Multi-Network Access: How It Can Benefit Network Operators and Consumers Alike

Submitted in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy in

Engineering and Public Policy

Nandi Zhang

B.S., Electrical Engineering, Purdue University
M.S., Engineering and Public Policy, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

December, 2019

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisors and thesis committee co-chairs, Dr. Jon Peha and Dr. Marvin Sirbu, without whose support I would not be where I am today. I am one of the lucky students who had two Ph.D. advisors. There was no primary or secondary, because both Jon and Marvin devote tremendous time and effort to my success. I get to meet with both of them nearly every week (or more frequently if needed), which I do not take for granted since there are students unable to get hold of their advisor on a regular basis. In almost every meeting there is a fruitful discussion of previous results and next steps. I say almost because there were times when I just was not productive, but Jon and Marvin always kept their patience. Sometimes opinions differ, but Jon and Marvin never put me in a position where I had to side with one of them and worry about offending the other. They are always focused on *my* success. Over the years I have been inspired by their vast knowledge of communications, computing and the economics and public policy thereof. Not only that, Jon and Marvin can do back-of-the-envelope math so much quicker than I do that it makes me question my Chinese heritage (and age). Joking aside, I am even more motivated by their scientific rigor and unwavering curiosity for the world, especially given how much they know about it already! What I learned from shadowing them as researchers is perhaps more valuable than my diploma, and I know I will be drawing from that experience for many more years down the road.

I am also profoundly grateful to other members of the thesis committee, Dr. Carlee Joe-Wong from Electrical and Computer Engineering and Dr. Martin Weiss from University of Pittsburgh (and EPP alumni). Carlee's many constructive comments encourage me to dig deeper about why certain trends appear in our results; her expertise on cross-carrier MVNOs helped me think about how novel MNA arrangements affect consumer behavior and operators' bottom line. Martin drew parallels between our economic analyses and earlier research on peering and efficient component pricing rule; his insights put MNA in a broader context for me.

I certainly got my tuition money's worth out of a school as expensive as CMU when I had two faculty members looking after me whereas most Ph.D. students only have one. Well, technically those were not my money. This research would not be possible without the generous support from the National Science Foundation Awards 1040801, 1343359, 1345305 and Portuguese Foundation for Science and Technology under the Carnegie Mellon | Portugal program (Grant A022078-PSTF-PEHA).

I had amazing company in Pittsburgh. I especially thank my officemates – Baojiang, Brock, Fan Tong, Ty, Vedran, Yu, Alex, Aviva, Francisco, Jeff, Aman, Octavio, Turner, and Joan – who made the time I spent on campus a joy, especially back in the Wean dungeon days. I would also like to thank other EPP students – Matthew, Sara, Sunhee, Travis, Rebecca, Barry, Michael, Daniel, Sonia, Serban, Carol, Nirajan, Paul, Xiaochen, Jiqiao and Guannan – You have all been a positive influence and thank you for the support and encouragement over the years. Similarly, everyone in the EPP front office – Adam, Vicki, Debbie, Patti, Steve, and Kim – whose help and patience throughout my time here are greatly appreciated. My roommates over the years – Cheng, Haisu, Botong, PK and Fan Zhang - have also been a delight, bringing happiness and gourmet dishes. I should also recognize my distinguished list of Chinese lunchbox suppliers, chief among them Sun Penang, Sushi Too and Chengdu Gourmet. Without them I would probably have resorted to {insert fast food franchises here} more often than I should.

Friends outside Pittsburgh, some working on their own Ph.D., have provided tremendous moral support – Yixing, Youtai, Jiali, Liangyu, Mingjun, Lingjie and Yubo. One friend set me up with surplus x86 server parts that made the Monte Carlo simulations underpinning this research

feasible within reasonable time frames. Thanks to Fred, I have been traveling in international premium cabins for a fraction of the cost when visiting home; Fred also offered words of wisdom from time to time.

I want to thank the O'Banions family: Peter, Melissa, Mr. and Mrs. O. The O'Banions are like a surrogate family to me. I still have no idea to what I owe this privilege. It takes a big kind heart to love and care for a complete stranger in the way Mrs. O did for me. They helped me settling into this foreign country; they invited me to Thanksgiving dinners; they took care of my car when I went back home for vacation. I struggled with English when I first landed in the states 10 years ago; Peter was one of the very few people who was always willing to slow down speaking when I needed it so that I would be able to understand. He also did not mind spending extra time to explain things to me, although he more often than not gave Urban Dictionary definitions rather than Merriam-Webster. The O'Banions are just wonderful people to hang out with. There is constant laughter in the family; much of it came after a witty comment or sarcasm and they really go at each other! Those who can take a joke are never bad people. The O'Banions are all good sports for sure.

On a quiet Saturday morning some five years ago, I heard these two guys with a weird accent on the radio talking about cars. Next thing I know they have grown on me. I almost did not want to admit this, given the care-free attitude of these MIT-graduates-turned-into-mechanics (one of them a Ph.D., no less), but when I am anxious I turn to Click and Clack the tappet brothers from Car Talk. For some reason, only a few minutes of hearing them insulting listeners on the radio is enough to calm me down. In all seriousness, it is not difficult to spot the scientific rigor in their advice to callers. They are both genius in talking to complete strangers and finding common

grounds, which is a skill I wish I had. My only issue with them is that they, like so many Pittsburghers, did not seem to understand the benefit of zipper merge. But I digress.

I mostly listen to Car Talk during the long drives to and from New York. Why am I always driving to NY you ask? For a girl, of course. Yun and I have gone from complete strangers to a married couple expecting their first daughter in the last six years, but to this day we are still in a long-distance relationship and always have been. (Not for much longer, though, after I turn this baby in.) What a miracle. She must be the right one, I guess. Yun made many sacrifices in the last six years to make this relationship work, and for that I am forever grateful.

I am blessed with a wonderfully supportive and caring family. My parents lifted the low and shared the high moments during this bumpy ride. Their unconditional love, belief in education and belief in me, make them my role models as I will soon become a father myself. My in-laws have been firmly supportive and thoughtful over the years, too, taking many burdens off me so that I have more time for research.

Abstract

Consumers today are often simultaneously covered by multiple wireless communications networks that vary in throughput, latency, price, coverage, etc. Despite that, today's wireless services and mobile devices are set up to use the same network as long as that network is available, or to only communicate via Wi-Fi as long as a known Wi-Fi network is present. In this thesis, we show that there can be huge benefits to the alternative scheme, multi-network access (MNA), where a mobile device may use the infrastructure owned by any one of multiple network operators, at any point in space and time.

MNA can greatly improve the spectral efficiency of cellular networks. Given uniformly distributed tower location, MNA can increase network capacity by over 50% without additional infrastructure or spectrum; or equivalently, it can reduce infrastructure and spectrum spending each by 20% to produce the same capacity. The benefit of MNA is accessible - it does not require large separation distance between the transmitters belonging to different network operators, nor does it require many non-colocated transmitters. Moreover, colocation does not necessarily negate the benefit of MNA; they can be deployed together to achieve even greater cost efficiency.

The economic impact of MNA on individual network operators is a result of both business decisions and technical parameters. When mobile network operators (MNOs) participate in MNA, the distribution of traffic volumes and the associated revenue among partner MNOs with MNA is not necessarily equal to that without MNA, and is not necessarily commensurate with the distribution of investment on infrastructure and spectrum resources among partner MNOs.

There exists MNA arrangements that make all stakeholders, including consumers, better off; exactly how these benefits are distributed among all stakeholders comes down to negotiations.

MNA-capable users can benefit from higher-layer protocols and features that enable the fast switching between networks without disruption to ongoing sessions and the better utilization of multiple networks. Flow-level migration allows fine-grained load balancing and QoS matching for multihomed devices running heterogeneous applications. Locators based on directed-acyclic-graphs allow quick response to link failure and enhances the resiliency of multihomed hosts and networks. Self-certifying identifiers facilitate authentication of migration signaling messages and helps reduce handoff latency.

MNA already exists today in the form of a multi-operator mobile virtual network operator (MO-MVNO) like Google Fi, and there are signs of growing momentum in other forms such as through dual-SIM phones. The advent of 5G and network slicing thanks to software defined networking (SDN) and network function virtualization (NFV) technologies will make MNA even more relevant in the future. MNA may have a profound impact on our wireless internet services by bringing a fundamental shift in control from service providers to end users. Consumers will no longer be limited by a single service provider in price, coverage, performance and customer service, and can always receive the service that best meets current needs. While practical barriers to MNA remain, the potential economic and social benefits provide strong incentives for its consideration.

Table of Contents

List of Tables

List of Figures and Illustrations

1 Introduction

Nowadays, we are often simultaneously covered by multiple wireless communications networks, all capable of IP datagram transport for accessing the abundant internet services and content. These networks may run on the same communications technologies or different (e.g. LTE, Wi-Fi, etc.), and may vary in throughput, latency, cost, coverage, etc., depending on the user's location and time of day. Despite the availability of many networks, today's wireless services and mobile devices are set up to use the same network as long as that network is available. Cellphones will not roam onto a foreign network unless in a location where the home operator does not have infrastructure at all; most mobile devices are programed to communicate via their Wi-Fi network interface as long as a known or open Wi-Fi access point is present.

There are missed opportunities in the single-network access scheme described above. A user's achievable data rate depends highly on signal-to-interference-plus-noise ratio (SINR), which decreases rapidly as one moves further from the transmitter. Today, an Operator A subscriber who is far from any Operator A transmitter continues to be served by Operator A even when there is an Operator Z transmitter closer by. It would have required fewer spectrum resources to achieve the same data rate for that subscriber or, equivalently, generated higher data rate with the same amount of spectrum resources, if Operator Z served that subscriber. The same goes for an Operator Z subscriber who is closer to an Operator A transmitter than any Operator Z transmitter.

In this thesis, we will show that there can be huge benefits to the alternative scheme, multi-network access (MNA), where a mobile device may use the infrastructure owned by any one of multiple network operators, at any point in space and time. As long as network operators do not colocate all their transmitters, access to multiple networks should, on average, reduce the distance between a user and a transmitter, and thus improve average SINR. Network operators can therefore produce higher capacity with less spectrum and/or fewer transmitters. This thesis will take both of the following perspectives:

- How much can MNA increase the network capacity on existing infrastructure and spectrum resources?
- How much infrastructure and spectrum resources can be saved by adopting MNA to produce the capacity, and how does that affect consumer welfare and the financials of network operators?

The answers to these questions will depend on many factors, including the fraction of MNA-capable UEs, the location diversity of the transmitters belonging to different networks, the algorithm for choosing which network to use, the networks' resource allocation schemes, etc. We will explore all these dimensions and investigate how they affect the benefit of MNA.

While MNA has promising upsides, it also creates unique challenges. Today, much of the wireless communications take place when we are on the move. Given a user's mobility and networks' diverse characteristics, how to keep a user on the best available network all the time without disrupting ongoing sessions? With single-network access, it is often sufficient to solve the mobility problem within a network. For example, mechanisms in LTE allow a user to move from one base station (BS) to another within the same operator's radio access network (RAN) without interrupting ongoing sessions. With MNA, because the quality-of-service (QoS) provided by a network varies with geographic location, it becomes more likely that a user wants to move onto a different network as she moves. When that happens, the internet locator (e.g. IP address in the case of TCP/IP) typically changes. A change in the internet locator could break an ongoing session, which is usually (at least partially) identified by the internet locators of both endpoints. Therefore, a key requirement for effective MNA is a migration protocol that transfers ongoing sessions from using one internet locator to another.

The need to migrate is not necessarily due to handoff, i.e. leaving the coverage area of the previously used transmitter. Migration may also be warranted when a newly available network provides better QoS and/or lower cost for an ongoing session, even if the previous network is still accessible. Using a migration protocol, MNA-capable devices would be able to move different application traffic to different networks on the fly, either to balance load or to match an application to the network that best meets its QoS requirement. The state-of-the-art internet architecture, TCP/IPv6, supports migration poorly. We have designed a new migration protocol in the context of the eXpressive Internet Architecture (XIA), a clean-slate Future Internet Architecture [1] that features expressiveness, intrinsic security and the use of directed acyclic graphs (DAGs) as locators. Along these lines, we seek to answer the following questions:

- How does the XIA Migration Protocol compare with IPv6 migration solutions such as Mobile IPv6 regarding handoff latency and security?
- How does XIA as an architecture compare with IPv6 regarding the ability to make use of multiple networks to improve resiliency, match network characteristics to QoS requirements and to balance load?

This dissertation is the first interdisciplinary study on MNA, covering both technical and engineering economics aspects of MNA. The main contributions are as follows:

- We quantified the spectral efficiency gain from MNA in a wide variety of base station layouts, including scenarios with partial colocation.

- We determined how the benefits of MNA depend on the fraction of MNA-capable users among all users.
- We measured how much MNA can reduce network costs, improve operator profitability, and lower end user prices.
- We demonstrated that there can be synergy between colocation and MNA that further improves cost-efficiency.
- We showed how different types of network selection algorithms used by MNA-capable devices affect the benefit of MNA.
- We discovered the potential discrepancy in the quality-of-service for subscribers of different operators in an MNA arrangement and discussed the implications thereof on the incentives and disincentives to adopt MNA.
- We evaluated migration and multihoming support in XIA and compared them with IPv6; we identified key protocol design choices and architectural features that facilitate mobility and utilization of multiple access networks.

There are signs that MNA is slowly gaining momentum. As radical as it may seem, MNA already exists in limited capacity today in the form of Google Fi, a multi-operator mobile virtual network operator (MO-MVNO) that partners with multiple MNOs (T-Mobile, Sprint and US Cellular in the US) and chooses among them for “the network that our analysis shows will give you the best Fi experience at your current location” [2]. However, research showed that the current Google Fi implementation does not effectively switch network when performance is poor [3]. MNA may be more relevant as 5G rolls out, which promises two pertinent features. One is lower overhead of maintaining a connection to the network, which may facilitate having simultaneous attachment to multiple cellular networks and in turn faster switching between different cellular networks. Network slicing is also reported to be part of 5G, which should make it easier for one MNO to rent capacity from another MNO, and for a potential MO-MVNO to launch its service.

Other industry players have been slowly gearing towards network architectures that is ready for MNA but stop short of the kind of MNA envisioned in this thesis that would realize the full potential of a pool of network infrastructure and spectrum resources belonging to different operators. For example, two of the three largest cable TV and internet providers in the US have begun offering wireless internet services using a combination of their own infrastructure and those of a cellular MNO. XFINITY Mobile by Comcast taps into the 19 million XFINITY Wi-Fi hotspots (partially fueled by home routers turned into a public access point) in addition to an MVNO hosted by Verizon. Spectrum Mobile by Charter Communications employs a similar business model, while also considering deploying its own cellular infrastructure using equipment built to work on cable plants in the CBRS spectrum band [4]. For Comcast and Charter, it is likely that the additional Wi-Fi hotspots and cellular infrastructure are used to lower leasing charges paid to their host MNOs instead of increasing spectral efficiency. The fourth largest cable provider in the US, Altice, has also launched its Altice Mobile MVNO hosted by Sprint, with AT&T as a roaming partner [5]. In this case, it is likely that AT&T is only used to complement Sprint's coverage gaps.

MNA may have a profound impact on our wireless internet services. MNA can bring a fundamental shift in control from service providers to end users. One common thread in the MNA and semi-MNA examples described above is that in these arrangements, it is the service provider (the MVNO in these cases) that has the control over which networks are used and when. However, we could be on the brink of a fundamental shift in control from service providers to end users. Comcast, Charter and Google Fi are all reportedly testing dual-SIM technology, presumably to facilitate access to and the switch between the different networks operating on separate infrastructure [4], [6], [7]. Dual-SIM phones were exactly the vision we had in our first work on MNA [8], and they could put the control back in the hands of end users. A subscriber with a single service provider is limited by that service provider in price, coverage, performance and customer service. If the end user could freely choose its MNO at any point in space and time with little switching cost, the end user would always receive the service that best meets current needs, even though needs may change as the end user changes location, or switches from web browsing to YouTube. This will also increase pressure on MNOs to compete, perhaps driving down prices.

Despite the practical limitations that may exist in current implementations, this thesis lays out what is possible, and presents a new vision for how consumers use wireless internet services.

This thesis is structured as follows. Chapter 2 reviews related works on MNA. Chapter 3 looks at how MNA can increase the capacity of cellular networks without additional infrastructure or spectrum, and how that effect varies with base station offset, fraction of MNA-capable UE, resource allocation schemes and network selection algorithms. Chapter 4 examines how the cost-effectiveness of MNA in the presence of colocation. Chapter 5 investigates how MNA can change the economic viability of network operators and the welfare of consumers and discusses incentives and disincentives for network operators to adopt MNA. Chapter 6 introduces migration and multihoming support in the eXpressive Internet Architecture (XIA) and compares them against TCP/IPv6 solutions. We conclude in Chapter 7 and discuss future research direction in Chapter 8.

2 Related Works on Multi-Network Access

In the context of MNA, network capacity and user data rate depend on many factors, including, but not limited to, the fraction of MNA-capable users, the spatial diversity of the cell towers (e.g. the extent of colocation), resource allocation schemes and network selection algorithms. Previous works on MNA have explored a small subset of these dimensions, which will be discussed in the following subsections, but generally lack the following perspectives:

- Previous studies on the performance implications of MNA have assumed that all users are MNA-capable [9]–[16], while the most relevant form of MNA in the real world today remains a small fraction of all user base.
- Previous studies on the performance implications of MNA either considered only the case of complete colocation [11]–[13], [16] or did not consider colocation at all, while in reality the extent of colocation often falls somewhere between these two extremes.
- Previous studies have not examined the potential infrastructure and/or spectrum resource savings as a result of MNA involving two facility-based cellular service providers. The implications on operator profitability and market prices from MNA's higher efficiency is not well understood.

Related works on the efficiency improvements from MNA are reviewed in Section 2.1. Section 2.2 discusses related works concerned with improving a UE's network selection algorithm. Most are designed in the context of Heterogeneous Networks (HetNets), improving the logic of switching between cellular and Wi-Fi. Only a few studies are conducted for UEs that can freely access multiple cellular networks. Our work considers different selection algorithms and highlight their efficiency implications in the context of MNA, although we do not attempt to optimize the algorithm for a specific objective.

Few works have investigated the cost savings as a result of MNA and how the saving is related to parameters like network selection algorithms and operators' market share, although related issues have been explored. Some examined the effect of MNA on capacity without cost analyses [9]–[12], [15], [16] as will be discussed in Section 2.1. Others addressed the dynamics between an MVNO and MNOs either in a traditional context without MNA [17], or in the MNA context but without considering the efficiency improvements from MNA [18]. These will be discussed in Section 2.3.

2.1 Efficiency Improvements from Multi-Network Access

The possibility of first responders roaming to multiple MNOs was discussed in [19], [20]. The potential for greater total capacity when UEs connect to the BS with the strongest signal regardless of its provider was discussed in [21]–[23]. These works did not perform quantitative analyses on the topic of MNA.

Several studies [9], [14], [15] on the performance implications of MNA use the stochastic geometry model proposed by [24], and they reported a consistent 40%-50% gain in spectral efficiency given no colocation when all users are MNA-capable and attach to the nearest tower. This figure is also consistent with our result in a similar scenario. These studies did not explore alternative network selection algorithms as we did in this thesis, such as algorithms that consider not only distance but also SINR and expected data rate. All users were assumed to be capable of MNA. What we call MNA was referred to as "infrastructure sharing" or "flexible roaming" in these works.

Previous studies on the performance implications of MNA have mostly assumed a single resource allocation scheme. The exception is [16], which considered what we call "Equal-Allocation" and "Equal-Throughput" resource allocation schemes. Reference [16] examined

MNA in the form of subscribers having cell phone plans from multiple MNOs and simulated an urban hot zone (e.g. a train station). Each user reevaluates available BS's every one second and attaches to the one that provides higher expected data rate. Due to drastically different base station topology, their results are not easily comparable with ours. What we call MNA was referred to as "end user network switching".

Reference [12], [13] investigated the performance improvement from MNA in two particular cell layouts as part of a comparison between MNA, spectrum sharing and sharing on virtualized infrastructure. The reported performance gain of less than 10% was much smaller than our results, likely because the author stipulated that only lightly loaded cells can engage in MNA. The author did not consider alternate network selection algorithms – UEs were assumed to attach to the BS that gives the highest signal power. All users were assumed to be capable of MNA. What we call MNA was referred to as "capacity sharing".

References [10], [11] quantified the performance improvement from MNA as a function of the offset in tower locations and sector orientation using a hexagonal grid model. Most notably, MNA was found useful even with tower colocation because the authors argue that sectors are rarely aligned on colocated cell towers. The authors assumed all UEs were MNA-capable, and did not consider alternate resource allocation schemes or scenarios where MNOs have different tower densities. What we call MNA was referred to as "user swapping" and "smart roaming".

This thesis is the first to consider the relationship between the fraction of MNA-capable users and the benefit of MNA, and the first to examine scenarios where MNOs colocate some of their base stations but not others.

2.2 Network Selection Algorithms

The question of how to choose among multiple BS's arises in the context of the radio access technology (RAT) selection problem in heterogeneous networks and is treated in [25]–[31]. The objectives of the proposed algorithms vary, e.g. to maximize throughput, fairness or some measure of utility. The typical scenario of RAT selection problems is a UE attempting to optimize selection of a network interface, usually between Wi-Fi and cellular. This thesis applies to the selection of cellular carriers regardless of whether they use the same RAT. Only a few works are available in this context. Reference [32] identified the deficiencies in the built-in network selection algorithm of Google Fi, and designed solutions that improve throughput and switching latency. Our work builds on this research to explore a range of selection algorithms, and how they affect the cost efficiency of cellular networks that support MNA.

2.3 Interactions between MNOs and MVNOs

Reference [18] discussed the partnering strategy and optimal pricing in a market with two MNOs and one MVNO. Using a Stackelberg game theoretic model, the authors derived profit-maximizing wholesale and retail pricing, and characterized the conditions for which both MNOs choosing to partner with the MVNO is the unique Nash equilibrium. The potential spectral efficiency gain as a result of partnering with more than one MNOs were not explored. The additional profits to be shared among operators were the result of the MVNO's assumed ability to offload traffic to free Wi-Fi hotspots and to earn side revenue (e.g. ads) from its customers. Our work quantifies the potential efficiency gain from such an arrangement and how to divide up the cost savings in a way that benefit all stakeholders.

Reference [33], [34] explored the pricing strategy and profitability of an MO-MVNO. The MO-MVNO was assumed to offer a pricing structure that attracted low usage consumers and had access to free Wi-Fi hotspots whereas the MNOs did not. It was concluded that an MO-MVNO was only viable in the short term, as consumers' monthly data consumption would increase in the long run. While [33] showed improvements in UE data rate, the effect of MNA was not isolated from the effect of additional Wi-Fi hotspots. The implications on cost were not considered. Neither were the effects of the resource allocation scheme and network selection algorithm.

3 MNA on Cellular Network Capacity

3.1 Background

The increased spectral efficiency from MNA can be put to use in two ways.

1. Network operators that adopt MNA can get immediate increase in network capacity from existing infrastructure and spectrum resources.
2. Network operators dimensioning a new network to meet a given capacity target can use less infrastructure and/or spectrum resources to do so.

In Chapter 3, we take the first perspective and seek to quantify how much network capacity gain can be achieved by adopting MNA in a variety of scenarios. Chapters 4 and 5 will take the second perspective and seek to quantify the potential infrastructure and spectrum resource savings as a result of MNA and the values thereof.

MNA may take many forms. One example is dual-SIM[1] phones. A more recent example is a multi-operator mobile virtual network operator (MO-MVNO) [35], such as Google Fi, which leases capacity from two or more facility-based MNOs and intelligently assigns each subscriber to the operator that currently offers the better performance. Going forward, software defined networking (SDN) and network functions virtualization (NFV) in 5G will simplify the leasing of a slice in another operator's network [36] and perhaps make MNA more prevalent. One could imagine a capacity sharing arrangement between two or more MNOs without an intermediary MO-MVNO, in which all parties agree to let each other's subscribers use the network, possibly

[1] Dual-SIM phones have traditionally been used to facilitate international travel, personal phone-work phone separation, etc., and are probably not designed for the intelligent MNA envisioned in this thesis. Network selection is often performed manually by the end user. Nevertheless, if equipped with appropriate software, dual-SIM phones can be a great platform for MNA.

in exchange for a fee. Throughout Chapters 3 to 5, we will focus on the MO-MVNO form of MNA.

The benefits of MNA, whether they are additional capacity or resource savings, depend on many factors. Throughout Chapters 3 to 5, we will examine the following dimensions to the problem.

3.1.1 How do the benefits of MNA vary with the fraction of MNA-capable UEs?

In the context of an MO-MVNO, the market share of the MO-MVNO(s) represents the fraction of UEs that can use the infrastructure belonging to different MNOs, i.e. "MNA-capable". Intuitively, we expect that the more MNA-capable UEs, the more spectrum resources will be used efficiently and the greater the benefit of MNA will be. Our goal is to define the relationship more precisely.

3.1.2 How do the benefits of MNA vary with resource allocation scheme(s)?

The resource allocation scheme refers to the method by which an MNO BS decides which resource blocks are to be assigned to which UE in each time interval. Resource allocation must trade off between throughput fairness and total throughput. Because MO-MVNO subscribers are on average closer to a BS than MNO subscribers, MO-MVNO subscribers need fewer resource blocks to realize the same data rate. The resource allocation scheme determines how those resource savings are distributed among UEs. A fairness-oriented scheme will distribute saved resources differently from a throughput-oriented scheme.

3.1.3 How do the benefits of MNA vary with an MO-MVNO's MNO selection algorithm?

The MNO selection algorithm refers to the process by which an MO-MVNO UE chooses which BS to attach to at any point in space and time. Globally, the most efficient use of spectrum resources occurs when UEs connect to the BS that offers the highest SINR (typically the closest), but that is not necessarily the best strategy for an individual UE. A UE's nearest BS may offer lower throughput than a BS that is farther away but less congested, in which case the UE might attach to the farther BS.

3.1.4 Who benefits more from MNA, MNOs or MO-MVNOs?

How are the efficiency gains of MNA distributed between MNO and MO-MVNO customers? If an MNO treat all UEs in the same fashion, MO-MVNO UEs will realize higher average data rates than MNO UEs, because they are, on average, closer to the serving BS. We will consider how the MNO may adjust its resource allocation scheme so that both groups of customers realize the same average data rate.

3.1.5 How do the benefits of MNA vary with the spatial diversity of base stations belonging to different MNOs?

The benefit of MNA hinges on the participating MNOs not colocating all BS's, so that UEs have a larger set of distinctively located BS's to choose from in the same geographic area. Exactly how much MNA can increase the capacity of cellular networks depends on how far apart the BS's of different MNOs are from each other. This question is of great interest to an MNO trying to decide whether to adopt MNA. Tower layouts vary widely from city to city and from operator to operator [37]. Depending on how its infrastructure is complemented by that of the potential partner MNO, the decision may be different in different cities, or even different parts of the same city, e.g. MNO A may choose to partner with MNO B in one city but MNO C in another.

This question will be treated with two approaches. In Chapter 3, we will use a hexagonal grid model for BS layout, vary the offset between the BS's of participating MNOs and observe how it affects network capacity. In Chapter 4, we will use a model of uniformly distributed BS's, vary the extent of tower colocation and observe how it affects the network capacity and potential resource savings from MNA.

3.1.6 What happens when two partner MNOs are unequal in size?

Facility-based MNOs have varying market share, and they invest differently in spectrum and BS densities in order to serve their share of the market. For a given pair of resource allocation scheme and MNO selection algorithm, the available infrastructure and spectrum resources at each MNO can affect which and how many MO-MVNO subscribers are served by each MNO, how much spectrum resource is devoted to MO-MVNO subscribers by each MNO, and how much of the overall capacity is sold to the MO-MVNO by each MNO. These metrics, in turn, can change how much each MNO is paid by the MO-MVNO(s), depending on the nature of the business agreements between them. We will investigate whether an MNO always get returns commensurable with the spectrum and BS resources it devotes to MO-MVNO subscribers.

The rest of the chapter is structured as follows. We give an overview of the research methods and assumptions shared by chapters 3 through 5 in Section 3.2. We detail the simulation model used for Chapter 3 in Section 3.3. Results are presented in Section 3.4 and we conclude this chapter in Section 3.5.

3.2 Overview of Methods and Simulation Models

In this subsection, we describe the method and key assumptions that apply to Chapter 3 through Chapter 5.

3.2.1 Methods and Key Assumptions

We constructed simulation models in MATLAB with a preset distribution of BS's and UEs, and computed the downlink network capacity when MNOs serve only their own subscribers (i.e. no

MNA) and when some or all UEs can use any MNO's BS's (i.e. MNA-capable). We will examine the case of one MO-MVNO that partners with two MO-MVNOs.

Stationary UEs are uniformly distributed throughout the area. We assume that UEs have an infinite amount of data to receive, i.e. a full-buffer traffic model. The simulated area is wrapped around on all four edges (effectively a torus) to account for edge effects.

We assume that the downlink data rate approaches the Shannon limit. The data rate r_i in bits per second for UE i at distance d from the BS is given by:

$$r_i = s_i B \, log_2\left(1 + \frac{P_r^i}{I^i + P_n^i}\right) \tag{1}$$

Here, s_i is the share of spectrum resource blocks assigned to UE i; B is total available bandwidth of each BS in Hz; P_r^i is the received signal power across $s_i B$ at UE i; I^i is the sum of interference power across $s_i B$ from all other BS's, where we have assumed a frequency reuse factor of 1. P_n^i is the noise power across $s_i B$, where we have assumed a fixed noise power spectral density. We assume BS's transmit at a fixed power level.

3.2.2 MNO's Resource Allocation Schemes

Each UE is apportioned a share s_i of resource blocks, governed by the MNO's resource allocation scheme, which is discussed in the next section. We constructed an abstract model of resource allocation schemes that allows for varying the tradeoff between throughput fairness and total throughput. A UE's share of spectrum resources, s_i is given by:

$$s_i = \beta[log(1 + SINR_i)]^\gamma \tag{2}$$

Here, β is a constant and γ is a parameter that controls the tradeoff between throughput fairness and total throughput. The larger γ, the more high-SINR UEs are favored (though cell edge UEs may be starved), which increases total capacity. The smaller γ, the more low-SINR UEs are favored, which reduces the difference in data rate between UEs, thereby increasing throughput fairness. We name three special cases:

- "Equal-Allocation": $\gamma = 0$. Every UE receives an equal fraction of resources. As a result, each UE's throughput can be different due to their varying distances to the BS, though total throughput is higher.
- "Equal-Throughput": $\gamma = -1$. All UEs in a cell get the same data rate. Each UE's share of spectrum resources is inversely proportional to its spectral efficiency at its location.
- "Balanced": $\gamma = -0.5$. This compromise scheme most closely resembles real world MNO practices.

Our resource allocation scheme also considers fairness in the QoS experienced by those capable of MNA and by those incapable. The model based on (2) allocates the same amount of resources to all UEs with the same SINR. However, as shown in Figure 1 as an example, MNO UEs (or those incapable of MNA in general) are on average farther from a BS than MO-MVNO UEs (or those capable of MNA in general), and thus are likely to have lower SINR than MO-MVNO UEs. Consequently, MNO UEs on average experience lower data rates than MO-MVNO UEs, except when the "Equal-Throughput" scheme is used. Such performance disparity may put the MNO at a disadvantage, encourage its subscribers to switch to the MO-MVNO, and potentially discourage MNOs from partnering with an MO-MVNO.

To remedy the situation, MNOs could seek monetary compensation from the MO-MVNO for delivering higher throughput. Alternatively, MNOs could modify their resource allocation scheme to give its own subscribers proportionally more resources in order to restore throughput parity, so that average data rate for all MNO UEs equals average data rate for all MO-MVNO

UEs. In this work, we adopt the latter solution and introduce another dimension in the resource allocation scheme.

- "Impartial": A BS allocates the same amount of resources to all UEs with the same SINR based on (2).
- "Rate-Adjusted": A BS adjusts resource allocation constant β to equalize the mean data rate among UEs of different operators, given the different distributions of distance and thus SINR.

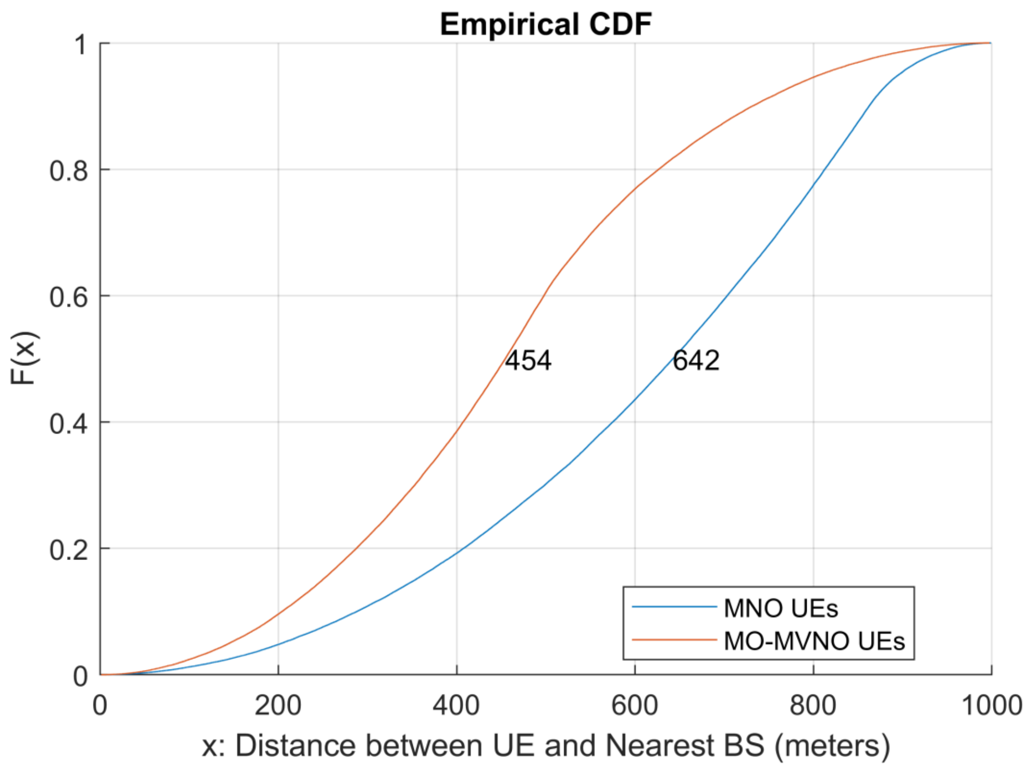

Figure 1 Cumulative Distribution Function of distances between UEs and their nearest BS. There is one MO-MVNO that partners with two MNOs. Each MNO has its BS's laid out on a hexagonal grid. The two grids are offset by one cell radius. The 50th percentile distances are marked. MNO UEs are on average farther from a BS than MO-MVNO UEs, and thus are likely to have lower SINR than MO-MVNO UEs.

3.2.3 MO-MVNO's MNO Selection Algorithm

We investigate two algorithms an MO-MVNO UE might use to select an MNO.

- "Maximum-SINR": An MO-MVNO UE will attach to (the MNO of) the BS that provides the highest SINR.
- "Maximum-Throughput": An MO-MVNO UE will first find BS with the highest-SINR belonging to each MNO, and then attach to the one that provides higher expected throughput.

We assume that the UE knows precisely the data rate it would receive, as if it were able to test and observe the performance on each MNO BS.

For the "Maximum-Throughput" MNO selection algorithm, we assume that UEs select their MNO one at a time, and that they cannot change their selection afterwards. However, the order in which UEs make the selection can change the MNO selected by each UE, and subsequently the data rate. To reduce the effect of selection order on the final data rates, we randomize the order in which MO-MVNO UEs select MNOs, and take the average data rate resulting from each selection order as the final data rate. Figure 2 is a flowchart of this process.

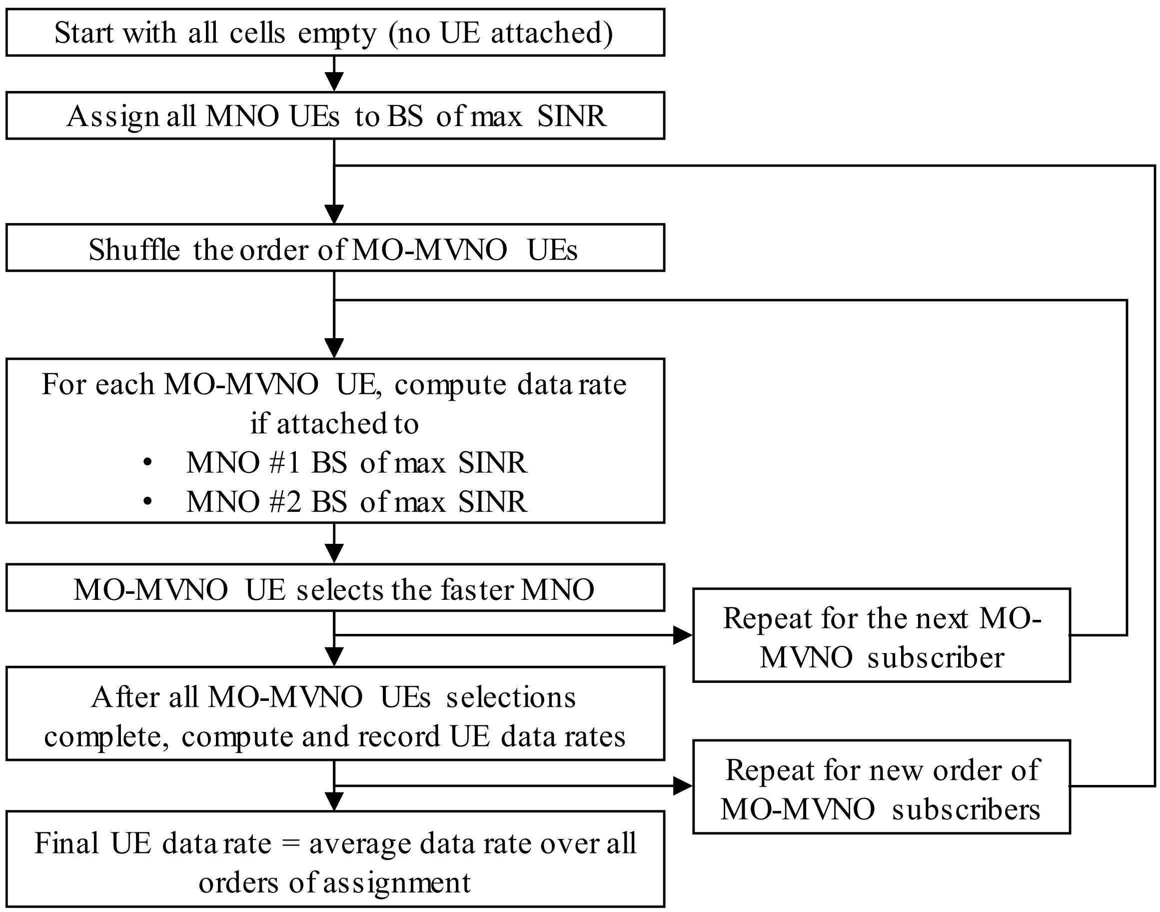

Figure 2 UE-cell association flowchart for Maximum-Throughput MNO selection algorithm.

3.3 Simulation Model (Chapter 3)

This subsection describes the simulation model used for Chapter 3 analyses.

3.3.1 Overview

The received power at reference distance d_0 from the BS is P_t. The received power P_r at distance d is calculated by (3), where α is the path loss exponent. Our model does not consider fading.

$$P_r(d) = P_t \left(\frac{d}{d_0}\right)^{-\alpha} \tag{3}$$

Each MNO has 48 BS's placed on a hexagonal grid in an approximately 10 km by 12 km area that wraps around at the edges. Distance from the BS to cell edge (a hexagon vertex) is 1 km.

Each BS has $B = 10$ MHz of downlink spectrum available. Received power P_t at $d_0 = 1$ meter is 10 Watts. The path loss exponent choice of $\alpha = 3.5$ accounts for both the effects of clutter and long-term fading; short-term fading is not likely to significantly affect the long-term average data rates we present here. Noise power over 10 MHz $P_n = 10^{-13}$ Watts, or -170 dBm/Hz.

We assume there are 4000 customers in total, and each customer subscribes to one of the two MNOs or the MO-MVNO. In the baseline scenario, there is no MO-MVNO subscriber, and each MNO has 50% of the customers. We then vary the fraction of MNA-capable UEs, i.e. MO-MVNO's market share, from 0% to 100% percent, with the remaining customers split evenly between the two MNOs.

In Chapter 3, the maximum-SINR selection algorithm is equivalent to connecting to the nearest BS, because all BS's transmit at the same power level, noise power is constant and we do not consider fading.

3.3.2 Base Station Offset

We add an adjustable offset between the BS's of the two MNOs. Starting with just two BS's on top of each other, an offset is created by shifting the second BS away from the fixed, first BS by an "offset distance" ρ in the direction defined by the "offset angle" θ as shown in Figure 3.

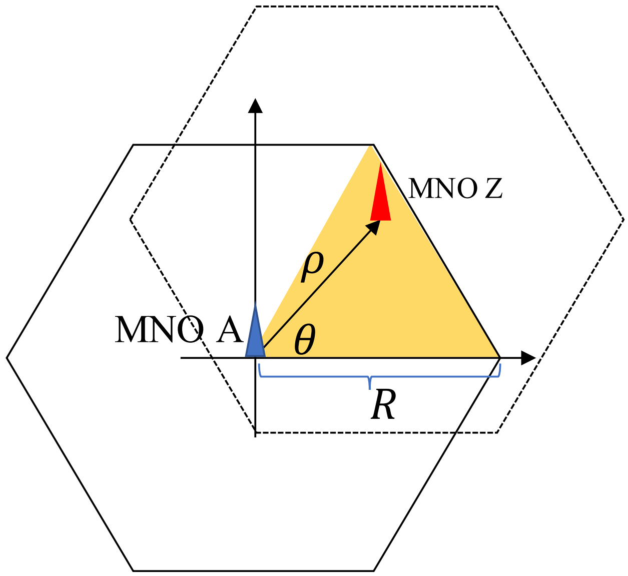

Figure 3 Definition of base station offset.

Examples of the resultant BS map are shown in Figure 4.

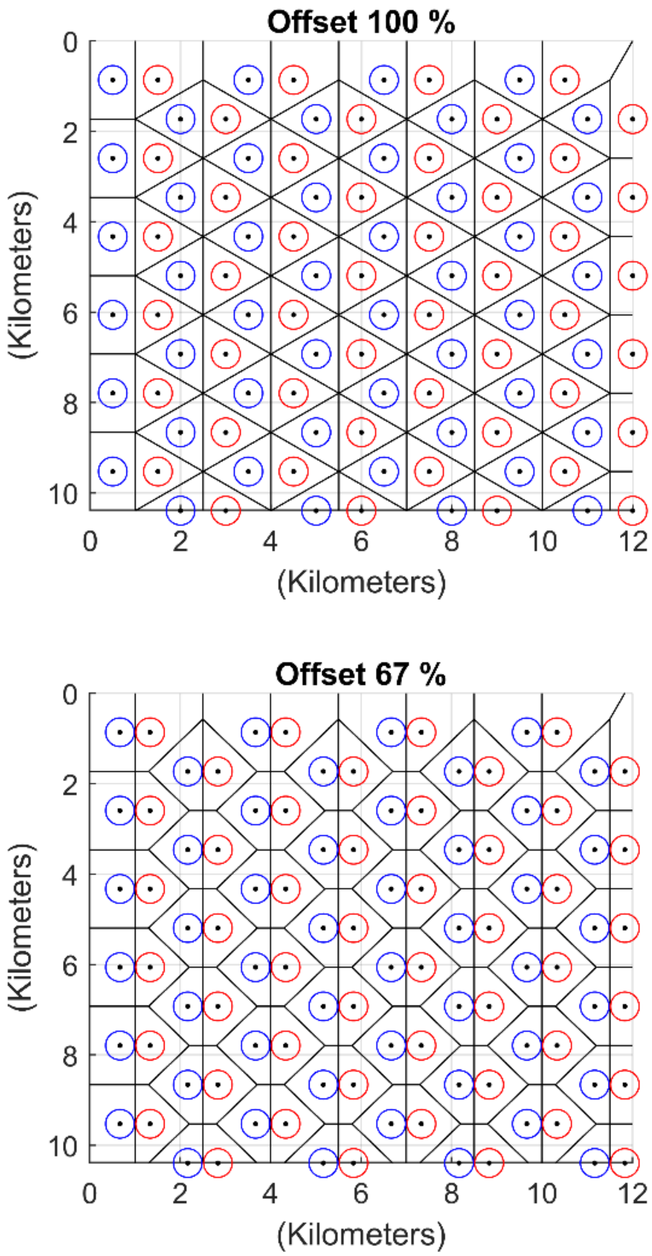

Figure 4 Base station offset with full and 2/3 of maximum offset distance at 0° offset angle, respectively.

In the real world, the distance between the BS's belonging to different operators varies greatly and cannot be fully captured by a single value. To get a realistic estimate of MNA's impact on network capacity under varying BS offset, we will perform a Monte Carlo simulation with randomized BS offset and observe the distribution of the resulting network capacity and UE data rate. A random offset is generated as follows. Start with one MNO (MNO A) whose BS's are located on a hexagonal grid. Randomly pick a location within each cell of MNO A, and drop an MNO B BS at that location. For each random draw, we apply the same offset in all cells, so that MNO B BS's also end up on a hexagonal grid. In other words, we create an offset between two hexagonal grids that are aligned and are of the same size. Each random offset can be represented by an offset angle and offset distance as defined in Figure 3.

Given the symmetry in a hexagon, we only need to consider θ in the range of $[0,60°)$ as indicated by the shaded region in Figure 4. Furthermore, we only need to consider offsets that make the MNO B BS fall within the cell of the MNO A BS. The maximum offset distance ρ is a function of offset angle θ and is given by the following when $\theta \in [0,60°)$:

$$\rho = \frac{R}{\cos\theta + \frac{\sin\theta}{\sqrt{3}}} \tag{4}$$

3.4 Results

3.4.1 MNA's Effects on Total Capacity

MNA can greatly increase network capacity regardless of which resource allocation scheme and which MNO selection algorithm are used. Figure 5 plots the change in total network throughput of one MNO for various resource allocation schemes and MNO selection algorithms. When the MO-MVNO has 100% market share, each MNO can carry as much as 78% more traffic than in the base case where no UE is capable of MNA. The MNO's network capacity monotonically increases as the MO-MVNO's market share grows, and it increases faster than linearly. The convex shape is because of the rate-adjustment discussed in 3.2.2. As more and more users become MNA-capable, those who remain incapable of MNA will be assigned increasingly disproportionately more resources so that they still achieve the same data rate on average as MNA-capable users do. Therefore, the marginal benefit of converting one more MNA-incapable user increases as fewer and fewer of them remain.

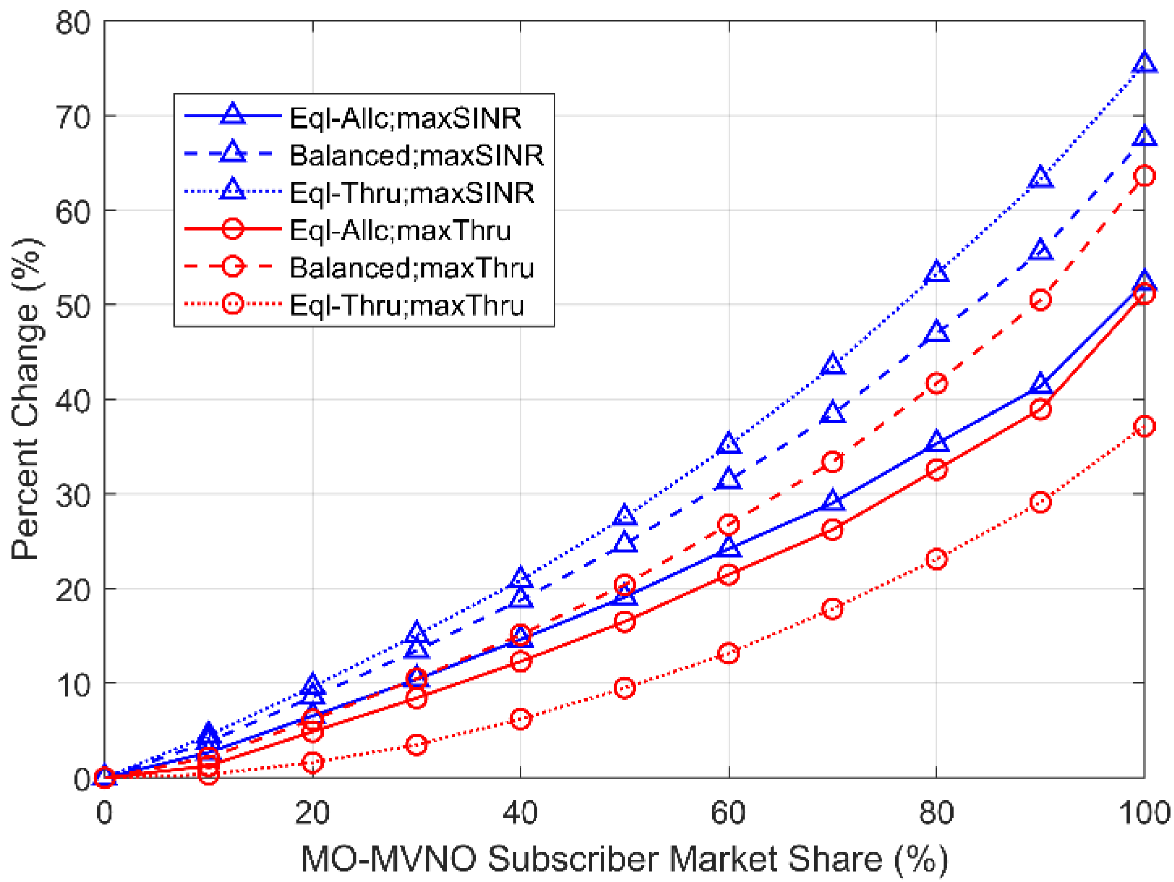

Figure 5 Percent change in MNO network throughput vs. MO-MVNO market share. MNA can greatly increase network capacity regardless of which resource allocation scheme and which MNO selection algorithm are used.

3.4.2 Effects of MNO's Resource Allocation Schemes

The effects of resource allocation scheme choices are modulated by the MNO selection algorithm being used. When the MO-MVNO's selection algorithm maximizes SINR, resource allocation schemes that favor fairness over total throughput experience higher capacity gains (Figure 5). This is expected, because the MO-MVNO removes from an MNO's network UEs that are spectrally least efficient, who thus consume extensive resources when MNOs strive for throughput fairness. That said, even though the capacity improvement is smaller with the "Equal-

Allocation" scheme than "Equal-Throughput", MNA still achieves a healthy gain of as much as 50% over the base case.

Under the "Maximum-Throughput" MNO selection algorithm, the relative positions of the three resource allocation schemes are altered. The "Balanced" scheme dominates, and "Equal-Throughput" now experiences the least relative benefit from MNA.

Note that, while networks that adopt the "Equal-Allocation" scheme tend to see a smaller relative capacity gain from MNA, they still lead in absolute network capacity, as shown in Figure 6 , because "Equal-Allocation" by design devotes more resource blocks than the other two schemes to UEs with better channel conditions.

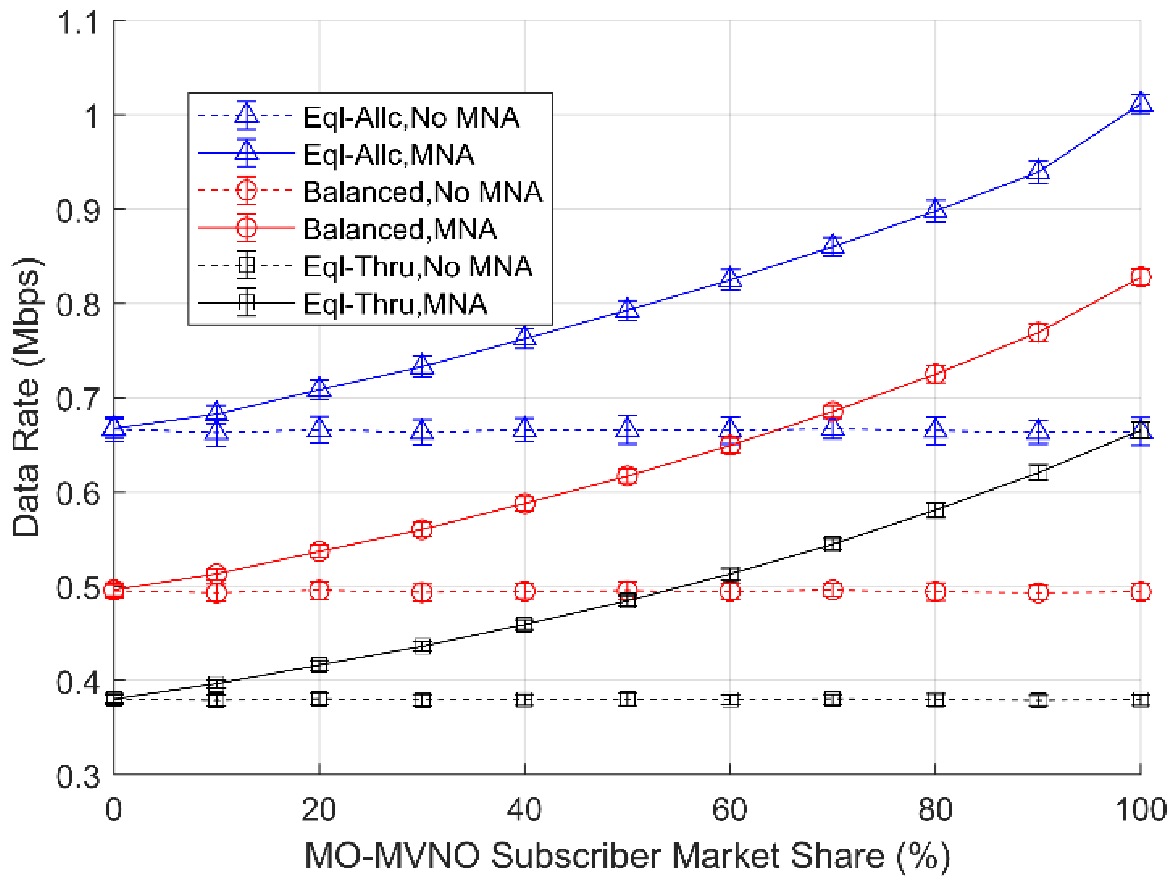

Figure 6 Average UE data rate with various resource allocation schemes. While networks that adopt the "Equal-Allocation" scheme tend to see a smaller relative capacity gain from MNA, they still lead in absolute network capacity.

3.4.3 Disproportionate Gains by Operator

Next we demonstrate how MO-MVNO customers may capture the lion's share of the benefits from MNA and why MNOs may want to preferentially apportion radio resources to their own subscribers to reduce this effect.

Figure 7 plots the change in average subscriber data rate over the base case as a result of MNA when the MNOs do not distinguish their own subscribers from MO-MVNO subscribers. Because MO-MVNO subscribers are on average closer to a BS and enjoy higher spectral efficiency, they stand to receive a significantly greater performance boost than MNO subscribers, unless the MNOs employ the "Equal-Throughput" resource allocation scheme (not shown) which enforces uniform UE data rate in each cell. In particular, under "Equal-Allocation", MNO subscribers see no data rate increase while MO-MVNO subscribers claim all the benefits.

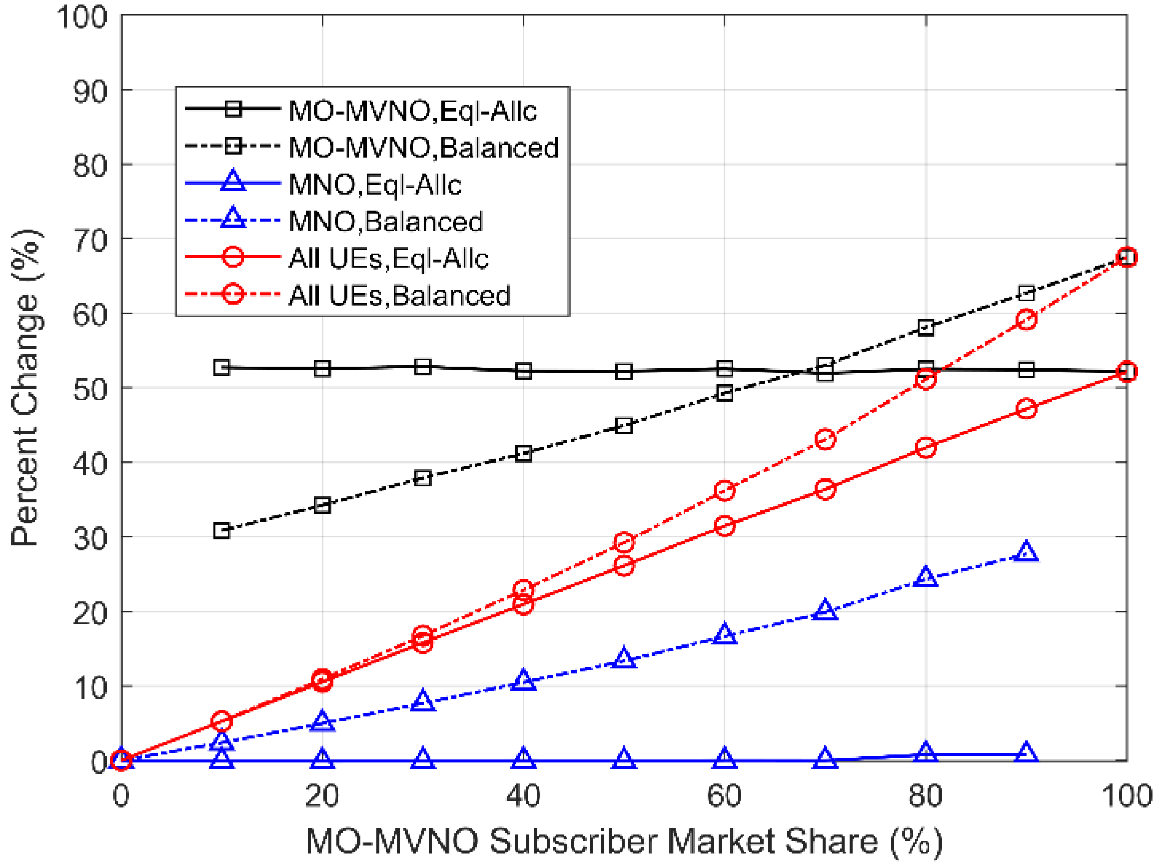

Figure 7 Change in average subscriber data rate by operator ("Impartial" resource allocation, Maximum-SINR MNO selection). Because MO-MVNO subscribers are on average closer to a BS and enjoy higher spectral efficiency, they stand to receive a significantly greater performance boost than MNO subscribers, unless the MNOs employ the "Equal-Throughput" resource allocation scheme (not shown) which enforces uniform UE data rate in each cell.

Figure 8 plots the average subscriber data rates resulting from the "Impartial" and "Rate-Adjusted" resource allocation schemes as discussed in 3.2.2. If the schedulers do not take a UE's subscribed operator into account (solid curves), the MO-MVNO subscribers would receive about 25% higher data rate on average than an MNO's own subscribers. While MNOs may find such disparity troublesome, their subscribers nonetheless see higher data rates than in the base case. In other words, MNA always benefit all operators, although not necessarily equally.

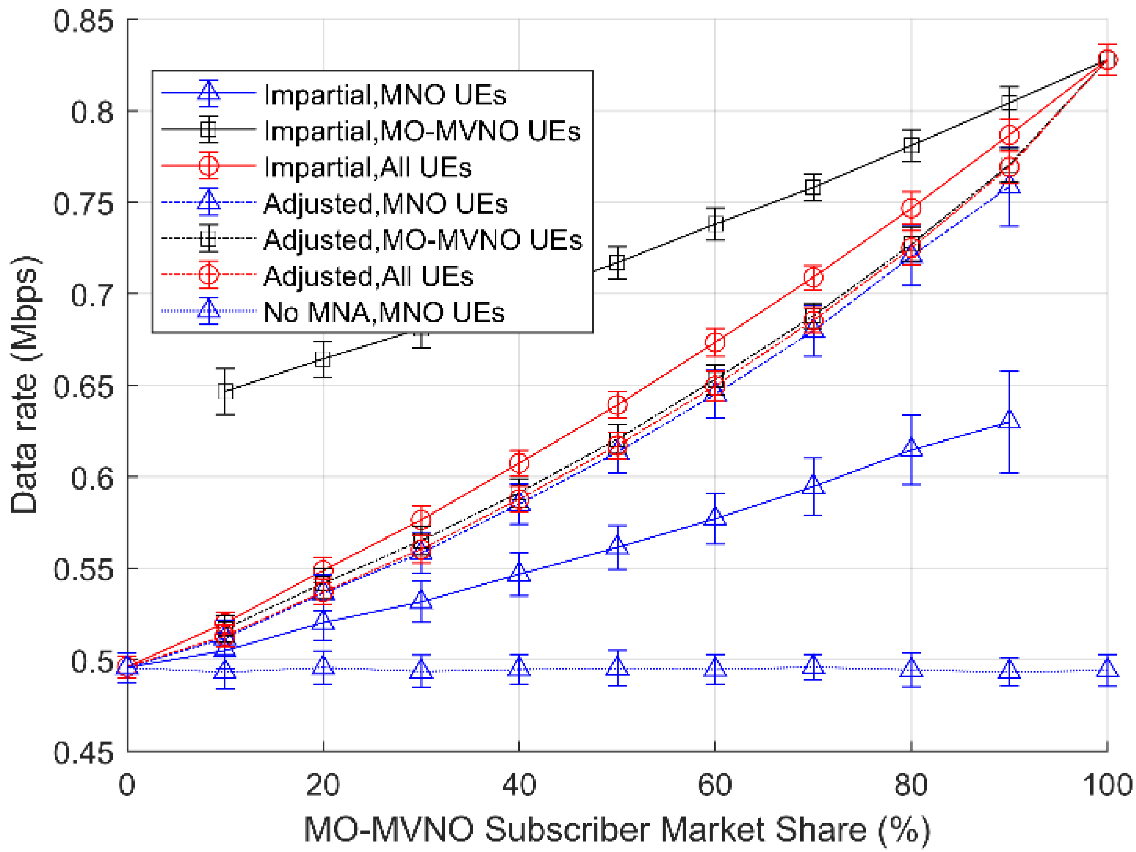

Figure 8 Average subscriber data rate by operator vs. MO-MVNO market share (Maximum-SINR MNO selection; "Balanced" resource allocation). If the schedulers do not take a UE's subscribed operator into account (solid curves), the MO-MVNO subscribers would receive about 25% higher data rate on average than an MNO's own subscribers.

MNOs can adjust their schedulers to provide their own subscribers proportionally more resources to avoid a potential data rate disadvantage (dash-dot curves). However, such an adjustment would incur a slight capacity penalty as it shifts resources to UEs with lower SINR.

3.4.4 Effects of MNO Selection Algorithm

Figure 9 shows that, perhaps surprisingly, the "Maximum-Throughput" selection algorithm consistently yields smaller data rate improvement for any operator than "Maximum-SINR". This is most evident when MNOs enforce throughput fairness, and holds true for all other resource allocation schemes tested as well. By behaving rationally as an individual UE by choosing the operator that provides the higher data rate, an MO-MVNO subscriber actually hurts the other MO-MVNO subscribers. Always choosing a BS with the highest SINR makes the most efficient use of spectrum. Thus, while an individual MO-MVNO subscriber may be better off forgoing the nearest BS due to local variance in congestion, doing so sacrifices overall system efficiency.

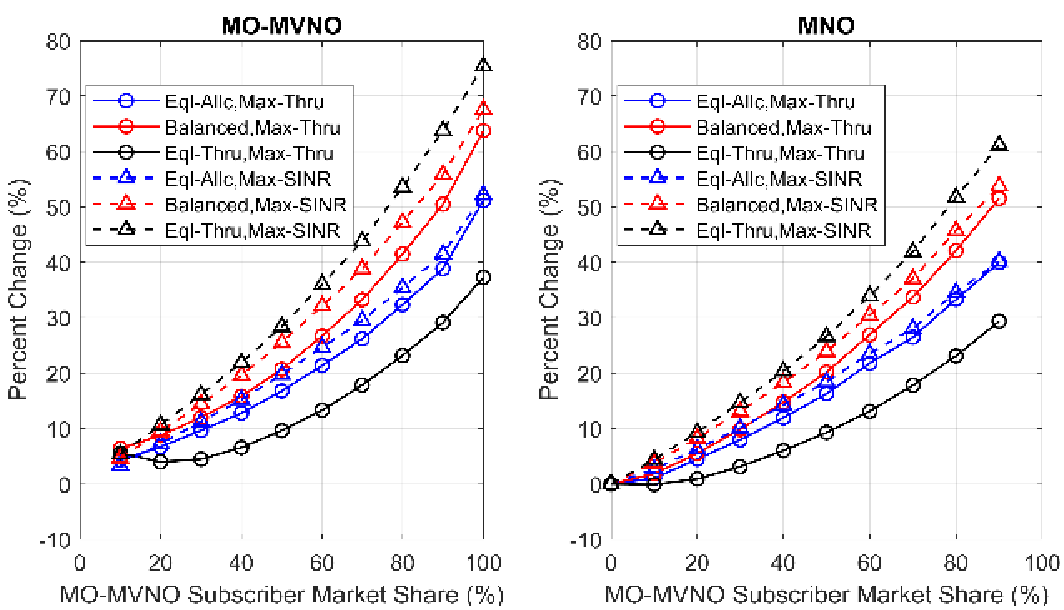

Figure 9 Relative change in average subscriber data rate by operator. The "Maximum-Throughput" selection algorithm consistently yields smaller data rate improvement for any operator than "Maximum-SINR".

While "Max-SINR" is more efficient than "Max-Throughput", it comes at the expense of data rate fairness. Figure 10 shows the probability density function of the data rate of all users in the network. “Max-SINR” causes higher variance in data rate among users; there are more users getting very high data rates and more users getting very low data rates.

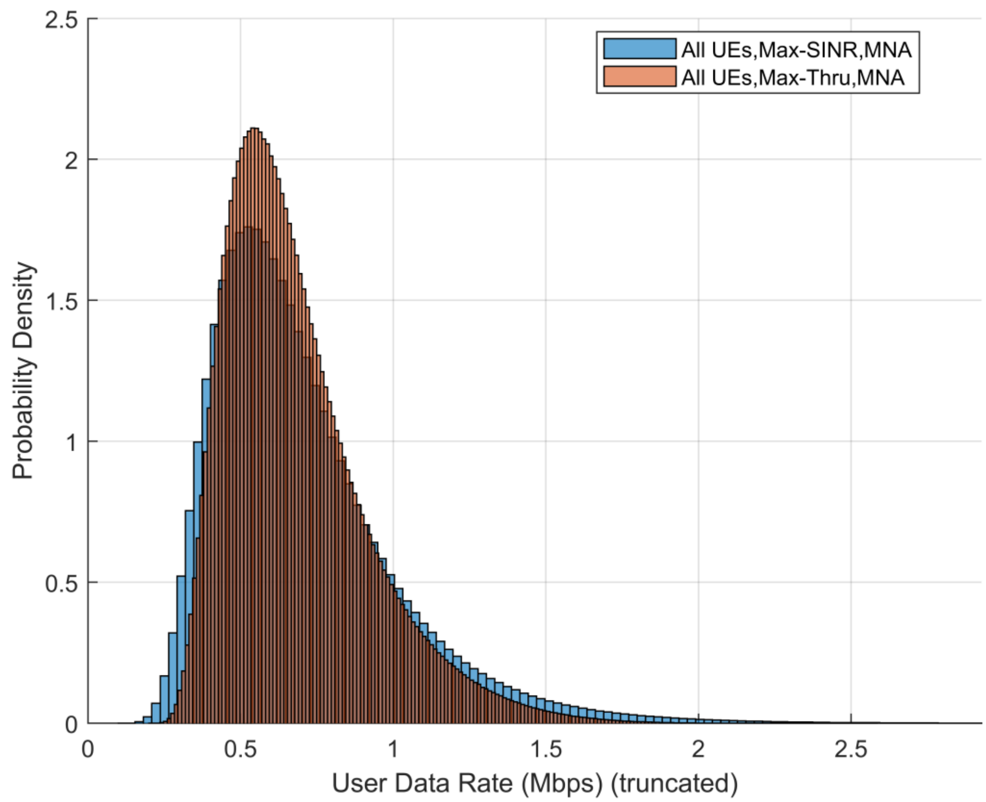

Figure 10 Probability density function of user data rate (Uniformly distributed towers; MNOs have same tower density and spectrum bandwidth; “Balanced” resource allocation)

3.4.5 Capacity Improvement from MNA Given Random BS Offset

Results in previous sections have assumed the most favorable BS offset. Because the distance between two BS’s in actual deployments varies in practice, we want to get a sense of how much MNA can increase capacity in the real world. To do that we calculated the mean UE data rate resulting from 16000 different BS layouts, each corresponding to a random BS offset described in 3.3. We assumed 100% MO-MVNO market share and "Maximum-SINR" MNO selection algorithm.

Figure 11 scatters circles whose color represents the mean data rate averaged over all UEs against offset angle on the x-axis and offset distance on the y-axis, given "Balanced" resource allocation scheme. First, note that for any given offset angle, as one moves up the y-axis one also moves up on the color scale, meaning that increasing the offset distance raises the mean UE data rate. We will further explore this relationship in the next section. Additionally, for any given offset distance on the y-axis, the color across the x-axis is largely uniform (hence the horizontal color strata), which implies that the offset angle does not significantly affect network capacity under MNA. Therefore, the direction in which two BS's are situated relative to each other will not greatly influence how useful MNA will be. The hollow white space at the top of Figure 11 is due to the geometry of a hexagon, as the distance from the center of a hexagon to its edges changes as one moves along the edges.

Figure 11 also shows a higher concentration of circles toward the upper y-axis. This is because when we randomly select a location in a hexagonal cell (BS at the center), the probability of that location being y meters away from the BS is greater than the probability of being x meters away if y is greater than x, again due to geometry. This property suggests that the offset distance will more likely to be large when we randomize the offset between two aligned hexagonal grids, the implications of which we will demonstrate in Figure 12 and Figure 13 that follow.

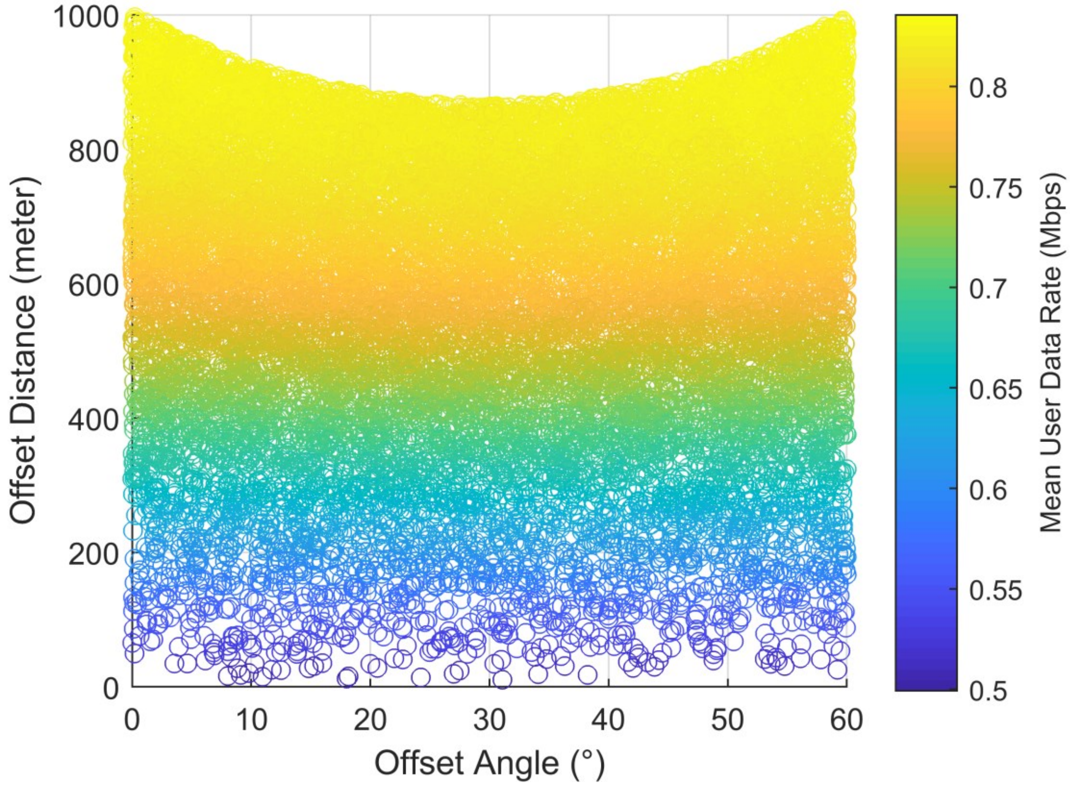

Figure 11 Mean UE data rate resulted from randomized offset (Maximum-SINR MNO selection; "Balanced" resource allocation). The offset angle does not significantly affect network capacity under MNA.

Figure 12 plots the histogram of the 16000 mean UE data rate given randomized BS offset. Each mean data rate is the average over all UEs in one particular BS layout from randomized offset. The histogram shows that the bulk of the distribution falls into the upper end of the data rate range, and that the mode of the distribution sits near the maximum value.

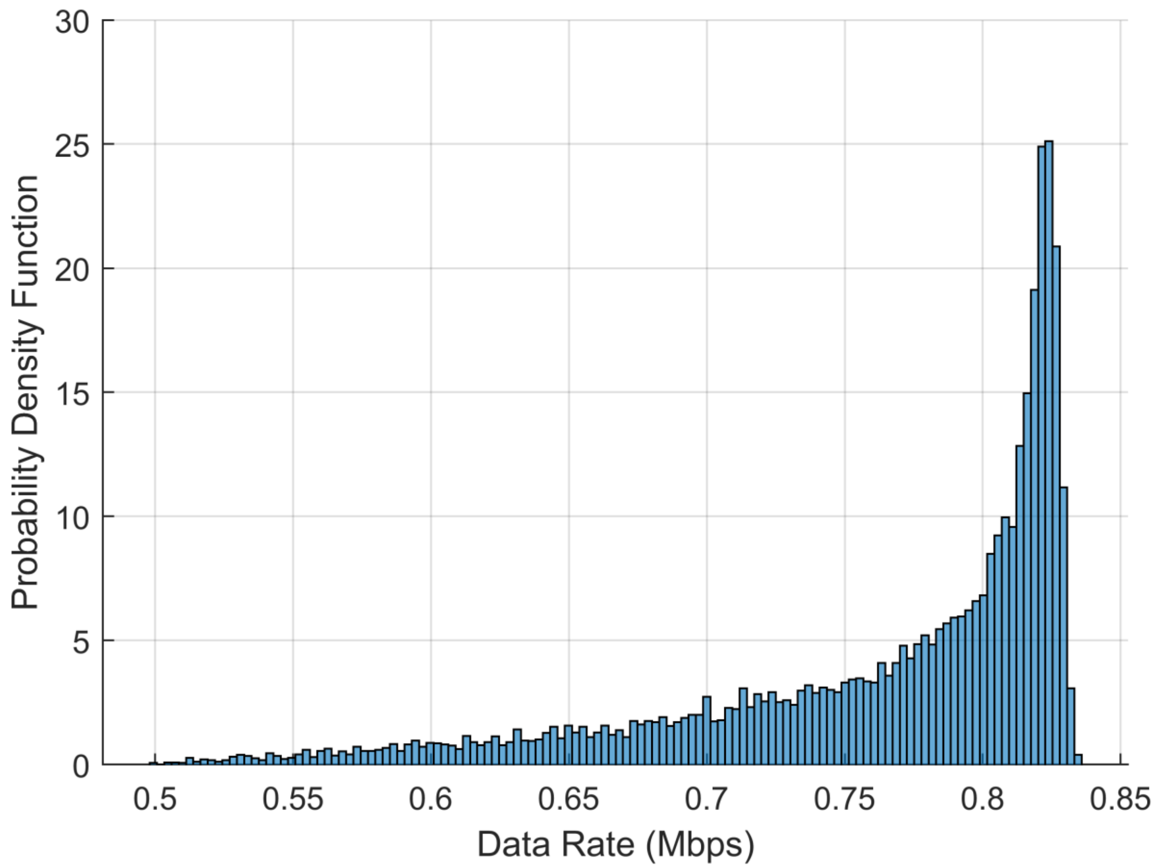

Figure 12 PDF of mean UE data rate given randomized offset (Maximum-SINR MNO selection; "Balanced" resource allocation). The bulk of the distribution falls into the upper end of the data rate range, and that the mode of the distribution sits near the maximum value.

Additionally, according to Figure 13, which presents the empirical cumulative distribution function (CDF) of the same sample, the 50th percentile mean UE data rate falls within 10% of maximum mean UE data rate. In other words, even if the offset is random, half of the time the resultant network capacity is close to the best-case scenario, as long as the BS's are not colocated. These observations about the histogram and the CDF hold for the "Equal-Allocation" and "Equal-Throughput" schemes, as well. While one might be concerned that the BS's of two MNOs need to be far enough apart for MNA to generate significant capacity gain, these results suggest that the benefit of MNA is in fact rather accessible and does not hinge on the BS's of participating MNOs being perfectly interspersed between each other.

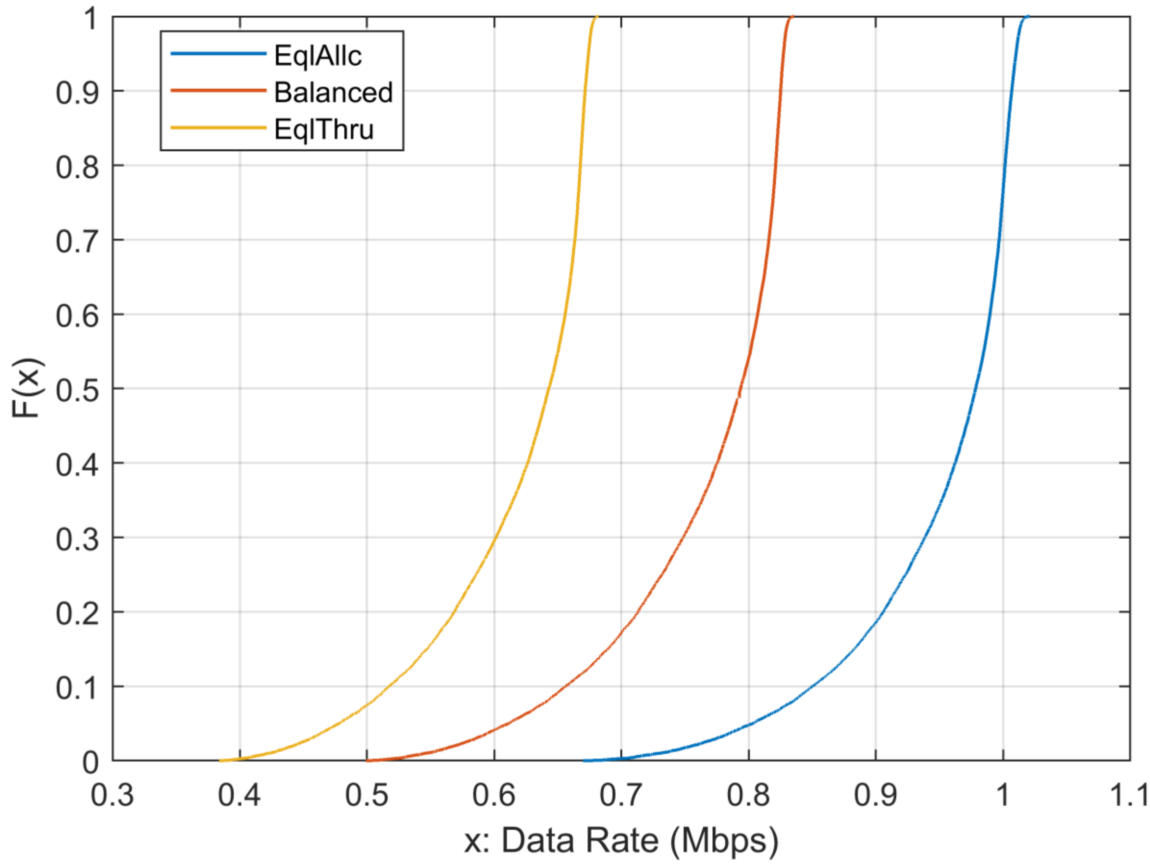

Figure 13 CDF of mean UE data rate (Maximum-SINR MNO selection; "Balanced" resource allocation). The benefit of MNA is accessible and does not hinge on the BS's of participating MNOs being perfectly interspersed between each other.

3.4.6 Effects of Varying Offset Distance on Capacity

MNOs are always looking for cost-effective ways to increase their network capacity. With the presence of MNA, capacity is a function of BS offset. In this section, we further explore the relationship between network capacity under MNA (in terms of mean UE data rate) and BS offset. This relationship allows an MNO to gauge how much capacity gain can be realized by adopting MNA in a given region based on where each operator's BS's are located in that region.

MNOs are also interested in the quality of service provided to individual subscribers, although not necessarily equally. For example, MNOs may pay more attention to those subscribers experiencing lower data rates and see if the service they receive meets a minimum requirement. To that end, in addition to the mean UE data rate, we will also rank all UEs based on the data rate they receive, and report the data rate of different percentiles as a function of offset distance.

Figure 14 plots on the left the UE data rate of the 5th, 10th, 20th, 50th, 75th and 95thpercentiles and the mean, and on the right the relative change in these data rates in percentage terms with respect to these data rates in the baseline scenario, with "Maximum-SINR" MNO selection and "Balanced" resource allocation. MNA generally improves the lower-percentile data rate more than it does high-percentile data rate. It can be a great tool for operators to improve the minimum level of service provided to customers. That said, even the 95^{th} percentile data rate can get as much as 50% improvement.

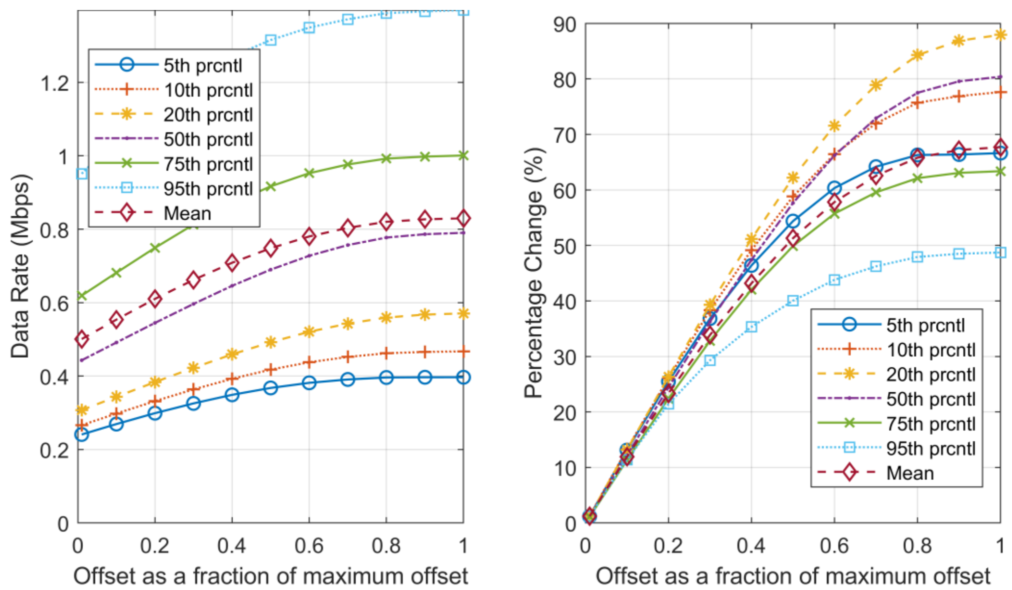

Figure 14 UE data rate and change in UE data rate relative to the baseline (Maximum-SINR MNO selection; "Balanced" resource allocation). MNA generally improves the lower-percentile data rate more than it does high-percentile data rate.

Increasing offset distance leads to higher network capacity, but experiences diminishing returns as all curves level off once beyond roughly 70% of the maximum offset. This partially explains why a perfect layout is not required for MNA to be useful as we have shown in Figure 12 and Figure 13.

Figure 15 plots the two different CDFs of UE data rates without MNA and with MNA (maximum offset distance at 0° offset angle). MNA is particularly effective in reducing the number of UEs experiencing low data rate. For example, in this experiment the fraction of UEs getting less than 0.5 Mbps is reduced from 60% to 15%.

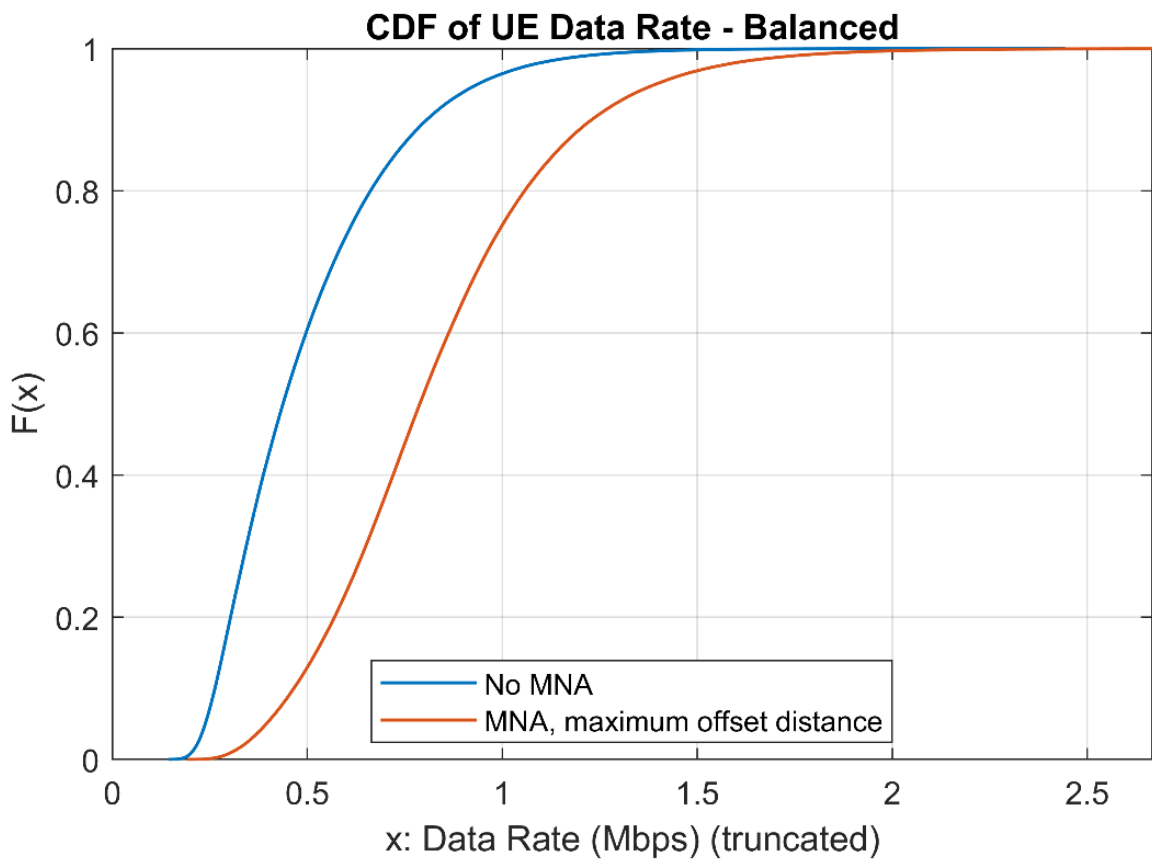

Figure 15 CDF of UE data rate with and without MNA (Maximum-SINR MNO selection; "Balanced" resource allocation). MNA is particularly effective in reducing the number of UEs experiencing low data rate.

3.4.7 When Partner MNOs Are Unequal in Size

We continue our analysis in a setting of two MNOs and one MO-MVNO, where each MNO has 48 BS's on a hexagonal grid and the two grids are offset by the maximum offset distance at 0° offset angle. The MNOs' market shares and spectrum holdings in the baseline scenario are varied while the total spectrum holdings and subscriber count are kept constant at 20 MHz and 4000, respectively. We assume that MNOs have identical capacity per subscriber. Therefore, each MNO's market share is varied in proportion to its spectrum holding, i.e. when MNO #1 has 40%

of the 20 MHz spectra, it also has 40% of the 4000 customers. In the MNA scenario, we assume 100% MO-MVNO market share, i.e. all MNO subscribers switch to the MO-MVNO.

Figure 16 plots the combined MNO capacity as a function of one MNO's share of spectra and subscribers for two different MNO selection algorithms. We found that the MNO selection algorithm plays a pivotal role in determining the total capacity with MNA when the two partner MNOs are unequal in size. With "Maximum-SINR", total capacity is independent of the relative spectrum holdings and market shares between the two MNOs. However, with "Maximum-Throughput", the more the two MNOs differ in size, the less total capacity can be realized. In a corner case where one MNO has 9x the spectrum and initial market share of the other, "Maximum-Throughput" results in 20% lower total capacity than the "Maximum-SINR" MNO selection algorithm. The discrepancy manifested in previous results between these two MNO selection algorithms becomes even more pronounced when the two partner MNOs are unequal in size. It is crucial that MNOs understand the consequences of the way MO-MVNO subscribers choose which network to use.

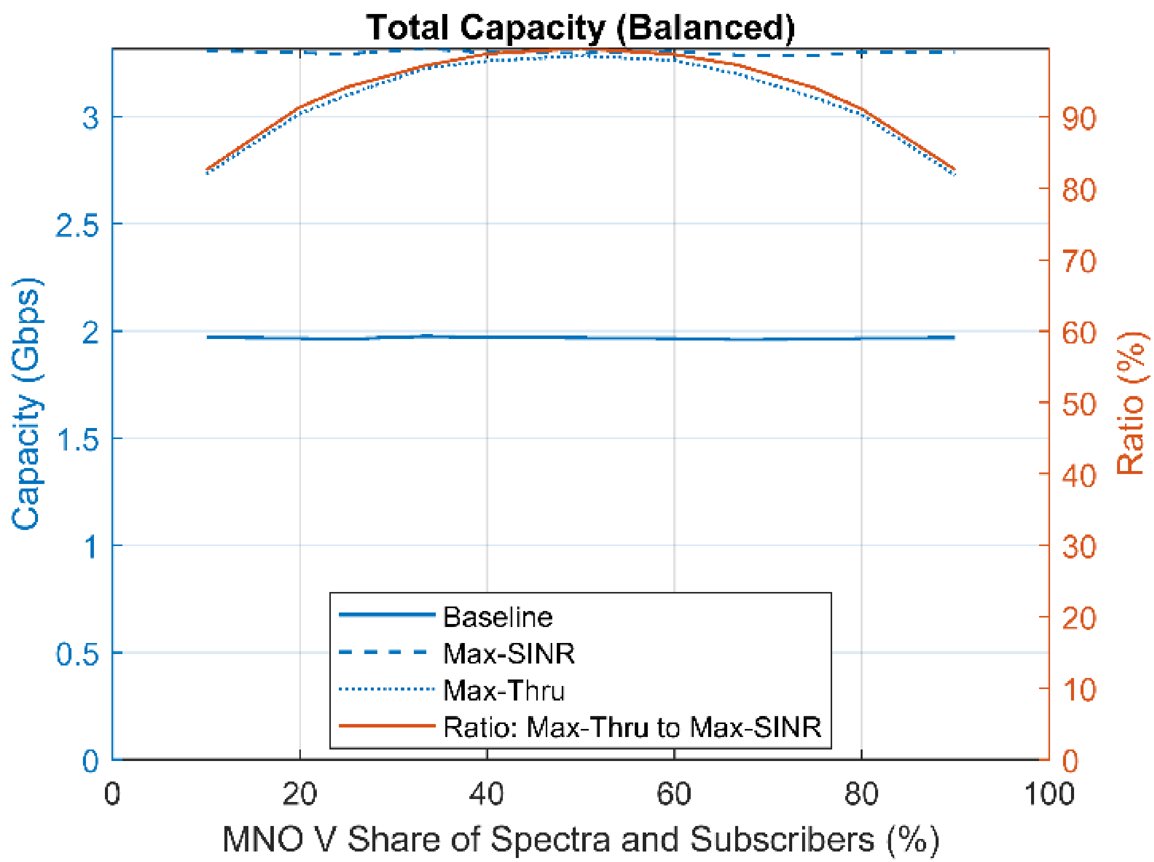

Figure 16 Combined capacity vs. one MNO's share of spectra and initial market ("Balanced" resource allocation). With "Maximum-Throughput", the more the two MNOs differ in size, the less total capacity can be realized.

While "Maximum-SINR" has advantages over "Maximum-Throughput" in terms of total capacity, it may come at the expense of throughput fairness. Figure 17 plots the 80th percentile, mean and 20th percentile subscriber data rate for each operator for "Maximum-SINR" and "Maximum-Throughput" MNO selection algorithms. With "Maximum-Throughput" (right column), subscribers across the board experience similar gains in data rate. With "Maximum-SINR" (left column), the mean data rate under MNA does not vary with one MNO's share of spectrum and initial market share, which is consistent with Figure 16. However, the 80th and 20th percentile data rates under MNA vary dramatically with one MNO's spectrum holdings and initial market share. Specifically, when there is a large imbalance (4x difference or more) in the partner MNOs' spectrum holdings and initial market share, the 20^{th} percentile data rate with MNA can be lower than that in the baseline scenario if MO-MVNO subscribers choose BS's that provide the highest SINR. On the other hand, the 80^{th} percentile data rate with MNA sees disproportionately large jumps from the baseline scenario. This is because the "Maximum-SINR" MNO selection algorithm assigns on average an equal number of UEs to two MNOs that have unequal amount of spectrum, causing unbalanced congestion. The share of UEs served by the smaller MNO is more than its share of spectrum. Therefore, the simplistic "Maximum-SINR" MNO selection algorithm, while effective in maintaining overall spectral efficiency, is detrimental to throughput fairness when partner MNOs differ in size.

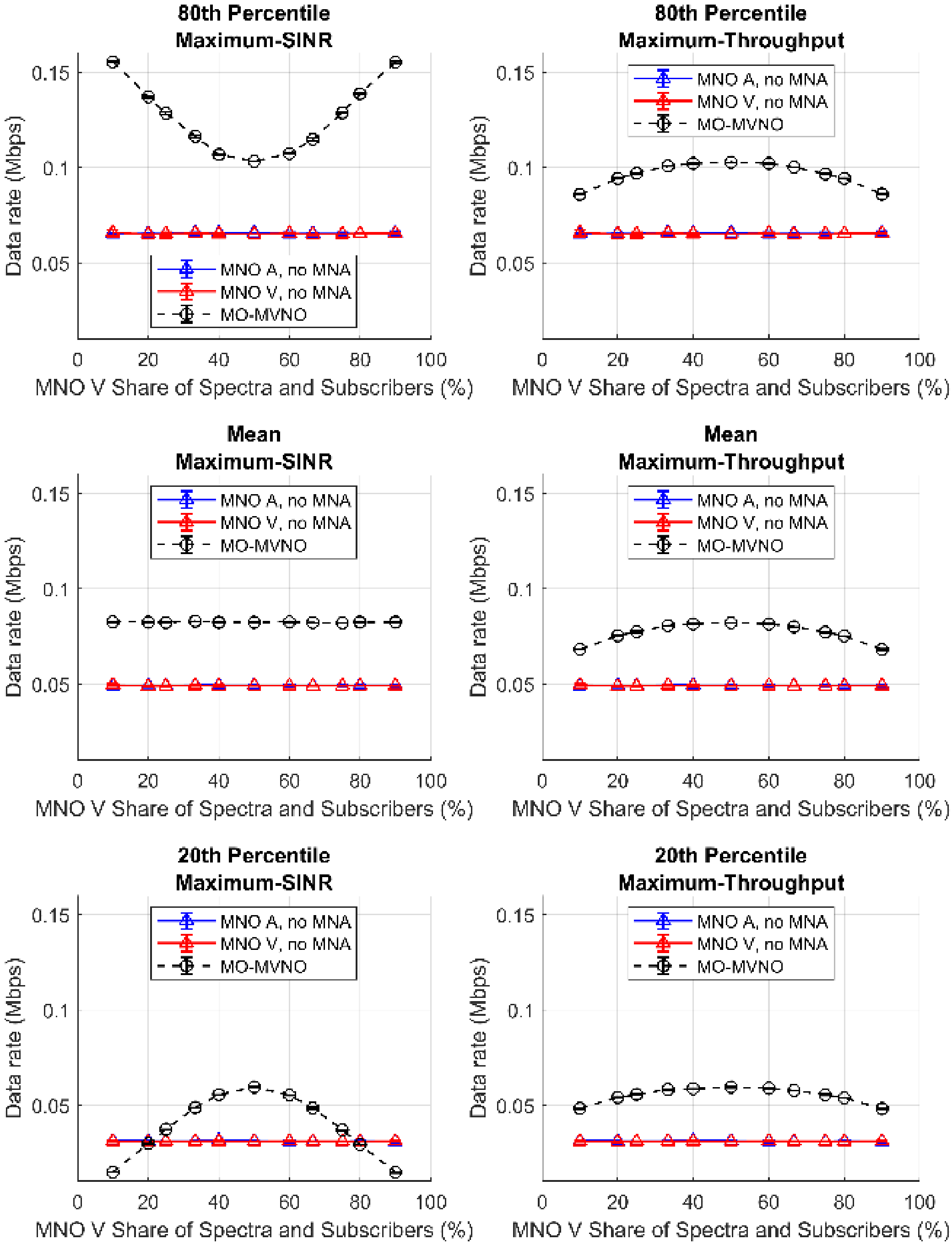
80th Percentile
Maximum-SINR
80th Percentile
Maximum-Throughput
Mean
Maximum-SINR
Mean
Maximum-Throughput
20th Percentile
Maximum-SINR
20th Percentile
Maximum-Throughput
Data rate (Mbps)
MNO V Share of Spectra and Subscribers (%)
MNO A, no MNA
MNO V, no MNA
MO-MVNO
0.15
0.1
0.05
0
20
40
60
80
100

Figure 17 Subscriber data rate vs. one MNO's share of spectrum and initial market ("Balanced" resource allocation). While "Maximum-SINR" has advantages over "Maximum-Throughput" in terms of total capacity, it may come at the expense of throughput fairness.

Figure 18 plots the share of traffic carried by one of the two partner MNOs as a function of that MNO's share of spectrum and initial market for two different MNO selection algorithms. With "Maximum-SINR", the share of traffic carried by each MNO is the same with or without MNA, and is equal to the MNO's share of spectrum and initial market. With "Maximum-Throughput", the MNO with less spectrum and lower initial market share will carry a larger share of traffic with MNA than without MNA, and more than its share of total spectrum. The discrepancy originates from the fact that, due to BS's having unequal amounts of spectrum, the BS that provides the highest SINR does not necessarily provide the highest data rate. Certain UEs would therefore attach to a BS that is farther away but has more spectrum available. This, however, means that the larger MNO is going out of its way to serve more UEs, while the smaller MNO only serves those UEs close to its BS's. The smaller MNO would have a smaller effective cell size than the larger MNO and therefore, it can extract more bps per Hz of spectrum than the larger MNO. Such discrepancy may hurt a larger MNO's incentive to partner with an MO-MVNO.

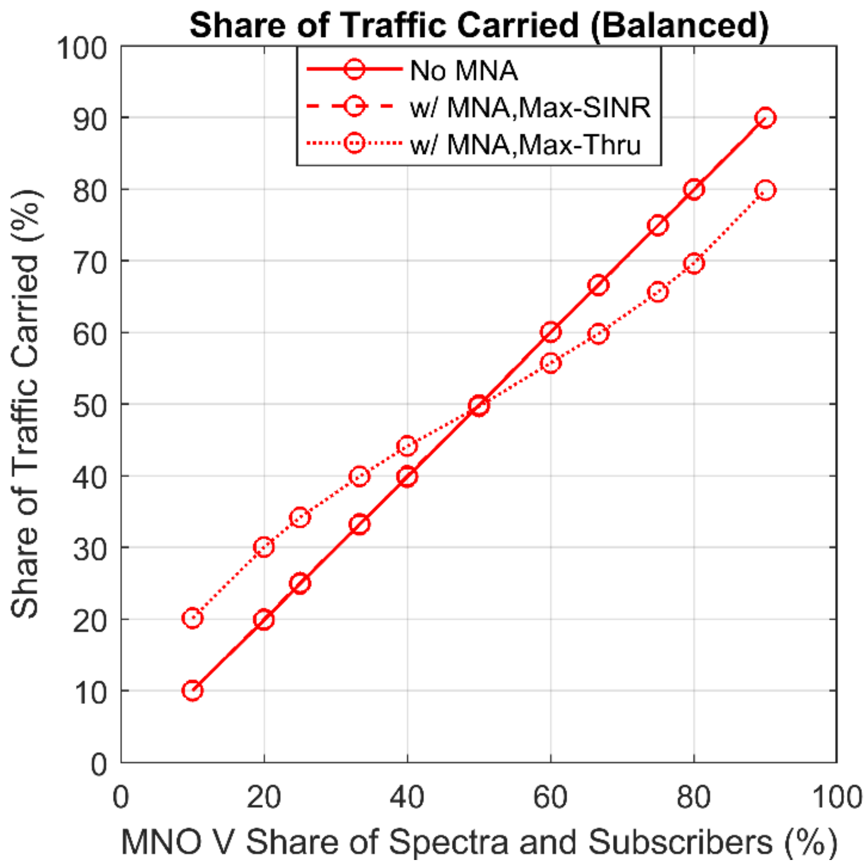

Figure 18 Share of traffic carried by one MNO vs. its share of spectra and initial market (Balanced resource allocation). With "Maximum-Throughput", the MNO with less spectrum and lower initial market share will carry a larger share of traffic with MNA, and more than its share of total spectrum.

3.5 Conclusions

In this chapter we explored the implications of MNA on the capacity of cellular networks. MNA can greatly increase network capacity. When all UEs can utilize multiple networks, the gain in network capacity ranges from 40% to 80%, depending on the exact resource allocation scheme and MNO selection algorithm being used. This benefit monotonically increases with a growing fraction of MNA-capable UEs as is expected. MNA can expand network capacity without additional spectrum or additional infrastructure, making it more cost-effective for operators to provide a given capacity. However, we note that in an MNA arrangement, operators rely on each other to serve customers, which could reduce competition.

We studied how base station layout affects the capacity of cellular networks with MNA. We found that the capacity increases as the offset between the BS's of partner MNOs increases. However, the benefit of MNA does not depend on maximum offset between the BS's of partner MNOs. The median value of the mean UE data rate given a randomized offset between two hexagonal grids is almost on par with the mean UE data rate given the most favorable offset between two hexagonal grids.

We investigated how the network capacity with MNA is influenced by the selection algorithm by which an MNA-capable UE selects a BS to attach to. We found that if MNA-capable UEs attach to BS's that offer the highest data rate, the average subscriber data rate of any operator is lower than if MNA-capable UEs attach to BS's that provide the highest SINR. In other words, the optimal choice for each individual MO-MVNO subscriber hurts not only the MNOs, but also the MO-MVNO itself.

This effect is even more apparent when the two partner MNOs differ in size. In scenarios where partner MNOs have the same BS density but varying spectrum holdings in proportion to their initial market share, the MNO selection algorithm takes on an even greater role in shaping the outcome of an MNA arrangement. The greater the size differential between the two partner MNOs, the lower total capacity can be realized by a selection algorithm that chooses the BS offering higher data rate. However, when the size differential is large, a selection algorithm that favors the higher data rate will yield much greater fairness, and a much higher 20th percentile data rate than an algorithm that chooses maximum SINR. For that reason, when there is a large spectrum imbalance, an MO-MVNO that cares about both average data rate and fairness may maximize SINR when its market share is small, but not when its market share is large.

When MO-MVNO subscribers choose MNO based on data rate, the MNO with greater spectrum holdings may carry a share of traffic that is less than its share of spectrum contributions in an MO-MVNO partnership. Depending on the wholesale pricing structure, this can make larger MNOs reluctant to partner with an MO-MVNO.

The fact that a simple maximum-SINR selection algorithm is highly effective is encouraging from an implementation perspective, since obtaining precise expected data rate is difficult if not impractical for an MO-MVNO UE. However, this simple algorithm may not be desirable when the two partner MNOs differ greatly in size, as the MNO with less spectrum may be disproportionally burdened, reducing fairness among UEs. Under these circumstances, a more sophisticated MNO selection algorithm may be necessary.

Even though MNA can improve the data rate when averaged over all UEs, MNO subscribers may see considerably lower data rate than MO-MVNO subscribers, which may be detrimental to the MNOs' business. For an extreme example, when a resource-fair allocation scheme is used, MNO subscribers do not see a higher data rate from MNA at all. MNOs can avoid such a disadvantage if they grant their own subscribers proportionally more resource blocks.

4 MNA on Infrastructure and Spectrum Resource Usage in The Presence of Colocation

4.1 Background

MNA has tremendous potential in improving spectral efficiency by taking advantage of the location diversity of base stations belonging to different MNOs. Chapter 3 quantifies how the additional spectral efficiency translates into higher capacity out of existing network infrastructure and spectrum holdings with little additional cost. We now switch to the perspective of dimensioning a new network to meet a future capacity requirement. We will investigate the potential infrastructure and spectrum resource savings as a result of adopting MNA as compared to building networks that produce the same capacity without MNA.

In this chapter we also address the issue of colocation, which has been a popular method employed by networks operators to reduce infrastructure cost. The benefit of MNA on network capacity hinges on the location diversity of the transmitters belonging to different MNOs. Colocation reduces that diversity, and in so doing may temper the efficiency gain from MNA. We will examine if MNA is still useful given the common practice of colocation in the industry.

Our method, assumptions and simulation model are described in Section 4.2. We present the results in 4.3 and conclude this chapter in Section 4.4.

4.2 Method

To determine how many tower and spectrum resources can be saved by adopting MNA, we built a MATLAB simulator that computes the amount of resources needed to produce a given downlink capacity with MNA and without MNA, for a preset distribution of tower and UE locations. The MATLAB simulator used in Chapter 3 has been updated in many ways, which we discuss below.

4.2.1 Key Assumptions

Like in Chapter 3, we assume that UEs are stationary and uniformly distributed throughput the simulated area. We also assume a full-buffer traffic model, i.e. UEs always have pending data to receive from BS's.

Unlike in Chapter 3, we assume that towers are uniformly distributed throughout the simulated area. We choose this assumption because it allows us to easily examine situations in which partner MNOs have different tower densities, and because it naturally includes cells of various shapes and sizes. These advantages are balanced by a weakness: because tower locations are generated independently of UE locations, there will be areas with many UEs and no towers nearby, and areas with towers but very few UEs to serve.

To see how useful MNA is in the presence of tower colocation, we compare the downlink capacity with MNA and without MNA, while varying the extent of tower colocation in the network. To model different extents of colocation, we keep the number of cells owned by each MNO constant while varying the number of cells that share a tower with the other MNO. Once the numbers of colocated and non-colocated towers are set, the x/y coordinates representing the locations of each site are randomly generated from a uniform distribution. Consistent with our previous assumption, the towers that are colocated are chosen independently of UE densities.

We assume each BS transmits at a fixed power P_t across the available bandwidth B. The path loss model follows the 3GPP E-UTRA RF specifications Release 15 [38] and is given by (5) in dB, where 700 MHz carrier frequency and 30-meter antenna height are assumed.

$$L(d) = 113.2 + 35.2\,log_{10}\,d + LogF \quad (5)$$

The term $LogF$ is a normal random variable with zero mean and $\sigma = 10$ dB standard deviation, and models large-scale fading (shadowing due to clutter). Additionally, the fading in the channels between a UE and two non-colocated transmitters has a correlation coefficient of 0.5; the fading in the channels between a UE and two colocated transmitters has a correlation coefficient of 1.0.

We assume that the downlink data rate approaches the Shannon limit. The data rate r_j in bits per second for UE j at distance d from the BS is $r_j(d) = s_j B\, log_2\left(1 + \frac{s_j P_r^j(d)}{s_j I^j + s_j B N_0}\right)$ or

$$r_j(d) = s_j B\, log_2\left(1 + \frac{P_r^j(d)}{I^j + BN_0}\right) \quad (6)$$

Here, s_j is the share of spectrum resources assigned to UE j in its cell (further discussed later); $P_r^j(d)$ is the received signal power at UE j across available bandwidth B at distance d from the BS; I^j is the sum of interference power across B from all other co-channel BS's, where we have assumed a frequency reuse factor of 1. We assumed a fixed noise power spectral density N_0.

For a given communications standard (e.g. LTE), network capacity depends primarily on tower density and the available spectrum bandwidth. With increased spectral efficiency, an MNO can carry the same traffic with less infrastructure, less spectrum or less of both. In computing potential resource savings, we assume that MNOs adjust tower density and spectrum bandwidth in tandem, as a cost-minimizing MNO would do [39] when cost per tower and cost per MHz of spectrum are fixed.

4.2.2 Simulation Model (Chapters 4 & 5)

This subsection describes the simulation model used for analyses in Chapter 4 and 5.

We simulated an area of approximately 7 km by 6 km that is wrapped around at all four edges, creating effectively a torus. In the case where both partner MNOs have the same amount of tower and spectrum resources, each MNO has 16 towers and 60 MHz of spectrum bandwidth, of which 30 MHz is for downlink. If the 16 towers were distributed on a hexagonal grid in this area, each cell would have a radius of roughly 1 km.

Transmit power P_t is 40 Watts or 46 dBm. Noise power spectral density $N_0 = -174$ dBm/Hz.

We assume that there are 3200 active UEs in total across the two partner MNOs and the one MO-MVNO. We vary the fraction of UEs that are subscribers of the MO-MVNO, while the remaining UEs are assumed to be split between the partner MNOs in proportion to the MNOs' capacities.

4.2.3 Resource Allocation Scheme and Network Selection Algorithm

The resource allocation schemes and network selection algorithms remain the same as those used in Chapter 3 analyses. Readers can refer to 3.2 for the details.

Unless otherwise specified, all the results that follow assume "Rate-Adjusted" resource allocation to balance the QoS of those capable of MNA and of those incapable.

4.3 Results

4.3.1 MNA on Capacity

Given the results in Chapter 3 based on the hexagonal grid model as a reference, we first present the MNA's effect on capacity given random tower locations. The relative change in the aggregate capacity of both partner MNOs (Figure 19) is slightly smaller than those under a hexagonal grid model with the most favorable offset (Figure 5). The difference is on the order of 5 to 10 percentage points at baseline, depending on the resource allocation scheme and network selection algorithm. Two factors are at play here. The randomness of tower locations reduces the effectiveness of MNA because unlike the case of two hexagonal grids with the most favorable offset, the locations of each MNO's infrastructure are independent of each other and not designed to complement each other. The second factor is that fading rewards the ability to choose from a larger pool of distinctively located transmitters. Fading was not considered in the analyses based on the hexagonal grid model whereas it was considered in the random tower model.

The relative standings of the gains under each combination of resource allocation scheme and network selection algorithm remain changed from the hexagonal grid model with the most favorable offset. "Max-SINR" selection algorithm consistently produces higher capacity gains than "Max-Throughput" selection algorithm, and the difference between them is larger when the resource allocation scheme prioritizes throughput fairness rather than total capacity.

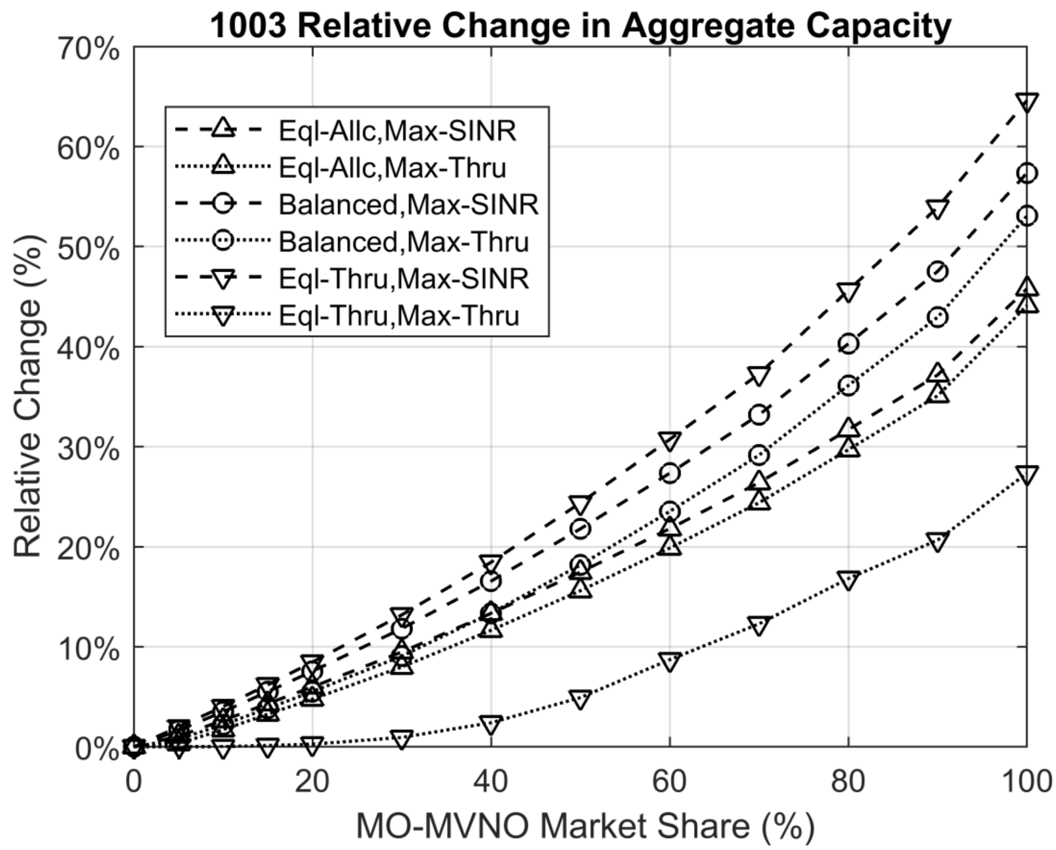

Figure 19 Relative change in aggregate network capacity vs. fraction of MNA-capable users. (Note: figure legends have changed from Chapter 3 figures. MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

4.3.2 Infrastructure and spectrum resources savings without colocation

If the fraction of MNA-capable UEs is high, such as when MNOs share capacity directly with each other (referred to by some as "flexible roaming" [9] or "smart roaming" [11]), the same network capacity can be realized with significantly fewer tower and spectrum resources. Figure 20 shows the fraction of tower and spectrum resources saved as a function of this fraction, which also equals MO-MVNO market share. At baseline, when all UEs are MNA-capable and given a resource allocation scheme that balances total capacity and throughput fairness, 20% of resources can be saved if MNA-capable UEs attach to towers of higher SINR, or 19% if choosing towers of higher expected data rate.

MNA has economies of scale in the fraction of MNA-capable UEs. As shown in Figure 20, the amount of resources saved increases with the fraction of MNA-capable UEs faster than linearly. With retail prices held constant, the combined profit of the MO-MVNO and MNOs increases faster than linearly as well. Thus, as an MO-MVNO gains market share, it can demand lower wholesale prices, or MNOs are more incentivized to work with MVNOs, or both.

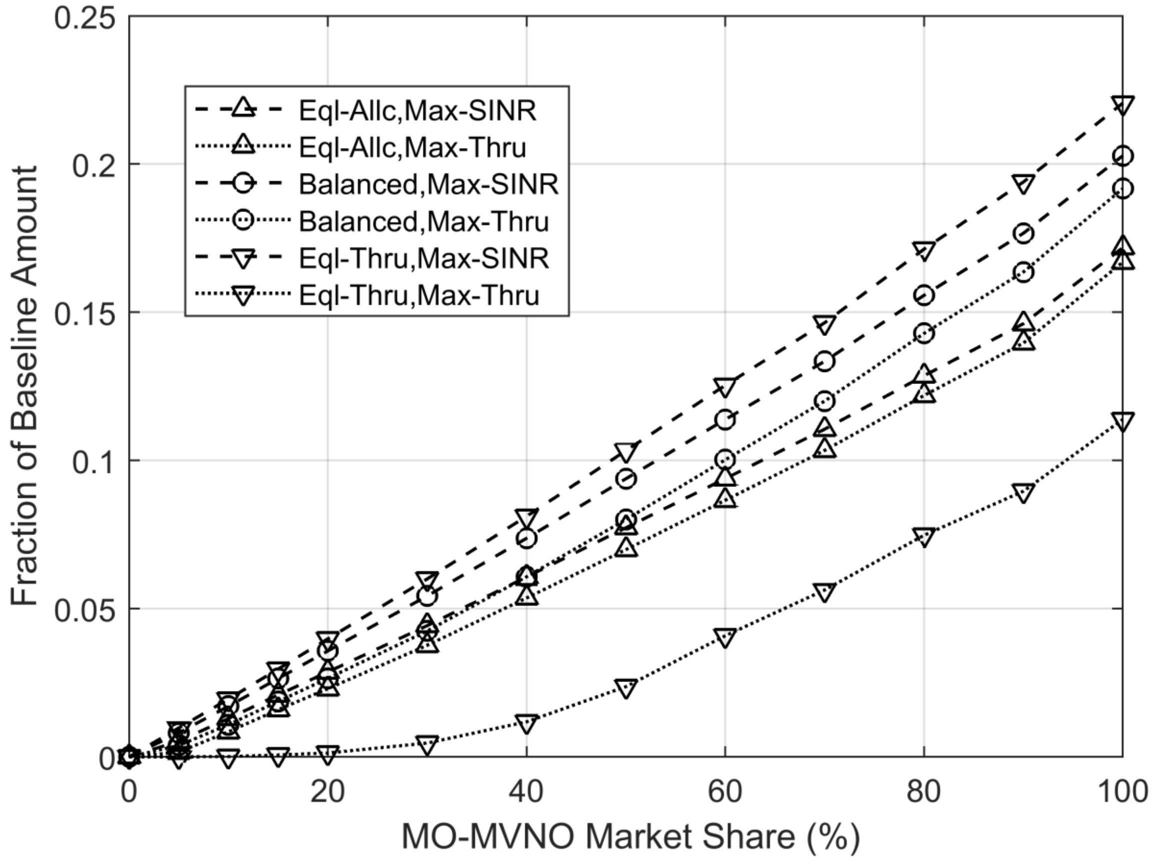

Figure 20 Tower and spectrum resources saved with MNA as a fraction of baseline amounts (MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

The network selection algorithm employed by MNA-capable UEs plays a significant role in determining the cost-efficiency of MNA networks. MNA is more cost-effective (as measured in tower and spectrum cost per bit of traffic carried) when MNA-capable UEs attach to towers of higher SINR rather than higher expected data rate, because Shannon's theorem predicts a higher spectrum efficiency for a channel with higher SINR. Figure 20 shows that this difference is large at baseline. Most notably, when only a small fraction of UEs are MNA-capable (e.g. 10%), the selection algorithm makes a huge difference. There are still tangible resource savings if UEs attach to towers of higher SINR. However, if they attach to towers of higher expected data rate, there may be negligible resource savings. From a business perspective, an MVNO such as Google Fi could make the case to its partner MNOs for why they should receive a discount on wholesale rates if their subscribers instead attach to towers of higher SINR, as that makes the MNO networks more efficient.

Consistent with previous observations, the difference in cost-efficiency between the two selection algorithms is larger when MNO resource allocation prioritizes throughput fairness, and is smaller when MNO resources allocation prioritizes total capacity.

4.3.3 Infrastructure and spectrum resources savings in the presence of colocation

The previous results assume no colocation. In reality, MNOs colocate their cell sites to various extents, which potentially reduces the effectiveness of MNA.MNA can still be useful even with moderate (<50%) levels of colocation. Figure 21 shows the fraction of tower and spectrum resources saved as a function of the fraction of an MNO's cells that are colocated with the other MNO. As expected, when more base stations are colocated, the fraction of resources saved decreases. The concavity of the curves indicates that the drop is small unless colocation is widespread. For example, between no colocation and 50% colocation, the fraction of resources saved only decreased from 20% to 15%. Once colocation surpasses 50%, however, the benefit of MNA diminishes more quickly.

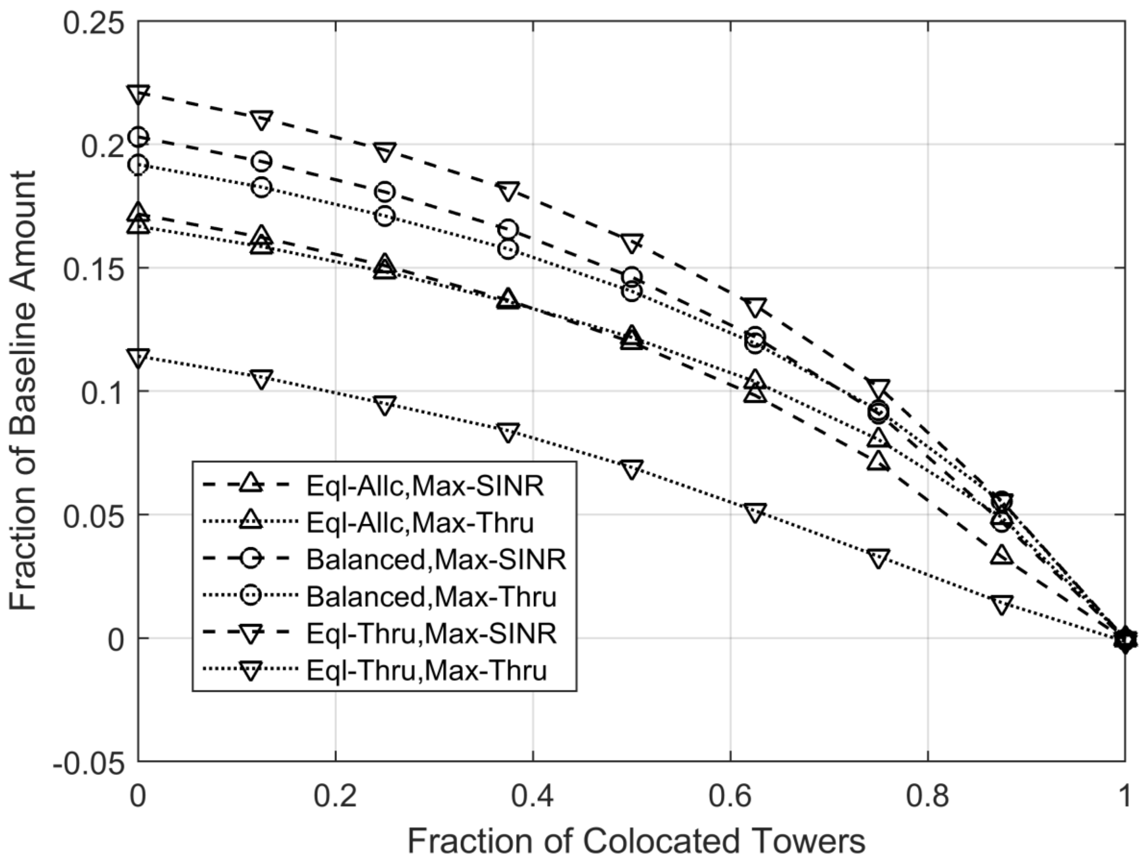

Figure 21 Tower and spectrum resources saved with MNA as a fraction of baseline amounts vs. extent of colocation (Partner MNOs have equal tower density and spectrum bandwidth; 100% MNA-capable UEs; log-normal fading with 10 dB standard deviation)

The concavity can be explained by the relative change in the number of distinctively located cell towers. In this case, for example, moving from 0 to 1/16 colocation is a relatively smaller decrease (from 32 to 30) in the number of distinctvely located cell towers than moving from 15/16 colocation to complete colocation (from 18 distinctively located cell towers to 16). Therefore, the rate at which colocation reduces capacity is slow at first and then gradually increases.

The effect of the network selection algorithm with colocation is different from previous results. Most notably, when the resource allocation scheme is one that favors total capacity like "Equal-

Allocation”, choosing a network with a higher SINR produces lower total capacity under MNA than choosing higher expected data rate, as shown in Figure 22; the difference is small, though.

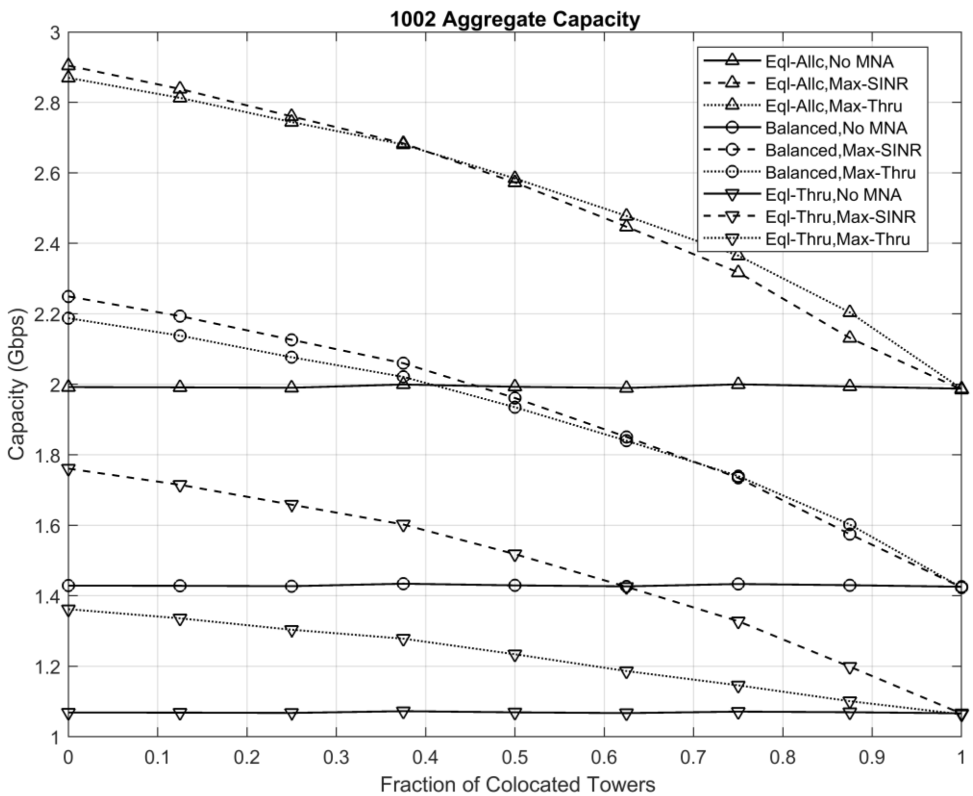

Figure 22 Parnter MNOs' combined network capacity vs. extent of colocation (Partner MNOs have equal tower density and spectrum bandwidth; 100% MNA-capable UEs; log-normal fading with 10 dB standard deviation)

The reason may be uneven user SINR distribution for the two base stations colocated at the same cell tower. We discuss this in more detail in 0.

4.4 Conclusions

When MNA is widely adopted, a given network capacity can be realized with much fewer tower and spectrum resources than if MNOs build standalone networks to produce the same combined capacity without MNA. The amount of resource savings increases faster than linearly with increasing fraction of MNA-capable users, due to the resource allocation adjustment that compensates the remaining MNA-incapable users in the system. Having MNA-capable users served by the network that provides higher SINR generally produces higher total capacity than having them choose the network that provides higher expected data rate. But the opposite is true when colocation is prevalent. The network selection algorithm is particularly influential when the fraction of MNA-capable UEs is low, as is likely the case for a newly established MO-MVNO, and when MNOs' resource allocation scheme prioritizes throughput fairness over total throughput. This is consistent with Chapter 3 results.

The benefit of MNA is robust against moderate levels of colocation. Even with half of all base stations colocated, MNA can still provide significant spectral efficiency improvements. Given that the extent of colocation varies widely across MNOs and geographic areas, there is no single answer to how useful MNA can be; whether MNA can be beneficial should be determined on a city-by-city and operator-by-operator basis. While network capacity under MNA always decreases with more colocation, colocation is a cost-saving measure in and of itself. The net effect of MNA and colocation on the cost per unit capacity will require further analysis, which will be addressed in the next chapter.

5 The Economics of MNA

5.1 Background

By allowing user equipment (UEs) to be served by a larger set of distinctively located base stations (BS's), MNA improves the spectral efficiency of wireless networks [8]. As shown in previous chapters, MNA can increase network capacity by over 60% without additional spectrum or infrastructure. As the world moves to build networks that can accommodate the ever-increasing demand for mobile data, adopting MNA can reduce the resources needed to achieve that goal, lower cost and increase social welfare. In this chapter, we examine the economics of MNA.

We are first concerned with how much MNA can reduce the cost of cellular data services. Tower infrastructure and spectrum licenses account for significant expenses for facility-based MNOs, but fewer of these resources are needed to produce a given capacity if MNA is adopted, thanks to the higher spectral efficiency it enables. In this chapter, we will take the perspective of dimensioning a new network to meet a future capacity requirement, and investigate the potential cost savings from incorporating MNA into the system as compared to building networks without MNA that produces the same capacity. Building upon the results from Chapter 4, we will determine the economic value of the potential resource savings in infrastructure and spectrum.

With MNA, the total network capacity depends on many factors, including the fraction of UEs that are capable of MNA, the relative tower density and spectrum bandwidth between partner MNOs, the network selection algorithm employed by MNA-capable UEs, and MNOs' resource allocation schemes. We will examine the potential cost savings from MNA in a variety of scenarios along these dimensions. We will then determine how the cost savings can translate into lower prices for consumers and/or higher profits for operators.

The benefit of MNA on network capacity hinges on the location diversity of the transmitters belonging to different MNOs. Colocation reduces that diversity, and in so doing moderates the spectral efficiency gain from MNA as was shown in Chapter 4. That said, colocation is a cost-saving measure itself, allowing MNOs to share the often time-consuming and expensive processes of site acquisition, tower construction and land lease. With colocation but no MNA, each cell is less expensive, but more cells are needed to produce a given capacity. With MNA but no colocation, each cell is more expensive to build, but fewer cells are needed to produce a given capacity. This tradeoff motivates the following questions: which is more cost-effective, colocation or MNA? Assuming MNA is adopted, is colocation still useful in improving cost-effectiveness?

Besides Google Fi, the adoption of MNA has been quite limited to the best of our knowledge, which prompts a question about incentives. Under what conditions are MNOs willing to partner with an MO-MVNO? Given the potential cost savings, when would an MNO offer wholesale discounts, and by how much? Building upon the cost analyses, we also seek to characterize the business arrangements that allow both MNOs and MO-MVNOs to benefit from MNA.

From an MNO's perspective, partnering with an MO-MVNO can be a double-edged sword. On one hand, MNOs may save costs and gain wholesale revenue, some of which may come from previous subscribers of a competing MNO. On the other hand, an MNO's retail revenue may diminish as some subscribers switch to an MO-MVNO. Unless the net effect is positive on the bottom line, an MNO is unlikely to adopt MNA. From an MO-MVNO's perspective, it also needs to be profitable to stay in business, which creates another constraint on the range of wholesale prices that would make MNA attractive for all operators involved. As we will demonstrate, an MNA arrangement should factor into not only business decisions, like how much each MNO invests in infrastructure and spectrum resources, but also technical parameters, like resource allocation schemes and network selection algorithms. We will discuss how these aspects influence the range of wholesale prices that can incentivize MNOs and MO-MVNOs alike to adopt MNA. We will identify realistic scenarios where wholesale prices may appear unusual, and yet still benefit all stakeholders.

The rest of the chapter is structure as follows. Section 5.2 describes our method and engineering economic model. Section 5.3 presents and discusses the results. Section 5.4 concludes this chapter.

5.2 Method

In Chapter 4, we have built a MATLAB simulator that computes the amount of resources needed to produce a given downlink capacity with MNA and without MNA for a preset distribution of tower and UE locations, in order to determine how many tower and spectrum resources can be saved by adopting MNA. The outputs of the simulator feed into an engineering economic model that calculates an operator's revenue and costs. We determine if an operator has the incentive to adopt MNA by comparing the operator's profit with MNA against that without MNA. We focused on the MO-MVNO form of MNA, and looked at an MO-MVNO that partners with two MNOs.

In comparing the economic outcomes of all operators before and after they participate in an MNA arrangement with an MO-MVNO, we assume that the MNOs' combined network capacity is the same with MNA as it is without MNA. We also assume that the total number of customers is the same with or without MNA. Given that total capacity, total number of users and expected user data rate are the same with and without MNA, we can reasonably assume that the retail price for mobile data service is also the same with or without MNA. In other words, the total revenue opportunity is the same with or without MNA. This allows us to compute the change in an operator's profitability independent of a specific (and potentially inaccurate) demand function and price elasticity. Another way to look at it is that we are modeling MNA as an option for operators to meet a future capacity requirement which is exogenously given, rather than a way to extract more capacity from existing tower and spectrum resources.

While not explicitly considered in our model, there are other costs associated with bringing MNA to fruition. For example, devices that support more spectrum bands and communications standards in order to take advantage of MNA might be more expensive than those that do not. There might be a cost on the network side if the base station scheduler software in existing infrastructure needs to be updated to maintain certain fairness objectives under MNA, as we will discuss later. There is also a transaction cost in developing an MNA agreement that satisfies all stakeholders, be it an MVNO wholesale agreement or a capacity sharing agreement. We assume these costs are insignificant relative to infrastructure and spectrum expenses.

5.2.1 Engineering Economic Model

We assume that an operator's retail and wholesale revenues are proportional to the retail and wholesale traffic volumes it carries, respectively. Total revenue for MNO i ($i = \{A, Z\}$) R_{ik} is given by (4), where p^r is the retail revenue per unit of traffic, and p_{ik}^w is the wholesale price per unit of traffic MNO i charges. Subscript k denotes whether MNA is adopted. $k = 0$ if MNA is not adopted. $k = 1$ if MNA is adopted.

$$R_{ik} = p^r v_{ik}^r + p_{ik}^w v_{ik}^w \tag{7}$$

This resembles the pricing model of today's operators, which usually charge more for a higher data quota. While many operators also sell "unlimited" plans, most of them include a hidden data consumption threshold beyond which a subscriber's peak data rate will be significantly throttled. There may be intangible benefits from having a direct relationship with end users that are not necessarily reflected in the revenue from selling cellular data, like better ad targeting. Such benefits are not explicitly considered in our model.

We assume that an MNO's cost for spectrum C_{ik}^{S} is proportional to its spectrum holdings bandwidth B_{ik}, and that all MNOs face the same cost per unit of spectrum per unit of time per unit of area k^{S} ($\frac{\$}{km^2 \cdot Hz \cdot Sec}$).

$$C_{ik}^{S} = k^{S} B_{ik} A \tag{8}$$

Since spectrum licenses are routinely renewed at a minimal cost, they are considered "indefinite-lived" assets [40]. We calculate the cost of spectrum as the interest paid on the initial capital expenditure to acquire the spectrum as follows, where $p_{auction}$ is the price per MHz-Pop , ρ is population density ($\#/km^2$), and d is the interest rate or cost of capital:

$$k^{S} = p_{auction} \rho d \tag{9}$$

Operators also incur a cost to perform marketing, customer support, billing and other retail functions (hereinafter referred to as simply "marketing cost"). We assume that a fixed share k^{M} of retail revenue is used to perform marketing.

$$C_{ik}^{M} = k^{M} p^{r} v_{ik}^{r} \tag{10}$$

Tower cost is proportional to the number of towers, with coefficient k^T. A portion of the tower cost may depend on the capacity per tower or the spectrum bandwidth available at each tower. We assume that any such variation is insignificant compared to overall tower costs.

$$C_{ik}^T = k^T t_{ik} \tag{11}$$

Let $p^{rnm} = p^r - k^M$. For MNO i: Profit = retail revenue + wholesale revenue – spectrum costs – tower costs – marketing costs.

$$\pi_{ik} = p^{rnm} v_{ik}^r + p_{ik}^w v_{ik}^w - k^S B_{ik} A - k^T t_{ik} \tag{12}$$

The profit for the MO-MVNO:

$$\pi_{MO-MVNO,k} = \sum_{i=A,Z} v_{ik}^w (p^{rnm} - p_{ik}^w) \tag{13}$$

A summary of variables is provided in Table 1.

5.2.2 Feasible Wholesale Prices

We assume that MNOs would only partner with an MO-MVNO if they could earn more profits, and that an MO-MVNO must earn positive profits to remain in business. A pair of partner MNO wholesale prices is considered feasible if the following two conditions are satisfied:

1. Both partner MNOs earn more profits if they partner with the MO-MVNO than if they do not.
2. The MO-MVNO is profitable.

Condition #1 yields the lower bound of feasible wholesale prices. Setting $\pi_{ik} \geq \pi_{i0}$ yields:

$$p_{ik}^w \geq p^{rnm} \frac{v_{i0}^r - v_{ik}^r}{v_{ik}^w} - \frac{k^S (B_{i0} - B_{ik}) A + k^T (t_{i0} - t_{ik})}{v_{ik}^w} \tag{14}$$

The right-hand side of (11) is a function of an operator's retail and wholesale traffic volumes with and without MNA, and the tower and spectrum resources it deploys with and without MNA. As our results will show, MNA can change how traffic volume and the associated revenue are distributed between partner MNOs, depending on the relative amounts of tower and spectrum resource as well as choices among resource allocation schemes and network selection algorithms. These parameters need to be considered for an MNA agreement, because they can influence whether and how much one partner MNO gains or loses revenue, and the extent to which that partner MNO can save on costs.

Solving Condition #2 yields the upper bound of feasible wholesale prices. Set $\pi_{MO-MVNO,\mathrm{k}} > 0$ yields:

$$\mathrm{p}_{\mathrm{Ak}}^{\mathrm{w}} < -\frac{\mathrm{v}_{\mathrm{Zk}}^{\mathrm{w}}}{\mathrm{v}_{\mathrm{Ak}}^{\mathrm{w}}}\mathrm{p}_{\mathrm{Zk}}^{\mathrm{w}} + \mathrm{p}^{\mathrm{rnm}}\left(1 + \frac{\mathrm{v}_{\mathrm{Zk}}^{\mathrm{w}}}{\mathrm{v}_{\mathrm{Ak}}^{\mathrm{w}}}\right) \tag{15}$$

We define a feasible region to be the collection of feasible wholesale prices for a given distribution of traffic volume and resource investment between partner MNOs.

Table 1 Summary of Variables

i	MNO index
j	UE index
k	$\mathrm{k} = 0$ indicates MNA is not adopted. $\mathrm{k} = 1$ indicates MNA is adopted.
v_{ik}	Total traffic volume of MNO i (bits)
$\mathrm{v}_{\mathrm{ik}}^{\mathrm{r}}$	Retail traffic volume of MNO i (bits)
$\mathrm{v}_{\mathrm{ik}}^{\mathrm{w}}$	Wholesale traffic volume of MNO i (bits)
A	Area of coverage, same for all MNOs (km^2)

B_{ik}	Spectrum holdings of MNO i (Hz)
t_{ik}	Number of towers of MNO i in the area A
k^T	Constant for tower cost ($\frac{\$}{\text{tower}\cdot\text{sec}}$)
k^S	Constant for spectrum cost ($\frac{\$}{\text{Hz}\cdot\text{sec}\cdot\text{km}^2}$)
k^M	Share of retail revenue used for marketing and other retail operating expenses
p^r	Price per unit of retail traffic carried ($\frac{\$}{\text{bit}}$)
p^w_{ik}	Price per unit of wholesale traffic carried ($\frac{\$}{\text{bit}}$)

5.2.3 Parameters for Numerical Evaluation

To calculate the amount of traffic carried, we assume that the throughput at the peak hour is 80% of the total network capacity and that the throughput averaged over the busy hour is five times the throughput averaged over all hours [41]. Revenue per unit of time from data traffic is then capacity*(peak usage rate)/(peak to average ratio)*(price per unit of traffic). We set retail market price $p^r = \$10/\text{GB}$ based on current Google Fi pricing [42].

We set spectrum price at $\frac{\$2}{\text{MHz}*\text{Pop}}$, based on recent results from FCC auctions on low-band spectrum [43]. Population density is set at $\frac{2000\ \text{Pop}}{\text{km}^2}$, similar to that of a medium-sized city like Pittsburgh. We use a discount rate of 7%. Under these assumptions, the spectrum cost coefficient $k^S = \frac{\$2}{\text{MHz}*\text{Pop}} * \frac{2000\ \text{Pop}}{\text{km}^2} * 7\% = \frac{\$280}{\text{MHz}*\text{km}^2*\text{Year}}$.

We use the following estimates for the major components of infrastructure cost. Land lease, amortization of the tower construction, maintenance and utility for a tower site are estimated to cost $6000 per month, based on the average rent of a tower company and the average number of tenants per site [44]. Base station electronics per cell site with 3 sectors are estimated to cost about $1600 per month, based on $100,000 purchase price [45], 5-year useful life and $0 salvage value. Backhaul is estimated to cost $3000 per month per cell site with 3 sectors, based on the per-customer revenue of fiber infrastructure companies [44], [46]. Lastly, we assume an additional 10% markup on RAN and backhaul costs to account for upgrading and maintaining the core network [47]. Since our model uses omnidirectional antennae, we divide these costs by 3 to get infrastructure cost per cell in our model.

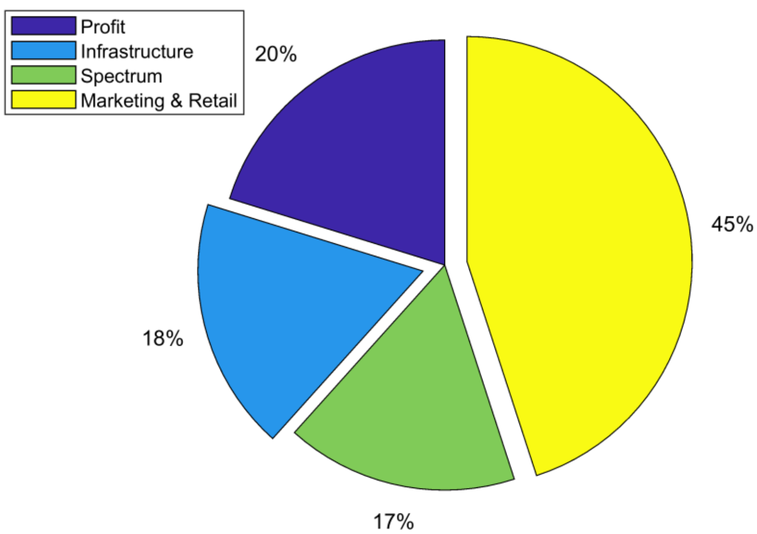

Figure 23 Breakdown of an MNO's costs and profits without MNA. (16 towers; 60 MHz spectrum; "Balanced" resource allocation; log-normal fading with 10 dB standard deviation)

Finally, we set marketing cost coefficient $k^M = 45\%$ based on T-Mobile's recent income statement, where selling, general and administrative expenses have consistently been about 45% of retail revenue during 2016-2018.

Under these assumptions, the breakdown of the costs and profits for an MNO with 16 towers and 60 MHz of spectrum without MNA is shown in Figure 23.

5.3 Results

This section presents our main results. Section IV.A quantifies how much MNA can reduce the cost to provide cellular data services, and how the cost savings vary with various technical parameters. We characterize the wholesale prices that would make all operators willing to participate in MNA in Section 5.3.2. Section 5.3.3 describes the implications of MNA on consumer welfare and operators' profitability. Section 5.3.4 explains why, in addition to financial considerations, network operators should also pay attention to the QoS experienced by subscribers of different operators when drafting an MNA agreement that benefit all stakeholders.

5.3.1 Cost-Efficiency: MNA vs. Colocation

Complete MNA (i.e. all UEs are capable of MNA) without colocation is likely more cost-effective than complete colocation without MNA. Figure 24 shows RAN cost (tower, base station electronics & backhaul) per unit capacity produced, as a function of the extent of colocation. Given the numerical assumptions in Section III.E, colocation alone can reduce the RAN cost per unit capacity by up to 28%, while MNA alone can improve the same metric by up to 33%. All things considered, complete MNA would be about 10% cheaper in RAN cost than complete colocation for the same network capacity.

Without MNA, additional colocation always improves cost efficiency. However, in the presence of MNA, colocation becomes a tradeoff between capacity and cost. The optimal level of colocation (one that minimizes cost for a given network capacity) depends on how widely adopted MNA is. As we have seen in Figure 21, the rate at which colocation reduces capacity under MNA is not constant. If the rate at which additional colocation reduces cost is faster than the rate at which it reduces capacity, the net effect on cost efficiency would still be positive. The convex curves in Figure 24 are a few examples of this dynamic. When MNA is widely adopted, moderate levels of colocation can preserve the benefit of MNA while further reducing RAN cost.

Given our baseline numerical assumptions, the inflex point in terms of fraction of MNA-capable users is about half. Therefore, before MNA-capable users become the majority, it is probably more economical for MNOs to colocate base station as much as needed; as the fraction of MNA-capable users approaches 50%, MNOs may want to consider locations complementary to existing infrastructure when building new cell sites.

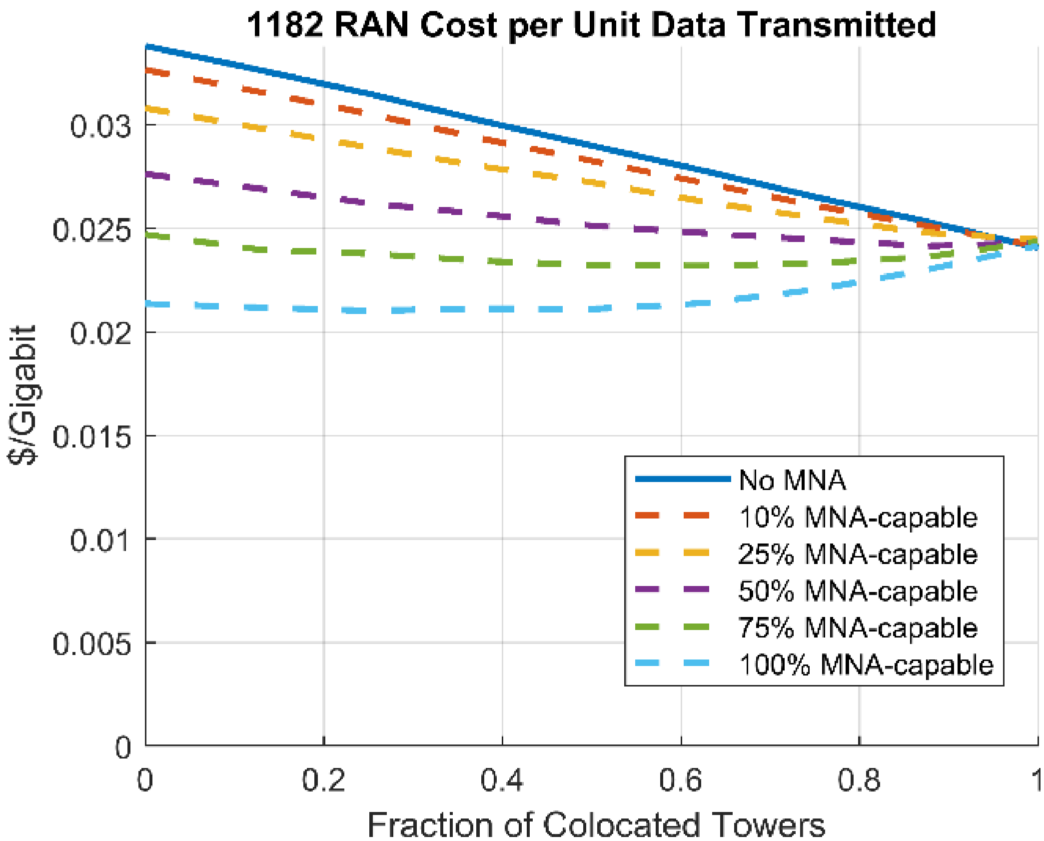

Figure 24 RAN cost per second per unit capacity produced vs. extent of colocation. (Balanced resource allocation; max-SINR selection algorithm; MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

5.3.2 Feasible Wholesale Prices

The fact that MNA reduces the aggregate investment on tower and spectrum resources that produces a given network capacity does not guarantee the willingness of an individual operator to participate in MNA. Partner MNOs require wholesale prices high enough so that they can benefit from partnering with an MO-MVNO, and the MO-MVNO requires wholesale prices low enough so that it, too, can profit. While the exact prices that all operators agree on are the result of negotiations, we can find the range of wholesale prices that would incentivize each operator to participate in MNA. In this section, we discuss the lower and upper bounds of feasible wholesale prices, and how these prices are influenced by the relative amount of tower and spectrum resources between partner MNOs, the share of traffic carried by each partner MNO and the network selection algorithm used by MNA-capable UEs.

If the traffic share between MNOs does not change with MNA, both partner MNOs will be willing to accept a wholesale price below the retail price net of marketing cost. Assuming that the total traffic volume carried with MNA is the same as that without MNA, when an MNO carries the same share of traffic with MNA as it does without MNA, we have:

$$v_{ik}^{w} + v_{ik}^{r} = v_{i0}^{r} \tag{16}$$

The minimum feasible wholesale price (14) can then be simplified to:

$$p_{ik}^{w} \geq p_{0}^{rnm} - \frac{k^{S}(B_{i0}-B_{ik})A+k^{T}(t_{i0}-t_{ik})}{v_{ik}^{w}} \tag{17}$$

That is, absent any change in revenue, as long as a partner MNO can produce the same capacity with fewer tower and spectrum resources with MNA as it does without MNA (i.e. $B_{i0} > B_{ik}$ and $t_{i0} > t_{ik}$), that MNO should be able to offer the MO-MVNO a discount off of the retail price net of marketing cost and still make the same profits as it would without MNA.

There are realistic scenarios where traffic share does not change with MNA. One example is when partner MNOs have comparable tower density and spectrum bandwidth between them with or without MNA, absent any mechanism that purposely steers more traffic onto one MNO or the other. The solid triangle in Fig. 3 shows the range of feasible wholesale prices for the scenario where an MO-MVNO partners with two MNOs that have equal tower density and spectrum bandwidth, with all UEs being MNA-capable, under the numerical assumptions in Section III.E. Here, price is expressed as a fraction of the retail price net of marketing cost per unit of traffic. The minimum feasible wholesale price would be lower if tower and spectrum costs represents a higher fraction of revenue. Under our baseline retail price and cost assumptions, infrastructure and spectrum costs represent 45% of revenue (Figure 23). In this case, both partner MNOs require only 85% of the retail price net of marketing cost (the lower left corner of the triangle) per unit of traffic from the MO-MVNO in order to maintain the same profit with MNA as that without MNA.

If the traffic share between MNOs does change with MNA, the MNO gaining traffic share will be willing to accept a wholesale price below the retail price net of marketing cost, but the one losing traffic share might demand a higher wholesale price, possibly close to or even above the retail price net of marketing cost in order to participate in MNA. This happens when a partner MNO loses so much traffic share with MNA that the savings on tower and spectrum costs are comparable to or even less than its lost retail revenue. Nevertheless, MNA can still benefit all operators in this case.

A significant change in traffic share can happen under realistic conditions. For example, partner MNOs may maintain a fixed division of investments on tower and spectrum resources with or without MNA. If that division is lopsided, traffic share could change considerably and that affects minimum feasible wholesale prices. Given that the relative tower density and spectrum bandwidth between partner MNOs with MNA is the same as that without MNA, if MNA-capable UEs attach to towers of higher expected data rate, the partner MNO with higher tower density and more spectrum bandwidth may carry a smaller share of all traffic with MNA than it does without MNA; if MNA-capable UEs instead attach to towers of higher SINR, the partner MNO with higher tower density and more spectrum bandwidth may carry a larger share of all traffic with MNA than it does without MNA.

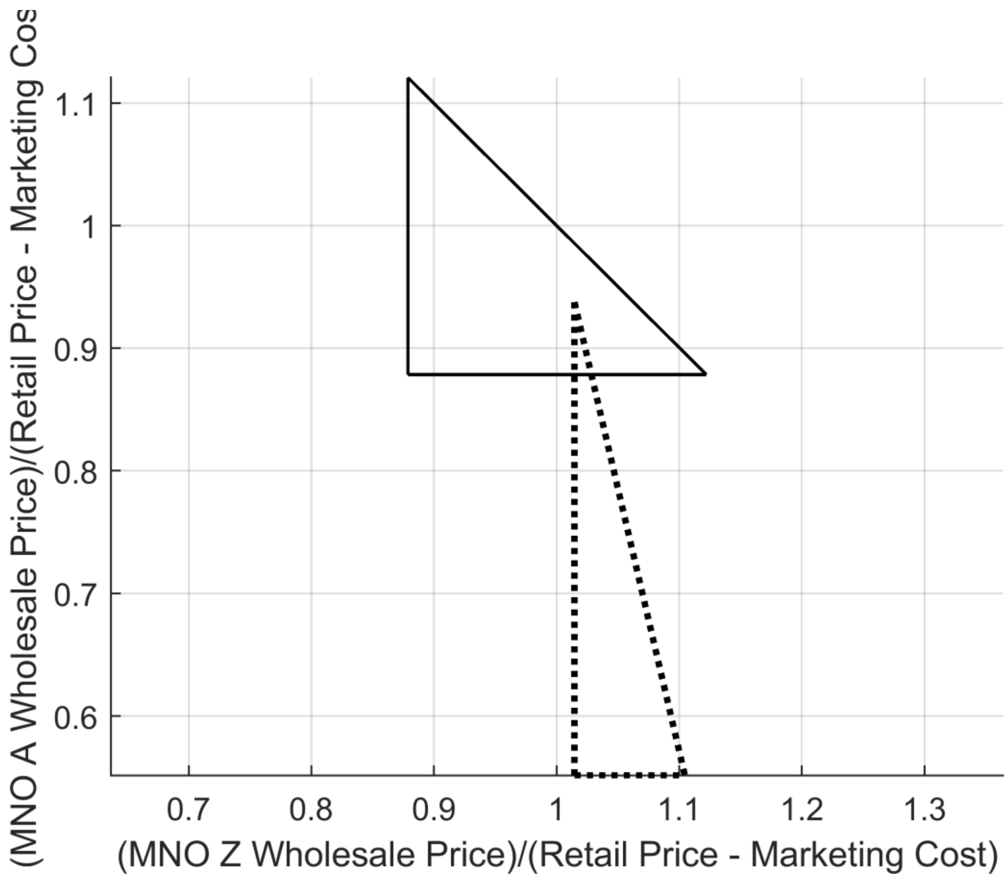

Figure 25 Feasible wholesale prices. Price is shown as a fraction of the retail price net of marketing cost per unit of traffic. Triangle with solid edges: when the two partner MNOs have equal tower density and equal spectrum bandwidth. Triangle with dotted edges: MNO Z has 3x the tower density and 3x spectrum bandwidth of MNO A. ("Balanced"; "Max-Throughput";100% MO-MVNO market share; log-normal fading with 10 dB standard deviation)

The dotted triangle in Figure 25 shows the range of feasible wholesale prices corresponding to a scenario where one partner MNO has three times the tower density and three times the spectrum bandwidth of the other partner MNO both with and without MNA, under the numerical assumptions in Section III.E. It was assumed that all UEs are MNA-capable and they attach to towers of higher expected data rate. The MNO with more resources in this example must charge almost the retail price net of marketing cost to maintain the same profit as it would earn without MNA. Even if a partner MNO needs to charge a wholesale price higher than the retail price net of marketing cost, for example, when infrastructure and spectrum costs are relatively small compared to other costs, that does not mean such a wholesale price renders MNA undesirable – all three operators are still better off with MNA than without MNA – but in this scenario the wholesale prices that can distribute the cost savings in a way that benefits all three operators may look somewhat unusual.

There is no realistic scenario where neither MNO could offer a wholesale price that is lower than the retail price net of marketing cost. We prove this by contradiction. Assume, to the contrary, that both partner MNOs charge a wholesale price equal to or higher than the retail price net of marketing cost and make the same profits as they do without MNA. The total profits made between partner MNOs are equal to (retail price – marketing cost per unit of traffic)*(total retail traffic) + (wholesale price 1)*(wholesale traffic 1) + (wholesale price 2)*(wholesale traffic 2) – (total tower + spectrum cost). We have shown that MNA always reduces the amount of tower and spectrum resources needed to achieve a given capacity, so total tower + spectrum cost must decrease. We have also assumed that total retail traffic + wholesale traffic 1 + wholesale traffic 2 remain the same. Therefore, if both partner MNOs charge a wholesale price equal to or higher than the retail price net of marketing cost, the MNOs must have made higher profits collectively. That means at least one partner MNO makes more profit than it would without MNA, which contradicts the initial assumption.

An MO-MVNO is more sensitive to the wholesale price charged by one MNO than to that of the other when its traffic is distributed unequally between partner MNOs. In the case of two partner MNOs, as shown in Figure 25, the line on which the MO-MVNO makes zero profit makes up the hypotenuse of the feasible region. The slope of the hypotenuse is equal to the negative ratio of the traffic volumes carried by the two partner MNOs (12). If the MNO carrying more traffic raises its wholesale price by one unit, for the MO-MVNO to maintain the same profit the MNO carrying less traffic must lower its wholesale price by more than one unit.

5.3.3 Consumer Welfare and Operator Profitability

As MNA reduces the investment needed on infrastructure and spectrum resources to provide a given capacity, consumers could enjoy lower prices, operators could make more profits, or both. In this section, we discuss the potential impact of MNA on consumers, MO-MVNOs and partner MNOs by examining how an operator's profit could change if it were able to drive a hard bargain and capture all the cost savings from MNA, and how the retail price could change if operators pass all the cost savings on to consumers.

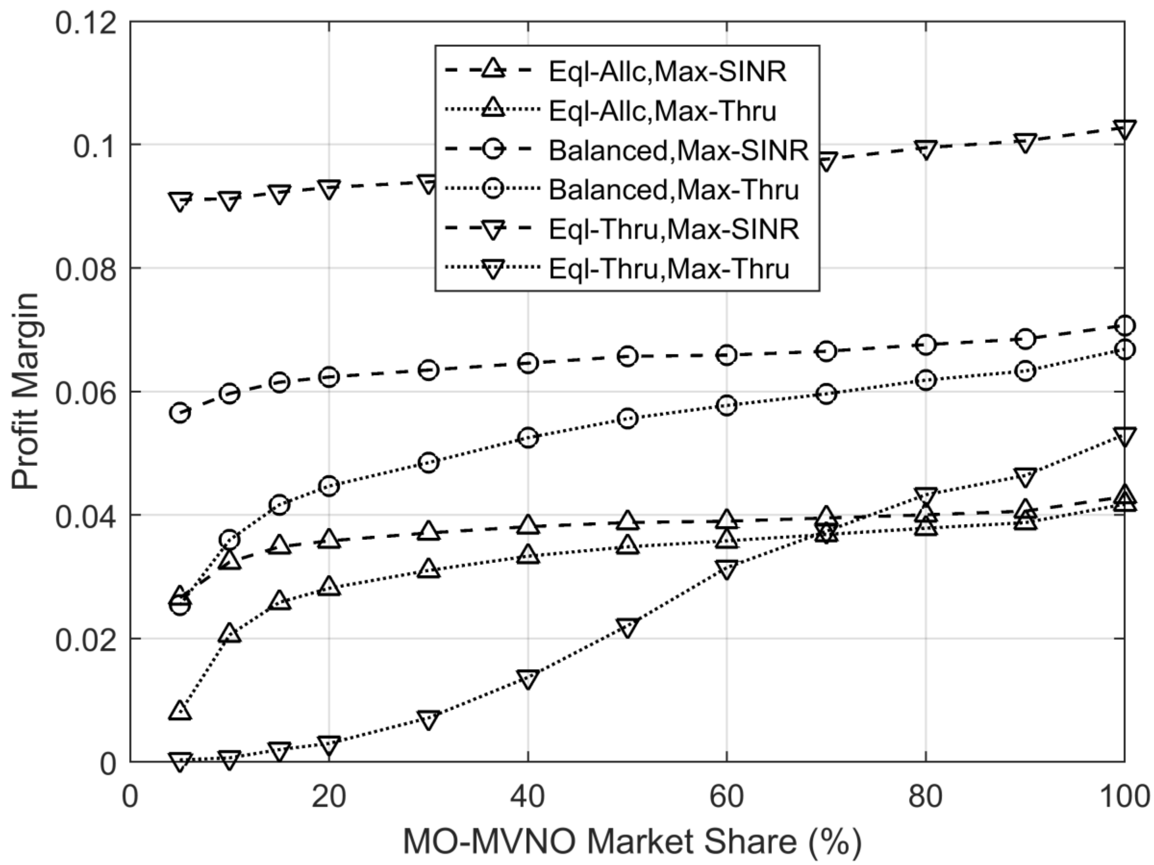

Figure 26 Potential MO-MVNO profit margin if MNOs passed all cost savings on to it. (MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

When partner MNOs charge their respective minimum feasible wholesale prices, the MO-MVNO can enjoy a healthy profit margin regardless of its market share, provided that it uses an appropriate network selection algorithm. As shown in Figure 26, when MO-MVNO subscribers attach to towers of higher SINR, the best-case profit margin for the MO-MVNO starts at a decent 6% at low market share and increases gradually to 7% as its market share increases, given "Balanced" resource allocation and the numerical assumptions in III.E. Because the amount of resources saved from MNA is largely commensurate with the fraction of MNA-capable UEs, as we saw in Figure 20, the profit margin, which measures the ratio of the value of the resources saved to revenue, does not vary with MO-MVNO market share as quickly as the amount of resources saved does.

To put the profit margin in perspective, imagine a hypothetical business which, once it reaches saturation market share, earns a profit margin m in each time period in perpetuity, i.e. $\frac{R-C}{R} = m$ where R is the revenue earned in each time period and C is the cost incurred in each time period. Further assume that the business reaches its saturation market share in time period $\tau + 1$, and for the first τ time periods, the business only incurs only marketing and retail cost $k_m R$ without earning any revenue[2]. The cash flow is shown in Table 2.

Table 2 Cash flow of hypothetical business

Time	0	1	2	…	τ	$\tau + 1$	$\tau + 2$	…
Cash flow	0	$-k_m R$	$-k_m R$	$-k_m R$	$-k_m R$	$R - C$	$R - C$	$R - C$

The net present value (NPV) of this business with the said cash flow is then

$$NPV_{business} = \frac{R}{i}\left(\frac{k_m + m}{(1+i)^{\tau}} - k_m\right) \tag{18}$$

Derivations are given in Appendix C. We want to find out for how long this hypothetical business can endure zero revenue and still turn in a positive NPV, so we set $NPV_{business} > 0$ and derive the range of τ.

$$\frac{R}{i}\left(\frac{k_m + m}{(1+i)^{\tau}} - k_m\right) > 0$$

[2] Real businesses probably go through a ramp-up period and gradually pick up market share, as opposed to the step function-like transition to equilibrium as we assume here.

Assuming positive discount rate and revenue, we have $\frac{k_m+m}{(1+i)^\tau} - k_m > 0$, which can be rearranged into:

$$\tau < \log_{1+i} \frac{k_m+m}{k_m} \tag{19}$$

For $i = 7\%$, $k_m = 0.45$ and $m = 6\%$, $\tau < 1.85$ years or about 22 months. That is, the business can tolerate zero revenue for about two years if it can earn a profit margin of 6% every year after that. However, with i and k_m unchanged but $m = 1\%$, $\tau < 0.32$ or less than 4 months. Once again, it is important for an MO-MVNO to assign users based on higher SINR rather than higher expected data rate, or otherwise

As a reference, it takes up to one year to launch an MVNO, and four to six years for investment payback [48]; the major facility-based MNOs have a profit margin of around 10%. All things considered, an MO-MVNO that depends solely on the cost-savings from MNA can be a reasonably attractive business. In practice, an MO-MVNO may not be able to capture all the cost savings for itself; on the other hand, it may also tap into other revenue opportunities besides selling cellular data, but they are outside the scope of this thesis.

Even if an MO-MVNO is unable to secure favorable wholesale prices, it can still be a viable business in terms of return-on-investment (ROI), and remain an additional choice of service providers for consumers. While a profit margin of 10% seems nothing out of the ordinary, the capital investment required for an MVNO is typically much lower than facilities-based MNOs. Therefore, cost savings that are otherwise minor for an MNO generate a sizable ROI for an MVNO which requires fewer assets to operate. Even if its market share is low, or if it is unable to obtain the most generous wholesale prices, an MO-MVNO should still be encouraged to enter the market.

MNA could bring meaningful reduction in the price of cellular data services if operators pass the cost savings on to consumers. At baseline, consumers could see a retail price reduction between 4% and 10% when all UEs are MNA-capable, as a result of resource savings that vary from 11% to 22% depending on the resource allocation scheme and network selection algorithm. The more expensive tower and spectrum are relative to other costs, the more MNA can lower the prices faced by consumers, increase operators' profits, or both.

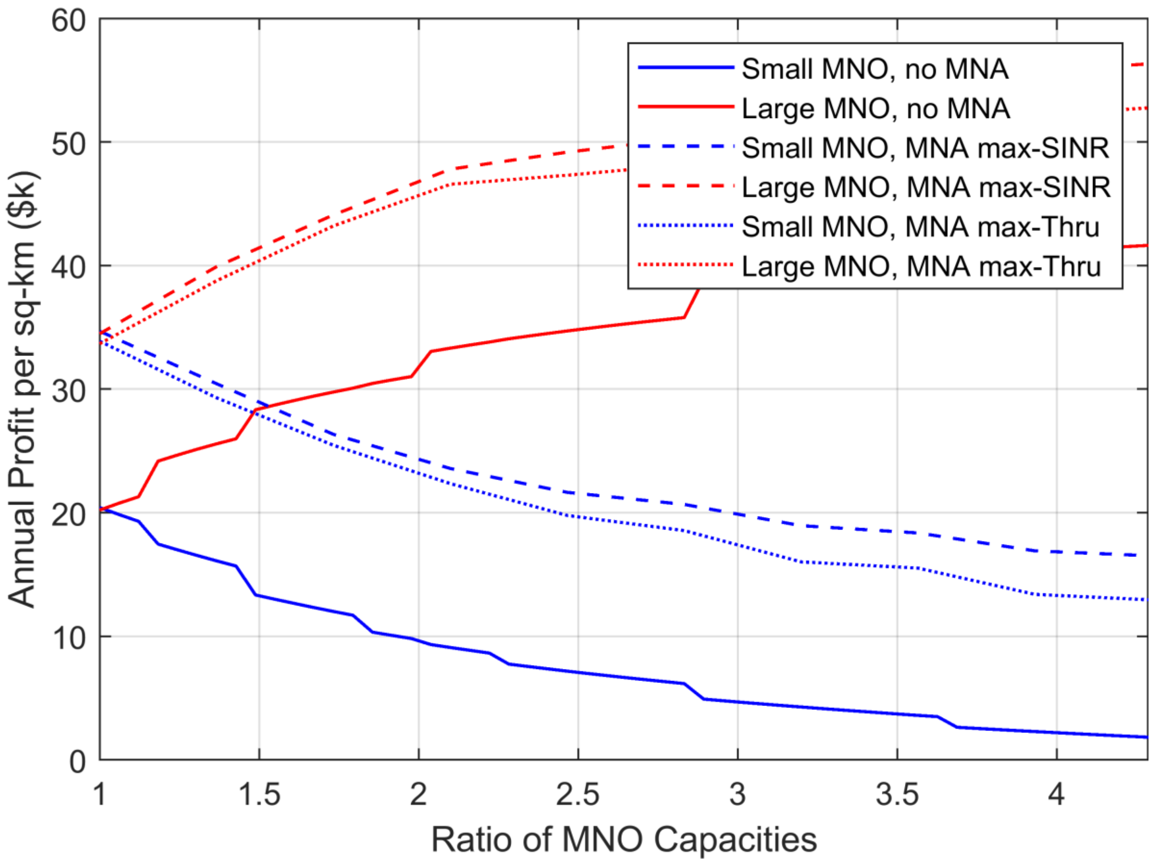

Figure 27 Best-case MNO profit per square kilometer vs. MNO capacity ratio. (Balanced resource allocation; all UEs are MNA-capable; each MNO carries the same traffic volume with MNA as it does without MNA; log-normal fading with 10 dB standard deviation)

For partner MNOs, MNA can be an opportunity for tremendous profit growth. Figure 27 shows each MNO's profit without MNA, and their maximum possible profit with MNA (if that MNO were to capture all the cost savings) for partner MNOs of various sizes. At baseline, for MNOs that both have 16 towers and 60 MHz of spectrum, each makes an annual profit of $20k per square kilometer without MNA. When all UEs are MNA-capable, either MNO can see annual profit increase to $34k per square kilometer if UEs attach to towers of higher expected data rate, or $35k if UEs instead attach to towers of higher SINR.

MNA could have a relatively greater impact on a smaller MNO (lower tower density and lower spectrum bandwidth) than it does for an MNO of larger scale. Absent any change to the combined traffic volume carried between partner MNOs, the maximum possible additional profit is the same for all partner MNOs and is equal to the saved investments in aggregate tower and spectrum resources as a result of MNA. But those same additional profits are more consequential for an MNO with smaller profit margins without MNA. As shown in Figure 27, due to cellular economies of scale [39], an MNO with higher tower density and higher spectrum bandwidth is more profitable than an MNO with lower tower density and lower spectrum bandwidth. While MNA could bring a healthy uplift in profits for either MNO, the cost savings from MNA would be more significant for the relatively smaller MNO, or even turn around an MNO that is unprofitable without MNA.

5.3.4 Balancing Quality of Service

While MNA can greatly reduce the tower and spectrum resources needed to provide a given capacity and substantially increase a partner MNO’s profits, it may put that partner MNO at a disadvantage when it comes to the QoS experienced by its subscribers compared to MO-MVNO subscribers.

It was shown previously that, in the absence of resource allocation schemes that treat MO-MVNO subscribers differently from MNO subscribers, the two groups will get vastly different

QoS because MNA-capable UEs are on average closer to a tower [8]. Another reason not previously reported is that MNA-capable UEs tend to end up in cells with fewer UEs if they choose towers of higher expected data rate. This does not happen when MNA-capable UEs use a selection algorithm that considers only SINR.

That means it is not enough to balance the QoS experienced by subscribers of different operators by balancing QoS in each cell. We experimented with a simplistic adjustment scheme that gives MNO subscribers proportionally more resources so that in each cell, the data rate averaged over MNO subscribers is equal to that averaged over MO-MVNO subscribers. Figure 28 shows the mean subscriber data rates across all cells with the said adjustment. When the fraction of MNA-capable UEs is low, such as when an MO-MVNO has just entered the market, even though in each cell MO-MVNO subscribers are constrained to the same mean data rate as that of MNO subscribers, over the entire networks they can still have a slight advantage on the order of 10% over MNO subscribers at baseline. If MO-MVNO subscribers attach to towers of higher SINR, there was minimal inequality in the mean subscriber data rate. This phenomenon disappears at high fraction of MNA-capable UEs, when the load is likely more even among different cells.

MNA agreements should take QoS into account, and there are a few ways. QoS disparity can be mitigated with an appropriate resource allocation scheme or selection algorithm, justified with financial compensation, or a combination of both. For example, a partner MNO may tolerate a QoS for its retail customers that is inferior to the QoS for MVNO customers if the MNO can charge a higher wholesale price to match. Conversely, if the MNO wants to make its own retail service more attractive, a partner MNO may even "overcompensate" through technical means so that its subscribers get better QoS than MO-MVNO subscribers, in order to protect its own market share.

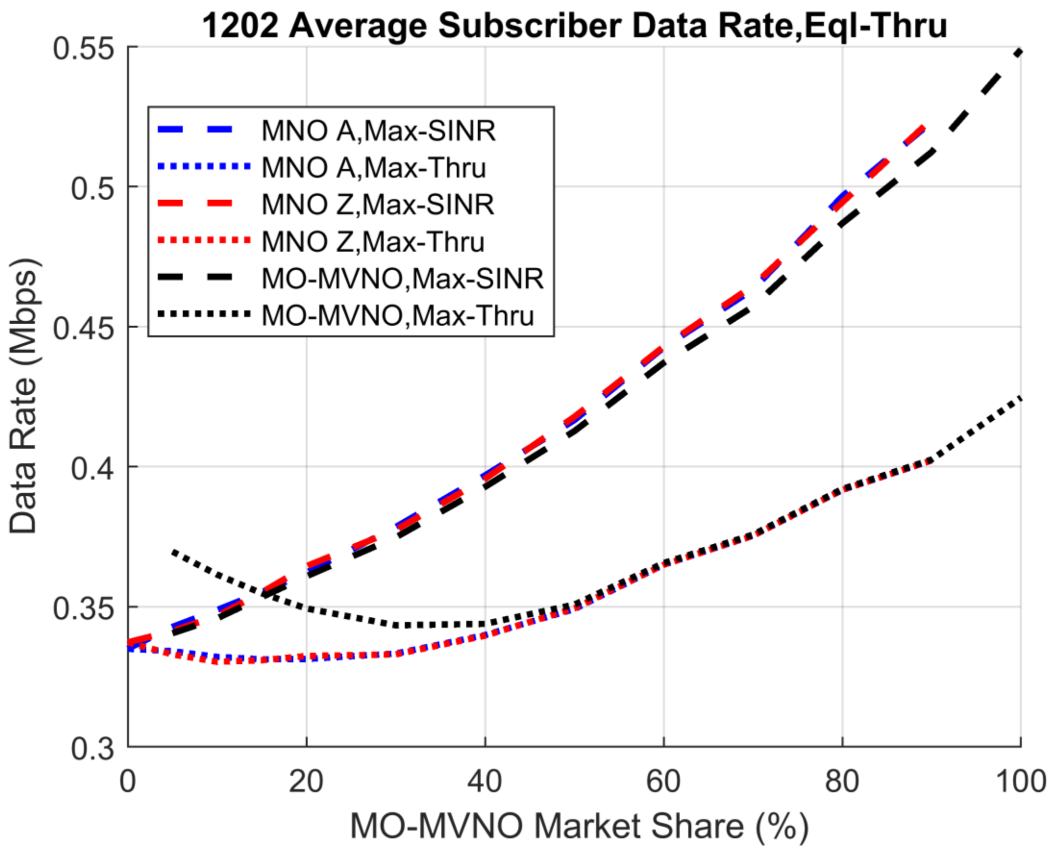

Figure 28 Average subscriber data rate by operator ("Equal-Throughput" resource allocation; MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

5.4 Conclusions

The benefit of MNA is robust against moderate levels of colocation. In fact, colocation and MNA can be deployed together to achieve even greater cost savings. Before MNA-capable users reach a critical mass (about half given our baseline assumptions), more colocation always improves cost-effectiveness. Given MNA, the extent of location that minimizes cost is neither 100% nor 0; it varies with the fraction of users capable of MNA. The higher the fraction of MNA-capable users, the less the optimal extent of colocation because there are more users who can utilize the spatial diversity. While MNA thrives on spatial diversity, colocation does not necessarily negate the benefit of MNA.

Because MNA-capable devices use spectrum more efficiently, MNOs can potentially entice an MO-MVNO with discounted wholesale prices. The lowest wholesale price it can offer, or, equivalently, the minimum wholesale price needed to incentivize an MNO to partner with an MO-MVNO, is closely related to whether, and, if so, how, the share of total traffic carried by that MNO changes. There are realistic scenarios where all partner MNOs are willing to offer comparable wholesale prices. But, if an MNO loses traffic share substantially, the value of its saved infrastructure and spectrum resources may not make up for its lost retail revenue; that MNO may demand a wholesale price much higher than those of the other partner MNOs, possibly close to or even above the retail price net of marketing cost.

An MNO can set its wholesale rate just low enough to keep the MO-MVNO in business and pocket most cost savings itself. The highest wholesale price an MNO can charge depends on what the other partner MNO(s) charges, because an MO-MVNO's bottom line depends on the wholesale price and traffic volume of each of its partner MNOs. Theoretically, then, any one partner MNO can eat into the cost savings of all the other partner MNOs if it can drive a hard bargain; that MNO may be able to charge a wholesale price higher than the retail price net of marketing cost. In other words, an MO-MVNO can theoretically lose money on some (but not all) of its wholesale contracts and still come out profitable. Note that even in cases where an MO-MVNO loses money on some wholesale contracts, all participants are still better off adopting MNA than not; it is just the distribution of MNA's benefits may be imbalanced. While not the only way to implement MNA, an MO-MVNO is the most prevalent form of MNA in the real world today and it can be a reasonably profitable business. Given our baseline revenue and cost assumptions, MNOs' cost savings as a result of MNA amount to up to 10% profit margin for the MO-MVNO if all passed on to the MO-MVNO. This allows an MO-MVNO to endure up to 3 years of only incurring marketing and retail costs without revenue and still turn in a positive NPV for the business if it can earn a profit of 10% in perpetuity afterwards, which should be enough time to prepare and launch an MO-MVNO and to build up a customer base.

To maximize profit potential, an MO-MVNO should let its subscribers attach to base stations that provide higher SINR. Using a network selection algorithm that picks higher expected data rate would significantly reduce potential profitability, especially when the MO-MVNO has a small market share. As an MO-MVNO gains market share, it will be able to demand better wholesale prices from partner MNOs, because the resource savings from MNA increase faster than linearly with rising fraction of MNA-capable users. If all cost savings are passed on to consumers, an MO-MVNO can deliver meaningful savings on the order of 5-10% to its subscribers. It can do so even at a small market share, given an appropriate network selection algorithm.

The cost savings from MNA could be more consequential for an MNO with fewer tower and spectrum resources than its competitors because it benefits less from economies of scale, which would make that MNO more inclined to participate in MNA. But exactly how the cost savings are distributed between all participants comes down to negotiation. For any network operator to decide whether to adopt MNA, it involves complex considerations concerning not only business decisions like investment on infrastructure and spectrum resources, and wholesale pricing, but also technical parameters like network selection algorithms and resource allocation schemes. Besides economics, there could be regulatory and strategic headwinds to MNA's uptake. Regulators may be wary of MNA for the risk of reduced competition when MNOs rely on each other to serve consumers. From an operator's perspective, providing its core service by relying on a competitor may sound like a discomforting proposition. Notwithstanding these caveats, the infrastructure and spectrum savings resulting from MNA provide a significant financial incentive for its consideration.

6 Migration and Multihoming Support for MNA-Capable Devices

6.1 Background

In previous chapters, we have shown that the way an MNA-capable cell phone chooses an MNO could have profound implications on the overall efficiency of cellular networks. We demonstrated that the best MNO for a cell phone depends on its location, which determines the distances between the cell phone and the base stations of different MNOs. Because users are mobile, their location may change during the process of a data transfer session or a call. As a result, the best MNO as determined at the beginning of a session may no longer be the best one later. It may even become unavailable if the user moves outside its coverage area. While a cell phone may be quick to attach to a new MNO, it typically receives a different Internet locator, e.g. IP address in the case of TCP/IP networks, when it switches MNOs. A change in the locator can disrupt ongoing data transfer sessions between a cell phone and its correspondent if the cell phone is no longer reachable at the previous locator while the correspondents do not recognize the new locator. Therefore, a protocol is needed that transfers ongoing communications from using one Internet locator to using another, a process which we call migration.

A migration protocol is significant for multihoming. Today's mobile devices, such as cell phones, and mobile networks, such as Internet-connected cars, are often equipped with multiple network interfaces, meaning that they can multihome. For example, a car may transmit to a roadside unit (RSU) using the Dedicated Short Range Communications (DSRC) protocol [49] while simultaneously transmitting to a cellular base station. Using a migration protocol, multihomed devices and networks would be able to move different application traffic to different access networks on the fly, either to balance load or to match an application to the network that best meets its QoS requirement. Migration may also be warranted if a newly available network operator provides better QoS and/or lower cost for an ongoing session, even if the previous network is still accessible.

Migration can be performed at different granularities, namely flow-level, host-level, and network-level. Flow migration involves a change in the locators associated with an individual flow, which is useful when one would like to migrate some flows but not others. Host migration changes the locator associated with a host, which in turn can change the locators of all flows afforded by the host, making it possible to migrate all flows at once. One can also perform migration at the network level by changing the network locator, e.g. network prefix.

Today's state-of-the-art Internet architecture, TCP/IPv6, supports migration poorly. Flow-level migration support is limited to multi-path TCP (MPTCP) [50] and the Stream Control Transmission Protocol (SCTP) [51]. The Mobile IPv6 (MIPv6) [52] host migration protocol is complex and depends on a fixed mobility anchor point. Network migration in the form of the Network Mobility (NEMO) Basic Support Protocol (BSP) [53] suffers from long packet propagation delays.

The eXpressive Internet Architecture (XIA) [54] is a next-generation Internet architecture that features expressiveness, intrinsic security, and the use of directed acyclic graphs (DAGs) as locators. XIA allows users to express their "intent" to the network, so the ultimate endpoint of a communication is explicit. Intrinsic security ties an identifier to the public key of the corresponding entity, which facilitates authentication of migration signaling messages. DAG addressing creates flexibility for the network in fulfilling an intent, and provides redundancy for multihomed devices and networks.

In the context of XIA, we have designed a new migration that has the following properties:

- Allows fast migration for environments where connectivity changes frequently.
- Supports flow-level migration. Flow migration allows fine-grained load balancing for multihomed devices and networks, and allows an application to always use its preferred access network.
- A successful migration results in data traffic taking a direct path between the endpoints without routing through an off-path intermediary.
- Provides resiliency for multihomed devices and networks, with low overhead.

We compared it against its TCP/IP counterparts with respect to handoff performance and security, as well as support for multihoming use cases.

This chapter is structured as follows. Section II reviews migration support in IPv6. Section III – V are the main contributions of this paper. Section III introduces XIA and the XIA Migration Protocol. Section IV presents a comparative evaluation of the XIA Migration Protocol and its IPv6 counterparts with respect to handoff performance and security. Section V discusses multihoming. We conclude in Section VI.

6.2 Migration Support in IPv6

TCP/IPv6 features Mobile IPv6 (MIPv6) [55] for host migration, Network Mobility (NEMO) Basic Support Protocol (BSP) [56] for network migration, and Multipath TCP (MPTCP) [57] and Stream Control Transmission Protocol (SCTP) [51] for flow migration. MIPv6 is further divided into three variants.

6.2.1 Host Migration with Mobile IPv6

6.2.1.1 Overview

MIPv6 uses a special router called "home agent" to maintain the mapping between a mobile host's "Home Address" and its current IP address acquired from a visited network. This allows a mobile host to appear to retain the same IP address as it changes point of attachment to the Internet, and thereby preserves session continuity with correspondents. One major weakness of basic MIPv6 is triangular routing, where all traffic to and from a mobile host traverses a home agent that might be topologically far from the actual endpoints. An enhancement scheme called Route Optimization [55] is introduced to circumvent triangular routing. However, Route Optimization brings with it additional signaling overhead and handoff latency. A further enhancement called Enhanced Route Optimization [58] makes use of cryptographically generated addresses (CGAs) [59]. The additional security provided by CGAs simplifies the authentication of signaling messages between a mobile host and its correspondent, and thereby reduces the signaling overhead due to Route Optimization.

6.2.1.2 Base Mobile IPv6

IPv6 [60] does not specifically support mobility. When a mobile node (MN) leaves an access network and joins a new one, it is typically assigned a new IP address. Existing communications cannot continue because packets would be routed to the previous IP address where the MN no longer resides. The core idea of MIPv6 is to allow a MN to change its point of attachment to the Internet while retaining the same IP address. In this way, mobility is transparent to the correspondent node (CN) and to the transport and application layer.

The key enabler of transparent mobility is a component called the home agent (HA). When a MN joins a foreign network, it obtains a care-of address (CoA) from the foreign network. To perform a migration, the MN sends a Binding Update (BU) message to the HA to register the CoA. The CN is unaware of the MN's address change and continues to send packets to the MN's home address (HoA). The home agent intercepts these packets and tunnels them to the CoA, which is the MN's actual Internet location.

Traffic originating at the MN uses "reverse tunneling". The MN encapsulates an outgoing packet inside a new IPv6 packet with the source address set to the CoA and the destination address set to the home agent's address. The home agent decapsulates the packet and forwards it to the CN.

Bidirectional Tunneling has a major disadvantage, referred to as "triangular routing", where all traffic is forced to detour through an anchor point, i.e. the HA, instead of going directly between the MN and CN, thereby adding packet propagation delay.

6.2.1.3 Mobile IPv6 Route Optimization

Route Optimization in MIPv6 eliminates triangular routing. In Route Optimization (RO) mode, the MN sends a BU message to both the CN and HA. Now that the CN recognizes the MN's new CoA, the MN can directly communicate with the CN using its CoA. Packets no longer need to be encapsulated and traverse a HA that might be topologically far from the MN and/or CN, which reduces both packet propagation delay and header overhead.

The challenge with employing RO is to authenticate the signaling messages between the MN and CN. For example, an attacker may spoof a BU message with a false CoA to deceive an unsuspecting CN, and the victim MN would be unable to receive packets from the CN until the false binding expires or is removed. Therefore, the Route Optimization protocol needs to provide the CN some assurance that the claimed HoA and CoA indeed belong to the sender of the BU. This authentication is accomplished through the "return routability procedure" [55], [61]. A successful return routability check gives the CN reasonable confidence (though not guarantee) that the sender of subsequent BUs is the legitimate owner of the claimed HoA and CoA.

6.2.1.4 Mobile IPv6 Enhanced Route Optimization

The return routability procedure introduces four additional one-way messages, which adds to the handoff latency. A further enhancement to MIPv6 aims to reduce this additional delay using cryptographically generated addresses (CGAs).

Under regular RO, the MN must perform a home test for each migration to obtain a home keygen token (HKGT), the knowledge of which implies HoA ownership. CGAs provide an alternative method to perform this authentication. If a MN's home address is a CGA, it means that the host ID portion of the address is the hash of the MN's public key. Under Enhanced RO, the CN encrypts the HKGT with the MN's public key. Knowledge of the HKGT implies knowledge of the MN's private key, which in turn implies HoA ownership because CGAs bind a MN's HoA to its public key. A HKGT exchanged in this way is considered "permanent" and can be used across multiple migration events. As a result, the home address test becomes unnecessary except during the very first migration to verify the validity of the subnet prefix. This saves one round trip time (RTT) between the MN and HA, plus one RTT between the HA and CN, and thus reduces handoff delay. In addition, because this method of authentication does not rely on the HA, mobility operation can continue even when the HA is temporarily unavailable. The MIPv6 Enhanced Route Optimization include additional optimizations, too, such as proactive home test, concurrent care-of test and parallel home and correspondent registration [58].

6.2.2 Network Migration with NEMO BSP

NEMO BSP is the IPv6 standards track protocol for network migration. It is an extension of MIPv6 and works in a similar fashion. A mobile network has a mobile router that serves as the default gateway for all hosts inside the mobile network. A home agent maintains the mapping between the current IP address of the mobile router and the prefix assigned to the mobile network. Due to the use of home agents, NEMO BSP suffers from the same routing inefficiencies as basic MIPv6. Route optimization for NEMO has not been standardized by the IETF, although there have been a number of proposals [62].

6.2.3 Flow Migration with MPTCP and SCTP

MPTCP is an experimental TCP extension that allows one end-to-end connection to make use of multiple paths. MPTCP introduces a shim layer between the TCP layer and application layer, and is designed to be transparent to both layers. To an application, MPTCP presents a standard TCP interface. To the network layer, a MPTCP connection looks like multiple independent, standard TCP flows.

Clients that support MPTCP can start a MPTCP connection in the same way they initialize a standard TCP flow, i.e. through a three-way handshake. MPTCP capability is signaled in the form of a TCP Option, and the last ACK in the initial handshake confirms MPTCP usage for the current connection. To take advantage of a second link, the client initiates a second subflow with another three-way handshake. The SYN packet in this second handshake includes an identifier, exchanged during the first handshake, of the MPTCP connection it wishes to "join". After a second sub-flow is established, the first can be terminated, mimicking a "soft" migration.

The connection being joined may or may not have any currently active subflow. To support "break-before-make" scenarios, as would be the case when a single-homed host needs to disconnect from one network before attaching to another, MPTCP introduces connection-level close [57]. MPTCP thus offers a complete mobility solution at the transport layer, and permits both soft and hard migration modes. However, it requires both endpoints to support MPTCP, and does not work with non-TCP communications.

The Stream Control Transmission Protocol (SCTP) [51] is an IETF Standards Track reliable data transfer protocol, and a complement to TCP at the transport layer. It natively supports multihoming, as a SCTP port can be associated with multiple IP addresses. Standard SCTP supports static multihoming, while a later extension enables dynamic address reconfiguration [63], which can be used to manage mobility [64]. However, similar to MPTCP, SCTP has not been widely used for migration.

6.3 Migration in XIA

We propose the XIA Migration Protocol for migration of mobile clients with active sessions. In this section, we start with a review of XIA core concepts, followed by detailed protocol operation and message formats. Whenever applicable, we use the same terminologies as in MIPv6 and NEMO BSP to describe the protocol. Lastly, we briefly discuss a rendezvous service for migration of mobile servers.

6.3.1 XIA Background

XIA refers to communicating entities as principals, and names them with eXpressive identifiers (XIDs) [65]. Four main XID types are host XID (HID), service XID (SID), content XID (CID) and network XID (NID). XIDs are cryptographically derived to achieve intrinsic security. HID, SID and NID are hashes of the public keys of the corresponding host, service and network, respectively. A CID is the hash of the content itself.

XIA separates identifier (XID) and locator (DAG), which facilitates mobility [66]. An XIA locator is a DAG of XIDs, as shown in Figure 29. The dot represents a conceptual source of the packet. The terminating XID represents the intent, or the end principal of a locator. We believe that mobile networks will be sufficiently numerous so that flat routing, where each mobile network is represented separately in the global routing table, is infeasible due to excessive burden on router hardware. Therefore, we assume that the locator for a mobile network uses "scoping" [65], as shown in Figure 29b. That is, an application running on Host A attached to a mobile network served by Internet service provider (ISP) #1 will include ISP1 in its locator, e.g. ·→NIDISP1→NIDmobile→HIDA→SID, as opposed to ·→NIDmobile→HIDA→SID. Continuing with this example, if the upstream access network becomes ISP2, the locator should be updated to ·→NIDISP2→NIDmobile→HIDA→SID.

Another feature of DAGs is fallbacks. Fallbacks are alternative routes to reach the intent, as shown by the dashed edge in Figure 29c. DAG-based addressing in XIA brings immediate benefit for multihomed devices and networks because fallbacks can be used to expose the availability of multiple upstream access networks. XIA allows multihomed hosts and networks to build redundancy through DAG-based addressing, without burdening routing tables or adding middleboxes such as NAT [67].

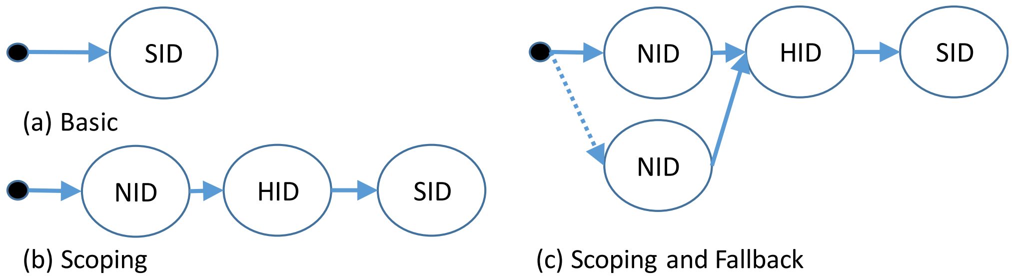

Figure 29 Examples of DAGs

6.3.2 XIA Migration Protocol

6.3.2.1 The need for a migration protocol

Hierarchical DAGs (network_prefix:endpoint_id) resemble the format of IP addresses. An endpoint thus has different DAGs when attached to different access networks. An endpoint may need to update its DAG over the course of a conversation with its corresponding endpoint, either because of mobility, i.e. when a device moves to a different access network, or for traffic engineering purposes, e.g. a multihomed device prioritizing one access network and not the other. A change in the DAG may disrupt ongoing communications, which is typically identified by source and destination identifier and locators. Along these lines, a migration protocol is needed to provide two functions: to certify the migrating endpoint's identity and to ensure that traffic can be rerouted to the new Internet location.

XIA inherently separates identifier and locator, allowing an endpoint to maintain the same identifier during a migration. MIPv6 achieves the same goal with two IP addresses, using home address as the identifier and care-of address as the locator. MIPv6 essentially manages the mapping between the two.

To ensure that the migrating endpoint can still receive packets from the ongoing conversation, MIPv6 either uses indirection (Bidirectional Tunneling mode) or sends the new locator directly to a correspondent (Route Optimization mode). The XIA Migration Protocol adopts a philosophy similar to Route Optimization mode and allows an endpoint to communicate its new DAG directly to its correspondent.

6.3.2.2 Message formats

The design of the XIA Migration Protocol is fundamentally similar to the approach by Snoeren and Balakrishnan [68]. We will use a mobile vehicular network in Figure 30 as an example to describe the protocol. The protocol consists of two messages, MIGRATE and MGRTACK (**migr**at**e** **ack**nowledgement). Assume that the endpoint wants to migrate from ISP1 to ISP2. The MIGRATE message is sent from Application A1 (SID_{A1}) to its correspondent Application B1 (SID_{B1}). The MGRTACK message is the response to a valid MIGRATE message, confirming a successful migration.

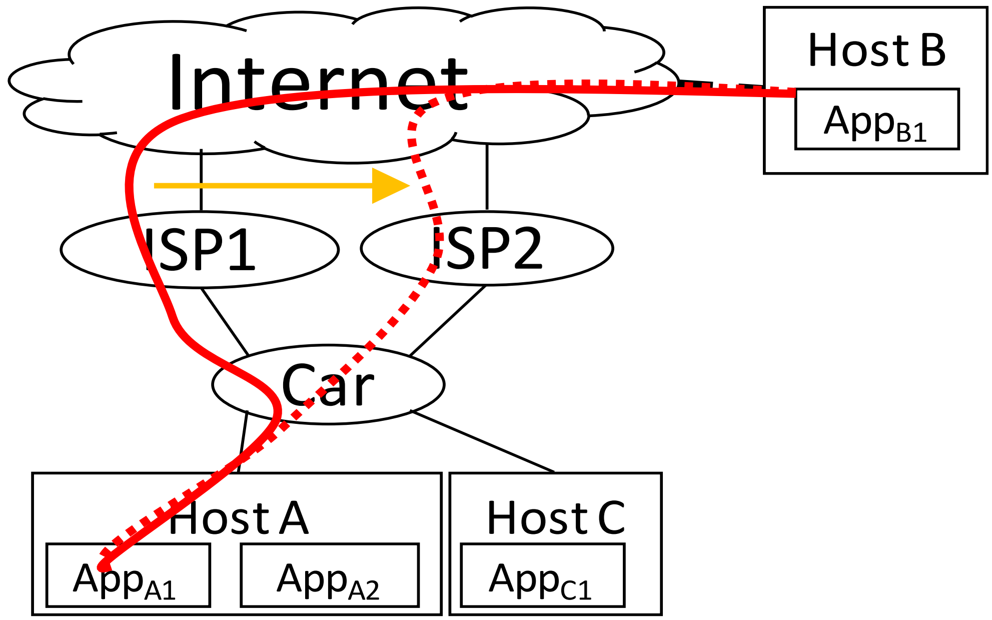

Figure 30 Example Topology for Protocol Specification

6.3.2.2.1 MIGRATE message

Note that the XIA Migration Protocol is a flow migration protocol. It lets an application notify its correspondent of a change in its locator. The MIGRATE message contains the old and new locator of Application A1, a sequence number (or, alternatively, a timestamp) and Application A1's public key. Upon receipt of a MIGRATE message, Application B first checks the sequence number to see if this is a replayed MIGRATE message. If the MIGRATE message is fresh, it then verifies the signature on the MIGRATE message. If the signature is correct, Application B1 sends back a MGRTACK message. If any of these checks fails, Application B1 should discard the MIGRATE message in question.

Table 3 Definition of MIGRATE Message

Destination	DAG_{B1}
Source	$\cdot\rightarrow NID_{ISP2} \rightarrow NID_{car} \rightarrow HID_{A} \rightarrow SID_{A1}$

Body	$\{DAG_{A1,old}, \mathrm{DAG_{A1,new}}, SEQ, K_{SID_{A1}}\}^{K^{-1}_{SID_{A1}}}$
Example	$\begin{Bmatrix} \cdot\rightarrow \mathrm{NID_{ISP1}} \rightarrow \mathrm{NID_{car}} \rightarrow \mathrm{HID_A} \rightarrow \mathrm{SID_{A1}}, \\ \cdot\rightarrow \mathrm{NID_{ISP2}} \rightarrow \mathrm{NID_{car}} \rightarrow \mathrm{HID_A} \rightarrow \mathrm{SID_{A1}}, \\ seq, \\ key \end{Bmatrix}^{K^{-1}_{SID_{A1}}}$

6.3.2.2.2 MGRTACK message

The MGRTACK message contains the Migration Acknowledgement flag indicating this is a confirmation of receipt of a MIGRATE message, an echo of the sequence number, and the Application B1's public key.

The protocol is designed to be robust against accidental loss of protocol messages and during overlapping migration. After sending a MIGRATE message, Application A1 starts the MGRTACK timer. If Application A1 does not receive a MGRTACK message at its new locator before timeout, it should assume that the MIGRATE message is lost and resend the MIGRATE message. The number of migration retries is up to the implementation.

It might occur that a MIGRATE message from an earlier handover arrives later than the MIGRATE message of a current handover. The sequence number in MIGRATE and MGRTACK messages enables endpoints to identify stale messages. The sequence number and signature also serve to defend against spoofing and replay attacks, which we discuss in Section IV.

Table 4 Definition of MGRTACK message

Destination	$\cdot\rightarrow \mathrm{NID_{ISP2}} \rightarrow \mathrm{NID_{car}} \rightarrow \mathrm{HID_A} \rightarrow \mathrm{SID_{A1}}$
Source	$\mathrm{DAG_{B1}}$

Body	$\{\text{MGRTACK flag, SEQ, } K_{SID_{B1}}\}^{K_{SID_{B1}}^{-1}}$
Example	$\{\text{MGRTACK flag, seq, key}\}^{K_{SID_{B1}}^{-1}}$

A time sequence diagram of the XIA Migration Protocol is shown in Figure 31.

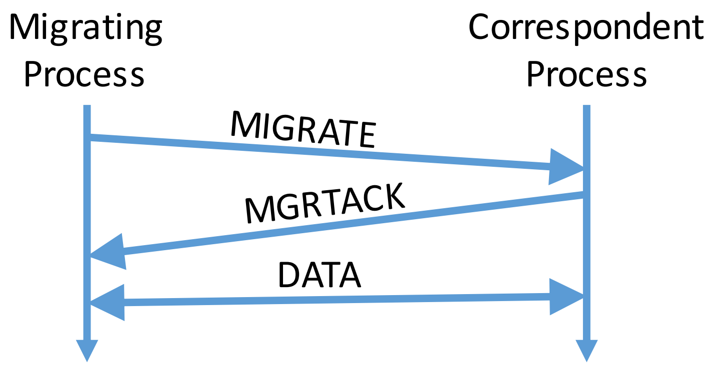

Figure 31 Time Sequence Diagram of XIA Migration Protocol

6.3.1 Interaction with X-Host Configuration Protocol

The XIA Migration Protocol provides a method for resuming a flow after migration to a new access network. There must be a complementary mechanism that notifies applications of the need to migrate in the first place. In XIA, X-Host Configuration Protocol (XHCP) [69] provides this functionality. An XHCP Server broadcasts a periodic beacon containing the default router DAG along with other configuration information for bootstrapping a newly-joined host. Additionally, an XHCP Server broadcasts a BEACON whenever there is a change in upstream connectivity to notify hosts of the need for a migration.

It is assumed that, within a router, there is a mechanism that notifies the XHCP Server process when the router detects a new DAG on any of its WAN-side interfaces. When that happens, the router XHCP Server broadcasts an XHCP BEACON with the new router DAG(s), regardless of its position in the default broadcast cycle. In the case of a mobile router, the mobile router acts as an XHCP Client to the XHCP Server in the upstream access network and as an XHCP Server to hosts within its own network.

An XHCP BEACON also contains a CONFIG_CHANGED flag. A router sets CONFIG_CHANGED = 0(1) if none (any) of the network configuration information has changed since the last BEACON. XHCP Servers are required to cache a copy of the most recently-sent BEACON. Before issuing a new BEACON, an XHCP Server compares the network configuration information in the current BEACON against that in the previous BEACON. If they are the same, the CONFIG_CHANGED flag in the outgoing BEACON should be set to 0. If they are different, the CONFIG_CHANGED flag in the outing BEACON should be set to 1. If an XHCP Server has just booted up and does not have a copy of BEACON, it should always set the flag.

An XHCP BEACON contains the router DAG of the router sending the BEACON. For a multihomed network, i.e. when a router has multiple WAN-side interfaces, a router DAG may include fallbacks through different NIDs. The default router DAG is determined by the network's own policies in terms of performance, cost or other criteria.

The CONFIG_CHANGED flag is intended to be a quick indicator for XHCP Clients to tell whether they need to process the BEACON further due to a network configuration change. It saves a compare operation of potentially hundreds if not thousands of bits (each XID in a DAG is 160 bits). If CONFIG_CHANGED=0, XHCP Clients can discard the BEACON. If CONFIG_CHANGED=1, XHCP Clients should continue to examine the rest of the BEACON and process it as appropriate.

6.3.2 Supporting Mobile Services

The above solution allows a mobile client to maintain a connection. In order for a mobile server to be reachable for incoming connections, XIA uses a rendezvous service to track the current locator of a mobile server [70].

Conceptually, a rendezvous service may appear similar to an MIPv6 home agent. Whereas an MIPv6 home agent is required for both mobile clients and mobile servers, in XIA, a rendezvous service is optional for mobile clients, which means less infrastructure is needed to support client mobility.

6.3.3 Implementation

The XIA Migration Protocol has been implemented [71]. Active session migration for mobile clients and rendezvous services for mobile servers both perform as expected. A demonstration of migrating a video stream in a moving vehicle is available [72].

6.4 Comparing Migration in IPv6 With XIA Migration Protocol Method

This section presents a comparison between migration support in IPv6 and in XIA. We start with a performance evaluation followed by a discussion on the security design of the migration protocols. The related issue of multihoming support is discussed separately in the next section.

6.4.1 Performance Comparison

We compare the handoff latency of each protocol, defined as the time period during which a packet sent by a correspondent host cannot be delivered to the migrating endpoint. This time period can be further divided into two phases. Assume that a mobile host does not multihome and that migration signaling is performed after a mobile host has associated with the new access network. Immediately after the mobile host disconnects from its previous upstream access network, there is a period when the mobile node is connected to no network before physical and link layer association with the new access network can be completed. The first phase contributes equally to the overall handoff latency regardless of the migration protocol being used, so it will not change the relative standing when we compare the handoff latency of each migration protocol. Of greater interest is the time period after the mobile host has associated with its new network, when the migrating endpoint executes the migration protocol. The time to complete migration signaling is protocol-specific, and thus affects handoff latency differently. More specifically, the contribution to handoff latency from running a migration protocol is the time it takes for a correspondent node/home agent to be informed of a migrating endpoint's new locator after the migrating endpoint gets the new locator.

We will compare performance in three scenarios corresponding to three levels of migration: flow migration, host migration, and network migration. We assume that there are ff flows talking to cc correspondent hosts per migrating host, h migrating hosts per mobile network.

In each scenario, we ask the following quantitative questions:

- What is handoff latency?
- What is the signaling overhead? i.e. what is total number of migration signaling messages?

We also ask the following qualitative questions protocol:

- Does data traffic take a direct path between the endpoints, or is it routed through an intermediary? In other words, is it route-optimized?
- Will a mobile service continue to be reachable after performing a migration using the protocol in question?
- Does the protocol support flow migration?

Table 5 Summary of MIPv6, NEMO and XIA Migration Protocol

	MIPv6 Bidirectional Tunneling	MIPv6 Route Optimization	MIPv6 Enhanced RO	NEMO Basic Support	XIA
Protocol Message Set	BU, BA	BU, BA, HoTI*2, HoT *2, CoTI, CoT, BU, (BA)	BU, BA, Early BU+CoTI, Early BA+CoT, BU, (BA)	BU, BA	MIGRATE, MGRTACK
Route-optimized?	No	Yes	Yes	No	Yes
Flow migration?	No	No	No	No	Yes
# of messages Host migration:	2	$2+7c$	$2+3c$	n/a	$2+2f$
# of messages Network migration:	$2h$	$(2+7c)h$	$(2+3c)h$	2	$(2+2f)h$
Latency Host migration:	$0.5RTT_{MNHA}$	$1RTT_{MNHA} + 1RTT_{MNHA} + 1RTT_{HACN} + 0.5RTT_{MNCN}$	$0.5RTT_{MNCN}$	n/a	$0.5RTT_{MNCN}$
Latency Network migration:	$0.5RTT_{MNHA}$	$1RTT_{MNHA} + 1RTT_{MNHA} + 1RTT_{HACN} + 0.5RTT_{MNCN}$	$0.5RTT_{MNCN}$	$0.5RTT_{MRHA}$	$0.5RTT_{MNCN}$

Table 5 summarizes our findings. Starting with the qualitative comparison, among the five migration protocols, MIPv6 Bidirectional Tunneling mode and NEMO Basic Support Protocol are not route-optimized. All five protocols allow a mobile service to remain reachable using the same identifier before and after migration. The rendezvous service in XIA can be viewed as the equivalent of a home agent in MIPv6 in this regard. Among the five migration protocols, only the XIA Migration Protocol supports flow migration. Flow migration allows fine-grained and flexible management of an application's access network usage, giving XIA a clear advantage over MIPv6 and NEMO BSP in the ability to improve allocative efficiency in multihomed environments. The XIA Migration Protocol holds a clear advantage over its IPv6 counterparts in our qualitative comparison.

Quantitatively, in terms of handoff latency, all protocols except MIPv6 Route Optimization are comparable. They can notify a correspondent node (in the case of XIA Migration Protocol and MIPv6 Enhanced Route Optimization) or a home agent (in the case of MIPv6 Bidirectional Tunneling and NEMO BSP) of a migrating endpoint's new locator with one one-way message, hence the $0.5\text{RTT}_{\text{MNCN}}/0.5\text{RTT}_{\text{MNHA}}$. MIPv6 Route Optimization is significantly slower due to its complex protocol message set.

In terms of the number of signaling messages consumed to perform a migration, MIPv6 Bidirectional Tunneling and NEMO Basic Support Protocol are clear winners for a host migration and a network migration, respectively. MIPv6 Route Optimization consumes the most signaling messages among the three MIPv6 variants. There is no clear winner between MIPv6 Enhanced Route Optimization and the XIA Migration Protocol. For MIPv6 Enhanced Route Optimization to consume fewer messages than the XIA Migration Protocol, it must be that $2+3c<2+2f$, or $f>1.5c$. That is, when there are more than 1.5 flows per correspondent node on average, the XIA Migration Protocol will use more messages than MIPv6 Enhanced Route Optimization does in migrating a host or a mobile network.

MIPv6 Bidirectional Tunneling and NEMO Basic Support Protocol both have a simple message set and low handoff latency, but they incur higher packet propagation delay due to the absence of Route Optimization.

In summary, we conclude that the XIA Migration Protocol and MIPv6 Enhanced Route Optimization are the better migration protocols among the five candidates, for their low handoff latency and route efficiency. The XIA Migration Protocol has the additional advantage of supporting flow migration, although it might come at the expense of higher signaling overhead when there are many flows to the same correspondent host, also referred to as "Signaling Storm" [73].

6.4.2 Security Comparison

To compare the security of migration protocols, we assess their respective abilities to withstand all possible attacks that involve either forging or replaying any of the messages in a migration protocol. Migration protocols provide a mechanism to redirect packets from one destination locator to a new destination locator. An attacker may exploit this capability in an attempt to disrupt a normal packet flow, causing denial-of-service, data breach and/or falsified information.

Section 6.4.2.1 discusses the spoofed or replayed Binding Update (BU)/MIGRATE messages, since they would be the most harmful ones within the protocol message set. Section 6.4.2.2 discusses all other spoofed/replayed protocol messages.

6.4.2.1 Spoofed/Replayed BU/MIGRATE Messages

6.4.2.1.1 Threats and Consequences

Spoofed and replayed BU/MIGRATE messages present the biggest concern. Availability, confidentiality and message integrity would be at risk if BU/MIGRATE messages are not authenticated.

Availability. An attacker may forge a BU/MIGRATE message with the "migrate-to" locator set to any locator other than the victim's true locator. If a home agent or correspondent node accepts such a BU/MIGRATE, it would stop sending packets to the victim's true locator.

Confidentiality. An attacker may forge a BU/MIGRATE message with the "migrate-to" locator set to its own locator. If a home agent or correspondent node accepts such a BU/MIGRATE message, it would send packets destined to the victim to the attacker instead, potentially exposing the content of the message. An attacker could even spoof BU/MIGRATE messages to both endpoints of a session and stage a man-in-the-middle attack.

Message integrity. A man-in-the-middle could also modify the content of the packets received before forwarding them to the actual recipients. There are other mechanisms to defend against confidentiality and message integrity attacks, such as encryption. Those methods are separate from the migration protocols themselves.

6.4.2.1.2 Defenses in Mobile IPv6

The primary defense mechanism against spoofing attacks in MIPv6 Route Optimization is the return routability procedure. The return routability procedure tests the reachability of the sender of a BU message at both the claimed home address and care-of address. It does not defend against attackers who are able to receive both the Home Test message and Care-of Test message, but practically limits the location of attackers to the path between a correspondent node and the home agent [61]. Attackers may attempt to replay a BU message to get around the return routability procedure, but a stale BU message would be identified by the correspondent node using the sequence number.

MIPv6 Enhanced Route Optimization builds upon the same reachability-probing principle, and adds an extra layer of protection through the use of CGAs. Because the home keygen token is transmitted encrypted, assuming that the mobile node's private key is not compromised, an attacker cannot steal the token to forge a BU message even if he is able to eavesdrop on the path between the correspondent node and the home agent, and receive the Home Test message.

6.4.2.1.3 Defenses in XIA Migration Protocol

Aside from provider-independent addresses [74], an IPv6 address is typically network-dependent. Successful completion of a return routability procedure in MIPv6 is a strong indication that a mobile node is entitled to use the claimed IPv6 addresses, including both the interface identifier and the subnet prefix. The XIA Migration Protocol, on the other hand, examines a network-independent field, typically an SID, when authenticating MIGRATE messages. The "subnet prefix" in the XIA locator is not authenticated. Therefore, when assessing the XIA Migration Protocol, we consider two types of spoofed locators separately, locators with spoofed intent XID and locators with spoofed intermediate XID(s).

The first type, spoofed intent (XID), is easy to spot. In the XIA Migration Protocol, a MIGRATE message must be signed with the sender's private key that is in turn tied to the sender's intent XID, similar to the use of CGA in MIPv6 Enhanced Route Optimization. Without the corresponding private key, a spoofed MIGRATE message will fail the signature check at the receiver and therefore will not be accepted. A replay of a MIGRATE message will be identified by the correspondent node from the sequence number.

The other type of spoofed locator is to use a legitimate intent XID but spoofed intermediate XIDs. Such attacks are harder to protect against. XIA Migration Protocol by default only authenticates the intent XID (typically the SID of the migrating process), even though each XID in the locator might be cryptographically generated. This makes the protocol vulnerable to malicious locators. For example, an attacker may initiate a flow with "$\mathrm{NID_{attacker}}$: $\mathrm{HID_{attacker}}$: $\mathrm{SID_{attacker}}$", and then use "$\mathrm{NID_{victim}}$: $\mathrm{HID_{attacker}}$: $\mathrm{SID_{attacker}}$" as the new locator in a MIGRATE message. The MIGRATE message would look legitimate to a correspondent node because the attacker can correctly sign it. However, when packets are sent to the new locator, they will be forwarded to the victim network, which becomes a model for a flooding attack against $\mathrm{NID_{victim}}$.

Due to the use of DAGs in the XIA architecture, this kind of attack can occur with XIA that would be stopped in IPv6. However, such an attack can occur in any architecture using DAGs as locators, with or without a migration protocol, so it is not specifically a limitation of the XIA Migration Protocol itself. DAGs were adopted in XIA despite this known limitation, because DAGs have other advantages such as flexibility in packet forwarding [65]. DAGs also provide an inexpensive way for multihomed hosts and networks to build redundancy, which will be further discussed in Section V.

In summary, MIPv6 Enhanced Route Optimization is less vulnerable to spoofed BU/MIGRATE message attacks than the XIA Migration Protocol as a result of the thorough verification of both interface identifier as well as subnet prefix in MIPv6. Both protocols can defend against replay attacks, although they use different mechanisms to do so. MIPv6 Enhanced Route Optimization avoids public-key cryptography for every migration by design due to performance concerns. Increasing computational power in end devices will make universal adoption of public-key cryptography less of an issue.

6.4.2.2 Spoofing/Replaying Other Protocol Messages

Section 6.4.2.1 reviewed how MIPv6 and XIA Migration Protocol cope with spoofed or replayed BU/MIGRATE messages. This section will go over the remaining messages types in each protocol and show that spoofing or replaying those messages is not harmful.

6.4.2.2.1 Mobile IPv6 Enhanced Route Optimization

Spoofed or replayed HoTI and CoTI messages. Such messages will trigger the correspondent node to send HoT and CoT messages containing the tokens needed for authenticating a BU message. By themselves they are not harmful.

Spoofed or replayed HoT and CoT messages. Upon receiving a HoT or CoT message, a mobile node will extract the contained token from such messages and proceed to send a BU message, if the mobile node has a corresponding entry in its Binding Update List. If the mobile node has not previously sent a HoTI or CoTI message, it would not find the corresponding entry in that list, and would simply ignore the HoT and CoT messages.

Spoofed BA message. Such a message becomes an issue if the correspondent node has not received the BU message. It would indicate to the mobile node a successful registration that did not actually happen. However, the probability of such an event is small.

Replayed BA message. Such messages will be identified through a stale sequence number by the mobile node.

6.4.2.2.2 XIA Migration Protocol

Spoofed MGRTACK message. In a MGRTACK message, a correspondent node is required to sign the MGRTACK. An attacker needs to compromise the correspondent node's private key in order to forge a MGRTACK.

Replayed MGRTACK message. Such messages will be identified by the mobile node via the stale sequence number.

In summary, spoofing or replaying protocol messages other than the BU/MIGRATE message does not pose a major threat.

6.5 Comparing Multihoming Support in IP and XIA

Today's mobile devices are often equipped with multiple network interfaces. They frequently multihome and possess multiple Internet locators, a classic example being a smartphone connected to both a LTE network and a Wi-Fi network. Primary use cases of multihoming include increased redundancy, load balancing, and QoS matching.

This section highlights the role of migration protocols in multihomed environments. IPv6 and XIA differ greatly in terms of how hosts and networks can make use of multihoming. Migration is closely related to multihoming in XIA, while the two are largely independent under IPv6.

6.5.1 Using Multihoming to Improve Resiliency

One common incentive to multihome is to improve resiliency in case of upstream link failure. Hosts and networks that value uninterrupted connectivity can utilize multihoming because multiple link failures at the same time are less likely than a single link failure. In the event of a link failure, in-flight packets may be lost if they are forwarded onto the downed link. End users would like to shorten the reaction time to link failure, which we define as the time period during which in-flight packets cannot be delivered to the recipient after a link failure.

For network multihoming, the IP architecture has BGP-based and NAT-based solutions [74]. NAT-based solutions cannot preserve session continuity in the event of a link failure. For BGP-based, the reaction time depends on how quickly BGP converges when a link becomes unavailable. Reducing convergence time of dynamic routing protocols has proved difficult, and current solutions to do so come at the expense of router overhead and protocol complexity [75].

XIA provides a complementary failure-handling mechanism that would improve reaction time to link failure thanks to the fallback feature of DAGs. In XIA, applications have the option to construct their DAGs with fallbacks to expose multiple access networks. When a router makes a forwarding decision on a packet whose destination DAG contains fallback(s), if the link to the primary XID is unavailable, the router can simply switch to the appropriate fallback XID as the forwarding destination. As long as the router before the downed link is aware of the link failure and incoming packets contain usable fallbacks, those packets will be redirected to their final destination via an alternative route when the primary path has failed. In other words, a network multihomed with BGP relies solely on the routing system to provide fault tolerance. With XIA, every packet can carry failover information in the form of fallbacks.

This benefit of fallbacks extends to multihomed hosts as well, such as cell phones with both Wi-Fi and LTE connections. While a large network might afford running BGP, the same does not apply for smaller networks such as in-vehicle networks. For multihomed hosts, BGP-based multihoming is not an option.

6.5.2 Using Multihoming to Improve Allocative Efficiency

Access networks often provide different QoS's and/or charge different prices. For example, cellular networks are usually expensive in terms of cost per unit of data transmitted, but relatively reliable as long as one is in the coverage area. Vehicle-to-vehicle and vehicle-to-infrastructure communications using the DSRC protocol are relatively cheap, but their availability might be limited. On the other hand, applications' QoS requirements and willingness to pay vary.

The set of available access networks of a multihomed mobile host or network, and thus the available QoS's and prices, may change as the device moves. Applications with ongoing flows need to perform a migration, during which they choose which access network to migrate to. Because of diverse QoS requirements and willingness to pay, the preferred access network varies across flows. In order to satisfy all flows, it must be that individual flows can migrate to different access networks if they wish to. Therefore, our criterion for assessing Internet architectures in terms of ability to utilize multihoming is:

- Does the architecture support flow-level migration?

The MIPv6 protocol family performs migration typically at the host level, as flows on the same host usually share a single IP address in current implementations. NEMO BSP works at the network level. Only the XIA Migration Protocol can perform flow-level migration, making XIA particularly desirable for reaping multihoming benefits.

6.5.3 Load Balancing for Multihomed Hosts and Networks

Lastly, a multihomed host or network may wish to balance incoming load across all of its links to avoid congestion on one particular link. Assigning traffic to a particular link (corresponding to a particular locator) during flow initiation stage is well-understood [76]. To shift an existing flow between access networks, a flow migration protocol is needed. Granular flow-level migration protocols, including the XIA Migration Protocol, have a natural advantage over host-level solutions.

6.6 Conclusions

In this paper, we briefly reviewed the Mobile IPv6, Network Mobility Basic Support Protocol and Multipath TCP. We presented the XIA Migration Protocol in the context of eXpressive

Internet Architecture and compared it with the IPv6 mobility solutions. We then discussed the related issue of multihoming and presented a comparison between XIA and IP with respect to three multihoming use cases.

From a migration support perspective, both IPv6 and XIA provide acceptable solutions. The handoff latency of XIA Migration Protocol and Mobile IPv6 Enhanced Route Optimization are expected to be comparable, consuming one one-way message propagation time. In both cases, data traffic will take direct paths between the endpoints after a migration. Neither protocol creates significant vulnerability to spoofing or replay attacks. However, the two protocols differ slightly in terms of the infrastructure needed to support mobility. Home agents are required in MIPv6, regardless of whether the mobile node is a client or a server. A mobile node must establish a trust relationship with one or more home agents a priori, which represents higher overhead to engage in mobility, and the mobile node is dependent on the home agent when migration occurs. In contrast, a rendezvous service is optional in XIA if a mobile node only engages in client activities.

From a multihoming perspective, XIA has advantages over IP for two reasons. Regarding fault tolerance, existing IP solutions either forgo scalability or sacrifice resiliency. BGP-based multihoming is only suitable for larger networks, and mobile hosts and smaller networks like vehicular networks may not afford BGP. NAT-based multihoming works for hosts and networks of all sizes, but it does not preserve session continuity when a link fails. XIA's fallback-based multihoming achieves resiliency while remaining scalable as it does not rely on a heavy-weight protocol like BGP, or a middle box like NAT that breaks session continuity. Moreover, the reaction time to link failure is shorter in XIA due to DAG-based locators; in an XIA network with DAGs, packets can be rerouted to a fallback path after a failure on the preferred path long before new routes could be established using a distributed algorithm like BGP. Regarding load-balancing and allocative efficiency, XIA's flow-level migration allows fine-grained incoming load-balancing and QoS-matching, whereas in a typical IP network one IP address is associated with many flows and flow migration relies on specialty protocols such as MPTCP. In a world where multihoming is the norm, XIA brings substantial benefits.

The enablers of these benefits are twofold, which brings us to our lessons learned on specific features and design choices. From a protocol design perspective, flow-level migration allows fine-grained load balancing and QoS matching for multihomed devices running heterogeneous applications. However, it implies more signaling messages than host- or network-level migration where there is more than one flow per correspondent host, or more than one host per network. Multipath TCP does allow flow-level migration, but is limited to TCP flows.

From an architectural point of view, DAG-based locators, and fallbacks in particular, allow quick response to link failure and enhances the resiliency of multihomed hosts and networks. Identifier-locator separation plays an important role in mobility management, which is inherent to the XID-DAG setup of XIA and implicit in the HoA-CoA arrangement of MIPv6. However, DAGs are not without drawbacks. An attacker can construct a malicious DAG with XIDs that he is not entitled to use. Additional mechanisms are required to mitigate such threats. Lastly, cryptographically generated addresses, or self-certifying identifiers in general, facilitate authentication of migration signaling messages, which helps reduce the round trips consumed by authentication and thus handoff latency. They are a fundamental building block in both MIPv6 Enhanced Route Optimization and XIA. Internet architects may want to consider flow-level migration, DAG-based locators and/or self-certifying identifiers when designing the future Internet.

7 Conclusions

Multi-network access is an innovative scheme that can greatly improve spectral efficiency by taking advantage of the spatial diversity in the communications infrastructure belonging to different network operators. MNA can extract over 70% more capacity out of existing infrastructure and spectrum, or equivalently, save 20% in both infrastructure and spectrum and still produce the same capacity. The benefit of MNA is easily attainable; it does not require large separation distance between the transmitters belonging to different network operators, nor does it require a large number of non-colocated transmitters. With an appropriate network selection algorithm, MNA always improves spectral efficiency, unless all transmitters are colocated[3].

Colocation or MNA?

Colocation and MNA can work in tandem to reduce cost. The optimal level of colocation (one that minimizes cost for a given network capacity) depends on how widely adopted MNA is. In general, the higher the fraction of MNA-capable users, the less the optimal extent of colocation because there are more users who can utilize the spatial diversity. If the fraction of MNA-capable users is expected to be low, MNOs should continue to pursue colocation as much as needed. The fraction of MNA-capable users MO-MVNOs may be low because MO-MVNO cannot gain significant market share. Even if MNOs engage in capacity sharing, the fraction of MNA-capable users may remain small in the short term because it takes time for new and old devices to acquire the hardware and/or software capabilities to support MNA. If MNA-capable users

[3] Even when all transmitters are colocated, other researchers still found MNA beneficial if sectors do not align [11].

gradually become the majority of all users, then MNOs should consider separate locations when constructing new cell sites.

How should an MNA-capable device decide which network to use?

When participating MNOs have similar tower density and spectrum holdings, a simple algorithm which picks higher SINR strikes a good balance between total capacity and throughput fairness. At low MNA penetration in particular, an MNA-capable users should attach to the base station of highest SINR to maximize efficiency. Such an algorithm may not maximize capacity when there is high level of colocation, but the difference is small compared to an algorithm that picks higher expected data rate.

Using an algorithm that picks higher expected data rate can avoid some of the shortcoming of an algorithm that considers only SINR when there is high levels colocation, or when there is a large imbalance in the base station density and/or spectrum bandwidth of the participating MNOs. However, an algorithm that considers expected data rate may be more complex to implement as it may require additional coordination between user devices and base stations to work well, the mechanisms for which may not exist in current technologies, or may only be available to an MNO itself.

Network selection algorithm can also change the distribution of traffic volume among participating MNOs, which is also important to an MNO if revenue depends on traffic volume. For a given resource ratio between two participating MNOs, a network selection algorithm that considers only SINR tend to steer traffic to the MNO with more infrastructure and/or spectrum resources, while an algorithm that considers expected data rate would have the opposite effect. There is no right or wrong algorithm per se, but the financial terms of an MNA arrangement should reflect the traffic distribution as a result of the network selection algorithm chosen, among other things.

Is MNA a good idea for network operators?

To determine the economic impact of MNA on an individual network operator requires careful examination of both business decisions and technical parameters like the network selection algorithm used. When mobile network operators participate in MNA, the distribution of traffic volumes and the associated revenue among partner MNOs with MNA is not necessarily equal to that without MNA; and it is not necessarily commensurate with the distribution of investment on infrastructure and spectrum resources among partner MNOs. Imbalance in this regard can cause uneven distribution of the economic benefits of MNA among participating network operators. Arrangements exist that make all stakeholders, including consumers, better off; exactly how the economic benefits are distributed among all stakeholders comes down to negotiations. It is relatively straightforward for an MNO to partner with an equally-resourced MNO, as traffic distribution is unlikely to change. But if one partner MNO has significantly more/less resources, more research and development and higher transaction cost may be needed to find an arrangement that benefit all participant equitably.

An MO-MVNO can make a reasonably attractive profit margin if partner MNOs are willing to pass on most of their cost savings. Prospective MO-MVNOs should not necessarily be concerned about market share – while higher market share does increase potential profit margin, an MO-MVNO can already make a good profit margin at low market share. That said, MNOs may still be reluctant to risk its retail business despite the potential cost efficiency from partnering with an MO-MVNO, so having a strong brand appeal and being able to target market segments different from those of partner MNOs remain crucial for an MO-MVNO, as is the case of traditional MVNOs.

MNA in 5G and future networks

MNA may become more relevant in 5G as 5G brings about several innovations that could improve various aspects of MNA. The use of network slicing [77] could simplify the renting/trading of capacity between two facilities-based MNOs, and the setup of an MO-MVNO. Lower signaling overhead [78] could allow 5G networks to support more devices attached to multiple networks, and streamline the dynamic switching between different networks based on current application needs.

While capacity-limited macro cells are obvious candidates to leverage the power of MNA, other types of radio access infrastructure may not benefit as much. One example is neutral host small cells, which are localized cellular network inside a building that provides roaming access to MNOs to improve the service traditionally provided by outdoor macro cells in terms of coverage and capacity, possibly using unlicensed or shared access spectrum like the 3.5 GHz Citizens Broadband Radio Service (CBRS) spectrum band. If neutral host small cells turn out to be a popular business model, then MNA may be less applicable for in-buildings wireless services, because all MNOs would be using the same radio access infrastructure inside the building and the spatial diversity would be limited.

Policy implications

MNA presents a way to leverage economies of scale, but without the scale. There has been a constant tussle between MNOs and regulators when it comes to scale. MNOs are obsessed with greater economies of scale [79], especially in the form of a merger, while regulators are rightfully wary of possible ramifications for competition and prices from such a merger. MNA may be a viable middle ground and perhaps an argument against the merger between T-Mobile and Sprint. MNA effectively pools communications infrastructure and increases utilization of scarce spectrum resources, without necessarily transforming the competitive landscape.

That said, certain forms of MNA, such as when two or more facility-based MNOs partner with each other (through a flexible roaming agreement, buying network slices from each other, or some other technical means), may be of concern to antitrust regulators. As we have shown in the thesis, to make sure all participating MNOs get a fair share of MNA's benefits, MNOs may need to jointly make many business decisions, such as investment on infrastructure and spectrum; they may need to develop a new network selection algorithm and adjust resource allocation schemes. Such decision-making and research and development (R&D) may require exchanging sensitive information at the core of an MNO's business, and deep cooperation like that starts to blur the line between a roaming partner and a merger partner. Regulators may be wary of facility-based MNOs using MNA as an avenue to price fixing.

Data roaming rules may acquire a new interpretation in the context of MNA. In a 2011 ruling, the FCC required that "providers of commercial mobile data services offer data roaming arrangements on commercially reasonable terms and conditions" [80]. An MNA arrangement between two facilities-based MNOs may not be fundamentally different from a roaming partnership, except the criteria for when to roam is different. Can MNO A engage an otherwise unwilling MNO Z for MNA using the data roaming rule? The rule was originally intended to level the playing field for small/new MNOs with a coverage disadvantage. Now that roaming takes on a new meaning in the context of MNA (others have in fact called MNA "flexible roaming" and "smart roaming" [9], [11]), so may the data roaming rule.

The fact that an MO-MVNO is the most relevant form of MNA today also has policy implications. MO-MVNO arrangements may be more preferable to an MNA deal directly between two facilities-based MNOs due to the competitive concerns about the latter as discussed previously. Some regulators around the world have deliberated the idea of a wholesale/MVNO mandate on reasonable terms as an effort to increase competition, and have generally not found it necessary except in Spain and Canada [81]–[83]. However, there is only one MO-MVNO to date to the best of our knowledge [42]. Should regulators step in if an MNO conditions wholesale access on exclusivity, i.e. the MNO will host an MVNO but not an MO-MVNO? One could argue that the potential spectral efficiency benefit of MNA is a case for imposing access obligations to prevent a high barrier to entry for prospective MO-MVNOs. On the other hand, the past few years have witnessed prominent MVNO entries [4], [5], despite the lack of mandated access. It remains to be seen whether MNOs are more or less incentivized to partner with an MO-MVNO compared to a traditional MVNO, and whether policy interventions are necessary.

MNA may prompt a rethink of separating roaming rules and wholesale/MVNO mandate in the first place. They are both concerned with providing access to the radio access network, except traditional roaming is customarily incidental access and a traditional MVNO is customarily permanent access. In fact, one Canadian operator used a roaming deal to support an MVNO business [82], although it was deemed inappropriate by the Canadian regulators at the time. MNA is neither incidental nor permanent access; the optimal may be somewhere in between these two schemes. The rules should reflect the potential spectral efficiency gains from preserving opportunities to use a more intelligent roaming or wholesale access method.

MO-MVNO may be the most relevant form of MNA today, but end user-initiated MNA may unleash greater potential of MNA, e.g. lower cost, increased competition. However, MNOs' artificial restriction on technologies that facilitate switching between carriers like refusing to support eSIM [84] and Apple SIM [85] is a major barrier to end user-initiated MNA. The use of these non-harmful devices could have been protected by the FCC's 2015 Open Internet Order [86], but the relevant no-blocking provisions were eliminated in the 2017 order [87]. Restrictions that hinder wider adoption of MNA do not serve the public interest.

8 Future Work

Several directions can be explored in future research.

8.1 To obtain more accurate estimates of MNA's benefits

We made some simplifying assumptions when estimating the benefit of MNA. Although they are unlikely to affect our conclusions on MNA, more accurate estimates may be of interest to new and existing network operators, and to policymakers.

This thesis has used simulations based on hypothetical hexagonal grid model and random tower locations. These models may not represent real base station deployments. If the data of real base station locations are available, we can apply the same method and simulation model to the base station layout of production cellular networks.

This thesis has largely focused on low-band spectrum and assumes that MNOs operate in roughly the same frequency range. In practice, some MNO may own more low-band spectrum while others own more mid-band; some MNO may be deploying millimeter wave radio while others are not. It would be useful to examine the effect of MNA when different spectrum holdings and propagation characteristics are accounted for.

Real networks contain a mix of macro and small cells, i.e. are heterogeneous networks (HetNets). While our use of random tower locations did produce cells of different sizes, all base stations were assumed to transmit at the same power, and within a network a user was assumed attach to the base station of maximum SINR. In practical HetNets, transmit power varies; cell association policy pushes load onto small cells and does not strictly follow "max-SINR".

The features discussed above can be accounted for with modest changes to the existing simulator. Other characteristics, such as MIMO, beamforming, mobility, etc., may require more substantial modifications to the simulator or a more sophisticated software tool altogether like the Vienna simulators [88].

8.2 How to implement a "Maximum-Throughput" selection algorithm?

We have discovered in this thesis that there are realistic scenarios (e.g. uneven spectrum bandwidth, pervasive colocation, etc.) where simply attaching to the network that provides higher SINR can lead to suboptimal efficiency or fairness, and that a network selection algorithm that considers higher expected data rate can address some of those shortcomings. Therefore, an intelligent network selection algorithm is crucial for reaping the benefits of MNA.

However, developing such an algorithm may not be trivial, as it may require information not necessarily available to the end user device, such as a base station's load. With operator-initiated MNA (e.g. capacity sharing between MNOs, MNA through an MO-MVNO, etc.), operators may be able to build some intelligence into tailored device firmware and software, as is the case of Google Fi. For end user-initiated MNA with unlocked phones, though, it may be more difficult to rely on customizations like that. Having an intelligent network selection algorithm is important for end user-initiated MNA. Along those lines, it may be useful to investigate the following:

- What information can be useful for a distributed network selection algorithm to balance load in a multi-network access environment, while preserving the privacy of network operators?

- Absent coordination with network operators, can end user devices infer the expected data rate it would be receive from different MNOs from the information currently available on today's standard cell phones?

Existing literature on load balancing in heterogeneous networks and on radio access technology selection between Wi-Fi and cellular may shed some light on this topic. But MNA adds a new dimensions to this problem – previous works may not have considered choosing between multiple cellular networks with unbalanced resources and/or partially colocated infrastructure.

Other optimizations on the device side can also benefit MNA. For example, a common issue of today's mobile devices is high latency in network switching, i.e. the device keeps attached to a network that provides weak signal. How to improve the logic of automatically switching to the network of higher QoS? How does a device determine whether the weak signal is due to temporary fluctuation in fading, or is a prelude of losing the connection altogether because the user is moving out of the coverage area? The switching time between two cellular networks is currently on the order of 10 seconds [89]. How to improve that? Many of these challenges are not unique to MNA and solving them can benefit user experience in general.

8.3 How does MNA affect competition?

While multi-network access can significantly improve spectral efficiency and cost effectiveness, its implications on the competition in the mobile data services market is less clear. Without a good understanding of how MNA affects competition, it can be difficult for policymakers to act on MNA, to the extent that policy interventions are needed.

This question can be examined from multiple angles, depending on the form of MNA. For example:

- How does an MO-MVNO affect wholesale competition?
- How does end user-initiated MNA affect retail competition?
- How does capacity sharing between MNOs affect competition?

How does an MO-MVNO affect wholesale competition?

For an MNO, partnering with an MO-MVNO means getting wholesale payment rather than collecting retail rates for at least a portion of the traffic generated by the MO-MVNO. On the other hand, a partner MNO gets to serve users who may be previously a subscriber of a competitor MNO and gets paid for carrying their traffic; it will also be able to carry more traffic with existing infrastructure and spectrum.

In a market of multiple facilities-based MNOs and one MO-MVNO, would the MNOs all choose to partner with the MO-MVNO? What is the market equilibrium? One can use a coalitional game theoretical model [90] to analyze such an arrangement; a nonempty "core" would indicate that it is globally optimal for MNOs to join the coalition, i.e. partner with an MO-MVNO.

With increased number of network operators, marketing cost (advertisement, promotional plans, device subsidy, etc.) may also increase when network operators try to gain market share. Does the introduction of MO-MVNO(s) stimulate marketing expenses and if so, by how much? Would the increase in marketing offset the potential cost reduction from MNA? This is also worth a look

How does capacity sharing between MNOs affect competition?

One way to think about the effect of capacity sharing is: can smaller MNOs collectively achieve the economies of scale of a larger MNO? In other words, how much cost efficiency can two or more MNOs produce by enabling MNA compared to that of a full merger? The answer to these questions may be useful to MNOs considering a merger, or to antitrust regulators debating a merger. This analysis can utilize largely the same engineering economic model developed in this thesis.

8.4 Other benefits and costs of MNA

This thesis has focused on the benefit of MNA, e.g. how much MNA improves spectral efficiency and what is the monetary value of that improvement, etc. But there are costs associated with MNA, too. For example, what is the signal overhead of maintaining simultaneous connections to multiple cellular networks? How much additional energy is required for simultaneous attachment to multiple cellular networks, and how does that affect device battery life?

How much more does it increase the bill of materials (BOM) cost to equip devices with two or more sets of radio frontend? What is the additional cost of a second SIM card slot? How does the cost of eSIM compare to physical SIM? What is the environmental impact of phasing out physical SIM cards? Understanding the magnitude of these costs and benefits would paint a more complete picture of MNA.

9 References

[1] "NSF Future Internet Architecture Project." [Online]. Available: http://www.nets-fia.net/. [Accessed: 21-Nov-2019].

[2] Google, "Frequently asked questions – Project Fi." [Online]. Available: https://fi.google.com/about/faq/. [Accessed: 13-Nov-2017].

[3] L. Yuanjie *et al.*, "iCellular: Device-Customized Cellular Network Access on Commodity Smartphones," *13th USENIX Symp. Networked Syst. Des. Implement.*, pp. 643–656, 2016.

[4] "Charter, Comcast could deploy strand-mounted wireless small cells | FierceWireless." [Online]. Available: https://www.fiercewireless.com/tech/charter-comcast-could-deploy-strand-mounted-wireless-small-cells. [Accessed: 21-Nov-2019].

[5] J. Baumgartner, "Altice USA Signs AT&T to Nationwide Roaming Deal | Light Reading." [Online]. Available: https://www.lightreading.com/mobile/altice-usa-signs-atandt-to-nationwide-roaming-deal-/d/d-id/753161. [Accessed: 21-Nov-2019].

[6] L. Hardesty, "Comcast taps its 19 million Wi-Fi hotspots to offload Xfinity Mobile traffic | FierceWireless." [Online]. Available: https://www.fiercewireless.com/operators/comcast-taps-its-19-million-wi-fi-hotspots-to-offload-xfinity-mobile-traffic. [Accessed: 21-Nov-2019].

[7] "Better coverage with Fi's dual connect technology." [Online]. Available: https://www.blog.google/products/google-fi/better-coverage-with-fis-dual-connect-technology/. [Accessed: 21-Nov-2019].

[8] N. Zhang, J. M. Peha, and M. A. Sirbu, "Expanding Cellular Network Capacity with Multi-Network Access," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–6.

[9] S. Hua, P. Liu, and S. S. Panwar, "The urge to merge: When cellular service providers pool capacity," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 5020–5025.

[10] D. Fooladivanda and C. Rosenberg, "Why user swapping could be the best coordination mechanism in a cellular network?," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 4895–4901.

[11] B. Venkitesh and C. Rosenberg, "Smart Roaming: How Operator Cooperation Can Increase Spectrum Usage Efficiency at Practically No Cost," *IEEE Trans. Netw. Serv. Manag.*, vol. 16, no. 2, pp. 690–700, Jun. 2019.

[12] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile Network Resource Sharing Options: Performance Comparisons," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4470–4482, Sep. 2013.

[13] J. S. Panchal, "Inter-operator resource sharing in 4G LTE cellular networks," 2011.

[14] J. Kibilda, P. Di Francesco, F. Malandrino, and L. A. DaSilva, "Infrastructure and

spectrum sharing trade-offs in mobile networks," in *2015 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2015, pp. 348–357.

[15] J. Kibilda, N. J. Kaminski, and L. A. DaSilva, "Radio Access Network and Spectrum Sharing in Mobile Networks: A Stochastic Geometry Perspective," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 4, pp. 2562–2575, Apr. 2017.

[16] B. Finley and A. Basaure, "Benefits of mobile end user network switching and multihoming," *Comput. Commun.*, vol. 117, pp. 24–35, Feb. 2018.

[17] M. H. Lotfi and S. Sarkar, "The economics of competition and cooperation between MNOs and MVNOs," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 2017, pp. 1–6.

[18] N. Ben Khalifa, A. Benhamiche, A. Simonian, and M. Bouillon, "Profit and Strategic Analysis for MNO-MVNO Partnership," in *2018 IFIP Networking Conference (IFIP Networking) and Workshops*, 2018, pp. 325–333.

[19] J. M. Peha, "Improving Public Safety Communications," *Issues Sci. Technol.*, vol. 23, pp. 61–68, 2007.

[20] J. M. Peha, "Fundamental Reform in Public Safety Communications Policy," *Fed. Commun. Law J.*, vol. 59, no. 3, pp. 455-456,517-545, 2007.

[21] R. Hallahan and J. Peha, "Policies for Public Safety Use of Commercial Wireless Networks," in *TPRC 2010*, 2010.

[22] J. M. Peha, "A public-private approach to public safety communications," *Issues Sci. Technol.*, vol. 29, no. 4, pp. 37–42, 2013.

[23] R. Hallahan and J. M. Peha, "Enabling Public Safety Priority Use of Commercial Wireless Networks," *Homel. Secur. Aff.*, no. 9, p. Article 13, 2013.

[24] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[25] R. Mahindra, H. Viswanathan, K. Sundaresan, M. Y. Arslan, and S. Rangarajan, "A practical traffic management system for integrated LTE-WiFi networks," in *Proceedings of the 20th annual international conference on Mobile computing and networking - MobiCom '14*, 2014, pp. 189–200.

[26] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *2013 Proceedings IEEE INFOCOM*, 2013, pp. 998–1006.

[27] S. Deb, K. Nagaraj, and V. Srinivasan, "MOTA: engineering an operator agnostic mobile service," in *Proceedings of the 17th annual international conference on Mobile computing and networking - MobiCom '11*, 2011, p. 133.

[28] P. Coucheney, C. Touati, and B. Gaujal, "Fair and Efficient User-Network Association Algorithm for Multi-Technology Wireless Networks," in *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, 2009, pp. 2811–2815.

[29] K. Zhu, D. Niyato, and P. Wang, "Network Selection in Heterogeneous Wireless

Networks: Evolution with Incomplete Information," in *2010 IEEE Wireless Communication and Networking Conference*, 2010, pp. 1–6.

[30] W. Wang, X. Liu, J. Vicente, and P. Mohapatra, "Integration Gain of Heterogeneous WiFi/WiMAX Networks," *IEEE Trans. Mob. Comput.*, vol. 10, no. 8, pp. 1131–1143, Aug. 2011.

[31] D. D. Nguyen, H. X. Nguyen, and L. B. White, "Reinforcement Learning With Network-Assisted Feedback for Heterogeneous RAT Selection," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 9, pp. 6062–6076, Sep. 2017.

[32] Y. Li *et al.*, "Device-Customized Multi-Carrier Network Access on Commodity Smartphones," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2542–2555, Dec. 2018.

[33] L. Zheng, J. Chen, C. Joe-Wong, C. W. Tan, and M. Chiang, "An economic analysis of wireless network infrastructure sharing," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2017, pp. 1–8.

[34] L. Zheng, C. Joe-Wong, J. Chen, C. G. Brinton, C. W. Tan, and M. Chiang, "Economic viability of a virtual ISP," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[35] Department for Culture Media Sport, "Tackling Partial Not-Spots in Mobile Phone Coverage," 2014.

[36] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, May-2017.

[37] R. Jurdi, A. K. Gupta, J. G. Andrews, and R. W. Heath, "Modeling Infrastructure Sharing in mmWave Networks With Shared Spectrum Licenses," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 328–343, Mar. 2018.

[38] 3GPP, "3GPP TR 36.942 V15.0.0," 2018.

[39] J. M. Peha, "Cellular economies of scale and why disparities in spectrum holdings are detrimental," *Telecomm. Policy*, vol. 41, no. 9, pp. 792–801, Oct. 2017.

[40] T-Mobile US Inc., "Annual Report," 2018. [Online]. Available: http://d18rn0p25nwr6d.cloudfront.net/CIK-0001283699/f46f27cd-c5f9-447c-b6d1-97818a2d3d00.html. [Accessed: 01-Jul-2019].

[41] Cisco, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper - Cisco," 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html. [Accessed: 23-Jun-2019].

[42] Google, "Google Fi - Plan benefits & details," 2019. [Online]. Available: https://fi.google.com/about/plan/. [Accessed: 23-Jun-2019].

[43] Mi. Dano, "Special Report—25 charts on spectrum ownership in the United States |

FierceWireless," 2018. [Online]. Available: https://www.fiercewireless.com/wireless/25-charts-spectrum-ownership-united-states. [Accessed: 23-Jun-2019].

[44] Crown Castle International Corp., "Annual Report," 2018. [Online]. Available: https://investor.crowncastle.com/node/22611/html. [Accessed: 30-Jun-2019].

[45] Public Safety and Homeland Security, "A Broadband Network Cost Model: A Basis for Public Funding Essential to Bringing Nationwide Interoperable Communications to America's First Responders | Federal Communications Commission," 2010.

[46] Uniti Group Inc., "Annual Report," 2018. [Online]. Available: https://investor.uniti.com/node/9081/html. [Accessed: 30-Jun-2019].

[47] E. Oughton, Z. Frias, T. Russell, D. Sicker, and D. D. Cleevely, "Towards 5G: Scenario-based assessment of the future supply and demand for mobile telecommunications infrastructure," *Technol. Forecast. Soc. Change*, vol. 133, pp. 141–155, Aug. 2018.

[48] J. Lehikoinen, P. Pont, and / / Yannick, "Virtually mobile: What drives MVNO success," 2014.

[49] US DOT, "RITA - Intelligent Transportation Systems - DSRC: The Future of Safer Driving Fact Sheet," *United States Dep. Transp.*, 2015.

[50] M. Handley, O. Bonaventure, C. Raiciu, and A. Ford, "TCP Extensions for Multipath Operation with Multiple Addresses," *IETF*, 2013. .

[51] R. Stewart, "Stream Control Transmission Protocol," *IETF*, 2007. [Online]. Available: https://tools.ietf.org/html/rfc4960.

[52] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," *IETF*, 2011. .

[53] A. Petrescu, R. Wakikawa, P. Thubert, and V. Devarapalli, "Network Mobility (NEMO) Basic Support Protocol," *IETF*, 2005. .

[54] D. Naylor *et al.*, "XIA: architecting a more trustworthy and evolvable internet," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 50–57, Jul. 2014.

[55] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," *IETF*, 2011. [Online]. Available: http://tools.ietf.org/html/rfc6275.

[56] A. Petrescu, R. Wakikawa, P. Thubert, and V. Devarapalli, "Network Mobility (NEMO) Basic Support Protocol," *IETF*, 2005. [Online]. Available: https://tools.ietf.org/html/rfc3963.

[57] M. Handley, O. Bonaventure, C. Raiciu, and A. Ford, "TCP Extensions for Multipath Operation with Multiple Addresses," *IETF*, 2013. [Online]. Available: https://tools.ietf.org/html/rfc6824.

[58] J. Arkko, W. Haddad, and C. Vogt, "Enhanced Route Optimization for Mobile IPv6," *IETF*, 2007. [Online]. Available: https://tools.ietf.org/html/rfc4866.

[59] T. Aura, "Cryptographically Generated Addresses (CGA)," *IETF*, 2005. [Online]. Available: https://tools.ietf.org/html/rfc3972.

[60] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," *Network Working Grouproup*, vol. 6, no. 2460. pp. 1–40, 1998.

[61] J. Arkko, T. Aura, G. Montenegro, E. Nordmark, and P. Nikander, "Mobile IP Version 6 Route Optimization Security Design Background," *IETF*, 2005. [Online]. Available: https://tools.ietf.org/html/rfc4225.

[62] A. Z. M. Shahriar, M. Atiquzzaman, and W. Ivancic, "Route optimization in network mobility: Solutions, classification, comparison, and future research directions," *IEEE Commun. Surv. Tutorials*, vol. 12, no. 1, pp. 24–38, 2010.

[63] R. Stewart, X. Q., M. Tuexen, S. Maruyama, and M. Kozuka, "Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration," *IETF*, 2007. [Online]. Available: https://tools.ietf.org/html/rfc5061.

[64] S. Zeadally and F. Siddiqui, "An Empirical Analysis of Handoff Performance for SIP, Mobile IP, and SCTP Protocols," *Wirel. Pers. Commun.*, vol. 43, no. 2, pp. 589–603, Oct. 2007.

[65] D. Naylor *et al.*, "XIA," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 50–57, Jul. 2014.

[66] J. Pan, S. Paul, R. Jain, and M. Bowman, "MILSA: A Mobility and Multihoming Supporting Identifier Locator Split Architecture for Naming in the Next Generation Internet," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, 2008, pp. 1–6.

[67] X. Liu and L. Xiao, "A Survey of Multihoming Technology in Stub Networks: Current Research and Open Issues," *IEEE Netw.*, vol. 21, no. 3, pp. 32–40, May 2007.

[68] A. C. Snoeren and H. Balakrishnan, "An end-to-end approach to host mobility," in *Proceedings of the 6th annual international conference on Mobile computing and networking - MobiCom '00*, 2000, pp. 155–166.

[69] D. Barrett, "XIA Prototype Overview," *GitHub*, 2017. [Online]. Available: https://github.com/XIA-Project/xia-core/wiki/XIA-Prototype-Overview.

[70] N. Gupta, "XIA Mobility and Intrinsic Security," *GitHub*, 2017. [Online]. Available: https://github.com/XIA-Project/xia-core/wiki/XIA-Mobility-and-Intrinsic-Security.

[71] Carnegie Mellon University, "eXpressive Internet Architecture Project," *GitHub*, 2016. [Online]. Available: https://github.com/XIA-Project/xia-core.

[72] R. Meireles, "XIA Vehicular Video Demo," *YouTube*, 2016. [Online]. Available: https://youtu.be/msLZnPcNp2o.

[73] M. Watari, C.-W. Ng, F. Zhao, and P. Thubert, "Network Mobility Route Optimization Solution Space Analysis," *IETF*, 2007. [Online]. Available: https://tools.ietf.org/html/rfc4889.

[74] V. Gill, J. Abley, E. B. Davies, K. E. Lindqvist, and B. Black, "IPv4 Multihoming Practices and Limitations," *IETF*, 2005. [Online]. Available:

https://tools.ietf.org/html/rfc4116.

[75] M. Caesar, M. Casado, T. Koponen, J. Rexford, and S. Shenker, "Dynamic route recomputation considered harmful," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 2, p. 66, Apr. 2010.

[76] Fanglu Guo, Jiawu Chen, Wei Li, and Tzi-cker Chiueh, "Experiences in building a multihoming load balancing system," in *IEEE INFOCOM 2004*, vol. 2, pp. 1241–1251.

[77] K. Sparks *et al.*, "5G NETWORK SLICING WHITEPAPER FCC Technological Advisory Council 5G IoT Working Group."

[78] S. Hailu, M. Saily, and O. Tirkkonen, "RRC State handling for 5G," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 106–113, Jan. 2019.

[79] "T-Mobile and Sprint to Combine, Accelerating 5G Innovation & Increasing Competition | T-Mobile Newsroom." [Online]. Available: https://www.t-mobile.com/news/5gforall. [Accessed: 20-Nov-2019].

[80] Federal Communications Commission, "Roaming for Mobile Wireless Services." [Online]. Available: https://www.fcc.gov/wireless/bureau-divisions/competition-infrastructure-policy-division/roaming-mobile-wireless. [Accessed: 12-Dec-2019].

[81] A. Banerjee and C. M. Dippon, "Voluntary relationships among mobile network operators and mobile virtual network operators: An economic explanation," *Inf. Econ. Policy*, vol. 21, no. 1, pp. 72–84, Feb. 2009.

[82] Canadian Radio-television and Telecommunications Commission, "Telecom Notice of Consultation CRTC 2019-57 PDF version Notice of hearing," 2019.

[83] "Mobile virtual network operators and competition in mobile markets: Lessons from Canada | American Enterprise Institute - AEI." [Online]. Available: https://www.aei.org/technology-and-innovation/mobile-virtual-network-operators-and-competition-in-mobile-markets-lessons-from-canada/. [Accessed: 12-Dec-2019].

[84] Sue Marek, "GSMA revamps its eSIM standard to appease DoJ anti-collusion investigation | FierceWireless." [Online]. Available: https://www.fiercewireless.com/regulatory/gsma-revamps-its-esim-standard-to-appease-doj-anti-collusion-investigation. [Accessed: 13-Dec-2019].

[85] "Latest iPad Pro Makes It Even Easier to Switch Wireless Carriers - Vox." [Online]. Available: https://www.vox.com/2016/3/22/11587182/latest-ipad-pro-makes-it-even-easier-to-switch-wireless-carriers. [Accessed: 22-Nov-2019].

[86] Federal Communications Commission, "FCC Releases Open Internet Order | Federal Communications Commission." [Online]. Available: https://www.fcc.gov/document/fcc-releases-open-internet-order. [Accessed: 13-Dec-2019].

[87] Federal Communications Commission, "FCC Releases Restoring Internet Freedom Order | Federal Communications Commission." [Online]. Available: https://www.fcc.gov/document/fcc-releases-restoring-internet-freedom-order. [Accessed: 13-Dec-2019].

[88] "Vienna 5G Simulators – nt.tuwien.ac.at." [Online]. Available: https://www.nt.tuwien.ac.at/research/mobile-communications/vccs/vienna-5g-simulators/. [Accessed: 13-Dec-2019].

[89] "Network handover in Google Fi | Nicholas Armstrong." [Online]. Available: https://nicholasarmstrong.com/2015/08/network-handover-google-fi/. [Accessed: 16-Dec-2019].

[90] C. Singh, S. Sarkar, A. Aram, and A. Kumar, "Cooperative Profit Sharing in Coalition-Based Resource Allocation in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 69–83, Feb. 2012.

Appendix A. MNA in different base station topologies with and without fading

The hexagonal grid model with maximum offset predicts higher gains across the board, except for the case of "Equal-Throughput" resource allocation paired with "Max-SINR" network selection algorithm. This is mainly a result of the more consistent location diversity of towers in the hexagonal grid model.

Figure 32 shows the aggregate network capacity as a function of the fraction of MNA-capable users *without* fading.

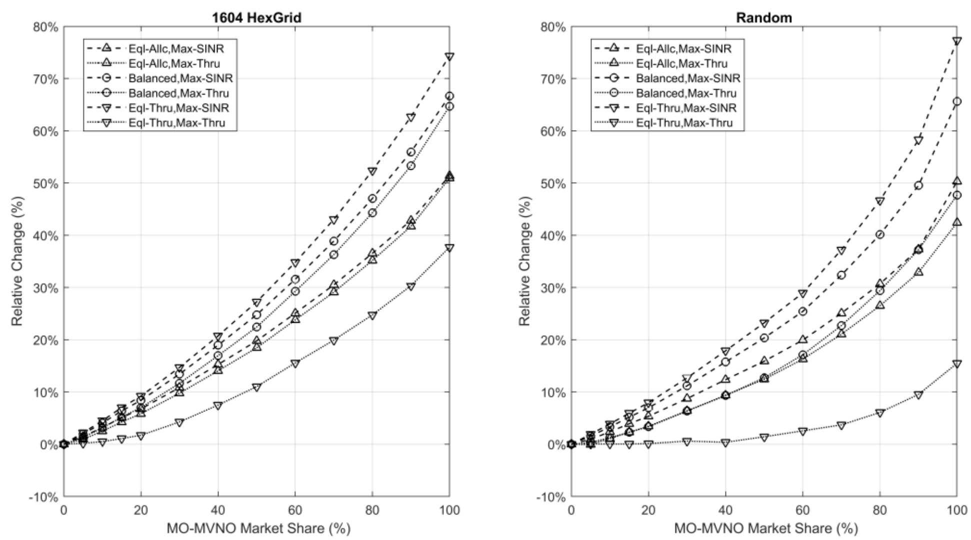

Figure 32 Relative change in aggregate network capacity vs. fraction of MNA-capable users. (MNOs have equal tower density and spectrum bandwidth; no fading with 10 dB standard deviation)

Figure 33 shows the aggregate network capacity as a function of the fraction of MNA-capable users *with* fading.

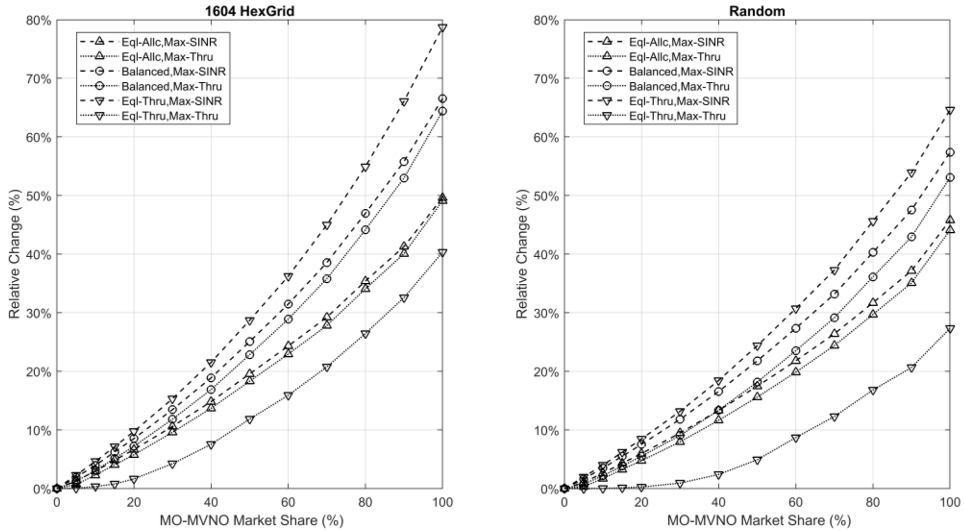

Figure 33 Relative change in aggregate network capacity vs. fraction of MNA-capable users. (MNOs have equal tower density and spectrum bandwidth; log-normal fading with 10 dB standard deviation)

Table 6 summarizes the differences in the capacity improvement between the hexagonal grid model and random tower location model. The two percentages correspond to change under "Max-SINR" and "Max-Throughput", respectively.

Table 6 Capacity Improvement from MNA in Hexagonal and Random Tower Models.

(Max-SINR/Max-Throughput)	Hexagonal grid No fading	Random No fading	Hexagonal grid With fading	Random With fading
Equal Allocation	50%/50%	50%/40%	50%/50%	45%/45%
Balanced	70%/65%	65%/45%	65%/65%	60%/55%
Equal-Throughput	75%/35%	75%/15%	80%/40%	65%/30%

Appendix B. Cell association with high level of colocation

When most base stations are colocated, the interference power is heavily influenced by the location of the few standalone (non-colocated) base stations. Consider a scenario where the standalone base station nearest to colocated tower #1 (with both MNO A and MNO Z base stations) belongs to MNO A, and most other towers in the vicinity are also colocated. Without fading, SINR is deterministic. Most users near the colocated tower #1 will experience more interference in the MNO A's spectrum band, and therefore will all choose MNO Z when "Max-SINR" selection algorithm is used. The MNO A base station may end up serving no users, while MNO Z's base station colocated at the same tower is heavily congested. In this case, MNO A spectrum is completely unused in this area, causing unusually in low capacity. An example of this phenomenon is shown in 0 Figure 34.

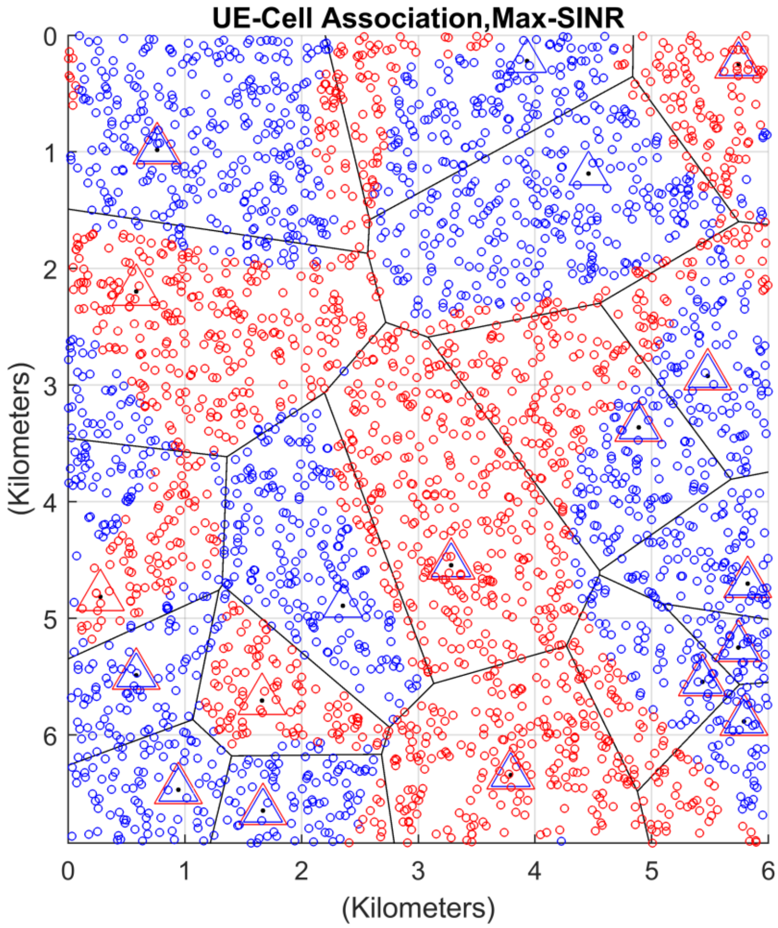

Figure 34 An example of user-cell association under high levels of colocation, without fading. "Equal-Allocation" resource allocation. Circles represent user locations. Triangles represent base station locations. Overlapping triangles indicate a colocated cell tower. The size of triangle is proportional to the average SINR of the users of that base station. Color blue corresponds to MNO A users/base stations; color red corresponds to MNO Z users/base stations. When colocation is prevalent and when there is no fading, one of the base stations at a colocated tower may never be chosen, wasting spectrum.

With fading, the extreme case of a base station serving no users at all is unlikely due to the randomness in interference power. However, unbalanced user SINR distribution can still occur with meaningful likelihood. It is possible that a MNO A base station will only serve users far enough away from it (and the standalone MNO A base station nearby), which causes inefficient utilization of MNO A's spectrum at this base station. That is, some base stations are used mostly by users of low SINR.

An example of such a scenario is shown in Figure 35. In the bottom left quadrant, most users near a colocated cell tower choose an MNO A (blue) base station because there is standalone MNO Z base station nearby but no standalone MNO A base station nearby (thus less interference in MNO A's spectrum band). The colocated MNO Z (red) base stations mostly serve users far away from the tower; there are very few MNO Z users located near those colocated MNO Z base stations.

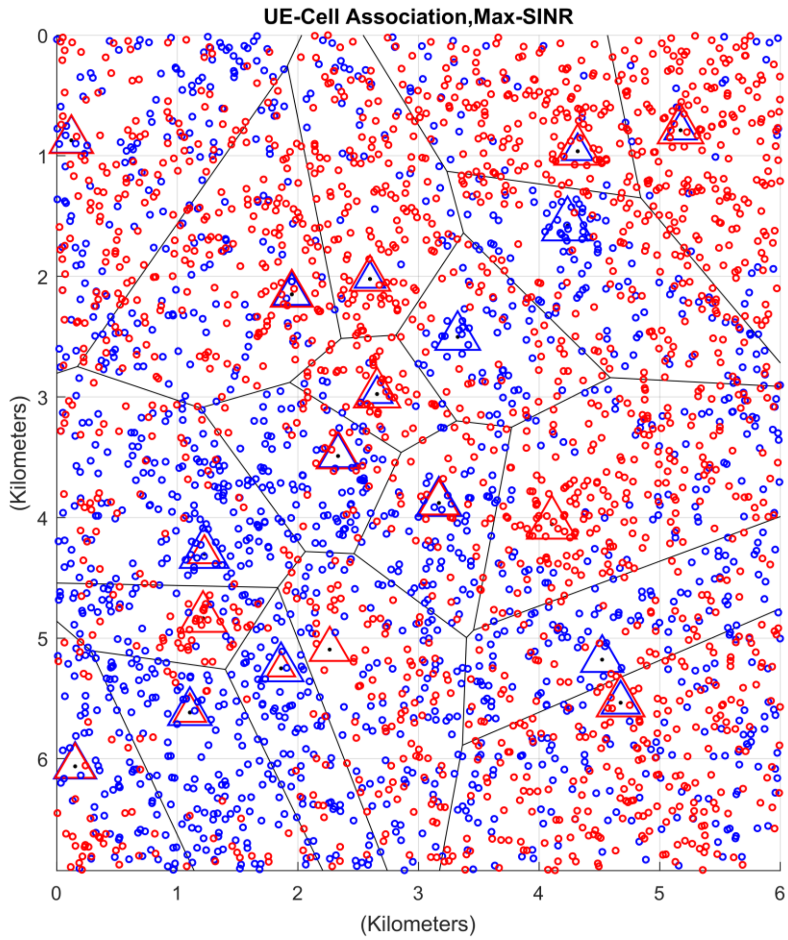

Figure 35 An example of user-cell association under high levels of colocation, with fading. "Equal-Allocation" resource allocation. Circles represent user locations. Triangles represent base station locations. Overlapping triangles indicate a colocated cell tow er. The size of triangle is proportional to the average SINR of the users of that base station. Color blue corresponds to MNO A users/base stations; color red corresponds to MNO Z users/base stations.

Appendix C. NPV Derivation

$$NPV_{business} = \frac{R-C}{i} + \frac{\frac{-k_m R - (R-C)}{1+i}\left[1 - \frac{1}{(1+i)^{\tau}}\right]}{1 - \frac{1}{1+i}}$$

$$= \frac{R - C + [-k_m R - (R-C)]\left[1 - \frac{1}{(1+i)^{\tau}}\right]}{i}$$

$$= \frac{R - C + [-k_m R - (R-C)] - \frac{-k_m R - (R-C)}{(1+i)^{\tau}}}{i} = \frac{-k_m R + \frac{k_m R + (R-C)}{(1+i)^{\tau}}}{i}$$

$$= \frac{-k_m R + \frac{k_m R + mR}{(1+i)^{\tau}}}{i} = \frac{R}{i}\left(\frac{k_m + m}{(1+i)^{\tau}} - k_m\right)$$

$$NPV_{business} = \frac{R}{i}\left(\frac{k_m + m}{(1+i)^{\tau}} - k_m\right)$$

Set $NPV_{business} > 0$ and derive the range of τ.

$$\frac{R}{i}\left(\frac{k_m + m}{(1+i)^{\tau}} - k_m\right) > 0$$

Assuming positive discount rate and revenue, we have $\frac{k_m+m}{(1+i)^{\tau}} - k_m > 0$, which can be rearranged into:

$$\frac{k_m + m}{(1+i)^{\tau}} > k_m$$

$$(1+i)^{\tau} < \frac{k_m + m}{k_m}$$

$$\tau < \log_{1+i}\left(1 + \frac{m}{k_m}\right)$$

Appendix D. Different mix of infrastructure and spectrum resources

In Chapter 5 we assumed that MNOs always adjust their base station density and spectrum bandwidth in proportion so as to minimize the cost of producing a given network capacity, which is a reasonable assumption for the long term (while also assuming the price of base stations and the price of spectrum stay constant). At any given moment, though, the ratio of base station density to spectrum bandwidth may be different for different MNOs. So, we also investigated scenarios where MNOs have the same base station density but different spectrum bandwidth, as well as scenarios where they have the same spectrum bandwidth but different base station densities. This appendix presents some preliminary results based on random tower layout and the same path loss model as used in Chapter 4 except fading was not considered.

a. Same base station density, different spectrum bandwidth

In this analysis, total spectrum bandwidth is held constant, but the spectrum bandwidth of each MNO is varied. The aggregate no-MNA capacity is also held constant.

When two MNOs with the same base station density but different spectrum bandwidth enables MNA, i.e. capacity sharing, "Max-Throughput" network selection algorithm yields lower capacity gain than "Max-SINR". The bigger the spectrum differential, the less efficient "Max-Throughput" becomes. This is because "Max-Throughput" pushes more users onto the MNO with more spectrum, which increases the effective cell size of the MNO with more spectrum and lowers its average SINR, while the opposite is true for the MNO with less spectrum.

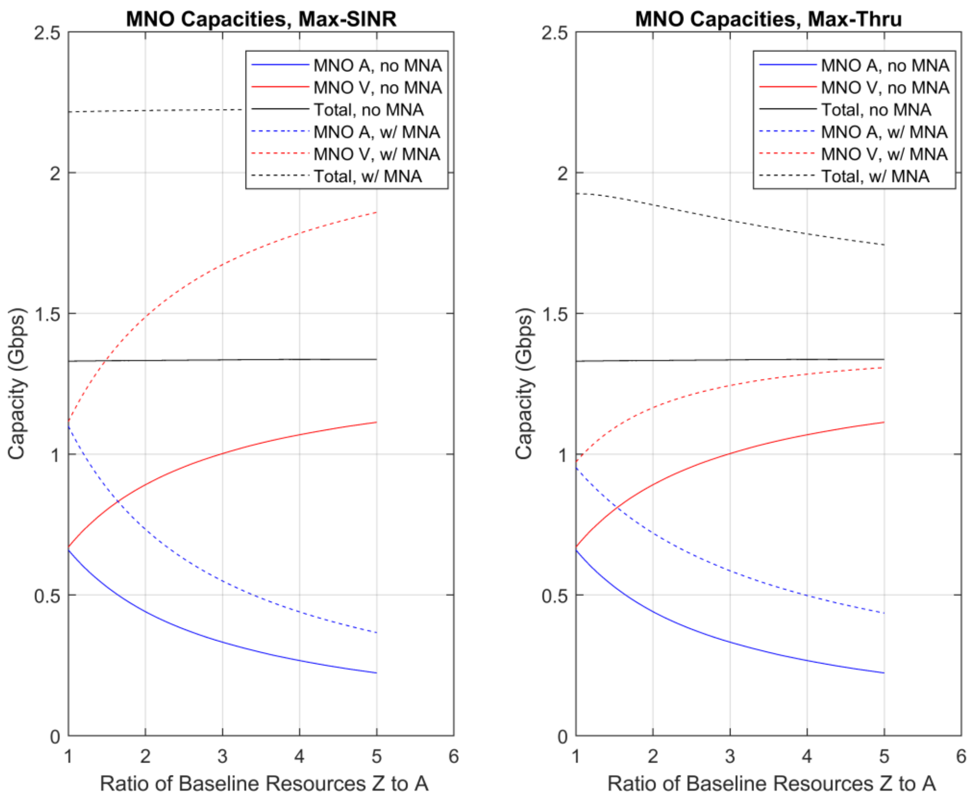

Figure 36 MNO capacity vs. varying relative spectrum bandwidth. Left: "Max-SINR" network selection; right: "Max-Throughput". (Partner MNOs have the same base station density; "Balanced" resource allocation)

As a result, the network selection algorithm heavily influences how traffic is distributed between partner MNOs. With "Max-SINR", one MNO's the share of traffic with MNA is the same as the share of traffic without MNA and is the same as that MNO's share of total spectrum bandwidth. With "Max-Throughput", the MNO with more spectrum will carry a smaller share of traffic if MNA is enabled. This is because the network selection algorithm pushes more users onto the MNO with more spectrum. The MNO with more spectrum uses it less efficiently, and therefor produces a share of capacity that is smaller than its share of spectrum bandwidth. This is consistent with Figure 18 in Chapter 3 where a similar analysis was performed with the hexagonal grid model.

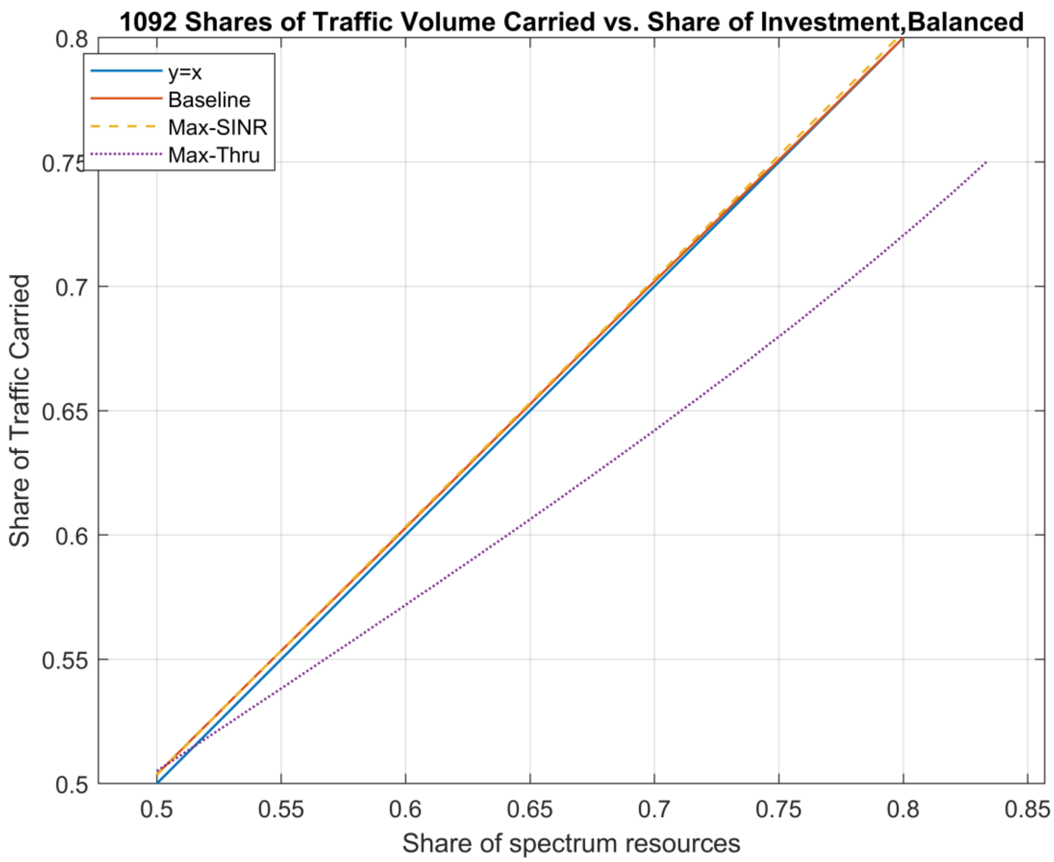

Figure 37 Share of traffic volume carried vs. share of spectrum bandwidth. (Partner MNOs have the same base station density; "Balanced" resource allocation)

b. Same spectrum bandwidth, different base station density

In this analysis, spectrum bandwidth is kept same and constant between the two partner MNOs; the relative base station density between them is varied while keeping aggregate no-MNA capacity constant.

As base station density differential increases, "Max-SINR" becomes more efficient while "Max-Throughput" becomes less efficient.

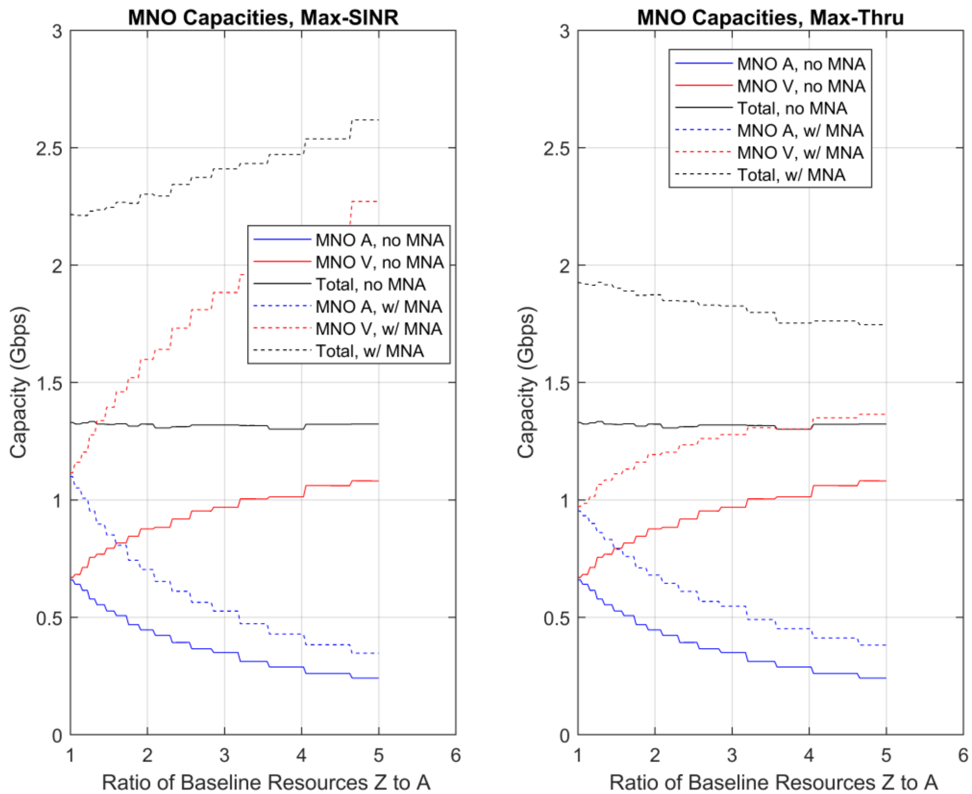

Figure 38 MNO capacity vs. varying relative base station density. Left: "Max-SINR" network selection; right: "Max-Throughput". (Partner MNOs have the same spectrum bandwidth; "Balanced" resource allocation)

When MNA is enabled on two partner MNOs with the same spectrum bandwidth but different base station density, if "Max-SINR" is used, the MNO with more base stations will carry a *larger* share of all traffic than before MNA is enabled; if "Max-Throughput" is used, the MNO with more base stations will carry a *smaller* share of all traffic than before MNA is enabled.

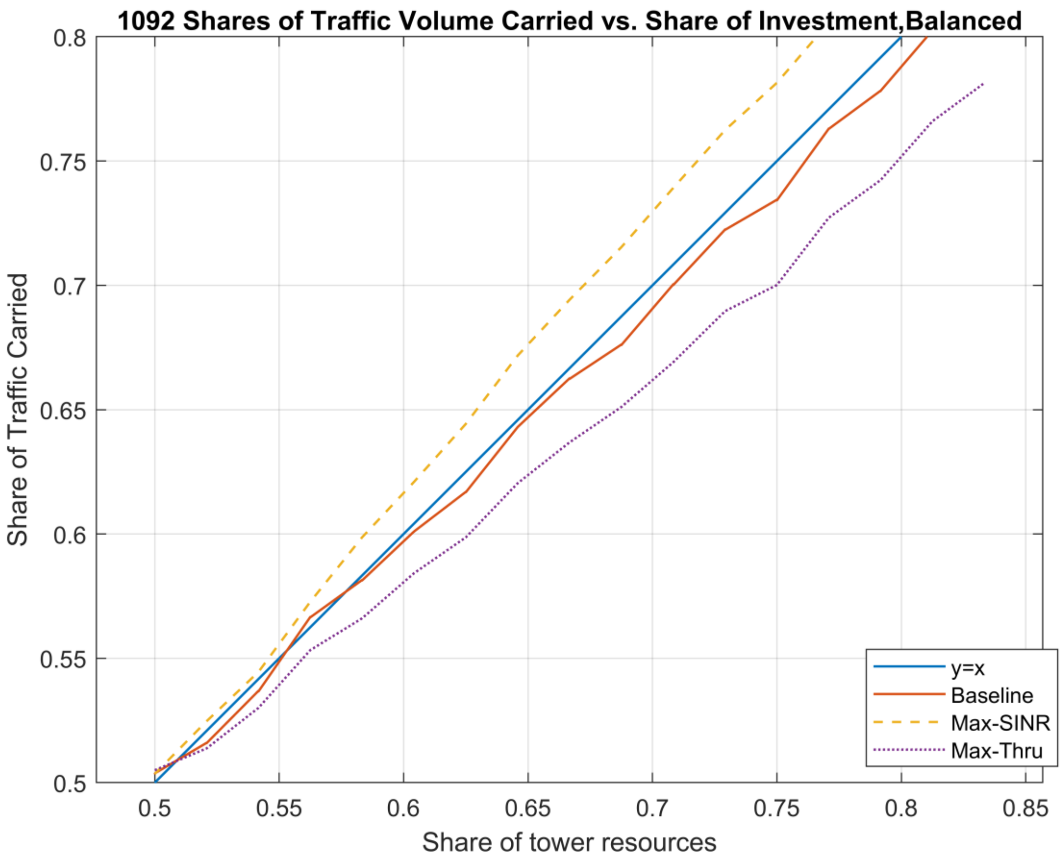

Figure 39 Share of traffic volume carried vs. share of base stations. (Partner MNOs have the same spectrum bandwidth; "Balanced" resource allocation)