

Max-Min Fair Bandwidth Allocation in Millimeter-Wave Radio Clusters

İdil Zeynep Alemdar

Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
idil.alemdar@metu.edu.tr

Ertan Onur

Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
eronur@metu.edu.tr

Abstract—Enabling ultra high speed wireless communication, Extreme High Frequency (EHF) or Millimeter Wave (mmWave) bands will play a significant role for the 5G. Apart from speed, 5G will be very useful for handling great amounts of data simultaneously and serving bandwidth hungry applications as well. Ultra high quality and ultra fast video streaming will be one of those applications that will be made possible by 5G. While serving bandwidth hungry applications with ease will be an important development and maximizing throughput is most of the time the main goal in a network, it is also important to make sure that no other application starves. In order to prevent such a situation, fair bandwidth allocation should be considered in wireless communications. We simulated a max-min fair bandwidth allocation scenario in a mmWave radio cluster, where a radio cluster is a set of base stations connected to a main hub over 60 GHz radio links. We ran experiments with different path loss exponent values with increasing number of base stations to examine the effects of topology complexity and radio signal loss on the optimization time and on the overall network throughput while maintaining max-min fair allocation. The results showed that as the topology becomes more complex, the problem takes longer to solve. However, the overall network throughput increases. In addition, our model has achieved a decent quantitative fairness level as shown by Jain’s index values, which are always more than 0.5 on a scale of 0 to 1 with respect to the topology complexity and the number users.

Index Terms—5G New Radio, Max-Min Fairness, Bandwidth Allocation

I. INTRODUCTION

The fifth generation of telecommunication technology (5G) is already on air with plenty of more research yet to conduct. Differently from the previous generations, 5G introduces a new paradigm in the wireless communication domain where the digital transformation of virtually every vertical industry is directly affected [1]. The main use cases of 5G can be grouped into three categories: Extreme Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC) [2]. In order to satisfy the requirements of various applications, 5G New Radio (NR) has been introduced [2]. 5G NR is made up of two frequency ranges (FR): FR1, or sub 6 GHz range, and FR2, or millimeter-Wave (mmWave) range [2]. Among the three main use case families, eMBB is mainly enabled by the mmWave communication, as mmWave introduces the usage of higher frequency bands up to 100 GHz which allow data

rates of Gbps more easily than the sub 6 GHz frequencies [3]. Bandwidth hungry applications such as full HD video streaming or AR/VR applications will be facilitated by 5G. While being able to serve bandwidth hungry applications in very high rates is a desirable situation in terms of resource utilization, maximizing throughput and increasing quality of experience (QoE) of end users, there is a risk that these applications could cause starvation of others, which do not require much bandwidth. As a result, some other metrics other than throughput or utilization maximization should also be taken into consideration. At this point, fairness in bandwidth allocation becomes a concept necessary to be addressed.

Fairness can be described as fair sharing or allocation of network resources in the context of wireless networking [4]. Although there are various fairness measures, all of them try to allocate network resources to the tenants of the network according to their needs and to prevent starvation of any of them. In this work, we implemented a mixed integer linear program (MILP) to simulate allocation of max-min fair (MMF) bandwidth to users who want to watch on-demand video streams via their 5G-supported devices. These users are connected to 5G base stations (BS) in a given area, which is also called a radio cluster. A radio cluster is a set of BSs connected to a hub (e.g., a BS with a fiber backhaul) over multi-hop 60 GHz links that are prone to environmental changes and more brutal conditions compared to wired connections. To the best of our knowledge, fair bandwidth allocation in radio clusters is not studied in the literature. Our work is not only the first example which focuses on fair bandwidth allocation in radio clusters, but it also improves a previous MILP formulation of MMF.

Before diving into our problem **Max-Min Fair Radio Cluster (MMFRC)**, we first give the mathematical definition of MMF: A feasible allocation of a resource to n tenants \vec{x} , $|\vec{x}| = n$ is max-min fair if and only if an increase in any of the allocation (within the feasible space) results in a decrease in (within the feasible space) an already smaller one. Formally, given any other feasible allocation \vec{y} , $|\vec{y}| = n$ and $i \in 1, 2, \dots, n$; if $y_i > x_i$ then $\exists i'$ such that $x_{i'} \leq x_i$ and $y_{i'} < x_{i'}$ [4] [5].

It can be stated that MMF in some sense protects smaller allocations against already greater allocations. Hence, it ensures

that each demand in the system has its minimum common share. From that point on, those who can be assigned more of a given network resource will get more according to the available resource capacity without effecting others.

In MMFRC, the network resource we aim to distribute in an MMF fashion is the bandwidth (or flow rate) for video streaming. Hence, the feasible allocation space will be determined by link capacities within the topology, as we cannot allocate more bandwidth to any link than the link capacity allows. At this point, an alternative definition of MMF considering link capacities comes handy while constructing a mathematical optimization problem, since the legacy definition of MMF is not straightforward to model as a linear program constraint. This alternative definition includes the bottleneck link concept [5]. A link e is a bottleneck for a path p if and only if

- 1) the link e is saturated, i.e. $\sum_i A_{e,i} x_i = c_e$ where x_i is the i^{th} element of the allocation vector \vec{x} , $A_{e,i}$ is a binary variable equals to 1 when x_i uses the link e and c_e is the capacity of the link e .
- 2) the flow allocated to path p is greater than or equal to any other flow on the link e .

Then, a feasible allocation is max-min fair if and only if each tenant's path includes at least one bottleneck link [5].

In this work, we implemented a radio cluster as a topology $G(V, E)$ where users are connected to the 5G BSs in this cluster. We allocated bandwidth to those users using the mathematical optimization model we developed. This optimization model ensures max-min fair bandwidth allocation, as well as it is robust and easy to understand and implement. We also showed the correlation between size and complexity of the topology and throughput and time to optimize.

The rest of the paper is organized as follows: In Section II, we will go over the related work on max-min fairness and its applications in wireless networks. Then, in Section III, we will define our problem MMFRC in detail. Section IV will be about conducting experiments and discussion of the results. Finally in Section V and Section VI we will discuss what can be done as future work to improve this study and conclude the paper.

II. RELATED WORK

In order to work on an MMF problem, firstly its mathematical meaning must be understood. Although almost every work in the literature, including this paper, makes the definition of MMF, it is more suitable to start with some tutorial. While Le Boudec makes an easy to understand introduction to the topic and relates it with rate adaptation and congestion control in [5], Nace and Pioro's tutorial [6] stands out as a comprehensive tutorial for those who want to learn mathematical details of MMF in depth and to gain a holistic view on MMF as a mathematical optimization problem.

Although there are various versions of MMF such as weighted MMF, distributed MMF or multiple path MMF [4], MMF is inherently applied as in a centralized fashion with a predetermined single, unsplitable path. This plain version

of MMF was formulated by Tomaszewski in [7] where paths between sources and destinations are assumed to be given. While we were formulating our own optimization problem, we also started from Tomaszewski's work. However we adopted a modified version of it, as we will show later in this paper. On the other hand, many of the existing works on MMF also found paths within the optimization problem. For instance, in [8] Coniglio et al. formulated a bilevel programming problem where the first level optimization problem would find a single path for each demand to maximize the network utility function after the second level optimization problem finds an MMF allocation. Treating MMF not as an objective but as a constraint, Amaldi et al.'s work [9] is another example where the routing is also found by the optimization algorithm itself.

In addition, there are also works where MMF was implemented or approximated in multiple path scenarios. Danna et al. presented a new concept, called Upward Max-Min Fairness (UMMF) [10], and converged to by a distributed algorithm, where each path may be MMF on its own but only one of them is the lexicographically greatest (Global MMF). Allalouf and Shavitt developed one centralized and one distributed multiple path MMF algorithms which find globally and locally optimal MMF allocations respectively in [11].

Implementing a max-min fair sharing scheme in wireless networks can be more challenging compared to the wired networks, because concepts like noise, interference, signal attenuation or even mobile nodes should be taken into consideration. In [12], Huang and Bensaou developed distributed algorithms to achieve max-min fair shares in the networks as the nodes are mobile and they lack the overall knowledge of the topology of an ad hoc network at a given time. Zhou and Maxemchuk developed a novel method, which is also independent of the number of nodes, to model naturally occurring bottlenecks (caused by obstacles etc.) in wireless ad hoc networks [13]. A weighted max-min fair scheduling policy has been introduced by Tassiulas and Sarkar in [14], where weights are dynamically adjustable according to the congestion. Finally, Sridharan and Krishnamachari developed a model for max-min fair rate allocation in wireless sensor networks in [15].

III. MAX-MIN FAIR RADIO CLUSTER (MMFRC)

Given a radio cluster modeled as a graph $G = (V, E)$ where $v \in V$ are the BSs and $e \in E$ are the 60 GHz links between the BSs, our aim is to distribute the bandwidth to the users connected to the BSs in a max-min fair fashion.

In our problem MMFRC, we assumed users want to watch videos on-demand and are connected to 5G BSs. We also assumed these 5G BSs as a *radio cluster*, where there are N BSs in a given area and only one of them, called the *main hub*, is connected to the core network with a high capacity link. All the other BSs are connected to any other BS in the radio cluster, including to the main hub, over 60 GHz radio links. Users may demand all of the videos via the main hub, however some of the BSs may have cached some of the

TABLE I
LIST OF SYMBOLS

Symbol	Unit	Definition	Domain or Value
R_x	mW	received signal power	\mathbb{R}^+
R_xS	mW	receiver sensitivity	-98 dBm
T_x	mW	transmitted signal power	10 dBm [16]
G_{tx}	dBi	transmitter antenna gain	15 dBi [16]
G_{rx}	dBi	receiver antenna gain	15 dBi [16]
$O(d)$	dB/m	oxygen attenuation	0.016 dB/m [16] [17]
c	meter/second (m/s)	speed of light	3×10^8 m/s
f	GHz	central frequency at which the connection operates on	60.48 GHz [18]
B	GHz	bandwidth around the central frequency	2.16 GHz [18]
d_0	meters (m)	reference point distance	1 m
d	meters (m)	distance between two BSs	\mathbb{R}^+
η	-	path loss exponent	2, 2.7, 3 [19]
k	Joules/Kelvin, J/K	Boltzmann constant	1.380649×10^{-23} J/K
T	Kelvin (K°)	temperature of the environment	293 K°(20 C°)
τ	seconds (s)	propagation delay in wireless communications	\mathbb{R}^+
e	-	edge (radio link) in the topology	$e \in E$
p	-	path between a source and destination pair	$p \in P$
δ_{ep}	-	binary matrix denoting whether e is on p	{0, 1} 1 when e is on p, 0 otherwise
c_e	Gbps	capacity of the edge (link) e	\mathbb{R}^+
x_p	Gbps	amount of bandwidth allocated to the path p	\mathbb{R}^+
y_e	Gbps	total capacity usage on the edge e	\mathbb{R}^+
z_e	Gbps	maximal amount of bandwidth among paths passing on the edge e	\mathbb{R}^+
θ_{ep}	-	binary matrix indicating whether the edge e is a bottleneck link for the path p	{0, 1} 0 \rightarrow yes, 1 \rightarrow no

videos. If there is a lesser cost path from the BS which the user is connected to, to the caching BS compared to the main hub, it thus makes more sense to connect to the caching BS instead of the main hub.

The first question to ask is whether or not two BSs are wirelessly connected with each other over a radio link. Existence of a radio link between two BSs depends on the transmission power of the signal, distance between them, signal power lost due to oxygen attenuation and antenna gains both on the transmitting and receiving sides. Using the simple path loss model [19], we compute the received signal strength and compare it to the receiver sensitivity (RxS) of the receiving BS. If the received signal strength exceeds the sensitivity, we conclude that the received signal strength is strong enough to make two BSs connected. We used -98 dBm for RxS. Summing up the parameters above, we get

$$R_x = T_x G_{tx} G_{rx} \left(\frac{c}{4\pi f} \right)^2 \left(\frac{d_0}{d} \right)^\eta \left(\frac{1}{O(d)} \right), \quad (1)$$

where R_x (mW) is the received signal power, T_x (mW) is the transmitted signal power, G_{tx} (dBi) and G_{rx} (dBi) are the transmitter and receiver antenna gains, respectively, c is the speed of light, f (60.48 GHz) is the center frequency at which the connection operates on, d_0 (1 m) is the reference point distance, d (m) is the distance between two BSs, $O(d)$ (dB) is the oxygen attenuation per meter (using the value Kim et al. used in [16]) and finally η is the path loss exponent (see Table I for definitions, units and domains of the variables or constants.) Attenuation due to other whether conditions like rain, snow etc. is omitted for simplicity.

In order to distribute bandwidth to users in an MMF way as well as to compute the least cost path between a source and destination pair, we need to know capacity of links and delay between two ends of a link, respectively. This is because link capacities set an upper bound to the flows which is important both for feasibility and fairness. On the other hand, delay values are assigned as edge weights while calculating the least cost path between a demanding and a caching BS (or the main hub). Shannon's Capacity equation [19],

$$C = B \log_2 \left(1 + \frac{R_x}{kTB} \right) \quad (2)$$

is used to compute the link capacities and the propagation delay is

$$\tau = \frac{d}{c}. \quad (3)$$

B is the bandwidth around the central frequency (2.16 GHz), kTB is the thermal noise and $\frac{R_x}{kTB}$ is signal-to-noise ratio (SNR), k is the Boltzmann constant and T is the temperature of the environment (293 K°) as shown in Table I. Once hop-by-hop delays are computed, least cost paths between each pair of BSs can be calculated using Dijkstra's shortest path algorithm and when the paths are known, MMFRC can be

formulated into a MILP as follows:

$$\begin{aligned}
& \underset{x_p}{\text{maximize}} && \Sigma_p x_p \\
& \text{subject to} && \\
& \Sigma_p \delta_{ep} x_p = y_e, \forall e \in E && (1) \\
& y_e \leq c_e, \forall e \in E && (2) \\
& c_e \theta_{ep} \geq c_e - y_e, \forall e \in E, \forall p \in P && (3) \\
& \Sigma_e \delta_{ep} (1 - \theta_{ep}) \geq 1, \forall p \in P && (4) \\
& x_p \delta_{ep} \leq z_e, \forall e \in E, \forall p \in P && (5) \\
& c_e \theta_{ep} \geq \delta_{ep} (z_e - x_p), \forall e \in E, \forall p \in P && (6) \\
& x_p \geq 0, \forall p \in P && (7) \\
& z_e \geq 0, \forall e \in E && (8) \\
& z_e \leq y_e, \forall e \in E && (9)
\end{aligned}$$

Given the topology $G(V, E)$, the space size of this MILP problem is $O(|V| \times |E|)$.

MMFRC formulation above was inspired from [7] and we used a modified version of [7]'s formulation where (1) means that all flows using edge e must sum up to the total flow allocated to link e , (2) satisfies that the total flow allocated to link e must not exceed the capacity of that link, (3) is the saturation constraint of the bottleneck link. This constraint means that if a link e is a bottleneck link for the path p ($\theta_{ep} = 0$), then also combined with the constraint above, $c_e = y_e$. If not, $\theta_{ep} = 1$, $c_e \geq c_e - y_e$ which is trivial, (4) is the constraint for the existence of at least one bottleneck link for each path, (5) means that if some flow uses edge e , then its value cannot exceed the maximal flow on e , (6) is the maximal flow constraint of the bottleneck link, that is, the x_p must be at least as much as any other flow sharing that link if e is a bottleneck link for p . (7) and (8) puts 0 as the lower bound to x_p and z_e respectively, as flow values must be nonnegative. Finally, (9) means that maximal flow on link e cannot exceed the total flow.

What we did differently from Tomaszewski [7] are the following:

- Instead of (4) above, he used $\Sigma_p \delta_{ep} (1 - \theta_{ep}) \geq 1$, which would actually mean that there should exist one bottleneck link on the overall topology but not for each path.
- (5) was originally $x_p \theta_{ep} \leq z_e$ in Tomaszewski's work. The reason why we wanted to modify it to (5) was because $\theta_{ep} = 1$ when e is not a bottleneck link for p . This includes the situations where p does not even go over e . In that case, it does not make sense to restrict x_p by some maximal flow value of an edge which does not belong to p .
- (6) was originally $c_e \theta_{ep} \geq z_e - x_p$ in Tomaszewski's work as well. We also modified this formula for the same reason above. Instead, we multiplied the right hand side of the inequality in order to make sure that this constraint would be applicable only if e is on p .
- (7) was not in the Tomaszewski's list of constraints. We needed to add it to give a proper upper bound to z_e .

IV. MMFRC SIMULATIONS & RESULTS

In this section, we present the simulation campaign.

A. BS Generation and Role Assignment

We modeled our MMFRC scenario as follows: We assumed a 1 km² (1000 m × 1000 m) urban area to deploy our radio cluster. There are n 5G BSs whose coordinates are uniform randomly generated in this area, where $n = 10, 15, \dots, 50$. Among these n BSs, the BS with cell id 0 is the main hub (in the middle in Fig. 1), while the remaining $n - 1$ BSs can either cache videos (called *caching BS*, the ones with a video icon on top in Fig. 1), or serve to the users who demand to watch a video (called *demanding BS*, the ones with users as U_1, U_2, \dots are connected to in Fig. 1) or act as a gateway between other BSs and hence provide the connectivity of the topology. These

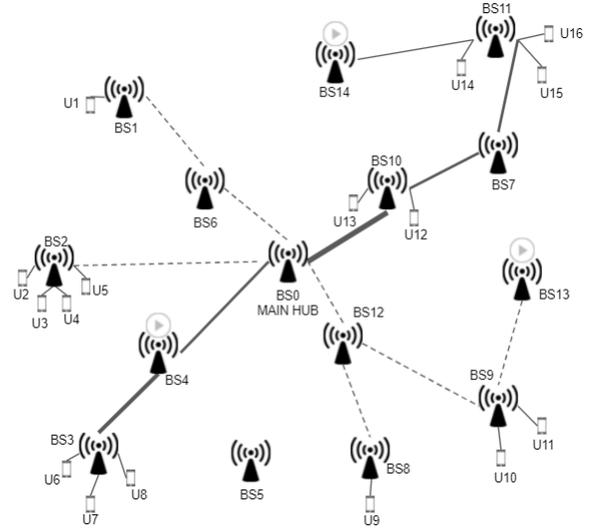


Fig. 1. Example topology with $n = 15$

three roles are assigned to the BSs uniform randomly, as we explain in sequel. There are 10 different videos that can be watched by the users in total. Except for the main hub, which is actually caching all 10 of the videos and hence a caching BS itself, during the generation process all of the BSs are assigned 3 videos to cache with a probability of $[0, 0.2]$. If a BS is chosen to be a caching BS, videos to cache are also determined uniform randomly. If, on the other hand, a BS is not a caching BS, we then check whether it is a demanding BS. With a probability of $(0.2, 0.7]$, a BS is assigned to be a demanding BS. The number of users connected to it, which is in the range $[1, 10]$ and the videos they want to watch are again assigned uniform randomly, where each user is assumed to demand a distinct video. Finally, a BS is not assigned any role with a probability of $(0.7, 1)$. Other than their roles (main hub, caching BS, demanding BS or plain BS), these BSs are assumed to be equipped with identical radio transceivers in terms of transmit power ($T_x = 10$ dBm), receiver sensitivity ($R_xS = -98$ dBm) and antenna gains (15 dBi for both transmitter gain (G_{tx}) and receiver gain (G_{rx})).

B. Topology Management

So far what we have is disconnected BSs scattered upon the area. In order to compute the radio links and capacity of them, we first used 1 to compute the received signal R_x to infer whether two BSs are connected with each other or not. In order to have them connected, R_x must be at least as much as -98 dBm. Here we assume an undirected graph where the radio link between two BSs is bidirectional. We gave three different values to the path loss exponent η . First value is 2, as it is the free space loss exponent [19]. Considering our scenario takes place in an urban area, in order to get more realistic results, we also set η to 2.7 and 3, respectively, where these are two different values η can have in an urban area [19]. Then with the help of 2, we calculated the link capacity and finally 3 gave us the delay between two BSs, where these delay values are assigned as edge weights to the corresponding links. Since we employ highly directional 60 GHz links, we implicitly assume there is no interference from other sources. Therefore, we are done with the generation of the topology $G = (V, E)$ where V is the set of nodes (BSs) and E is the set of weighted edges (radio links with capacity c_e and weight τ).

C. Calculating Paths and Optimizing

The least cost path from each user to the main hub is computed using Dijkstra's Shortest Path algorithm. Additionally, if some caching BS is caching the demanded video, a least cost path from the source BS to that caching BS is also computed and this way, least cost paths from the demanding BS to any caching BS who has the demanded video are computed. The costs of those least cost path candidates are compared and among them, the minimum one is chosen to be the actual simple, unsplittable, single flow path. Dashed lines in Fig. 1 denote least cost paths of demands. Some of them are connected to the main hub, while some others benefit from the caching BSs as they are closer. Straight lines are also same, however their thickness indicate the traffic density in the links. For instance, U_{14} 's demand is cached in BS_{14} and hence U_{14} is connected to BS_{14} . On the other hand, U_{15} 's and U_{16} 's demands are forwarded to the main hub and finally the link between the hub and BS_{10} becomes more dense after U_{12} 's and U_{13} 's demands are added. After paths are computed, the optimizer is made ready to run by defining optimization constants, variables, constraints and objective function.

D. Running Experiments

With all other parameters kept constant, we ran 500 experiments each for a given (n, η) combination where $n \in \{10, 15, \dots, 50\}$ and $\eta \in \{2, 2.7, 3\}$ as mentioned in Section IV-A and IV-B respectively. The reason we chose the mentioned values for the path loss exponent is to compare free space loss $\eta = 2$ with urban area losses $\eta = 2.7$ and $\eta = 3$ [19]. We observed how the number of BSs, path loss exponent as well as link density, which is defined as the number of existing links in a graph divided by the number of links in a complete graph, affected the average total throughput and the time optimizer had to spend to solve a problem. The machine

the experiments ran on is Ubuntu 20.04 LTS with 8 GB RAM and Intel Core i5-4200H processor. The software is written in Python3. Python's networkx library is used to model graphs we generated as well as to calculate paths and Gurobi 9.1.2 Python API was used as the optimizer. All of the experiment results are plotted with a 95% confidence interval.

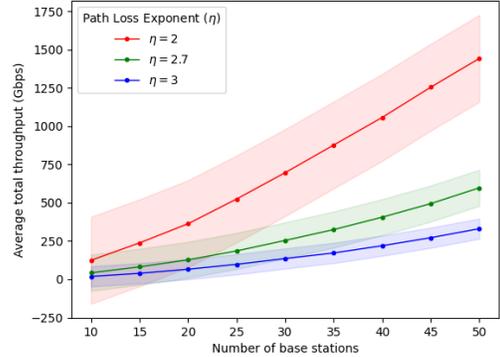


Fig. 2. The average total throughput vs the number of BSs

E. Results

As the path loss exponent becomes larger the received signal becomes smaller and at the same time, link capacity also becomes smaller according to the Shannon's capacity equation in (2). It is expected that as the average link capacities shrink, total average throughput of the network also becomes smaller as can be seen in Figure 2. The smallest path loss exponent, $\eta = 2$ results in an obvious difference in average total throughput compared to $\eta = 2.7$ and $\eta = 3$. This is because of the exponential decrease in capacity occurring due to increase in the path loss exponent.

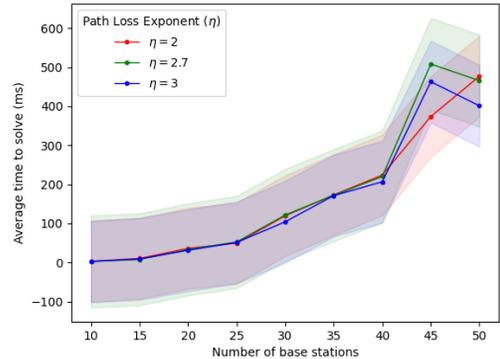


Fig. 3. The time to solve the problem vs the number of BSs

On the other hand, time to solve the optimization problem is not affected much by the value of the path loss exponent. As the topology becomes larger, time spent on solving the optimization problem grows and this phenomenon can be observed in Figure 3 as well. While the results for $\eta = 2$

follows a relatively stable curve, we can observe that the results of $\eta = 2.7$ and $\eta = 3$ make a peak at $n = 45$ and then drop again at $n = 50$. The reason of this peak may be that due to the higher path loss exponents, two BSs become less likely to be connected with each other and thus less number of radio links are created, where they also have relatively lower capacities compared to the ones created with $\eta = 2$. It is possible that after a certain topology complexity, the optimization time increases due to the more number of potential bottleneck links created by those low capacity links. Having more than one bottleneck link candidates makes the problem harder to solve as the optimizer has to check more cases.

When we inspect the effect of link density on the throughput and the time to solve, we can clearly see that the highest path loss exponent $\eta = 3$ causes significant amount of outlier values. The general order of curves are same in both Figure 2 and Figure 4 as well as Figure 3 and 5, however while the average throughput and average time to solve the problem increases almost linearly with respect to the number of BS, they both increase exponentially with respect to the link density, since link density is rather a ratio than a direct link count. In Figure 4 after a link density of 0.92, both the throughput and time values becomes unstable for $\eta = 3$. Again, this is due to the fact that having more low capacity links which create potentially more bottleneck links.

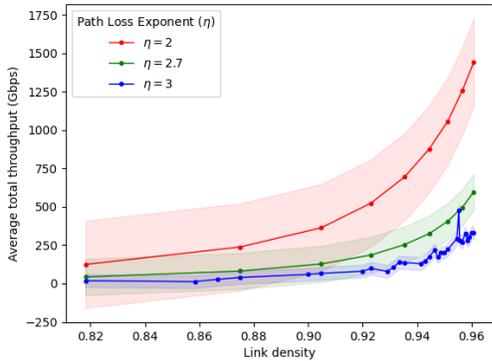


Fig. 4. Average total throughput vs link density

Finally, in order to make sure that our experiments have indeed resulted in a fair bandwidth allocation, we computed Jain's index as it is a quantitative fairness measure having a value between 0 and 1, where a value closer to 1 means a fairer allocation [4]. By observing Figure 6 and Figure 7, we conclude that our allocation has achieved Jain's values no less than 0.5 in any of the cases with a small confidence interval, which can be interpreted as that our bandwidth allocation is actually fair.

V. DISCUSSIONS & FUTURE WORK

Even though it represents realistic results and puts a solid mathematical optimization model for MMF resource allocation, our model is relatively simple and inspects only a small

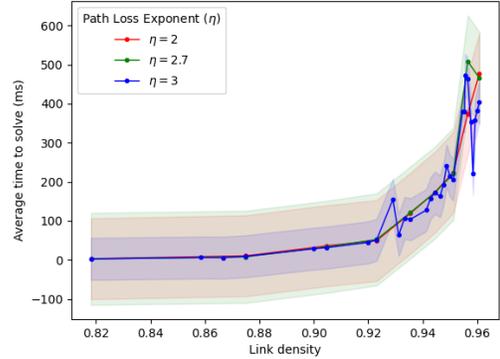


Fig. 5. Time to solve the problem vs link density

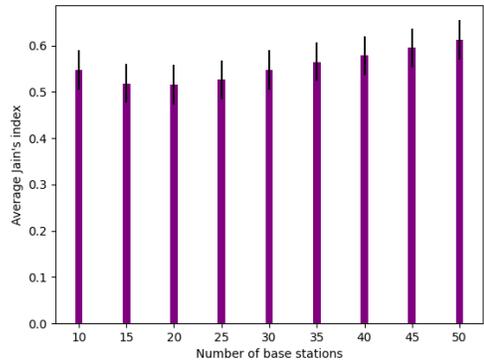


Fig. 6. Jain's index vs BS count

number of parameters' effects. Considering that there are many parameters like transmit power, antenna gains, temperature, effects of these different parameters can result in various scenarios worth paying attention to. Running experiments with different parameters should be the first thing to do as a future work. Simulations including rain attenuation, like in [16] or other weather conditions should also be carried out.

In order to get more realistic results, instead of distributing BSs on a square shaped area in a uniform random manner, more research can be done on how BSs are actually deployed in an urban area. Tree or mesh topologies can be used to inspect how the topology affects throughput and fairness.

Effects of using caches should also be inspected in more detail. Throughput obtained with and without using caching BSs can be compared with each other and hence we can see how caching influences throughput. Instead of assigning random caches and demands, some heuristics can be developed and caching and demanding can be done accordingly.

Considering the communication overhaul within a radio cluster in a real world scenario, a centralized algorithm may be burdensome. Besides, users will be mobile in a real world scenario and even a dynamic topology where new links may emerge or existing ones may disappear, will be possible. These

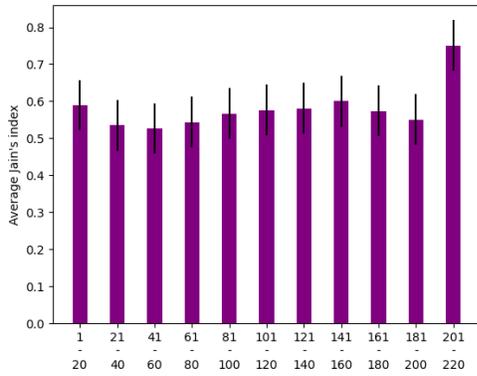


Fig. 7. Jain's Index vs number of users

facts make it even harder to develop a static and centralized algorithm. As a result, a distributed algorithm modeling user mobility and dynamic topology as well as approximating MMF allocation can be developed to get more realistic results and its results can then be compared with the centralized MMF allocation algorithm to see how much the distributed algorithm approximated the centralized algorithm.

It is also important to show that the algorithm truly allocated bandwidth fairly. While fairness in wireless networks is usually a qualitative measure solvable as a mathematical optimization problem, in order to show that the distribution is actually fair, we often need quantitative fairness measures. In this work, we used Jain's Index as a quantitative fairness measure, however other quantitative fairness measures like Entropy [4] should also be computed to get more insight on the fairness level of the bandwidth distribution.

VI. CONCLUSION

In this work we solved single path MMF bandwidth allocation problem where the topology is a mmWave 5G radio cluster and the flow in the network is created by users who want to watch high quality on-demand video streams. We first developed an improved version of an existing MMF MILP formulation and verified its the correctness. Then, we presented it in an easy to understand and easy to implement way. Finally, we implemented it on a 60 GHz radio cluster simulation where real world parameters like different path loss exponents, oxygen attenuation etc. were also taken into account. Our simulations showed that our model indeed distributes bandwidth in an MMF way and our model shows strong parallelism with real world scenarios. As future work, while we will develop a distributed algorithm which will try to allocate the bandwidth in an MMF fashion and observe how close we can get to the centralized algorithm, we will also enhance our simulation with more real world parameters. In addition, mobility of users and dynamic topology models are to be inspected.

ACKNOWLEDGEMENT

This study is supported by Turkcell within the framework of 5G and Beyond Joint Graduate Support Programme coordinated by Information and Communication Technologies Authority.

REFERENCES

- [1] 5G PPP Architecture Working Group, "5G empowering vertical industries." Brochure, 2016.
- [2] X. Lin, J. Li, R. Baldemair, J.-F. T. Cheng, S. Parkvall, D. C. Larsson, H. Koorapaty, M. Frenne, S. Falahati, A. Grovlen, and K. Werner, "5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology," *IEEE Communications Standards Magazine*, vol. 3, no. 3, pp. 30–37, 2019.
- [3] T. Inoue, "5G Standards Progress and Challenges," in *IEEE Radio and Wireless Symposium (RWS)*, pp. 1–4, Jan 2017.
- [4] H. SHI, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in Wireless Networks: Issues, Measures and Challenges," *IEEE Communications Surveys Tutorials*, vol. 16, pp. 5–24, First 2014.
- [5] J.-Y. le Boudec, "Rate adaptation, congestion control and fairness: A tutorial," Tech. Rep. December, 2020.
- [6] D. Nace and M. Pioro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys Tutorials*, vol. 10, pp. 5–17, Fourth 2008.
- [7] A. Tomaszewski, "A polynomial algorithm for solving a general max-min fairness problem," *European Transactions on Telecommunications*, vol. 16, no. 3, pp. 233–240, 2005.
- [8] S. Coniglio, L. G. Gianoli, E. Amaldi, and A. Capone, "Elastic Traffic Engineering Subject to a Fair Bandwidth Allocation via Bilevel Programming," *IEEE/ACM Transactions on Networking*, vol. 28, no. 6, pp. 2407–2420, 2020.
- [9] E. Amaldi, A. Capone, S. Coniglio, and L. G. Gianoli, "Network Optimization Problems Subject to Max-Min Fair Flow Allocation," *IEEE Communications Letters*, vol. 17, no. 7, pp. 1463–1466, 2013.
- [10] E. Danna, S. Mandal, and A. Singh, "A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering," in *2012 Proceedings IEEE INFOCOM*, pp. 846–854, 2012.
- [11] M. Allalouf and Y. Shavitt, "Centralized and Distributed Algorithms for Routing and Weighted Max-Min Fair Bandwidth Allocation," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1015–1024, 2008.
- [12] X. L. Huang and B. Bensaou, "On Max-Min Fairness and Scheduling in Wireless Ad-Hoc Networks: Analytical Framework and Implementation," in *Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking Computing, MobiHoc '01*, (New York, NY, USA), p. 221–231, Association for Computing Machinery, 2001.
- [13] C. Zhou and N. F. Maxemchuk, "Scalable Max-Min Fairness in Wireless Ad Hoc Networks," in *Ad Hoc Networks* (J. Zheng, S. Mao, S. F. Midkiff, and H. Zhu, eds.), (Berlin, Heidelberg), pp. 79–93, Springer Berlin Heidelberg, 2010.
- [14] L. Tassioulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 763–772 vol.2, June 2002.
- [15] A. Sridharan and B. Krishnamachari, "Maximizing Network Utilization with Max-Min Fairness in Wireless Sensor Networks," in *2007 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops*, pp. 1–9, April 2007.
- [16] J. Kim, L. Xian, and A. Sadri, "60 GHz Modular Antenna Array Link Budget Estimation with WiGig Baseband and Millimeter-Wave Specific Attenuation," *International Journal of Antennas and Propagation*, vol. 2017, pp. 1–9, 06 2017.
- [17] S. Singh, R. Mudumbai, and U. Madhoo, "Interference Analysis for Highly Directional 60-GHz Mesh Networks: The Case for Rethinking Medium Access Control," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1513–1527, 2011.
- [18] ITU-R, "Use of multiple gigabit wireless systems in frequencies around 60 GHz," Tech. Rep. M.2227-2, 2017.
- [19] C. Beard and W. Stallings, *Wireless Communication Networks and Systems Global Edition*. Pearson, 2016.