Article

Traffic Steering for eMBB in Multi-Connectivity Scenarios

Jesús Burgueño ¹ , Isabel de-la-Bandera ^{1,*} , David Palacios ² and Raquel Barco ¹

¹ Department of Communications Engineering, University of Malaga, 29071 Málaga, Spain; jesusbr@ic.uma.es (J.B.); rbm@ic.uma.es (R.B.)

² Tupl Spain S.L., Tupl Inc., 29010 Málaga, Spain; david.palacios@tupl.com

* Correspondence: ibanderac@ic.uma.es

Received: 11 November 2020; Accepted: 1 December 2020; Published: 3 December 2020

Abstract: Multi-connectivity (MC) is one of the most important features to be introduced in 5G networks, allowing User Equipment (UE) to simultaneously aggregate radio resources from several network nodes to enhance both data rates and reliability. Thus, this feature enables a further flexibility in the allocation of resources to the UEs in order to fulfil the users' requirements in more complex 5G scenarios. This paper takes advantage of this wide flexibility to present a traffic steering approach that determines the amount of traffic to be held by each of the serving nodes in a multi-connectivity scenario. In this sense, the proposed technique is based on network and UE performance metrics in order to maximize the users' perceived quality of experience (QoE) for enhanced Mobile Broadband (eMBB) services. It is then compared with a homogeneous traffic split among the serving nodes, with a single-connectivity approach and with state-of-the-art solutions. The benefits are analysed in terms of throughput and Mean Opinion Score (MOS), which is the main QoE metric. The analysis shows that a noticeable UE throughput improvement is reached when the proposed traffic steering method is applied. Consequently, this enhancement is noticed in the users' QoE, which can lead to a reduction of operating expenses (OPEX) of the network.

Keywords: traffic steering; multi-connectivity; eMBB; quality of experience

1. Introduction

The arrival of the Fifth-Generation (5G) standard has introduced new service categories such as Ultra-Reliable Low-Latency Communications (URLLC), which support services demanding high reliability and low latency, massive Machine-Type Communications (mMTC) services, which are characterized by supporting a high number of devices with low demands of throughput and enhanced Mobile Broadband (eMBB) services with the objective of providing high throughput connectivity [1]. Therefore, new 5G networks must address several requirements depending on the service demanded by the User Equipments (UEs).

In addition to traditional requirements that a service must meet, such as packet delay budget or packet error loss rate, next-generation networks will present a user-centric approach, which will focus on achieving the quality of experience (QoE) that the user demands for each service. The most relevant QoE metric is the Mean Opinion Score (MOS), whose value represents the subjective service quality that a user perceives on a scale of 1 (bad) to 5 (excellent).

In the context of eMBB services and in order to cope with their new requirements, new solutions have been proposed in the literature. This is the case of millimetre wave (mmWave), whose underused frequency bands become possible to boost the network capacity [2]. Moreover, the extremely short wavelength of the mmWaves allows a large number of antenna elements to be grouped together in a

small space. This enables carrying out massive MIMO (Multiple Input Multiple Output). This feature further improves the spectral and energy efficiency of the network.

Nonetheless, the solution that will be addressed in this paper and one of the most promising functionalities to be used in 5G networks is Multi-Connectivity (MC) [3], whose implementation will allow UEs to simultaneously aggregate radio resources from several network nodes. Hence, this feature will enable aggregating resources from different nodes or duplicate the delivery of packets through several links in order to increase throughput or reliability, respectively. In addition, MC can be used to reduce the latency of the mobility procedures or enhance mobility robustness [4]. Therefore, depending on the objectives of the use case, different evaluation metrics such as throughput, packet loss rate or latency can be taken into account.

Currently, MC is described by 3GPP as Multi-Radio Dual-Connectivity (MR-DC) [3]. Presently, this mechanism allows the simultaneous aggregation of radio resources from only two different network nodes. In this way, MR-DC allows UEs to simultaneously connect to an evolved Node B (eNB) and a 5G New Radio node B (gNB), as well as connecting to two gNBs. One of these nodes will adopt the role of the master node (MN), carrying the signalling between the core network (CN) and the UE and determining the UE radio resource control (RRC) state, as well as delivering data to the UE. The other node will assume the role of a secondary node (SN) whose functionality is only to deliver data to the user.

In addition to MR-DC, carrier aggregation (CA) allows a UE to aggregate several spectrum chunks from a single node in order to increase the bandwidth assigned to the same user. These spectrum chunks are called Component Carriers (CC) and each one is managed by a single scheduler. Consequently, CA allows MC-enabled UEs to use multiple CCs simultaneously at each of the serving nodes.

To bring significant benefits to eMBB services, MC is expected to include the simultaneous connection of a UE to more than two nodes in the upcoming 3GPP releases. In this way, an important throughput boost can be achieved as a result of using MC. Nonetheless, this improvement would imply that the resources available in the network nodes should be properly scheduled, since resource management is more complex when the UEs are simultaneously connected to several nodes. Hence, this paper proposes a technique to decide the amount of resources to be allocated to a UE by each of its serving nodes in a MC-enabled scenario, which has not been established by current 3GPP release.

Although 3GPP does not establish how many radio resources should be managed by each serving node in MC scenarios, 3GPP indicates that the selection of the serving nodes is made based on Reference Signal Received Power (RSRP) [5]. However, some studies propose alternatives to this standard selection. Although not in a MC scenario, authors from [6] propose a method of serving CC selection based on the assignment of the most downloaded CCs in a CA-enabled scenario. A more recent study provides a MC management scheme based on Channel State Information (CSI) and cell load to improve the throughput and decrease the probability of radio link failure [7]. This technique dynamically adds or removes serving nodes based on the introduced metrics. The authors in [8] propose a rule-based system that determines which CCs should be assigned to a specific UE based on quality and network load metrics. However, once the serving CCs are assigned, the amount of resources provided by each CC is the same in these studies, i.e., a homogeneous traffic split is made among the serving CCs. On the other hand, other studies propose a selection of the CCs based on RSRP since 3GPP establishes it but an uneven traffic split among the serving CCs. The authors in [9] present a technique in which the largest amount of traffic is provided by the nodes that serve a smaller number of users. This method aims at maintaining the quality of service required by old users when a new user requests access to network resources. Furthermore, authors from [10] propose a scheme in which the total traffic is unevenly distributed among the serving nodes in order to minimize the delay of the complete transmission. In this way, less traffic is served by the secondary nodes that provide a greater delay to the user. These two techniques will be compared with the performance of the method proposed in this paper to analyze the improvement achieved with respect to state-of-the-art studies.

In a completely different way, authors from [11] propose a packet duplication scheme to improve the performance of URLLC services in exchange for increasing resource usage in MC scenarios. Finally, the combined use of network slicing and MC is proposed in [12]. In this case, eMBB and URLLC traffic are allocated in different network slices to ensure that URLLC traffic reaches the minimum required latency.

Beyond the state-of-the-art, this paper proposes a novel traffic steering technique that aims at maximizing the QoE perceived by users of eMBB services in new MC scenarios. The objective of this study is to take advantage of the flexibility in resource allocation to further optimize the network performance. Thus, this technique focuses on deciding the amount of traffic that each of the serving CCs will provide to the UE. To this end, in addition to network load metrics, quality measurements will be used by the proposed technique. In this context, this study makes the following contributions:

- The technique proposed is especially optimized for eMBB traffic, so high throughput connectivity is required. Hence, quality measurements reported by UEs are also used, contrary to what happens in state-of-the-art studies.
- The MOS is used to evaluate the performance of the proposed technique in different eMBB services. In this way, the performance of the proposed technique is compared with the performance obtained by using state-of-the-art techniques, in addition to the performance of the single connectivity and MC with homogeneous traffic split approaches. This can demonstrate that the deployment of this technique can lead to a reduction of operating expenses (OPEX) of the network.
- The MC used in this study enables simultaneous connection with more than two nodes, addressing a solution for the upcoming 3GPP releases. In addition, several CCs can be assigned to a UE by a single node, thus the scenario is also CA-enabled.

The rest of the paper is organized as follows. Section 2 presents the proposed traffic steering technique for eMBB traffic in MC scenarios. Performance analysis is then presented in Section 3 followed by concluding remarks in Section 4.

2. Traffic Steering Solution for eMBB Traffic

This section details the technique that decides what percentage of the total traffic is held by each of a user's serving CCs. As indicated above, these CCs may belong to more than two nodes. This allows great flexibility in resource allocation. Thus, sub-optimal network performance may be reached if this flexibility is not used to carry out intelligent traffic steering. This means that the network operator may be wasting resources if a simple approach is deployed, i.e., each serving CC provides the same amount of resources to the UE regardless of network conditions. Therefore, the conditions of the UEs can be optimized if a heterogeneous traffic steering solution is considered.

The proposed solution should be based on the key factors that affect the metric to be optimized. In this study, the solution is focused on maximizing the QoE perceived by users of eMBB traffic. The QoE of eMBB services depends mainly on throughput, whereas reliability is not critical, i.e., data packets are not duplicated and delivered by different nodes in the proposed approach. Therefore, the best way to improve the QoE of eMBB users is to increase their throughput, which depends on the available resources that can be assigned to the UE and the signal quality. Specifically, the number of available physical resource blocks (PRBs) and the Reference Signal Received Quality (RSRQ) have been selected as inputs for the proposed algorithm. In addition, to adapt the proposed solution to the 5G standard, the RSRQ may be exchanged for new quality metrics, such as Channel State Information SINR (CSI-SINR) or Synchronization Signal SINR (SS-SINR) [13].

Before the traffic split can be performed and the MN distributes the data to the SNs, the selection of the CCs that will serve the UE must be made. As described above, several CCs may be selected at each serving node. As 3GPP describes for Dual Connectivity scenarios [5], in this paper it is assumed that the CCs selection is done according to the RSRP; i.e., a UE is allocated the strongest CCs. The number

of assigned CCs is actually based on network operator policies. Likewise, the maximum number of CCs that can be allocated to a UE was increased from 5 to 32 in Release 13 [14].

After a UE has been allocated CCs from different nodes, the SNs deliver the information of available PRBs in the serving CCs to the MN through the Xn interface. Also, the UE reports to the MN the RSRQ of the links with each serving CC. Once the MN knows these values, the proposed technique can be used to calculate the amount of total traffic that each CC should provide to the UE. Then, once the MN receives the full data traffic from the User Plane Function (UPF), the MN knows the amount of data to be delivered to each SN. Therefore, the proposed technique will be implemented within the Quality of Service flow handling block defined by 3GPP [15]. This block allows traffic to be split into multiple radio bearers. Thus, each radio bearer is responsible for carrying the traffic to be served by each serving CC.

Initially, the proposed technique defines two weights (W_{Load} and W_{RSRQ}). These weights represent the relevance of available bandwidth and signal quality for traffic split, respectively. Both weights are configured by the network operator and sum up to one, so that increasing the relevance of the available bandwidth (W_{Load}) implies reducing the relevance of the signal quality (W_{RSRQ}).

Later, as the MN knows the available bandwidth in each serving CC, it creates an array of normalized values ($\overline{PRB}[cc]$) in order to assign a score based on the resources availability to each serving CC. It is given by Equation (1):

$$\overline{PRB}[cc] = \frac{PRB[cc]}{max\_PRB} \quad for \quad cc \in [1, Number \ of \ serving \ CCs] \quad (1)$$

where max_PRB is the maximum number of available PRBs that a CC could provide individually to the UE, and $PRB[cc]$ is the array of available PRBs in each serving CC. On the other hand, Equation (2) carries out the same process but based on RSRQ values of each link as follows:

$$\overline{RSRQ}[cc] = \frac{RSRQ[cc]}{max\_RSRQ} \quad for \quad cc \in [1, Number \ of \ serving \ CCs] \quad (2)$$

where max_RSRQ indicates the link that provides the maximum RSRQ to the UE. In addition, $RSRQ[cc]$ indicates the array of RSRQ for the different links. Afterwards, the normalized arrays are multiplied by W_{Load} and W_{RSRQ} , respectively, and these new values are then added for each serving CC (Equation (3)). This is expressed as follows:

$$Score[cc] = W_{Load} \cdot \overline{PRB}[cc] + W_{RSRQ} \cdot \overline{RSRQ}[cc] \quad for \quad cc \in [1, Number \ of \ serving \ CCs] \quad (3)$$

where $Score[cc]$ is a new array consisting of an overall score for each serving CC. The traffic steering will be based on this array as defined in Equation (4):

$$\% \ traffic[cc] = 100 \cdot \left(\frac{Score[cc]}{\sum_{i=1}^N Score[cc]} \right) \quad for \quad cc \in [1, Number \ of \ serving \ CCs] \quad (4)$$

where $\% \ traffic[cc]$ indicates the amount of total data traffic provided by each CC, which corresponds to the percentage of total score with which it contributes. Thus, the MN will distribute the traffic flows among the SNs via the Xn interfaces based on the amount of data traffic that each CC will provide to the UE and considering that several serving CCs may belong to the same node. Finally, the traffic will be managed by the scheduler used in each serving CC. Figure 1 represents an example of the proposed method in which each step is detailed numerically for a single UE. It should be noticed that $\overline{PRB}[cc]$, $\overline{RSRQ}[cc]$, W_{Load} and W_{RSRQ} are randomly generated in this example. At the end, it can be seen how CC #2 provides almost a third of the total traffic.

Additionally, Figure 2 shows a use case of the proposed traffic steering solution. In this figure, the central UE (in red) is accessing a video service. Since it demands high-throughput connectivity,

three nodes are assigned to it simultaneously. At first, it is shown how the MN receives the full data traffic from the UPF through the N3 interface. Then, the MN uses the proposed technique to calculate the amount of data that has to deliver to each SN via the Xn interfaces. The figure shows how the MN is serving a large group of UEs, probably carrying a high load on all their CCs, thus the MN only provides the UE with a small part of the complete data traffic. Hence, the MN delivers most of the data traffic to the SNs. The SN #1 provides a low RSRQ to the UE due to a non-line of sight link. In contrast, the SN #2 serves a single UE with high RSRQ. These reasons imply that the SN #2 provides most of the data that the UE receives. Finally, the percentage of total data traffic that each serving node delivers to the UE is represented over the link with stripes.

Figure 1. Example of the traffic steering solution with random initial values.

Figure 2. Use case of the traffic steering solution.

3. Performance Analysis

3.1. Scenario Setup

In this section, the proposed method for traffic steering is tested. To that end, the performance of the proposed technique is compared with a homogeneous traffic distribution in which each of the serving CCs provides the same amount of data traffic to the UE, as well as with state-of-the-art proposals. In this way, Simsek [9] and Zhang's [10] approaches have been implemented into the simulator. For the sake of clarity, the last name of the first author of each study has been used to name its proposed method within this section where the performance of all approaches is analyzed. As indicated in the first section, more traffic is provided by nodes that serve fewer users in Simsek. On the other hand, Zhang minimizes the traffic delay difference among MN path and SNs paths. Finally, the results are also compared with a single-connectivity approach in which the MN delivers the complete data traffic to the UE. Hence, several system-level simulations have been performed to assess the benefits.

For these tests, a Matlab simulator with MC capabilities, based on [16], has been used. An urban macrocellular scenario has been considered, consisting of 12 tri-sectorized sites uniformly distributed. Both W_{RSRQ} and W_{Load} are configured with the same values for all sectors. They are set before each simulation starts and are kept constant during each execution. In addition, to reduce the computational load, only the downlink has been simulated. The main simulation parameters are detailed in Table 1, but the reader is referred to [16] for additional details about the simulation tool.

Each UE uses one of four possible eMBB services. These are real time video (RTV), video streaming (VS), whose functioning implies a non-real time service, file download (FD) and web browsing (WB) [17]. The distribution of services is uniform, i.e., the average number of UEs using each of the services is the same. However, the amount of traffic generated by each of these services is different. In this sense, video services generate a greater amount of traffic on the network, as can be seen in Table 1.

Finally, to show the benefits of the proposed method, the perceived QoE is calculated for each service, evaluating its corresponding MOS. The definition of the MOS for each of the four services is shown in [18,19]. Both the MOS for FD (Equation (5)) and WB (Equation (6)) are defined as:

$$MOS_{FD} = \max(1, \min(5, 0.0065 T - 0.54)) \quad (5)$$

$$MOS_{WB} = 5 - \frac{578}{1 + \left(\frac{T+541.1}{45.98}\right)^2} \quad (6)$$

where T is the average session throughput in kilobits per second. For VS service, the MOS is given by Equation (7):

$$MOS_{VS} = 4.23 - 0.0672 T_{init} - 0.742 F_{reb} - 0.106 T_{reb} \quad (7)$$

where T_{init} is the initial buffering time in seconds, F_{reb} is the average stalling frequency in seconds⁻¹ and T_{reb} is the average stalling duration in seconds. On the other hand, the MOS for RTV is modeled as in Equation (8):

$$MOS_{RTV} = \begin{cases} 5 & \text{if } BT \geq BTT \\ 1 + 4 \frac{BT}{BTT} & \text{if } 0 < BT < BTT \\ 1 & \text{if } BT \leq 0 \end{cases} \quad (8)$$

where the MOS depends on the playout time corresponding to the packets in the receiver buffer (BT , buffered time). In addition to BT , the buffered time threshold (BTT) indicates the minimum buffer size before playout starts and is set to 5 s [19]. Finally, the average user download throughput is analyzed too.

Table 1. Simulation assumptions .

Environment	Urban Macro network with 12 sites, 3 sectors/site and 5 CCs per sector
Carrier	10 MHz carrier bandwidth at 2 GHz
Propagation model	Pathloss: Okumura-Hata Slow fading: Log-Normal with $\sigma = 8$ dB, $d_{corr} = 20$ m Fast fading: EPA model
PHY numerology	15 kHz sub-carrier spacing 12 subcarriers/PRB (180 kHz)
Data channel MCS	QPSK to 64 QAM with same encoding rates as specified for LTE
Antenna configuration	SISO, $P_{TXmax} = 46$ dBm
Scheduler	Classical exponential/proportional fair [20]
Link adaptation	CQI-based
UE distribution	1200 eMBB UEs (100 UEs per site and 300 UEs per service on average)
FD traffic model	File size: log-normal distribution (avg. 2 MB)
WB traffic model	Web page size: log-normal distribution (avg. 3 MB) No. pages per session: log-normal distribution (avg. 4) Reading time: exponential distribution (avg. 30 s)
RTV traffic model	BTT = 5 s Video codec: H.264, Resolution: 480 p Video bitrate range: 700–900 kbps Video duration: uniform distribution [0, 540] s
VS traffic model	Video codec: H.264, Resolution: 720 p Video bitrate range: 1.2–1.5 Mbps Video duration: uniform distribution [0, 540] s

3.2. Results and Discussion

At first, the value of W_{RSRQ} has been set from zero to one for each of the simulations, making W_{Load} to be configured accordingly. In this sense, the horizontal axis of Figures 3 and 4 represents each weight's configuration. Thus, different weights configurations may cause differences in network performance when using the proposed technique. However, it is clear that the other approaches do not depend on weights. Hence, the same performance will be represented for these approaches although the weights configuration changes in order to the gains are noticeable in these figures.

Figure 3. Median of the download throughput per UE.

Figure 4. Median of the MOS per UE for each service.

Figure 3 shows the median of the UE download throughput in Mbit/s. On the other hand, Figure 4 consists of four subfigures that represent the median of the MOS per UE for the four services considered in this work. These figures mainly point out that the download throughput and MOS are highly correlated in eMBB traffic, as can be deduced from the above MOS equations. This means that the maximum throughput and the maximum of all MOS values appear roughly for the same weights configuration. This best case is achieved when W_{RSRQ} equals 30% and, consequently, W_{Load} is 70%. Given this configuration, these weights indicate that the proposed technique is carrying out a load balancing which is also quite dependent on the quality provided by the serving CCs. Hence, as the relevance of the nodes with more available resources is higher, the amount of radio resources provided by overloaded network nodes will be lower. Thus, this configuration allows more users to access network resources, so the median of the download throughput per UE will increase instead of reaching very high peak data rates for a few users. In addition, although not directly related, the number of users served is highly correlated with the number of resources occupied. Therefore, given the relevance of the load, the figures indicate that Simsek is the second best technique in terms of overall performance.

However, these figures also show how a traffic steering that depends only on the network load will not achieve an optimal network performance for eMBB traffic, i.e., the MOS and throughput values are not the maximum for $W_{Load} = 100\%$ and $W_{RSRQ} = 0\%$. This demonstrates the benefit of using quality metrics for resource allocation in MC scenarios rather than just load metrics for eMBB users that demand high-throughput connectivity. This fact is especially noticeable in the case of RTV, where the proposed technique obtains a worse performance than single connectivity for this configuration. In this case, the demanding requirement for real-time data cannot be met using downloaded CCs that provide poor quality. Therefore, using single connectivity decreases the data rate in exchange for minimizing the latency, which is more efficient in reaching a higher MOS compared with the proposed technique configured with $W_{Load} = 100\%$. Likewise, given this latency requirement in RTV service, Figure 4 shows how Zhang achieves the second highest MOS despite reaching a lower throughput than the rest of MC approaches.

Therefore, the proposed technique will be set to provide the maximum download throughput and, consequently, the maximum MOS values ($W_{RSRQ} = 30\%$ and $W_{Load} = 70\%$). Given this weights configuration, Figure 5 represents the relative gain of the UE download throughput and MOS achieved by each introduced technique with respect to the single-connectivity approach. This figure shows how the MOS obtained for all services are higher with the proposed method, as well as the download throughput. At first, it can be noticed how a high enhancement in the MOS of the FD service is achieved. This gain is greater for FD than for WB because WB users perceive good QoE even using only MC, whereas FD users consider that a huge improvement has been made with the proposed technique. The Equations (5) and (6) confirm that FD users are more demanding since the FD MOS is lower than WB MOS for the same average session throughput. In addition to the equations, this can be seen in Figure 4.

Figure 5. Relative gain of the UE download throughput and MOS achieved by each introduced technique with respect to the single-connectivity approach.

In the case of video services, a greater weight of resource availability allows more users to receive data without any interruption. However, the proposed configuration must also be quality-aware in order to reach high data rates that allow the video quality demanded to be correctly received. In this sense, VS users demand higher video quality, as can be seen in Table 1. This implies that a greater improvement in throughput will be more appreciated by VS users. On the other hand, the QoE of RTV users will be affected by the delay added when using MC, since RTV is the only considered real-time service. Based on these reasons, the MOS improvement is lower for RTV than for VS when using the proposed method as depicted in Figure 5.

In summary, Simsek takes into account the number of users served at each node, which correlates to the network load. This allows it to achieve significant improvement with respect to single connectivity and homogeneous split approaches. Nevertheless, the proposed method further optimizes network performance once links quality is considered. Thus, the proposed technique exceeds the throughput and the perceived MOS achieved by the rest of approaches. On the other hand, Zhang and homogeneous split approaches achieve a worse overall performance. Regarding Zhang, its performance is poor, except for video services. In this respect, it is better than Simsek for the RTV service, which is the most demanding service in terms of latency. Hence, the proposed system maximizes the QoE perceived by eMBB users, leading to a decrease in user complaints. This would imply a reduction in the number of incidents to be manually addressed by network management experts, decreasing network OPEX.

4. Conclusions

In this work, the current state of MC and its possibilities with the arrival of the fifth-generation networks were introduced. Next, one of the most important MC challenges is addressed: the radio resource allocation of the nodes that simultaneously serve a UE. For this purpose, this paper presents a solution focused on optimizing the QoE perceived by the eMBB users. Given the eMBB traffic requirements, the proposed method uses the RSRQ metric in addition to load measures. In this context, the MN uses this technique to steer the traffic that must be held by each serving CC of the SNs.

The results of the several system-level simulations show that eMBB users' QoE is maximized when the proposed technique works as a quality-conscious load balancing. This means that, in addition to the need for available resources, taking into account the quality provided by each serving node is also essential to reach higher MOS values in eMBB services that demand high-throughput connectivity. Its performance was then compared with the performance of state-of-the-art techniques, and the proposed method achieves a great improvement in eMBB users' QoE, as well as in the download throughput per UE. This can lead to a decrease in user complaints in new MC scenarios.

Author Contributions: The contribution of authors are: Conceptualization, J.B., I.d.-l.-B. and D.P.; methodology, J.B., I.d.-l.-B. and D.P.; software, J.B. and I.d.-l.-B.; validation, J.B. and I.d.-l.-B.; formal analysis, J.B.; investigation, J.B. and I.d.-l.-B.; resources, R.B.; data curation, J.B. and I.d.-l.-B.; writing—original draft preparation, J.B.; writing—review and editing, I.d.-l.-B., D.P. and R.B.; visualization, J.B.; supervision, I.d.-l.-B. and R.B.; project administration, R.B.; funding acquisition, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been performed in the framework of the Horizon 2020 project ONE5G (ICT-760809), receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CA	Carrier Aggregation
CC	Component Carrier
CN	Core Network
CSI	Channel State Information
CSI-SINR	Channel State Information SINR
eMBB	enhanced Mobile Broadband
eNB	evolved Node B
FD	File Download
gNB	5G New Radio node B
MC	Multi-connectivity
MIMO	Multiple Input Multiple Output
mMTC	massive Machine-Type Communications
MN	Master Node
MOS	Mean Opinion Score
MR-DC	Multi-Radio Dual-Connectivity
OPEX	Operating Expenses
PRB	Physical Resource Block
QoE	Quality of Experience
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTV	Real Time Video

SN	Secondary Node
SS-SINR	Synchronization Signal SINR
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra-Reliable Low-Latency Communications
VS	Video Streaming
WB	Web Browsing

References

1. 3GPP. TR 38.913, Study on Scenarios and Requirements for Next Generation Access Technologies; V14.3.0, Rel-14. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2996> (accessed on 2 December 2020).
2. Bogale, T.E.; Le, L.B. Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefits and Challenges. *IEEE Veh. Technol. Mag.* **2016**, *11*, 64–75. [CrossRef]
3. 3GPP. TS 37.340, NR; Multi-Connectivity; Overall Description; Stage-2; V15.7.0, Rel-15. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198> (accessed on 2 December 2020).
4. Suer, M.; Thein, C.; Tchouankem, H.; Wolf, L. Multi-Connectivity as an Enabler for Reliable Low Latency Communications—An Overview. *IEEE Commun. Surv. Tutor.* **2019**, *22*, 156–169. [CrossRef]
5. 3GPP. TR 36.842, Study on Small Cell Enhancements for E-UTRA and E-UTRAN; Higher Layer Aspects; Rel-12. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2543> (accessed on 2 December 2020).
6. Wang, Y.; Pedersen, K.I.; Sorensen, T.B.; Mogensen, P.E. Carrier load balancing and packet scheduling for multi-carrier systems. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 1780–1789. [CrossRef]
7. Ba, X.; Wang, Y. Load-Aware Cell Select Scheme for Multi-Connectivity in Intra-Frequency 5G Ultra Dense Network. *IEEE Commun. Lett.* **2019**, *23*, 354–357. [CrossRef]
8. Mahmood, N.H.; Laselva, D.; Palacios, D.; Emara, M.; Filippou, M.C.; Kim, D.M.; de-la-Bandera, I. Multi-Channel Access Solutions for 5G New Radio. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 15–18 April 2019; pp. 1–6. [CrossRef]
9. Simsek, M.; Höfler, T.; Jorswieck, E.; Klessig, H.; Fettweis, G. Multiconnectivity in Multicellular, Multiuser Systems: A Matching- Based Approach. *Proc. IEEE* **2019**, *107*, 394–413. [CrossRef]
10. Zhang, B.; Xu, X.; Zhang, K.; Zhang, J.; Teng, V. Goodput-Aware Traffic Splitting Scheme with Non-ideal Backhaul for 5G-LTE Multi-Connectivity. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp.1–6. [CrossRef]
11. Mahmood, N.H.; Karimi, A.; Berardinelli, G.; Pedersen, K.I.; Laselva, D. On the Resource Utilization of Multi-Connectivity Transmission for URLLC Services in 5G New Radio. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 15–18 April 2019; pp. 1–6. [CrossRef]
12. Zhang, K.; Xu, X.; Zhang, J.; Zhang, B.; Tao, X.; Zhang, Y. Dynamic Multiconnectivity Based Joint Scheduling of eMBB and uRLLC in 5G Networks. *IEEE Syst. J.* **2020**. [CrossRef]
13. 3GPP. TS 38.215, NR; Physical Layer Measurements; V15.4.0, Rel-15. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3217> (accessed on 2 December 2020).
14. Bhamri, A.; Hooli, K.; Lunttila, T. Massive carrier aggregation in LTE-advanced pro: Impact on uplink control information and corresponding enhancements. *IEEE Commun. Mag.* **2016**, *54*, 92–97. [CrossRef]
15. 3GPP. TS 38.300, NR; NR and NG-RAN Overall Description; Stage 2; V16.3.0, Rel-16. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3191> (accessed on 2 December 2020).
16. Muñoz, P.; de-la-Bandera, I.; Ruiz, F.; Luna, S.; Barco, R.; Toril, M.; Lázaro, P.; Rodríguez, J. Computationally-Efficient Design of a Dynamic System-Level LTE Simulator. *Int. J. Electron. Telecommun.* **2011**, *57*, 347–358. [CrossRef]

17. ICT-317669-METIS/D6.1, Mobile and Wireless Communications Enablers for the Twenty-Two Information Society (METIS). Available online: <https://cordis.europa.eu/docs/projects/cnect/9/317669/080/deliverables/001-METISD61v1pdf.pdf> (accessed on 2 December 2020).
18. Gijón, C.; Toril, M.; Luna-Ramírez, S.; Marí-Altozano, M.L. A Data-Driven Traffic Steering Algorithm for Optimizing User Experience in Multi-Tier LTE Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 9414–9424. [CrossRef]
19. Navarro-Ortiz, J.; Lopez-Soler, J.M.; Stea, G. Quality of experience based resource sharing in IEEE 802.11e HCCA. In Proceedings of the 2010 European Wireless Conference (EW), Lucca, Italy, 12–15 April 2010; pp. 454–461.
20. Rhee, J.H.; Holtzman, J.M.; Kim, D.K. Scheduling of real/non-real time services: Adaptive EXP/PF algorithm. In Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference, VTC 2003-Spring, Jeju, Korea, 22–25 April 2003; Volume 1, pp. 462–466.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).