Max-Min Fair Millimetre-Wave Backhauling

Rui Li and Paul Patras

Abstract—5G mobile networks are expected to provide pervasive high speed wireless connectivity, to support increasingly resource intensive user applications. Network hyper-densification therefore becomes necessary, though connecting to the Internet tens of thousands of base stations is non-trivial, especially in urban scenarios where optical fibre is difficult and costly to deploy. The millimetre wave (mm-wave) spectrum is a promising candidate for inexpensive multi-Gbps wireless backhauling, but exploiting this band for effective multi-hop data communications is challenging. In particular, resource allocation and scheduling of very narrow transmission/reception beams requires to overcome terminal deafness and link blockage problems, while managing fairness issues that arise when flows encounter dissimilar competition and traverse different numbers of links with heterogeneous quality. In this paper, we propose WiHAUL, an airtime allocation and scheduling mechanism that overcomes these challenges specific to multi-hop mm-wave networks, guarantees max-min fairness among traffic flows, and ensures the overall available backhaul resources are fully utilised. We evaluate the proposed WiHAUL scheme over a broad range of practical network conditions, and demonstrate up to $5 \times$ individual throughput gains and a fivefold improvement in terms of measurable fairness, over recent mm-wave scheduling solutions.

Index Terms—mm-wave, backhauling, multi-hop, max-min fairness.

1 INTRODUCTION

Market surveys confirm the number of mobile subscriptions and the popularity of bandwidth-intensive applications (including ultra high-definition video and virtual/ augmented reality) continue to grow at an unprecedented pace [2]. In response to the accelerating traffic demands, carriers are offering flat-rate unlimited data plans [3], which requires to substantially extend the capabilities of current mobile network infrastructure. Cell densification is a first step [4], but this entails revisiting existing backhauling practices, to be able to transfer vast volumes of data between the access and core networks. In particular, the cost of deploying traditional, fibre-based backhauls surges with network density, whilst reconfiguration of such solutions is limited. Wireless alternatives have been thus far confined to microwave spectrum (0.3–30GHz) of restricted capacity and already overcrowded with numerous applications, including Wi-Fi, digital video broadcast (DVB), cellular access, RADAR, and M2M communications.

The millimetre-wave (mm-wave) band (30–300GHz) is in contrast underutilised and exposes considerably wider spectral resources that could support an order of magnitude higher data rates [5]. Harnessing their potential is however only possible with electronically steerable highly directional antenna beams, which mitigate characteristic severe signal attenuation. Directionality intrinsically eliminates interference and enables better spatial reuse, though introduces the risk of link blockage due to moving obstacles and terminal deafness, i.e. receivers can hardly be aware of transmitters, unless their beams are mutually aligned [6]. The latter is particularly problematic in deployments with small form factor base stations (such as urban lamppost based infrastructure) that serve large numbers of end-users over Wi-Fi/cellular and communicate with gateways using single mm-wave transceivers, over multiple hops.

We exemplify these issues in Fig. 1, where 6 base stations communicate over mm-wave links with a wired gateway. Specifically, three high volume traffic flows are relayed by intermediary hops from the gateway towards base stations 1, 2, and 5 respectively. In this scenario, station 6 is locked out when attempting to transmit to station 4, if this station has its TX/RX beams steered towards station 5. In addition, the communication between stations 1 and 3 is partially blocked by a moving object, resulting in link quality degradation. Further, the three backlogged traffic aggregates traversing the backhaul in this example are relayed over different number of hops, and encounter different level of competition on heterogeneous links. Therefore, the airtime allocation strategy will impact on the distribution of

![Fig. 1: Mm-wave backhaul with 6 base stations. Three aggregate flows traverse the network in downlink direction (paths shown with continuous lines). TX/RX beams shown with blue/orange shades, possible beams with dashed lines. Link 6-4 subject to terminal deafness, 3-1 partially blocked. Link bit rates labelled.](image-url)

TABLE 1: Rate distribution, total throughput, and unfairness measure with different resource allocation schemes for the topology shown in Fig. 1. Numerical example.

| Scheme               | Flow 1 | Flow 2 | Flow 3 | Total throughput | Gini coefficient |
|----------------------|--------|--------|--------|------------------|-----------------|
| Max-throughput       | 3,378  | 0      | 0      | 3,378           | 0.6667          |
| Round-Robin (equal airtime) | 289    | 1,126  | 770    | 2,185           | 0.2554          |
| Proposed WiHAUL (max-min) | 763    | 763    | 1,504  | 3,030           | 0.1630          |
resources and lead to fairness issues and/or suboptimal network utilisation, unless all these aspects are carefully considered. Indeed, max-throughput strategies favour large volume flows traversing high capacity links, while round-robin schemes that allocate equal airtime are proportionally fair, but lead to wastage of network resources, as shown in Table 1. In the table we also indicate the performance of the WiHAUL max-min fair backhauling scheme we propose in this paper. This clearly yields the smallest level of unfairness, as quantified with the Gini coefficient [7], and only 10% lower total network throughput as compared to the greedy max-throughput strategy, which allocates all resources to a single flow.

Effective mm-wave backhauling is difficult and solutions designed for legacy multi-hop wireless networks operating in the 2.4/5GHz bands inappropriate. This is because every base station needs to decide when to beam steer to which neighbour and for how long, and to which flow to allocate more resources, subject to link rate heterogeneity (due to distance dependent path loss and potential blockages), dissimilar traffic demands, and fairness constraints. In particular, as the infrastructure has commercial value, it is essential to ensure resources are not left underutilised, while customers remain satisfied with the level of service provided. While the industry pursues standardisation of future 5G systems [8], the only protocol specification adopted with carrier-grade provisions, IEEE 802.11ad [9], leaves open the airtime allocation and scheduling tasks, which are crucial for backhauling.

In this paper we jointly solve the airtime allocation and per-link scheduling of aggregate traffic flows, i.e. flow bundles that originate/terminate at the same base station, which traverse multi-hop mm-wave backhauls. We explicitly take into account the distinct features of mm-wave technology, i.e. terminal deafness and susceptibility to link blockage, as well as realistic heterogeneous traffic demand regimes. Our goal is to achieve a good balance between overall network throughput performance and inter-flow fairness. In particular, we make the following key contributions:

1) We cast backhaul resource allocation as a max-min optimisation problem with specific terminal deafness and traffic demand constraints. We demonstrate that a max-min fair solution exists and is unique in scheduled-based multi-hop mm-wave networks.

2) We propose WiHAUL, a backhauling scheme comprising (i) a progressive filling algorithm that solves the max-min optimisation problem and computes per-hop airtime shares for each aggregate flow, and (ii) a light-weight scheduling protocol that works on top of IEEE 802.11ad, enforces the computed airtimes, and coordinates multi-hop transmissions, enabling spatial reuse.

3) We implement WiHAUL in the NS-3 simulator, building on preliminary mm-wave PHY measurements and incorporating 802.11ad specifications. We evaluate the performance of our solution over different network topologies, link dynamics, and traffic regimes. We demonstrate up to 5-fold throughput and fairness gains over previously proposed mm-wave access schemes.

1. Hereafter, whenever there is no scope for confusion, we use the terms 'flow' and 'aggregate flow' interchangeably.

2 System Model

We focus on dense mobile broadband deployments where $B$ fixed base stations provide wireless access to mobile users with different traffic demands. Base stations are connected via mm-wave links to wired Internet gateways, possibly over multiple hops. To achieve multi-Gbps transmission rates, base stations employ multi-carrier modulation and coding schemes (MCSs) and form very narrow beams between them. Following beam sweeping and beam-form training, base stations are aware of their neighbours, and time synchronisation is underpinned by GPS. To meet carrier-grade requirements, we consider time division multiplexing (TDM) is employed in the backhaul.

Our objective is to allocate the airtime resources available on the mm-wave backhaul links to aggregate traffic flows and co-ordinate transmissions among base stations. Flows either enter the network via gateways, are relayed by intermediary hops, before reaching the end users (downlink), or originate at different base stations and are forwarded externally by the gateways (uplink). The problem we pursue is challenging and fundamentally different to previous efforts in multi-hop wireless networks (e.g. [10]), since the backhaul system is free of secondary interference, but instead prone to terminal deafness. This is confirmed by recent measurements that reveal interference is negligible in mm-wave bands when employing highly-directional beams and links between any pair of nodes can be regarded as pseudo-wired [11]. Since we consider deployments with small form factor base stations equipped with a single mm-wave interface, intra-flow competition occurs and fairness issues arise as flows are relayed by base stations, unlike in multi-radio mesh networks [12]. Meanwhile, concurrent transmissions on non-interfering links is feasible, which allows for spatial reuse and appropriate network utilisation at a lower cost.

We envision a centralised architecture, whereby a controller has full knowledge of the network topology, periodically collects link rate and flow demand information, and subsequently performs airtime allocation and beam scheduling through the solution we introduce in this work. We denote $c_{i,j}$ the maximum achievable data rate between an $(i,j)$ base station pair (upper bounded by the Shannon capacity) and work with aggregate data traffic flows between base stations and the gateway. The demand $d_k$ of a flow $k$ represents the packet injection rate at the MAC queue. We assume the controller is also responsible for computing paths $p_k$ for all flows $k$ traversing the backhaul, which is orthogonal to the problems we attack and thus not considered herein.

The 802.11ad service period (SP) paradigm, which exploits the vast spectral resources available in the 60 GHz
We note that a flow segment can belong to multiple cliques be active simultaneously.

**Definition 1.** A ‘clique’ is the set of all flow segments that cannot
belong to a clique at a time. A clique is defined as a subset of the flows that cannot be simultaneously active.

\[ C \cap C' = \emptyset \quad \forall C, C' \in \mathcal{C} \]

where \( \mathcal{C} \) is the set of all cliques. We exemplify the conflict graph and clique notions with the simple topology depicted in Fig. 1, for which we can construct the equivalent conflict graph shown in Fig. 3. Observe that two cliques exist in this example and the segments of flows 1 and 2 over link \( l_{3,4} \), i.e. \( s_{1,3,4} \) and \( s_{2,3,4} \), simultaneously belong to both. Returning to our problem, by (3) we introduce a clique constraint that guarantees the total time consumed by all flow segments in a clique does not exceed 1, i.e. \( \sum_{s_{k,i,j} \in C_q} t_{k,i,j} \leq 1, \forall k \in \mathcal{F}_A, \forall C_q \in \mathcal{C} \).

In solving our problem, it will also prove useful to work with the notion of conflict node, defined on the actual network topology as below.

**Definition 2.** In a backhaul network, a ‘conflict node’ is a base station that forwards traffic on behalf of others.

For the example shown in Fig. 1, base stations 3 and 4 are conflict nodes.

**Table 1.** The rate region of a multi-hop mm-wave backhaul network is convex.

**Proof.** Since we consider transmissions between base stations are precisely scheduled, channel access in a clique can be seen as a single-hop time division multiplex (TDM) instance, which is known to have a convex capacity region [15].

Recall that there is no secondary interference between transmissions (pseudo-wired, point-to-point links) and the throughput of any sub-flow \( s_{k,i,j} \) in a clique \( C_q \) is upper bounded by the minimum between the throughput allocated in the clique \( C_{q-1} \) traversed previously and the total flow demand \( d_k \). The network rate region is obtained by the appropriate intersection of the rate regions of the component cliques. Thus it is convex.

The following key result follows.

**Corollary 1.** Max-min fair allocation in multi-hop mm-wave networks exists and it is unique.
Proof. We can prove by contradiction following the approach of Radunovic and Le Boudec that a max-min fair allocation vector is achievable on compact convex sets [16]. As per Lemma 1 above, the rate region of a scheduled mm-wave backhaul is convex, therefore a max-min fair allocation vector exists. By Theorem 2 in [17] and the constructive proof of Gafni and Bertsekas, p.1011 in [18], if any max-min allocation vector exists, then it is unique.

Hence, in the mm-wave backhaul scenario we consider, a max-min fair rate allocation vector exists and it is unique.

Finally, the rate region has the free disposal property [16] since each element of the rate vector $r = \{ r_k \mid k \in F \}$ is lower bounded by zero and any non-zero feasible allocation can always be decreased. It follows that a progressive filling algorithm can be employed to find the solution to the max-min fair allocation problem with mm-wave particularities.

4 WiHAUL: MAX-MIN FAIR BACKHAULING

In what follows we present a max-min fair multi-hop mm-wave backhauling mechanism, which we name WiHAUL. This consists of a progressive filling algorithm that solves the optimisation problem (1)-(3) in polynomial time, and a light-weight scheduling protocol that distributes airtime solutions among base stations, ensuring they communicate at the right time for the computed duratons.

4.1 Progressive Filling Algorithm

Algorithm 1 summarises the progressive filling procedure we propose to achieve max-min fair allocation of the backhaul resources under clique and demand constraints, and we detail its operation next. We start with all flow rates equal to zero and consider none of the aggregate flows have been allocated resources (lines 1–2). We call active flows, these flows for which an allocation was not performed. We gradually increases flow rates simultaneously, in steps of size $\epsilon$ kbps (line 4) until one or more flows either meet their demands (line 6) or activate a clique constraint (line 13). Note $\epsilon$ is a configurable parameter whose magnitude impacts on algorithm runtime. If a flow’s demand $d_k$ is satisfied, we freeze the allocated rate $r_k$ to the demand and remove that flow from the active set (line 8), thereafter considering it inactive and its resources frozen.

When a clique is fully utilised, we stop increasing the rates of the flows traversing it and proceed with computing from scratch the rates these should be assigned according to the remaining airtime budget. To this end, we subtract from the total available airtime, i.e. 1 (line 14), the fractions already reserved for inactive flows (line 18) and sum up the inverse of the link capacities corresponding to active flows in that clique (line 20). The latter will allow us to provide all active flows with the same rate $R$ (line 23), which under heterogeneous link rate conditions translates into allocating airtimes to each sub-flow that are inversely proportional to the traversed link’s capacity (line 25), i.e.

$$t_{k,i,j} = \frac{t_{\text{left}}}{c_{i,j} \sum_{s_{k,i,m} \in F \cap C_q} \frac{1}{c_{l,m}}}.$$ 

It is straightforward to verify that airtimes $t_{k,i,j}$ above sum to $t_{\text{left}}$, as required. Subsequently, we freeze the rates $r_k$ of flows in clique $C_q$ and remove them from the active set (line 26).

We repeat this procedure for the remaining active flows, until meeting their demand or activating other clique constraints. The progressive filling algorithm terminates when the set of active flows is empty (line 3). At that point we have obtained the airtimes to be allocated for each flow on each traversed backhaul link, in order to fulfil the max-min fair allocation of the rates.

Our algorithm’s runtime is a function of the highest flow rate divided by the step-length, which recall is configurable, and the total number of flows. Therefore the algorithm solves the max-min fairness optimisation problem posed in polynomial time.

4.2 Scheduling Procedure

Terminal deafness is a major challenge in mm-wave networks. Therefore, unless stations know to which neighbour to steer their beams, when, and for how long, they may be locked out, which would lead to frame loss and overall performance degradation. Algorithm 1 addresses the latter
and computes the airtimes for each flow segment, in order to attain max-min fair rates. To convey the computed airtimes and overcome the former issue (i.e. timing of TX/RX), WiHAUL employs a network-wide co-ordination procedure based on a scheduling hierarchy. This enables a centralised controller to dictate when nodes can transmit to others without conflict and in which order, so as to maximise spatial reuse.

We explain WiHAUL’s scheduling operation with the example topology shown in Fig. 1, considering the network operates with the 802.11ad SP mechanism. We assume a central controller (typically placed at the gateway; here node 6) has full knowledge of the network topology, including the hop distance to each base station, which of these are conflict nodes (i.e. have more than one neighbour), as well as their addresses, i.e.

1. \( H_i \): hop distance from node \( i \) to the gateway,
2. \( S_i \): node \( i \)’s conflict state,
3. \( A_i \): node \( i \)’s unique ID (e.g. its IP address).

With this information and the airtime shares computed by Algorithm 1, the controller constructs a hierarchy to establish when a node should transmit/receive and when it should schedule its neighbours, respectively. Specifically, WiHAUL first considers all conflict nodes as eligible candidates for acting as scheduling coordinators (in our example nodes 4 and 3). Among these, the one with the lowest hop distance \( H_l = \min_{j \in S_i} H_i \) is designated as the root coordinator and placed at the top of the scheduling hierarchy, namely at Level 0. In this example it is node 4 that acts as coordinator, while 6 (the gateway) is not a conflict node. The remaining nodes with \( S_i = 1 \) will be placed at a level that depends on the difference between their \( H_i \) value and that of the main coordinator (\( H_r \)) i.e. Level \( l = |H_r - H_i| \). Nodes with \( S_i = 0 \) will be placed at Level, below their neighbouring conflict node. As such, in our example nodes 5 and 6 reside at Level 1, while 1 and 2 at Level 2, as illustrated in Fig. 4.

At each level of the hierarchy, WiHAUL assigns airtime top–down, a node accepting the time allocated by its parent and assigning SPs to its children. In the considered example, the protocol first assigns SPs for 4 and then the nodes at Level 1, i.e. 3, 5 and 6. In turn, node 3 assigns SPs to 1 and 2, outside the interval when it is involved in communication with 4. This allows for spatial reuse, as links \( l_{4,5} \) and \( l_{3,1} \), and respectively \( l_{6,4} \) and \( l_{3,2} \) will be active simultaneously.

In case of multi-path routing, it may be happen that two or more nodes on the same level share the same neighbouring node that they could schedule. In such cases, the node with the smallest identifier \( A_i \) takes priority and will be the one scheduling. In turn, the child informs the other candidate parents of the assigned time, to resolve the tie and avoid conflicts. This process is repeated until all computed SPs have been disseminated to all stations.

Subsequently, nodes will periodically switch their beams towards the corresponding neighbours for transmission/reception during the assigned times. To adapt to the dynamics of physical channel conditions (e.g. link blockage) and the changing flow demands, the controller will periodically (e.g. every beacon interval) collect link quality and flow demand information, run the progressive filling algorithm, and re-schedule flow segments as appropriate.

### 5 Performance Evaluation

To evaluate the performance of WiHAUL, we implement this in NS-3 and conduct extensive simulations under different scenarios, comparing with recent scheduling schemes for mm-wave networks, including DLMAC [19], M-DMAC [20], and variations of these. We examine achievable gains in terms of flow throughputs and airtime distribution, and overall network throughput, and the level of fairness each approach attains over realistic multi-hop topologies. We further analyse WiHAUL’s behaviour in terms of allocated airtime distributions and airtimes, and give insight into the impact of link rates and flow demands on the partitioning of resources. Lastly, we evaluate our solution with real data traffic traces and examine end-to-end delay performance.

#### 5.1 Simulation Environment

To incorporate multi-hop frame relaying, the controller logic, and the progressive filling algorithm in NS-3, we extended the IEEE 802.11ad service period (SP) based MAC implementation of Facchi et al. [21]. We note that this employs a simple PHY, where antenna beams are wide and thus transmissions may be subject to secondary interference. While PHY design remains outside the scope of this work, we expect narrow beam-forming should be feasible with mm-wave transceivers. Therefore, we update the simulation model to eliminate inter-link interference. We compute receive powers using the Friis free space equation [23], then map SNRs onto the corresponding MCSs.

Given the switched operation of transmissions and receptions, and the high PHY bit rates employed on links, to avoid excessive delays and buffer overflows at relaying stations, we divide the airtime allotted to each sub-flow into multiple SPs each of shorter duration. In our implementation we work with 20 short SPs that sum up to computed airtime allocations. To maximise protocol efficiency, we employ AMPDU frame aggregation and upon each SP transmit up to 45KB of data preceded by a single PLCP. We work

2. The source code of our implementation is available at https://git.io/wihaul.
3. We assume beam-form training has been established upon simulation start and beam switching overheads are negligible [22].
with applications that generate fixed packets, except when experimenting with real traffic traces. We summarise the parameters used in simulation in Table 2.

### Table 2: Simulation settings.

| Parameter               | Value       |
|-------------------------|-------------|
| Tx power                | 10 dBi      |
| Tx Rx antenna gain      | 20 dBi      |
| BI duration             | 102400 µs   |
| BI overhead             | 10240 µs    |
| Progressive filling step length ($e$) | 10 kbps   |
| UDP payload             | 1470 B      |
| TCP MSS                 | 1460 B      |
| Time fraction allocated for TCP ACKs | 0.06 |
| TCP Initial Slow Start Threshold | 64 KB |
| TCP Tx/Rx Buffer Size   | 10 MB       |

5.2 Fairness Metrics

Note that max-min is a qualitative fairness criterion. That is, some allocation is max-min fair if increasing the rate of a flow is only possible by decreasing that of others [14]. Unlike e.g. Jain’s fairness index, this typically does not have a directly measurable value. Therefore, to quantify fairness, we first resort to the concept of inequality distribution used in economics, and compute Gini coefficients [7], using the following formula:

$$G = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1}^{n} |r_k - r_l|,$$

where $r_k$ is the rate allocated to flow $k$, and $n$ is the total number of flows. The lower this coefficient is with a certain rate allocation vector, the more fair the distribution of resources is.

To add further perspective and quantify to what extent the minimum flow rate in the network might be higher with WiHAUL than with other schemes, we employ the generalised measure of fairness defined in [24], as follows

$$M_\beta(r) = \text{sign}(1-\beta) \cdot \left\{ \frac{1}{n} \sum_{k=1}^{n} \left( \frac{r_k}{\sum_{l=1}^{n} r_l} \right)^{1-\beta} \right\}^{\frac{1}{1-\beta}},$$

where $\beta$ dictates different types of fairness measures. For max-min fairness $\beta \to \infty$, and $M_\beta(r)$ becomes

$$M_\beta(r) = \lim_{\beta \to \infty} \text{sign}(1-\beta) \left[ \sum_{k=1}^{n} \left( \frac{r_k}{\sum_{l=1}^{n} r_l} \right)^{1-\beta} \right]^{\frac{1}{1-\beta}} = -e \lim_{\beta \to \infty} \log \left[ \left\{ \sum_{k=1}^{n} \left( \frac{r_k}{\sum_{l=1}^{n} r_l} \right)^{1-\beta} \right\}^{\frac{1}{1-\beta}} \right].$$

We denote $y_k = (\sum_{l=1}^{n} r_l)/r_k$ and solve the limit above by applying l’Hôpital’s rule, which leads to

$$\lim_{\beta \to \infty} \left( \sum_{k=1}^{n} \beta^{-1} \log(y_k) \right)/\left( \sum_{k=1}^{n} \beta^{-1} \right).$$

As $\beta \to \infty$, the numerator is dominated by the highest $y_k$ term, i.e. $\max_k \{ y_k \log(y_k) \}$, hence the limit converges to

$$\max_k \left\{ \frac{\sum_{l=1}^{n} r_l}{r_k} \right\}.$$

5.3 Comparison with State-of-the-Art Solutions

We compare the performance of WiHAUL against that of recent mm-wave scheduling schemes DLMAC [19] and MDMAC [20] in terms of mean and total network throughput, and inter-flow fairness. We conduct the evaluation over several topologies generated with the Cerdà-Alabern model that captures the characteristics of real-world multi-hop wireless deployments [25]. The topologies considered comprise 10 to 15 stations (including the Internet gateway) and the number of aggregate flows traversing the network varies between 7 and 10. We illustrate four of these topologies in Fig. 5, where the X and Y axes represent the base stations’ coordinates, with base station 0 being the gateway. Link rates vary between 2.772-6.756 Gbps, depending on distance between stations.

We also compare against optimised DLMAC and MDMAC versions that seek to reduce gaps between transmissions (BinDLMAC) [19] and operate with slot sizes that maximise transmission efficiency respectively (OptMDMAC). We note all these are decentralised and do not explicitly consider fairness in their design. Each approach transports backlogged aggregate flows (unlimited demand) transmitted over UDP.

**Finding:** WiHAUL achieves the highest average flow throughput (and therefore total network throughput), irrespective of the number of hops flows traverse and with how many competing flows they share links.

Let us examine first Figs. 6a–6d, where we show the average and 95% confidence intervals of individual flow throughputs attained with WiHAUL, DLMAC, MDMAC, and their variations, in each topology considered. In these figures we also plot the average throughput performance over all flows as the last cluster of bars to the right of each plot. Observe in these clusters that the bars corresponding to WiHAUL are indeed the highest and the total network throughput ranges between ranges between 2.25-2.5 Gbps in all cases.

**Finding:** With WiHAUL, flows attain similar throughput as long as they share the same cliché, while additional underutilised network resources are equally divided among unconstrained flows.

Indeed, observe that flows which encounter less competition attain superior performance with our approach, without negatively impacting on the others. This can be observed in Figs. 6a and 6b, where with WiHAUL flows $f_0$ and $f_1$, and respectively $f_0-f_3$ achieve approximately 450 Mbps and 100 Mbps more throughput than the other flows traversing the backhaul. At the same time, we reduce the gross performance dissimilarity between flows (e.g. up to 1 Gbps between flows $f_1$ and $f_0$ with BinDLMAC in topology 1). In addition, the flows penalised by earlier approaches attain up to 5x higher throughput with WiHAUL (observe flow $f_4$ in Fig. 6b with WiHAUL and BinDLMAC).

4. The default MDMAC design works with a slotted channel where slot size is fixed to 20µs. The optimised version we consider works with slots that can accommodate exactly one transmission burst.
Finding: WiHAUL does not unnecessarily penalise flows that terminate/originate further away from gateways.

Note in Figs. 6c–6d that with WiHAUL all flows achieve the same throughput for topologies 3–4, unlike with DLMAC, MDMAC, and their variations, which largely favour flows terminating closer to the gateway and penalise those with end-points multiple hops away. (Opt)MDMAC is less prone to such behaviour, though has the disadvantage of requiring appropriate configuration of the slot size, which is impractical. Nonetheless, although the ‘optimised’ MDMAC version performs relatively well overall, it still carries
unfairness, as e.g. with this scheme flow $f_{7}$ in the third topology attains nearly half the throughput provided by WHAUL (Fig. 6c).

To examine closer the fairness properties of all schemes, in Fig. 7 we plot the Gini coefficients corresponding to the flow rate allocations each of these yields in the 5 topologies considered. Recall the Gini coefficient gives a numerical representation of inequality, with a lower value corresponding to a fairer allocation. Observe that although these values depend on the network topology, number of flows, and link rates, WHAUL outperforms the existing schemes, being in particular considerably more fair than the DLMAC variants. Precisely, the Gini coefficients when the network operates with BinDLMAC range between 0.2 and 0.5 and are the highest in all 4 topologies. DLMAC performs marginally better, while (Opt)MDMAC yields Gini coefficients in the 0.1–0.3 range. Our proposal leads to the lowest Gini coefficients in all topologies (0.004–0.2), being substantially less unfair than the others. These properties are further confirmed by the results we give in Table 3, which shows the fairness measure as derived in (4) for our approach and the benchmarks considered. Indeed $M_{3}$ is up to $5 \times$ higher with our approach, which also indicates WHAUL ensures superior performance for the smallest flow, yet remains fair to the others.

We conclude that existing decentralised approaches bias against flows with longer hop-distance and/or inferior link rates; in contrast, the proposed WHAUL not only achieves more fair partitioning of resources among all traffic flows, but also higher throughput for the smallest flow and overall higher mean throughput performance. This has important practical implication on cellular backhauls where WHAUL could provide superior and more homogeneous service guarantees to users.

5.4 Dynamic Conditions

Next we undertake an in-depth analysis of WHAUL’s operation, investigating the impact of link quality dynamics and flow demand variations on airtime allocation and end-to-end performance. For this we envision a lamppost based deployment in the Old Market Square of Nottingham as shown in Fig. 8, which we obtain from a publicly available data set [26]. This topology consists of 16 base stations (STAs) that communicate over mm-wave links and we envision 10 aggregate flows from the gateway (STA0). Also shown in the figure are three cliques of interest and, for ease of explanation, we consider the deployment as ‘partitioned’ into three regions.

Finding: Max-min fair backhauling requires a non-trivial partitioning of the available airtime resources, which depends on the demand of each flow, the paths traversed, and the capacities of the links these comprise.

5.4.1 Demand Variation

We first examine a scenario where the demand of a single flow (i.e. $f_{6}$ originating at STA0 and terminating at STA14) grows from 300 Mbps to 1.5 Gbps, while that of the others remains fixed to 400 Mbps. Our goal is to understand how this impacts on airtime allocations and verify that the rates of the smallest flows are unaffected. We illustrate the results of this experiment in Fig. 9, where we plot (a) the time evolution of the individual throughputs and (b) the fraction of airtime allocated to $f_{6}$ on link $l_{0,4}$, as well as the total airtime allocated in Clique $C_{0}$, which constrains $f_{6}$.

Observe that the throughput of $f_{6}$ increases with demand, up to 1 Gbps, when the clique constraint is activated (total airtime in $C_{0}$ reaches 1) and the throughput is capped despite further growth in demand. As intended, the throughput of the remaining flows stays at 400 Mbps, which indicates their demand is satisfied throughout. Note that the scheduling process is repeated every BI, link rate and demand updates are collected during BHIs, and takes one BI duration for the demand increase to propagate through the network. Examining the airtime utilisation in the bottleneck clique $C_{0}$ and the time allocated to the demand-varying flow, $t_{6,0.4}$, we see that sufficient resources exist to accommodate the entire demand growth up to 900 Mbps, as $t_{6,0.4}$ is tripled. Further increasing this demand does not result in a throughput increase above 1 Gbps, to protect the remaining flows, which complies with the max-min fair allocation paradigm proposed.

5.4.2 Shared Link Degradation

Next we examine the impact of link quality variation on the performance of all flows traversing such a link, when max-min fair allocation is performed. To this end, we simulate different degrees of link blockage between STA3 and STA0 (i.e. $l_{0,3}$), which results in signal attenuation between 5dB and 20dB. As a result, the modulation and coding scheme (MCS) employed is reduced from 4.9 Gbps to 598 Mbps, to preserve link reliability. In this scenario, we assume the bit rates of the other links remains constant and the demand of all flows is 400 Mbps.

Fig. 10 illustrates the results of this experiment, where we measure (a) the individual flow throughputs and (b) the total time utilisation in cliques $C_{0}$ and $C_{3}$, as well as the sum of airtime fractions allocated to all flow segments traversing $l_{0,3}$, from the perspective of these cliques. Note that the airtime allocation on $l_{0,3}$ is effectively fixed under each link quality condition, but it may well represent dif-
different fractions from cliques’ perspectives. When the link quality is high (i.e. \( c_{0,3} = 4.982 \) Gbps), the total airtime consumption in \( C_0 \) and \( C_3 \) is below 1, hence all flows are satisfied. This is indeed confirmed by the flow throughputs shown in Fig. 10a. Subsequently, when a 5 dB attenuation is introduced at the third BI, the throughputs of flows \( f_0-f_4 \) drop slightly, while those of \( f_5-f_9 \) remain satisfied. That is because \( C_0 \) still has sufficient resources (airtime consumed sums to 0.96), while the \( C_3 \) clique constraint becomes active (airtime reaches 1). Further attenuation on link \( l_{0,3} \) (yielding 2.776 Gbps bit rate), leads to the activation of the \( C_0 \) constraint, and consequently to a decrease in the throughput of all flows. However, as \( C_3 \) becomes constrained before \( C_0 \), flows \( f_0-f_4 \) attain slightly lower (approx. 30 Mbps) throughput than \( f_5-f_9 \). Lastly, this performance gap shrinks as link \( l_{0,3} \) degrades further (BI 9 onward) and additional degradation would completely close the gap to meet the max-min fairness criterion. Meanwhile, the total time consumed by link \( l_{0,3} \) to transport all flows is increasing to as much as 0.9 at the end of the simulation.

We conclude that degradation of an intensively shared link (and clique) has a significant impact on the throughput performance of the entire network. Nevertheless, WiHAUL guarantees max-min fair allocation of the flow rates.

5.4.3 Heterogeneous Demands and Cascaded Cliques

Finally we consider more complex circumstances where the demands of flows in regions 1–3 as shown in Fig. 8 are 500, 400, and 600 Mbps respectively, while the quality of link \( l_{5,8} \) varies. Signal attenuation decreases and capacity grows from 598 to 4,982 Mbps on this link after every third BI. As \( l_{5,8} \) only carries flow \( f_2 \), we investigate in Fig. 11b the changes in time allocation within all the cliques that \( f_2 \) traverses, i.e. \( C_5, C_3, \) and \( C_0 \), and show the time evolution of individual flow throughputs in Fig. 11a.

Note that as \( c_{5,8} \) increases, more airtime is made available for both \( f_1 \) and \( f_2 \), as they share the same clique \( C_5 \). In effect, the constraint of this clique is removed (total airtime consumption drops from 1 to 0.5) and this also impacts on the flows with which \( f_1 \) and \( f_2 \) share cliques \( C_3 \) and \( C_0 \), i.e. \( f_0, f_3 \) and \( f_4 \). Precisely, the throughput of these drops to 415 Mbps after the third BI. As the quality of \( l_{5,8} \) further increases, the total airtime allocated to \( f_2 \) on this link, i.e. segment \( s_{2,5,8} \), decreases, though the flows in region 1 are together constraint by \( C_3 \). This confirms the proposed max-min fair allocation strategy ensures \( f_2 \) is not allocated more resources in cliques \( C_3 \) and \( C_0 \), as this would come at the cost of a decrease in \( f_0, f_3 \) and \( f_4 \)’s throughput. Lastly, observe that the throughput of the other flows remains
unaffected, as the demand of \( f_5, f_6, \) and \( f_7 \) is the smallest among all (i.e. 400 Mbps) and changes in \( c_{5,8} \) do not affect clique \( C_0 \), which is shared by all flows.

### 5.5 Real-Time Traffic

We complete our evaluation of WiHAUL by conducting experiments with real-time traffic potentially subject to latency constraints. We are particularly interested in the delay packets experience while traversing multi-hop mm-wave backhauls, where cascaded queues could have a negative impact on user experience. To this end, we emulate dynamic adaptive streaming over HTTP (DASH) by extracting meta-data from mobile traffic traces collected in New York City [27]. We replay 100 such video sessions in parallel towards different base stations (download) in the topology shown in Fig. 8. The distribution of the session bit rates is shown in Fig. 12, where observe that individual bit rates vary between 100 Kbps and 3.4 Mbps.

Under these circumstances, we measure the packet round-trip-time (RTT) for each aggregate flow over 30 seconds, as well as the average throughputs. We plot the RTT experienced by TCP segments in Fig. 13a, where observe this is below 30 ms, with median values for all aggregates falling between 8 and 15 ms. This complies with the NGMN Alliance specifications for end-to-end delay (20 ms) in small cell backhauls [28]. As expected, RTTs are proportional to the number of hops traversed, however, their distribution also depends on how frequently they are served. Precisely, note that the slope of the CDFs decreases with the number of aggregates traversing the first hop from the gateway and thus the latency in different regions is only scaled up by the number of hops each aggregate traverses. For instance, flows \( f_1, f_2, \) and \( f_4 \) are 3 hops away from the gateway (node 0) and share \( l_{0,3} \) with \( f_0 \). As such, the RTTs they experience are identical (overlapping curves). Flow \( f_0 \) also traverse 3 hops, but only shares \( l_{0,1} \) with \( f_8 \), hence their RTT distributions start at \( \sim 5 \)ms, but quickly diverge (medians 11 and respectively 15ms).

Turning attention to aggregate flow throughputs, we show the average and 95% confidence intervals of this metric in Fig. 13b. We see that overall performance is homogeneous (despite flows traversing different number of hops and experiencing different link rates), fluctuating around 100 Mbps for each aggregate. Note that in this scenario all flows are satisfied and cliques are not constrained.

### 6 Related Work

To the best of our knowledge, the proposed WiHAUL scheme is the first to perform airtime allocation in mm-wave backhauls, while explicitly addressing the distribution of flow rates and quantifying max-min fairness. In what follows, we review related work of direct relevance to our contribution.

**Mm-wave Characterisation:** Recent empirical studies confirm the millimetre-wave band (30–300GHz) will be able to support multi-Gbps link rates [5], hence it becomes a promising candidate to accommodate bandwidth intensive small-cell wireless backhauling solutions [29]. Channel measurement efforts also confirm that beamforming necessary to mitigate attenuation in mm-wave bands also drastically reduces interference, and links can be regarded as pseudo-wired [11]. As a consequence, terminal deafness becomes a key challenge when scheduling transmissions/receptions [30].

**Medium Access & Scheduling in mm-wave Networks:**

The IEEE 802.11ad standard [9] specifies contention based
and service period (SP) driven channel access mechanisms for communications in the unlicensed 60GHz band. However, the standard leaves open the airtime allocation and multi-hop transmission coordination tasks. Hemanth and Venkatesh analyse the performance of the SP mechanism in terms of frame delay [31]. Several works build upon the standard and specify MAC protocol improvements for single-hop WLANs [32], [33], [34]. Chandra et al. employ adaptive beamwidth to achieve improved channel utilisation [32]. Sim et al. exploit dual-band channel access to address terminal deafness and improve throughput [33]. Optimal client association and airtime allocation is pursued in [21] to maximise the utility of enterprise mm-wave network deployments.

A directional cooperative MAC protocol is introduced in [34], where user devices select intermediate nodes to relay the packets to the AP, in order to establish multi-hop paths that exhibits higher SNR than direct links. Mandke and Netttles propose a dual-band architecture for multi-hop 60 GHz networks where scheduling and routing decisions are communicated at 5.2 GHz [35]. Despite considering the implications of terminal deafness, these designs do not tackle the airtime allocation problem.

Distributed opportunistic transmission schemes for multi-hop scenarios have been proposed to achieve network-wide scheduling [19], [20]. MDMAC operates with a slotted channel whereby a station’s transmission can occupy one or multiple slots, but the slot duration remains fixed for all participants (20 µs by default), which may harm efficiency [20]. Unslotted approaches (Bin)DLMAC are introduced in [19] to improve protocol efficiency and ‘learn’ when to transmit in the presence of terminal deafness. Both schemes do not explicitly consider inter-flow fairness, as each node seeks to transmit as much as possible. Our results confirm this leads to poor performance for flows encountering lower capacity links. In contrast, the WiHAUL mechanism we propose in this work not only improves throughput performance, but is also significantly more fair, as we take into account all flow demands, link rates, and level of competition.

**Fairness in Multi-hop Wireless Networks:** Bertsekas and Gallager consider max-min fairness for flow control in wired networks [14] and subsequently Le Boudec and Radunovic demonstrate this is a geometric property of the set of feasible allocations [16]. The 802.11 rate region is proven log-convex, and station attempt probabilities and burst sizes in 802.11 mesh networks are derived for max-min fair regimes in [12]. This however only holds in multi-channel mesh topologies where stations employ multiple interfaces, which is impractical with small form factor mm-wave devices equipped with a single interface. Wang et al. argue that channel time rather than flow rate should be used with the max-min allocation criterion in wireless multi-hop networks and accordingly propose a new definition of max-min fairness [10]. Unfortunately, under this definition, flows traversing more hops will, by design, obtain considerably smaller throughput than those close to gateways. This implies inferior service performance for distant users, hence the approach is ill-suited to carrier-grade the backhauls.

Lan et al. propose a unified fairness measure that enables to explicitly quantify max-min fairness, which is largely perceived as qualitative [24]. We use their general measure of fairness to derive a max-min fair metric and evaluate the gains achieved by our proposal. To add further perspective we also resort to economic notions of inequality, i.e. the Gini coefficient [7].

7 Conclusions

By supporting multi-Gbps link rates, mm-wave technology is becoming a promising enabler of wireless backhauling solutions in ultra-dense cellular deployments. Highly directional beam-forming is mandatory to combat severe signal attenuation specific to these frequencies, though gives rise to cumbersome terminal deafness issue that must be tackled to fully exploit vast bandwidth resources. In this paper, we built upon the scheduled access paradigm of the IEEE 802.11ad standard and proposed WiHAUL, a network-wide airtime resource allocation and scheduling mechanism, which explicitly guarantees inter-flow max-min fairness in mm-wave backhauls. We validated our solution over a broad range of network conditions and demonstrated via extensive simulations that WiHAUL achieves up to 5× higher measurable fairness as compared to existing
mm-wave MAC proposals, improving up to fivefold the throughput of otherwise limited flows, while attaining superior overall network throughput. Further, we demonstrated that our approach is able to meet typical delay constraints of real-time applications.

**ACKNOWLEDGEMENTS**

We thank Pablo Serrano for his valuable feedback that helped improving this manuscript. We also thank Nicolò Facchi for helping us extend the NS-3 simulator, which we used for evaluation purposes.

**REFERENCES**

[1] Rui Li and Paul Patras. WiHaul: Max-min fair wireless backhauling over multi-hop millimetre-wave links. In Proc. ACM HotWires, New York City, NY, USA, Oct. 2016.

[2] Patrik Cerwall (ed). Ericsson Mobility Report. https://www.ericsson.com/assets/local/mobility-report/documents, 2017/ericsson-mobility-report-june-2017.pdf, Jun. 2017.

[3] Forbes. With T-Mobile Mixing It Up Again, helped improving this manuscript. We also thank Nicolò Facchi for helping us extend the NS-3 simulator, which we used for evaluation purposes.

[4] Theodore S. Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for wireless evolution into 5G. IEEE Communications Magazine, 52(2):82–89, 2014.

[5] Jeffrey Jaffe. Bottleneck flow control. IEEE Transactions on Communications, 29(7):954–962, Jul 1981.

[6] Eli Gafni and Dimitri Bertsekas. Dynamic control of session input rates in communication networks. IEEE Transactions on Automatic Control, 29(11):1009–1016, Nov 1984.

[7] Gek Hong Sim, Rui Li, Cristina Cano, David Malone, Paul Patras, and Joerg Widmer. Learning from experience: Efficient decentralized scheduling for 60GHz mesh networks. In Proc. IEEE WoWMoM, pages 1–9, 2016.

[8] Sumit Singh, Raghuraman Mudumbai, and Upamanyu Madhow. Distributed coordination with deaf neighbors: Efficient medium access for 60GHz mesh networks. In Proc. IEEE INFOCOM, pages 1–9, 2010.

[9] Nicolò Facchi, Francesco Gringoli, and Paul Patras. Maximising the utility of enterprise millimetre-wave networks. arXiv preprint arXiv:1706.04278, 2017.

[10] G. R. MacCartney and T. S. Rappaport. A flexible millimeter-wave channel sounder with absolute timing. IEEE Journal on Selected Areas in Communications, 35(6):1402–1418, June 2017.

[11] Theodore S Rappaport (ed). Wireless communications: principles and practice. 2nd edition, Prentice Hall, 2002.

[12] G. R. MacCartney and Theodore S. Rappaport. 73GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City. In Proc. IEEE ICC, pages 4862–4867, 2014.

[13] Evangelos Haleplidis, Kostas Pentikousis, Spyros Denazis, James Hadi Salim, David Meyer, and Odysseas Koufopavlou. Software-defined networking (SDN): Layers and architecture terminology. RFC 7426, Jan. 2015.

[14] Jeffrey Jaffe. Bottleneck flow control. IEEE Transactions on Communications, 29(7):954–962, Jul 1981.

[15] Qian Chen, Jiqiang Tang, David Tung Chong Wong, Xiaoming Gek Hong Sim, Thomas Nitsche, and Joerg C Widmer. Addressing MAC layer inefficiency and deafness of IEEE 802.11ad millimeter wave networks using a multi-band approach. In Proc. IEEE CCNC, pages 458–464, 2014.

[16] Eli Gafni and Dimitri Bertsekas. Dynamic control of session input rates in communication networks. IEEE Transactions on Automatic Control, 29(11):1009–1016, Nov 1984.

[17] Gek Hong Sim, Rui Li, Cristina Cano, David Malone, Paul Patras, and Joerg Widmer. Learning from experience: Efficient decentralized scheduling for 60GHz mesh networks. In Proc. IEEE WoWMoM, pages 1–9, 2016.

[18] Sumit Singh, Raghuraman Mudumbai, and Upamanyu Madhow. Distributed coordination with deaf neighbors: Efficient medium access for 60GHz mesh networks. In Proc. IEEE INFOCOM, pages 1–9, 2010.

[19] Nicolò Facchi, Francesco Gringoli, and Paul Patras. Maximising the utility of enterprise millimetre-wave networks. arXiv preprint arXiv:1706.04278, 2017.

[20] G. R. MacCartney and T. S. Rappaport. A flexible millimeter-wave channel sounder with absolute timing. IEEE Journal on Selected Areas in Communications, 35(6):1402–1418, June 2017.

[21] Theodore S Rappaport (ed). Wireless communications: principles and practice. 2nd edition, Prentice Hall, 2002.

[22] G. R. MacCartney and Theodore S. Rappaport. 73GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City. In Proc. IEEE ICC, pages 4862–4867, 2014.

[23] Qian Chen, Jiqiang Tang, David Tung Chong Wong, Xiaoming Gek Hong Sim, Thomas Nitsche, and Joerg C Widmer. Addressing MAC layer inefficiency and deafness of IEEE 802.11ad millimeter wave networks using a multi-band approach. In Proc. IEEE CCNC, pages 458–464, 2014.

[24] Gek Hong Sim, Thomas Nitsche, and Joerg C Widmer. Addressing MAC layer inefficiency and deafness of IEEE 802.11ad millimeter wave networks using a multi-band approach. In Proc. IEEE PIMRC, pages 1–6, 2016.

[25] Qian Chen, Jiqiang Tang, David Tung Chong Wong, Xiaoming Peng, and Youguang Zhang. Directional cooperative MAC protocol design and performance analysis for IEEE 802.11ad WLANs. IEEE Transactions on Vehicular Technology, 62(6):2667–2677, 2013.

[26] Ketan Mandke and Scott M Nettes. A dual-band architecture for multi-Gbps communication in 60GHz multi-hop networks. In Proc. ACM International Workshop on MmWave Communications: From Circuits to Networks, pages 9–14, 2010.