On Optimal Hotspot Selection and Offloading

Vandana Mittal*, Sanjit Krishnan Kaul* and Sumit Roy†

* Wireless Systems Lab, IIIT-Delhi, India, † University of Washington, Seattle, WA
\{vandan13162, skkaul\}@iiitd.ac.in, sroy@u.washington.edu

Abstract—Devices like smartphones come with 3G/4G and WiFi radios, which creates possibilities of heterogeneous network access. We investigate scenarios where internet access to a device is available only via the cellular network. However, not every user may connect directly to it. Users in the network may be split into hotspots and clients. Hotspots are the users that connect directly to the cellular network and may provide connectivity to the internet to other users by allowing them to connect to their WiFi interface. Clients connect to the cellular network only via hotspots. The optimization problem is to find the split of hotspots and clients, and the association between clients and hotspots, that maximizes the sum of the link rates of users. Importantly, the users must get at least the link rate they get when all are directly connected to the cellular network. In this paper, we formulate the optimization problem. We provide insights into the interplay of WiFi connectivity amongst users, their link rates to the cell tower, and the split that maximizes sum rate. We propose a novel heuristic approach to split the network. Median gains of 1.5× are observed over networks of up to 40 nodes.

I. INTRODUCTION

Smartphones come with WiFi radios and can connect to cellular networks. Further, their WiFi radios can be configured as hotspots. That is, a smartphone while connecting to the internet over the cellular network, can function as a WiFi access point for other smartphone users in its proximity. These other users become clients of the WiFi hotspot and access the internet via the hotspot’s connection to the cellular network instead of using their own cellular connections.

Prior works [1] [2] have looked at offloading the data traffic generated by cellular users to WiFi AP(s) in their vicinity. However, the WiFi AP chosen for offloading a user accesses the internet independently of the cellular network to which the user is connected, say via cable or fiber at home or a cafe. In this work, we consider scenarios where internet access is only via the cellular network.

We define the baseline network configuration to be the one in which all users access the internet using their own link to the cell tower. The possibility of gains in link rates of users on leveraging the said heterogeneity in available networks is exemplified by the network topology in Figure 1a. The sum rate of the baseline network is the sum of the rates of the three links. We abstract the cellular MAC as a perfectly fair time-sharing or TDMA system, as is reasonable for a scheduled LTE-like system [3]. Each smartphone gets a rate that is the link’s Shannon rate divided by the total number of links connected to the cell tower. Links with 30 dB SINR (signal-to-interference-and-noise-ratio) will thus get a baseline rate of 1/3 (log₂(1 + 1000)) = 3.32 bits/sec/Hz each. The sum rate of the baseline network is 1/3 log₂(1 + 1000) + 1/3 log₂(1 + 10) = 5.63 bits/sec/Hz.

Now consider the alternative where one of the smartphones with 30dB link SINR turns into a hotspot to which all other phones connect as WiFi clients. All access to the internet is now via the chosen hotspot’s link to the cellular tower. Therefore, the maximum achievable sum rate is the Shannon rate of this link, which is log₂(1 + 1000) = 9.96 bits/sec/Hz, which is a rate improvement of 77%. Under the assumption that the WiFi links between the clients and the hotspot are not rate bottlenecks, this improvement can be achieved while ensuring that all nodes get at least their baseline rates. Finally, note that the increase in sum rate would not occur if a smartphone that has a SINR of 10 dB was chosen to be a hotspot.

Our specific contributions are as follows: (a) We formulate the sum rate maximization problem for a network of nodes (smartphones/users) that can either connect to the internet directly via the cellular network or can connect via other nodes that are configured as WiFi hotspots. We enforce the constraint that all nodes must get at least their baseline rates. (b) We provide a detailed exposition of the attributes of the problem and insights into how gains in rate can be obtained from the said heterogeneity. These motivate a novel heuristic approach to split a network into hotspots and clients. (c) We evaluate the heuristic approach over different WiFi connectivities amongst nodes, and cellular SINR(s) of the nodes. We obtain median gains of 1.5× in sum rate over the baseline network configuration.

We will consider the case when all users connect to the internet via a single cell tower. Extensions to multiple cell...
towers are deferred to the future. We will use simple abstractions for WiFi connectivity amongst nodes. This helps derive insights into the offloading problem, which should apply to other out-of-band device-to-device networks too. Aspects like mobility, energy consumption, and implementation, are outside the scope of this work.

We describe related works in Section II. In Section III we formulate the optimization problem. In Section IV we exemplify the main attributes of the problem. Section V details the proposed heuristic approach. Section VI details the evaluation methodology and results. We conclude in Section VII.

II. RELATED WORK

The work in [4] focuses on the placement of a fixed number of Mobile backbone nodes (MBN) and the assignment of regular nodes (RN) to MBN(s). Unlike our problem, the regular nodes in their network cannot directly connect to the internet. Hence, they cannot act as MBN(s) for other regular nodes. Also, we do not a priori fix the number of hotspots (similar to MBN(s) in [4]).

Authors in [5] propose an approach that enables collaboration by using bandwidth of available access points to serve local as well as non-local users in high access point density wireless LAN(s). The access points are known a priori and users are connected to these access points optimally.

There are several works on offloading cellular data traffic using small cells [1] [6] [7] and WiFi [2] [8] [9]. Unlike our work, they aim to offload cellular connections to networks that have an independent connection to the internet.

In [10], mobile hosts that suffer from poor channel quality use peer-to-peer links to access proxy clients with better channel quality. While a proxy can be up to three hops away, the work does not capture the general problem of network reorganization and how it must be done such that it guarantees benefits to all nodes in the network.

Authors in [11] propose to form clusters among cellular users who are in vicinity and can use device-to-device communication. The cluster heads (hotspots) are selected opportunistically. Also, they do not ensure that no nodes suffer as a result of cluster formation.

III. OPTIMIZATION PROBLEM

Consider a network $N$ of $N$ users/nodes (smartphones), indexed as $i = 1, 2, \ldots, N$. User $i$ has a cellular link (to the tower) with a signal-to-interference-and-noise ratio of $\text{SINR}_i$. We will not distinguish between the cellular uplink and downlink of a user. Also, we will assume that all users have access to the same amount of cellular bandwidth. Without loss of generality, let the bandwidth be 1 Hz. Each user gets access to the bandwidth for a fraction $1/N$ of the total time [3]. The baseline rate $R_i^{(B)}$ of user $i$ is therefore

$$R_i^{(B)} = \frac{1}{N} \log_2(1 + \text{SINR}_i).$$

Figure 1b illustrates how the link of node $i$ (a hotspot) to the cell tower is shared by itself and another node $j$ that is a client of $i$. Let $R_{ij}$ be the rate reserved for user $j$ on the link to the cell tower of user $i$, when $i$ is configured as a hotspot and $j$ is a client of $i$. Implicitly, $R_{ii}$ is the rate reserved for hotspot $i$ on its own link. A hotspot is its own client.

Note that a node $j$ can get a rate $R_{ij}$ to the tower via its WiFi link to $i$ only if the WiFi link supports the rate. WiFi uses a carrier sense and collision avoidance based multiple access. The rate that can be supported between any two nodes will therefore depend on the number of hotspots that occupy interfering channels, and the number of clients connected to the hotspots. In this work, we will assume a simplified model for a WiFi link, in which a WiFi link has a binary state. It is either ON or OFF and is a known constant for a link.

Specifically, if the WiFi link between nodes $i$ and $j$ is ON, then if node $j$ connects as a client to node $i$, its rate share $R_{ij}$ will always be supported by the WiFi link between $i$ and $j$, irrespective of how nodes are split into hotspots and clients. That is the WiFi link between these will never be a bottleneck. Else, if the WiFi link is OFF, it supports a rate share of zero and is always a bottleneck. Only if the WiFi link between nodes $i$ and $j$ is ON, node $j$ can be a client of node $i$ and vice-versa. We will assume that the forward and reverse WiFi links between any two nodes have the same state.

Let $C$ be the set of all $(i, j)$ such that the WiFi link between node $i$ and node $j$ is ON. In Figure 1b node $j$ is a client of node $i$. Thus $(i, j) \in C$. A node $j$ must connect to the internet either via its own link to the tower or via exactly one other node’s link to the tower. That is if a node is a client, it must be a client of exactly one hotspot. Let $a_{ij}, i = 1, \ldots, N, j = 1, \ldots, N$, be variables that indicate assignment of clients to hotspots. Specifically, for $i \neq j$, $a_{ij} = 1$ when $i$ is a hotspot.
and $j$ is its client. It is zero otherwise. Also, if $i$ is a hotspot, $a_{ii} = 1$, else $a_{ii} = 0$. Finally, if $(i,j) \notin C$, then $a_{ij} = 0$.

The rate $R_j$ of user $j$ can thus be written as $R_j = \sum_{i=1}^{N} a_{ij} R_{ij}$. If $j$ is chosen to be a hotspot, we have $a_{ij} = 1$, $a_{kj} = 0$ for all $k \neq j$, and $R_j = R_{jj}$. That is the rate that $j$ gets is equal to its share on its link to the cell tower. If instead $j$ is chosen to be a client of $i$, then $a_{ij} = 1$, $a_{kj} = 0$ for all $k \neq i$, and $R_j = R_{ij}$. That is the rate that $j$ gets is equal to its share on the link of $i$ to the cell tower. We want to maximize the sum rate $\sum_{j=1}^{N} R_j$ of the network, while ensuring that all users get at least their baseline rates. The problem is

Maximize: $\sum_{j=1}^{N} \sum_{i=1}^{N} a_{ij} R_{ij}$, \hspace{1cm} (2)

subject to: $R_j \geq R_{j}^{(B)} \forall j$, \hspace{1cm} (3)

$\sum_{j=1}^{N} R_{ij} a_{ij} = \frac{a_{ii}}{H} \log_2(1 + \text{SINR}_i) \forall i$, \hspace{1cm} (4)

$\sum_{i=1}^{N} a_{ij} = 1 \forall j$, \hspace{1cm} (5)

$a_{ij} \in \{0,1\} \forall i,j \text{ s.t. } (i,j) \in C$, \hspace{1cm} (6)

where the number $H$ of hotspots is given by $H = \sum_{i=1}^{N} a_{ii}$.

The variables of optimization are $R_{ij}$ and $a_{ij}$ for all $(i,j) \in C$. The SINR(s), $\text{SINR}_i$, $i=1,\ldots,N$, of the cellular links and the set $C$ are assumed to be known. Constraint (3) enforces that the optimal solution must be such that each user gets at least as much rate as the user was getting in the baseline case.

Constraint (4) ensures that all clients of a hotspot get rates such that their sum is equal to the rate of the cellular link of the corresponding node. Constraint (5) enforces that a client must be connected to exactly one hotspot and that a hotspot cannot be connected to another hotspot.

Constraints (3)-(5) further ensure that every client is connected to a hotspot and not to another client. That is if $j$ and $k$ are clients then $a_{jk} = 0$.

It is instructive to note that the utility function in (2) is equal to the sum, over all nodes $i$ in the network, of the right-hand side of equation (4). Thus, we are choosing hotspots such that the total rate to the cell tower is maximized, while all nodes get at least their baseline rate. Also, note that if the total rate to the cell tower via hotspot $i$ is greater than or equal to the sum of the baseline rates of all nodes $j$ for which $a_{ij} = 1$, then for such nodes $j$, $R_{ij}$ can be selected so as to satisfy constraints (3) and (4).

In the following, we will call the network that results from this optimization as the hotspot network.

### IV. Problem Attributes

The problem (2)-(6) is a mixed integer non-linear program. In this section, we will illustrate the characteristics of the problem and its solution, using networks of small size. These will motivate our heuristic approach explained in Section V.

Example networks are shown in Figure 2. The networks are shown as graphs. Vertices correspond to nodes in the network. The label on a vertex is the SINR (dB) of the cellular link of the corresponding node. We will also use this label to refer to a node. The graphs with circular vertices, see Figures 2a-2e, show example WiFi connectivity graphs amongst nodes in the network. An edge between two vertices implies that the WiFi link between the two users is ON. Else, the WiFi link between the users is OFF. The graphs with square vertices, see Figures 2a′-2e′, show hotspot networks. Vertices shaded green are the hotspots. They connect directly to the cell tower. The unshaded vertices are the clients. Edges connect (over WiFi) hotspots and their clients. The graphs in Figures 2a′, 2b′, 2c′, 2d′, 2e′ are the optimal hotspot network configurations obtained via an exhaustive search, respectively, for the networks in the Figures 2a, 2b, 2c, 2d, 2e.

When nodes in the network are a single cluster of good WiFi connectivity: Figure 2a shows a network of users that have good WiFi connectivity (all WiFi links are ON) amongst each other. The set of users is $N = \{2, 3, 7, 8, 9, 10, 12, 17\}$. The cellular SINR(s) of the users range from 2 dB to 17 dB. In the absence of constraints (3)-(6), the sum rate to the cell tower is maximized by simply allowing the user with the largest cellular SINR (user indexed 17) to access the cellular bandwidth all the time. The resulting sum rate is $\log_2(1 + 10^{17/10})$ bits/sec/Hz.

Baseline rates of the nodes $R_{i}^{(B)}$ can be calculated using equation (1). We have $N = 8$. The baseline rates (rounded to two decimal places) of the nodes are, in the increasing order of node index, 0.17, 0.20, 0.32, 0.36, 0.40, 0.43, 0.51, 0.71 bits/sec/Hz. The sum rate of the baseline configuration is 3.0976 bits/sec/Hz. Since all WiFi links are ON, user 17 can act as a hotspot for the rest. The resulting hotspot network is shown in Figure 2a′. The resulting sum rate is $\log_2(1 + 10^{17/10}) = 5.68$ bits/sec/Hz, which is also the maximum sum rate achieved in the absence of constraints. Thus the hotspot network with 17 as a hotspot and all others its clients is sum rate optimal. The percentage gain in sum rate over baseline is 83%. Gains in sum rate are due to the fact that while the total time share available to the nodes is the same as in baseline, they now have access to a higher SINR link to the tower.

We now verify that the constraints (3)-(6) are satisfied. We have $a_{17,ij} = 1$ for all $j$ and $a_{ij} = 0$ for all $j$ and for $i \neq 17$. Also, $R_{ij} = 0$ for all $i \neq 17$ and the number of hotspots $H = 1$. For $i = 17$ the $R_{ij}$ for all $j$ can be chosen such that $\sum_{j \in N} R_{ij} = \log_2(1 + 10^{17/10})$ (constraint (4)) and $R_{ij} \geq R_{ij}^{(B)}$ (constraint (3), note $R_{ij} = R_{j}$). The $R_{ij}$ may be selected in multiple ways. An example selection of $R_{ij}$ is as follows.

Reserve a rate of $R_{j}^{(B)}$ for all nodes $j$. A total of 3.0976 bits/sec/Hz, which is equal to the sum of baseline rates, will be reserved. Split what remains, which is $\log_2(1 + 10^{17/10}) - 3.0976 = 2.59$ bits/sec/Hz, equally amongst all nodes. The resulting rates are, in increasing order of node index, 0.5, 0.5, 0.6, 0.7, 0.7, 0.8, 0.8, and 1.0 bits/sec/Hz. All nodes see rates larger than their baseline rates. We will use this method of selection for the examples that follow.
When we have more than one cluster of good WiFi connectivity: Consider the network in Figure 2b. We have two clusters, \( C_1 = \{2, 3, 7\} \) and \( C_2 = \{8, 9, 10, 12, 17\} \), of good WiFi connectivity. WiFi links amongst all nodes within a cluster are ON, where as those between different clusters are OFF. Note that constraints can be satisfied only if we have two or more hotspots (at least one per cluster). Figure 2b’ shows the optimal hotspot network configuration, arrived at via an exhaustive search, which has nodes with SINR(s) 7 dB and 17 dB as hotspots. Both hotspot selections have maximum SINR(s) within their clusters. In the optimal network (Figure 2b’), hotspot 17 and its clients get half the cellular time resource. In the baseline configuration they got a larger share of 5/8 of the time. Hotspot 7 and its clients get half the time instead of 3/8 of time they got under baseline. The time share of \( C_2 \) reduces. However, the SINR of 17 is large enough to satisfy baseline requirements of all nodes in \( C_2 \). The time share of \( C_1 \) increases and the nodes now have access to higher SINR link of node 7. Thus the sum rate of the network increases. In fact, no other feasible (those that satisfy all the constraints (3)-(6)) hotspot selections can do as well. This is because any other feasible selection of hotspots will involve either more hotspots (smaller time shares to the tower) and/or hotspots with smaller SINR(s).

The sum rate of the optimal hotspot network is 4.13 bits/sec/Hz, which is \((5.0 \log_2(1 + 10^{17/10})) + 0.5 \log_2(1 + 10^{7/10}))\). It is a 33% improvement over the sum rate (3.0976 bits/sec/Hz) of the baseline configuration. Selecting \( R_{ij} \) as before (performing the procedure shown earlier, for each hotspot), users get the rates, in increasing order of node indices, 0.4, 0.4, 0.4, 0.5, 0.5, 0.5, 0.6, and 0.8 bits/sec/Hz.

We note that the above result in which the optimal hotspot network is obtained by choosing the highest SINR node in each cluster as a hotspot for nodes in the cluster is not true in general even for networks with two clusters. Figure 2c slightly modifies the network in Figure 2b. While the WiFi connectivity remains unchanged, the SINR(s) of the users in the larger cluster \( C_2 \) have a smaller spread. The hotspot configuration in Figure 2b’ can no longer satisfy all the constraints. Specifically, it is easy to verify that the rate of node 17 to the cell tower, when it gets access to the cellular bandwidth half the time as in Figure 2b’, is smaller than the sum of the baseline rates of all the nodes in its cluster. That is \( \sum_{j \in C_2} R_j^{(B)} < 0.5 \log_2(1 + \text{SINR}_{17}) \). Nothing has changed with respect to the smaller cluster \( C_1 \) and so node 7 can act as a hotspot for it as long as 7 gets half the time.

Since, given half the time, 17 is not a feasible hotspot for the larger cluster, none of the other nodes in the cluster, given that their SINR(s) are smaller than 17, can be the only hotspot for the larger cluster. It turns out that in the optimal network the three users 17, 16, and 7 are hotspots. In this configuration, the cluster containing 17 gets 2/3 of the time as it has two hotspots 17 and 16, instead of the 1/2 it got when 17 was chosen to be the only hotspot for the cluster. So while on average the nodes in \( C_2 \) have access to a smaller SINR link (16 dB and 17 dB instead of all accessing a link with 17 dB SINR), they have access to a larger time share. Also, 7 as a hotspot can support at least baseline rates of the nodes in the smaller cluster, even when it gets only 1/3 of the time. The network is shown in Figure 2c’. The resulting gains in sum rate over the baseline are 26%. The user rates can be calculated as before.

Unlike Figure 2c, the network in Figure 2d has a node 11 that is common to both the clusters. While the optimal network for Figure 2c is feasible for the network in Figure 2d, it is not optimal. An optimal configuration is shown in Figure 2d’. The choice of 17, 16, and 11 as hotspots, instead of the feasible choice of 17, 16, and 7, leads to sum rate gains of 37% over baseline instead of 26%. Therefore, assigning 11 as a hotspot instead of making it a client improves the gains by about 60%.

Finally, consider the network in Figure 2e. It retains the larger cluster of Figure 2c. However, its smaller cluster has nodes with larger cellular SINR(s) and a smaller spread in the SINR(s). Using arguments we have made for the examples above, it can be shown that the baseline configuration is sum rate optimal. However, it is instructive to consider the feasibility of choosing two hotspots (6 and 7) for the smaller cluster and three for the larger cluster (13, 16, and 17). Each hotspot gets 1/5 of the time. The sum rate of hotspots of the larger (resp. smaller) cluster is greater than the sum of baseline rates of the nodes in the larger (resp. smaller) cluster. However, no feasible assignment of nodes in the smaller cluster to the hotspots 7 and 6 exists. For a feasible assignment to exist, we must be able to assign 5 to one of 6 and 7. At 1/5 of the time, node 7 has a rate of 0.518 bits/sec/Hz to the tower and node 6 has a rate of 0.463 bits/sec/Hz. The baseline rates of 5, 6, 7 are 0.257, 0.29, 0.324 bits/sec/Hz respectively. Node 5 cannot be assigned to 7 (and hence 6) as the link of 7 to the tower has a rate of 0.518 bits/sec/Hz, which is less than the sum 0.58 bits/sec/Hz of baseline rates of 5 and 7.

In summary, clusters (groups of nodes with good WiFi connectivity) can not be optimized independently of others, as configuring a cluster into hotspots and clients impacts the possible configurations of other clusters and vice versa. For a node to be a hotspot for other nodes, not only its SINR, but also the SINR of the other nodes and the total share of time the nodes get in the final network configuration are important. Finally, when a node can play the role of either a hotspot or a client, its assigned role may significantly impact the sum rate.

V. Heuristic Approach

Our basic approach is summarized by the following three steps. (a) Fix the number \( H \) of hotspots. Start with \( H = 1 \). This fixes the fraction of time \((1/H)\) and allows us to calculate the rate of a node’s link to the cell tower if it were to be one of the \( H \) hotspots. Note that since \( H \) is fixed, this sum rate is only dependent on the cellular SINR of a node. (b) Choose a \( H \)-hotspot network as follows. For every node, calculate the set of nodes (clients) for which it can be a hotspot, given that it has a fraction \( 1/H \) of the time to the tower. We propose a simple cellular SINR-based heuristic to do the same. Next, greedily pick \( H \) nodes as the hotspots in decreasing order of
The algorithm accepts the network hotspots, which are indexed approach. Let \( i \in H \) the selected number feasible hotspot networks found by the approach. the number of clients they can support. The workings of the Algorithm 2: Configure-Network

Input: \( N, W \)
Output: \( H' \) - Hotspot Network Configuration
for \( H = 1 \) to \( |N| \) do
\[ H = H \cup H'(1); \] \( H' \) is the hotspot corresponding to the set \( U_h \in V_h : \)
\[ \{ H, R, U_c \} \rightarrow 0 \] Select hotspots (\( N', W, H, H' - 1 \)).
do
1 = \( c + 1 \);
\( H = H \cup H'; \) if \( R = 0 \) then
1) \( R_h = R \) of hotspots and clients, the algorithm also returns
\( \mathcal{O} = \{ \rho : W_{\rho} = 1, \rho \in N' \}; \) \( R = \sum_{\rho \in \mathcal{O}} \log_2 (1 + \text{SINR}_\rho); \)
while \( A < R \) do
\( W_{\rho}^{H} = \emptyset; \) \( \mathcal{O} = \{ \rho : W_{\rho}^{H} = 1, \rho \in N' \}; \) \( R = \sum_{\rho \in \mathcal{O}} \log_2 (1 + \text{SINR}_\rho); \)
end
\( H' = \arg \max_{H \in \{1, \ldots, |N|\}} R_h; \)
\( H' = \{ h \in \mathcal{H} \mid \text{SINR}_h > \text{SINR}_h \}; \) \( | N' | \) to \( N \).
for \( h = 1 \) to \( H' \) do
\( n^* = \arg \max_{n \in N' \setminus \{ \}} \{ \rho : W_{\rho} = 1 \}; \) \( H = H' \cup n^*; \)
\( U_h = \{ u : W_{\rho}^{H} = 1 \}; N' = N' - U_h; \)
end
Algorithms: Algorithm 1 is the starting point of our approach. Let \( W \) be the WiFi connectivity matrix. We have \( W_{ij} = 1 \) for \( (i, j) \in \mathcal{C} \) and when \( i = j \), otherwise \( W_{ij} = 0 \). The algorithm accepts the network \( N \) and \( W \) as inputs. For the selected number \( H \), \( \mathcal{H} \) is the corresponding set of selected hotspots, which are indexed \( 1, \ldots, H \). \( \mathcal{U}_h \) is the set of clients of hotspot \( i \). Note that \( i \in \mathcal{U}_h \). \( R_h \) is the sum rate of the selected \( H \)-hotspot network. The set \( \mathcal{R}' \) stores the nodes that were not assigned to any hotspot. For a given \( H \), the sets \( \mathcal{H}, \mathcal{U}_1, \ldots, \mathcal{U}_H \) are initially empty and the set \( \mathcal{R}' = N \).

The algorithm attempts to configure a \( H \)-hotspot network by up to \( H \) calls of Select-Hotspots (Algorithm 2). The need for multiple calls is explained later. The first time it is called for a given \( H \), we set \( H' = H \) and \( N' = N \).

Select-Hotspots takes a given set \( N' \subseteq N \) of nodes, the target number \( H \) of hotspots in the hotspot network, the number \( H' \) that remains to be selected, and \( W \). It uses the SINR-based heuristic to decide for every node \( n \in N' \) the nodes in \( N' \) that can become its clients when it has a time-share of \( 1/\text{H} \) to the cell tower. The set \( \mathcal{O} \) is initialized to contain \( n \) and all nodes that have an edge to \( n \) in the WiFi connectivity graph. The heuristic helps prune this set so that the sum of the baseline rates of all nodes in the resulting set are smaller than the rate of \( n \)'s link to the tower. As per the heuristic, nodes (in \( \mathcal{O} \) but not including \( n \) itself) that have a larger cellular SINR are removed first from the set \( \mathcal{O} \). This is motivated by the fact that, everything else equal, a node with larger cellular SINR is a more desirable hotspot.

After prospective clients for every node in \( N' \) have been finalized, we look for \( H' \) nodes that together provide internet connectivity to all nodes in \( N' \). The \( H' \) nodes are chosen greedily in decreasing order of the number of their prospective clients. In addition to returning the set \( \mathcal{H} \) of hotspots and the corresponding sets of clients, the algorithm also returns the set of nodes \( \mathcal{R}' \) that were not covered in the above sets. If the set is nonempty, then the selected hotspots do not provide connectivity to all nodes in \( N' \). That is the selection of hotspots is infeasible. This infeasibility may often result from the greedy decision making. To improve the chances of finding an existing feasible selection, Select-Hotspots is called multiple times in the manner explained next.

In general, the \( c^\text{th} \) time Select-Hotspots is called, it returns \( H' = H - c + 1 \) hotspots in the set \( \mathcal{H}' \) and their set of clients \( U_c', \ldots, U_H' \). Note that prior to the \( c^\text{th} \) call, \( c < H \) hotspots and their clients have already been selected. All the selected hotspots are stored in \( \mathcal{H} \). Let \( \mathcal{R}' \) be the set of nodes in \( N' \) that are not provided connectivity by the \( H' \) hotspots. If \( \mathcal{R}' = \emptyset \), the hotspot that was selected first amongst the \( H' \) hotspots, denoted by \( \mathcal{H}'(1) \), and its clients, that is the set \( U_{c-1} \), are added to the existing hotspot network configuration. In the next iteration, the rest of the hotspots and clients (the set \( \cup_{H=H}^{H=H} U_h \)) and the nodes in \( \mathcal{R}' \) are sent back to Select-Hotspots. If all nodes are not covered after \( H \) calls of Select-Hotspots, a \( H \)-hotspot network is assumed infeasible.

Figures 3b-3d exemplify the workings of the proposed algorithm for the network of nodes in Figure 3a, which shows the WiFi connectivity amongst the nodes. Figure 3b shows the results of setting \( H' = H = 1 \). It turns out that each node (top row) can support all nodes (bottom row) that have edges to it in the WiFi graph. Also, 13 can provide connectivity to all nodes in the network. So \( H = 1 \) is feasible. However, \( H = 2 \) gives a larger sum rate. Figure 3c shows the result of calling Select-Hotspots the first time for \( H = H' = 2 \). For each node in the network (top row), nodes (bottom row) that are connected...
by a dashed or a solid line are nodes that have good WiFi connectivity with the selected node. Nodes that have a dashed line are the ones that are removed from the set of prospective clients using the SINR-based heuristic. The greedy method of selecting hotspots will first select 13 as a hotspot, since it supports the largest number (two) of clients, followed by 15 (largest SINR amongst those supporting one client). The selection of 13 and 15 will leave 14 unassigned. It will be in the set $R'$. We choose 13 as one of the hotspots with 8 and 10 as its clients and call Select-Hotspots over the nodes 15 and 14, with $H = 2$ and $H' = 1$, with the goal of adding one more hotspot. In Figure 3d, to the left of the arrow, we have a hotspot network configuration created in part, together with nodes 14 and 15 with their prospective clients. To the right, we have the final hotspot network configuration. Since 15 and 14 have the same connectivity, 15 is chosen as a hotspot due to its larger SINR.

VI. EVALUATION METHODOLOGY AND RESULTS

We evaluate our proposed approach over simulated networks consisting of 5, 8, 10, 20, and 40 nodes. Having chosen the number $N$ of nodes in the network, we next choose (a) WiFi connectivity between them, (b) their cellular SINR(s).

WiFi networks typically have small coverage. We simulate WiFi connectivity between the $N$ nodes by first splitting them into clusters of good WiFi connectivity and then adding ON WiFi links between nodes in different clusters with a certain chosen maximum probability of connectivity between clusters. For 5 nodes, we choose single WiFi clusters. For 8, 10, 20 and 40 nodes, respectively, we choose up to 2, 2, 4, and 8 clusters. Nodes are assigned randomly to clusters.

We choose the maximum probability of connection between clusters for a simulated network to be one of 0, 0.1, 0.2. If the chosen maximum is $p$, then each WiFi link between nodes in different clusters is switched ON with probability chosen uniformly and randomly between (0, $p$).

The cellular SINR(s) are selected by first selecting a mean SINR in the range of (5, 30) dB for the network. The chosen mean is then perturbed for each node in the network, independently and randomly, by adding to it a normal random variable (as in the log-normal shadowing model) with zero mean and a chosen standard deviation $\sigma$ (chosen from 4, 8, 12 dB).

We evaluate our approach over a total of 90, 190, 400, and 940 networks, respectively, containing 5, 10, 20 and 40 nodes. Figure 4a shows the gains in sum rate over baseline obtained for networks of size 5, 10, 20, and 40. Median gains of 1.5× are obtained. Figure 4b shows the gains obtained for networks of 40 nodes as a function of mean cellular SINR and Figure 4c shows the gains as a function of the SINR spread $\sigma$. We observe that percentage gains increase as mean SINR decreases and the spread increases. Nodes in networks that have small SINR(s) are more likely to benefit from hotspot configurations as the hotspot configuration effectively leads to better and fewer SINR links to the cell tower, and given the small SINR of a node, the resulting gains in SINR (power gain) lead to large improvements in sum rate to the cell tower.

Similarly, in networks with large SINR spreads, a hotspot network configuration gives the nodes that have smaller SINR(s) the large rate benefits of becoming clients of those that have much larger SINR(s). Finally, Figure 4d shows gains as a function of number of clusters. As is expected gains reduce as the number of clusters increase. However, median gains of 45% are obtained even when the number of clusters is 8, which on an average is 5 nodes per cluster.

VII. CONCLUSIONS

We formulated the sum rate optimization problem that splits users connected to the cellular network into users who will be configured as hotspots and users who will access the internet by becoming clients of the hotspots, under the constraint that all users must at least get the rate they were getting when each one of them was directly connected to the cellular network. We proposed a heuristic approach to solve the problem. Median gains of 1.5× were obtained over networks of up to 40 nodes.

REFERENCES


