WiHaul: Max-Min Fair Wireless Backhauling over Multi-Hop Millimetre-Wave Links

Rui Li and Paul Patras
School of Informatics, The University of Edinburgh, UK

ABSTRACT
The mobile networking community is pursuing densification of small cell deployments to address the capacity crisis inherent to the projected exponential increase in mobile data traffic. Connecting massive numbers of access points to the Internet using optical fibre is however both very complex and expensive. In this paper we tackle small cell backhauling wirelessly, building upon recent advances in millimetre-wave technology. We propose a resource allocation algorithm for aggregate data flows traversing such multi-hop backhauls, and specify WiHaul, a light-weight hierarchical scheduling protocol that enforces the computed airtime shares and coordinates multi-hop transmissions effectively. To achieve high throughput performance while ensuring low demand flows are satisfied, we adopt a max-min fair allocation strategy. Results we present show our solution attains max-min fairness through a non-trivial partitioning of the airtime budget available in cliques of sub-flows, which depends on flow demands, their paths, and the capacities of the links traversed.

CCS Concepts
• Networks → Network resources allocation;

Keywords
mm-wave; multi-hop; backhauling; max-min fairness

1. INTRODUCTION
Recent market surveys highlight increasing user preference for wireless Internet access and growing popularity of bandwidth-intensive applications, which substantially accelerate mobile data traffic demand worldwide [1]. In response, mobile service providers are shrinking cell coverage, while increasing the number of base stations deployed [2]. Such densification requires efficient backhauling, to transfer vast volumes of data between the access and core networks. Traditional fibre based high speed backhauls are only a short-term solution, since CAPEX grows with network size, while reconfiguration flexibility is limited in intensive deployments.

Wireless backhauling alternatives have been thus far based on technologies that operate over ultra high or microwave (0.3–30GHz) frequency bands. These resources, however, have limited capacity and are already overcrowded with numerous applications, including Wi-Fi, television broadcast, cellular access, RADAR, baby monitors, and cordless phones. In contrast, the underutilised millimetre-wave (mm-wave) band (30–300GHz) exposes considerably wider spectrum that could support an order of magnitude higher data rates, which prompted regulators to mandate license-exempt use of the 60GHz band [3]. The shortcoming is that signals attenuate severely at these frequencies, which imposes to use electronically steerable highly directional beams, to mitigate loss.

Directionality intrinsically eliminates interference and enables spatial reuse, though introduces terminal deafness, i.e. receivers can hardly be aware of transmissions, unless their beams are aligned with those of transmitters. To minimise base stations' form factor and cost, these are expected to operate with single transceivers, which further exacerbates link blockage issues, especially when relaying packets over multiple hops. Unfortunately, the IEEE 802.11ad standard only specifies scheduled, service period (SP) based medium access [4], to meet carrier-grade requirements, but leaves open the SP allocation and beam scheduling tasks. Given the unique characteristics of mm-wave networks, mechanisms designed for legacy multi-hop wireless networks operating in the 2.4/5GHz bands cannot overcome these problems. Simultaneously rethinking airtime allocation and beam scheduling for mm-wave backhauling is required instead, to ensure network resources are utilised efficiently. This is particularly difficult when network density increases and base stations need to schedule the TX/RX of a set of nodes, while being scheduled in turn by others. Link rate heterogeneity due to distance dependent path loss, dissimilar flow demands, and fairness constraints complicate this problem further.

Contributions: In this paper, (i) we propose a resource allocation algorithm for aggregate data flows traversing multi-hop backhauls, specifically addressing the distinct features of mm-wave networks employing 802.11ad, as exemplified in Fig. 1; (ii) we then detail WiHaul, a light-weight scheduling protocol that enforces the computed airtime shares and coordinates multi-hop transmissions.

Our goal is to achieve a good balance between total throughput performance and inter-flow fairness, and ensure all users experience a satisfactory level of service. Therefore, we take a max-min fair allocation approach, whereby flows encoun-
terating low capacity links and increased competition are not unnecessarily throttled. This contrasts with max-throughput allocation strategies that favour large volume flows traversing high capacity links, as well as with round-robin schemes that allocate equal resources and lead to wastage in the presence of low demand flows. We expect max-min fair allocation will ensure flows with inferior demands receive as many resources as possible and assign no less resources to flows with high demands.

We show that when working with service periods (SPs), max-min fair allocation indeed exists and it is unique. We thus give a progressive filling algorithm that takes into account properties specific to multi-hop mm-wave networks, to compute per-hop airtime shares for each aggregate flow. The proposed WiHaul protocol then distributes the computed shares and enables concurrent transmissions between non-interfering links, to improve spatial reuse. Finally, we present numerical results that suggest max-min fair allocation of backhaul resources is achieved with a non-trivial partitioning of the airtime budget available in cliques of sub-flows, which depends on flow demands, their paths, and the capacities of links traversed.

2. MAX-MIN FAIR ALLOCATION

Our focus is on dense mobile broadband networks where LTE base stations and Wi-Fi access points serve end users, and are connected to the Internet wirelessly through multi-hop mm-wave links to gateways. We aim distribute backhaul resources among aggregate flows that originate at different base stations and are forwarded externally by the gateways (uplink), or enter gateways, are relayed by intermediary hops, and reach users through last hop access (downlink).

A key challenge specific to this task is ensuring aggregate flows are treated fairly while traversing the backhaul. Working with a certain type of fairness criterion is often contentious, as operators may favour equal throughput, equal airtime, or mixed allocations. To ensure less intensive flows are not starved, here we work with the max-min fair criterion [5] and thus seek to ensure flow demands are fulfilled in increasing order, where possible, while any remaining network capacity is shared among those with higher demands. This is equivalent to maximising the allocation of each aggregate flow, subject to two constraints, namely (i) the allocated throughput does not exceed the flow’s demand, and (ii) any increase in the allocation of that flow would not cause a decrease in that of others with already smaller or equal throughputs. In what follows we show that a max-min fair allocation in 802.11ad based backhaul exists and give a progressive filling algorithm to compute this. Having found the fractions of airtime to be allocated to each flow segment, we specify a protocol that coordinates the transmissions of all nodes and enforces these allocations.

System Model: We consider a network with $B$ base stations that form narrow beams to communicate with their neighbours and denote $C_{i,j}$ the maximum achievable data rate between an $(i, j)$ base station pair. According to recent measurements, interference is negligible at 60GHz when employing highly-directional beams and links between any pair of nodes can be regarded as pseudo-wired [6]. Therefore the problem we pursue is different to previous efforts in multi-hop wireless networks, since the system is free of secondary interference, but instead prone to terminal deafness. Further, given the envisioned small form factor of base stations, we expect these to be equipped with a single mm-wave interface and thus, unlike in multi-radio mesh networks [7], intra-flow competition occurs as base stations relay flows.

We denote $\mathcal{F}$ the set of flows traversing the backhaul and $p_k$ the path of flow $k$, i.e. the sequence of links this follows from source to destination. We are concerned with carrier-grade backhauling and thus consider multi-hop 60GHz networks that operate with the SP based paradigm, whereby some access points assign transmission times (SPs) in their neighbourhood [4]. By this mechanism a flow $k$ is assigned an SP of duration $t_{k,i,j}$ on link $l_{i,j}$ and we find the vector

$$
t = \{t_{k,i,j} | k \in \mathcal{F}, l_{i,j} \in p_k\},$$

that achieves max-min fair allocation of flow throughputs.

We note that single transceiver stations can only send to or receive from one neighbour at a time, and construct a conflict graph $G(V, E)$ to model which flow segments $s_{k,i,j}$ cannot be activated simultaneously, i.e. an edge exists between any two such vertices. We group conflicting sub-flows into cliques and note that a sub-flow can belong to multiple cliques. We also define conflict nodes in the original network topology, as those base stations that forward traffic on behalf of others. In the network example shown in Fig. 1, stations 3 and 4 are conflict nodes and we can build the equivalent conflict graph shown in Fig. 2.

Consider a flow $k$ has a demand $d_k$ and the end-to-end
throughput to be allocated is \( r_k \). We will also incorporate a demand constraint to ensure the allocated rate does not exceed the demand of the aggregated flow, i.e.,

\[
    r_k \leq d_k, \forall k \in \mathcal{F},
\]

so that no resources allocated to a flow will be left unused.

**Feasibility of Max-min Fairness:** To verify whether a max-min fair allocation exists in a multi-hop mm-wave network, we first characterise the network’s rate region. We use the notion of conflict graph and denote \( \mathcal{C} \) the set of cliques. Any feasible max-min fair allocation must satisfy:

\[
    \sum_{k,i,j \in C_q} t_{k,i,j} \leq 1, \forall k \in \mathcal{F}, \forall C_q \in \mathcal{C}.
\]

This clique constraint guarantees the total time consumed by all flow segments in a clique does not exceed 1, and can also be expressed as a constraint on the rates, i.e.

\[
    \sum_{k,i,j \in C_q} \frac{r_k}{c_{i,j}} \leq 1, \forall k \in \mathcal{F}, \forall C_q \in \mathcal{C}.
\]

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

**Proof.** Since we consider a SP-based method to schedule transmissions, the channel access in a clique can be seen as single-hop time division multiplex (TDM) instance, which is known to have a convex capacity region [8].

Given that there is no secondary interference between transmissions and the throughput of a sub-flow \( s_{k,i,j} \) in a clique \( C_q \) is upper bounded by the throughput allocated in the preceding clique \( C_{q-1} \) or by 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 is unique.

**Proof.** By [9], max-min fair allocation exists in compact convex sets and if any max-min allocation vector exists, then it is unique.

Finally, the network rate region has the free disposal property [9] since each element of the rate vector \( \mathbf{r} = \bigcup_r k \) 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 in polynomial time.

**Progressive Filling Algorithm:** To achieve max-min fair allocation of the backhaul resources under clique and demand constraints, we specify a progressive filling mechanism which we summarise in Algorithm 1 and whose operation we detail 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 refer to these as active flows. We gradually increase flow rates simultaneously, in steps of size \( \epsilon \) (line 4) until one or more flows either meet their demands (line 6) or activate a clique constraint (line 13). 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 the fractions already reserved for inactive flows (line 18) and sum the inverse of the link capacities corresponding to active flows in that clique. The latter will allow us to provide all active flows with the same rate \( R \) (line 23), while allocating airtimes to each sub-flow that are inversely proportional to the traversed link’s capacity (line 25), i.e.

\[
    t_{k,i,j} = \frac{t_{\text{left}}}{c_{i,j} \sum_{k,l,m \in \mathcal{F}} c_{l,m} (1/c_{l,m})}.
\]

With the above it is straightforward to verify that airtimes \( t_{k,i,j} \) 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).

Having obtained the airtimes to be allocated for each flow on each traversed backhaul link, the remaining task is to coordinate the transmission of base stations to ensure they steer for TX/RX towards the right neighbour at appropriate times and for the computed durations. Next we introduce, WiHaul, a scheduling protocol that implements this.

3. **WIHAUL PROTOCOL**

Terminal deafness is a major challenge in mm-wave networks and unless stations know to which neighbour to steer their beams, when, and for how long, some will be locked out and frame loss will increase, leading to overall performance degradation. To overcome these issues and to convey the airtimes allocated to each sub-flow on each link, and attain max-min fair rates as computed previously, we pro-

```plaintext
Algorithm 1 Progressive Filling

1: \( r_k = 0, \forall k \) \( \triangleright \) Initialisation
2: \( F_0 := \mathcal{F} \) \( \triangleright \) Set of active flows
3: while \( F_0 \neq \emptyset \) do \( \triangleright \) Loop until all flows allocated
4: \( r_k := 1, \forall k \in F_0 \) \( \triangleright \) Increase rates of all active flows with same step
5: for \( \forall f_k \in F_0 \) do \( \triangleright \) Flow satisfied
6: if \( r_k \geq d_k \) then \( \triangleright \) Flow satisfied
7: \( r_k := d_k \)
8: \( F_0 := F_0 \setminus \{ f_k \} \) \( \triangleright \) Remove from active set
9: end if
10: end for
11: for \( q = 1 : |\mathcal{C}| \) do \( \triangleright \) All cliques
12: \( t_{k,i,j} := r_k/c_{i,j} : \forall k \in C_q \) \( \triangleright \) Time consumed by each flow segment in \( C_q \)
13: if \( \sum_{k,i,j \in C_q} t_{k,i,j} \geq 1 \) then \( \triangleright \) Clique constraint
14: \( \Delta t_{\text{left}} = 1 \) \( \triangleright \) Total airtime budget
15: \( S = 0 \) \( \triangleright \) To obtain airtime shares
16: for \( \forall k \in C_q \) do \( \triangleright \) All sub-flows
17: \( t_{\text{left}} = t_{\text{left}} - t_{k,i,j} \) \( \triangleright \) Airtime reserved
18: end for
19: \( S = S + 1/c_{i,j} \) \( \triangleright \) Weight by capacity
20: end if
21: end for
22: \( R = t_{\text{left}}/S \) \( \triangleright \) To use for all active flows
23: for \( \forall f_k \in F_0 \) do \( \triangleright \) All sub-flows
24: \( r_k := R \times t_{k,i,j} \) \( \triangleright \) Rate & airtime
25: \( t_{k,i,j} := t_{\text{left}} / c_{i,j} \) \( \triangleright \) Rate & airtime
26: Freeze \( r_k, F_0 := F_0 \setminus \{ f_k \} \)
27: end for
28: end for
29: end for
30: end while
```
pose a network-wide service period (SP) assignment protocol based on a scheduling hierarchy.

To decide the position of a base station \( i \) in the scheduling hierarchy, WiHaul takes into consideration the following:

1. Hop distance to the gateway, \( H_i \);
2. Conflict state, i.e., \( L_i = 1 \), if node \( i \) is a conflict node, and \( L_i = 0 \), if it is a leaf node;
3. Node’s unique ID (this can be the IPv6 address).

The protocol first considers all conflict nodes eligible candidates for acting as scheduling coordinators. Among these, the one with the lowest hop distance value (\( H \)) is designated as the root coordinator and placed at the top of the scheduling hierarchy, namely at Level 0. During announcement transmission intervals (ATIs) of beacon intervals, the root coordinator collects information including flow demands and link capacities, runs the progressive filling algorithm, and computes the airtimes to be assigned to each flow on the links forming their path. Remaining nodes with \( L_i = 1 \) will be involved in the scheduling, and their level depends on the difference between their \( H \) value and that of the main coordinator, \( H_c \), i.e., \( \text{Level}_i = |H_i - H_c| \).

At each level of the hierarchy, a node accepts the time allocated by its parent and assigns SPs to its children. If two or more nodes on the same level share the same child, the node with the smallest ID 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 periodically switch their beams towards the corresponding neighbours for TX/RX during the assigned SPs.

We illustrate WiHaul’s operation in Fig. 3 for the simple topology in Fig. 1, to which we add a 7th station that could communicate with both 1 and 2 (potential scheduling tie). Here, 4 has the lowest \( H \) value, so it is the root coordinator and a Level 0 node. The adjacent neighbours of 4, i.e., 6, 3, and 5, form up Level 1. Then, both 1 and 2 are connected to one of the Level 1 nodes, so they will be placed on the next level. Finally, the last \( L_i = 0 \) node, i.e., 7, is at the lowest level. WiHaul assigns airtime from the top to the bottom of the hierarchy. In this case, 4 computes the time allocation for each flow segment and assigns SPs to Level 1 nodes. 5 and 6 simply accept the assigned time, and 3 will further allocate time for 1 and 2, avoiding overlapping with the schedule assigned between 3 and 4. On the next level, since 1 has a lower index compared to 2, it will make scheduling decisions for 7, which will inform 5 to avoid conflicts.

In practice, to adapt to the dynamics of physical channel conditions and the changing flow demands, the root coordinator will periodically collect the link and flow information, run the progressive filling algorithm and re-schedule flow segments. This can take place across multiple ATIs, whose periodicity depends on the beacon interval (typically 100ms). The progressive filling algorithm’s runtime is a function of the highest flow rate divided by the step-length, which is configurable, and the total number of flows.

4. RESULTS

To evaluate the performance of the proposed max-min fair allocation scheme for multi-hop mm-wave, we consider the topology shown in Fig. 1 and examine the end-to-end flow throughputs and airtime fractions allocated, using Matlab.

![Hierarchical scheduling structure built for the topology in Fig. 1, to which a 7th node is added, that can communicate with both 1 and 2. SPs allocated within beacon intervals shown on the right. Links (4,6) and (2,3) are simultaneously active.](image-url)

**Table 1: 802.11ad PHY bit rates.**

<table>
<thead>
<tr>
<th>Bitrate [Mb/s]</th>
<th>25</th>
<th>125</th>
<th>626</th>
<th>1250</th>
<th>6756</th>
</tr>
</thead>
<tbody>
<tr>
<td>885</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>1155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>626</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2503</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6756</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We work with three aggregate flows, as follows: Flow 1 between 1 and 6, Flow 2 between 6 and 2, and Flow 3 between 5 and 6. We investigate the effects of heterogeneous demands, dissimilar link qualities, and shared links. Links employ one of the modulation and coding schemes (MCS) specified by the 802.11ad standard and listed in Tab. 1, and we neglect overheads associated with beamform training procedures.

Growing flow demand: First we consider all links operate with MCS 24, except \( l_{1,3} \) and \( l_{4,5} \) which employ MCS 4 and respectively 12. We fix the demand of flows 1 and 2 to 1Gb/s and vary that of flow 3 between 0.5–2Gb/s in 500Mb/s increments. We plot the flow throughputs and fractions of airtime allocated for these in each clique in Fig. 4. Observe that flows traversing low capacity links (e.g. \( f_1 \) and \( f_2 \)) activate the clique constraint in \( C_1 \) before their demands are satisfied. With \( C_1 \) fully utilised, increasing the demand of \( f_2 \) (which is not a member of \( C_1 \)) from 500 to beyond 1,500Mb/s increases channel utilisation in \( C_2 \) from 63% to 100%. Further when a clique is fully utilised, active flows traversing that of assigned equal rates, even if encountering different link capacities. In this scenario, \( f_1 \) and \( f_2 \) receive the same data rates while the first links on their paths, i.e. \( l_{1,3} \) and \( l_{2,3} \) have different capacities. Here \( f_1 \) consumes approximately 80% of the time available in \( C_1 \).

Bottleneck links: Next we show that the quality of the poorest link on a flow’s path has a major impact on the rate distribution in the entire backhaul, while eliminating such a bottleneck may create others. To this end, we consider the same MCSs used in the previous example, but now vary the capacity of \( l_{1,3} \) and examine again the flow throughputs and airtime allocations. We keep demands fixed to 1, 1, and 2Gb/s. The results are depicted in Fig. 5. First, observe that \( f_1 \) and \( f_2 \) initially traverse different links (\( l_{1,3} \) and \( l_{4,5} \)) and then the same sequence before the GW (\( l_{3,4} \) and \( l_{4,6} \)). With high capacity on the latter, the achievable data rate of \( f_1 \) directly depends on the capacity of \( l_{1,3} \) since \( C_1 \) is always fully utilised. In effect \( f_1 \) and \( f_2 \) are assigned equal throughputs. On the other hand, when the data rate on \( l_{1,3} \) increases, the channel time consumed by \( f_1 \) in \( C_1 \) drops from 0.7 to 0.2, which gives \( f_2 \) more opportunities to communicate. The throughputs of \( f_1 \) and \( f_2 \) increase simultaneously with this link capacity increase, until the clique constraint in \( C_2 \) activates. Further increasing \( l_{1,3} \)’s capacity results in less channel time available for \( f_2 \) and as a consequence \( r_3 \) decreases from 2 to 1.5Gb/s, while \( f_3 \)’s demand is no longer met.
flows. For this purpose we consider links have implications on the airtime distribution among and 2Gb/s. Though as some flow segments belong to both cliques, which is shared by which in turn leads to a larger fraction of time allocated to \( f_3 \), and hence larger improvements, \( f_1 \) and \( f_2 \) experience identical link rates, they equally share the airtime available in both cliques. As the quality of \( l_{3,4} \) improves, \( f_1 \) and \( f_2 \) consume less resources in \( C_2 \), which in turn leads to a larger fraction of time allocated to \( f_3 \), and hence larger throughput not only for \( f_1 \) and \( f_2 \), but also for \( f_3 \).

We conclude that max-min fair allocation of end-to-end throughputs in multi-hop mm-wave backhauls is achieved with a non-trivial partitioning of the airtime budget within cliques, which depends on flow demands, their paths, and the capacities of the links traversed.

5. RELATED WORK

Recent studies suggest highly directional mm-wave links can be modelled as pseudo-wired, since collision induced losses are negligible [6]. Consequently, terminal deafness is a key challenge when scheduling transmissions/receptions [3]. Chen et al. propose a directional cooperative MAC protocol, in which the user device selects an intermediate node to relay the packets to the AP, when the multi-hop path exhibits higher SNR than the direct link [10]. However, the deafness issue is not specifically addressed therein, while resource allocation in multi-hop mm-wave backhauls with fairness guarantees is yet to be investigated.

Max-min fairness was first considered for flow control in wired networks [5]. It was subsequently observed that the existence of max-min fairness is a geometric property of the set of feasible allocations, and if a max-min fair vector exists, then it is unique [9]. 802.11 rate region is proven log-convex, and station attempt probabilities and burst sizes in 802.11 mesh networks were derived for max-min fair regimes [7]. However, this holds in multi-channel mesh points communicating with multiple neighbours simultaneously via different interfaces, which is unfeasible with small form factor mm-wave devices equipped with a single interface.

6. CONCLUSIONS

In this paper we addressed max-min fair rate allocation in mm-wave backhauls built upon IEEE 802.11ad by designing a progressive filling algorithm and a hierarchical scheduling protocol to enforce the computed airtime shares and coordinate multi-hop transmissions. Preliminary results suggest max-min fairness requires non-trivial partitioning of the airtime budget in cliques of sub-flows, which depends on flow demands, their paths, and the capacities of traversed links.

7. REFERENCES