Approaches to Joint Base Station Selection and Adaptive Slicing in Virtualized Wireless Networks

Kory A. Teague

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Electrical Engineering

Allen B. MacKenzie, Chair

Luiz DaSilva

R. Michael Buehrer

Mohammad J. Abdel-Rahman

June 1, 2018 (TBD)

Blacksburg, Virginia

Keywords: TBD

Approaches to Joint Base Station Selection and Adaptive Slicing in Virtualized Wireless Networks

Kory A. Teague

(ABSTRACT)

This work received support in part from the National Science Foundation via work involved with the Wireless @ Virginia Tech research group.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ever increasing desires of users to remain connected to one another and to access the wealth of knowledge and content available on the internet have driven explosive growth in both the volume of users and the demands of those users within mobile carrier networks. Previously unforeseen data-driven applications, such as audio/video streaming, social networking, and the Internet of Things have also placed increasing demand upon the networks. In 2016, the amount of IP data handled by mobile networks exceeded 86 Exabytes; it is projected to reach almost 200 Exabytes in 2018, and 580 Exabytes in 2021 [1]. Due to this exponential growth, incremental approaches to improve the network will fail to satisfy demand. As this growth continues to progress into the near future, new architectures like 5G and its associated technologies will be needed to keep pace with the demand.

However, deployment of these technologies and networks can be a costly, prohibitive venture. To meet these demands requires a similar increase in capital (CapEx) and operational expenditures (OpEx). As volumes and costs rise and margins shrink, approaches to reduce these expenditures become increasingly necessary. Resource infrastructure sharing has been a common practice for MNOs going back to 2G and 3G networks. First, MNOs needed to offer coverage for their users in regions where they had no infrastructure, leading to the creation of roaming agreements, eliminating the need for deploying new infrastructure in that region and reducing CapEx. Second, by sharing passive elements of the infrastructure, such as physical sites, tower masts, power, and air-conditioning, the CapEx of deploying new backhaul and radio access networks (RANs), such as cellular base stations (BSs), has decreased [2].

CapEx reductions from passive resource sharing drove an interest in resource sharing of the active elements of the network. For example, MNOs might share RANs, core networks, BSs, antenna systems, or reuse backhaul, which lead to reductions in both CapEx and OpEx. The ability to share these active resources removed the necessity for network operators to own and maintain a physical network while providing actual MNO-like mobile services. These mobile virtual network operators (MVNOs) functioned similarly to MNOs, but operating a virtualized wireless network (VWN) comprised of virtual resources instead of physical resources and their associated CapEx. It has been shown that virtualization in this manner can increase overall demand satisfaction of a set of VWNs while decreasing overall cost (i.e., OpEx) by decreasing the idle capacity of the networks [3].

With an increase of virtualizable resources and competition for those resources, a specific problem must be solved: how to select the set of virtual resources to form a VWN that meets its demands with necessary or maximum demand satisfaction at minimum cost. The solution to this problem is further complicated in the context of multiple MVNOs, each with one or more VWNs with unique demands, within a large pool of available virtual resources that can be adaptively allocated as demands shift. This thesis addresses this topic of resource selection and adaptive slicing within cellular VWNs through the lens of a stochastic optimization problem and a couple approaches to efficiently reach a solution.

1.1 Trends in Wireless Networks and Network Virtualization

IP traffic is increasing across all types of network archetypes, and is trending to become more mobile focused. According to the Cisco Visual Networking Index, global IP traffic will increase nearly threefold over the 2016-2021 time period, reaching 3.3 Zetabytes (ZB) annually in 2021 from 1.2 ZB annually in 2016. Traffic across the fixed internet backbone is projected to match this threefold pace, growing from 790 Exabytes (EB) to 2.2 ZB. However, mobile data traffic is projected to have twice the growth of fixed internet over the same period, increasing almost sevenfold from 86 EB in 2016 to 580 EB in 2021. By the end of 2021, 17% of global IP traffic will be mobile data or internet traffic generated by handsets, notebook cards, and mobile broadband gateways, up from 7.5% in 2016. More broadly, traffic from wireless and mobile devices combined will reach 63% of total IP traffic by 2021, up from 49%

in 2016. By 2021, smartphone IP traffic (33% of global IP traffic) will alone outnumber PC IP traffic (25%). Overall, these projections show that population data use and the number of devices attached to the networks is increasing; in 2021, global IP traffic will reach 35 GB per capita associated with 3.5 networked devices per capita, up from 13 GB per capita and 2.3 networked devices per capita in 2016. [1]

Consumer IP traffic comprises a majority share of global IP traffic, and this share will continue to grow. From 2016 to 2021, business IP traffic will grow at an average of 21% compound annual growth rate (CAGR), below the growth rate of global IP traffic (24% CAGR). Similarly, business mobile data traffic (41% CAGR) will lag behind that of global mobile data traffic (46% CAGR). [1]

In both the consumer and business markets, this demand includes enormous growth in video applications, specifically that of video streaming. By 2021, global IP video traffic will reach 82% of all consumer internet traffic, up from 73% in 2016. This is represented by a threefold increase of all global IP video traffic over the period and a fourfold increase on just the internet backbone. By 2021, internet live video streaming will account for 13% of this video traffic, growing 15-fold over the period. Similarly, internet video surveillance will occupy 3.4% of internet video traffic, increasing sevenfold, and internet video to tv will increase by a factor of three and a half, growing to comprise 26% of internet video traffic. While these are all values for increases on the internet backbone, mobile data should see similar, but smaller, growth in these areas. Virtual and augmented reality uses will see the largest increase, growing at a 82% CAGR, and expected to reach a 20-fold increase between 2016-2021. [1]

The technology underlying the mobile data network has needed to evolve with these changing trends and growth. The primary focus of the 5G cellular standard has been to meet these targets in an effective and robust manner. Of specific interest is that of aggregate data rate (e.g., area capacity, the available amount of data a network can facilitate over a unit area) and edge rate (e.g., 5% rate, the minimum data rate that can be reasonably provided to all but 5% of users) of the network. For 5G, the general consensus is that aggregate data rate and edge rate must be 1000x and 100x that of 4G, respectively [4]. To supply these rates, several strategies are being investigated, with three primary technologies being (1) the continuing of cellular densification and offloading, (2) increased bandwidth by expanding into new spectra like WiFi and millimeter wave, and (3) increasing spectral efficiency through advances such as those in massive multiple-input multiple-output (MIMO).

The first strategy is extreme densification and offloading. By making network cells smaller, the number of active nodes increases for the same unit area. This is a common strategy across cellular generations, and a large impetus behind the use of smaller range RANs like microcells and femtocells [5]. Cell sizes have shrunk, dropping from the order of hundreds of square kilometers to now fractions of a square kilometer. The most important benefit of cell densification is that it increases spectral reuse, which reduces the amount of users competing for the same resources. Theoretically, since signal-to-interference ratio is maintained as the cell shrinks, such densification can be repeated indefinitely as deployments allow [4, 6].

The second strategy is to increase bandwidth through the use of previously unused spectra like millimeter wave (mmWave) and WiFi. Cellular networks have utilized microwave frequencies ranging from a few centimeters to about a meter in wavelength; this range has become thoroughly occupied and to generate new bandwidth would require expanding to new frequencies *cite (1)*. Up to now, mmWave has been unused and in some cases unlicensed due to very poor propagation properties and high equipment costs. However, equipment costs are falling rapidly due to technological maturation. Further, the propagation qualities are increasingly surmountable as cell sizes shrink *cite (2)*.

The third strategy involves the use of massive MIMO to increase spectral efficiency. MIMO uses multiple transmit and receive antennas to exploit multipath signal propagation, multiplying the capacity of a given radio link. The technology has been used for over a decade as a component of WiFi before being introduced into the 3G, 4G, and 4G LTE standards [4]. A new approach to be used in 5G is that of "massive MIMO", where the number of transmit antennas at the BS greatly outnumber the number of active users [7]. For example, a given BS might have hundreds of antennas while maintaining data links for tens of users. This provides several benefits, most importantly that of vastly improving spectral efficiency.

5G must supply these rates at much higher energy and cost efficiencies, ideally matching or exceeding the capacity increases to avoid increasing overall network energy use and OPEX. However, these technologies that have been investigated to adequately increase the capacity of the network have several major hurdles to meet in order to be implemented at the desired energy and cost efficiencies. Massive MIMO requires the deployment of a vast number of antennas, which requires new BS architectures that have issues with scalability and cost. Millimeter-wave is more expensive than the more mature hardware of typical cellular bandwidths. Decreasing cell size for cellular densification allows for smaller, cheaper BSs, but this decreased cost may not keep pace with the required number of increased deployments. Additional measures to decrease expenditures on top of that of the aforementioned technologies is worth investigating. [4]

3GPP (the Third-Generation Partnership Project), is currently working on finalizing the standard for 5G implementations. In December 2017, 3GPP froze the first half of release 15 of the 5G standard, establishing the specifications of non-standalone 5G which utilizes existing LTE networks. It is expected that 3GPP will freeze the second half of release 15, covering 5G New Radio (5G-NR), which establishes specifications for new 5G deployments.

1.2 Virtualization, Virtualized Wireless Networks, and the Networks without Borders Paradigm

Discuss the concept of virtualization in a wider lens than in the intro lede. The terminology as it exists. The benefits it can provide.

Talk about the NwoB paradigm (Doyle paper). Lead into the specific virtualization model we use, it's various mechanisms and entities within it, and how they interact.

Through efficient sharing of radio resources, wireless virtualization is one the most promising approaches for reducing expenditures in next-generation mobile networks [4].

1.3 Review of Optimization Methods

Discuss the previously existing literature that I have been reading with regards to optimization methods. Discuss optimization and stochastic optimization. Discuss metaheuristic algorithms, especially genetic algorithms, and how they have been used for and to simplify optimization.

Considering the uncertainty of user equipment (demand point) locations, in this paper we consider solving two problems jointly. The first problem is to optimally orchestrate a set of BSs from a set of RPs to meet the demands of a set of SPs. The second problem is to slice the set of selected BSs adaptively (in response to the change in the users distribution) between the demand points of various SPs. Stochastic programming provides a powerful mathematical tool to handle optimization under uncertainty. It had been recently exploited to optimize resource allocation in various types of wireless networks operating under uncertainties (examples include [3, 8–13]). We establish a stochastic optimization problem from the perspective of the VNB to determine the optimal selection of resources to be leased from the RPs and adaptively sliced and allocated to form VWNs that satisfy the demands of the SPs. Then, we develop two approaches that would be run in the VNB to reach a near-optimal solution of the stochastic optimization problem. We consider two optimality criteria: maximizing demand satisfaction of SP users and minimizing the cost of the BS resources to satisfy the demand. Finally, we consider the efficacy of the approaches with a single SP with log-normal spatially-correlated demand modeled to mimic real cellular networks.

1.4 Thesis Objective

The objective of this thesis is to develop two approaches of joint resource allocation to construct a set of a VWNs and adaptively slice the selected resources to the individual VWNs. A model will be presented as the context for these approaches. The validity of this model will be restricted to the scope of cellular networks using generic base stations as its resources with perfect connections to demand points within range. The two proposed approaches will be performed within the VNB. Accordingly, efficacy of these approaches will be measured primarily by the optimality of the solutions and the run time, providing the VNB with a sufficient solution in a reasonable amount of time.

1.5 Thesis Outline

This thesis is organized as follows. Chapter 1 establishes the motivation for investigating resource selection for virtual network construction. Chapter 1 also presents the associated background information and literature review regarding virtualization, wireless networks, and optimization. Chapter 2 defines the model used for the resource allocation methods explored in this thesis. Further, Chapter 2 also details the two-stage stochastic optimization problem which optimally performs resource selection and slicing as a basis of approaches presented within this work. Chapter 3 establishes the two approaches investigated to provide solutions to the optimization problem posed in Chapter 2: a sampled Deterministic Equivalent Program which solves the problem as a whole and a genetic algorithm that simplifies the problem by providing an estimated optimal resource selection. Chapter 4 tests these two approaches by presenting four data sets that mimic real world cellular networks and evaluates the results. Chapter 5 contains the conclusions and proposed future work in this area.

Chapter 2

Virtual Network Builder Model

Replace with an introduction paragraph describing the work of this chapter. This chapter discusses the model upon which the approaches that are run within the VNB are defined. It discusses the assumptions that the VNB runs within and lays out the optimization problem that the VNB is tasked with solving.

Begin breaking down the model used as the basis for the work. Start with a lead in, then start defining the model in the first subsection. Expound on some of the definitions and descriptions that I moved past in the conference paper (perhaps describing the SSLT model more fully). Make sure that the definitions and work are generalized for writing about a constructed VWN (or constructed VWNs, depending on the terminology I intend to use) for multiple RPs and SPs.

2.1 Network Area Definitions

Change "we"s to "I"s or alternate voice. Expand SSLT description: a mention of ρ^S, the specific autocorrelation which is dependent on $\omega_{\max}$, the differences between my implementation and Lee's, and any other pertinent information (review "Lee, Zhou, and Niu" [14]). Expound directly on the assumption I make that demand has a perfect connection within range and no connection out of range (i.e., $u_{ms} = 1$ if DP m is less than or equal to b_s distance from BS s and $u_{ms} = 0$ otherwise). Expound on nonstationary PPP procedure; provide more information laying out the process, and include the possibility of generating a set number of demand points if desired - this does alter the PPP slightly, in that the PPP no longer has a specific overall density of points, but it does maintain the structure of the overall density function, scaled to a desired degree. Ensure equations are spaced well for full page, and devoid of unnecessary horizontal and vertical spacing (from when the conference paper was trimmed). A figure demonstrating continuous SSLT model and a possible scenario of demand points would be appropriate here. Improve on the overall wording/phrasing regarding the capability for multiple RPs and SPs to be present; that is BSs come from the overall pool of RP resources which the VNB is selecting to build its VWNs of interest, then slices those selected resources into forming the specific, individual VWNs for each SP. When referring to "Problem"s (e.g., Problem 1 from section 2.2), refer to the equations that make up that problem; see first reference of Problem 1 in Chapter 3 for reference.

We consider a geographical area of width X (m) and length Y (m) that contains a set

$\mathcal{S} \stackrel{\text{def}}{=} \{1,\, 2,\, \ldots,\, S\}$ of BSs available to be leased to the VNB by a set of $\mathcal{N} \stackrel{\text{def}}{=} \{1,\, 2,\, \ldots,\, N\}$ RPs. The rate capacity of BS $s \in \mathcal{S}$ is denoted by r_s, its cost is denoted by c_s, and its coverage radius is denoted by b_s.

A Service Provider (SP) seeking a virtualized wireless network from the VNB is assumed to know the distribution of traffic demand within the region the VWN would cover. It has been shown that a log-normal distribution or a mixture of log-normal distributions can approximate traffic demand in real-world cellular networks [15, 16]. It has also been shown that traffic distribution is spatially correlated [16, 17]. We model the spatial traffic demand of a single SP using a similar, continuous form of the SSLT (Scalable, Spatially-correlated, and Log-normally distributed Traffic) model as proposed by Lee, Zhou, and Niu [14].

To generate this spatial distribution over the area of consideration, an initial Gaussian field, $\rho^G = \rho^G(x,\, y)\,,\, x \in [0,\, X]\,,\, y \in [0,\, Y]$, is generated by

$$\rho^G(x,\, y) = \frac{1}{L} \sum_{l=1}^{L} \cos\left(i_l x + \phi_l\right)\, \cos\left(j_l y + \psi_l\right) \tag{2.1}$$

where $\mathcal{L} \stackrel{\text{def}}{=} \{1,\, 2,\, \ldots,\, L\}$ is a set of the products of two cosines with angular frequencies $i_l,\, j_l \sim \mathcal{U}\left(0,\, \omega_{\max}\right),\, l \in \mathcal{L}$ and phases $\phi_l,\, \psi_l \sim \mathcal{U}\left(0,\, 2\pi\right),\, l \in \mathcal{L}$. As L increases, ρ^G approaches a Gaussian random field with a spatial autocorrelation dependent on $\omega_{\max}$ according to the central limit theorem.

The approximate Gaussian distribution ρ^G is then normalized to a standard normal distribution. The final log-normal distribution, $\rho = \rho(x,\, y)\,, x \in [0,\, X], y \in [0,\, Y]$, is determined by assigning location and scale parameters

$$\rho(x,\, y) = \exp\left(\frac{\sigma}{\sqrt{\mathrm{Var}\left(\rho^G\right)}}\, \rho^G(x,\, y) + \mu\right) \tag{2.2}$$

where $\mathrm{Var}\left(\rho^G\right)$ is the variance of ρ^G.

$\rho(x,\, y)$ can be sampled over the space into individual pixels as per Lee with each pixel's value indicating the number of homogeneous demand points within the pixel [14]. In contrast, we allow $\rho(x,\, y)$ to provide a continuous, spatially-correlated log-normal distribution depicting the demand density over the region for the SP.

Let $\mathcal{M} \stackrel{\mathsf{def}}{=} \{1,\, 2,\, \ldots,\, M\}$ be the set of the SP's demand points seeking to connect to the VWN; the value of total traffic demand at each point is denoted by d_m. Further, let $u_{ms} \in [0,\, 1]$, $m \in \mathcal{M}$, $s \in \mathcal{S}$, represent the normalized capacity (with respect to r_s) of BS s at point m, i.e., the normalized maximum rate that a user can receive at point m from BS s. $u_{ms} = 0$ when m is outside the coverage area of s and $u_{ms} = 1$ when m is within a small distance of s. The specific position of the points in $\mathcal{M}$, and therefore the values of u_{ms}, is determined via a non-stationary 2D Poisson point process (PPP) with M points using the demand field, ρ, as the spatial intensity function. To generate this non-stationary PPP, we use an acceptance-rejection method. Each point of a stationary PPP with an intensity of $\rho_{\max} = \max_i \rho(x_i,\, y_i)$ is retained with probability $\frac{\rho(x_i,\, y_i)}{\rho_{\max}}$, where x_i and y_i are the x- and y-coordinates of the i^{th} point of the stationary PPP.

We assume that a BS $s \in \mathcal{S}$ can be allocated between multiple demand points, and $\delta_{ms} \in$

$[0,\, r_s]$, $m \in \mathcal{M}$, $s \in \mathcal{S}$, represents the rate of BS s that is allocated to point m.

Throughout this paper, stochastic variables will be differentiated from deterministic variables with a tilde ($\sim$) placed above the symbol.

2.2 Stochastic Optimization

Replace "we"s to "I"s or find alternate wording/tense/voice. Ensure equations are spaced appropriately for full page column and that there are no unnecessary vertical or horizontal spacing. Expounding on the various components of the stochastic optimization problem might be worthwhile. Might be worthwhile to also mention that the stochastic nature of this specific formulation is limited to handling stochastic demand point locations. As with 2.1, modify wording and phrasing accordingly to accommodate the possibility for multiple RPs and SPs in the model. When referring to "Problem"s (e.g., Problem 1 from section 2.2), refer to the equations that make up that problem; see first reference of Problem 1 in Chapter 3 for reference.

We formulate the presented problem as a two-stage stochastic optimization problem. We introduce $z_s, s \in \mathcal{S}$ as a binary decision variable defined as

$$z_s \stackrel{\mathsf{def}}{=} \begin{cases} 1, & \text{if BS } s \text{ is selected for the created VWN,} \\ 0, & \text{otherwise.} \end{cases}$$

To balance the interest of maximizing demand satisfaction against minimizing cost, we introduce the positive real number α as a weighting coefficient between the two stages.

Problem 1 (Two-Stage Stochastic Optimization Problem)

$$\underset{\{z_s,\, s \in \mathcal{S}\}}{\text{minimize}} \left\{ \sum_{s \in \mathcal{S}} c_s \; z_s + \alpha \mathbb{E}\left[h\left(z,\, u\right)\right] \right\} \tag{2.3}$$

subject to:

$$z_s \in \{0, 1\}, \forall s \in \mathcal{S} \tag{2.4}$$

where $h(z, u)$ is the optimal value of the second-stage problem, which is given by:

$$\underset{\{\delta_{ms}, m \in \mathcal{M}, s \in \mathcal{S}\}}{\text{minimize}} \left\{ -\sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \delta_{ms} \; \tilde{u}_{ms} \right\} \tag{2.5}$$

subject to:

$$z_s = \mathbb{1}_{\left\{\sum_{m \in \mathcal{M}} \delta_{ms} > 0\right\}}, \forall s \in \mathcal{S} \tag{2.6}$$

$$\sum_{s \in \mathcal{S}} \delta_{ms} \; \tilde{u}_{ms} \leq d_m, \forall m \in \mathcal{M} \tag{2.7}$$

$$\sum_{m \in \mathcal{M}} \delta_{ms} \leq r_s, \forall s \in \mathcal{S}. \tag{2.8}$$

The first stage objective function (2.3) minimizes the total cost of the selected network with respect to that network's ability to satisfy the demand contained within the region. The second stage objective function (2.5) maximizes demand satisfaction by maximizing the total demand allocated to the resources comprising the network, as specified by δ_{ms} as the decision variable of the second stage.

Constraints (2.4), (2.6), and (2.8) implement the defined ranges and values of the decision variables z_s and δ_{ms}, with (2.6) ensuring that demand is allocated only to selected resources. For constraint (2.6), $\mathbb{1}_{\{*\}}$ is defined by

$$\mathbb{1}_{\{*\}} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if condition } \{*\} \text{ is true,} \\ 0, & \text{otherwise.} \end{cases}$$

Constraint (2.7) ensures a demand point $m \in \mathcal{M}$ is not allocated more resources than it demands.

Chapter 3

Approximation Approaches

Replace with introduction paragraph to this chapter. This chapter lays out the two major approaches I am using in this thesis: the sampled DEP and genetic algorithm, and the foundation those approaches are built on. These approaches are built on the stochastic optimization problem as laid out in 2.2, and meant to provide a solution (i.e., the DEP) or an estimate (i.e., the sampled DEP and genetic algorithm) as the original problem is not directly solvable.

In this chapter, I work on defining the approximation approaches used in my work. Lead in to discussing the need to approximate the stochastic optimization problem from section 2.2 to adequately solve my work, then introduce the two approaches I used to approximate the optimization problem: the DEP/its sampling/generalized post-selection slicing and the genetic algorithm as a selection method.

3.1 Deterministic Equivalent Program

Replace all "we"s with "I"s or with alternate phrasing/tense/voice. When referring to "Problem"s (e.g., Problem 1 from section 2.2), refer to the equations that make up that problem; see first reference of Problem 1 in Chapter 3 for reference.

Introduce the idea of a DEP as an approach for solving the original stochastic problem. Present the solved problem here in the form of the true deterministic equivalent program - as in, it is actually an equivalent to the original stochastic problem - with all the necessary expansions and additional variables. Focus on how this formulation no longer includes any stochastic variables and is purely deterministic. Mention that the trade off is that the deterministic variables are part of a infinitely large set of potential scenarios.

In order to solve the two-stage stochastic optimization formulation (Problem 1, eqs. (2.3)–(2.8)), we need to convert it to a deterministic equivalent program (DEP) that does not contain any stochastic variables (only deterministic variables) [18].

Let Ω be defined as the sample space, i.e., the set of all scenarios. Let $\hat{\Omega} \stackrel{\text{def}}{=} \{1, 2, \ldots, O\}$ be a discrete set containing sampled scenarios. The probability a given scenario $\omega \in \hat{\Omega}$ occurs is denoted by $p^{(\omega)}$, $\omega \in \hat{\Omega}$, where $\sum_{\omega \in \hat{\Omega}} p^{(\omega)} = 1$. Variables that are dependent on the scenario are shown with a superscript (ω) with the specific scenario it is dependent on indicated by ω.

Problem 2 (Deterministic Equivalent Program of Problem 1)

$$\underset{\left\{\substack{z_s,\,\delta_{ms}^{(\omega)},\\ s\in\mathcal{S},\,m\in\mathcal{M},\\ \omega\in\hat{\Omega}}\right\}}{\text{minimize}} \left\{\sum_{s\in\mathcal{S}} c_s\ z_s - \alpha \sum_{\omega\in\Omega} p^{(\omega)} \left(\sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \delta_{ms}^{(\omega)}\ u_{ms}^{(\omega)}\right)\right\} \tag{3.1}$$

subject to:

$$\sum_{s\in\mathcal{S}} \delta_{ms}^{(\omega)}\ u_{ms}^{(\omega)} \leq d_m,\ \forall m \in \mathcal{M},\ \forall \omega \in \hat{\Omega} \tag{3.2}$$

$$\sum_{m\in\mathcal{M}} \delta_{ms}^{(\omega)} \leq r_s\ z_s,\ \forall s \in \mathcal{S},\ \forall \omega \in \hat{\Omega} \tag{3.3}$$

$$z_s \in \{0,\ 1\},\ \forall s \in \mathcal{S}. \tag{3.4}$$

The objective function (3.1) combines both objective functions (2.3) and (2.5) of the initial formulation into a deterministic form. Constraints (3.2) and (3.3) ensure demand is not overallocated and is only allocated to selected resources and within capacity for all scenarios.

Problem 2 provides an equivalent deterministic form of Problem 1 for the (finite) sampled state space, $\hat{\Omega}$, containing O scenarios. With sufficiently large O, $\hat{\Omega}$ approaches a tight approximation of the original sample space. Within each scenario $\omega \in \hat{\Omega}$, the SSLT demand field ρ is sampled to provide a set of M discrete demand points. Each sampling of ρ is generated by creating a non-stationary 2D PPP with M points as described in Section **??**.

3.1.1 Sampling Approaches

As the infinitely large set of scenarios renders the problem unable to be solved, it needs to be sampled into a finite set to be solved. Present the structure and nomenclature used to imply a sampled set of scenarios, and describe the structure of how the scenarios are sampled into a truncated set. Might be worth mentioning that there are other methods that might be better for sampling beyond the completely random sampling approach I am using. Worth consideration?

Sample Average Approximation

At what point is the sampling enough? As the set of scenarios considered within the sampled DEP increases, it more closely compares to the original DEP and the stochastic optimization problem, but it also becomes increasingly difficult to solve as the number of scenarios considered increases. So, it is beneficial to understand that a certain known number of scenarios provides a reasonably tight - what does reasonable mean? - solution to the original DEP to avoid being unnecessarily computationally expensive to solve. Finding this minimum necessary number of scenarios can be done via a sample average approximation (SAA) analysis, which should not be too complicated to do.

3.1.2 Adaptive Slicing

Now that we have a (close) approximation to the DEP and the original stochastic optimization problem, we have a method for deriving the minimum cost BS selection and adaptive slicing for the desired VWN. However, this selection is overly time consuming to constantly run, and the BSs selected for the VWN(s) by the VNB are fairly constant, so all that is needed is to dynamically (read: adaptively) slice the selected BSs to the various SPs. To do this, we simplify the sampled DEP such that it has only one scenario - ostensibly, the current scenario in time - and the BSs selected set to be a constant rather than a decision variable. The resulting problem is a single stage linear program that is much simpler to solve. This is used to adaptively slice resources to the demand.

After the solution to the sampled DEP of Section **??** has been found, the VNB has determined the joint BS selection that forms the VWN and a proposed resource slicing of considered possible scenarios, $\hat{\Omega}$, that allocates the resources to the SP's demand points. Since O is not infinite, any given scenario present in the formed VWN is unlikely to be an element of $\hat{\Omega}$. Further, as demand points move between BSs or enter or exit the VWN, a new scenario $\omega \notin \hat{\Omega}$ is formed. The VWN must adapt its resource slicing to these new demand points to maintain maximal demand satisfaction. With the VWN built, the joint BS selection, z_s, becomes a constant of the network, simplifying Problem 2 to a single-stage optimization problem.

Problem 3 (Deterministic Adaptive Slicing)

$$\underset{\{\delta_{ms},\, s\in\mathcal{S},\, m\in\mathcal{M}\}}{\text{maximize}} \left\{ \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \delta_{ms}\, u_{ms} \right\} \tag{3.5}$$

subject to:

$$\sum_{s\in\mathcal{S}} \delta_{ms}\, u_{ms} \leq d_m,\ \forall m \in \mathcal{M} \tag{3.6}$$

$$\sum_{m\in\mathcal{M}} \delta_{ms} \leq r_s\, z_s,\ \forall s \in \mathcal{S}. \tag{3.7}$$

It is worth noting that Problem 3 is more tractable than Problem 2 as it only contains the single continuous decision variable for resource slicing, simplifying the objective function (3.5) and constraint (3.7) from a mixed integer linear program to a linear programming problem.

3.2 Genetic Algorithm

Now that the first approach - DEP and its sampling - has been tackled, and the necessary tool to evaluate it has been derived from it - the simplified adaptive slicing program - move on to the genetic algorithm approach for approximating the BS selection process. Discuss the core algorithm of a genetic algorithm, then the various approaches that I used in its process (e.g., binary chromosomes, elitism, uniqueness, uniform crossover, bitwise mutation).

The Problem 2 formulation becomes intractable as O, S, or M increases. Most importantly, the accuracy of the sampled DEP is directly dependent on the size of $\hat{\Omega}$, O, directly causing a trade off between the accuracy of the sampled DEP and its computability in a reasonable amount of time. In this subsection, we reformulate the problem of joint BS selection for the VWN as a genetic algorithm, circumventing the need to discretize demand or to establish $\hat{\Omega}$, thereby simplifying the original problem into a more scalable form.

A genetic algorithm is an iterative metaheuristic in which an approximate solution to a given optimization problem is arrived at via a series of progressive generations. Each generation contains a number of candidate solutions, called individuals, each of which is defined by a chromosome. During a given generation, a fitness heuristic is assessed for each individual based on its chromosome. Then individuals are selected at random, with more fit individuals being selected with higher probability. Pairs of selected individuals will crossover with probability p_{xov}, a process similar to genetic recombination in biology. The resulting chromosomes then have probability p_{mut} to mutate, altering the chromosome slightly. Once enough new individual chromosomes have been selected and possibly undergone crossover and mutation, this set of new individuals, called children, forms the next generation to repeat the process.

For the genetic algorithm, ρ is not sampled for discrete demand points. Instead, we assume that all demand over the region is allocated to the closest resource. The subset of $\mathcal{S}$, $\mathcal{S}'$, that is selected for a given possible VWN forms a Voronoi tessellation from the point locations of the selected resources. The total demand allocated to a selected resource $s \in \mathcal{S}' \subseteq \mathcal{S}$ is $\iint_{V_s} \rho\,(x,\, y)\; dx\, dy$, where V_s is the region bounded by the cell of resource s in the Voronoi tessellation. If the total demand allocated to s exceeds r_s, s is considered to be *overcapacity.*

If V_s is not wholly contained within the coverage area of resource s, s is considered to be *overcoverage.*

Let $\mathcal{G} \stackrel{\text{def}}{=} \{1,\, 2,\, \ldots,\, G\}$ be the set of generations used in the genetic algorithm and $\mathcal{I}_g \stackrel{\text{def}}{=} \{1,\, 2,\, \ldots,\, I\}\,, g \in \mathcal{G}$ be the set of individuals within generation g. Each individual $i \in \mathcal{I}_{g\in\mathcal{G}}$ has a binary chromosome $z^{\{ig\}}$ of length S. $z_s^{\{ig\}}, s \in \mathcal{S}$, denoting each individual bit of the chromosome, is defined as follows:

$$z_s^{\{ig\}} = \begin{cases} 1, & \text{if BS } s \text{ is selected for the VWN forindividual } i \text{ in generation } g, \\ 0, & \text{otherwise} \end{cases}$$

The fitness heuristic of each individual chromosome, $z^{\{ig\}}$, is assessed as the reciprocal of the chromosome's cost, which is defined as

$$\text{fitness}\left(z^{\{ig\}}\right) = \frac{1}{\text{cost}\left(z^{\{ig\}}\right)} \tag{3.8}$$

$$\begin{aligned} \text{cost}\left(z^{\{ig\}}\right) = \sum_{s\in\mathcal{S}} \Bigg(& c_s\; z_s^{\{ig\}} + c_{\text{cov}}\; \mathbb{1}_{\{V_s \not\subseteq R_s\}} + \\ & \left(c_{\text{cap}}^g - 1\right)\; \max\left(0,\; \iint_{R_s} \rho\left(x,\, y\right)\, dx\, dy - r_s\right)\Bigg) \end{aligned} \tag{3.9}$$

where R_s is the coverage area region of resource $s \in \mathcal{S}$.

The cost function (3.9) indicates cost increases not only based on the cost of the resources selected, but also with imperfection costs c_{cov} and c_{cap}, the costs of a selected resource being overcoverage or overcapacity, respectively. The overcapacity cost grows with each successive generation. For early generations, this allows for imperfect solutions to temporarily exist to seed later generations and improve diversity to increase the probability of finding a better final approximate solution.

Elitism is used, where the n most fit individuals of a given generation are automatically selected without crossover or mutation to be the first children of the next generation. Selection occurs via the roulette wheel selection method. Every individual i of a given generation g has a probability of being selected given by

$$\frac{\text{fitness}\left(z^{\{ig\}}\right)}{\sum_{i\in\mathcal{I}}\text{fitness}\left(z^{\{ig\}}\right)}$$

When crossover is performed on selected individuals, it is via the uniform crossover method with a mixing ratio of 0.5. That is, if two selected parent individuals crossover, each equivalent bit in the parents will swap with a probability of 50%. Mutation occurs on a bit-by-bit level, with each bit mutating (i.e., flipping) with probability $\frac{1}{S}$. The uniqueness property is then enforced on the resulting children to ensure diversity; if a child chromosome is identical to another child chromosome in the next generation, the child is discarded and a new child generated, ensuring that each individual of any given generation is unique within that generation.

The genetic algorithm iterates for a number of generations G. If the genetic algorithm settles on a single individual for a number of continuous generations, G_{halt}, it will halt and present that individual's chromosome as the final approximate solution for z_s. Otherwise, the chromosome of the fittest individual of generation G determines z_s.

The genetic algorithm only determines an approximate solution to the BS selection forming the VWN, informing the VNB of which BSs to obtain from the RPs. With this selection, z_s, the SP's demand points can be dynamically allocated resource slices as described by Problem 3 in Section 3.1.2.

Chapter 4

Testing and Simulations

In this chapter I will be introducing four different cases to test the provided approximation approaches. The first will be the test case used in my conference paper (one SP, with homogeneous resources). The second will be an expansion of the test case used in my conference paper, but with heterogeneous resources. The third will extend to service multiple similar cellular SPs. The fourth will extend to a case with multiple SPs with various, specialized demands.

4.1 VWN Construction for a Single SP

Lead into the first two cases, which test the approaches while using a single SP.

4.1.1 Case I: Homogeneous Urban Cellular Network

Basically as presented in my conference paper. One SP, homogeneous resources within the RPs. Might need to use a new data set, though, with a larger data set.

4.1.2 Case II: Impact of Heterogeneous Resources

Same as Case I, but with heterogeneous resources within the RPs. Need to understand how this changes the approaches.

4.2 VWN Construction for Multiple SPs

Lead into the second two cases (should I have more?), which test using multiple SPs to satisfy from the same set of resources.

4.2.1 Case III: Two Similar Urban Cellular Networks

First consider a case with two SPs with similar demands. Overlapping cellular networks. Could see how the approaches behave while two SPs partially overlap.

Homogeneous Resources

If it appears that the difference between Case I and Case II (sections 4.1.1 and 4.1.2) is worth further consideration, then analyze here with homogeneous resources. Otherwise, a single comparison should be sufficient.

Heterogeneous Resources

As for the previous subsubsection (4.2.1), but consider with heterogeneous resources.

4.2.2 Case IV: SPs with Specialized Demands

This is the major case that is the extension of my work. Case I (4.1.1) analyzed what happens with a single SP, Case II (4.1.2) expanded that to heterogeneous resources, and Case III (4.2.1) added an additional similar SP, but Case IV considers when there are several SPs and with their own considerations and unique demands. Need to consider what these SPs look like. One would be a cellular network like in Case I (moderate to high number of users, moderate demand). Another could be a streaming service (few users, high individual demand). Another an emergency service (very low number of users and demand, but requiring virtually 100% demand satisfaction - see note below). What other SPs should I consider?

***Note**: I need to consider how to accurately label demand satisfaction within the approaches. In effect, this would be controlled by α for the (sampled) DEP and controlled by β or some such for the genetic algorithm. I should investigate this at some point of the thesis, probably* *within their appropriate sections in chapter 3 (DEP: 3.1 and GA: 3.2).*

Homogeneous Resources

As for Case III (4.2.1), if a considerable difference was detected between Cases I and II (4.1.1 and 4.1.2), consider analyzing the case with homogeneous resources and

Heterogeneous Resources

also with heterogeneous resources.

Chapter 5

Conclusions

Consider conclusions of my work. I don't think this chapter would be long, but condense my findings into some coherent thoughts, and redirect to what they are. Also expound on some of the further work that my research could be expanded to (e.g., further use cases investigating my approaches, use of (meta)heuristics other than a genetic algorithm to approximate the optimization problem, improve the basic capacity function used in my optimization model).

Bibliography

[1] Cisco, "Cisco visual networking index: Forecast and methodology, 2016-2021," [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf, June 2017, white paper at Cisco.com.

[2] C. Beckman and G. Smith, "Shared networks: making wireless communication affordable," *IEEE Wireless Communications Magazine*, vol. 12, no. 2, pp. 78–85, April 2005.

[3] M. J. Abdel-Rahman, K. Cardoso, A. B. MacKenzie, and L. A. DaSilva, "Dimensioning virtualized wireless access networks from a common pool of resources," in *Proceedings of the IEEE CCNC Conference*, January 2016, pp. 1049–1054.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[5] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, September 2008.

[6] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, “Modeling and analysis of k-tier downlink heterogeneous cellular networks,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, April 2012.

[7] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[8] M. J. Abdel-Rahman and M. Krunz, “Stochastic guard-band-aware channel assignment with bonding and aggregation for DSA networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3888–3898, July 2015.

[9] N. Y. Soltani, S. J. Kim, and G. B. Giannakis, “Chance-constrained optimization of OFDMA cognitive radio uplinks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1098–1107, March 2013.

[10] K. V. Cardoso, M. J. Abdel-Rahman, A. B. MacKenzie, and L. A. DaSilva, “Virtualization and programmability in mobile wireless networks: Architecture and resource management,” in *Proceedings of the Workshop on Mobile Edge Communications (MECOMM’17)*, 2017, pp. 1–6.

[11] M. J. Abdel-Rahman, M. AbdelRaheem, A. B. MacKenzie, K. Cardoso, and M. Krunz, “On the orchestration of robust virtual LTE-U networks from hybrid half/full-duplex Wi-Fi APs,” in *Proceedings of the IEEE WCNC Conference*, April 2016.

[12] M. J. Abdel-Rahman, M. AbdelRaheem, and A. B. MacKenzie, "Stochastic resource allocation in opportunistic LTE-A networks with heterogeneous self-interference cancellation capabilities," in *Proceedings of the IEEE DySPAN Conference*, September/October 2015, pp. 200–208.

[13] R. Atawia, H. Abou-zeid, H. S. Hassanein, and A. Noureldin, "Joint chance-constrained predictive resource allocation for energy-efficient video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1389–1404, May 2016.

[14] D. Lee, S. Zhou, and Z. Niu, "Spatial modeling of scalable spatially-correlated log-normal distributed traffic inhomogeneity and energy-efficient network planning," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 1285–1290.

[15] U. Gotzner and R. Rathgeber, "Spatial traffic distribution in cellular networks," in *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, May 1998, pp. 1994–1998, vol. 3.

[16] M. Michalopoulou, J. Riihijrvi, and P. Mhnen, "Towards characterizing primary usage in cellular networks: A traffic-based study," in *Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, May 2011, pp. 652–655.

[17] J. Reades, F. Calabrese, and C. Ratti, "Eigenplaces: analysing cities using the space - time structure of the mobile phone network," *Environment and Planning B: Planning and Design*, vol. 36, pp. 824–836, 2009.

[18] P. Kall and S. W. Wallace, *Stochastic Programming.* John Wiley and Sons, 1994.