Mobile Hotspot Selection and Offloading

Student Name: Vandana Mittal

IIIT-D-MTech-ECE-13162
July 30, 2015

Indraprastha Institute of Information Technology
New Delhi

Thesis Committee
Dr. Sanjit Krishnan Kaul (Chair)
Dr. Pravesh Biyani
Dr. Sumit Roy

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Electronics and Communication

This research was partially funded by ITRA project, funded by DeitY, Government of India, under grant with Ref. No. ITRA/15(57)/Mobile/HumanSense/01.

Keywords: Hotspot, Offloading, Throughput, Optimization

Certificate

This is to certify that the thesis titled **"Mobile Hotspot Selection and Offloading"** submitted by **Vandana Mittal** for the partial fulfillment of the requirements for the degree of *Master of Technology* in *Electronics and Communication & Engineering* is a record of the bonafide work carried out by her under my guidance and supervision at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Sanjit Krishnan Kaul
Indraprastha Institute of Information Technology, New Delhi

Abstract

The adoption of smartphones, their ability to configure as mobile hotspots (WiFi access points), together with access to cellular networks, creates novel possibilities of heterogeneous network access for users. Specifically, a smartphone user may access a cell tower directly using the phone's 3G/4G connection or may connect to another smartphone in its vicinity that is configured as a hotspot. Many prior works [18] [20] have looked at offloading the data traffic generated by cellular users to WiFi AP(s) in their vicinity. In these works the WiFi AP chosen for offloading a user, accesses the internet independently of the cellular network to which the user is connected. For example, the chosen WiFi AP could be in the user's home or in a cafe that may have connectivity to the internet via cable or fiber.

In our work, we restrict ourselves to scenarios where internet access 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 provide connectivity to the internet to other users by allowing them to connect to their WiFi interface. Clients do not access the cellular network directly. Instead each client connects to a hotspot. The optimization problem is to find the split of hotspots and clients, and the association between clients and hotspots, that maximizes network throughput while ensuring that users get at least the throughput they get when all are directly connected to the cellular network. In this paper, we formulate the optimization problem, detail its characteristics, and propose a novel heuristic approach to split the network. We evaluate the gains in throughput achieved by using the approach for networks containing up to 100 nodes. Median gains of 1.5× are observed.

Acknowledgments

I dedicate this work to my parents and grandparents.

After expressing gratitude towards God and my loving parents, I would like to thank my advisor Dr. Sanjit Krishnan Kaul who has always supported me throughout the thesis work with his knowledge and patience. His constant support, encouragement and belief in my abilities is the sole factor behind the successful completion of this work. His invaluable help, dedication and perfection for the work always inspired me to work hard. I could not have imagined having a better advisor and mentor for my Masters research. I will always be grateful to him throughout my life.

I sincerely thank Dr. Sumit Roy from University of Washington for accepting to be a part of my thesis committee as the external examiner and giving the valuable feedback and constructive comments. I would also like to thank Dr. Pravesh Biyani for accepting to serve as the internal examiner on my thesis committee and for discussions.

I thank Dr. Rajiv Raman and Dr. P.B. Sujit for discussing problems. I would also thank Dr. Amar Prakash Azad for his valuable inputs.

I thank Dr. Pankaj Jalote and all founding members of IIIT Delhi for creating such a wonderful institute and giving me an opportunity to study here.

I would like to thank all my friends and PHD members of MUC lab for being with me at each step when I need their support. Special thanks to Rekha Tokas and Manoj Gulati for their invaluable help and moral support. In the end, I would like to once again thank the Almighty for adding this wonderful chapter to my life.

Contents

List of Figures

Chapter 1

Introduction

Increasing penetration of smartphones [7] and access to cellular networks provides users a near ubiquitous opportunity of heterogeneous access to the internet, especially in regions where users are in proximity of each other (for example, public transport). Smartphones come with WiFi radios. They also have the ability to connect to cellular networks. Further, their WiFi radios can be configured as mobile 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 can 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.

The possibility of gains in network throughput on leveraging the said heterogeneity, is exemplified by the network topology in Figure 1.1. We will assume that a link can achieve rates given by the Shannon's capacity formula and that a link always has a packet to send. Also, in this work, we will not distinguish between the uplink and the downlink. The throughput of the network is the sum of the throughputs of the five links. Each smartphone gets a throughput that is the link's Shannon rate divided by the total number of links connected to the cell tower [19]. Links with 30 dB SINR (signal-to-interference-and-noise-ratio) will thus get a throughput of $1/5$ $(\log_2(1 + 1000)) = 1.99$ bits/sec/Hz each. The network throughput is 6.06 bits/sec/Hz.

Now consider the alternative where one of the smartphones with 30dB link SINR turns into a mobile 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 network throughput is the Shannon rate of this link, which is $\log_2(1+1000) = 9.96$ bits/sec/Hz, which is a throughput improvement of 64%. Under the assumption that the WiFi links between the clients and the hotspot are not throughput bottlenecks, this improvement can be achieved. Finally, note that the increase in throughput would not occur if a smartphone that has a SINR of 10 dB was chosen to be a mobile hotspot.

In general, we want to partition the set of nodes (smartphones/users)[1] in a network into mobile

---

[1] We will use nodes and users interchangeably.

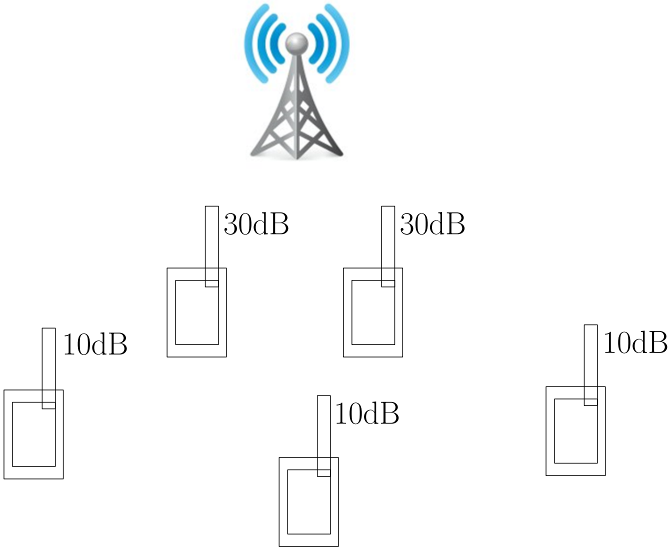


Figure 1.1: Example network topology of 5 smartphones connected to a cell tower. The signal-to-interference-and-noise ratio of each phone's link to the tower is mentioned besides it.

hotspots and clients of hotspots, and assign clients to hotspots, such that the network throughput is maximized. Our specific contributions include:

1. We formulate the network throughput maximization problem for a network of nodes that can either connect to the internet directly via the cellular network or can connect via other nodes that are configured as a mobile WiFi hotspot. The maximization is carried out under the constraint that all users must get at least the throughput they were getting when all of them were directly connected to the cellular network. The problem is a mixed integer non-linear program.
2. We propose a novel heuristic approach that is motivated by a detailed exposition of the attributes of the problem. First, we use the approach to optimize a network of nodes connected to the same cell tower. Next, we propose an extension that leverages WiFi connectivity between nodes connected to different cell towers to optimize a network in which clients are connected to more than one cell tower.
3. We evaluate the proposed heuristic approach over a range of WiFi connectivity between nodes, and cellular SINR(s) of the nodes in the network. In our evaluations, we obtain median gains of 1.5× in network throughput, for networks of up to 100 nodes and one cell tower, over the default network configuration in which all nodes connect to the internet using their own connection to the cellular network.

We assume that the network operator will solve the throughput maximization problem and will configure nodes connected to the cellular network into hotspots and clients. Specifically, in this work, we do not consider the possibility of user cooperation. Also, assessing the impact of mobility and energy consumption, while very important aspects of the problem, are outside the scope of this work. We defer such investigations to the future.

The rest of the thesis is organized as follows. Chapter 2 describes related works. In Chapter 3 we formulate the optimization problem. In Chapter 4 we exemplify the main attributes of the problem. This is followed by the proposed heuristic approach for nodes connected to a single cell tower in Chapter 5. Chapter 6 extends the approach to the case when nodes are connected to more than one cell tower. Chapter 7 details the evaluation methodology and the results obtained from it. We conclude and discuss possible extensions to the work in Chapter 8.

Chapter 2

Related Work

The work in [21] focuses on the placement of a fixed number of Mobile backbone nodes (MBN) and the assignment of regular nodes (RN) to MBN(s). They provide algorithms that achieve the placement and assignment to maximize the minimum node throughput and to maximize the sum throughput. 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, in our problem we do not a priori fix the number of hotspots (similar to MBN(s) in [21]). Finally, in our work we want to maximize the network throughput while ensuring that at least the throughputs in the baseline case are achieved. Such a baseline is not possible in their work as the RN(s) cannot directly connect to the internet.

Authors in [12] 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 fixed and users are connected to these access points optimally. Instead in our work, the access points (mobile WiFi hotspots) are selected from the given nodes and remaining nodes access the internet via these hotspots.

Works [23] [15] [1] are related to offloading of cellular data to other wireless networks to achieve load balancing. The traffic may be offloaded to the less congested networks even if the user sees a low signal-to-noise-ratio from the networks.

There are several works on offloading cellular data traffic using small cells [18] [11] [6] and WiFi [13] [20] [8] [16] [2]. These works aim to offload cellular connections to networks that have an independent connection to the internet. In our work, the connectivity to the internet is still via the cellular network. Just that not all users may connect directly to it. Some users may connect to it via other users (hotspots).

D2D communication [14] [9] is also a good way of efficiently utilizing the spectrum. There are two kinds of D2D communication. One is inband (underlay and overlay D2D communication) [17] [24] [22] in which same spectrum is used by both the cellular users as well as D2D users. But here comes the problem of interference between cellular users and D2D users. So, to mitigate the interference between cellular and D2D users people started looking at second type of D2D communication which is the outband D2D communicaton [10] [3] in which orthogonal spectrums are used for cellular users and D2D users. Authors in [3] propose to form clusters among cellular users who are in WiFi range and one of the cluster member called cluster head (we can say hotspot in our case) communicates with the BS and the remaining members (clients in our scheme) of cluster communicate to base station via cluster head. Scheduling algorithm consist of two parts. One is to schedule clusters and other is the selection of cluster head opportunistically within each cluster. But here it is not ensured that the cluster head can fulfill the throughput requirement of all the members belong to the cluster. The work related to D2D communication in detail is given in survey paper [4].

Chapter 3

Optimization Problem

Consider a network $\mathcal{N}$ of N users/nodes (smartphones), indexed as $i = 1, 2, \ldots, N$, each of whom want to connect to the internet and can do so by connecting to a cell tower. We will formulate the problem for the case when all users get access to the internet via a single cell tower. This allows for a clearer exposition of the problem. Extending the formulation to the case when there are more than one cell tower is straightforward. User i has a cellular link (to the tower) with a signal-to-interference-and-noise ratio of SINR_i. We will not distinguish between the cellular uplink and downlink of a user. Let the *baseline* network configuration be the one in which all users access internet using their own connection to the cell tower. 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. While all users access the same amount of bandwidth, each user gets access to the bandwidth for a fraction $1/N$ of the total time [1]. The baseline throughout $T_i^{(B)}$ of user i is therefore given by

$$T_i^{(B)} = \frac{1}{N} \log_2(1 + \mathrm{SINR}_i). \tag{3.1}$$

Let a_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, N$, be variables that indicate assignment. Specifically, $a_{ij} = 1$ when i is a hotspot and j is its client. It is zero otherwise. Also, if i is a hotspot, $a_{ii} = 1$, else $a_{ii} = 0$. This case when $i = j$ is further discussed after the definition of the problem (3.3)-(3.8).

Figure 3.1 illustrates how the link of node i to the cell tower is used by a node j that is a client of i. In the figure, node j connects to the tower via i. Let T_{ij} be the throughput reserved for user j on the link to the cell tower of user i, when i is configured as a mobile hotspot and j is a client of i. Let W_{ij} be the throughput of the WiFi link between hotspot i and its client j. Clearly, $T_{ij} \leq W_{ij}$. Also, T_{ij} must be greater than or equal to its baseline throughput. That is $T_{ij} \geq T_j^{(B)}$. Finally, T_{ii} is the share of the throughput that the hotspot i gets on its own link to the internet.

Note that any node j must connect to the internet via exactly one node. If it uses its own internet connection, it gets a throughput of T_{jj}. Also, $a_{jj} = 1$ and $a_{ij} = 0$ for all $i \neq j$. On the other hand, if it connects to the internet via another node i, its throughput is T_{ij}. Also, $a_{kj} = 1$

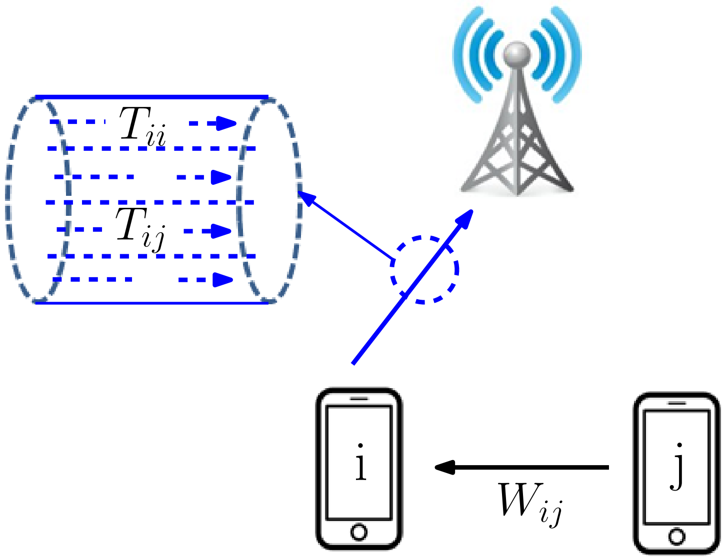


Figure 3.1: A node's link to the cell tower and how it is shared by other nodes $j \in \mathcal{N}$.

for $k = i$ and $a_{kj} = 0$ for all $k \neq i$. Therefore, the throughput of user j is given by

$$T_j = \sum_{i=1}^{N} a_{ij}\, T_{ij}. \tag{3.2}$$

We want to maximize the sum throughput $\sum_{j=1}^{N} T_j$ of the network, while ensuring that all users get at least their baseline throughputs. The optimization problem is

$$\begin{aligned}
\text{Maximize:} \quad & \sum_{j=1}^{N}\sum_{i=1}^{N} a_{ij} T_{ij}, && && (3.3)\\
\text{subject to:} \quad & T_{ij} \leq W_{ij} && \forall i, j, && (3.4)\\
& T_j \geq T_j^{(B)} && \forall j, && (3.5)\\
& \sum_{j=1}^{N} T_{ij} a_{ij} = \frac{a_{ii}}{H} \log_2(1 + \mathrm{SINR}_i) && \forall i, && (3.6)\\
& \sum_{i=1}^{N} a_{ij} = 1 && \forall j, && (3.7)\\
& a_{ij} \in \{0, 1\} && \forall i, j, && (3.8)
\end{aligned}$$

where H is the number of users connected to the cell tower (hotspots). It is given by

$$H = \sum_{i=1}^{N} a_{ii}. \tag{3.9}$$

The variables of optimization are T_{ij} and a_{ij} for $i = 1, \ldots, N$ and $j = 1, \ldots, N$. The SINR(s), SINR_i, $i = 1, \ldots, N$, of the cellular links are assumed to be known. Equation (3.4) ensures that the throughput reserved for any user j on the link to the cell tower of i is not greater than the throughput of the WiFi link between them. The equations (3.5)-(3.8) specify the other constraints under which we perform the optimization. Constraint (3.5) enforces that the optimal solution must be such that each user gets at least as much throughput as the user was getting in the baseline case. This is to ensure that no user suffers as a result of enabling heterogeneous access to the internet.

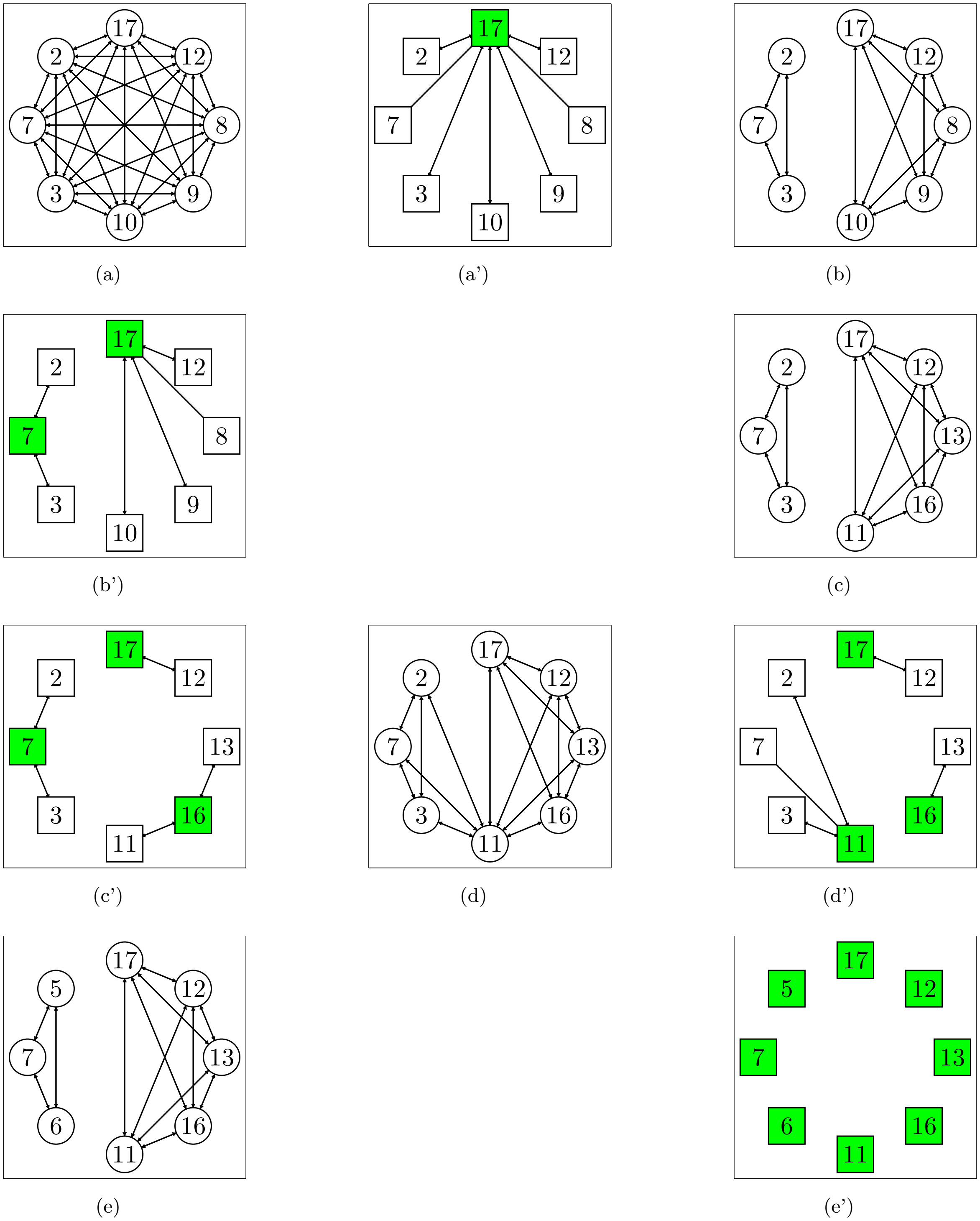
Constraint (3.6) ensures that all users (clients) connected to a hotspot and the user configured as the hotspot get throughputs such that their sum is equal to the throughput of the cellular link of the hotspot. Note that a_{ii} gets set to 1 when i is a hotspot. Also, note that in case the optimal solution includes a hotspot node i that has no clients connected to it, then (3.5)-(3.6) will not be satisfied unless $a_{ii} = 1$. Constraint (3.7) enforces that a client must be connected to exactly one hotspot. For a node i that is selected to be a hotspot this constraint is satisfied by setting $a_{ii} = 1$.

Constraints (3.5)-(3.7) further ensure that every client is connected to a hotspot and not to another client. That is if j and k are clients then $a_{jk} = 0$. If for the clients $a_{jk} = 1$, then even if (3.7) is satisfied, constraint (3.5) and (3.6) cannot both be satisfied as $a_{kk} = 0$ for client k. Also, note that (3.7) ensures that a hotspot cannot be connected to another hotspot.

It is instructive to note that the utility function in (3.3) is equal to the sum, over all nodes i in the network, of the right-hand side of equation (3.6). As $a_{ii} = 1$ if i is a hotspot and $a_{ii} = 0$ otherwise, we are choosing hotspots such that the total throughput to the cell tower is maximized, while all nodes get at least their baseline throughput. Also, note that if the a_{ij} are set such that the total throughput to the cell tower via a hotspot i is greater than or equal to the sum of the baseline throughputs of all nodes j for which $a_{ij} = 1$, then for such nodes j, T_{ij} can be selected so as to satisfy constraints (3.5) and (3.6).

WiFi is a CSMA based random access mechanism [5]. All nodes in a WiFi network that are on the same or overlapping channels (for example, channel 9 and 11 in 2.4 GHz ISM band used by 802.11g) interfere with each other and must contend with each other for channel access. There are a total of only three non-overlapping channels in the 2.4 GHz ISM band. In general, the WiFi throughput W_{ij} will change as a function of the number of hotspots, the channels in which they operate, and the number of clients connected to a hotspot (more clients lead to greater contention and less throughput per client). In this work, we will make a simplifying assumption that a WiFi link is either ON or OFF. If a link between i and j is ON, we set $W_{ij} = \infty$. Such a link will always satisfy (3.4), irrespective of the network configuration obtained after the optimization. We will also assume that $W_{ij} = W_{ji}$. Thus, $W_{ij} = \infty$ implies that nodes i and j have good WiFi connectivity between each other. On the other hand, if the WiFi link is OFF, we have $W_{ij} = 0$. The nodes have poor WiFi connectivity between each other. We set $W_{ii} = \infty$ for all i. This is to enforce that the throughput T_i of a hotspot i is not affected by its WiFi connectivity to self.

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



(a) (a') (b)

(b') (c)

(c') (d) (d')

(e) (e')

Figure 3.2: Examples of WiFi connectivity and the corresponding optimal hotspot networks. WiFi connectivity is shown by graphs that represent nodes by circles. Optimal hotspot networks are shown by graphs with square nodes. The nodes in green are the hotspots. Figure 3.2(x') is the optimal hotspot network corresponding to Figure 3.2(x).

Chapter 4

Problem Attributes

The problem (3.3)-(3.8) 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 proposed heuristic approach, explained in Sections 5 and 6.

Example networks are shown in Figure 3.2. The networks are shown as graphs. Each vertex of a graph corresponds to a user 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 3.2a- 3.2d', show the WiFi connectivity amongst the users in the network. An edge between two vertices implies that the WiFi link between the two users is ON. Lack of an edge implies that the WiFi link between the users is OFF. These users may be configured into a throughput optimal network of hotspots and clients. The graphs with *square vertices* show connectivity of the users to the internet in the optimal network configuration. Vertices in green are the hotspots. They connect directly to the cell tower and may have zero or more vertices connected to them via WiFi. The graphs in Figures 3.2a', 3.2b', 3.2c', 3.2d', 3.2e' are the optimal network configurations, respectively, of the networks in the Figures 3.2a, 3.2b, 3.2c, 3.2d, 3.2e.

Figure 3.2a shows a network of users that have good WiFi connectivity (all WiFi links are ON) amongst each other. The set of users is $\mathcal{N} = \{2, 3, 7, 8, 9, 10, 12, 17\}$. The cellular SINR(s) of the users range from 3 dB to 17 dB. In the absence of constraints (3.4)-(3.8), the network's throughput (to the cell tower) is maximized by simply allowing the user with the largest cellular SINR (user indexed 17) to use the access the cellular bandwidth all the time. The resulting network throughput is $\log_2(1 + 10^{(17/10)})$ bits/sec/Hz. Baseline throughputs of the nodes $T_i^{(B)}$ can be calculated using equation (3.1). We have $N = 8$. The baseline throughputs (rounded to two decimal places) of the nodes indexed $2, 3, 7, 8, 9, 10, 12, 17$ are, respectively, $0.17, 0.20, 0.32, 0.36, 0.40, 0.43, 0.51, 0.71$ bits/sec/Hz. The sum throughput of the baseline configuration is therefore 3.0976 bps. Since all WiFi links are ON, user 17 can act as a hotspot for the rest. The resulting hotspot network is shown in Figure 3.2a'. It is easy to verify that the constraints (3.4)-(3.8) are satisfied. We have $a_{17,j} = 1$ for all j and $a_{ij} = 0$ for all j and $i \neq 17$. Also, $T_{ij} = 0$ for all $i \neq 17$, number of hotspots $H = 1$, and for $i = 17$ the T_{ij} for all j

can be chosen such that $\sum_{j\in\mathcal{N}} T_{ij} = \log_2(1+10^{(17/10)})$ and $T_{ij} \geq T_j^{(B)}$. The percentage gain in network throughput over baseline is 83%.

Now consider the network in Figure 3.2b. 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. We have $W_{ij} = \infty$ when both nodes i and j belong to either C_1 or C_2. Else, $W_{ij} = 0$. In the absence of constraints 3.4- 3.8, network throughput is still maximized by allowing the link with a cellular SINR of 17 dB to use the entire resource. However, constraints can be satisfied only if we have two or more hotspots (at least one per cluster). Figure 3.2b' 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. No other feasible hotspot selections can do as well. This is because any other feasible selection of hotspots will involve either more hotspots and/or hotspots with smaller SINR(s). In the optimal network (Figure 3.2b'), hotspot 17 and its clients get half the cellular time resource. The remaining half will be used by hotspot 7 and its clients. The sum throughput of the network is the sum of the bits/sec $(.5\log_2(1+10^{(17/10)}) + 0.5\log_2(1+10^{(7/10)}))$ that the nodes 17 and 7 get to the cell tower. It is 4.13 bps, which is a 33% improvement over the sum throughput (3.0976 bps) of the baseline configuration. To summarize, in the optimal hotspot network, we have $a_{ij} = 1$ for $i = 7, j \in C_1$ and for $i = 17, j \in C_2$. Else, $a_{ij} = 0$. Also, for each node $j \in C_1$ ($j \in C_2$), $T_{7,j}$ ($T_{17,j}$) can be set such that all nodes get at least their baseline throughputs.

Figure 3.2c slightly modifies the network in Figure 3.2b. While the WiFi connectivity remains unchanged, the SINR(s) of the users in the larger cluster have a *smaller spread*. The baseline throughputs of the users $2, 3, 7, 11, 12, 13, 16, 17$ are respectively $0.17, 0.20, 0.32, 0.47, 0.51, 0.55, 0.67, 0.71$ bps. The hotspot configuration in Figure 3.2b' can no longer satisfy all the constraints. Specifically, it is easy to verify that the total throughput of node 17 to the cell tower, when it gets access to the cellular bandwidth half the time as in Figure 3.2b', is smaller than the sum of the baseline throughputs of all the nodes in its cluster. Note that nothing has changed with respect to the smaller cluster 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 a 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. Also, 7 as a hotspot can support at least baseline throughputs of the nodes in the smaller cluster, even when it gets only 1/3 of the time. The network is shown in Figure 3.2c'. The resulting gains in sum throughput over the baseline are 26%.

Figure 3.2d has a network similar to that in Figure 3.2c. They differ in WiFi connectivity of the nodes. Unlike Figure 3.2c, the network in Figure 3.2d has a node (11) that is common to both the clusters. While the optimal network for Figure 3.2c is feasible for the network in Figure 3.2d, it is not optimal. One of the optimal configurations is shown in Figure 3.2d'. The choice of 17, 16, and 11 as hotspots, instead of the feasible choice of 17, 16, and 7, leads to throughput gains of 37% instead of 26%. Therefore, assigning 11 as a hotspot instead of making it a client improves the percentage gains by about 60%.

Finally, consider the network in Figure 3.2e. It retains the larger cluster of Figure 3.2c. However, its smaller cluster has nodes with larger cellular SINR(s) and a smaller spread in the SINR(s). It turns out, as is detailed next, that the baseline configuration is in fact throughput optimal.

Clearly, the hotspot network must contain at least two hotspots (one per cluster). Also, note that the larger cluster is the same as in Figure 3.2c. Thus, as shown when discussing the Figure 3.2c, the larger cluster needs at least two hotspots. This implies that the smaller cluster can get only 1/3 of the time. Given 1/3 of the time, the throughput of the link from 7 to the cell tower is $(1/3)\log_2(1 + 10^{(7/10)}) = 0.86$ bps. This is smaller than the sum 0.87 bps of baseline throughputs of the nodes $5, 6, 7$ in the smaller cluster. Thus 7 cannot be the only hotspot for the smaller cluster. This means that we need at least two hotspots per cluster. At two hotspots per cluster, each hotspot gets 1/4 of the time. The hotspots 16 and 17 can no longer support the baseline throughputs of the larger cluster. The throughputs of their links to the cell tower become $(1/4)\log_2(1 + 10^{(16/10)})$ and $(1/4)\log_2(1 + 10^{(17/10)})$ respectively, which together can provide 2.757 bps. This is smaller than the sum 2.91 bps of the baseline throughputs of the nodes $11, 12, 13, 16, 17$ in the larger cluster.

Thus, two hotspots for the smaller cluster must be accompanied by at least 3 hotspots for the larger cluster. At 1/5 of the time, nodes $6, 7$ as hotspots together provide a throughput to the tower that is larger than the sum of baseline throughputs of $5, 6, 7$. Node 7 has a throughput of 0.518 bps to the tower and node 6 has a throughput of 0.463 bps. The baseline throughputs of $5, 6, 7$ are $0.257, 0.29, 0.324$ bps respectively. Node 5 must be assigned as a client to either one of 6 and 7. It cannot be assigned to 7 (and hence 6) as the link of 7 to the tower has a throughput of 0.518 bps, which is less than the sum 0.58 bps of baseline throughputs of 7 and 5.

Thus, we need at least three hotspots for each of the smaller and the larger cluster. Note that the smaller cluster cannot have more than three hotspots. Given the resulting 1/6 of time per hotspot, the selection of $13, 16, 17$ as the three hotspots cannot support the baseline throughputs of all the nodes in the larger cluster. The total throughput of $13, 16, 17$ to the cell tower is 2.57 bps, which is smaller than the sum 2.9 bps of the baseline throughputs of all the nodes in the larger cluster.

Thus we need 3 hotspots for the smaller cluster and at least 4 for the larger cluster. This, at 1/7 time per hotspot, is also infeasible, as no 4 nodes in the larger cluster can support the baseline throughput requirements of the 5 nodes in it. The only configuration that remains is the one in which we have three hotspots in the smaller cluster and five in the larger cluster. The hotspots use their own cellular connection and have no clients connected to them. This configuration is equivalent (with respect to the network throughput) to the baseline configuration.

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. Specifically, for a network that consists of clusters of varied sizes, often consideration of more than just the cellular SINR of the nodes is required when selecting hotspots and clients. 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, the role it is assigned may significantly impact the network throughput.

In the examples above we did not assign specific values to T_{ij}, which is the share of node j on the link to the cell tower of node i. An example assignment that satisfies the constraint (3.6) is as follows. Note that this assumes that each node selected as a hotspot is able to support the sum of baseline throughputs of all nodes connected to the internet via it. Define $\pi_j = T_j^{(B)}/(\min_{j \in \mathcal{N}} T_j^{(B)})$. Let the throughput of the link of hotspot i to the cell tower be $T_i^{(H)}$. Let $\mathcal{U}_i$ be the set of clients of hotspot i. For a hotspot i and all nodes j connected to the internet via it (which is the set of nodes $\{\mathcal{U}_i \cup i\}$), let $T_{ij} = (\pi_j / \sum_{j \in \{\mathcal{U}_i \cup i\}} \pi_j) T_i^{(H)}$.

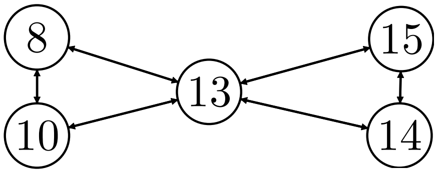
Chapter 5

One Cell Tower: Heuristic Approach

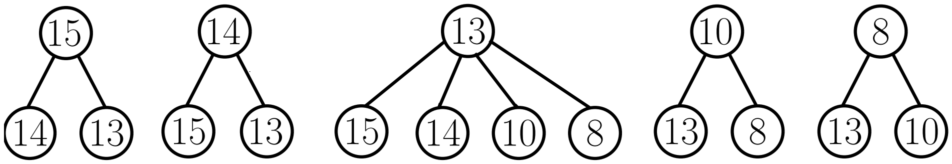
Algorithm 1: Configure-Network

```
Input: $\mathcal{N}$: // Set of nodes in the network
Input: $W$: // WiFi Connectivity Matrix
Output: Hotspot Network Configuration
Compute $i_1, i_2, ...., i_{\mathcal{N}}$ $s.t.$ $\text{SINR}_{i_1} \leq \text{SINR}_{i_2} .... \leq \text{SINR}_{i_{\mathcal{N}}}$
for $H = 1$ $to$ $|\mathcal{N}|$ do
    $\mathcal{H} = \emptyset$; $\mathcal{U}_h = \emptyset,\ 0 \leq h \leq H$;
    $\mathcal{H}' = \emptyset$; $\mathcal{R}' = \mathcal{N}$; $c = 1$
    while $\mathcal{R}' \neq \emptyset$ && $c \leq H$ do
        $\mathcal{H} = \mathcal{H} \cup \mathcal{H}'(1)$;
        // $\mathcal{H}'(1)$ is the hotspot corresponding to the set $\mathcal{U}_{c-1}$
        $\mathcal{U} = \cup_{h=c}^{H} \mathcal{U}_h$;
        $\mathcal{N}' = \mathcal{R}' \cup \mathcal{U}$;
        $[\mathcal{H}', \mathcal{R}', \mathcal{U}_c, \ldots, \mathcal{U}_H]$
                = Select Hotspots$(\mathcal{N}', W, H, H - c + 1)$;
        $c = c + 1$;
    end
    $\mathcal{H} = \mathcal{H} \cup \mathcal{H}'$;
    if $\mathcal{R}' == \emptyset$ then
        $T_H = \frac{1}{H} \sum_{i \in \mathcal{H}} \log_2(1 + \text{SINR}_i)$;
        save $\mathcal{H}, \mathcal{U}_1, \ldots, \mathcal{U}_H$ as the $H$-hotspot network configuration.
    else
        $T_H = 0$;
        save $\emptyset$ as the $H$-hotspot network configuration.
    end
    if $H < |\mathcal{N}|$ $AND$ $\frac{1}{H+1} \sum_{i \in i_1, \ldots, i_{H+1}} \log_2(1 + SINR_i) < T_H$ then
        break;
    end
end
$H^* = H$;
Return the saved $H^*$-hotspot network configuration;
```

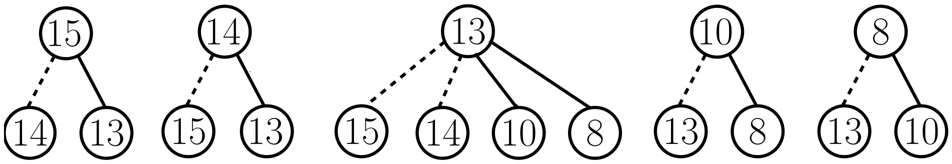
Our proposed approach, under the assumption that all nodes are connected to the same cell tower, is detailed in Algorithms 1 and 2 and is exemplified by Figures 5.1b-5.1d for the network in Figure 5.1a. In summary, it proceeds by first fixing the number, say H, of hotspots. This fixes the fraction of time ($1/H$) for which any node selected to be a hotspot can access the cellular



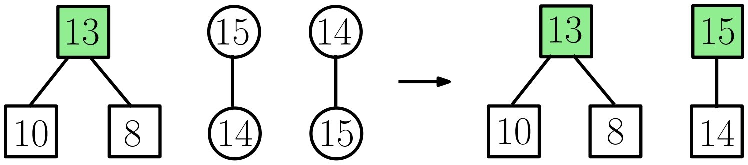
(a) Example network.



(b) First run of Select-Hotspots for $H = 1$.



(c) First run of Select-Hotspots for $H = 2$.



(d) Second run of Select-Hotspots for $H = 2$.

Figure 5.1: An example of the working of the proposed approach.

Algorithm 2: Select-Hotspots

```
Input: $\mathcal{N}'$ // Given set of nodes
Input: $W$
Input: $H$: // Target number of hotpsots
Input: $H'$: // Number of hotpsots that remain to be selected
Output: $\mathcal{H}'$,$\mathcal{N}'$,$\mathcal{U}_1, \ldots, \mathcal{U}_{H'}$
$\mathcal{H} = \emptyset$; $\mathcal{U}_h = \emptyset,\ 1 \le h \le H'$;$W' = W$
$\mathcal{H}' = \emptyset$;
for $n \in \mathcal{N}'$ do
    $\mathcal{O} = \{o : W'_{no} = \infty, o \in \mathcal{N}'\}$;
    $R = \sum_{i \in \mathcal{O}} T_i^{(B)}$;
    $A = \frac{1}{H} \log_2(1 + \mathrm{SINR}_n)$;
    while $A < R$ do
        $k^* = \arg\max_{k \in \mathcal{O}} \mathrm{SINR}_k$;
        $W'_{nk^*} = 0$;
        $\mathcal{O} = \{o : W'_{no} = \infty, o \in \mathcal{N}'\}$;
        $R = \sum_{i \in \mathcal{O}} T_i^{(B)}$;
    end
end
for $h = 1$ to $H'$ do
    $n^* = \arg\max_{n \in \mathcal{N}'} |\{k : W'_{kn} = \infty\}|$;
    $\mathcal{H}' = \mathcal{H}' \cup n^*$;
    $\mathcal{U}_h = \{u : W'_{n^*u} = \infty\}$;
    $\mathcal{N}' = \mathcal{N}' - \mathcal{U}_h$;
end
```

bandwidth. Thus, for every node in the network, we can calculate the throughput it will get to the cell tower if it were to be one of the H hotspots. Note that since H is fixed, this throughput is only dependent on the cellular SINR of a node.

Next, for every node, we calculate the set of nodes for which it can be a hotspot. Note that a node i can serve as a hotspot to only those nodes with whom it has good WiFi connectivity, that is the set of nodes $\mathcal{W}_i = \{j : W_{ij} = \infty\}$. The cellular SINR of i and the available time share of $1/H$ may further constrain it to serve only a fraction of the nodes in $\mathcal{W}_i$. We propose a simple cellular *SINR-based heuristic* to decide the order in which nodes are removed from the set $\mathcal{W}_i$. Specifically, nodes that have larger cellular SINR(s) are removed first. The motivation is that, everything else remaining equal, a node with a larger cellular SINR is a better candidate for becoming a hotspot than a node with a smaller cellular SINR. The node with higher SINR will accommodate more clients and will have better throughputs to the cell tower. Nodes are removed till those that remain can be supported by the cellular throughput of i when there are a total of H hotspots. Once we have for every node in the network, the set of nodes for which it can be serve as a hotspot, we greedily pick the H nodes that together connect all nodes in the network to the cell tower. In case, H such nodes are not found, we declare that a hotspot network configuration with H hotspots is infeasible and the resulting network throughput is zero. The above is performed for $H = 1, \ldots, N$. We set $H < N$ as the optimal number of hotspots H^* if the largest possible sum throughput for any selection (feasible or not) of $H + 1$ hotspots is less than the throughput T_H obtained from the feasible selection of H hotspots. If

$H = N$, then the baseline configuration is the returned as the optimal network configuration.

5.1 Algorithmic Details

The starting point of our approach is the algorithm Configure-Network (Algorithm 1). Note that the number of hotspots can range from 1 to the size of the network $\mathcal{N}$ of users. For every selection of number of hotspots, the algorithm first invokes Select-Hotspots described in Algorithm 2, and then checks the feasibility of the returned H-hotspot network configuration. Finally, it picks the H^*-hotspot configuration, which gives a throughput larger than that given by other H-hotspot configurations.

In the algorithms, $\mathcal{H}$ is the set of hotspots, indexed $1, 2, \ldots, H$, selected for a given number H of hotspots. The sets $\mathcal{U}_1, \ldots, \mathcal{U}_H$ are the sets of clients that are connected to the H hotspots in $\mathcal{H}$. As before, W is the WiFi connectivity matrix. T_H is the throughput of the selected H-hotspot network.

Select-Hotspots takes any given set $\mathcal{N}'$ of nodes, the target number H of hotspots, the number H' that remains to be selected, and the WiFi connectivity matrix. Note that Select-Hotspots may be called up to H times by Configure-Network for a given H. The first time it is called for a given H, $H' = H$ and $\mathcal{N}' = \mathcal{N}$. The need for multiple calls is explained later. Inside Select-Hotspots, the first for-loop calculates for each node in $\mathcal{N}'$, using the proposed SINR-based heuristic, the nodes which it can serve as a hotspot. A is the cellular throughput available to the selected node. $\mathcal{O}$ is the set of nodes that are in the set of prospective clients of the node. This set is initialized to all nodes that have good WiFi connectivity with the node. Nodes are removed from the prospective set by setting the corresponding entry in W' to zero and updating $\mathcal{O}$.

After prospective clients for each node have been finalized, in Select-Hotspots, we look for the H' nodes that provide connectivity to the given set $\mathcal{N}'$ of nodes. 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 $\mathcal{U}_h$, $1 \leq h \leq H'$, of clients, the algorithm also returns the set of nodes 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 $\mathcal{N}'$. That is the selection of hotspots is infeasible.

Figure 5.1b-5.1d exemplify the workings of the two algorithms for the network of users in 5.1a. Figure 5.1b shows the results of setting $H' = H = 1$. It turns out that each node (top row) can support all nodes (bottom row) that have good WiFi connectivity with it. Also, 13 can provide connectivity to all nodes in the network. So $H = 1$ is feasible. However, as we will see, $H = 2$ gives a larger throughput. Figure 5.1c shows the result of calling Select-Hotspots the first time for $H = H' = 2$. For each node in the network (shown in the 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 result is that node 15 and 14 can both be a hotspot for 13, 13 can be a hotspot for 10 and 8, 10 and 8 can be hotspots for each other. The greedy method of selecting hotspots will first select 13 as a hotspot, followed by 15. Note that while 15, 14, 10, and 8 have the same number of selected clients, 15 has the largest SINR, and so is selected as the second hotspot. The selection of 13 and 15 will leave 14 without a connection. The issue with 14 is that it has a high cellular SINR and so it gets eliminated as a client. Also, the only client it supports gets selected as a hotspot. Infeasibility of such kind, which is due to the greedy decision making by Select-Hotspots, can be averted by calling Select-Hotspots again, however, over nodes that do not include the first selected hotspot and its clients.

For our example, this would mean calling Select-Hotspots over the nodes 15 and 14, with $H = 2$ and $H' = 1$, with the goal of adding one more hotspot. We started with $H = H' = 2$ and have already chosen 13 as one of the hotspots. This gives us Figure 5.1d. To the left of the arrow, we have a hotspot network configuration created in part, together with nodes 14 and 15 with their feasible clients. It is worthy of note that while we want to select $H' = 1$ hotspot from 14 and 15, their available capacity (A in Select-Hotspots) is calculated for the final target of $H = 2$ hotspots. To the right of the arrow, we have the final hotspot network configuration. Note that since 15 and 14 have the same connectivity, 15 was chosen as a hotspot due to its larger SINR.

More generally, for a given H, Configure-Network calls Select-Hotspots up to a maximum of H times. The c^{th} time Select-Hotspots is called, it returns $H' = H - c + 1$ hotspots in the set $\mathcal{H}'$ and their set of clients $\mathcal{U}_c, \ldots, \mathcal{U}_H$. Note that prior to the c^{th} call, $c - 1 < H$ 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 $\mathcal{N}'$ that are not provided connectivity by the H' hotspots. If $\mathcal{R}' \neq \emptyset$, the hotspot that was selected first amongst the H' hotspots, denoted by $\mathcal{H}'(1)$, and its clients, that is the set $\mathcal{U}_c$ and its hotspot, are added to the existing hotspot network configuration. In the next iteration, the rest of the hotspots and clients (now the set $\cup_{h=c}^{H} \mathcal{U}_h$) are sent back to Select-Hotspots. If all nodes are not covered even after H calls of Select-Hotspots, a H-hotspot network is declared infeasible.

Chapter 6

More Than One Cell Tower

We will now extend our approach to a network that consists of more than one cell tower. While a node is still connected to exactly one cell tower, different nodes in the network may be connected to different cell towers. As before, our network consists of the set of nodes $\mathcal{N}$. In addition we have $M \geq 1$ cell towers indexed $1, \ldots, M$. The cell towers occupy orthogonal and equal bandwidths (as before, assumed to be 1 Hz). We assume that a node that is chosen to be a hotspot in the resulting hotspot network is always connected to the cell tower it was connected to in the baseline configuration. However, a client may be connected to a hotspot that is connected to a cell tower which is not the one to which the client was connected to in the baseline configuration.

The modifications to the problem 3.3- 3.8 required to incorporate multiple cell towers are limited to the constraint 3.6. Recall that (3.6) enforces, for all hotspots connected to the only cell tower, that the sum of throughputs of clients connected to a hotspot i in the network must be equal to the throughput that the hotspot gets to the cell tower. In the single cell tower case, this throughput is a function of SINR_i and the number H of hotspots connected to the cell tower. When there are more than one cell towers, the constraint (3.6) must be enforced per cell tower. That is if there are $M > 1$ cell towers, we must enforce (3.6) separately for each of the M cell towers.

Let the connectivity of nodes to cell towers be given by the matrix C, which has M rows and N columns. If node i is connected to cell tower j, then $C_{ji} = 1$, else $C_{ji} = 0$. Also, for any node i, $\sum_{j=1}^{M} C_{ji} = 1$. That is a node can only connect to one cell tower. Algorithm 3 summarizes our approach to finding a hotspot network when there are more than one cell towers serving the nodes in the network. We first optimize per cell tower. The resulting set $\mathcal{H}$ of H hotspots and the sets of clients $\mathcal{U}_1, \ldots, \mathcal{U}_H$ connected to them are further rearranged by Algorithm Rearrange (Algorithm 4).

Rearrange is motivated by the following observations. Leveraging WiFi connectivity between nodes connected to different cell towers can reduce the number of hotspots in the hotspot network that was returned after optimizing per cell tower, and this can increase network throughput. However, a reduction of the number of hotspots that satisfies the constraints (3.4)- (3.8) does

Algorithm 3: Optimizing over multiple cell towers.

```
Input: $\mathcal{N}$: // Nodes in the network
Input: $W$: // WiFi Connectivity of all nodes
Input: $C$: // Connectivity of nodes with cell towers
Input: $M$: // Number of cell towers
Output: Hotspot Network Configuration
$H = \emptyset, \mathcal{H} = \emptyset, \mathcal{U} = \emptyset$;
for $c$ = 1 to $M$ do
    $\mathcal{N}_c = \{i : C_{ci} = 1, i \in \mathcal{N}\}$;
    $[\mathcal{H}', \mathcal{U}']$ = Configure-Network($\mathcal{N}_c, W$) ;
    $\mathcal{H} = \mathcal{H} \cup \mathcal{H}', \mathcal{U} = \mathcal{U} \cup \mathcal{U}'$;
end
$[\mathcal{H}^*, \mathcal{U}^*]$ = Rearrange($W, \mathcal{H}, \mathcal{U}$);
$\mathcal{H}^*, \mathcal{U}^*$ is the final configuration.
```

not always increase the network throughput. For instance, reducing the number of hotspots connected to a cell tower to zero will necessarily reduce network throughput. This is because the cell towers have access to orthogonal bandwidths. Also, rearranging such that a node with a large cellular SINR becomes a client, may reduce network throughput.

Define TOWER(i) to be a function that returns the cell tower to which node i is connected. Note that TOWER(i) can be evaluated using the matrix C. Also define HOTSPOTS($c, \mathcal{S}$) as a function that returns the hotspots that are in the set $\mathcal{S}$ and are connected to the cell tower c. This function can also be evaluated using the matrix C.

Rearrange begins by first ordering the hotspots in $\mathcal{H}$ in the increasing order of their cellular SINR. In the selected order, it tries to move a hotspot (h in the algorithm) and its clients to become clients of other hotspots in the network. First it tests if the hotspot h is eligible to be moved. For a hotspot h to be eligible to be moved, it must not be the only hotspot connected to its cell tower, it must have good WiFi connectivity with at least one other existing hotspot, and moving it must lead to a non-negative gain in the throughput being obtained via the cell tower to which it is connected.

Note that, in the algorithm, $\mathcal{H}_c$ is the set of hotspots that are connected to the cell tower of h. Therefore, its size must not be equal to 1. After removal of h, the set of hotspots connected to c is given by $\mathcal{H}'_c$. The improvement in throughput obtained via the cell tower c when h is removed as one of its hotpots is calculated as G. For h to be eligible to be moved, we need it to satisfy $G >= 0$. The last condition ($\sum_{j \in \mathcal{H}'} W_{jh} == 0$) tests if h is not connected to any other hotspot in the current network.

If h passes the initial eligibility tests, we try and assign it and its clients (the set $\{h \cup \mathcal{U}_h\}$) to other hotspots in the network (the set $\mathcal{H}'$). To do so, we first calculate the capacity A_i available at each hotspot in $\mathcal{H}'$, under the assumption that h is no longer a hotspot. Note that the available capacity is the difference between the throughput the hotspot gets to its cell tower and the sum of the baseline throughputs of all nodes (its clients and itself) using its connectivity to its cell tower. Note that A_i for those hotspots i that are not connected to the cell tower of h is not changed under the assumption that h is not a hotspot. The change in A_i is at the hotspots that are connected to the same cell tower as h.

What follows is an attempt to accommodate h and its clients as clients of the hotspots in $\mathcal{H}'$. The one with the larger SINR is accommodated first. We change the hotspot network by updating $\mathcal{H}$ and $\mathcal{U}_i, 1 \leq i \leq |\mathcal{H}|$ only if all of them can be accommodated. Else, we retain h as a hotspot and all its clients.

Algorithm 4: Rearrange

```
Input: $W$: // WiFi Connectivity
Input: $\mathcal{H}$: // Selected hotspots
Input: $\mathcal{U}$: // Clients of hotspots
Output: Hotspot Network $\mathcal{H}$,$\mathcal{U}$
$H = |\mathcal{H}|$;
Compute $i_1, i_2, ...., i_H$ $s.t.$ $\text{SINR}_{i_1} \leq \text{SINR}_{i_2} .... \leq \text{SINR}_{i_H}$
for $h = i_1, i_2, ...., i_H$ do
    $c = \text{TOWER}(h)$;
    $\mathcal{H}' = \{\mathcal{H} - h\}$;
    $\mathcal{H}_c = \text{HOTSPOTS}(c, \mathcal{H})$;
    $\mathcal{H}'_c = \text{HOTSPOTS}(c, \mathcal{H}')$;
    $\text{G} = \frac{\sum_{j \in \mathcal{H}'_c} \log_2(1+\text{SINR}_j)}{|\mathcal{H}'_c|} - \frac{\sum_{j \in \mathcal{H}_c} \log_2(1+\text{SINR}_j)}{|\mathcal{H}_c|}$;
    if $(|\mathcal{H}_c| == 1)$ $OR$ $(G < 0)$ $OR$ $(\sum_{j \in \mathcal{H}'} W_{jh} == 0)$ then
        continue;
    end
    $A_i = \frac{\log_2(1+\text{SINR}_i)}{\text{HOTSPOTS}(\text{TOWER}(i),\mathcal{H}')} - \sum_{j:a_{ij}=1} T_j^{(B)}, \forall i \in \mathcal{H}'$;
    $\widetilde{\mathcal{U}}_d = \emptyset, \ \forall d \in \mathcal{H}$;
    Compute $k_1, k_2, ...., k_f$ $s.t.$ $\text{SINR}_{k_1} \geq \text{SINR}_{k_2} .... \geq \text{SINR}_{k_f}$, where $k_1, k_2, ...., k_f \in \{h \cup \mathcal{U}_h\}$;
    for $k = k_1, k_2, ...., k_f$ do
        $p = \{e : W_{ek} = \infty, e \in \mathcal{H}'\}$;
        if $p = \emptyset$ then
            $k = -1$; break;
        end
        $p^* = \arg\max_p(A_p)$;
        $A_{p^*} = A_{p^*} - T_k^{(B)}$;
        if $A_{p^*} \geq 0$ then
            $\widetilde{\mathcal{U}}_{p^*} = \widetilde{\mathcal{U}}_{p^*} \cup p^*$;
        else
            $k = -1$; break;
        end
    end
    if $k \neq -1$ then
        $\mathcal{H} = \mathcal{H}'$;
        $\mathcal{U}_d = \mathcal{U}_d \cup \widetilde{\mathcal{U}}_d; \forall d \in \mathcal{H}$
    end
end
Return the saved $\mathcal{H}$ and $\mathcal{U}$ as the hotspot network configuration.
```

Chapter 7

Evaluation Methodology and Results

We will begin by describing the simulation setup and results for the case when all nodes are connected to just one cell tower. Later we will describe the same for the case when there are more than one cell towers in the network.

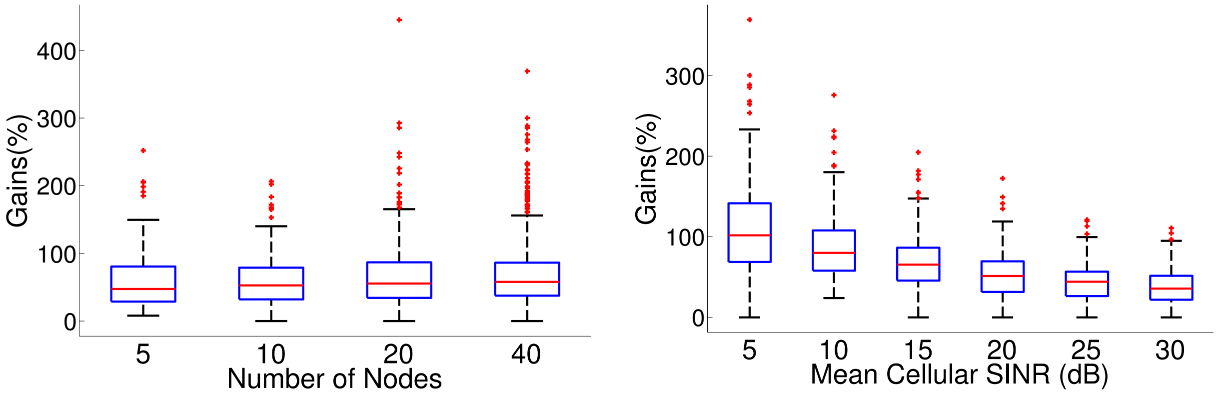
7.1 Single Cell Tower

We evaluate our proposed approach over simulated networks of 5, 10, 20, and 40 nodes. In the absence of real world data, we create networks that capture different cellular spreads for its nodes and varied WiFi connectivity between nodes. We assume that WiFi connectivity, given WiFi's small coverage, occurs in clusters of good connectivity. We also assume that cellular links suffer from log-normal shadowing. 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 10, 20 and 40 nodes, respectively, we choose up to 2, 4, and 8 clusters. Having selected the number of clusters, nodes are assigned to them randomly. If the random assignment results in one or more clusters being empty, the number of clusters for the resulting network is updated to the number of clusters that are non-empty after assignment.

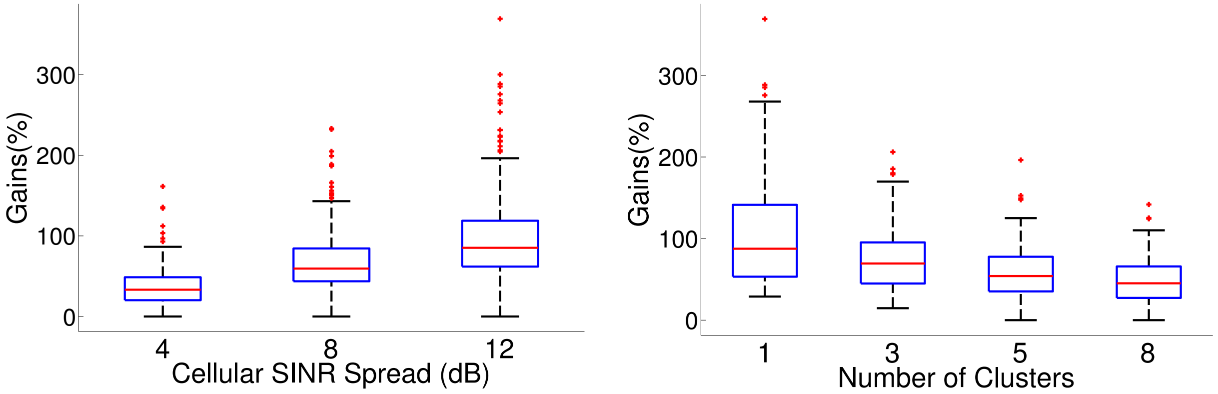
We choose the maximum probability of connection between clusters for a simulated network to be one of $0, 0.1, 0.2$. If the chosen maximum is p, then for the simulated network which now consists of some number of clusters of good WiFi connectivity, each WiFi link between nodes in different clusters is 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



(a) Gains vs. number of nodes.

(b) Gains vs. mean cellular SINR for 40 node networks.



(c) Gains vs. spread for 40 node networks.

(d) Gains vs. number of WiFi clusters for 40 node networks.

Figure 7.1: Boxplots of percentage gains in throughput.

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 σ, to give us the cellular SINR(s) of nodes in the network. We have considered the values of $\sigma = 4, 8, 12$ dB.

We evaluate our approach over a total of 90, 190, 400, and 940 networks, respectively, containing 5, 10, 20 and 40 nodes. Figure 7.1a shows the gains in throughput over baseline obtained for networks of size $5, 10, 20,$ and 40. Median gains of 1.5× are obtained. Figure 7.1b shows the gains obtained for networks of 40 nodes as a function of mean cellular SINR and Figure 7.1c shows the gains as a function of the SINR spread σ. 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 throughput to the cell tower. To exemplify consider two network with two nodes each. Network 1 has nodes with cellular SINR(s) of 17 and 20dB. Network 2 has cellular SINR(s) of 5 and 10 dB. Assuming good WiFi connectivity in both the networks, network 1 must select 20 as a hotspot and network 2 must select 10 as a hotspot. The resulting gains are 7.96% and 25.4%, respectively.

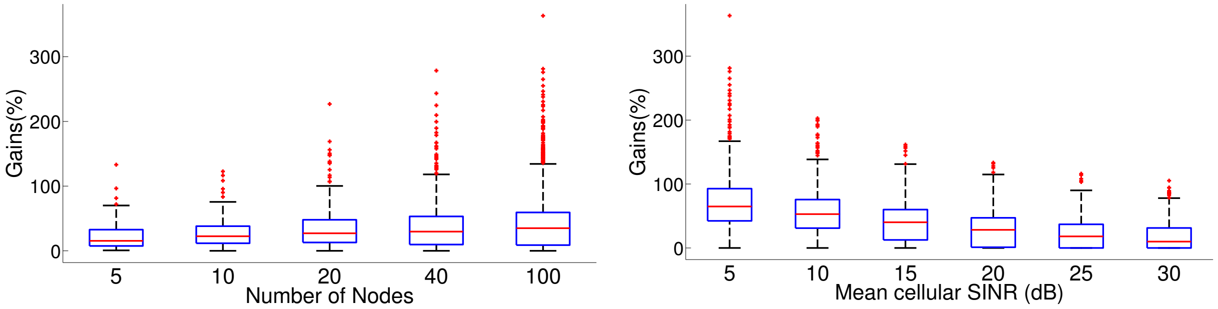
Similarly, in networks with large SINR spreads, a hotspot network configuration gives the nodes that have smaller SINR(s) the large throughput benefits of becoming clients of those that have much larger SINR(s). Finally, Figure 7.1d shows gains as a function of number of clusters. As is expected gains 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.

Finally, we did not see significant variations in gains as a function of maximum probability of connect between clusters.

7.2 More Than One Cell Tower

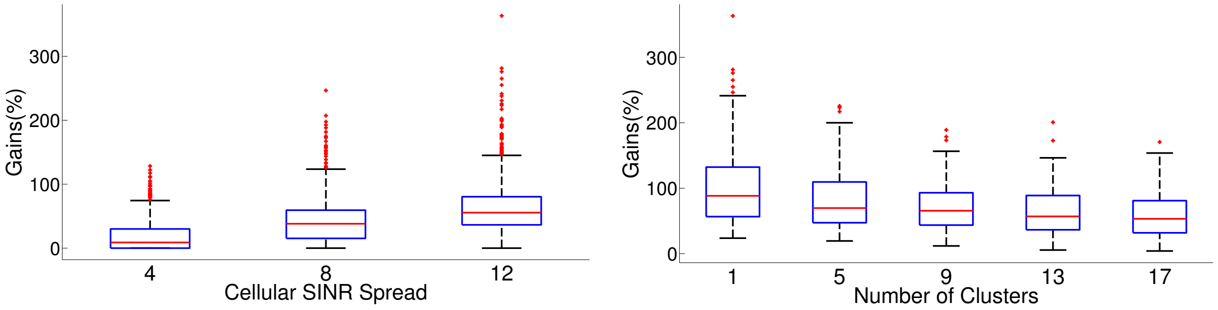
The simulation set up for multi cell tower is similar to the one for the single cell tower case. In addition to the network sizes for the single tower case, we simulated networks of 100 nodes. Also, the networks were assigned three cell towers. Each node in the network was assigned one of the three towers independently and randomly. Number of networks simulated with 5, 10, 20, 40, and 100 nodes were, respectively, 90, 190, 400, 940, and 3250.

Figure 7.2 shows the results obtained when each cell tower is optimized individually, as if the other cell towers are not present. That is the algorithm Rearrange (Algo 4) is not used. Figure 7.3 shows results obtained when using Algorithm 3, which calls Rearrange. Finally, in Figure 7.4, we show the percent improvement obtained when using Rearrange over optimizing each cell tower individually.



(a) Gains vs. number of nodes.

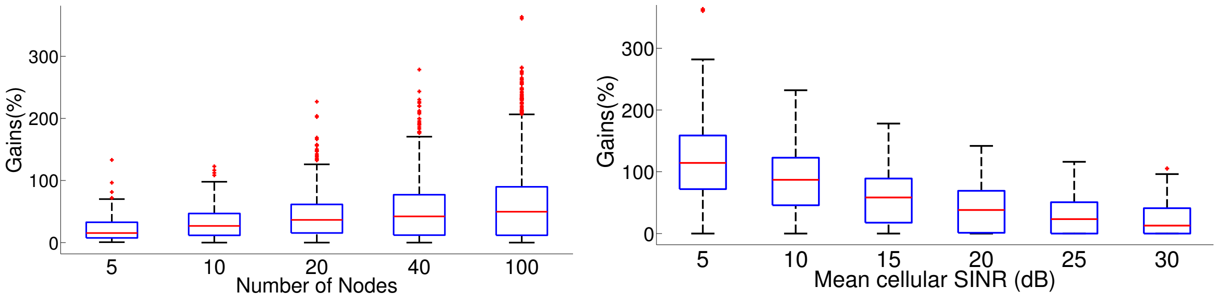
(b) Gains vs. mean cellular SINR for 100 node networks.



(c) Gains vs. spread for 100 node networks.

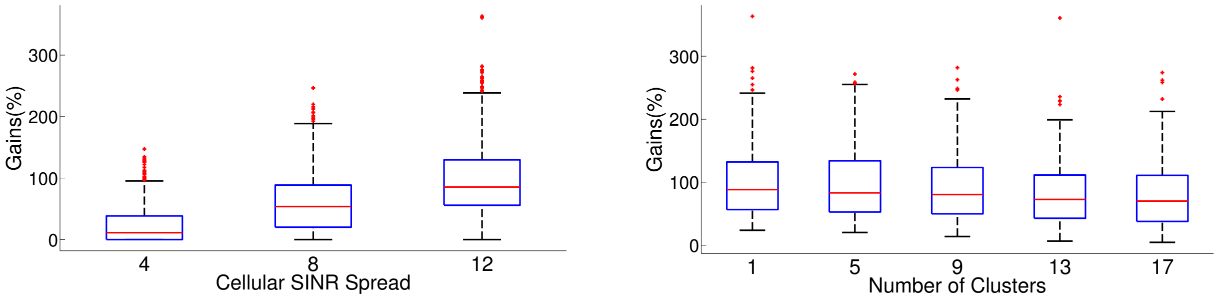
(d) Gains vs. number of WiFi clusters for 100 node networks.

Figure 7.2: Boxplots of percentage gains in throughput for a network having multiple towers. Each cell tower is optimized independently of each other. WiFi connectivity between nodes connected to different cell towers is not leveraged.



(a) Gains vs. number of nodes.

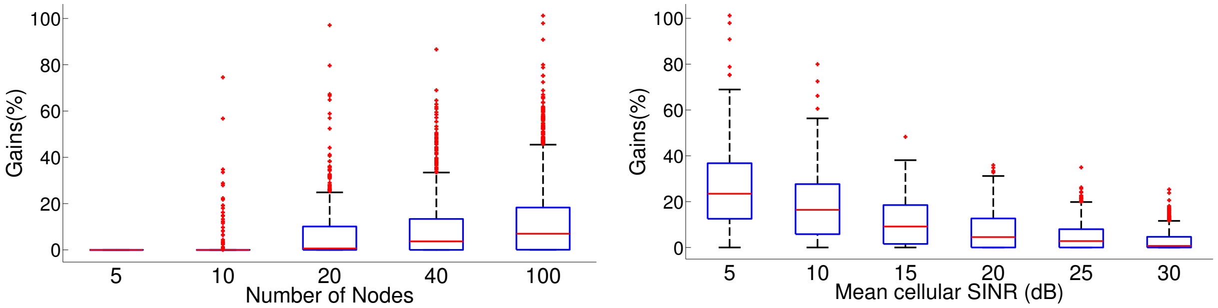
(b) Gains vs. mean cellular SINR for 100 node networks.



(c) Gains vs. spread for 100 node networks.

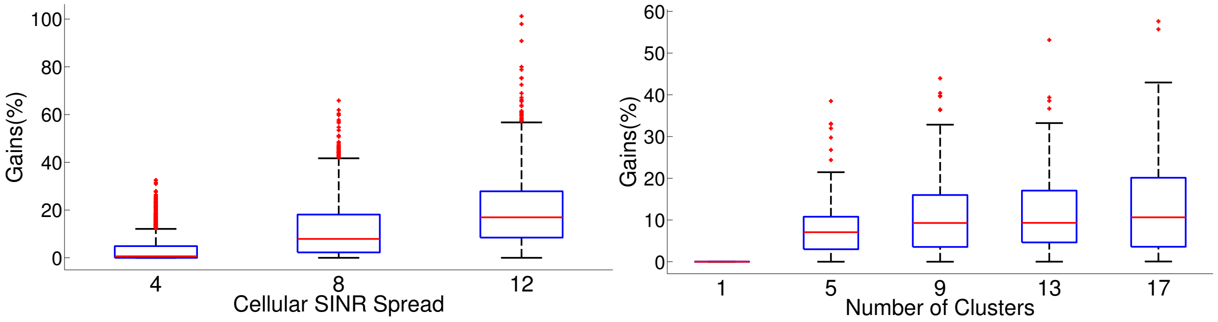
(d) Gains vs. number of WiFi clusters for 100 node networks.

Figure 7.3: Boxplots of percentage gains in throughput using Algorithm 3, which first optimizes cell towers individually and then invokes Rearrange (Algorithm 4).



(a) Gains vs. number of nodes.

(b) Gains vs. mean cellular SINR for 100 node networks.



(c) Gains vs. spread for 100 node networks.

(d) Gains vs. number of WiFi clusters for 100 node networks.

Figure 7.4: Boxplots of percentage improvement in throughput on using Rearrange over only optimizing cell towers independently of each other.

Chapter 8

Conclusions and Future Work

We formulated a network throughput optimization problem that splits users connected to the cellular network into users who will be configured as mobile hotspots and users who will access the internet by becoming clients of the mobile hotspots. We solved the problem under the constraint that all users must at least get the throughput they were getting when each one of them was directly connected to the cellular network. We proposed heuristic approaches to solving the problem for networks with one or more cell towers. The approach was evaluated a varied set of networks of up to 100 nodes. Median throughput gains of 1.5× were obtained.

In the future, we plan to incorporate a more accurate model of WiFi throughput.

Bibliography

[1] Andrews, J., Singh, S., Ye, Q., Lin, X., and Dhillon, H. An overview of load balancing in hetnets: Old myths and open problems. *Wireless Communications, IEEE 21*, 2 (2014), 18–25.

[2] Apostolaras, A., Iosifidis, G., Chounos, K., Korakis, T., and Tassiulas, L. C2m: Mobile data offloading to mesh networks. *University of Thessaly/CERTH-ITI, Tech. Rep* (2014).

[3] Asadi, A., and Mancuso, V. On the compound impact of opportunistic scheduling and d2d communications in cellular networks. In *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems* (2013), ACM, pp. 279–288.

[4] Asadi, A., Wang, Q., and Mancuso, V. A survey on device-to-device communication in cellular networks. *Communications Surveys & Tutorials, IEEE 16*, 4 (2014), 1801–1819.

[5] Bianchi, G. Performance analysis of the ieee 802.11 distributed coordination function. *Selected Areas in Communications, IEEE Journal on* (2000).

[6] Chen, X., and et al. Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks. *Selected Areas in Communications, IEEE Journal on* (2015).

[7] Cisco, C. V. N. I. Global mobile data traffic forecast update, 2013–2018. *white paper* (2014).

[8] Ding, A. Y., and et al. Enabling energy-aware collaborative mobile data offloading for smartphones. In *IEEE SECON 2013*.

[9] Fodor, G., Dahlman, E., Mildh, G., Parkvall, S., Reider, N., Miklós, G., and Turányi, Z. Design aspects of network assisted device-to-device communications. *Communications Magazine, IEEE 50*, 3 (2012), 170–177.

[10] Golrezaei, N., Mansourifard, P., Molisch, A. F., and Dimakis, A. G. Base-station assisted device-to-device communications for high-throughput wireless video networks. *Wireless Communications, IEEE Transactions on 13*, 7 (2014), 3665–3676.

[11] Iosifidis, G., and et al. An iterative double auction for mobile data offloading. In *IEEE WiOpt, 2013*.

[12] Karimi, O. B., Liu, J., and Rexford, J. Optimal collaborative access point association in wireless networks. In *INFOCOM* (2014), IEEE.

[13] Lee, K., Lee, J., Yi, Y., Rhee, I., and Chong, S. Mobile data offloading: how much can wifi deliver? In *Proceedings of the 6th International COnference* (2010), ACM, p. 26.

[14] Lin, X., Andrews, J. G., and Ghosh, A. Spectrum sharing for device-to-device communication in cellular networks. *Wireless Communications, IEEE Transactions on 13*, 12 (2014), 6727–6740.

[15] Nguyen-Vuong, Q.-T., and Agoulmine, N. Efficient load balancing algorithm over heterogeneous wireless packet networks. *REV Journal on Electronics and Communications* (2011).

[16] Nicholson, A. J., and Noble, B. D. Breadcrumbs: forecasting mobile connectivity. In *Proceedings of the 14th ACM international conference on Mobile computing and networking* (2008), ACM, pp. 46–57.

[17] Pekka, J., Chia-Hao, Y., Cassio, R., Carl, W., Klaus, H., Olav, T., and Visa, K. Device-to-device communication underlaying cellular communications systems. *Int'l J. of Communications, Network and System Sciences 2009* (2009).

[18] Qutqut, M. H., Al-Turjman, F. M., and Hassanein, H. S. Mfw: Mobile femtocells utilizing wifi: A data offloading framework for cellular networks using mobile femtocells. In *ICC* (2013), IEEE.

[19] Singh, S., and et al. Offloading in heterogeneous networks: Modeling, analysis, and design insights. *arXiv:1208.1977 [cs, math]* (Aug. 2012).

[20] Siris, V. A., and Kalyvas, D. Enhancing mobile data offloading with mobility prediction and prefetching. *ACM SIGMOBILE Mobile Computing and Communications Review* (2013).

[21] Srinivas, A., and Modiano, E. Joint node placement and assignment for throughput optimization in mobile backbone networks. In *INFOCOM 2008*, IEEE.

[22] Wei, L., Hu, R., He, T., and Qian, Y. Device-to-device (d2d) communications underlaying mu-mimo cellular networks. In *Global Communications Conference (GLOBECOM), 2013 IEEE* (2013), IEEE, pp. 4902–4907.

[23] Ye, Q., Rong, B., Chen, Y., Al-Shalash, M., Caramanis, C., and Andrews, J. G. User association for load balancing in heterogeneous cellular networks. *Wireless Communications, IEEE Transactions on 12*, 6 (2013), 2706–2716.

[24] Yu, C.-H., Doppler, K., Ribeiro, C. B., and Tirkkonen, O. Resource sharing optimization for device-to-device communication underlaying cellular networks. *Wireless Communications, IEEE Transactions on 10*, 8 (2011), 2752–2763.