End-user Benefits of LTE Dual Connectivity in Heterogeneous Networks

John Camilo Solano Arenas and Torsten Dudda
Radio Research Germany, Ericsson
52134 Herzogenrath, Germany
{john.camilo.solano.arenas, torsten.dudda}@ericsson.com

Johannes Schmitz and Rudolf Mathar
Chair for Theoretical Information Technology, RWTH Aachen University
52074 Aachen, Germany
{schmitz, mathar}@ti.rwth-aachen.de

Abstract

In response to the ever increasing traffic demand in mobile communication networks, e.g. LTE, features for radio resource aggregation are discussed. LTE Dual Connectivity (DC), standardized in 3rd Generation Partnership Project (3GPP) Release 12, is envisaged to improve the end-user performance by increasing the user throughput. This is achieved by aggregating resources from two separate eNBs operating on different carrier frequencies. In a heterogeneous network these eNBs are typically a Macro eNB providing large scale coverage, and a Pico eNB to boost capacity locally. The serving eNB can directly transmit a part of the buffered packets to the User Equipment (UE) while a second part is offloaded to a second eNB for transmission to the UE from there. We evaluate the end-user benefits of LTE DC in a realistic heterogeneous network deployment considering typical internet traffic types of today, e.g. webpage downloads. Our results show gains for the end-users with DC depending on the traffic type. This indicates that not only the achievable user throughput, but also the amount of data buffered in the eNBs and the latency dependency of the considered traffic determine the end-user experience. This work shows that DC is mostly beneficial when buffering in the eNBs occurs. Furthermore, for the worst users in the system, i.e. the cell edge users, DC enables to significantly reduce the webpage download times especially at medium system load. Therefore, DC can be considered an interesting feature to be added to LTE networks to cope with the increasing traffic demand of the upcoming years.

1 Introduction

Overall, mobile data traffic is exponentially growing within the next years, with a ten-fold increase forecast by the end of 2021 [1]. This ever increasing traffic demand in mobile communication networks is mostly (and will continue to be) due to the high spread of connected devices, especially smartphones and the increased usage of mobile data. For that, mobile operators have started to improve and densify the existing LTE radio networks with the deployment of small cells within the macro cell coverage, i.e. heterogeneous network deployments. The small cells are deployed for offloading the macro cells and improving outdoor and indoor coverage as well as cell edge user performance. Therefore, new LTE network features, such as Dual Connectivity (DC), have been introduced in order to provide an enhanced end-user experience, e.g., by aggregating radio resources from different cells and thus increasing the user throughput.

In DC, introduced in 3rd Generation Partnership Project (3GPP) Release 12 [2] and depicted in Figure 1, a user equipment (UE) may utilize radio resources aggregated from two separate eNBs operating on different carrier frequencies and connected with a non-ideal backhaul channel, i.e. a typical backhaul such as DSL access which comprises a large delay [3]. With a non-ideal backhaul channel between the eNBs, not only the deployment of the nodes becomes more flexible but it also allows the schedulers to operate independently. The involved eNBs, typically a macro cell and a pico cell, are denoted as Master eNB (MeNB) and Secondary eNB (SeNB).

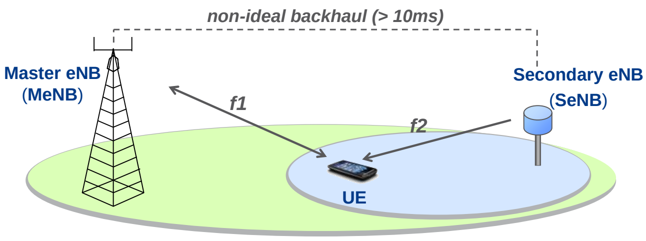

Figure 1 LTE Dual Connectivity (DC)

Two different DC user plane architectures are described in [2]: one in which the user plane data terminates in the MeNB and it can be transferred from the MeNB to the SeNB via the backhaul channel, and a second architecture where the user plane data can terminate in the SeNB as well. The first option, which is of interest in this work, is called DC split bearer architecture and it is depicted in Figure 2. The MeNB is the controlling node of the UE within this architecture, i.e. it terminates the control plane connection towards the UE. For the downlink user plane transmission, the UE may utilize aggregated radio resources from the MeNB as well as the SeNB. The aggregation point is at the MeNB Packet Data Convergence Protocol (PDCP) layer [4]. There, the MeNB may transmit PDCP packets to the UE via itself or by offloading part of the traffic via the non-ideal backhaul channel to the SeNB for further transmission to the UE. On the other hand, in Release 12 for the uplink user plane, the UE can transmit packets to the network only via one of the eNBs.

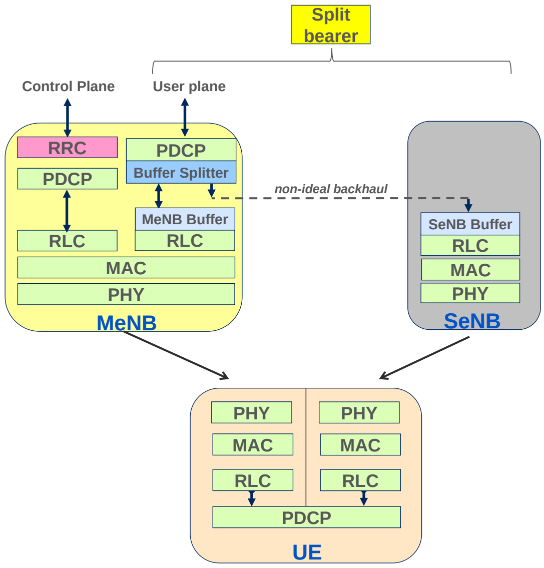

Figure 2 Dual Connectivity split bearer architecture

The PDCP packets are encapsulated IP packets (e.g. from internet traffic). Since these IP packets typically originate from a latency dependent traffic type, e.g. web browsing traffic, it is critical that the MeNB transmits the packets as fast as possible to the UE. For that, we introduce in this work a method to take a forwarding decision for each required packet at the MeNB. The decision relies on selecting the path with the lowest estimated packet delay to the UE.

To evaluate and analyze the end-user benefits of LTE DC, we take into consideration typical internet traffic types of today: large file transmissions and multiple small file transmissions. Large file transmissions, e.g. FTP large file size downloads, can be considered as full-buffer traffic, whereas multiple small file transmissions, e.g. webpage downloads, as a bursty traffic. Depending on these traffic types, the eNB may require to buffer packets before its MAC scheduler finds the next transmission opportunity for the respective UE (as depicted in Figure 3). For DC, the buffering of packets may occur at the eNB PDCP layer.

The paper is outlined as follows. Section 2 describes the PDCP buffer splitting method which is proposed within this work to efficiently forward the packets to the UE via either

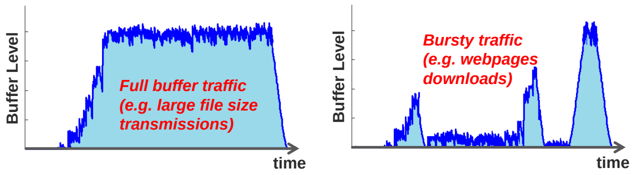

Figure 3 Buffering of Packets at the eNB

the MeNB or the SeNB. Simulation assumptions are given in Section 3. Section 4 provides performance evaluation results of the potential end-user benefits of DC. Finally, Section 5 concludes the paper.

2 Delay Compensated PDCP Buffer Splitting

The MeNB needs to forward the PDCP packets, Protocol Data Units (PDUs), to the UE via either the MeNB itself or via the non-ideal backhaul channel and the SeNB. For that, the MeNB performs a PDCP buffer splitting, i.e. the MeNB takes a forwarding decision for each packet in order to select the transmission path to the UE. Therefore, we propose within this work a new method based on a deterministic fluid approximation and feedback information in order to select the fastest path for each packet to reach the UE.

The deterministic fluid approximation is used in the MeNB to estimate the delays of a packet being transmitted to the UE at a time t via the MeNB or the non-ideal backhaul channel and the SeNB. Based on Figure 2, the estimation of the delays is done by the MeNB buffer splitter and it relies on the queuing delay in the MeNB buffer and the SeNB buffer, as well as the backhaul channel delay. Thereby, the MeNB takes the forwarding decision based on the estimation of the packet delays. In order to accurately take the forwarding decisions, this method comprises the estimation of the SeNB queuing delay at the time when the desired packet would arrive at the SeNB buffer, i.e. at $t+T_{DL}$, wherein T_{DL} is the backhaul downlink delay. The forwarding decision is defined as follows:

$$\text{Forward packet via} \begin{cases} MeNB, & \text{if } D_M(t) \leqslant D_S(t+T_{DL})+T_{DL} \\ SeNB, & \text{if } D_M(t) > D_S(t+T_{DL})+T_{DL}. \end{cases} \quad (1)$$

On the one hand, $D_M(t)$ is the MeNB queuing delay (in seconds) of the desired packet at the current time t

$$D_M(t) = \frac{q_M^Q(t)}{\mu_M(t)}, \quad (2)$$

where $q_M^Q(t)$ is the current MeNB buffer level (in bits) and $\mu_M(t)$ is the current MeNB throughput (in bit/s).

On the other hand, $D_S(t+T_{DL})$ is the SeNB queuing delay (in seconds) of the desired packet at the time $t+T_{DL}$

$$D_S(t+T_{DL}) = \frac{q_S^Q(t+T_{DL})}{\mu_S(t+T_{DL})}, \quad (3)$$

where $q_S^Q(t+T_{DL})$ is the SeNB buffer level (in bits) and $\mu_S(t+T_{DL})$ is the SeNB throughput (in bit/s) at the time $t+T_{DL}$.

To estimate the SeNB queuing delay, the MeNB requires information about the SeNB buffer state. Therefore, this method comprises that the SeNB sends periodically a feedback report via the backhaul channel to the MeNB. This feedback report includes the current SeNB buffer level $q_S^Q(t)$ (in bits), the current SeNB throughput $\mu_S(t)$ (in bit/s) and the backhaul downlink delay T_{DL} (in seconds).

As the feedback is transmitted over the backhaul channel, it is delayed by the backhaul uplink delay T_{UL}. Hence, the PDCP buffer splitting needs to be delay compensated. For that, the MeNB considers the backhaul delays T_{DL} and T_{UL} to adjust the SeNB feedback information.

2.1 Deterministic Fluid Approximation based on a Recursive Feedback Loop

Complex networks are very demanding to model and analyze with methods of queuing theory [5]. Therefore, other methods have been analyzed and developed for communication network analysis and flow control techniques, e.g. rate-based flow control mechanisms based on fluid approximations [6],[5],[7]. Based on this, we consider a fluid approximation to analyze and determine the queue states involved during the PDCP buffer splitting. The main idea of a fluid approximation is to define a deterministic process for the evolution of a stochastic process [6], i.e. assuming no randomness in the new states of the system. Thereby, we can replace all those discrete incoming and outgoing data packets by continuous deterministic data rates. The SeNB queue state analysis is depicted in Figure 4. In this case, the MeNB buffer splitter needs to estimate the SeNB buffer level in order to take decisions about forwarding packets over the backhaul channel.

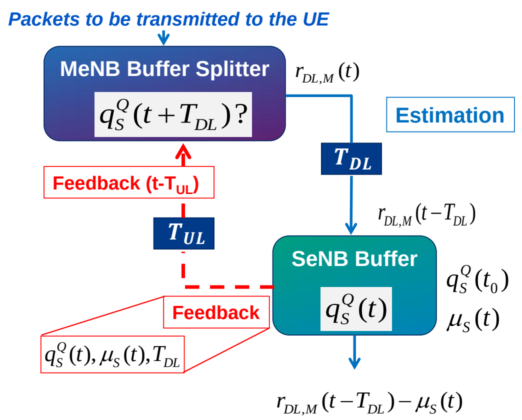

Figure 4 SeNB Queue State Analysis

Based on Figure 4, we define the fluid approximation relying on the downlink data rate $r_{DL,M}(t)$ currently outgoing from the MeNB buffer splitter to the SeNB buffer, the current data rate $r_{DL,M}(t-T_{DL})$ arriving at the SeNB buffer, the initial SeNB buffer level $q_S^Q(t_0)$ and $\mu_S(t)$ as follows:

$$q_S^Q(t) = max\left\{0,\ q_S^Q(t_0) + \int_{t_0}^{t}(r_{DL,M}(\tau - T_{DL}) - \mu_S(\tau))d\tau\right\}. \quad (4)$$

Taking into consideration the feedback report and (4), the MeNB can estimate the current SeNB buffer level by adjusting the received SeNB buffer level $q_S^Q(t-T_{UL})$ based on the incoming packets and the already transmitted packets at the SeNB from $t-T_{UL}$ to t. Similarly, the MeNB can estimate the SeNB buffer level at the desired time $t+T_{DL}$. Moreover, since the throughput $\mu_S(t)$ changes over the time according to the SeNB scheduler, it can be demanding for the MeNB to estimate its behavior. Therefore, a further aspect of this method is to avoid cross-layer operations and to assume $\mu_S(t)$ as constant from $t-T_{UL}$ to $t+T_{DL}$. Thus, the estimation of the SeNB buffer level is defined as follows:

$$\begin{aligned} \hat{q}_S^Q(t) = max\{0,\ q_S^Q(t-T_{UL}) + \int_{t-T_{UL}}^{t} & r_{DL,M}(\tau - T_{DL})d\tau \\ & -\mu_S(t-T_{UL}) * T_{UL}\}, \\ \hat{q}_S^Q(t+T_{DL}) = max\{0,\ \hat{q}_S^Q(t) + \int_{t}^{t+T_{DL}} & r_{DL,M}(\tau - T_{DL})d\tau \\ & -\mu_S(t-T_{UL}) * T_{DL}\}, \end{aligned} \quad (5)$$

where $\mu_S(t-T_{UL})$ is the received SeNB throughput at the time t.

Finally, the MeNB estimates the SeNB queuing delay of the desired packet at the time $t+T_{DL}$ relying on (3) and (5). Likewise, the MeNB can estimate its own queuing delay based on its current buffer level and throughput. This estimation is straightforward taking into account that there is no backhaul channel between the splitter and the buffer. With the estimated queuing delays, the MeNB takes the forwading decision for the desired packet based on (1).

3 Simulation Assumptions

The main objective of this work is to evaluate the potential end-user benefits of LTE networks with the Dual Connectivity (DC) feature. The LTE DC performance is compared with the LTE single connectivity performance, i.e. for users connected only to either a macro cell or a pico cell.

The performance evaluation has been done with an LTE system simulator which explicitly models each protocol layer involved in the data transfer between the application server and the UEs through LTE. This simulator comprises a realistic heterogeneous network deployment for a dense urban scenario in an Asian city inspired by downtown Tokyo and Seoul. The network deployment is well-tuned (related to coverage and pathloss geometry) with real measurements from the deployed network in these cities.

The center area of this Asian city ($\sim 0.5km^2$) is covered by 10 macro sites (with three macro cells per site) and an inter-site distance (ISD) of 200m. Within the center area, 30 pico nodes (3 pico/macro site) are deployed. The macro base station transmit power reaches 46 dBm, whereas the pico nodes transmit at 30 dBm. The network deployment is an inter-frequency scenario with two 10 MHz carriers: macro cells operating at a carrier frequency of 900 MHz and pico cells operating at a carrier frequency of 2 GHz. The users are generated in the simulation area according to a Poisson process with an indoor user probability of 90%.

DC is activated only between cells operating on different carrier frequencies in order to avoid interference, according to [2]. The backhaul channel between the macro and pico nodes is defined with a typical throughput of 100 $Mbit/s$ [3]. In order to analyze the backhaul channel delay impact, DC was evaluated for different delays: 0ms (ideal backhaul channel), 10ms and 30ms. The PDCP buffer splitting is based on the delay compensated PDCP buffer splitting method described in Section 2.

For the evaluation, we initially generate the traffic with the widely used File Transfer Protocol (FTP) traffic model (16 MB file size transmissions). Subsequently, we also consider a web traffic model, which was developed within this work, to further analyze the LTE DC performance with a very bursty and latency dependent traffic (several small file transmissions). The web traffic model is based on traffic measured from the current Top 25 most popular web sites [8]. This model consists of a web client and a web server supporting all the Hypertext Transfer Protocol (HTTP) versions within the Top 25, i.e. the HTTP versions (HTTP/1.1 [9], HTTPS [10] and HTTP/2 [11]) are explicitly modeled. The FTP users transfer a single file, whereas the web users load a web page of the Top 25. Once the respective transfer or download has finished, the user disappears from the LTE system. For the simulations, mobility is not considered for the users.

4 Simulation Results

The average of the user observed bit rate achieved for 16 MB FTP file transfers and different traffic loads is shown in Figure 5. The traffic load is defined based on the average of the user data volume per month: **low load** between 0 to 4 GB/month/user, **medium load** between 4 to 9 GB/month/user and **high load** above 9 GB/month/user.

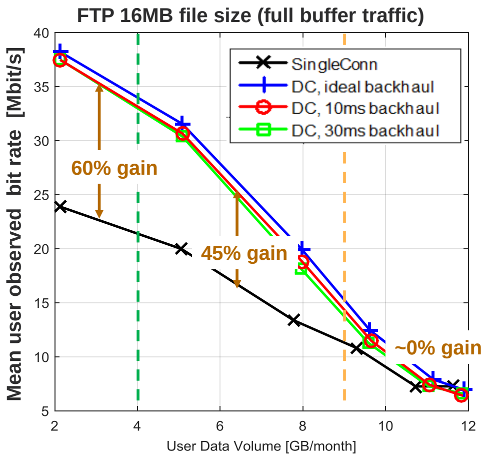

Figure 5 Average user observed bit rate and gains for 16 MB FTP file transfers

As can be observed, at low and medium loads DC has an enhanced performance compared to single connectivity due to the traffic offloaded via the aggregated resources, i.e. the more available resources per UE, the higher the throughput. Therefore, the buffer becomes empty faster. It also shows that DC does increase the user throughput even though a non-ideal backhaul channel is implemented between the eNBs. All the DC results show gains above 50% for low loads and above 30% for medium loads. The impact of the backhaul channel delay is not significant since there are many packets to be potentially buffered. When the system is highly loaded, single connectivity as well as DC become worse due to limited available resources, i.e. low throughputs.

On the other hand, the average of the user observed bit rate achieved for downloading the Top 25 web sites and for different traffic loads is shown in Figure 6. The results show also an enhanced user experience by the aggregation of a secondary-cell. Nevertheless, the gains of DC are smaller (compared to the obtained results with large file transmissions). It means that when the packets arrive to the eNB in a bursty manner, the aggregation of resources via a secondary-cell (i.e. the SeNB) is less beneficial. In other words, less packets need to be buffered at the MeNB for bursty traffic and therefore the MeNB buffer queuing delay becomes lower than the delay of forwarding the packets to the UE via the backhaul channel and the SeNB.

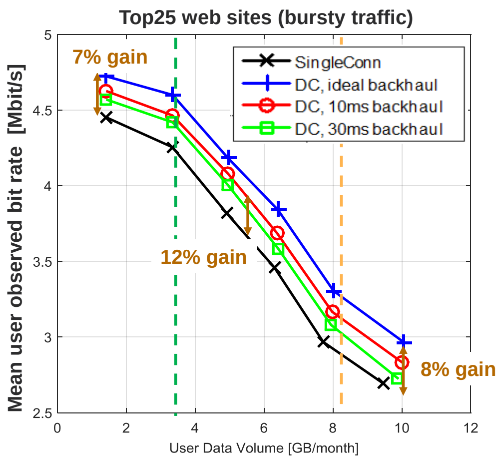

Figure 6 Average user observed bit rate and gains for downloading randomly web sites from the Top 25

Nevertheless, analyzing further potential end-user benefits of DC, the average of the user observed bit rate achieved by the cell edge users downloading the same web page (*www.youtube.com*) is depicted in Figure 7. The results show that the end-user experience is significantly improved, even for a backhaul channel delay of 30ms. DC thus provides very high gains for cell edge user,s especially at medium loads (a peak gain of more than 80% compared to single connectivity). The reason for the highest gains at medium loads is because the system has less available resources and therefore the throughput for those UEs become lower, i.e. more buffering may occur and thus with DC more packets can be potentially offloaded via the SeNB. This means that the system can aggregate throughput to those cell edge users.

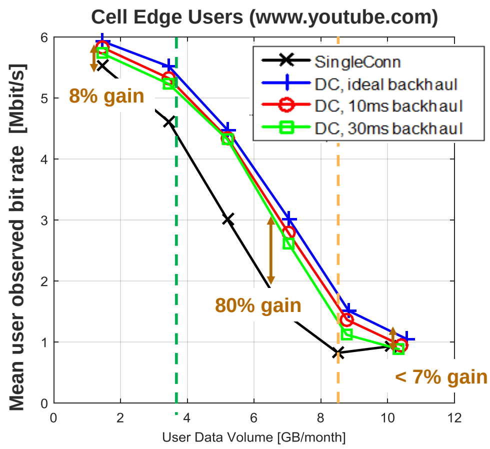

Figure 7 Cell Edge Users - Average user observed bit rate and gains for downloading *www.youtube.com*

5 Conclusion

This paper has evaluated and analyzed the potential end-user benefits of LTE DC in a realistic heterogeneous network deployment considering two types of traffic: full-buffer traffic, e.g. large file transfers, and bursty traffic, e.g. small file transfers as in webpage downloads. For the evaluation, a new method has been developed and considered within this work to efficiently route packets to be transmitted to the UE at the PDCP layer via either the MeNB or via the backhaul channel to the SeNB. Furthermore, the performance evaluation has been done using the FTP traffic model and a web traffic model which has been developed based on traffic measured from the current Top 25 most popular web sites.

The results have shown that the end user performance improves by offloading loaded eNBs buffers with LTE DC. For large file transmissions, DC provides high gains for the end-users since there are always packets in the buffer and therefore full potential of offloading traffic via the SeNB exists, i.e. the users can significantly increase the throughput by the resources aggregated from two eNBs. At low and medium loads, the DC system gains are in the order of 50% on average compared to single connectivity. On the other hand, the results have shown that for bursty traffic like web browsing, buffering in the network nodes occurs only at medium to high system loads and specially for the cell edge users. There, DC is able to provide higher user achievable throughput and therefore the buffers can be offloaded and emptied faster. The results have indicated that those users with the worst quality within the network gain more than 80% in throughput, which means e.g. that a website is downloaded in half the time with DC.

References

[1] Ericsson, “Ericsson Mobility Report, On the pulse of the Networked Society,” 2015. [Online]. Available: http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf

[2] 3GPP, “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description (Release 12),” 3rd Generation Partnership Project (3GPP), TS 36.300, Jun. 2015.

[3] 3GPP, “Technical Specification Group Radio Access Network; Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN (Release 12),” 3rd Generation Partnership Project (3GPP), TR 36.932, Mar. 2013.

[4] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification,” 3rd Generation Partnership Project (3GPP), TS 36.323, Feb. 2015.

[5] K. Grochla and T. Czachórski, “Fluid flow approximation model of TCP connection on wimax link with ugs scheduler,” in *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on.* IEEE, 2009, pp. 476–479.

[6] J.-C. Bolot *et al.*, “Analysis of a fluid approximation to flow control dynamics,” in *INFOCOM’92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE.* IEEE, 1992, pp. 2398–2407.

[7] D. Armbruster, E. Lefeber, and J. Rooda, “On the stability of queueing networks and fluid models,” 2010.

[8] “The top 500 sites on the web.” [Online]. Available: http://www.alexa.com/topsites

[9] R. Fielding, J. Gettys, J. C. Mogul, H. Frystik, L. Masinter, P. Leach, and T. Berbers-Lee, “Hypertext Transfer Protocol – HTTP/1.1,” HTTP Network Working Group, RFC 2616, Jun. 1999.

[10] E. Rescorla, “HTTP over TLS,” Internet Engineering Task Force (IETF), RFC 2818, May 2000.

[11] M. Belshe, R. Peon, and M. Thomson, “Hypertext Transfer Protocol Version 2 (HTTP/2),” Internet Engineering Task Force (IETF), RFC 7540, May 2015.