that easy to decide whether we should use one congestion control or the other. If the decision of the CCA is taken upon the evaluation of this performance metric, in the scenarios with similar outcome, the final CCA would be picked based on other circumstances. For instance, the decision is taken with: (a) developer preference; (b) the by-default CCA prevails over the others; (c) if at the point of selection there is one CCA already established and it is among the group with similar outcome, the CCA could remain selected. In contrast and still selecting the most appropriate CCA in forward movement pattern, the 300 km/h scenario is clearly dominated by CUBIC. CUBIC not only achieves a great goodput, but shows a more general and extensive good performance.

Regarding the results in backward movement scenarios, there is a clear pattern that reflects that CUBIC is more appropriate. The outcome contrasts the evaluation based on goodput that suggests the selection of BBR, proposing to pick CUBIC as a candidate that improves the overall performance based on the three selected performance aspects: goodput, delay and impact of retransmissions.

http://dx.doi.org/10.5772/intechopen.71779

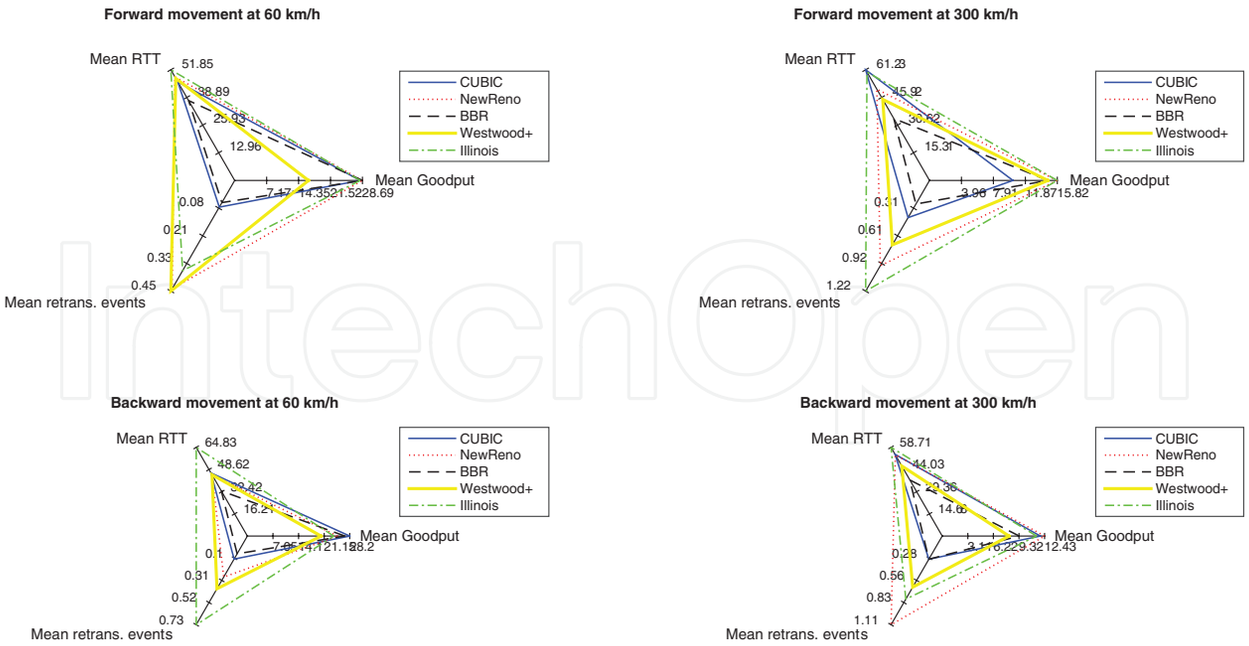

Figure 5. Performance spider plot of five CCAs while moving at different speeds under low latencies: forward movement on top; backward movement on bottom line.

5.2. Overall performance under mobility under low latency

Once explained the performance of the selected CCAs over the mobility scenarios under 4G latencies and having evaluated the best TCP candidate considering the goodput and the performance metric, this subsection covers the analysis under low latencies. **Figure 5** shows the spider plot results in mobility scenarios for CUBIC (straight line), NewReno (dotted line), Westwood+ (straight thick line), Illinois (dash-dotted line) and BBR (dashed line).

The most important details in **Figure 5** are the following:

- As a general performance and comparing with 4G latencies, Illinois still suffers due to an excessive delay and number of retransmissions, whereas BBR shows a more controlled behavior with a good performance in some scenarios in terms of goodput but also avoiding massive delay and retransmitted packets.
- In forward movement pattern: (1) there still exist huge similarities among CCA from the goodput's point of view. Under low latency circumstances, even the CCAs that in principle are less aggressive (i.e. NewReno in both scenarios or Westwood+ at 300 km/h), demonstrate to perform very similar to more aggressive candidates, thus achieving a great goodput performance but suffering from excessive delay and retransmissions; (2) moving at 300 km/h where the achievable capacities are low, aggressive solutions such as CUBIC obtain worse results than others due to their injection pattern. Even though the mean retransmission events are lower than Illinois or NewReno, the periodicity is different. In each "big" retransmission event of Illinois or NewReno, many retransmitted packets are involved, whereas in the case of CUBIC the "big" retransmission events are more but with less packets involved. Thus, CUBIC due to its aggressiveness, finds

more times "big" retransmission events (increasing the suffered mean RTT), reducing in more occasions the CWND and therefore, being less capable of achieving the available capacity.

- In backward movement pattern: (1) Westwood+ and Illinois are not able to perform better than CUBIC or NewReno in any case. The former primarily due to its poor ability to scale. The latter, due to its self-inflicted effects, suffering by far the greatest delay; (2) in the case of BBR, as mentioned earlier, the behavior under low latency looks more conservative. In this regard, in backward movement pattern that requires scalability, in terms of goodput BBR performs worse than CUBIC in the mobility scenario at 60 km/h and underperforms in the scenario at 300 km/h earning very poor performance results.

Once again and even more noticeable than with 4G latencies, we have detected cases in which a similar goodput is achieved but significantly more delay and retransmissions are suffered by different CCAs. Those examples require an evaluation of the protocols based on different application-layer requirements so as to appropriately select the best candidate whose performance matches the application requirements and network conditions.

Table 2 only covers the representation of the goodput with average values of the transmission for each mobility scenario and the average confidence interval for different independent tests.

Table 2 demonstrates that depending on the scenario, the selection of the most appropriate CCA is capable of allowing better performance, achieving greater capacities. All in all, under low latency conditions, the most suitable CCA regarding the goodput-based evaluation are the following:

- In forward movement, CUBIC is damaging in scenarios with drastic fluctuation in low available bandwidths comparing with other candidates and its aggressiveness looks as if does not properly fit for such scenarios. Due to the low end-to-end latency and taking into account that forward movement is an easier movement pattern to handle by TCP (because it allows having almost every time packets in-flight), we can barely decide one specific CCA in each scenario. The selection should be carried out among a group of CCA that have reported a very similar outcome in terms of goodput.

1. NewReno and Illinois have shown a great performance in variable scenarios with low available capacities (mobility scenario at 300 km/h).

2. Under variable conditions with big capacities and big jumps between resource assignments, all CCAs but Westwood+ have proven to achieve very similar performance. Comparing with 4G latency scenarios, these similarities clearly come due to the reduction in the end-to-end delay.

3. In backward movement, even with similarities among the CCAs, it is still clear that one or two candidates prevail over the others. Variable scenarios are better handled by CUBIC when the capacities are higher due to its superior aggressiveness, whereas

http://dx.doi.org/10.5772/intechopen.71779

NewReno improves the performance of variable scenarios with low available bandwidths. These results demonstrate that, yet, there is value in the performance of by-default CCA (CUBIC) and classic CCA (NewReno) in current and more importantly, future networks with low latency scenarios.

Once we have evaluated and selected based on goodput the best candidates for low latency mobility scenarios, it is crucial to evaluate the performance of the TCP flavors from the point of view of the performance metric. **Figure 6** depicts the ECDF results of the performance metric in all scenarios.

Taking into account that the current evaluation considers three different aspects of the overall performance, the results in **Figure 6** widely differ from the outcome merely based on goodput. There is a clear pattern that confirms BBR as the best candidate in forward movement scenarios. Previous evaluation has shown that based on goodput in all scenarios, BBR is within the group of best performers. Its capacity to reduce the delay close to the baseline delay in a movement pattern that strongly suffers due to the contrary, makes BBR the best candidate considering the performance metric.

In backward movement pattern, there are two clear candidates, one per each speed and fading combination:

Context	Forward movement		Backward movement	
	Mean	Mean CI	Mean	Mean CI
60 km/h with EVA60 under low latencies				
CUBIC	28.18	8.84	28.21	9.02
NewReno	28.68	9.4	24.89	10.36
BBR	28.69	8.29	26.84	8.41
Westwood+	16.57	10.83	20.3	12.27
Illinois	28.19	9.1	23.81	11.78
300 km/h with HST300 under low latencies				
CUBIC	10.33	5.98	11.91	5.4
NewReno	15.79	5.86	12.43	4.97
BBR	14.66	5.38	9.35	5.32
Westwood+	14.66	4.76	8.11	5.58
Illinois	15.82	5.48	11.46	4.92

Table 2. Goodput-based evaluation of the selected CCAs in mobility scenarios under low latencies.

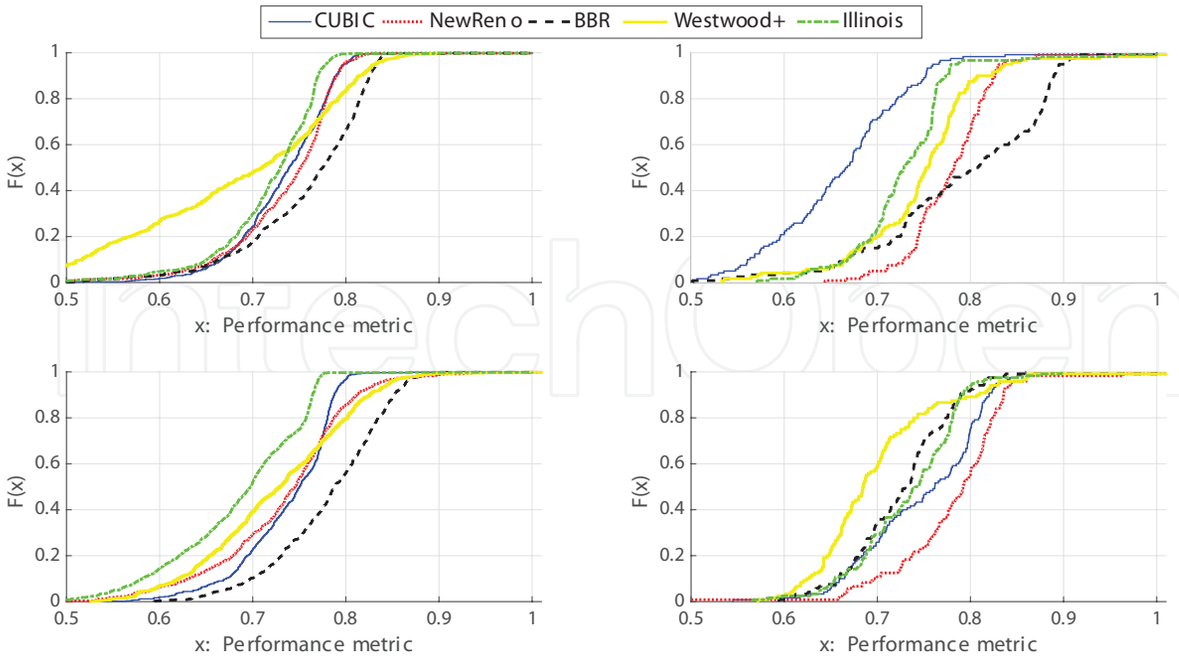

Figure 6. Comparison of five CCAs performance metric while movement at different speeds under low latencies: forward movement on top; backward movement on bottom line.

1. BBR is able to better handle the required scalability under variable fading conditions with big available capacities and jumps between consecutive assignment samples. As under 4G latencies, BBR demonstrates great scalability. In addition, under low latency circumstances BBR is capable of better managing the induced delay, achieving a considerably better performance than the rest of the TCP implementations in the commented network use-case.

2. The combination of low latency together with low available bandwidths makes the performance playground very suitable for the selection of NewReno, showing great performance both in achieved goodput and based on the performance metric.

Overall the results in **Figure 6** show two important facts in comparison with the mobility scenarios under 4G latencies: (1) the performance of BBR shows a consistent good performance in terms of goodput. The main difference is that in low latency scenarios the behavior of the CCA is more controlled and is able to take full advantage of its features, achieving great goodput while the delay is kept close to the baseline latency. This outcome shows that, the shorter response times (RTT) in each packet, the better for the accurate estimation of BBR in mobility scenarios; (2) based on the performance metric, it is clear that the reduction in the end-to-end latency makes Illinois perform worse in comparison with other candidates. While in 4G latencies, Illinois is eligible in many mobility circumstances that take advantage of it aggressiveness, the same scenarios under shorter latencies do not require its aggressiveness. Actually, this feature only increment the injected delay as well as the number of retransmitted packets.

http://dx.doi.org/10.5772/intechopen.71779

6. Conclusion

After evaluating the QoS of selected mobility scenarios, based on goodput and the performance metric, under both 4G latencies and low latency that mimic the potential delays in 5G networks, we have been able to select the most appropriate CCA for each situation. The selected CCAs are gathered in **Table 3** with an asterisk for the cases in which more than one CCA could be picked.

The mobility scenarios under 4G latencies require more aggressive TCP solutions to overcome the high variability with an increased BDP in comparison with low latency conditions. With reduced end-to-end latencies that model the proximity service provisioning in 5G, the most suitable CCAs are a compendium of classic NewReno for low available capacity circumstances that move in-cell, BBR as the most balanced CCA that allows both high bandwidth achievement and low delay and retransmissions, CUBIC when scalability is required in presence of big changes between consecutive samples of assigned radio capacity and equality in goodput-based outcome in out-cell movement. In this sense, in our pseudo-5G mobility scenarios, it is important to highlight the usability of even weak CCAs such as NewReno in future low latency deployments. The explanation of the selected TCP candidates and their implications in the mobility scenarios is followed CCA-wise.

Illinois has shown an aggressiveness that is suitable for application with goodput requirements in both movement patterns in 4G latency scenarios, but only partially appropriate for forward movement under low latency conditions. The reduction in the end-to-end delay increases the responsiveness of TCP in general and Illinois in particular. This effect makes Illinois suffer greater mean delays and retransmitted packets due to the aggressive injection and ramp-up, leading to deteriorate results (i.e. achieving poor results in backward movement).

BBR has demonstrated good performance in terms of goodput under 4G latencies. However, it also induced a long delay and provokes a great number of transmitted packets, demonstrating an unbalanced behavior in terms of the performance metric. In low latency scenarios instead, BBR has a more controlled and balanced behavior, achieving a sufficiently good performance in goodput but also preserving a low delay.

Context	**Low latency**		**4G latencies**	
	Goodput	**Metric**	**Goodput**	**Metric**
60 km/h with EVA60				
Forward	*	BBR	Illinois	*
Backward	CUBIC	BBR	BBR	CUBIC
300 km/h with HST300				
Forward	*	BBR	CUBIC	CUBIC
Backward	NewReno	NewReno	BBR	CUBIC

Table 3. Wrap-up of most appropriate CCAs for each mobility scenario and QoS metric.

Westwood+ has demonstrated a poor performance in majority of the scenarios due to its aggressive back-off policy. However, in scenarios with low achievable capacities (at 300 km/h) Westwood+ is capable of slightly bridging the performance gap with other CCAs.

NewReno, under low latency, provides with a TCP implementation able to achieve close to the maximum available capacity, being especially significant its performance and scalability in backward movement pattern with low achievable rates.

CUBIC has shown a balanced performance with low available bandwidth under 4G latencies. Apart from that, under low latency circumstances, CUBIC is appropriate for forward movements with variable big available bandwidths and scenarios that depend upon scalability.

The future work will progress in three main lines. Firstly, the analysis will be extended with the measurement of similar mobility scenarios in a selection of real-word use-cases. Secondly, the work would re-measure and compare the evolution of BBR in terms of adaptability and management of losses. Finally, we will evaluate pseudo-random mobility scenarios that combine both forward and backward movement as well as static positions.

Acknowledgements

This work is partially funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644399 (MONROE) through the open call project MARiL and by the Spanish Ministerio de Economía y Competitividad (MINECO) under grant TEC2016-80090-C2-2-R (5RANVIR). The views expressed are solely those of the author(s).

Author details

Eneko Atxutegi*, Jose Oscar Fajardo and Fidel Liberal

*Address all correspondence to: eneko.atxutegi@ehu.eus

University of the Basque Country (UPV/EHU), Bilbao, Spain

References

[1] Koksal CE, Jamieson K, Telatar E, Thiran P. Impacts of channel variability on link-level throughput in wireless networks. In: Joint International Conference on Measurement and Modeling of Computer Systems, ser. SIGMETRICS '06/Performance '06; 2006. New York: ACM; 2006. pp. 51-62

[2] Nguyen B, Banerjee A, Gopalakrishnan V, Kasera S, Lee S, Shaikh A, Van der Merwe Jacobus J. Towards understanding TCP performance on LTE/EPC mobile networks. In: 4th Workshop on All Things Cellular: Operations, Applications & Challenges; 2014. ACM; 2014. pp. 41-46

[3] European Telecommunications Standards Institute (ETSI). Mobile Edge Computing (MEC); Technical Requirements [Internet]. 2016 [Updated: 2016]. Available from: http://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/01.01.01_60/gs_MEC002v010101p.pdf [Accessed: 2017]

[4] The Chromium Projects. QUIC, a multiplexed stream transport over UDP [Internet]. Available from: https://www.chromium.org/quic [Accessed: 2017]

[5] Grinnemo KJ, Jones T, Fairhurst G, Ros D, Brunstrom A, Hurtig P. Towards a flexible internet transport layer architecture. In: 2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN); 2016;**2016**:1-7

[6] Jones T, Fairhurst G, Perkins C. Raising the datagram API to support transport protocol evolution. In: Workshop on Future of Internet Transport (FIT 2017); 2017. 2017

[7] Hurtig P, Alfredsson S, Bozakov Z, Brunstrom A, Evensen K, Grinnemo K-J, Fosselie Hansen A., Rozensztrauch T. A NEAT approach to mobile communication. In: The ACM SIGCOMM 2017 Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2017); 2017. ACM Press; 2017

[8] Jiang H, Liu Z, Wang Y, Lee K, Rhee I. Understanding bufferbloat in cellular networks. In: 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design; 2012. New York. ACM; 2012. pp. 1-6

[9] Alfredsson S, Del Giudice G, Garcia J, Brunstrom A, De Cicco L, Mascolo S. Impact of TCP congestion control on bufferbloat in cellular networks. In: World of Wireless, Mobile and Multimedia Networks (WoWMoM); 2013. IEEE; 2013. pp. 1-7

[10] Garcia J, Alfredsson S, Brunstrom A. A measurement based study of TCP protocol efficiency in cellular networks. In: 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt); 2014; 2014. pp. 131-136

[11] Garcia J, Alfredsson S, Brunstrom A. Examining TCP short flow performance in cellular networks through active and passive measurements. In: 5th Workshop on All Things Cellular: Operations, Applications and Challenges, ser. AllThingsCellular '15; 2015; ACM; 2015. pp. 7-12

[12] Alfredsson S, Brunstrom A, Sternad M. Cross-layer analysis of TCP performance in a 4G system. In: Software, Telecommunications and Computer Networks (SoftCOM); 2007; IEEE; 2007. pp. 1-6

[13] Huang J, Qian F, Guo Y, Zhou Y, Xu Q, Mao ZM, Sen S, Spatscheck O. An in-depth study of lte: effect of network protocol and application behavior on performance. In: ACM SIGCOMM 2013 Conference on SIGCOMM; 2013; ACM; 2013. pp. 363-374

[14] Zhang L, Okamawari T, Fujii T. Performance evaluation of end-to-end communication quality of LTE. In: Vehicular Technology Conference (VTC Spring); 2012; IEEE; 2012. pp. 1-5

[15] Xu Y, Wang Z, Leong WK, Leong B. An end-to-end measurement study of modern cellular data networks. passive and active measurement. Lecture Notes in Computer Science. 2014;**8362**:34-45

[16] Merz R, Wenger D, Scanferla D, Mauron S. Performance of LTE in a high-velocity environment: A measurement study. In: 4th Workshop on All Things Cellular: Operations, Applications, & Challenges, ser. All Things Cellular '14; 2014; New York. ACM; 2014. pp. 47-52

[17] Li L, Xu K, Wang D, Peng C, Xiao Q, Mijumbi R. A measurement study on TCP behaviors in HSPA+ networks on high-speed rails. In: 2015 IEEE Conference on Computer Communications (INFOCOM); 2015; 2015. pp. 2731-2739

[18] Pedersen KI, Niparko M, Steiner J, Oszmianski J, Mudolo L, Khosravirad SR. System level analysis of dynamic user-centric scheduling for a flexible 5G design. In: 2016 IEEE Global Communications Conference (GLOBECOM); 2016; 2016. pp. 1-6

[19] Sarret MG, Berardinelli G, Mahmood NH, Mogensen P. Impact of transport control protocol on full duplex performance in 5G networks. In: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring); 2016; IEEE; 2016. pp. 1-5

[20] Henderson T, Floyd S, Gurtov A, Nishida Y. The NewReno Modification to TCP's Fast Recovery Algorithm [Internet]. 2012 [Updated: 2012]. Available from: https://tools.ietf.org/html/rfc6582 [Accessed: 2017]

[21] Ha S, Rhee I, Xu L. CUBIC: A new TCP-friendly high-speed TCP variant. SIGOPS Operating Systems Review. 2008;**42**(5):64-74

[22] Mascolo S, Casetti C, Gerla M, Sanadidi MY, Wang R. TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links. In: 7th Annual International Conference on Mobile Computing and Networking; 2001; New York. ACM; 2001. pp. 287-297

[23] Liu S, Başar T, Srikant R. TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks. In: 1st International Conference on Performance Evaluation Methodolgies and Tools, ser. Valuetools '06; 2006. New York: ACM; 2006

[24] Cardwell N et al. BBR: Congestion-based congestion control. Communications of the ACM. 2017;**60**(2):58-66

[25] Cardwell N, Cheng Y, Yeganeh S, Jacobson V. BBR Congestion Control [Internet]. 2017 [Updated: 2017]. Available from: http://www.ietf.org/internet-drafts/draft-cardwell-iccrg-bbr-congestion-control-00.txt [Accessed: 2017]

[26] fiercewireless. 3G/4G wireless network latency: How did Verizon, AT&T, Sprint and T-Mobile compare in Q3 2015? [Internet]. 2016 [Updated: 2016]. Available from: http://www.fiercewireless.com/special-report/3 g-4 g-wireless-network-latency-how-did-verizon-at-t-sprint-and-t-mobile-compare-q3 [Accessed: 2017]

[27] Atxutegi E, Liberal F, Grinnemo K-J, Brunstrom A, Arvidsson Å, Robert R. TCP behaviour in LTE: impact of flow start-up and mobility. In: IFIP Wireless and Mobile Networking Conference - WMNC; 11-13 July 2016; Colmar, France