Impact assessment of Dual Connectivity for Multicast Transmissions in 5G New Radio

Sadaf ul Zuhra*, Prasanna Chaporkar* and Abhay Karandikar*†
*Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
Email: {sadafshah,chaporkar,karandi}@ee.iitb.ac.in
†Director and Professor, Indian Institute of Technology Kanpur, Kanpur, India
Email: karandi@iitk.ac.in

Abstract—In this paper, we have investigated the use of multicast transmission with Dual Connectivity (DC) for the fifth generation (5G) wireless networks. DC has emerged as an important enabler for meeting the requirements of 5G New Radio. However, its potential for improving the performance of multicast transmission has not been explored in the existing literature. Using DC alongside multicast can help provide high quality video streaming services to users in a lesser number of resources. This will especially benefit the cell edge users who usually do not experience very good channel conditions. Allowing users to receive content from multiple base stations leads to a significant gain in the quality of videos received by them. We assess the impact of DC on the performance of multicast streaming services. Even with simple decentralized resource allocation schemes, DC provides a significant performance improvement in terms of the system throughput, fairness and the user quality of experience.

Index Terms—Multicast, 5G New Radio, Video streaming, IMT 2020

I. INTRODUCTION

The Fifth Generation (5G) of communication networks aims at catering to an extremely dense network of User Terminals (UTs) and providing them high quality video streaming services at high speeds [1]. With nearly 80% of the data traffic comprising of videos [2], we need better ways of catering to video streaming services. Multicast transmission provides an excellent means of serving a large number of users in a limited bandwidth [3], [4]. A large number of end-users requesting the same content can be served on the same resources over a single multicast stream. Multicast can, therefore help meet the rate and service requirements of the 5G networks in an efficient manner.

Dual Connectivity (DC) is another technology that has grown to form an integral part of the next generation communication systems. DC provides the networks with the flexibility of delivering content to the end users from multiple sources and over multiple Radio Access Technologies (RATs). DC is expected to be a key enabler in 5G wireless networks [5]. The high data rate, ultra reliable low latency and high mobility requirements of 5G necessitate the reduction of radio link failures due to mobility. Use of DC makes it possible to avoid such failures and ensure seamless connectivity for mobile users [6]. A high level description of various kinds of DC for 5G has been given in [5]. Various options of simultaneously connecting users to multiple technologies in 5G, the associated benefits and architecture requirements have been discussed in [7].

Using DC along with multicast provides an interesting combination that has the potential to significantly ease the burden on our networks. Even though DC has received considerable attention from the research community in the past few years for throughput and handover improvement [8]–[11], it has never been considered to work with multicast transmissions. The only context in which delivery of multicast content from multiple sources has been considered is in the use of Multimedia Broadcast Multicast Single Frequency Networks (MBSFN) wherein multiple Base Stations (BSs) in an MBSFN area transmit the same content in synchronization [12]. It, however, requires strict synchronization between all the BSs in the area and an extended Cyclic Prefix (CP) so that the edge users can successfully combine the content from multiple BSs. The extended CP results in a loss of throughput and tight synchronization requires a lot of control signaling between the BSs. Also, MBSFN transmissions can only be used for transmitting a single stream at a time over the entire MBSFN area [12]. This greatly restricts the scope of MBSFN transmissions.

In this paper, we introduce the use of DC for multicast transmission. UTs are capable of DC and can connect to multiple gNodeBs (gNBs) at a time. Multiple multicast sessions could be going on in the cells simultaneously. The UTs can subscribe to the same multicast stream from multiple gNBs and receive the content from both the gNBs it is connected to. There is no need for the gNBs to use the same resource blocks for transmitting a particular multicast stream and no need for an extended CP. The proposed DC multicast can be implemented in the existing framework of networks with very minimal additional control overhead.

The rest of this paper is organized as follows. We discuss the concept of DC multicast being introduced in this paper in Section II. The system model used is explained in Section III. We elaborate on the resource allocation problem for DC multicast and explain the proposed heuristics in Sections IV and V respectively. We have verified the feasibility of DC multicast using extensive simulations in Section VI. Finally, Section VII concludes the paper.
II. Dual Connectivity Multicast

DC refers to the capability of a UT to connect to and receive content from multiple BSs at a time. In Long Term Evolution (LTE), DC typically refers to a user being connected to a primary macro BS and a secondary micro or femto BS [8]. In 5G, DC refers to a UT being connected to a primary LTE eNB and a secondary gNB or vice versa [13]. Various terms are used for these different forms of DC depending upon which BS acts as the primary/master BS and which acts as the secondary BS e.g. DC with an Evolved Packet Core (EPC) and an LTE eNB as the master node is referred to as Evolved Universal Terrestrial Radio Access (E-UTRA) NR DC (EN-DC) [13]. In LTE as well as 5G DC, the content being delivered to the UT is split between the two BSs. In this configuration, the primary BS acts as the master node through which all control signaling is conveyed to and from the UT [13]. The secondary BS is controlled by the primary BS for transmitting content to the dual connected UT.

While DC in the traditional form discussed has received considerable attention and has been extensively researched, the use of DC for multicast transmission has not been considered in the current literature. In this paper, we introduce the use of DC for multicast streaming for the very first time. The use of DC multicast introduced here involves a different dynamic between the primary and secondary BSs. Instead of being connected to a macro and a micro BS or to an LTE eNB and a gNB, the UT is connected to two neighboring macro BSs and receives the multicast streaming content from both. The two BSs are not even required to be streaming the multicast content over the same Physical Resource Blocks (PRBs). Each BS allocates PRBs to various multicast streams independently. The streaming content that the BSs receive comes from the same Broadcast Multicast Service Centre (BM-SC) and no additional synchronization is needed between the primary and secondary gNBs. The details of the functioning of such a DC enabled multicast streaming system are discussed in the next section.

Receiving the streaming content from an additional gNB significantly improves the quality of experience for the cell edge UTs which otherwise experience a higher packet loss. Through extensive simulations, we show that even with no coordination between the neighboring gNBs, using DC in the proposed form provides significant performance gains for multicast streaming. Such an uncoordinated streaming requires a very minimal additional control overhead. We also explore the use of centralized resource allocation mechanisms that can be implemented using a centralized SDN controller. Centralized resource allocation provides even greater performance gains because the SDN controller can optimize over the entire multicast region. In the next section, we discuss the system model used in this paper.

III. System Model

We consider a seven cell 5G system. Each cell has a gNB located at the center. There are $M$ multicast users in the system that are uniformly distributed through the cells. Every UT is capable of DC. There are $L$ multicast sessions available for streaming in all the cells. Let $[n] = \{1, \ldots, n\}$ and let $|A|$ denote the cardinality of a set $A$. Thus, $[L]$ denotes the set of multicast streams. Each multicast UT is subscribed to one of the $L$ sessions at a time. Once the session starts, the UTs are notified so they can begin receiving the streaming content. The UTs subscribed to a multicast stream in a cell form a single multicast group and receive the streaming content over the same PRBs. We denote the group corresponding to stream $i$ as $G_i$. The group to which UT $k$ belongs is denoted by $i(k)$.

In our model, the cell edge users are connected to a gNB from one of the neighboring cells in addition to their parent gNBs. We refer to this second gNB as the secondary gNB of this UT and the corresponding cell as the secondary cell of the UT. This UT can receive the multicast content from both these gNBs. We denote the primary cell of a UT $k$ as $c(k)$ and it’s secondary cell as $c'(k)$. For a UT that is not dual connected, $c'(k) = 0$. A cell edge UT, therefore, belongs to two multicast groups streaming the same content, one in it’s parent cell and the other in it’s secondary cell. Every multicast stream in a cell is allocated one PRB in every sub-frame $t$. The resource allocation to the various multicast streams can either be done by every gNB independently or by a central SDN controller that manages the gNBs in a region. The multicast data stream in the primary and secondary cells of a UT may or may not be scheduled on the same PRB.

Every multicast stream has a certain rate at which the content needs to be streamed to the subscribed UTs. We denote this rate for $G_i$ as $R_i$. The content of stream $i$ is streamed at this rate $R_i$ whenever the multicast service is active. The channel states of the UTs vary across time and frequency. As a result, a UT experiences a different channel in different sub-frames and also across different PRBs in a sub-frame. Depending on the channel state of a UT, there’s a certain maximum rate it can successfully decode in a PRB. We denote this rate as $r_{kj}[t]$ for UT $k$ in PRB $j$ of it’s primary cell in sub-frame $t$. We use $r_{kj}[t]$ to denote the corresponding rate for $k$ in it’s secondary cell. If data is transmitted at rates greater than the maximum rate decodable by a UT, the data is not delivered successfully and the UT remains unserved in that sub-frame. For instance, if $G_i$ is allocated PRB $j$ in the primary cell of a dual connected UT $k \in G_i$ and PRB $j'$ in it’s secondary cell, then, $k$ successfully receives data in sub-frame $t$ if it can decode the content from either of the two gNBs i.e., if $r_{kj}[t] \geq R_i$ or $r_{kj'}[t] \geq R_i$. A UT $k$ that is not dual connected would successfully receive data only if $r_{kj}[t] \geq R_i$. We now discuss and formally define the resource allocation problem for DC multicast.

IV. Resource Allocation in DC Multicast

The problem of resource allocation in a DC multicast system is aimed at serving as many UTs successfully in a sub-frame as possible. Since the cell edge UTs are connected to two different gNBs, they can receive the streaming content from either of the gNBs. As a result, the performance of a UT depends on the PRB allocated to it’s group in the secondary
cell as well as in its own primary cell. The optimal resource allocation for a region must, therefore, optimize over all the cells in that region. Optimal allocation of resources on a per cell basis is no longer the globally optimal allocation. Let us now mathematically define the optimal resource allocation problem.

Let $C$ denote the number of cells in the multicast region under consideration. Consider $NC$ sets $P_{jc}[t], j \in \{1, 2\}, c \in \{C\}$, each containing $L$ elements e.g., $P_{jc}[t] = \{p_{jc}^{1}[t], p_{jc}^{2}[t], \ldots, p_{jc}^{L}[t]\}$. The constituent elements of $P_{jc}[t]$ are also sets, $p_{jc}^{i}[t] \in P_{jc}[t]$ is the set of users that would be served if in cell $c$, PRB $j$ were allocated to group $G_i$ in sub-frame $t$. Note that the set $p_{jc}^{i}[t]$ includes users from cell $c$ as well as the users from other cells who might be dual connected to $c$. We define indicator $x_{jc}^{i}[t]$ as follows:

$$x_{jc}^{i}[t] = \begin{cases} 1, & \text{if PRB } j \text{ is allocated to } G_i \text{ in } c \text{ in } t, \\ 0, & \text{otherwise.} \end{cases}$$

Since resource allocation takes place every sub-frame, we fix a sub-frame $t$ and omit the sub-frame index from the notations from this point onwards. Using the defined terms, the optimal resource allocation problem in a sub-frame for the multicast region can be written as follows:

$$(B^*) : \max \left| \bigcup_{j,c,i} p_{jc}^{i} \right| x_{jc}^{i}$$

subject to:

1. $\sum_{i \in L} \sum_{j \in [N]} x_{jc}^{i} = L, \forall c \in [C], \quad (1)$
2. $\sum_{i \in L} x_{jc}^{i} = 1, \forall c \in [C], j \in [N]. \quad (2)$

The product in $(1)$ between $p_{jc}^{i}$ and $x_{jc}^{i}$ indicates that the set $p_{jc}^{i}$ is included in the union iff the corresponding indicator is 1. The objective function in $(1)$ maximizes the number of users who are successfully served under this allocation. The constraint in $(2)$ ensures that every PRB is allocated to every group in every cell and $(3)$ ensures that a PRB in a cell is not allocated to more than one group. The optimization problem $(B^*)$ is an NP-hard problem. Hence, no polynomial time algorithms exist for determining the optimal resource allocation for DC multicast. Therefore, in the next section, we present various efficient heuristics for resource allocation in DC multicast.

V. RESOURCE ALLOCATION ALGORITHMS

Resource allocation in DC multicast systems can be done in either an uncoordinated or a centralized manner. In uncoordinated allocation, every cell allocates resources to the multicast streams independently. In a dual connected system, this type of allocation does not fully reap the benefits of DC. We illustrate this with the following example. Consider a 2 cell system containing cells $c_1$ and $c_2$. There are two PRBs available for allocation in each cell. We denote these as $P_1$ and $P_2$. $c_1$ contains four users, $\{u_1, u_2, u_3, u_4\}$ and $c_2$ has two users $\{u_5, u_6\}$. All the users are subscribed to the same multicast stream. $u_1$ has a good channel only in $P_1$ and can successfully receive content only on $P_1$. Users $u_3, u_4, u_5$, and $u_6$ have a good channel only in $P_2$ and can, therefore, successfully receive content only on $P_2$. $u_2$ has a good channel in both the PRBs and would be happy with either of them. Users $u_1, u_3, u_4$ are dual connected to both the cells and can receive content from either of them.

Let us now look at the allocations that will be done by an uncoordinated policy that is maximizing the number of users served in every cell individually. $c_1$ will look at the users in it and allot $P_2$ to the stream because it serves the maximum number of users ($u_2, u_3, u_4$). Now $c_2$ optimizes independently and also allocates $P_2$ to the stream and all its UTs are served successfully. Under this allocation, $u_1$ remains unserved even though it was dual connected, since it could only receive the content over $P_1$. On the other hand, $u_3$ and $u_4$ received content from both the cells.

Let us now see how an iterative centralized policy solves this problem. In this policy, cells also look at the users dual connected to them in addition to their own users while making the allocation decisions. Allocation is done in the cells one by one and a user once satisfied is not taken into consideration while allocating resources in the rest of the cells. Using this policy, we start with $c_2$. $c_2$ now has to look at $u_5, u_6, u_3$, and $u_4$ while allocating resources. So, it chooses $P_2$ and users $u_5, u_6, u_3$, and $u_4$ are successfully served. Moving to $c_1$, users $u_3, u_4$ are already satisfied and therefore, are not taken into consideration by $c_1$. $c_1$ now looks at $u_1$ and $u_2$ and allocates $P_1$ to the stream, satisfying both the users. Thus, the centralized policy manages to successfully serve all the users in the system.

Lemma 1. Any centralized resource allocation policy always encounters a lower packet loss than an uncoordinated policy.

Lemma 1 claims that any centralized allocation policy, even if it is sub-optimal, will always do better in terms of the number of users successfully served than a policy which allocates resources in an uncoordinated manner. A centralized policy does not necessarily mean that the policy is optimizing over the entire system. Any form of centralization that looks beyond just the individual cell will reap a better performance than a completely uncoordinated allocation.

The above example and lemma establish the need for a centralized policy that can better serve the needs of the multicast users. In this paper, we propose a greedy and a Proportional Fair (PF) algorithm in both an uncoordinated and a centralized manner. These algorithms are discussed in detail below.

A. Uncoordinated Allocation

Uncoordinated allocation algorithms do not optimize over multiple cells. Under these algorithms, each gNB allocates resources to the multicast streams by optimizing over it's own cell only. The fact that some of the users from neighboring cells may be dual connected to this gNB is not taken into consideration in uncoordinated algorithms. We propose the use of the following uncoordinated greedy and PF algorithms for resource allocation in DC multicast:
1) Uncoordinated Greedy Allocation: Uncoordinated Greedy Allocation (UGA) solves the optimization problem $B^*$ for every cell individually. In every sub-frame, a gNB allocates PRBs to the multicast streams such that maximum number of users in it’s cell are served. When optimizing over a single cell, the problem can be solved in polynomial time using a maximum weight bipartite matching. The pseudo code for this algorithm is given in Algorithm 1. The set $q^j_{jc}$ in Algorithm 1 is the set of users of cell $c$ that would be successfully served if PRB $j$ were allocated to $G_i$ in $c$. Note that it is not the same as $p^j_{jc}$ defined in the previous section which also contained the dual connected users from other cells.

Algorithm 1: UG Allocation

Input: Sets $q^j_{jc}$ for all $i \in [L]$, $c \in [C]$ and $j \in [N]$

1 for $c = 1 : C$
2 Initialize: $N = [N]$, $L = [L]$ and $x^j_{jc} = 0$ for every $i, j$
3 while $L \neq \phi$
4 Assign $(i^*, j^*) = \arg \max_{(i,j) \in N \times L} \{ q^j_{jc} \}$
5 $x^j_{jc} \leftarrow 1$, $N \leftarrow N \setminus \{j^*\}$, $L \leftarrow L \setminus \{i^*\}$
6 end
7 end

2) Uncoordinated PF Allocation: The existing literature does not consider the use of PF allocation for multicast transmissions. However, for the constant rate video streaming system considered in this paper, using a PF scheme can ensure that UTs gracefully experience loss according to their channel conditions. Let $r_k$ be the historical average data rate of UT $k$ till the current sub-frame. The Uncoordinated PF Allocation (UPA) maximizes the following metric in a sub-frame:

$$\sum_{i,j} \sum_{k \in q^j_{jc}} \frac{x^j_{jc}}{r_k} \tag{4}$$

UPA can also be implemented in polynomial time using a maximum weight bipartite matching. The pseudo code for this algorithm is given in Algorithm 2.

Algorithm 2: UP Allocation

Input: Sets $q^j_{jc}$ for all $i \in [L]$, $c \in [C]$ and $j \in [N]$

1 for every sub-frame $t$
2 for $c = 1 : C$
3 Initialize: $N = [N]$, $L = [L]$, $x^j_{jc} = 0$ for every $i, j$ and $r_k = 1 \forall k$
4 while $L \neq \phi$
5 Assign $(i^*, j^*) = \arg \max_{(i,j) \in N \times L} \sum_{k \in q^j_{jc}} \frac{1}{r_k}$
6 $x^j_{jc} \leftarrow 1$, $N \leftarrow N \setminus \{j^*\}$, $L \leftarrow L \setminus \{i^*\}$
7 $r_k \leftarrow (r_k \times (t - 1) + R_i(k))/t$ for every $k \in q^j_{jc}$
8 end
9 end

The resource allocation in such iterative allocation mechanisms also depends on the order in which we iterate over the cells in the multicast region. Therefore, we also need to optimize over the order of iterations in these allocation mechanisms. Even for a seven cell system, the possible orders of iteration are huge (7!). Therefore, for the purpose of performance evaluation of these algorithms, we have considered ten different random orders of cells and chosen the best allocation from among them. Even if we make use of a random order to iterate over the cells, the policy will still do better than a completely uncoordinated approach (Lemma 1). We now discuss two such centralized algorithms.

1) Iterative Centralized Greedy Allocation: In Iterative Centralized Greedy Allocation (ICGA), the SDN controller first allocates resources to the multicast streams in one cell by maximizing the number of UTs served in the cell. Unlike UGA, ICGA considers the DC UTs as well while allocating resources in a cell. Once resource allocation for a cell is done, the UTs successfully served by the resource allocation are not considered while allocating resources in the remaining cells. This procedure is repeated for resource allocation in every cell. The resource allocation algorithms proposed in this section.

2) Iterative Centralized PF Allocation: The Iterative Centralized PF Allocation (ICPA) follows a similar procedure as ICGA. However, instead of maximizing the number of UTs successfully served, we use a PF allocation in every iteration. The PF approach here maximizes the quantity $\sum_{i,j} \sum_{k \in q^j_{jc}} \frac{x^j_{jc}}{r_k}$ in every cell. The pseudo code for this algorithm is given in Algorithm 3.

In the next section, we present the simulations carried out for evaluating the performance of DC multicast and it’s resource allocation algorithms proposed in this section.

VI. Simulations

We simulate a seven cell urban macro scenario with an inter site distance of 500 m [14]. UTs in the system are distributed uniformly at random through the cells. Each UT is subscribed to a multicast streaming service. Some other
We also compare their performance with Single Connectivity (SC) multicast to establish the performance gain obtained by using DC.

Figure 1a shows the average percentage of packets successfully received by the UTs under the greedy policy. We observe from the plot that even under UGA policy, use of DC significantly decreases the packet loss rate for the multicast streams and hence improves the quality of the stream received by the end users. This performance gain is achieved without any additional control overhead or need for coordination between the gNBs. Using the centralized allocation enables the central controller to optimize further over multiple gNBs and reduces the packet loss even more.

Figures 1b shows the average number of users left unserved under greedy allocation. Without DC, nearly 14 users are left unserved on an average. Using DC with uncoordinated allocation, the number comes down to 9. The average number of unserved users further decreases to around 8 when the resource allocation is centralized. These results (Figure 1a and 1b) for PF allocation algorithm show a similar trend and so have not been included here.

Figures 2a and 2c show the plots of the average percentage of packets successfully received by the UTs under the greedy and PF allocations respectively as a function of the cell radius. On an average, DC shows 10% greater success rate than SC. As expected, the overall successful reception of packets reduces as the cell sizes increase. However, for increasing cell sizes, the performance gap between SC and DC also shows an increasing trend. Figures 2b shows how the average number of users left unserved in a sub-frame changes under greedy allocation as the cell radius increases. Figure 2d shows the same plot for PF policy. DC succeeds in serving nearly 10% more users on an average. The centralized policies provides a further improvement in performance over the uncoordinated policies. Once again, we see that the performance gap between SC and DC increases as the cell radius increases.

Figures 3a and 3c show the variation of the percentage of packets successfully received by the UTs as a function of increasing group sizes under greedy and PF allocation respectively. Recall that the group size indicates the number of UTs subscribed to the same multicast stream in a cell. Using DC multicast with uncoordinated allocation successfully de-
Multicast transmission provides an efficient technique for catering to the exploding demand for high quality video streaming services. Using DC along with multicast can further increase the efficiency of multicast transmissions. In this paper, we have explored the use of DC multicast for the very first time. We have carried out an impact assessment of DC on the performance of multicast transmission. The optimal resource allocation problem in a DC multicast system is an NP-hard problem. Therefore, we have proposed various uncoordinated and centralized heuristic resource allocation algorithms. Through extensive simulations, we have shown that the use of DC with multicast transmission greatly improves the system performance. Even with the use of sub-optimal uncoordinated resource allocation policies, a dual connected multicast system performs significantly better than a single connected system. Moreover, these uncoordinated policies do not require any additional control overheads. Thus, in this paper, we have established the feasibility of using DC for multicast video streaming. As a future research direction, more efficient resource allocation algorithms can be developed which can further enhance the performance of DC multicast and pave the way for its integration into 5G NR.

REFERENCES


[14] “3GPP TS 38.901 : 5G; Study on channel model for frequencies from 0.5 to 100 GHz, v.15.0.0 Rel. 15,” 2018.