UNIVERSITY OF CALIFORNIA SAN DIEGO

Revealing Regional Access Network Design Differences to Evaluate User Network Performance and Resilience

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science

by

Zesen Zhang

Committee in charge:

Professor Aaron Schulman, Chair
Professor Dinesh Bharadia
Professor KC Claffy
Professor Xinyu Zhang
Professor Jingbo Shang

2025

The Dissertation of Zesen Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

DEDICATION

To people I met and decisions I made

EPIGRAPH

The closer you look at something,
the more complex it seems.

Vinton Cerf

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I feel incredibly fortunate to have had a joyful and fulfilling PhD journey, made possible by the support of so many amazing people. When I began this journey, I was filled with uncertainty and, to be honest, some fear. Without the encouragement and guidance of these incredible individuals, achieving this level of success would have been nearly impossible, and my path would have been far more challenging. I would like to express my deepest gratitude to everyone I have met along the way, who has made this journey enjoyable and rewarding. While I may not be able to fully convey my appreciation in this acknowledgement, please know that I am truly grateful for each and every one of you.

To begin with, I would like to express my sincere thankfulness Professor Aaron Schulman for his support as my advisor. Through multiple drafts and many long nights, his guidance has been indispensable. Aaron has been not only an important advisor to me professionally but also a mentor who has profoundly impacted me on a personal level. As an academic advisor, Aaron has been a true collaborator, encouraging me to explore my own ideas and interests while actively engaging with the research alongside me. He is a hands-on mentor who is always willing to dive deep into the field, offering insightful feedback and unwavering support. His efforts in connecting me with multiple research groups and introducing me to a vast network of incredible people have been instrumental in shaping my academic journey. His forever supports and encouragement is also really crucial to me as it inspires me to always be motivated and confident. Beyond academia, Aaron has also taught me valuable life lessons. Through his guidance, I have learned about equality, how to be a genuinely kind person, and how to communicate with others with respect. I joined during wired time and nearly half of my PhD careers were working from home. Without Aaron's support, my journey would have been significantly more difficult, and I would not have experienced the smooth progression I have today. There is an old Chinese saying: 'A teacher for a day is a father for a lifetime.' Aaron has truly embodied this role during one of the most important phases of my life, teaching me not only how to be a researcher but also how to navigate adulthood with integrity and kindness.

Throughout my PhD journey, I had the privilege of collaborating with several outstanding faculty members. First and foremost, I would like to express my deepest gratitude to Professor KC Claffy. KC was involved in nearly every project I worked on during my PhD, to the point that it is almost fair to say she co-advised me. Through countless collaborations with her, I learned how to conduct Internet measurement research and analyze data effectively. Beyond academics, KC also welcomed me to nearly every CAIDA Thanksgiving and Christmas party, which gave me a strong sense of belonging and community. I am also immensely grateful to Professor Dinesh Bharadia for introducing me to the world of wireless networks. Dinesh taught me most of the fundamental wireless techniques, and his ability to explain complex wireless concepts made learning both easy and enjoyable. Additionally, he used his connections to help me promote and find use cases for the final project I developed. This support was incredibly important, as it motivated me to continue refining my system and exploring ways to improve it. I would also like to thank Professor Kirill Levchenko for introducing me to PacketLab and giving me the opportunity to work on such an impactful large-scale project. This experience helped me understand how to build large-scale systems and apply my networking knowledge in practice—something I had struggled with for a long time. My gratitude also extends to Alex Marder, Ricky Mok, with whom I closely collaborated through my PhD. Our continued cooperation and discussions throughout my PhD made this journey even more enriching. And I would also want to give thank to Geoffrey M. Voelker for his forever kindness and smile made me feel home at sysnet group. Lastly, I want to give special thanks to my manager, Deon Whitlock, at AT&T. He provided me with the invaluable opportunity to work in the telecom industry and offered tremendous support in helping me understand their access network infrastructure. His guidance broadened my perspective and deepened my knowledge of real-world networking systems.

I also want to thank for all my committee members KC Claffy, Dinesh Bharadia, Xinyu Zhang and Jingbo Shang, for their invaluable advices and comments helped shape my thesis.

My PhD careers also give me a chance to work with multiple amazing researchers and

faculties in my project. I want to express my huge thanks to Bradley Huffaker, Matthew Luckie, Ramakrishna Padmanabhan, Alberto Dainotti, Alex C. Snoeren and Kyle Jamieson for their help to improve my project. I also got chance to work multiple amazing student collaborators. I want to thank Jon Larrea, Haoran Wan, Jarrett Huddleston, Rohith Reddy Vennam, Maiyun Zhang, Yunxiang Chi, Tzu-Bin Yan, Leila Scola and Jiting Shen for their amazing efforts they put for the informative discussions on all the myriad of technical topics and the collaborative work that made my dissertation becomes possible.

I also want to thank for all the awesome labmates. I want to first specifically thank Nishant Bhaskar. He played a really import role in the beginning of my PhD careers who literally teaches me how to get along with advisor as well as discuss with me together about my research problems. I would also like to thank Moein Khazraee, Amanda Tomlinson, Alex Yen, Ish jain, Ben Du, Sumanth Rao, Thomas Krenc, Shivani Hariprasad, Audrey Randall, Stewart Grant, Yibo Guo, Chengcheng Xiang, Bingyu Sheng, Vector Li, Lixiang Ao, Gautam Akiwate, Ariana Mirian, Anil Yelam, Rob McGuinness, Alex Forencich, Raghav Subbaraman, Gavin Roberts, Max Gao, Alex Bellon, Nishit Pandya, Alex Liu, Luoxi Meng, Yutong Huang, Wenhao Chen, Wenyi Zhang for all their mindful help and discussion offered during my PhD life. I am so lucky to have you guys in my career.

I also want to acknowledge all of my friends in my personal life Haolan Liu, Meihan Li, Yizhou Shan, Hanwen Yao, Huili Chen, Zixuan Wang, Mingyao Shen, Tianyi Shan, Zhankui He, Shihan Ran, Ke Chen, Jonathan Chang, Meng Zeng, Bryant Cao. Thank you for all your accompanies during the weird COVID time and my really important five years. Life can be so much harder without all of you.

At the end, I want to thank my parents Kelong Zhang and Xiangmei Su for their endless supports. I cannot be such a person today without their love, supports and encouragement.

I want to thank my girlfriend You Li, for her unwavering support and help during my dark time.

I also want to thank my cats Pengpeng, Lily, Wangcai for their snoring and accompanying

with my life.

Chapter 3, in part, is a reprint of the material as it appears in *Internet Measurement Conference 2021*. Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, k claffy, Aaron Schulman. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the materials as it apears in *USENIX Security Symposium 2023*. Alexander Marder, Zesen Zhang, Ricky Mok, Ramakrishna Padmanabhan, Bradley Huffaker, Matthew Luckie, Alberto Dainotti, KC Claffy, Alex C. Snoeren, Aaron Schulman. The dissertation author was the co-primary investigator and author of this paper.

Chapter 5, in part, is being prepared for submission for publication of material. Zesen Zhang, Jon Larrea, Jarrett Huddleston, Haoran Wan, Ricky Mok, Bradley Huffaker, KC Claffy, Kyle Jamieson, Alexander Marder, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

Chapter 6, in part, is currently being prepared for submission for publication of material. Zesen Zhang, Rohith Reddy Vennam, Maiyun Zhang, Yunxiang Chi, Dinesh Bharadia, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

VITA

2015-2019 Bachelor of Science in Information Security, Shanghai Jiao Tong University

2019-2022 Master of Science in Computer Science, University of California San Diego

2019-2025 Doctor of Philosophy in Computer Science, University of California San Diego

PUBLICATIONS

Tzu-Bin Yan, Zesen Zhang, Bradley Huffaker, Ricky Mok, kc claffy, Kirill Levchenko, "Marionette Measurement: Experimentally Testing the PacketLab Hypothesis" In Proceedings of *Passive and Active Measurement (PAM) 2025*

Zesen Zhang, Jiting Shen, Ricky Mok, "Empirical Characterization of Ookla's Speed Test Platform: Analyzing Server Deployment, Policy Impact, and User Coverage" In Proceedings of *Computing and Communication Workshop and Conference(CCWC) 2024*

Tzu-Bin Yan, Zesen Zhang, Bradley Huffaker, Ricky Mok, kc claffy, Kirill Levchenko, "Poster: Empirically Testing the PacketLab Model", In Proceedings of ACM *Internet Measurement Conference 2023 (IMC)*

Zesen Zhang, Leila Scola, Aaron Schulman, "Investigating the Sustainability of the 5G Base Station Overhaul in the United States" In Proceedings of the *ICT for Sustainability 2023. (ICT4S)*

Alexander Marder, Zesen Zhang, Ricky Mok, Ramakrishna Padmanabhan, Bradley Huffaker, Matthew Luckie, Alberto Dainotti, KC Claffy, Alex C. Snoeren, Aaron Schulman, "Access Denied: Assessing Physical Risks to Internet Access Networks" In Proceedings of *USENIX Security Symposium 2023*

Tzu-Bin Yan, Michael Chen, Anthea Chen, Zesen Zhang, Bradley Huffaker, Ricky Mok, Kirill Levchenko, kc claffy; "Poster: PacketLab - Tools Alpha Release and Demo" In Proceedings of ACM *Internet Measurement Conference 2022. (IMC)*

Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, k claffy, Aaron Schulman, "Inferring Regional Access Network Topologies: Methods and Applications" In Proceedings of ACM *Internet Measurement Conference 2021. (IMC)*

ABSTRACT OF THE DISSERTATION

Revealing Regional Access Network Design Differences to Evaluate User Network Performance and Resilience

by

Zesen Zhang

Doctor of Philosophy in Computer Science

University of California San Diego, 2025

Professor Aaron Schulman, Chair

ISP access networks are critical to the Internet. They aggregate millions of users through diverse infrastructure (e.g. cellular base stations and fiber), and route traffic through central offices to connect to the Internet backbone. Although access networks are fundamental to Internet connectivity, their design is not standardized. Each ISP builds its access networks differently with different vendors and designs across regions, and multiple ISPs deploy distinct access networks within the same region to compete for customers. These variations can lead to differences in access network performance and resilience across providers and regions.

However, accurately measuring access network performance and resilience is a challenge

for end-users, regulators and even ISPs themselves. Several problems contributes to this difficulty. First, regional access network infrastructure often blocks external probes to prioritize customer traffic, limiting active measurements. Second, measuring wireless last-mile infrastructure requires close proximity to capture signals above the noise floor. Third, field measurements are not always reliable—ordinary customers using the service and environmental noise can significantly affect performance analysis. These problem arises from diverse components in access networks and region-specific deployment strategies. Network measurement approaches that rely on limited vantage points and inference techniques, can fail to capture regional variations comprehensively.

In this dissertation, I introduce new methodologies to overcome these challenges by using public resources to create vantage points in different regions. Through extensive field measurements, I reveal that ISPs' access networks exhibit highly diverse designs across regions, providers, and vendors. I evaluate how these design disparities impact latency, traffic performance, and network resilience. My contributions provide a foundation for improving the transparency of regional access networks and evaluating their resilience and performance.

In summary, I defend the following thesis statement: *The diversity and opacity of regional access network infrastructure hinders accurate evaluation of its performance and resilience (e.g. network outage time, downlink throughput, wireless SNR), which can be addressed through: (1) Using local public Wi-Fi networks and public transit mobile phones to reveal access network topology and assess physical risks including the root cause of access network outage, (2) using controlled mobile phone experiments to uncover LTE base station scheduler design variations across vendors and evaluate the impact on downlink throughput, and (3) integrating mobile devices with software-defined radios to evaluate wireless signals under a variety of interference in the field.*

Chapter 1

Introduction

Internet service provider's access networks are one of the most essential components in the Internet. They bridge millions of last-mile users to access core Internet infrastructure. ISPs strategically aggregate last-mile customers by utilizing various infrastructures, such as cellular base stations and fiber networks. All access network traffic is routed through their central offices (COs) (Fig. 1.1) to reach the Internet backbone. Access networks must be carefully designed to balance the trade-off between performance, reliability, and cost.

Despite its critical role in providing last-mile customers access to the Internet, the design of access networks is neither standardized nor unified. First, building an access network involves multiple components, including base stations, switches, transponders, routers, and fiber infrastructure. These components are developed by various vendors. Different components from different vendors exhibit varying performance characteristics, which can impact overall network performance. Second, ISPs structure their access networks differently across regions based on the number of customers and cost considerations. This includes variations in network topology and the selection of components used in deployment. Third, multiple ISPs may operate in the same region, each independently designing and building its own network. As a result, regional access networks in the same area can differ in design, leading to local variations in performance and resilience. Accurate assessment of these factors is important for ISPs, customers, and regulators, as it helps understand differences in performance and reliability of network services and can help

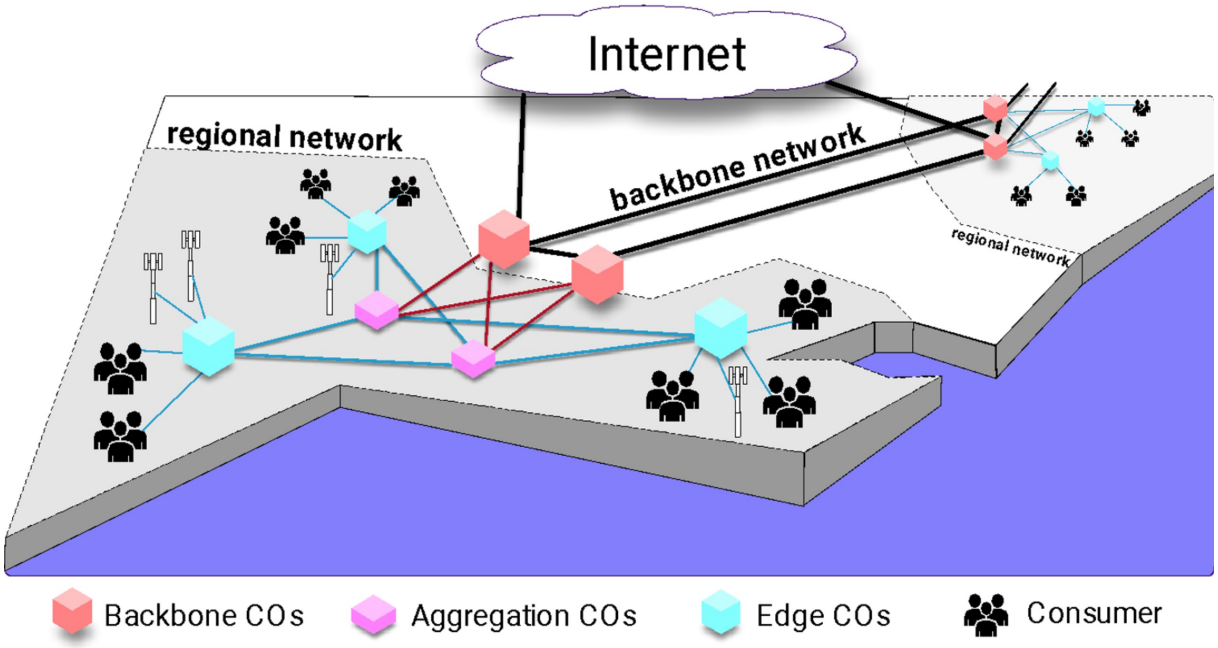


Figure 1.1. An Internet Service Provider (ISP) network can be divided into backbone and regional access network.

customers select a product that has the overall network quality that they require.

However, accurately measuring access networks is challenging, primarily due to their lack of transparancy—not only to third parties such as customers and regulators, but even to ISPs themselves. The lack of transparency exists mainly due to three factors:

Complexity of Network Infrastructure. Access networks consist of millions of interconnected components, including base stations, switches, routers, transponders, and fiber links. These components are developed by a wide range of vendors, such as Cisco, Juniper, Ericsson, and Samsung. While devices from different vendors are designed to serve similar purposes, their real-world performance can vary significantly. Even for ISP, despite deploying and operating the infrastructure, can struggle to fully assess their own operational performance.

Heterogeneity of Network Architecture. Access networks are proprietary and their designs vary significantly across regions. While many components adhere to standardized protocols and policies, their overall architecture can differ among providers and regions.

Evolution of Access Networks. As Internet performance demands rapidly increase, it

is difficult for access networks to keep pace with the evolution. While ISPs may design their networks effectively at the outset, they often implement partial upgrades when adopting new infrastructure rather than fully optimizing their full access network. This piecemeal approach can result in heterogenous networks, which are more difficult to measure because they consist of many generations of components that are interoperability.

These factors drive the per-region and per-provider differences in access network performance and resilience. Previous studies [205, 17] have assumed uniform behavior of the access network infrastructure within providers and regions. Users, although, may not experience significant differences when switching between across different access networks, the incorrect assumption of heterogeneity can make research results not universally applicable on many different access networks. To accurately analyze and understand access networks, each region must be studied separately, and any performance related strategies must be tailored to each network's specific characteristics.

Accurately inferring and analyzing access networks is critical for:

- Enhancing network resilience against failures and disasters.
- Reducing latency and improving Quality of Service (QoS) for end-users.
- Informing infrastructure investment decisions.
- Understanding the role of different ISPs and vendors in network performance.

Therefore, a comprehensive understanding of the structure and behavior of these networks is crucial for evaluating and improving their resilience, performance, and efficiency. In the next section, I introduce the challenges associated with measuring these propensities of access networks.

1.1 Challenges

Understanding and measuring regional access networks present significant challenges due to opacity of regional access network, data collection barriers, and the scarcity of vantage points. One of the primary difficulties is that access network infrastructure is typically neither visible nor measurable remotely. ISPs will constrain probing of their access networks infrastructure to only respond to their customers. Several factors contribute to this challenge:

- **Opaque Network Architectures**: Many ISPs provide limited visibility into their internal routing, making it difficult to validate inferred topologies.
- **Active Measurement Limitations**: The reliance on traceroute and alias resolution techniques for topology measurement introduces potential inaccuracies, especially when dealing with MPLS tunnels or stale infrastructure DNS records.
- **Wireless Services Range Constraint**: Wireless signals attenuate over distance. For instance, cellular wireless access network infrastructure is not able to serve more than a 10km range. Therefore, we need many local wireless vantage points to capture differences in infrastructure.

However, achieving this presents additional challenges:

- **Regional Presence**: Effective measurement of regional access networks requires deploying vantage points that are physically located within the network region and operating under the specific ISP's service.
- **Sufficient Number of Vantage Points**: Relying on data from only one or two vantage points in a region introduces significant biases and fails to capture the full picture of access network behavior.

- **Protocol Flexibility**: Access networks employ multiple protocols for service delivery. Vantage points must be capable of supporting measurements across different protocols to provide a comprehensive analysis.
- **Data Accuracy**: Wireless measurements are sensitive to signal-to-noise ratio (SNR). Vantage points must be positioned close enough to signal sources while avoiding interference to ensure accurate data collection.

In addition to visibility issues, the instability and heterogeneity of access network infrastructure across different regions and vendors creates further challenges. Large-scale data collection is essential to account for these variations, but scalability remains challenging:

- **Dynamically Changing Network Conditions**: Real-world measurement is susceptible to network variations. To reduce inaccuracies from outliers we need to perform multiple repeated measurements of each regional access networks.
- **Heterogeneous Aggregation Strategies**: ISPs structure their networks differently according to many factors including number of customers and service priority they need to provide in a region, making it difficult to generalize findings across regions. We must measure each region individually to understand its behavior.
- **Vendor-Specific Policies**: The lack of standardization in specific performance decision areas among access network infrastructure introduces performance discrepancies that affect end-user experience.

These challenges highlight the need for improved methodologies in access network measurements, motivating this dissertation's focus on scalable, accurate, and vendor-agnostic approaches to regional access network measurement.

1.2 Thesis

Although the above challenges make it difficult to answer the question about how to accurately evaluate strategies that providers and vendors use to build regional access network infrastructure. This thesis illustrates that the significant heterogeneity in access network deployments across different regions and vendors can be resolved by deploying a sufficient number of vantage points in each region. I demonstrate that it is possible to reveal regional access network infrastructures differences and to assess their performance and resilience.

To achieve this, I combine existing topology measurement techniques with new access network-specific probing and analysis techniques to perform large-scale field studies of real-world regional access network infrastructure. First, I describe how I combine traceroute with public Wi-Fi networks to conduct active measurement (McTraceroute), and place mobile phones inside ground shipping boxes to automatically send measurements across a wide areas(ShipTraceroute). I explain how these vantage points enable revealing ISP's regional access network topologies and assessing regional access networks for physical risks to reliability. Next, I introduce how network performance measurement techniques under the controlled measurement setup can reveal how vendors developed different LTE base stations scheduling strategies. I also show how the scheduling strategy differences affect download throughput for users. Lastly, I introduce how I integrate software-defined radio with portable mobile devices (MobileSDR) to capture local wireless signals from regional access network infrastructure and evaluate their performance under a variety of interference in the field.

In summary, I defend the following thesis statement: *The diversity and opacity of regional access network infrastructure hinders accurate evaluation of its performance and resilience (e.g. network outage time, downlink throughput, wireless SNR), which can be addressed through: (1) Using local public Wi-Fi networks and public transit mobile phones to reveal access network topology and assess physical risks including the root cause of access network outage, (2) using controlled mobile phone experiments to uncover LTE base station scheduler design variations*

across vendors and evaluate the impact on downlink throughput, and (3) integrating mobile devices with software-defined radios to evaluate wireless signals under a variety of interference in the field.

1.3 Contributions and Organization

The remainder of this dissertation is organized as follows.

In Chapter 2, I first introduce the existing measurement platforms and commercial datasets related to access network performance and resiliency. I show the benefits of using these platforms and datasets but I also explain why they are not enough to measure and evaluate regional access network. By doing so, I emphasis the importance of combining existing techniques with new access network-specific probing and analysis techniques to understand regional access network.

In Chapter 3, I introduce novel measurement techniques to widely reveal access networks, including McTraceroute, which leverages fast-food restaurant (e.g. McDonald's, Starbucks) Wi-Fi networks to measure regional access network topology, and ShipTraceroute, which places mobile phones inside shipping boxes to issue measurements across a wide area during ground transportation. I show that by using a sufficient number of vantage points, we are able to reveal ISP's wireline and wireless regional access network topology. Our findings reveal that network topology varies significantly among providers, and even within the same provider across different regions. As a result, we observe that network latency is not solely determined by geographical distance but also by detours within the network topology.

In Chapter 4, we discuss the security risks inherent in the network topologies observed across different ISPs. Our findings indicate the existence of severe vulnerabilities within these networks that could disrupt customer services. We analyze these risks in detail and propose mitigation strategies to enhance network security and resilience.

In Chapter 5, I explore the differences in strategies employed by vendors when developing access network infrastructure. Specifically, I analyze LTE base station scheduling strategies

across different vendors, demonstrating that LTE base stations from different manufacturers adopt distinct scheduling methodologies that may impact network performance.

In Chapter 6, I then introduce MobileSDR, a wireless measurement platform designed to separate the task of designing signal processing measurement programs from the process of data collection.

Chapter 3, in part, is a reprint of the material as it appears in *Internet Measurement Conference 2021*. Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, k claffy, Aaron Schulman. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the materials as it apears in *USENIX Security Symposium 2023*. Alexander Marder, Zesen Zhang, Ricky Mok, Ramakrishna Padmanabhan, Bradley Huffaker, Matthew Luckie, Alberto Dainotti, KC Claffy, Alex C. Snoeren, Aaron Schulman. The dissertation author was the co-primary investigator and author of this paper.

Chapter 5, is currently being prepared for submission for publication of material. Zesen Zhang, Jon Larrea, Jarrett Huddleston, Haoran Wan, Ricky Mok, Bradley Huffaker, KC Claffy, Kyle Jamieson, Alexander Marder, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

Chapter 6, in part, is currently being prepared for submission for publication of material. Zesen Zhang, Rohith Reddy Vennam, Maiyun Zhang, Yunxiang Chi, Dinesh Bharadia, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

Chapter 2

Limitations of Existing Platforms and Datasets

In this chapter, I provide an overview of related access network measurement platforms and commercial measurement datasets. I will describe the advantage and shortage of these platforms in measuring access networks in a large scale.

The first key requirement is having a sufficient number of Vantage Points (VPs) within the corresponding ISP in each region to have visibility of all the regional access networks infrastructure. Previous work place static bare metals into regions which made it hard to scale into multiple areas and share endpoints. In my work, I instead deploy software-based measurement platforms and mobile platforms to generate many VPs in and across different access network regions.

Access network measurement requires low-level detailed metrics to understand performance and reliability that are not typically available in prior Internet measurements. For instance, cellular and LoRA network analysis requires raw IQ samples to capture physical layer wireless link behavior.

Measurements also need to be annotated with geolocation information. Geolocation tagging provides context to which regional access network, and where inside that access network (e.g., which CO) the measurements correspond to.

Access networks typically are only measured with speed test downlink and uplink through-

put. However, they only provide end-to-end results, providing limited internal infrastructure visibility. To gain internal observability, we need probes that can cause internal infrastructure to reply, like traceroutes, and also we will need to resolve missing information in these replies using techniques often used in Wide Area Networks, like IP alias resolution. Therefore, access network measurement needs platforms that offer flexible probing and capture capabilities.

Historical measurements are particularly important for observing access network reliability. The reason is that one of the common faults in an access network is an outage of one or more links. With historical measurements we can compare the network topology and behavior before, after, and during the fault, to observe the network's resilience.

Table 2.1 presents an overview of the features supported by different network measurement platforms and datasets. In general, I categorize these platforms into three types: (1) server-oriented wired platforms, (2) user-oriented wired platforms, (3) wireless measurement systems and (4) commercial datasets.

- **Server-oriented Wired Measurement Platforms** function as clients that allow researchers to send active measurement packets to network infrastructure.
- **User-oriented Speed Test platforms** operate as servers that passively receive packets from users.
- **Wireless Measurement Systems** are designed to capture and analyze wireless signals in the field.
- **Commercial Datasets** are data collected and organized by commercial parties and publish online.

Additionally, several commercial datasets provide infrastructure-related information. Table 2.1 demonstrates that no single platform fully supports all the key features required for regional access network measurement. In the following sections, I will examine each platform in detail and discuss its limitations.

Table 2.1. Comparison of network measurement platforms, speed test platforms, wireless measurement platforms, and commercial datasets based on key features. ✓ indicates it has the corresponding features while ✗ means negative. ^is suggesting the platform has related features but not enough for revealing access networks. – is showing the corresponding feature is not indicated.

Platform Name	VPs in Regions /ISPs	Software Based	Raw Data	Geo-location Info	Perfor-mance Metrics	Flexible Tests	Public Datasets	Mobile
Server-Oriented Wired Measurement Platforms								
RIPE Atlas	^	^	✗	✓	✗	^	✓	✗
Ark	^	✗	✓	✓	✗	✓	✗	✗
BISmark	^	✗	✗	✓	✓	^	✓	✗
ICLab	✗	✓	✓	✓	✓	✗	✓	✗
Dasu	^	✓	✗	✓	✓	✓	✗	✗
Netalyzr	^	✓	✗	✓	✓	^	✗	✗
MITATE	^	✓	✗	✓	✓	^	✗	✓
Scriptroute	^	✓	✗	✗	✓	✓	✗	✗
User-Oriented Speed Test Platforms								
M-Lab	✗	✗	✓	✗	✓	✗	✓	✗
Comcast	✗	✓	✗	✗	✓	✗	✗	✗
Cloudflare	✗	✗	✗	✗	✓	✗	✗	✗
Fast.com	✗	✗	✗	✓	✓	✗	✗	✗
Ookla	^	✗	✗	✓	✓	✗	✓	✗
Wireless Measurement Platforms								
WebSDR	^	✗	✗	✓	✗	✗	✗	✗
Electrosense	^	✓	✓	✓	✗	✓	✓	✗
CloudSDR	✗	✗	✗	✗	✗	✗	✗	✓
LS Observer	^	✗	✗	✗	✓	✗	✗	✓
SpecNet	✓	✗	✗	✓	✓	✗	✗	✗
FieldFox	✗	✗	✗	✗	✓	✗	✗	✓
MaiaSDR	✗	✗	✓	✗	✗	✓	✗	✓
Android GR[1]	✗	✓	✓	✗	✓	✗	✗	✓
MagicSDR	✗	✓	✓	✗	✗	✗	✗	✓
Commercial Datasets								
CellInfo DB[2]	–	–	✗	✓	✗	–	✓	✓
WiGLE	✓	✗	✗	✓	✗	✗	✓	✓
PeeringDB	–	–	✗	✗	✓	✗	✓	✗
Rapid7	–	–	✗	✗	✗	✗	✓	✗

2.1 Server Oriented Platforms

Server oriented measurement has been an active area of research, with several well-established platforms include RIPE Atlas[141], BISmark[169], FCC's Measuring Broadband America (MBA)[1], CAIDA Ark[30], ICLab[81], Dasu[148], Netalyzr[90], MITATE[67], Script-

[1] Android GR refers to Android GNU Radio.

[2] CellInfo DB includes multiple different datasets like OpenCellID, GeoTel, Cellinfo, NetMonster

route[164], PEERING[153], and CAIDA Periscope[86]. These platforms have significantly contributed to network measurement research, but they exhibit fundamental limitations when it comes to measuring access networks.

Many of these platforms were developed to address specific research needs rather than to provide a comprehensive view of regional access networks. For instance, RIPE Atlas[141] was primarily designed to support the operational needs of the RIPE community. Even though they have multiple VPs inside of their platforms but they are from multiple random providers and some are from cloud providers instead of resides in access networks. And they are lacking of VPs in each region as well. CAIDA's Ark[30], infrastructure serves as a community platform for active Internet measurement. BISmark[169] and MBA[1] were specifically developed for broadband speed measurements, making them unsuitable for broader access network analysis. While some platforms, such as Ark, BISmark, and MBA, allow researchers to conduct arbitrary experiments, they require extensive vetting and direct involvement from platform operators, making it difficult to conduct large-scale or customized measurements. The challenges of experiment development, testing, and deployment further complicate their use for access network research.

For example, Scriptroute[164] allows researchers to execute scripts on measurement endpoints but applies strict local policy filters to limit the types of traffic an endpoint can generate. Similarly, RIPE Atlas restricts measurements to a predefined set of tests such as ping, traceroute, DNS queries, SSL/TLS, and limited HTTP interactions. While this conservative approach enables large-scale deployment (with nearly 10,000 active endpoints), it prevents researchers from conducting in-depth measurements of access network infrastructure, where more flexible data collection is required.

Moreover, measurement platforms often struggle with portability and integration issues. For instance, when researchers attempted to extend an inter-domain congestion measurement experiment originally designed for CAIDA Ark to BISmark, significant redesign was required due to fundamental differences in platform architectures. The inability to seamlessly transfer experiments across platforms hinders large-scale access network measurements, which require a

diverse set of vantage points distributed across different ISPs and regions.

Additionally, several research platforms have focused on minimizing endpoint complexity while allowing for limited customization. MITATE[67], for example, targets mobile network measurements but restricts experiments to predefined sequences of packets with specific timing, ensuring that they do not introduce security risks. While this approach provides controlled experimentation, it lacks the flexibility needed to probe deeper into access network structures, which often require adaptive measurement techniques to account for heterogeneous network deployments and vendor-specific configurations.

Efforts have been made to unify measurement infrastructures, such as the MPlane and Tophat projects. These initiatives aimed to create standardized measurement interfaces, shared data storage formats, and unified toolsets. However, they primarily focus on inter-domain and backbone network measurement rather than access networks. The fundamental challenge remains: most existing platforms are designed for general-purpose Internet measurement rather than for the specific task of mapping and analyzing regional access network infrastructure.

2.2 User-oriented Platforms

The most common user oriented platforms for access network performance measurement are speed test measurement platforms. Speed test platforms all consist of many geographically distributed test servers. They differ in how close servers are to their users, both in terms of network topology and geography. Speed test platforms typically consist of many geographically dispersed test servers, strategically positioned to reduce the latency and shorten the network paths to end-users. Various platforms adopt distinct strategies for building their test infrastructure. For example, M-Lab [103] and Comcast speed test [44] install bare metal machines dedicated for this purpose. Speedof.me [4], Cloudflare speed test [42], and fast.com [57] leverage existing content delivery networks (CDNs). In most cases, the speed test platform has control over the underlying (virtualized) hardware or the network connectivity. Ookla, currently hosting the largest number

of speed test servers, employs a crowdsourcing model to grow its fleet of test servers. Instead of deploying boxes to different locations, network operators use their own hardware to install Ookla's software and submit requests to join the Ookla platform [129].

However, the limitations of these platforms are they only provide end-to-end performance results without internal infrastructure visibility. Ookla for example, masks the critical information in the measurement results, such as geolocation, IP addresses and anonymize downlink and uplink throughput results, in order to help keep privacy for users. Comcast or Speedof.me, they never releases any of their measurement results to public. M-Lab is the only platforms provide many detailed information of the results they got from users. However, M-Lab relies on random users using their services and their servers are outside of access network. It lacks of flexibility for researchers to further understand detailed infrastructure design of access network.

2.3 Wireless Measurement Platforms

There are several wireless platforms that can collect data from different locations. For example, WebSDR [194], allows users to connect SDRs deployed at many locations to the internet, so many users can collect wireless measurments from many locations. Electrosense [189] leverages crowdsourcing efforts with inexpensive commodity radio hardware to sense the spectrum in densely populated regions of the world. However, these platforms are primarily used for viewing signal spectrograms and do not allow customization of data analysis on the software side.

CloudSDR [173], on the other hand, permits the remote execution of GNU Radio programs with SDRs connected to cloud-based devices. Spectrum Observer [217] utilizes SDR to observe and analyze wireless radios in the field and uses a co-located PC to detect radio signals at different frequency bands. However, the endpoints in these platforms are not to move to specific location and provide data collection according to researchers' demands.

Some mobile wireless measurement systems like FieldFox [88] and TinySDR [72] require

specially designed programs and hardware to collect specific field data. They mainly focuses on hardware by designing them to be used in specific field like transmit and receive IoT packets or cellular signals, but offer limited flexibility.

Several mobile wireless measurement platforms exist for Android devices that can collect IQ samples and view the spectrum, such as MaiaSDR [109], the Wideband SDR tool [94] and SDRs that are wirelessly connected with phone through bluetooth [161], but they do not support packet detection or signal processing. There are some Android-based SDR platforms that allow arbitrary signal analysis using GNU Radio [23] and MagicSDR [106]. However, they only allow users to decode specific signals like FM, AM, and Wi-Fi using pre-designed apps. They do not provide researchers with the flexibility to run their own programs and require expertise to respond appropriately to the observed spectrograms.

2.4 Commercial Datasets

Several commercial datasets are publicly available for Internet measurement. However, regional access network measurement requires significantly more detailed information than these datasets provided. This made it difficult to draw conclusive results from existing datasets.

Multiple datasets provide geolocation information about wireless infrastructure. One example is the OpenCellID dataset[187], a collaborative community project that collects GPS coordinates of cell towers and their corresponding location area identity. Base station locations can be identified using MCC (Mobile Country Code), MNC (Mobile Network Code), LAC (Location Area Code), and Cell ID. The dataset contains over 35.5 million unique cells and more than 2.1 billion unique measurements. GeoTel[64] offers a comprehensive dataset of approximately 525,000 cell tower sites, enabling users to identify and assess wireless infrastructure, including towers, rooftops, water tanks, billboards, and rural land. This information facilitates the deployment and optimization of wireless coverage solutions. WiGLE[199] (Wireless Geographic Logging Engine) is a website that collects data on wireless hotspots worldwide, including GPS

coordinates, SSID, MAC address, and encryption type. In addition to Wi-Fi networks, WiGLE also gathers cell tower data, providing a broad perspective on wireless network deployments.

However, these datasets primarily provide location information along with basic identifying details about the infrastructure. Most are based on passive, automated scanning of nearby infrastructure through mobile devices, relying on system information without access to raw signal data. Additionally, they do not include performance-related data on last-mile infrastructure. To evaluate performance, measurement devices must actively interact with infrastructure or collect raw sample data. In Chapters 5 and 6, we demonstrate how we assess last-mile wireless infrastructure performance.

For wired network infrastructure, several datasets provide information about access networks. PeeringDB[95], for instance, is a freely available, user-maintained database that facilitates global interconnection at Internet Exchange Points (IXPs), data centers, and other interconnection facilities. It provides details about colocation facilities, networks, and IXPs, serving as a valuable resource for understanding interconnection dynamics.

These datasets have been useful in our research, such as leveraging reverse DNS (rDNS) names from the Rapid7 Open Dataset. However, a major limitation of these wired network datasets is their generality. Extracting performance and resilience insights directly from them is challenging because access networks are highly opaque and region-specific. Without extensive active measurements targeting specific locations and vantage points, general-purpose datasets cannot provide the necessary depth of information on access network performance and resilience.

In conclusion, while we utilize commercial open datasets as part of our access network measurements, they do not have enough detailed data to accurately evaluate access network performance and resilience.

Chapter 3

How Access Networks are Different Across Different Providers and Regions

3.1 Introduction

ISP regional access networks are an essential component of an ISP's infrastructure: they bridge millions of users' last-mile access links to the ISP's nearest backbone routers, which may be hundreds of miles away, to reach the Internet. Access networks strategically *aggregate* traffic in order to balance reliability and performance against the cost of providing connectivity over large regions (Figure 1.1). Fiber cuts or other hardware failures can lead to large-scale outages spanning neighborhoods, counties, or entire states. Risk of outages motivates ISPs to provide *redundancy* within and across levels of aggregation. But regional access networks are remarkably opaque, which makes it challenging for academics to quantitatively study their role in the continually evolving ecosystem.

We present a measurement-driven exploration of regional access network topologies, through the lens of aggregation and redundancy as the foundations of scalability. Building on advances in Internet measurement methods and tools over the last two decades, we first establish and demonstrate the ability for an independent third-party to infer the topologies of different regional access networks, including aspects of the underlying physical (layer-1) topology, using only active measurements. We then perform measurement campaigns to infer and compare how major U.S. wireline (Comcast, Spectrum, AT&T) and mobile (AT&T, Verizon, T-Mobile) ISPs

incorporate aggregation into their regional access network topologies. We show how analyzing these topological differences across providers and even across regions of the same access ISP can yield insights into the propagation of large correlated last-mile link failures [155, 59], sources of edge computing latency [150] and how to minimize it [135], performance limitations of metro-area fiber networks [110], and the evolving Internet ecosystem [54].

Our methodology synthesizes the state-of-the-art in three dimensions of Internet router-level topology analysis: combining traceroute paths from distributed vantage points; extracting semantics from DNS hostnames and IPv6 addresses of observed topology; and IP address alias resolution to further refine topology inferences. Our methodology leverages ideas that are well-established in the Internet measurement community, but we used them in different combinations, and with creative refinements, to accommodate the geographic scope, scale, and architectural richness of U.S. regional access ISPs. One of our contributions is effectively a recipe book of how to gain insight into network topology structure under different sets of constraints.

For example, the largest U.S. cable providers today tend to have near-universal reverse DNS on their router IPs, but hostnames are often stale. We devised heuristic methods to filter out stale or misleading hostnames, facilitating a comprehensive mapping of their regional topologies.

Networks without geographically meaningful internal hostnames require another way to infer geographic coverage of routers in the Central Offices (COs), such as probing from many geographically distributed VPs in the regional network. This requirement highlights the most common challenge in inferring internal network topologies, well-known to the research community: many networks provide more (or more accurate) visibility to internal vantage points than to external ones, especially for mobile networks. For wireline networks, we obtained internal vantage points by wardriving public WiFi networks in fast-food chains.

Mobile (cellular) access networks present many challenges: e.g., limited rDNS and blocking external probing. To capture this topology, we used IPv6 address structure and a new approach to gain county-wide internal visibility into mobile networks: cross-country shipping of mobile phones while they actively perform energy-efficient network measurements.

Our measurement methodology contributions are:

- Cost-effective approaches to procuring vantage points; energy-efficient approaches to sustaining mobile vantage points while devices are in transit.
- Analysis of DNS hostnames and IPv6 addresses to infer and geolocate topology, as well as strategically select probe targets to fill in coverage gaps.
- Heuristics that leverage active measurement techniques to filter noise (e.g., stale DNS information) or erroneously inferred hops, and infer missing (e.g., non-responding) IP hops.

Our empirical contributions result from applying our methods to previously unmapped parts of the infrastructure: wireline and mobile regional access networks. We gather enough data to ground the following discoveries:

- Topological redundancy—a metric of resilience—varies widely within and across levels of the hierarchy.
- Layer 3 topology information, including hostnames and IPv6 addresses, can reveal building locations and building-level redundancy within access networks.
- Regional access networks leverage a range of aggregation strategies to accommodate diverse markets, environments, and technologies. One result is substantial disparity in latency from some Edge COs to their Backbone COs.

These measurements inform analysis of critical infrastructure, including resilience to disasters, persistence of digital divide, and challenges for edge computing.

3.2 Background

Like any network, regional access networks must balance reliability and performance against the cost of deploying and operating an infrastructure. For these networks, an additional

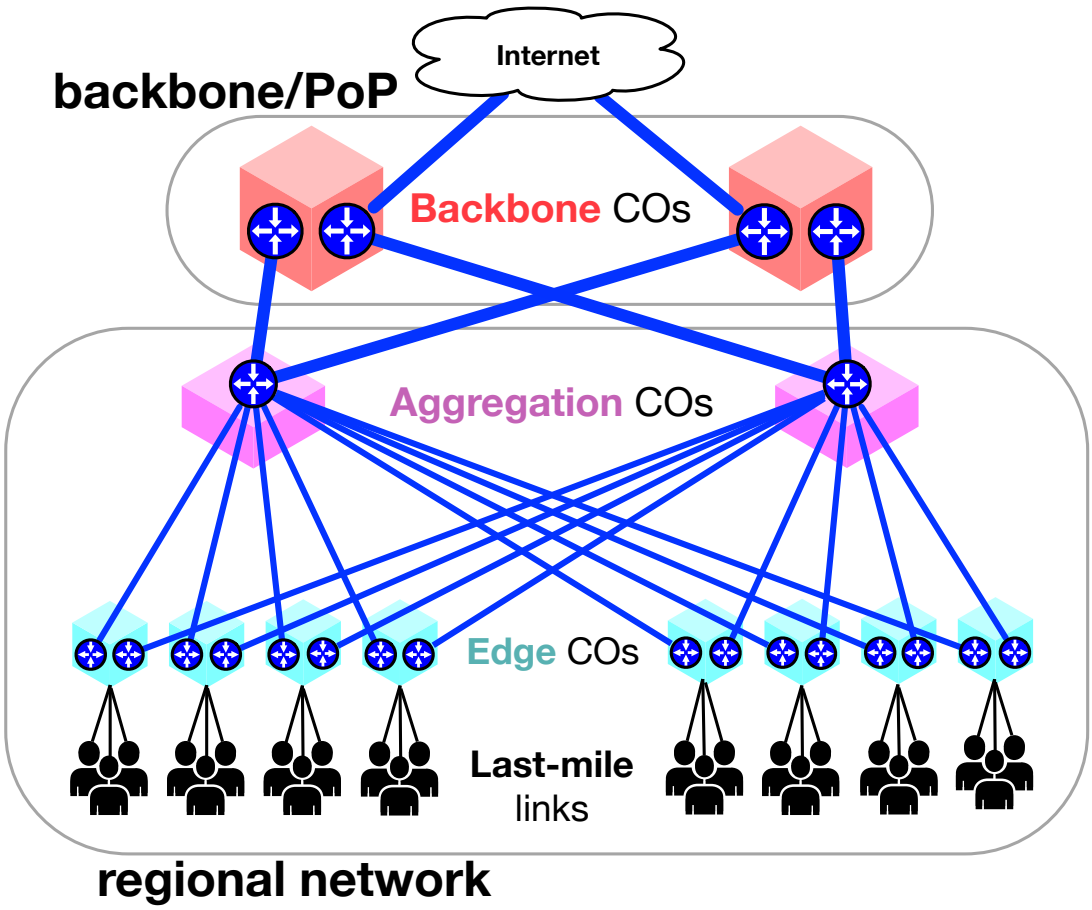


Figure 3.1. Routers in EdgeCOs aggregate users while routers in AggCOs aggregate EdgeCOs.

challenge is achieving all these aspirations at scale—in terms of market size and geographic scope—and adapting to evolution of technology and industry structure. The primary architectural mechanism used to achieve this scale is *aggregation*. Regional networks aggregate traffic in *Central Offices (COs)* through a hierarchy of routers (and switches). An *Edge CO* (or EdgeCO) aggregates traffic from many thousands of last-mile links. Similarly, an *Aggregation CO* (or AggCO) houses routers to aggregate traffic from dozens of EdgeCOs, often across metropolitan areas or entire states. *Backbone COs* (or BackboneCO) house equipment to aggregate traffic from AggCOs—and sometimes EdgeCOs, and provide transit services either via the access network provider's backbone network, or via other ISPs. Figure 4.1 illustrates how ISPs use *redundancy* across layers of aggregation to provide resiliency in case of link or node failures.

The topology of this regional network infrastructure frames the performance and resilience of the networks, as traffic must cross the regional access links to reach the nearest Internet PoP. If the nearest PoP is far away, users may experience significant minimum latency [144]. The level of redundancy in any component of the network similarly provides an upper bound on robustness in different parts of the network.

3.2.1 Evolution of access networks

We provide some historical background for context on the challenges and opportunities for measurement of these networks to study their performance and reliability.

A typical access network is physically constructed of several fiber rings (Figure 3.2). These networks generally use three hierarchical rings. Last-mile links (e.g., Cable/Passive Optical/DSL) are aggregated over *Last-mile Loops* that reach into neighborhoods and terminate at EdgeCOs. EdgeCO traffic is aggregated in *Edge Rings* that terminate at one or more AggCOs. Then AggCO traffic is aggregated on *Core Ring* to BackboneCOs. Early Internet access network architectures used shared SONET on these fiber rings. All traffic passed through every CO in reserved time slots to reach aggregation points at the higher layer, which resulted in suboptimal bandwidth and latency performance. In the 1990s, in response to exploding demand for Internet

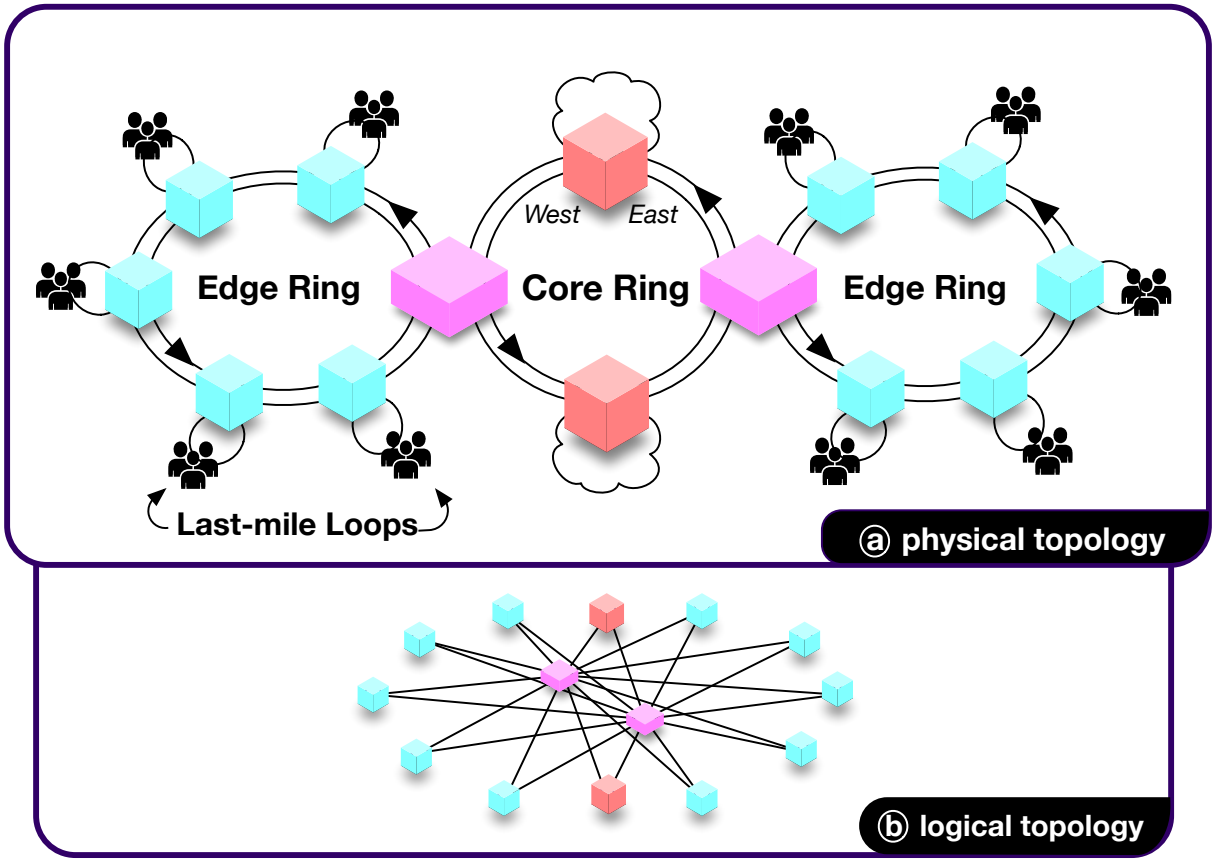


Figure 3.2. (a) Access networks are physically constructed of a hierarchy of fiber rings. (b) Overlapping Ethernet star topologies are built on top of these rings.

bandwidth in part due to IP convergence (moving all voice and data services to use IP), access networks gradually replaced these SONET ring topologies with Ethernet-based star topologies. Often repurposing the same physical plant that SONET used, ISPs deployed direct Ethernet links between EdgeCOs, AggCOs, and BackboneCOs using techniques such as Dense Wave Division Multiplexing (DWDM) [212].

With DWDM and the addition of reconfigurable optical switches, operators are now capable of flexibly configuring layer-1 topology on these physical fiber rings. However, there is no standard way to design such a topology, i.e., how many or which EdgeCOs connect to how many or which AggCOs. Designs are driven by demographic (population), geographic (or geological) and regulatory constraints, especially when crossing state boundaries. To improve resilience or accommodate high-traffic regions, some components may use a full mesh (e.g., between the backbone and aggregation layers in Figure 4.1) rather than star topology. These design choices have implications for accommodating future services, e.g., low-latency high-bandwidth edge services in COs [135].

Another driving force in network design is reliability and robustness in the face of inevitable failures of components. SONET was especially robust to fiber cuts and router failures, since traffic could travel in either direction around the ring. The Ethernet star topology does not have this feature. To compensate, ISPs add redundant routers and/or links over existing fiber rings, creating "dual ring/star" topologies (Figure 3.2).

Each network independently chooses how to implement redundancy: adding routers or links or entire COs, and within or across different levels of aggregation. Some regions connect with only one backbone CO; others have only one AggCO. Many COs have redundant routers, and fiber rings have inherent redundancy (i.e., "East" and "West" directions in Figure 3.2). Redundant backbone COs can dramatically improve regional reliability given that long-distance backbone also can be prone to failure [54, 65].

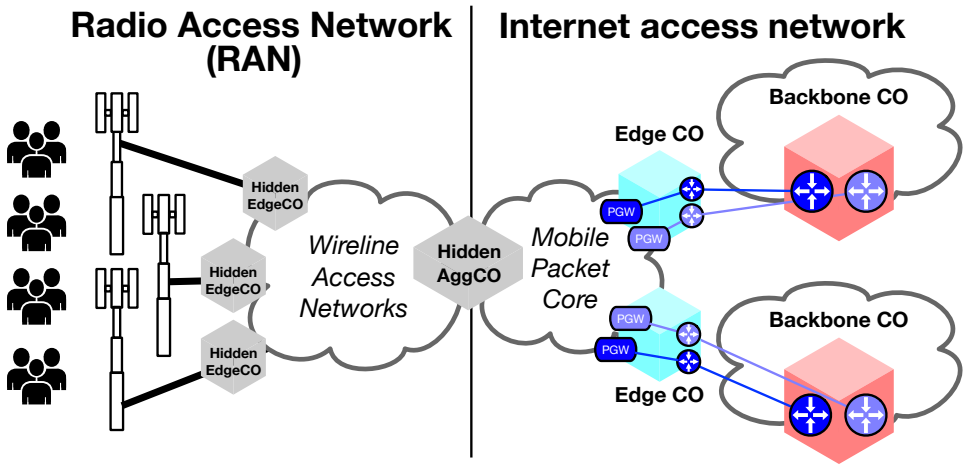


Figure 3.3. Mobile access networks are built of a combination of wireline access networks and mobile-specific networks.

3.2.2 Mobile access networks

Mobile regional access networks leverage wireline access networks to provide mobile Internet access over a large geographic area. Mobile access networks consist of two halves (Figure 3.3). Mobile devices communicate wirelessly with base stations, and the base stations aggregate user traffic over existing wireline networks and hidden mobile-specific AggCOs that connect to one or more EdgeCOs; these facilities can also be called Mobile Telephone Switching Offices (MTSO), Mobile Switching Centers (MSC), or mobile datacenters. This overlay network is called the Radio Access Network (RAN). These AggCOs serve as a bridge to connect the RAN to the mobile "Packet Core" which terminates the mobile network at one or more Packet Gateways (PGW) in an EdgeCO. Then these EdgeCOs connect directly to one or more regional BackboneCOs to connect with the rest of the Internet. EdgeCOs in mobile networks are the primary location where low-latency high-bandwidth IP-based edge services can be deployed; Verizon is already deploying edge services at these COs [49].

All mobile Internet traffic traverses both the wireline regional access networks described above and the mobile packet core. Therefore, to understand the aggregation and redundancy of mobile networks we need to also understand the wireline network. Unfortunately, the RAN—

and therefore the wireline access network—is not visible from probes sent by mobile devices. However, by observing the topology of wireline providers' regional networks using their wireline last-mile links, we get insight into the limitations of some of the RANs. Indeed, in this paper we mapped both the wireline and mobile networks of AT&T, and the cable providers we mapped provide backhaul for all major mobile carriers [183].

There can be significant differences in the topology of mobile access networks because each provider can make their own decision about how to aggregate their traffic to BackboneCOs. Indeed, they have significant flexibility because mobile networks are designed to be an overlay on other networks. The primary factors affecting mobile network topology are traedeoffs in performance, economics, and reliability.

3.3 Related work

Mapping Wired Networks: In 2002, Spring *et al.* developed Rocketfuel to construct router-level maps of individual networks using focused traceroutes, alias resolution, DNS hostnames, and BGP routing tables [120], and used it to map ten transit networks. Researchers used the Rocketfuel maps to study, interior routing [107], path inflation [165], and the maps sparked lively methodological discussions (e.g., [177, 200]). In 2007, Mao *et al.* developed NetworkMD [111], an approach to infer *failure groups* in the last-mile layer-1 network topologies—i.e., devices such as repeaters whose failure impact downstream modem connectivity through topological dependencies. In 2011, motivated by the incompleteness and methodological limitations of traceroute-based maps at the time, Knight *et al.* constructed the Internet Topology Zoo, parsing information that network operators published on their websites; the majority of their maps are at the PoP level (where a network interconnects with other networks) and half are research and education networks [89]. PoP-level maps are not sufficiently granular to study aggregation structures in access networks. Beginning in 2015, a growing body of research investigated the physical infrastructure behind networks, especially focused on fiber [54, 110],

the frequency and impact of fiber cuts [65], and the impact of fiber deployment on end-to-end latency [25]. In this work, we map the router-level aggregation structure of access networks.

Broadband Networks: Substantial work analyzed broadband networks by sending probe packets to user's gateway [51], deploying home routers or embedded devices with measurement scripts [172, 30, 141], embedding measurements into BitTorrent software [148], and crowdsourcing measurements to end-users [90, 128]. This work illuminated characteristics (e.g., latency, packet loss rate, throughput, and uptime) of the end-to-end [22, 32, 170] and last-mile [171, 16, 60] performance of residential broadband networks without understanding the logical and physical topologies of the access network. This paper leverages the aggregation structure that we discovered to understand the latency observed by end-users.

Mapping Mobile Networks: Previous work studied the geographic coverage of mobile regional networks using the correlation between IP prefixes and location [208] and locating the PoP used by mobile devices in traceroutes [213]. However, these analyses were performed on 3G networks, and do not reveal the underlying access network infrastructure that produces these behaviors. An extensive body of wireless network measurement research has investigated the behavior of network elements unique to wireless – everything from the end-user devices to the mobile-specific middleboxes [186, 55, 203, 35, 204, 124, 3, 201, 97]. Connectivity factors can also impede performance of mobile ISPs, e.g., legacy hierarchical routing [53], lack of direct interconnection with content providers [213], peering strategies between mobile virtual network operators (MVNOs) and the underlying network infrastructure [154], and poor selection of DNS servers [145, 215]. In this work we determine the sources of latency limitations in today's mobile networks, and if those limitations can be overcome by moving services into access network infrastructure. We also reveal new hints in IPv6 addresses that reveal the region, packet gateway, and CO serving a mobile carrier.

3.4 Methodology Overview

Broadly, our topology mapping methods require the ISPs to allow the following measurements of their access networks:

Traceroute: We need the ability to observe routers in each CO with traceroute to uncover access network topology. This is straight forward when the network uses IP routing; if the network uses MPLS to organize routing between COs, then our method requires the ability to observe routers in each CO using traceroute torwards exit routers, using the method in [190]. For wired access networks, we prune a small number of links because traceroute can produce false links. Our pruning assumes the access network has a ring/star topology (§3.2.1). Although we can not validate this assumption for all ISPs we study—ISPs rarely publish information about their internal topology—we found support for the use of ring/star topologies in access networks from Comcast [196], Deutsche Telekom [70], Cisco [39], and Juniper [83].

Alias Resolution: To accurately map IPs to routers, we require the ability to resolve aliases using active probing (e.g., with Mercator [69] and MIDAR [87]).

Reverse DNS or Structured Addressing: To accurately map routers to COs, we either require the operator to ether label some of their router IPs with hostnames in their reverse DNS (rDNS), or they must have clear structure in their router address space that corresponds to the structure of their access network.

How general is our approach? The remainder of this paper (§3.5, §3.6, §3.7) studies six different U.S. access networks – Comcast and Charter (wired), AT&T (wired and mobile), Verizon (mobile), and T-Mobile (mobile). We believe our method can be extended to other access networks outside of the U.S.; for example, China's top three providers have been shown to provide the measurements we need for topology mapping [182]. Also, Bell Canada and Shaw Communications Canada appear to provide the necessary measurement primitives for our method [100]. However, there are classes of access network where our method will not

work. Primarily, these are where traceroute does not observe routers in COs. This is common in countries where the access and retail functions of ISPs are separated; examples of these are New Zealand's UFB [47] and Australia's NBN [121]. In these scenarios, the access provider tunnels subscriber traffic to hand-over points where the retail provider is co-located, which might be in entirely different cities, so the aggregation structure in the access network is invisible.

3.5 Case Study: Comcast and Charter

Our first case study focuses on two networks amenable to external traceroute-style measurements that also provide CO information in their rDNS. We focus on Comcast and the former Time Warner regional networks acquired by Charter—the largest cable Internet providers in the U.S.—networks that typically include router, building, and network information in their rDNS. Figure 3.4a shows a traceroute into Charter's Southern California region. The rDNS for hop 13 ends with `tbone.rr.com`, indicating a BackboneCO. Each subsequent hop includes an rDNS tag for the regional network `socal`, showing the transition from Charter's backbone into the regional network. Each rDNS name includes a portion of a CLLI code geolocating the router. Figure 3.4b shows the same pattern in Comcast's Beaverton, OR region, using CO locations rather than CLLIs.

Similar to many backbone point-of-presence (PoP) rDNS labels, the regional CO tags indicate the location of the COs. In Charter, the CLLI codes uniquely identify a specific building. Comcast sometimes uses the street address for a CO, but more commonly uses neighborhood, or city, names that provide a general geographic location, along with the U.S. state. Inspired by other tools that extract information from rDNS [100, 78, 29, 120], we hand-crafted regular expressions (regexes) to map these CO addresses.

Our methodology for these types of networks—those with rDNS and that can be externally probed—proceeded in two phases:

(1) build and annotate CO-level topology graphs for these networks;

13	66.109.6.227	bu-ether15.**lsancarc**0yw-bcr00.**tbone**.rr.com
14	66.109.6.231	agg2.**lsancarc**01r.**socal**.rr.com
15	72.129.1.1	agg1.**sndhcaax**01r.**socal**.rr.com
16	72.129.1.141	agg1.**sndgcaxk**01h.**socal**.rr.com
17	76.167.26.170	agg1.**sndgcaxk**02m.**socal**.rr.com

(a) Charter traceroute with **CO CLLIs**.

14	96.110.41.226	be-1102-cr02.**sunnyvale**.ca.**ibone**.comcast.net
15	68.86.92.206	ae-72-ar01.**beaverton**.or.**bverton**.comcast.net
16	68.85.243.238	ae-1-rur201.**troutdale**.or.**bverton**.comcast.net
17	162.151.213.86	po-1-1-cbr01.**troutdale**.or.**bverton**.comcast.net

(b) Comcast traceroute with **CO locations**.

Figure 3.4. Paths into Charter's Southern California region and Comcast's Beaverton, OR region. Each hostname includes a CO identifier and regional network.

(2) heuristically refine the graphs to reflect the actual topology.

3.5.1 Phase 1: Build Router-Topology Graphs

This phase conducts traceroutes to reveal the CO interconnections in each regional network. We conducted our probing from 47 vantage points (VPs) distributed throughout the United States in access, cloud, and transit networks.

First, we tracerouted to an address in every /24 in each regional network to expose at least one router from each EdgeCO. Second, we tracerouted to every address with rDNS matching one of our regexes to find CO interconnections missed in the first step. We identified IP addresses with hostnames matching our regexes in the Rapid7 rDNS dataset [138] which queries for PTR records for every IPv4 address. Directly targeting CO router interfaces observed 5.3x and 2.6x more CO interconnections than the /24 traceroutes for Comcast and Charter, respectively, as some COs responded to the /24 probing using addresses without rDNS. Third, we tracerouted to every intermediate IP address observed in these traceroutes to identify links that are entry and exit routers for an MPLS tunnel [190], allowing us to discard false edges between these COs. This MPLS heuristic proved important in larger Charter regions, where top level AggCOs appeared directly connected to nearly all EdgeCOs, which contradicted information about the Charter topology in Maine that we recieved from a trusted source.

Finally, we used alias resolution (Mercator [69] and MIDAR [87]) to group IP addresses

according to their router. We included all IP addresses with rDNS matching our regexes, as well as all IP addresses routed by each regional network. We annotated each inferred router group with a CO tag, using the most common tag extracted by our regexes using rDNS names for the router's interfaces. If a router did not have a most common CO tag among the rDNS for its interface addresses, we removed the CO mapping from any address in the router group with rDNS, to avoid inconclusive and potentially inaccurate mappings. We provide more details for how we mapped IP addresses to COs in Appendix A.2.1.

3.5.2 Phase 2: Build CO-Topology Graphs

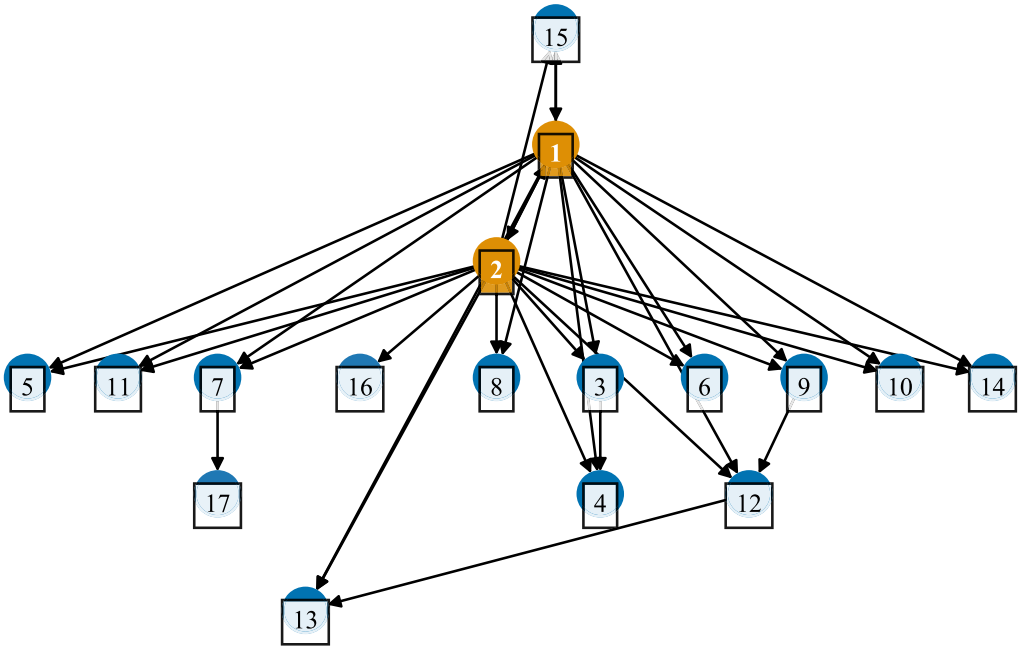
Using the CO mappings, we extract CO edges from traceroute paths, where immediately adjacent routers in a path map to different COs, and construct initial graphs of the topology for each region. The rest of this phase processes the topology graphs to more accurately reflect regional topologies. This phase (1) removes false inter-region edges; (2) identifies the AggCOs; (3) removes false edges between EdgeCOs; (4) adds missing edges from AggCOs to EdgeCOs; and (5) infers the entry points into each region.

Remove False Inter-Region Edges

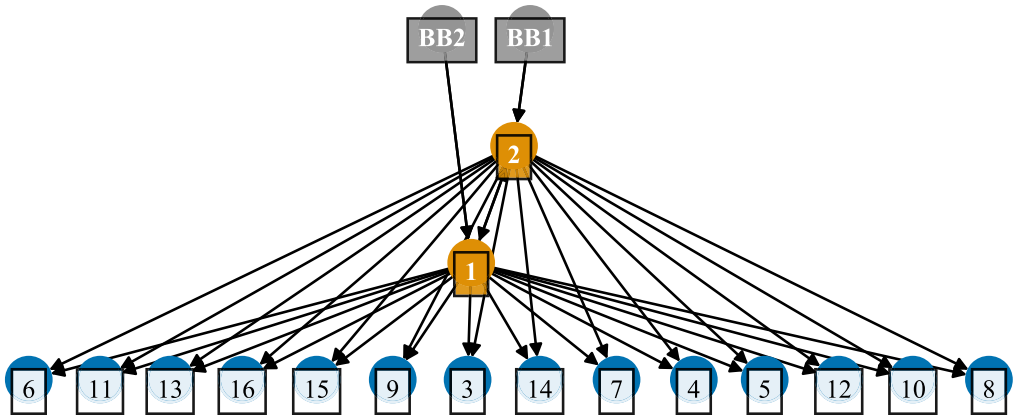
Large collections of traceroute paths likely contain some random noise [159], so we discard all edges that appear only in a single traceroute as anomalous. Next, we remove edges that appear to interconnect COs in different regions. While some links cross region boundaries (Section 3.5.2), many of these links result from outdated rDNS that our alias resolution did not catch. Further details are provided in Appendix A.2.2.

Identify AggCOs

Visually inspecting the regional graphs, such as the graph in Figure 3.5a, showed two features of interest. First, they signaled a hierarchical structure (partially obscured by extraneous intra-region edges), where a few COs appeared responsible for aggregating connectivity to the rest of the Internet for other COs in the region. We devised a heuristic to distinguish AggCOs



(a) Graph after removing external edges.



(b) Modified graph that accurately represents the topology.

Figure 3.5. Initially (a), the regional network graph has extraneous and missing edges. We identify the AggCOs (orange), heuristically refine the graph to reflect the regional network (b), and add the BackboneCO connections (grey).

from EdgeCOs based on the number of outgoing edges for each CO in the graph. In each region, we infer AggCOs as those with a higher out-degree than the average out-degree for that region plus one standard deviation. This heuristic identified the two orange COs in Figure 3.5a, COs 1 and 2, as the AggCOs in this region.

Second, the graphs naturally grouped these AggCOs; when a EdgeCO had more than one incoming edge, the two preceding AggCOs had outgoing edges to nearly identical sets of EdgeCOs. Topologically, this looks like a dual-star topology, with the implication that each AggCO in a subregion directly connects with every EdgeCO in the same subregion. The natural groupings also provide new insights into the physical topology, indicating which geographic regions rely on the same AggCOs, and the level of redundancy to each EdgeCO. Appendix A.2.3 details how we identified AggCOs, removed false edges, and added missing edges.

Remove False Edges between EdgeCOs

We knew that regional networks use a ring to connect an AggCO to its EdgeCOs (Section 4.3), so the star topology indicated that the fiber ring running from an AggCO to its EdgeCOs bundles separate fiber pairs for each AggCO-to-EdgeCO connection. These fiber pairs create two separate point-to-point connections between an EdgeCO and its AggCOs, bypassing all other EdgeCOs on the ring and eliminating the need to directly connect EdgeCOs to other EdgeCOs. While the graph in Figure 3.5a shows that most EdgeCOs only connect to AggCOs, some appear connected to other EdgeCOs, such as edges $9 \rightarrow 12$ and $3 \rightarrow 4$. These edges from EdgeCOs likely result from uncorrected stale rDNS, and we remove them to conform with a fiber ring (Figure 3.5b).

Add Missing Edges From AggCOs to EdgeCOs

When an EdgeCO lies along a fiber ring with one of the AggCOs, it will connect directly with the other AggCO on the ring as well. Otherwise, the fiber pairs would bypass the EdgeCO in only one direction. We therefore assume that missing edges, e.g., from AggCO 1 to node 16

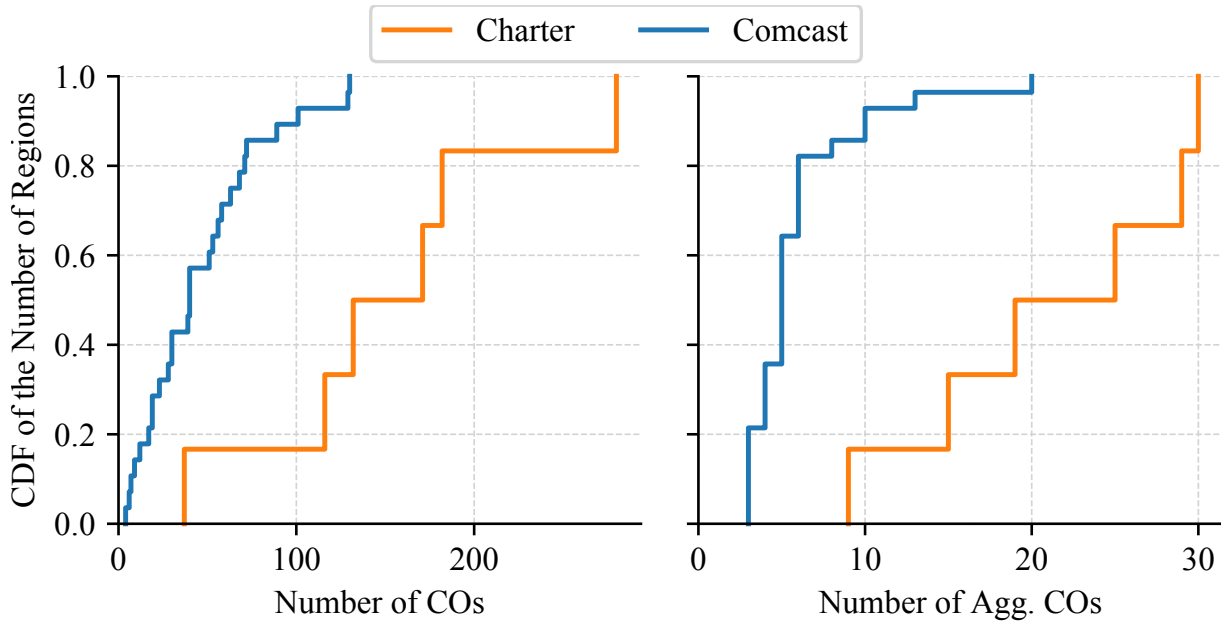
in Figure 3.5a, likely result from missing rDNS.

Resolving missing edges first requires us to identify the AggCOs on the same fiber rings. Our intuition is that AggCOs on the same ring will directly connect with the same EdgeCOs, so we devised a heuristic that looks for AggCOs that overlap at least 75% of their connections with EdgeCOs, strongly suggesting that they aggregate traffic on behalf of the same EdgeCOs. We then add edges to the graph such that all AggCOs on the same last-mile fiber ring connect to the same set of EdgeCOs. In Figure 3.5b, we add the missing edge from AggCO 1 to node 16.

Infer Entry Points Into Each Region

Finally, we add edges back into the graph that cross regional network boundaries, such as BackboneCO entry points and entry points from other nearby regions, but only when overwhelming evidence implies their existence. Returning to the traceroute paths, we extract all triplets of the form $(\text{CO}_i, \text{REGION}_1) \rightarrow (\text{CO}_j, \text{REGION}_2) \rightarrow (\text{CO}_k, \text{REGION}_2)$, where CO_i and CO_j appear in consecutive hops, indicating they directly connect. Given the hierarchical structure of the regional topologies, we only include potential entry points when they appear to lead to EdgeCOs in the region. To avoid misinterpretations caused by stale rDNS, we only include an entry point if we observe it leading to two or more COs in the same region.

Our analysis reveals that all regions in Charter, and all but three of the Comcast regions, connect to at least two BackboneCOs. A Comcast network operator told us that nearly every Comcast region directly connects to two BackboneCOs, so we likely missed three entry points in addition to the 57 backbone entry points we observed across the Comcast regions. In some regions we observe backbone connections and a direct connection to another region; e.g, the Central California region in Comcast appears to connect to two BackboneCOs and the San Francisco regional network. We did not observe direct inter-region connections in Charter.



(a) Total COs per region.

(b) AggCOs per region.

Figure 3.6. The 6 Charter regions include more COs than the 28 Comcast regions.

3.5.3 Contrasting Comcast and Charter

The key difference between Comcast and Charter is the number of regions they use, impacting the size of the regions and the extent of the aggregation inside each region. We observed only six Charter regions compared to 28 Comcast regions, but the Charter regions tend to cover more geographic area than the Comcast regions; e.g., Charter's Midwest region appears to touch 10 different U.S. states. Thus, a Charter region contains far more COs than a Comcast region (Figure 3.6a). Charter also uses more aggregation, and far more AggCOs per region (Figure 3.6b), than Comcast, where we define an AggCO as any CO with outgoing edges.

Figure 3.7 and Table 3.1 show the different types of aggregation we observed in Comcast and Charter. The smaller regions often used a single AggCO, small to mid-size regions used two AggCOs for greater redundancy, and the largest regions used multi-layer aggregation where lower aggregation levels might include one or two AggCOs. In the multi-layer topologies, Comcast nearly always connects EdgeCOs to multiple AggCOs, while Charter uses a mix. Charter's choices in aggregation lead to less redundancy to the EdgeCOs than in Comcast; 37.7% of

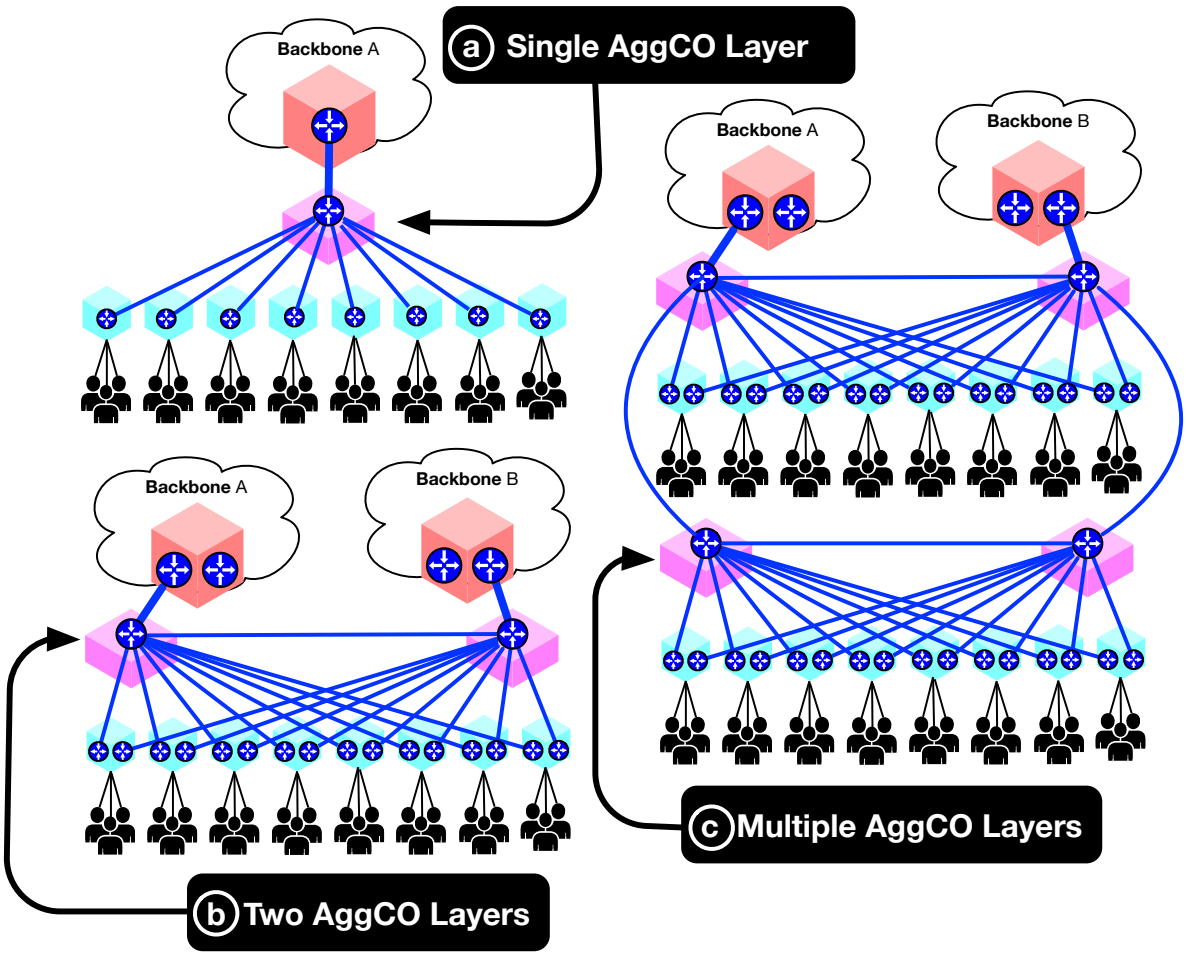


Figure 3.7. Three regional access network types.

EdgeCOs in Charter connect to only one upstream CO vs 11.4% in Comcast (see Appendix A.2.4 for important context).

Outside of one Charter region and one Comcast region, the difference in region size does not appear to manifest in greater entry points per regions. In Section 3.5.5, we find higher latency to the COs in the Charter regions, likely due to the combination of fewer entries per EdgeCOs and more aggregation lengthening the distance from the backbone to EdgeCOs. The fewer entries, additional aggregation, and less redundancy to the EdgeCOs that we observe in Charter could also increase the potential points of failure that could disconnect EdgeCOs, and customers connected to those EdgeCOs.

Table 3.1. Network types observed in Comcast/Charter.

Aggregation Type	Comcast	Charter
Single AggCO (Figure 3.7a)	5	0
Two AggCOs (Figure 3.7b)	11	0
Multi-level aggregation (Figure 3.7c)	12	6

3.5.4 Validating with Network Operators

We spoke with a network operator at Comcast and an operator at Charter to discuss our topology graphs and interpretations. The two people we spoke to are not experts for each region in their networks, however, both Comcast and Charter use a rough template for all of their regions, with AggCOs connected to EdgeCOs via fiber rings, and use similar technologies in each region.

The Comcast operator had extensive knowledge of one of the largest Comcast regions. We showed the operator our graph of that region, along with a list of the COs that we discovered, and the backbone entries. The operator confirmed that our inferred graph of the region was correct; the graph contained the COs, the second region with its own AggCOs that connect to the first region's AggCOs but not to the backbone, and the correct PoPs connected to the AggCOs. Finally, the operator confirmed that the largest Comcast regions often have two sets of AggCOs—one set connected to the backbone and another set connected to the first set—where each set connects to different EdgeCOs.

The Charter operator was not an expert for any specific Charter regions, but understood their design and general topology from the operator's experience with the Charter backbone. The operator thought we provided a reasonable representation of the regions and the regional topologies, but could not indicate if any COs were missing or superfluous. Importantly for our analysis, the operator confirmed that the Charter regions are vast, with layers of fiber rings with their own AggCOs.

We confirmed with both operators that they use fiber rings with star topologies—separate fiber pairs from AggCOs to EdgeCOs—as we inferred in Section 3.5.2, rather than a ring

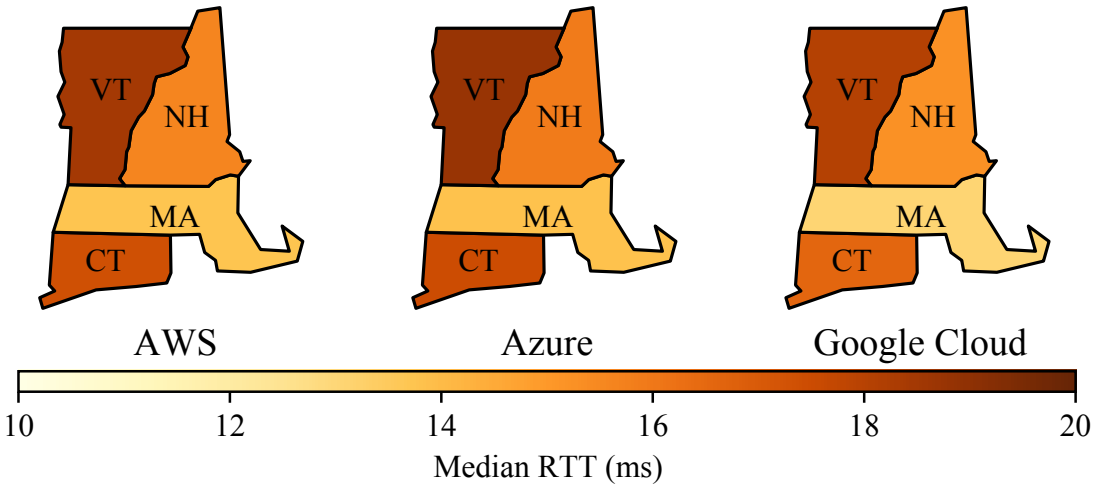


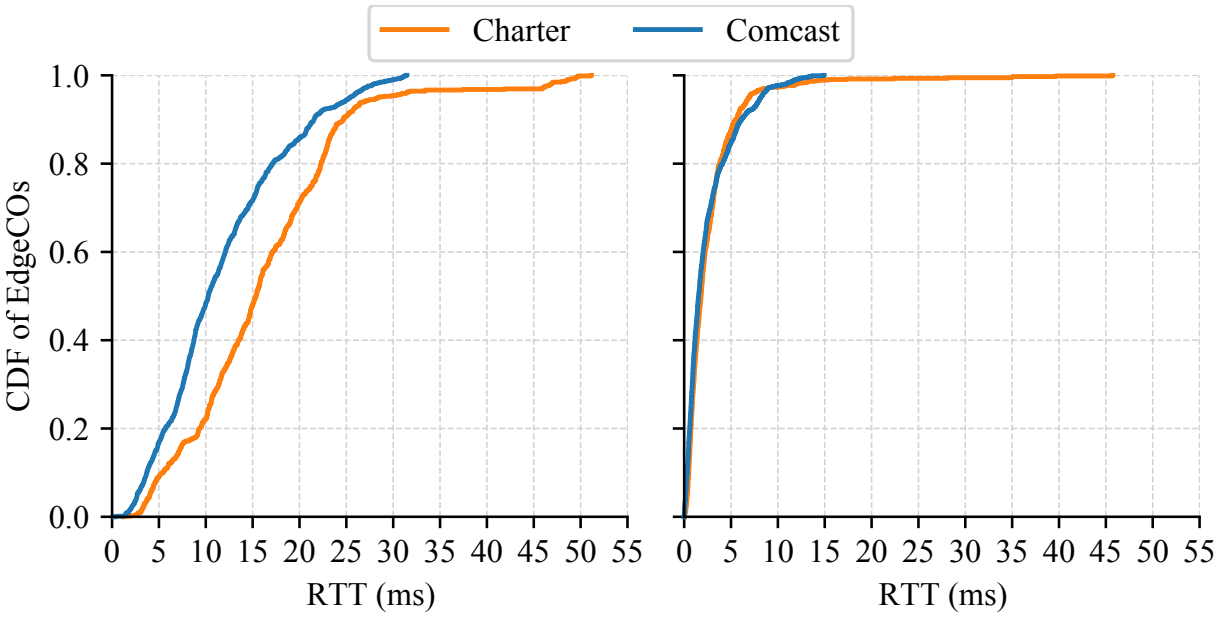
Figure 3.8. The median RTTs to Massachusetts, Connecticut, Vermont, and New Hampshire in the U.S. from the largest public cloud providers. Vermont has higher latency being geographically closest to the cloud datacenters.

topology. One network operator informed us that they chose this physical topology because it makes network upgrades simpler. We also asked both operators if the regions contain backup paths that traceroute might not observe. Both operators confirmed that all paths and COs are active, and cited the prohibitive cost of maintaining backup fiber paths or COs as the reason. This implies that traceroute can reveal all of the paths through the regional network, provided the VPs can exhaust the possible entries into the region.

3.5.5 Impact of Aggregation on Latency

The regional topologies help us better understand the inherent latency limitations imposed by the location of entry points and the aggregation in the regional topologies. To observe RTTs to different EdgeCOs, we conducted 100 pings from a VM in every U.S. cloud region for Amazon AWS, Microsoft Azure, and Google Cloud to every EdgeCO IP address included in our graphs. Then, we identified the closest location with the lowest minimum RTT to the highest number of EdgeCOs in a region.

Figure 3.8 provides the median of the minimum RTTs from the clouds to Comcast EdgeCOs in four states in the Northeast U.S.; in all three clouds the closest location was in



(a) RTT from cloud VM.

(b) RTT from agg. CO.

Figure 3.9. Although more than 80% of EdgeCOs are more than 5 ms RTT from the nearest cloud VM (a), more than 80% of the EdgeCOs are within 5 ms RTT of their AggCO (b).

Northern Virginia. Massachusetts, Vermont, and New Hampshire all use the same AggCOs in the Boston area, connected to BackboneCOs in New Jersey and New York, and a special purpose PoP in Boston. Surprisingly, although Connecticut is geographically closer to Northern Virginia than the other states, it has worse latency than Massachusetts and New Hampshire. The reason is that the Connecticut regional network does not have its own backbone entries; instead, its AggCOs connect to the backbone through the Massachusetts AggCOs, resulting in a 3.5 ms to 4 ms RTT penalty.

While the aggregation tends to increase latency to EdgeCOs, it presents opportunities to to bring cloud applications closer to users without placing edge computing infrastructure in every EdgeCO. Conventional wisdom holds that certain classes of applications, such as augmented or virtual reality, require less than 5 ms of latency [115], but more than 80% of the Comcast EdgeCOs and 90% of the Charter EdgeCOs have an RTT greater than 5 ms (Figure 3.9a) from the nearest cloud location. One approach is to push edge computing to the EdgeCOs, ensuring nearly all users are within the latency constraints, but increasing the cost and complexity of

Figure 3.10. EdgeCO in a Comcast regional network.

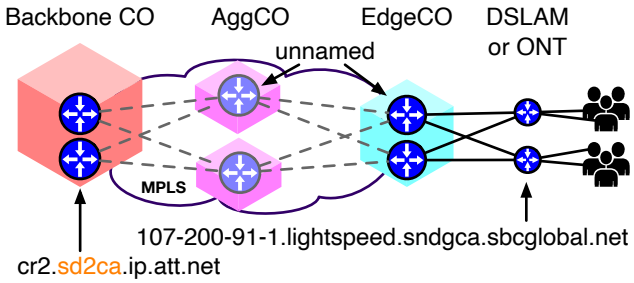


Figure 3.11. Architecture of AT&T's access network. Routers in AggCOs and EdgeCOs are unnamed. The backbone router is in the "sd2ca" region.

deployment. Another approach could exploit the hierarchy in the regional topologies and place the edge computing infrastructure in the AggCOs. Counting any CO with an outgoing edge as an AggCO, we observe 7.7x as many EdgeCOs as AggCOs across all regions of Comcast and Charter. More than 80% of the EdgeCOs for Comcast and Charter are within 5 ms RTT of the AggCOs, likely bringing the vast majority of regional network customers within the 5 ms requirement (Figure 3.9b). Furthermore, the AggCOs are often substantial datacenters, with the security, power, and capacity to include edge computing infrastructure, while EdgeCOs might be houses on residential streets (Figure 3.10).

3.6 Case Study: AT&T

Next, we investigate the topology of AT&T's wireline regional access networks (Figure 3.11). Compared to the cable providers AT&T's network is relatively opaque. AT&T

provides rDNS for their BackboneCO routers but not for other CO routers, and provides rDNS that identify their last-mile IP-DSLAMs and ONTs. However, there is no straight-forward way to identify which IP addresses AT&T assigns to their wired customers or mobile customers, or which addresses they delegate to other networks. While AT&T's EdgeCO and AggCO router topology can be uncovered by traceroute, this is only possible within their respective regional networks. Our methodology for AT&T is similar to Section 3.5 at a high-level: we (1) build and annotate router-level topology graphs, and then (2) construct CO topology graphs for each regional network. The low-level methodology details in this section are tailored to the specific challenges presented by AT&T.

3.6.1 Phase 1: Build Router-Topology Graphs

We bootstrap our discovery of router IPs in the COs by tracerouting from 5 Ark VPs near the region we are mapping to the IP addresses of DSLAMs/ONTs in EdgeCOs. AT&T uses rDNS to label DSLAM/ONT IPs as `*.lightspeed.(CLLI).sbcglobal.net`, illustrated in Figure 3.11. We tracerouted to all 95,821 IPs matching this pattern rDNS in the Rapid7 rDNS dataset. AT&T's access networks use MPLS tunnels, so these traceroutes only discovered the entry router for the tunnel (the BackboneCO router), and the exit router (an EdgeCO router), missing most routers in AggCOs. Further, the bootstrap traceroutes do not observe all EdgeCO routers, because of MPLS as well as some of DSLAMs/ONTs do not reply and perhaps some do not have rDNS. However, we found that the EdgeCO routers were allocated out of a few prefixes per region. For example, there appear to be 7 /24s used for EdgeCO router IPs in AT&T's San Diego, CA region (indicated with "`sd2ca`" in the rDNS entry for the BackboneCO router in Figure 3.11). To uncover AggCO routers, we use the same technique as in Section 3.5—traceroute to the observed MPLS tunnel exit router [191] in the EdgeCO. Appendix A.3 includes further details about how we infer EdgeCOs.

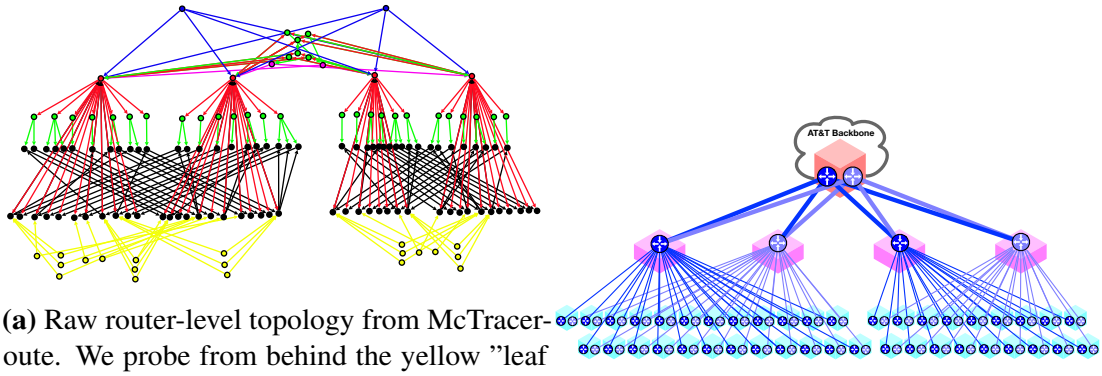
Because we can only traceroute to most EdgeCO router IPs from within the same region, we build per-region lists of EdgeCO /24s to probe by associating /24s with the region tag in

the BackboneCO router rDNS observed in bootstrap traceroutes. We then traceroute to all IP addresses in these prefixes in the region from a VP within the region. We also performed alias resolution to map individual IP addresses to routers, and then to EdgeCOs and AggCOs In total, we found 37 AT&T regional networks identified in rDNS, and CAIDA Ark and RIPE Atlas had VPs available in 35 of these regions. However, even in regions where we have many VPs, those VPs are insufficient to reveal the complete topology. Comprehensively revealing the regional network topologies requires finding VPs with different paths. This is particularly important because, as we will discuss in the next phase, mapping router IPs to EdgeCOs requires at least one VP served by each EdgeCO. However, finding topologically diverse AT&T VPs in a region we want to map is not feasible with existing sources of crowdsourced VPs (Atlas and Ark). Further, AT&T's looking glasses are not suitable VPs because they are located in AT&T's backbone network, which EdgeCO and AggCO routers do not respond to traceroute (ICMP) packets.

To significantly increase the number of VPs inside an AT&T region that we are mapping, we leverage existing network infrastructure that reveals many geographically distributed last-mile links in a region – public WiFi hotspots. Our insight is that many fast food restaurant chains (e.g., McDonald's, Starbucks, and Subway) have many geographically distributed last-mile links, to many EdgeCOs, serving their WiFi hotspots. We call this approach *McTraceroute*. We believe this is the first network topology measurement effort that has made use of geographically distributed WiFi hotspots.

To evaluate how well this technique improved our visibility of a region, we ran traceroutes from all 58 McDonald's in AT&T's San Diego region to all IP addresses in seven /24s that we inferred to contain AT&T's San Diego EdgeCOs and AggCOs. We found 23 McDonald's that used AT&T for their free WiFi services. The diverse location of McDonald's restaurants, whose locations are strategically selected to maximize coverage in an area, provided us opportunities to connect to, and perform measurements from, many EdgeCOs in the regional network.

Next, we investigate how many new paths we observed with each type of VPs in San



(a) Raw router-level topology from McTraceroute. We probe from behind the yellow "leaf nodes" (IP-DSLAMs) at the bottom.

(b) Inferred CO-level topology

Figure 3.12. AT&T San Diego Regional Network

Diego, to determine if McTraceroute significantly increases the number of paths we observe. Considering traceroute paths starting with the second hop, the eight Atlas and two Ark probes in AT&T's San Diego respective regions revealed only half of the IP paths we observed with McTraceroute. This indicates that increasing the number of VPs revealed many more paths, despite overlap in the McDonald's EdgeCOs. Note that because the network is opaque, we do not know the true number of paths, however McTraceroute provides a significant increase in observed paths over existing VPs.

3.6.2 Phase 2: Build CO-Topology Graphs

Phase 1 produced the router-level topology shown in Figure 3.12a. We inferred two backbone routers (blue), four aggregation routers (red), and 84 EdgeCO routers (black). We inferred the EdgeCO routers as connected redundantly to two aggregation routers each, and all aggregation routers connected to one backbone router. The router-level topology reveals a three-level structure, with two sub-regions that use different aggregation routers.

To infer the CO-level topology, we first map last-mile links to EdgeCO routers. Each last-mile link is served by a single EdgeCO, so if two routers are one hop away from the same last-mile link, we conclude they are both in the same CO. We observed each last-mile link connected to two EdgeCO routers, indicating that each EdgeCO has two routers.

We observed two backbone routers, and both appear fully connected to all aggregation

Table 3.2. Latency from Google Cloud VPs to EdgeCOs in San Diego. Two have >2x the average latency (4.3ms).

Latency:	3-4ms	4-5ms	5-6ms	6-7ms	9-10ms
EdgeCOs:	5	19	7	2	2

routers. This is unlike the cable networks, where we observed backbone routers connected to one aggregation router. We conclude from this inference that AT&T has only one BackboneCO in this region, and this office contains both core routers. We are less confident about the four aggregation routers, but the highest resilience design would have them operating out of four different COs. Figure 3.12b shows the inferred CO-level topology.

3.6.3 Analysis of AT&T's Topology

AT&T's regional network has a significantly longer history than cable networks, dating back to the early 1900s. Therefore, we expect its structure to reflect design choices constrained by the capabilities of early of telephone networks. AT&T's long distance network, called Long Lines, only reached a single CO in each region. These Long Lines COs now appear to serve as their BackboneCOs. In fact, the BackboneCO we inferred in San Diego still has a Long Lines microwave tower on its roof.

Aggregation

We observed significantly higher EdgeCO density in AT&T's network than in the cable providers we studied. In Charter's San Diego sub-region we observed 16 EdgeCOs, compared to 42 in AT&T's San Diego region. This CO deployment density is consistent with AT&T facing the constraint of local copper telephone service loop lengths. By the time cable networks emerged in the 1990s, Hybrid Fiber Coax allowed for much longer last-mile links from EdgeCOs to customers. We would thus expect, without considering other factors, that AT&T aggregates fewer last-mile links to each EdgeCO than do cable networks. This lower ratio of customers to EdgeCO helps to reduce the scale of outages when an EdgeCO fails (e.g., due to fiber cuts or

failed equipment).

To estimate latency differences between EdgeCOs in the San Diego region, we conducted traceroutes from a VM in a Los Angeles Google Cloud datacenter to all of the end user AT&T IP addresses we could find in the San Diego region. We used Measurement Lab data [102] to extract AT&T customer IP addresses from NDT measurements, and retained the subset of addresses located in San Diego or Imperial County according to NetAcuity [2], a commercial geolocation service. Using only traceroutes that passed through the BackboneCO in San Diego and reached the customer addresses, we inferred that the penultimate traceroute hop corresponds to a device in an EdgeCO.

We could not directly ping these devices, but we could elicit responses by sending an ICMP Echo packet to a customer IP address with the TTL field set to the penultimate probe TTL in the traceroute to that address. To measure latency from Google Cloud in Los Angeles to the EdgeCOs, we conducted 100 probes to each EdgeCO address observed in the traceroutes and used the minimum observed RTT (Table 3.2). These results show that some EdgeCOs have significantly less latency to the BackboneCO than other EdgeCOs in the region. Two distant EdgeCOs—with connected customers geolocated to Calexico and El Centro, CA—had over twice the average latency of 4.3ms to Google (9–10 msecs). This disparity suggests that some AT&T customers will suffer considerably higher latency to cloud services than other users in the region.

Redundancy

AT&T's network in San Diego has a similar lack of redundancy that we observed in some cable provider regions; namely, the use of only one BackboneCO. In AT&T's network, these BackboneCOs are fortified for natural disasters, such as Category 5 hurricanes. However, the Christmas 2020 attack on AT&T's Nashville office, which we assume is the lone BackboneCO in Nashville, took down the entire region, consistent with our inferred topology. Relative to the cable providers, AT&T appears to have more redundancy in their BackboneCO to AggCO

paths, with all backbone routers connecting to all Agg routers. Our measurements cannot detect whether these paths take diverse fiber paths.

Validation

Aspects of our inferences match historical documents describing AT&T's telephone network in San Diego. AT&T's access network was likely built using these same COs. The first document [14] states that AT&T operates one tandem building in San Diego (CLLI `SNDGCA02`), consistent with the single BackboneCO that we inferred. The documents also describe 42 subtending COs in San Diego, we believe these match the ∼40 EdgeCOs we inferred. A second document [152] shows four "Inter-office" COs in San Diego's network, we believe this term is AT&T's term for AggCOs.

3.7 Case Study: Mobile Carriers

For mapping the regional access networks of all three major mobile carriers, we focused on the portion of the network that bridges the mobile packet core with the rest of the Internet (and edge services). Some mapping challenges are similar those of AT&T's wireline network: they have no rDNS on routers, and probing requires internal vantage points. However, mobile networks face a significant additional challenge: they have no distributed VPs to provide internal views of the providers' regional networks. Although, they also present a unique opportunity to observe nationwide network topology: unlike the wireline transparent networks, we can physically move mobile VPs to probe inside different regions. Building on this insight, we introduce a new parcel-based measurement technique, ShipTraceroute, to obtain national coverage of mobile access network regions. Then, we use the large geographically-tagged dataset of traceroutes we collected to infer the topology of the networks.

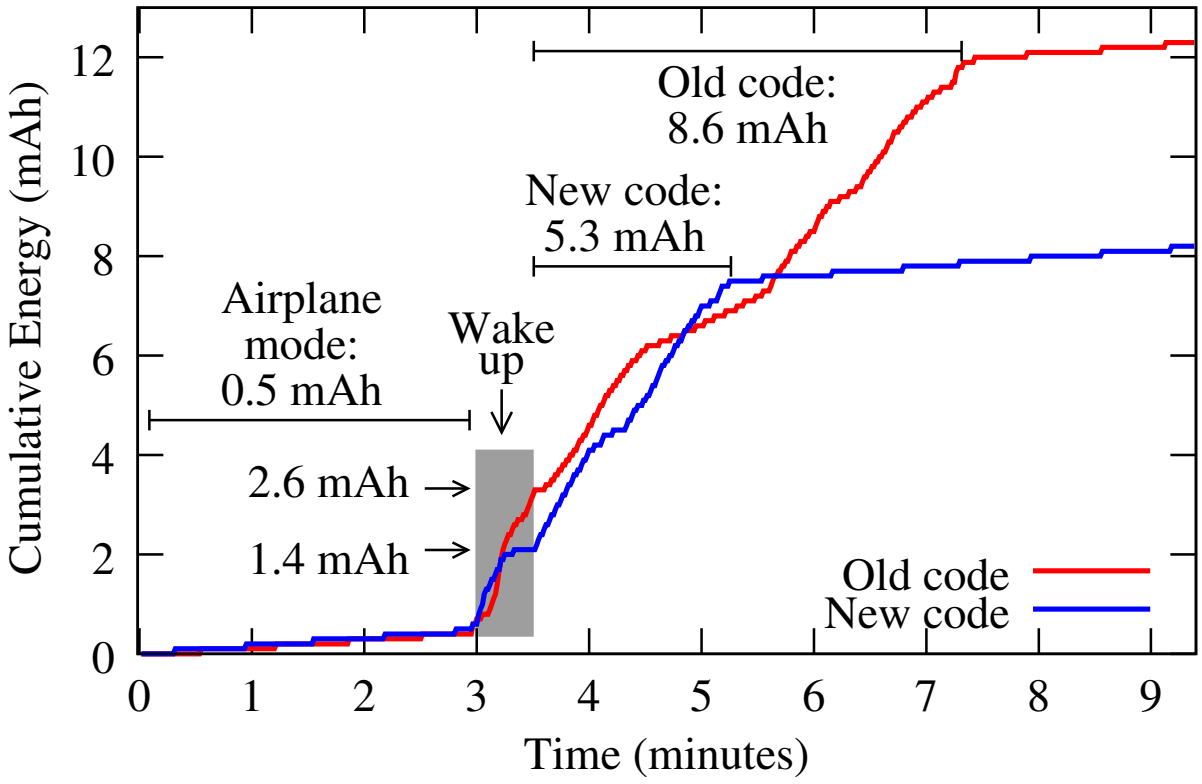


Figure 3.13. Improving scamper's traceroute efficiency

Figure 3.14. Shipping to 12 destinations covered 40 states

3.7.1 Phase 1: Collect router-level topology

We developed ShipTraceroute, a smartphone-based network measurement technique that can send traceroutes from a battery-powered Android device for a prolonged period of time while being shipped inside a truck or railcar. Appendix A.1.1 describes how shipment of a smartphone running this software complies with U.S. regulations for items shipped in a parcel inside of a truck or train.

Topology collection

We shipped three Samsung Galaxy A71 smartphones (one for Verizon, AT&T, and T-Mobile) to 12 locations in the U.S. The shipment paths traversed 40 states (Figure 3.14). During the shipments, the devices attempted to perform a round of traceroutes once per hour. However, signal conditions varied significantly along the routes. Some areas had too weak of a signal inside of the vehicle to perform the traceroutes, particularly in areas where there are no inhabitants. We observed the following success rates for rounds of traceroutes during the journey: 1592/1948 (82%) on AT&T, 1720/2054 (84%) on Verizon, and 872/1153 (75%) on T-Mobile.

The destinations for each round of traceroutes were IPv6 and IPv4 addresses in ASes neighboring the mobile carriers' networks.[1] The reason why we used external destinations (in neighboring ASes) to map mobile access networks, rather than internal addresses like we used in wireline networks, is because mobile networks block traceroutes to internal infrastructure. We used destinations in all neighboring ASes to try and traverse all of the carriers' BackboneCOs in each region (details are in Appendix A.4). However, quickly we discovered that traceroutes to all of the destinations took the same path inside each of the mobile access networks, allowing us to reduce to a single destination per provider.

We also observed that the path through the mobile network did not change as the phone moved within a region. We found we needed to force the phone to re-register with the core network by putting it into airplane mode before each round of traceroutes to route through all region's EdgeCOs and packet gateways.

Since GPS signals are rarely available inside of shipping vehicles, we logged the device's CELLID each time we started a round of traceroutes. We then converted the CELLID to a geolocation using the OpenCellID public cellular tower geolocation database [187].

Making mobile tracerouting energy efficient

We designed the measurement software on our smartphone to prolong battery life. The goal was to ship the phone by ground transport across the U.S.—a journey that takes about one week—while running measurements each hour, without the battery emptying.

We achieved this without sacrificing measurement fidelity by making two modifications to *scamper* [99]—ShipTraceroute's network probing tool. First, we modified scamper so that it could conduct measurements without without rooting the phone (rooting can disable thermal safeguards). Second, we reduced scamper's energy consumption by modifying its traceroute implementation to send probes to multiple consecutive hops in parallel. This significantly reduces the time that scamper spent waiting for unresponsive hops, and thus reduced the time the phone's

[1]We used Zayo's AS for T-Mobile because T-Mobile does not have its own IPv4 AS and T-Mobile's primary backbone provider is Zayo.

radio is fully powered.

We evaluated the energy efficiency of our modified scamper implementation by measuring the energy consumption of a Samsung Galaxy A71 5G performing traceroutes to the 266 IPv4 and IPv6 destinations in AT&T. To measure the device's energy consumption, we fully charged it, and instrumented with a USB-C power monitor on its charging port. This allowed us to measure the energy needed to operate the device. Figure 3.13 shows how much we improved energy efficiency: we achieved a 38% reduction in energy from 8.6 mAh with off-the-shelf scamper to 5.3 mAh with ShipTraceroute's scamper. As a result of these improvements, we calculated that our phone can perform hourly traceroutes for $\sim$12 days on one charge, a gain of $\sim$4 days over the off-the-shelf implementation. The other main contributors to power consumption are the energy consumption required to exit airplane mode when we start a measurement (1.4–2.6 mAh), and the trickle of energy consumed when the phone is asleep and in airplane mode between measurements. Although we put the device in airplane mode between traceroute rounds to force it to re-register in the packet core, it also has the additional benefit of reducing energy consumption (14.5 mAh vs. 9 mAh in airplane mode for every 55 minutes asleep).

3.7.2 Phase 2: Inferring CO-level topology

Each traceroute collected in phase one revealed a path from the mobile packet gateway (the first hop) until the packet reaches the BackboneCO. However, it is difficult to infer CO-level topology from these traceroutes because mobile networks have extremely limited rDNS (only Verizon has rDNS).

Fortunately, IPv6 is now widely deployed in cellular networks, and IPv6 addresses' are long enough that providers can encode information in them about where those addresses reside in the topology of their access network. Indeed, we found an early discussion about how to set IPv6 prefixes for LTE infrastructure that described how bits in addresses can be used to indicate what those addresses are used for—infrastructure or users—and what their location is in the network topology [40]. With the large number of geo-tagged samples of IPv6 router addresses in

the traceroutes we collected, we looked for patterns in how the bits in the addresses change as the mobile device moves.

AT&T

Figure 3.15(a) shows the patterns we observed in AT&T's addresses in their traceroutes. The user address and first hop (packet gateway) /32 prefix are consistent throughout the country, indicating it is the general AT&T mobile user prefix. User addresses also have a more specific /40 prefix that only changes 11 times as we move around the country. This prefix also changes simultaneously with bits 32-47 of the router addresses. We believe this prefix indicates the EdgeCO (and region) that is in use by the device, indeed this prefix can be used to route to the correct BackboneCO router to reach the user. For validation of this result, we discovered an AT&T document from 2014 that also lists 11 mobile datacenters in the U.S. [15]. However, we suspect that as they roll out their 5G network, they will add more EdgeCOs and thus reduce the size of their regions.

The /32 prefix of the rest of hops before leaving AT&T's mobile network are always the same, and different from the user address, so we infer they are the general prefix for AT&T infrastructure (i.e., routers). We observed bits 48–52 of these addresses cycling through several values inside each inferred EdgeCO, and they changed at the same time as bits 32–40 of the user address. Also, these bits changed each time we woke up from airplane mode and re-attached to the cellular network. Therefore, we infer these bits indicate the current packet gateway in the EdgeCO that user is attached to. Table A.5 (Appendix A.4) shows the inferred infrastructure in each region.

Verizon

Figure 3.15(b) shows the patterns we observed in Verizon's addresses. All of the first 10 hops are within Verizon's network, but only the first (packet gateway) and the last four hops appear in the traceroutes. The /24 prefix of the user address and first hop stays the same

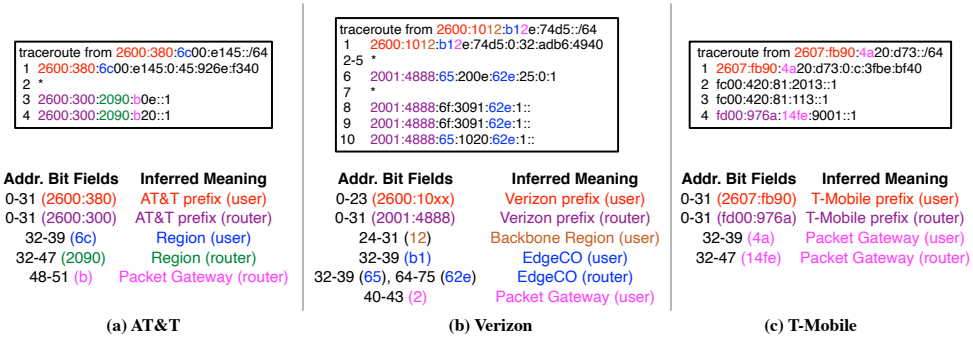


Figure 3.15. Topological hints for mobile networks encoded in IPv6 addresses.

throughout the country, indicating this is Verizon's user address prefix. As the device moves, more specific bits change. The /32 prefix changes 18 times, and the /40 prefix changed 32 times—both were stable within contiguous geographic regions. The /32 prefix changed less frequency in a geographic area than the /40 prefix. One plausible explanation for this behavior is that the /32 prefix identifies the BackboneCO and the /40 identifies the EdgeCO using that BackboneCO. We also observed bits 40–43 in in the user address can change when we cycle airplane mode, while other hops stay the same, indicating multiple packet gateways in each EdgeCO.

This explanation is supported by information from the rest of the hops (i.e., infrastructure). The /32 prefix in the user address is likely to represent the BackboneCO because it corresponds with changes in the rDNS of the Verizon backbone hop (i.e., `alter.net`). The /40 prefix is likely to be the EdgeCO because when it changes, so do the bits of the addresses in the other hops—the hops to reach the EdgeCO from the BackboneCO—namely, bits 64–75 in all of the infrastructure hops, and bits 32–39 in some of the hops. Table A.6 (Appendix A.4) shows the inferred infrastructure in each region.

Although we were unable to find documentation that validates our inferred topology, we performed several controlled experiments to test our inferences. First, we found Verizon Wireless deploys speedtest servers in their EdgeCOs which contain the names of the EdgeCOs in their rDNS. For example, `cavt.ost.myvzw.com` is the speedtest server in the **V**is**t**a, **Ca**lifornia

EdgeCO). We performed a controlled drive north from San Diego to Irvine while tracerouting to all of the speedtest servers, and we observed that when the shortest traceroute path switched from the Vista, CA to the Azusa, CA speedtest server, the expected bits in the traceroute hops changed at the same time. Additionally, we performed a long-running stationary experiment verify if the EdgeCO and BackboneCO address bits were stable in a location in San Diego. Indeed, they were generally stable across multiple days, however we did observe a small number of switches to the neighboring EdgeCO connected to the same BackboneCO. This implies the packet core connects to both EdgeCOs and it can switch between them if necessary for load balancing or redundancy.

T-Mobile

Figure 3.15(c) shows the patterns we observed in T-Mobile's addresses. Similar to the other two providers, the user IP prefix /32 stays the same across the entire country. The /40 prefix of the user IP can change each time it leaves airplane mode within a geographic area roughly the size of a city. These /40s are cycled through in a somewhat round-robin fashion, indicating that bits 32-39 likely represent the packet gateway. However, we observed that T-Mobile also cycles through different BackboneCO *providers*, suggesting that T-Mobile has a different mobile access network topology than the other providers. We infer that T-Mobile has a set of packet gateways in each region, possibly in different EdgeCOs, and with different backbone providers. These packet gateways are likely interconnected by the packet core.

We confirmed with T-Mobile that they have several backbone providers serving each region at different interconnection points within the region. Also a device in one location connects to different packet gateways at different sites (i.e., EdgeCOs), but that they prefer the closest site. Their network is designed in this distributed fashion for lower latency and resiliency. Therefore a device can wake up connecting to a different packet gateway than it connected to before it went to sleep.

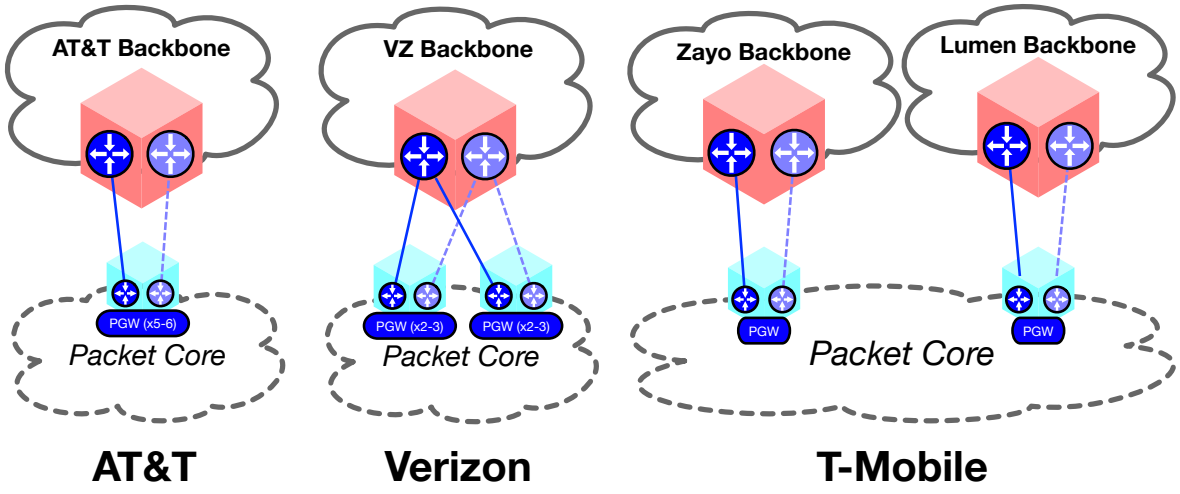


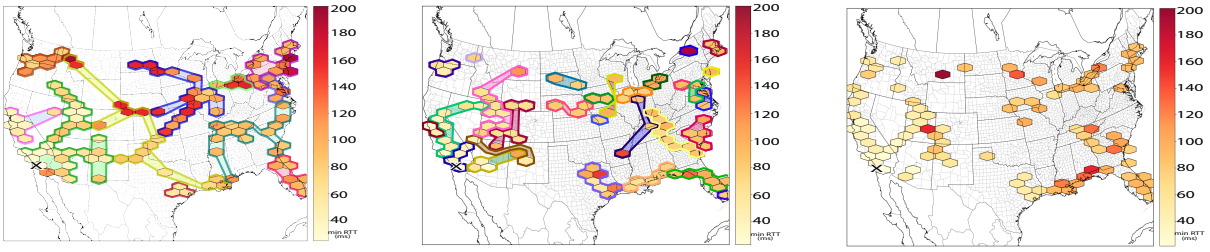
Figure 3.16. Inferred Internet topologies of U.S. mobile carriers

Summary

We infer topologies among the three providers' access networks (Figure 3.16). AT&T appears to have a single EdgeCO with multiple packet gateways connected to their nearest backbones. Verizon has multiple EdgeCOs sharing the same backbone CO, but the EdgeCOs cover non-overlapping regions. T-Mobile has multiple EdgeCOs in one region, but does not aggregate traffic to a single backbone, rather they aggregate to a variety of backbone providers directly connected to the EdgeCOs. These designs have different tradeoffs. AT&T's design may be more cost-efficient because equipment and links are centralized to a single EdgeCO per region. However, the lack of diverse CO locations may increase latency. Verizon and T-Mobile appear to have lower latency in part because they have multiple EdgeCOs per region.

3.7.3 Comparison of US Mobile Access Networks

The topology of mobile access networks has implications for network latency, because user traffic has to traverse to the backbone PoP of the region to reach other Internet hosts. Figure 3.17 shows the *minimum latency* we measured from our ShipTraceroute smartphone in different locations to a server located at CAIDA in San Diego. The hexagons indicate where we captured latency: the darker the color, the higher the minimum latency to the server from



(a) AT&T **(b)** Verizon **(c)** T-mobile

Figure 3.17. Minimum latency from each location to a single server in San Diego. Colored regions indicate the measurements were handled by the same EdgeCO (inferred from IPv6 addresses). T-Mobile does not aggregate traffic to a single EdgeCO.

that location. The colored regions containing multiple hexagons indicate those latency samples were taken from the same mobile access network region according to the IPv6 bit fields for that provider.

AT&T's regions are much larger than Verizon and T-Mobile, therefore some geographic areas (e.g., Montana and North Dakota) incur high latency to San Diego, due to circuitous paths to the BackboneCO. Verizon's network generally had lower latency (Figure 3.17b), because the larger number of EdgeCOs likely provided shorter average distance to BackboneCOs. As T-Mobile's distributed topology relies on several backbone providers (Figure 3.17c), they had latency similar to Verizon. However, we observed unusually high latency near the border of Florida and Louisiana (Figure 3.17c), because during the experiment the device in these regions attached to a distant EdgeCO in South Carolina.

3.8 Future Work

Resiliency. The tools and methods we have developed for inferring regional topologies enable a new approach to studying resilience across space and time. The topological differences that we have already observed across different regions have strong implications for resilience against disasters. A promising next direction is to combine these topologies with existing or future data sets on resilience of connectivity.

Edge Computing. Understanding the topology of these regional access networks, and

associated performance implications, may be the key to realizing the unachieved potential of the long-hyped edge computing paradigm [151, 75]. In addition to discovering the pyramid structure of the Edge CO and Agg CO topologies, our latency measurements suggest that the AggCO is typically less than 10 msecs from both the cloud and customers in the region, which meets the AR/VR latency requirement for edge computing [122]. This result suggests that putting edge computing infrastructure in Agg COs is the most efficient solution. Efforts to offload computation from mobile devices [104] can also leverage an understanding of the effect of distributed EdgeCOs on latency to the cloud.

Scalability of measurement methods. There is opportunity for improving scalability and manageability of our methods. For the AT&T study, we drove to each McDonald's location in San Diego, connected to their WiFi, and collected traceroutes. This approach is a fun adventure for a graduate student, but operationalizing such a measurement requires crowd sourcing. We could develop an app that connects to public WiFi spots (while the user waits for their food order), and provides a reward for uploaded results.

We also envision ways to improve the scalability of ShipTraceroute. Besides sending more cellular packets in parallel to save energy, we can arrange for the device to sleep even more between measurements. During a cross-country shipment, a device often stops at a hub for about a day. We could use the device's accelerometer to pause measurements when the device is at rest.

3.9 Conclusion

We have undertaken a comprehensive measurement study of the topology of U.S. regional access ISPs. Our motivation was to extract insights about architectural choices that ISPs make for how to aggregate traffic, and then empirically assess implications of those insights for the resilience and evolution of the Internet ecosystem. Growing interest in edge computing and 5G co-location, not to mention the pandemic-induced semi-permanent transition to working from home, is placing increasing pressure on these regional networks. We are now entirely dependent

on this infrastructure but there has been little attention to independent objective understanding of its resilience and reliability.

This dearth of attention is understandable. While perhaps not the most opaque part of the Internet, these networks are not amenable to straightforward measurement and analysis. Our tools have their limitations, but they allowed us to make surprisingly accurate maps in spite of considerable noise in our input signals, e.g., missing or incorrect DNS or traceroute hops. We were able to identify many different approaches to provisioning redundancy, across links, nodes, buildings, and at different levels of the hierarchy. These measurements can provide a basis for reasoning about sources of performance and reliability impairment in these networks. We believe that sharing our methods, lessons, and results will inform future analysis of critical infrastructure.

3.10 Acknowledgement

Chapter 3, in part, is a reprint of the material as it appears in *Internet Measurement Conference 2021*. Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, k claffy, Aaron Schulman. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Uncovering the Physical Risks in Access Network with Revealed Region Topologies

4.1 Introduction

Regional access networks are an essential component of the Internet infrastructure: they connect end users to the rest of the Internet. In order to balance reliability and performance against the enormous cost of providing last-mile connectivity to vast populations of geographically distributed users, access networks aggregate customer traffic into layers of central offices that are connected with varying degrees of redundancy. Unlike backbone networks, access networks often lack sufficient redundancy to withstand single-facility failures and a recent study showed that third parties can infer these points of limited redundancy [216]. Troublingly, physical attacks against regional access network infrastructure are becoming increasingly common [63, 160, 118, 195].

Today's regional access networks are far more critical than when they were first deployed, with ballooning real-world impacts of network outages. No longer just conduits of landline telephone and cable TV, modern access networks support 4/5G cellular phones, cloud services, hospital and financial services, and the remote work essential to the modern economy. Perhaps the most dramatic illustration of these inter-dependencies occurred in December 2020 when a bomb disconnected an AT&T network facility in Nashville, Tennessee [193]. This single event took the entirety of AT&T's wireline and wireless network in the Nashville area offline for several days. It also disconnected 911 emergency services [66], grounded flights by taking air

traffic control offline [168], prevented hospitals from reaching remote records and health-care providers [91], and even halted credit card processing [143]. We believe these circumstances demand a clear-eyed assessment of the threats to regional access network infrastructure and a reconsideration of the operational trade-offs occurring today.

In this paper, we evaluate the ways in which regional Internet access networks are at risk of physical attack in an effort to better inform the cost-benefit analysis of existing and future deployments. We perform a large-scale measurement campaign to study the impact of infrastructure failures on real-world access networks. Specifically, we continuously monitor users of the primary access networks in several regions of the United States for a year. When we observe large correlated outages, we identify the portion of the access network topology that likely failed using a technique we introduce in this work. To our knowledge, this is the first public study to assess the potential impacts of physical attacks on the regional access network infrastructure in the U.S.

Furthermore, we show how operational practices may facilitate targeted attacks. For example, regulations often require providers to record locations of their diesel fuel storage and battery backup power systems in local hazardous-materials registries. We demonstrate that an attacker often can identify the physical infrastructure serving a particular region based upon a set of design patterns: access networks typically have well-segregated coverage areas. As a result, an attacker can infer the infrastructure providing service to a particular target area by, e.g., wardriving nearby public WiFi hotspots.

We hope that our work will spur further analyses of this critical infrastructure. This paper makes the following contributions:

• **We identify concrete threats to operational regional access networks.** Through conversations with operators at the largest U.S. access networks and by analyzing recent results on mapping access network topology [216], we describe how the redundant power and packet-transport infrastructure currently in place to withstand natural events is insufficient for intentional attacks.

- **We study the root cause and impact of large access network outages.** We combine inferred network infrastructure maps with continuous reachability measurements to millions of access network customers to detect outages and identify the failed infrastructure. We investigate outages of different magnitudes in detail, including the Nashville bombing. These outages indicate that the scale of an attack's impact can be expected to range from thousands to hundreds-of-thousands of users, and the duration to span hours to days.

- **We show that targeted attacks can be launched without insider information.** By combining public hazardous-material datasets with targeted use of the ubiquitous traceroute tool, we show that an attacker can learn the location of infrastructure whose failure will disconnect specific areas. We demonstrate feasibility in three different networks.

- **We explore potential ways to mitigate risks.** Access networks must balance infrastructure security with manageability and cost, and we explore trade-offs associated with mitigating physical threats to the infrastructure.

Ethical considerations.

The Menlo Report [85, 52] explicitly addresses stakeholders such as network/platform owners in the context of revealing information about critical infrastructure that may provide advantages to adversarial actors. These principles, and feedback from network operators, guide our approach to anonymization and disclosure of details about networks. We anonymize details when we explore the attack surface of different networks (Section 4.6 and Section 4.7), but do not anonymize networks or locations in case studies (Section 4.5) when those details appear in the public press. All three operators we consulted were eager to understand what could be gleaned about their infrastructure by a capable independent third party and how they could raise the bar for attacks.

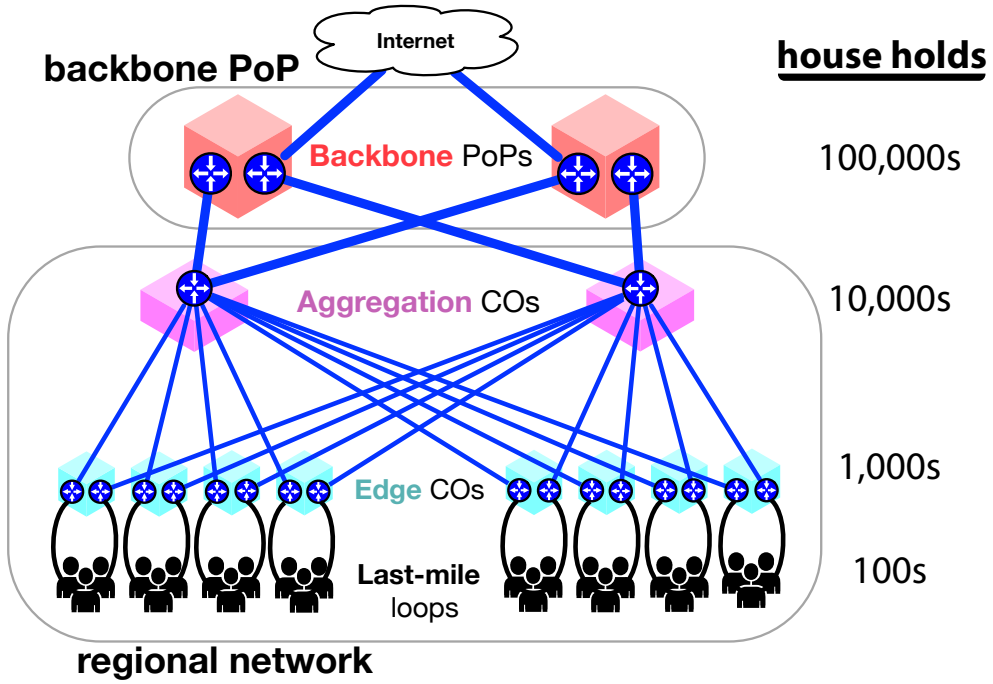


Figure 4.1. Access network hierarchy: EdgeCO routers aggregate customers and AggCO routers aggregate EdgeCOs.

4.2 Background: Access Network Topology

Internet Service Providers (ISPs) design access networks with significant redundancy to withstand common failures that occur through random chance, like trees falling on overhead fiber or mains power outages. This redundancy provides some protection against physical attacks as well: networks can continue to function as normal after incurring a fiber or power cut. However, physical attacks that damage the backup systems as well can lead to widespread outage, as we will show. To understand this risk, we describe the general architecture of Internet access networks (Section 4.2.1) and discuss where access networks deploy topological redundancy (Section 4.2.2).

4.2.1 Key Topological Elements

Access networks consist of dense deployments of fiber optic cables—and often also powered equipment—in nearly every neighborhood in the geographic regions where they provide service (e.g., a metropolitan area). To provide Internet access, each access network connects back to a small number of Internet backbone routers in one or more Internet Points-of-Presence (PoPs). Providers design their networks to achieve this connectivity efficiently by aggregating traffic through a hierarchy of facilities known as *COs*: these buildings aggregate traffic with last-mile link technologies and switches, and pass traffic up or down the hierarchy with inter-CO routers.

The general network topology of a typical access network is shown in Figure 4.1. An *Edge CO* (or EdgeCO) aggregates traffic from hundreds to thousands of customers over last-mile links; e.g., cable, DSL, and fiber. Similarly, an *Aggregation CO* (or AggCO) aggregates traffic from dozens of EdgeCOs providing service for hundreds of thousands of users—often across metropolitan areas or entire states. *Backbone Points of Presence* (Backbone PoPs) aggregate traffic from one or more AggCOs and provide Internet transit services over a backbone network operated by the ISP or another provider.

4.2.2 Redundant Infrastructure

Based on the topologies of major U.S. access networks revealed in recent work [216] and conversations with network operators, we explore differences in how ISPs deploy redundancy at different layers of regional access networks.

Some networks deploy redundant last-mile connections using fiber rings, letting them survive a single fiber cut to the ring. Well-provisioned networks may even terminate the ring at two different EdgeCOs to provide CO-level redundancy, although deploying and maintaining multiple last-mile connections is expensive. In most networks it is only economical to deploy a single last-mile link to each customer. Without redundancy, a single cut to a last-mile cable bundle

will disconnect all customers downstream from the EdgeCO on that fiber strand. Additionally, depending on the last-mile technology used in the network (DOCSIS cable, DSL, etc.), an attacker may be able to disconnect multiple users by cutting a single link in a neighborhood (e.g., DOCSIS feeder coax).

EdgeCOs aggregate thousands of last-mile links that terminate at specialized devices inside the CO; e.g., CMTS in cable networks or DSLAM in DSL networks. Often, adding redundant last-mile links to different EdgeCOs is cost prohibitive, so customers connect to a single EdgeCO. As a result, an EdgeCO outage will disconnect all downstream last-mile customers. A group of EdgeCOs connect to one or more AggCOs through a fiber ring. When a group of EdgeCOs connects to two or more AggCOs, each AggCO interconnects with each EdgeCO in one direction around the ring, allowing the EdgeCO to survive a single AggCO outage.

Smaller regional networks contain a single AggCO layer with one or two AggCOs. If there is only one AggCO, then an attacker can disconnect the entire region by attacking that one CO; if there are multiple, the network can survive one going down. Larger regions often employ multiple AggCO layers, where some AggCOs might only aggregate traffic from other AggCOs. Some providers split their aggregation layers into two or more subregions and use separate fiber rings with one or two AggCOs, so a failure of one ring will not take down all of the region's EdgeCOs.

At the top of the aggregation hierarchy, one or more AggCOs, which serve as entry points into the regional access network, connect to one or two Backbone PoPs, and occasionally interconnect with large transit ISPs as well. If a region only has one Backbone PoP and that PoP is taken offline, all customers in that region will be disconnected from the Internet. In regions that have more than one AggCO and Backbone PoP, each AggCOs usually connects to a different Backbone PoP. This configuration allows the the entire region to fail over to the other Backbone PoP if one Backbone PoP fails.

4.3 Threat Model

This section describes the physical attacks we consider on regional access networks, where the attacker's objective is to cause widespread connectivity outages. We first discuss how an attacker—without insider knowledge—can damage physical plant, such as fiber and power (Section 4.3.1). Then we discuss why existing redundancy insufficiently addresses the threat of intentional attack (Section 4.3.2).

4.3.1 Attacker Capabilities

In this work, we show how an attacker without insider knowledge can cause large-scale outages. We demonstrate that motivated attackers can combine network measurement tools with public information to identify minimum cuts in the access network dependency graph and target specific users.

Attackers can damage underground and overhead fiber.

Access networks are built out of fiber optic cables containing bundles of fiber optic strands that are deployed aerially along telephone poles or underground in cable vaults. In both cases, the fiber runs unprotected over large distances, and attackers can cut them using widely available wire cutters. Attackers can visually identify a provider's cables because they often use fiber ID tags on aerial lines, and marker poles and labeled cable vaults on underground lines. An attacker can reach aerial fiber by climbing telephone poles or damaging the poles themselves [96] and cut underground fiber with digging equipment or by accessing the cable vault. An individual attacker can also cut multiple fiber bundles in different locations before the ISP can repair the fiber. Simply detecting the location of damaged fiber can take minutes to hours [174], in part because the provider must dispatch repair crews to the fault location(s).

Recent examples demonstrate the risks for fiber deployments. For instance, between 2009 and 2016 there were more than a dozen incidents of vandals cutting fiber optic cables in California [118]. Two of the attacks disrupted AT&T's access network for hours and led them to

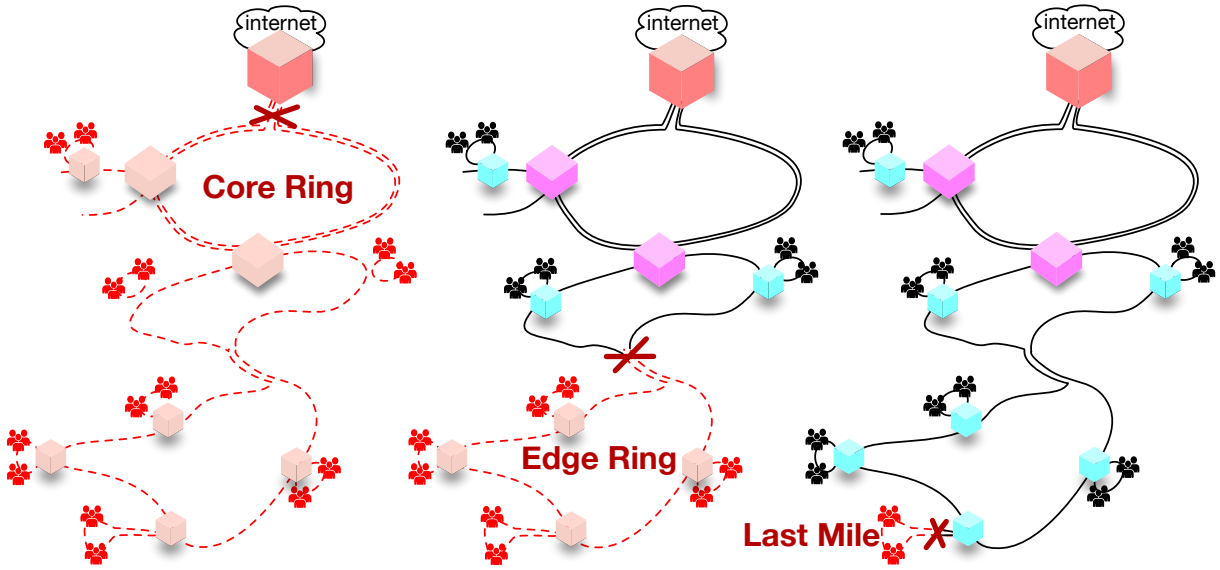


Figure 4.2. An attacker can easily cut fiber rings when both sides of the ring run in parallel.

offer a $250,000 reward for information about the culprits [63, 160].

Attackers can disrupt mains power and backup fuel.

Access networks require power inside facilities and out in the field to maintain network operations. An attacker can cut the mains power serving this infrastructure, forcing the network to rely on backup power, and that backup power may run out; e.g., due to lack of fuel. Also, an attacker can damage the mains and backup power simultaneously, which is what occurred in the Nashville bombing [143].

4.3.2 Threats to Fiber and Power Redundancy

ISPs design COs and last-mile links with redundancy to continue operating in the face of a single fiber cut or loss of power. Across ISPs, the conventional approach is duplicating nearly every piece of infrastructure related to power and network transport, such that if one component fails, the redundant component can seamlessly take over.

Fiber Rings.

ISPs physically deploy fibers in a ring topology to aggregate traffic from multiple COs to the CO in the next hierarchy level because rings are resilient to a single fiber cut at any location on the ring: traffic can route in the remaining direction around the ring to maintain connectivity until the fiber cut is repaired [212, 196, 70, 39]. If an attacker cuts a fiber ring in two places, however, it will disconnect all COs and customers downstream of both cuts. Some fiber rings are especially susceptible to this attack because economic or geographic constrains might force an ISP to run both sides of the ring close together (Figure 4.2).

Backup Power.

COs are typically provisioned with backup power that seamlessly maintains operation during a power outage for approximately 24 hours until the mains power returns. Although mains and backup power are largely independent, they generally meet at a central power control system. This presents an opportunity for an attacker to induce an outage that takes both systems offline.

A physical attack can proceed in three phases: (1) The attacker selects the customers they want to take offline, or the ISP they want to damage. (2) The attacker finds the fiber or power nearest to those customers or ISP's facilities by looking for markings on cables and vaults. (3) The attacker cuts fiber and/or or disables power. Figure 4.3 shows how an attacker can disconnect part of the access network by either cutting the fiber ring in two places or disabling all power input into powered network equipment.

4.4 Experiment Methodology

Having established that access network COs remain vulnerable to intentional attack, we empirically measure the expected impact of a CO outage. Specifically, after a successful attack, we examine how many users would likely experience an outage, and for how long. We design a measurement study of the three largest residential ISPs in the U.S.—Comcast, Spectrum, and

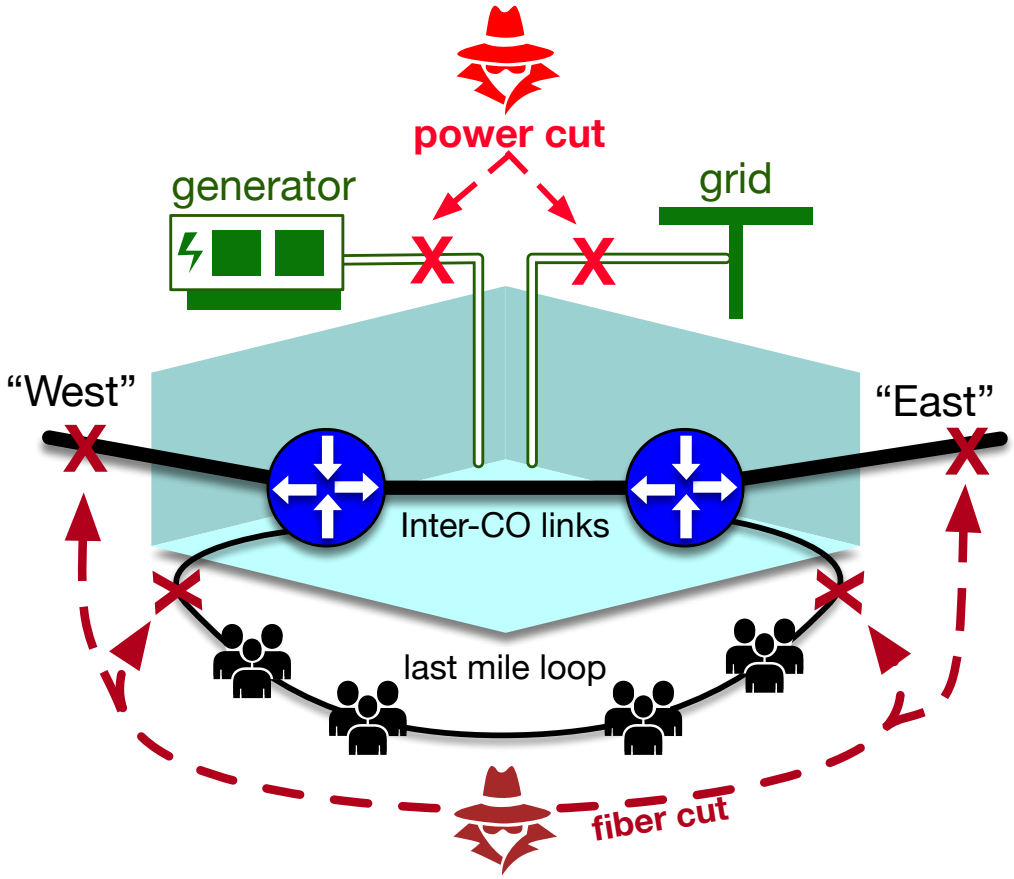


Figure 4.3. An attacker must disable either a CO's redundant power or redundant fiber to induce a failure.

AT&T—with two goals: (1) estimate the number of customers connected to COs across the U.S., and (2) leverage weather, accidents, and vandalism to empirically learn the scale and duration of CO outages.

Our analysis of CO outages proceeds in three stages. First, we create maps of each regional access network that capture the CO-level topology (Section 4.4.1). Second, we infer the customer IP address space connected to each CO (Section 4.4.2). Third, we continually send probes to customers of the access networks to observe when a CO experiences an outage and to measure the outage duration (Section 4.4.3). When possible, we add context from news stories to confirm that an attacker could intentionally recreate the failures we observe.

4.4.1 Mapping Regional Access Topologies

Our experiment touches 22 of the regional access networks that Comcast, Spectrum, and AT&T deploy across 14 U.S. states. We conduct large-scale measurements to create CO-level maps of these regional access networks.

At the core of our technique, we use the traceroute tool to reveal router IP addresses between a measurement VP and an arbitrary destination. Traceroute induces a single response from each router along the path containing the IP address assigned to an interface on the router. To increase the likelihood that our path measurements reveal all active paths through the regional networks, we use measurement VPs distributed across the U.S. Our VPs conduct traceroutes to customers connected to the networks, revealing the IP topology of each regional access network. We use the same techniques as Zhang *et al*. [216] to infer CO interconnections and aggregation hierarchies in each access network from the IP topologies.

4.4.2 Mapping Customers to COs

The techniques from Zhang *et al*. [216] reveal CO interconnections, and substantial prior work observed last-mile outages [131, 132, 155, 137], but no prior work has tied those outages to network facilities. To support tying outages to COs in Comcast and Spectrum, we also create

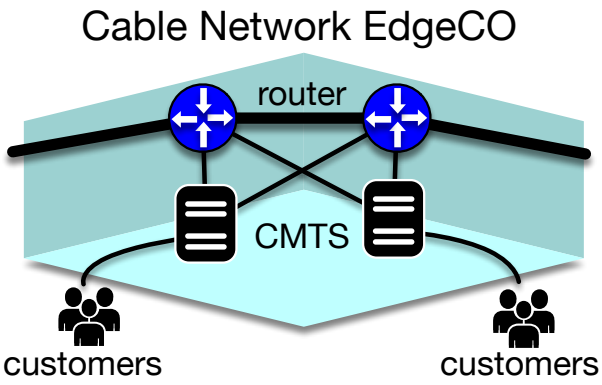


Figure 4.4. The routers and CMTSes inside EdgeCOs appear in traceroute paths.

mappings from customer address space to COs; i.e., the IP address ranges used by customers attached to a given CO. In the access networks, each CO assigns addresses to customer devices from a pool of addresses allocated to that CO. That pool consists of hundreds-to-thousands of IPv4 /26 subnets, and we infer the pool of residential /26s for each CO in Comcast and Spectrum.

Mapping from customer IPs to COs would be trivial if DNS names always indicated the CO for the IP address immediately before the customer in a traceroute path, but many of those addresses either lack a CO identifier or lack DNS names entirely. Instead, we leverage technical details of cable access-network infrastructure to infer comprehensive CO-to-address mappings. In particular, the cable modem termination systems (CMTS) housed inside cable-network COs respond to traceroute probes, so one hop prior to the customer is the CMTS, and two hops prior is a router in the same CO [74] (Figure 4.4). By sending traceroutes to every residential customer IP address, we construct a directed interface graph with edges between immediately adjacent hops. We cluster each customer IP address with all preceding addresses within distance two, allowing us to infer a CO mapping for the cluster rather than separate mappings for the individual IP addresses. The transitive closure of each cluster includes the customers, CMTS devices, and routers that all map to the same CO. Finally, we use the CO identifiers that Comcast and Spectrum include in many—but not all—hostnames for their router and CMTS IP addresses to map clusters to COs.

When EdgeCOs have multiple CMTS devices, we might observe different routers prior

to disjoint sets of CMTSes, creating two different clusters for a single CO. We evaluated this potential problem on Spectrum's access networks, which have good hostname coverage for EdgeCO router IP addresses. Clustering the IP addresses created 860 clusters where a hostname let us infer the CO identifier. Only 7.2% of the clusters received an identifier that was also assigned to another cluster, indicating a partial CO cluster. Our approach appears to work well for the other 92.8% of the CO clusters.

For AT&T, traceroutes to most residential customers failed to induce responses from routers within the access network. As a result, we only mapped AT&T customers to COs in one regional network and partially mapped customers in another region. We used the same technique as Zhang *et al*. [216] to estimate the customers connected to EdgeCOs by conducting traceroutes from various locations within the access network.

4.4.3 Detecting CO Outages

To detect CO outages, we continuously test reachability to the residential customers in each regional access network. Testing reachability of customers—rather than routers in the COs—ensures that any event we detect actually disconnected customers; i.e., the redundancy in the network failed to prevent an outage. We detect CO outages when all customers that depend on the CO experience an outage simultaneously.

We test reachability for Comcast and Spectrum by pinging access network customer addresses every ten minutes from three different VPs. We ping a static set of customer addresses consisting of 50% of the customer addresses for each network across 14 different U.S. states. Using this customer sample allows us to comprehensively detect outages at 10-minute granularity while bringing the financial cost of virtual machines and egress traffic from the cloud within our constraints. Three VPs ping each customer in our set in every ten-minute round, and we consider a customer responsive in a round if it responds to any of the three pings.

To detect CO-level outages, we find 10-minute rounds where all customers of a CO failed to respond to all three VPs. First, we compute the median number of responses for each

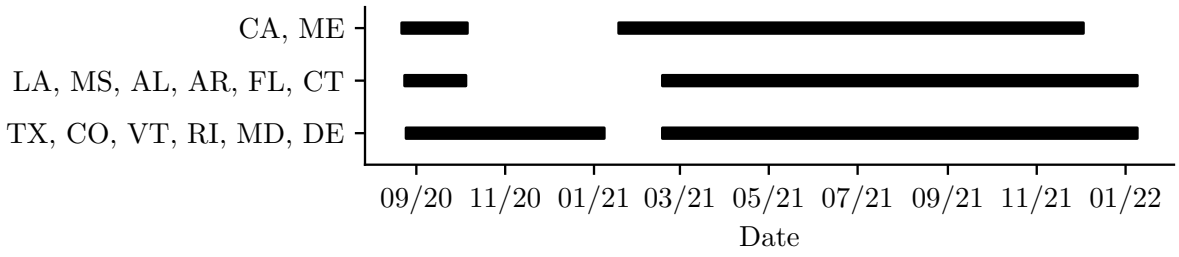


Figure 4.5. Bars indicate the measurement period for different regions in our study of Comcast and Spectrum. Gaps corresponds to configuration errors that prevented data collection.

CO and /26 subnet across all 10-minute rounds in each week of data. To reduce the likelihood of misclassifying last-mile failures, dynamic IP address reassignment, or transient customer device unresponsiveness as CO-level failures, we only consider COs with a median of at least 100 responding customers spread across 20 or more /26 subnets. Next, we iterate over each 10-minute round to identify COs without any responding customers, and the number of consecutive 10-minute rounds with no responding customers quantifies the outage duration. Using the CO interconnection maps we can also infer failures higher up in the access network aggregation hierarchy, when all EdgeCOs dependent on a set of AggCOs fail simultaneously.

We cannot detect outages in AT&T with the same granularity, since AT&T customer devices generally did not respond to our pings. Instead, we use traceroutes toward AT&T customers to observe when portions of an access network disappear at the same time; i.e., when previously observable COs disappear from the traceroutes. CAIDA's Ark [30] measurement platform uses globally distributed VPs to continually send traceroutes to every IPv4 /24 multiple times a day [28]. To detect outages, we look for periods of time where all traceroutes from Ark VPs fail to observe one or more COs. Ark conducts traceroutes less frequently than we conduct our pings for Comcast and Spectrum, so we can only observe AggCO outages that last for several hours in AT&T.

Table 4.1. Observed outages in Comcast, Spectrum, and AT&T.

CO Type	ISP 1	ISP 2	ISP 3	Total
Backbone PoP	0	0	1	1
AggCO	4	1	0	5
EdgeCO	40	24	0	64
Total	44	25	1	70

4.5 Outage Case Studies

We collected outage data for Comcast and Spectrum between August 2020 and December 2021 (Figure 4.5) and looked for AT&T outages in the December 2020 Ark traceroutes. We observe 70 outages where our reachability tests failed to reach any customer behind a CO (Table 4.1). Five outages affected all EdgeCOs downstream of a set of AggCOs, indicating problems either at or near the AggCOs. We observed at least one CO outage in 11 out of the 14 states we probed. The outages mostly lasted between 50–200 minutes, with the median outage lasting 1 hour and 10 minutes (Figure 4.6), and typically affecting 4,800–34,000 customers. The longest outage lasted nearly 3 days following Hurricane Ida in Louisiana, and the largest outage disconnected an entire access network in California that serves over 2M customers for 50 minutes.

Our approach cannot distinguish scheduled maintenance outages from failures, and ISPs cannot reroute customers during scheduled maintenance that requires disconnecting a CO. Because networks often perform scheduled maintenance between 00:00–05:59 local time [43, 163], we classify an outage as *overnight* if it occurs within that time window. Figure 4.6 shows that while overnight outages tend to be short, they can cover many customers.

The remainder of this section discusses specific outages (Table 4.2) that suggest the potential impact of successful physical attacks against access networks. We withhold CO locations when not revealed in news stories.

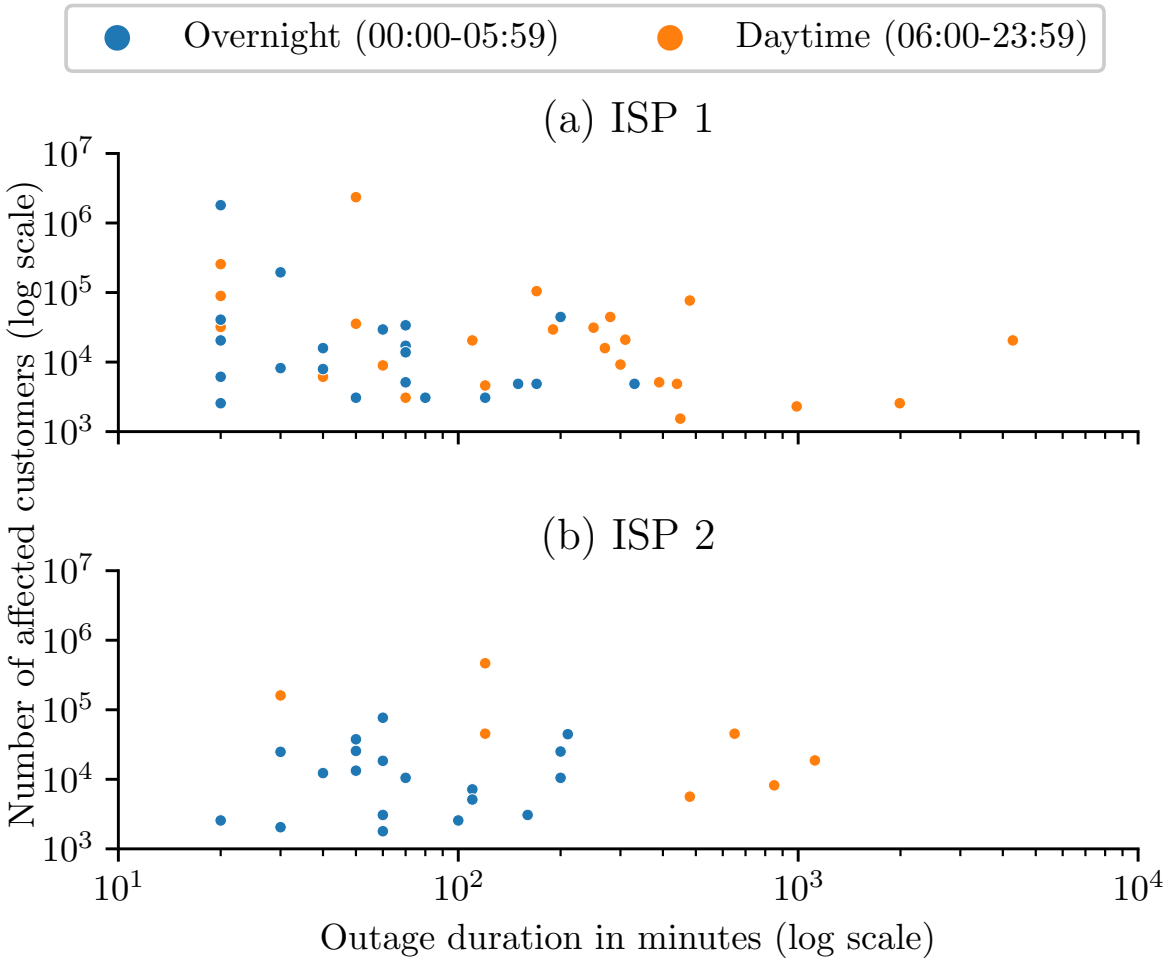


Figure 4.6. Outage duration and number of affected customers. We classify each outage as either overnight, when scheduled maintenance is common [43, 163], or daytime.

4.5.1 Case Studies: Backbone PoP Outage

The AT&T backbone PoP failure in Nashville, Tennessee caused widespread outages. On December 25, 2020, a van exploded on the street outside the AT&T Nashville backbone PoP. The explosion disconnected the facility from mains power and caused the backup generators to fail [143]. Battery backups maintained operations for several hours but the PoP went offline when they exhausted.

The PoP failure disconnected all AT&T wireline customers in the greater Nashville metropolitan area, but AT&T provides more than residential Internet access over the wireline

Table 4.2. Our case studies suggest the potential duration and scale of successful attacks against access network COs.

Failure Type	ISP	COs	Duration	Customers	Location	Date	Time
Backbone PoP Outage (Section 4.5.2)							
Single PoP	AT&T	41	31h	229,632	Nashville, TN	2020-12-25	07:10
AggCO Outages (Section 4.5.2)							
Multiple AggCOs	Spectrum	44	2h	388,608	Maine	2021-04-05	17:20
Degraded Service	AT&T	0	16h	0	San Diego, CA	2020-12-20	08:16
EdgeCO Outages (Section 4.5.3)							
Multiple EdgeCOs	Spectrum	12	30m	294,400	Los Angeles, CA	2021-02-22	18:00
Single EdgeCO	Comcast	1	40m	3072	Rio Vista, CA	2021-02-25	16:20

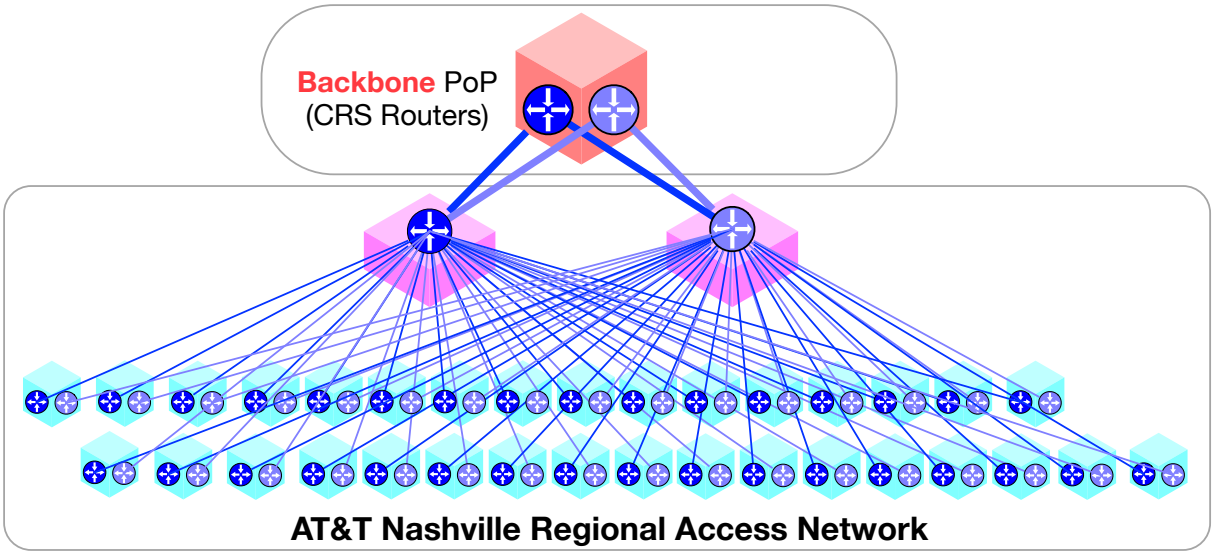


Figure 4.7. The AT&T Nashville access network relies on a single Backbone PoP. When that facility failed, it disconnected this entire access network from the Internet.

access network. AT&T wireless also used the access network facilities to reach the AT&T backbone [143]. Worse still, 911 emergency services [66], air traffic control [168], and hospitals [91] all relied on that AT&T access network for communication.

News reports explain how the PoP failed [143], but not why the single PoP failure disconnected hundreds of thousands of AT&T customers in and around Nashville, as well as vital services in the area. To understand why, we generate a topology map of AT&T's Nashville access network with Ark traceroutes (Figure 4.7), and compare that to the observable topology during the 24 hours after the explosion. The maps reveal that all traffic into the Nashville access network

passed through two core routers (i.e., CRS routers). During the outage, these two routers—and all routers previously observed downstream of the CRS routers—disappeared from the topology, indicating that the explosion took the CRS routers offline. This explanation is congruent with AT&T outage reports indicating that both CRS routers in Nashville experienced an outage [92], and an operator at AT&T confirmed that a single facility houses both CRS routers.

We confirmed that all AT&T customers throughout the greater metropolitan area relied on those CRS routers (Figure 4.8), explaining the geographic scale of the outage. Like Zhang *et al*. [216], we revealed AT&T's CO-topology in Nashville by conducting traceroute probing from publicly available WiFi access points in April 2021. We connected to three McDonalds' and eight business WiFi networks available through Instabridge [5] around the city that are AT&T customers, sending traceroutes from each location to destinations outside AT&T's network. Every traceroute from the eleven customers passed through one of the CRS routers in the Nashville PoP.

The Ark traceroutes let us retroactively watch as AT&T restored the Nashville access network, and discover that the PoP required only one CRS router. Starting at 16:14 on December 26, the Nashville PoP appeared in paths forwarding traffic to other backbone PoPs. Finally, on the morning of the 27th, we again observed downstream access network COs in the traceroute paths. Consistent with AT&T recovery reports [92], it appears that AT&T initially restored only one CRS router in the PoP along with its fiber connectivity, the minimum needed to restore connectivity to the regional network. We finally observed the second CRS router at 12:00 on December 28th, more than three days after the outage began.

While the bombing likely did not intentionally target the AT&T facility [193], it suggests that intentional attacks could similarly disrupt access network connectivity. AT&T appears to use a single PoP to reach other regional access networks, for instance Zhang *et al*. [216] found one entry PoP housing the two CRS routers in another regional network as well. The outage in Nashville also illustrates the risk of relying on a single access network for many different critical services: a single outage can disrupt nearly all communications in a geographic area. Outages

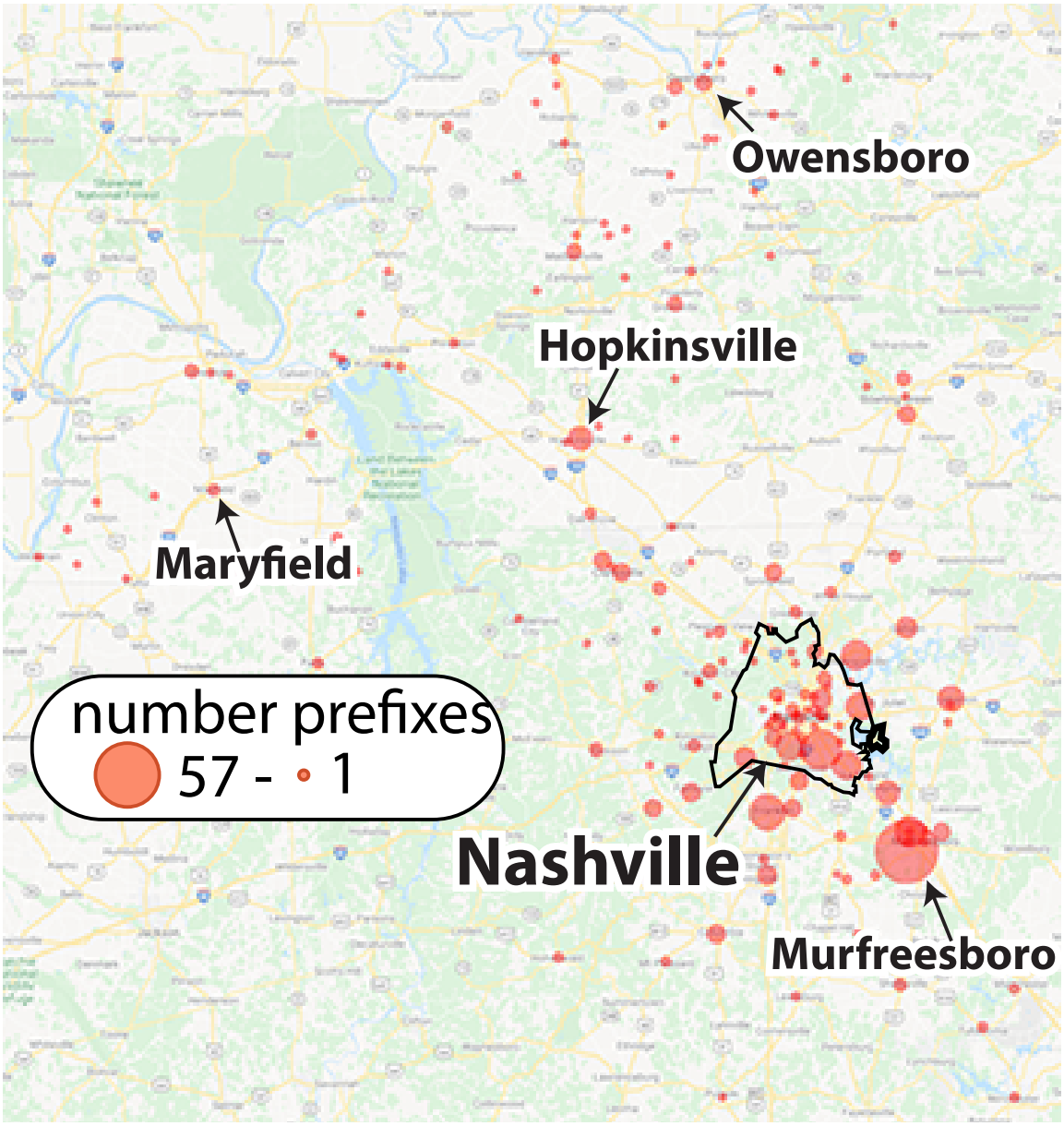


Figure 4.8. Ark traceroutes reached AT&T customers outside the city limits (black border) via the Nashville CRS routers. /24 prefixes (red dots) geolocated with NetAcuity.

that affect entire regional networks can even disconnect mobile networks [143], so LTE backup might not provide the redundancy that many expect.

4.5.2 Case Studies: AggCO Outages

Next, we discuss an outage that disconnected all Spectrum customers in the state of Maine, and another that degraded service in AT&T's San Diego access network.

Two Fiber Cuts Disconnect All Spectrum Customers in Maine.

On April 5, 2021, all Spectrum customers in Maine stopped responding to our pings for two hours (Figure 4.9). The outage included 1518 /24 subnets, indicating a maximum of 388,608 residential customers. Spectrum disclosed that two separate fiber cuts caused the state-wide outage:

> We've identified two separate fiber breaks in our network, impacting services for Spectrum customers in Maine and New Hampshire... These separate breaks have impacted our redundant path, which normally serves as backup... [117]

This explanation is precisely consistent with our assumptions, since it requires two fiber cuts to disconnect COs.

However, the press release does not indicate why two fiber cuts could disconnect all Spectrum customers in Maine. Spectrum is the largest broadband ISP in the state of Maine, but includes Maine in its larger Northeast regional access network. From our map of Spectrum's Northeast region (Figure 4.10), we learn that any IP packet sent to residential customers in Maine must pass through one of two AggCOs in upstate New York. From there, it goes to one of the two AggCOs in Maine. All EdgeCOs in Maine connect to both of the Maine AggCOs, and an EdgeCO needs a connection to only one of the two AggCOs to remain connected to the access network.

The map reveals that fiber cuts between the Maine AggCOs and the entry AggCOs are the only scenario that could disconnect all Maine customers from the Internet but not affect the rest of Spectrum's Northeast regional access network. Without that fiber connectivity, Spectrum

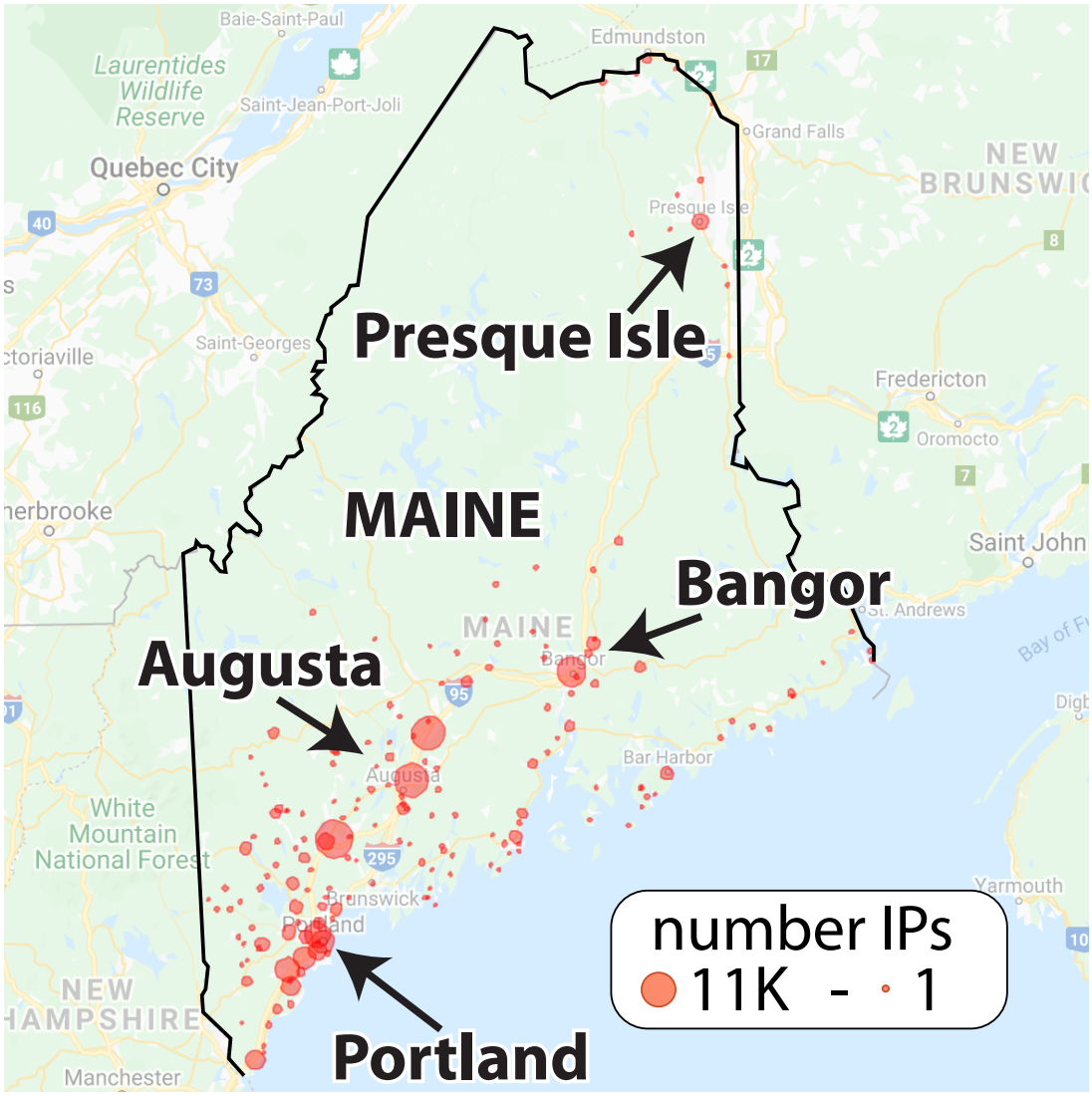


Figure 4.9. Spectrum customer IP addresses (red dots) were disconnected throughout Maine. Geolocated with NetAcuity.

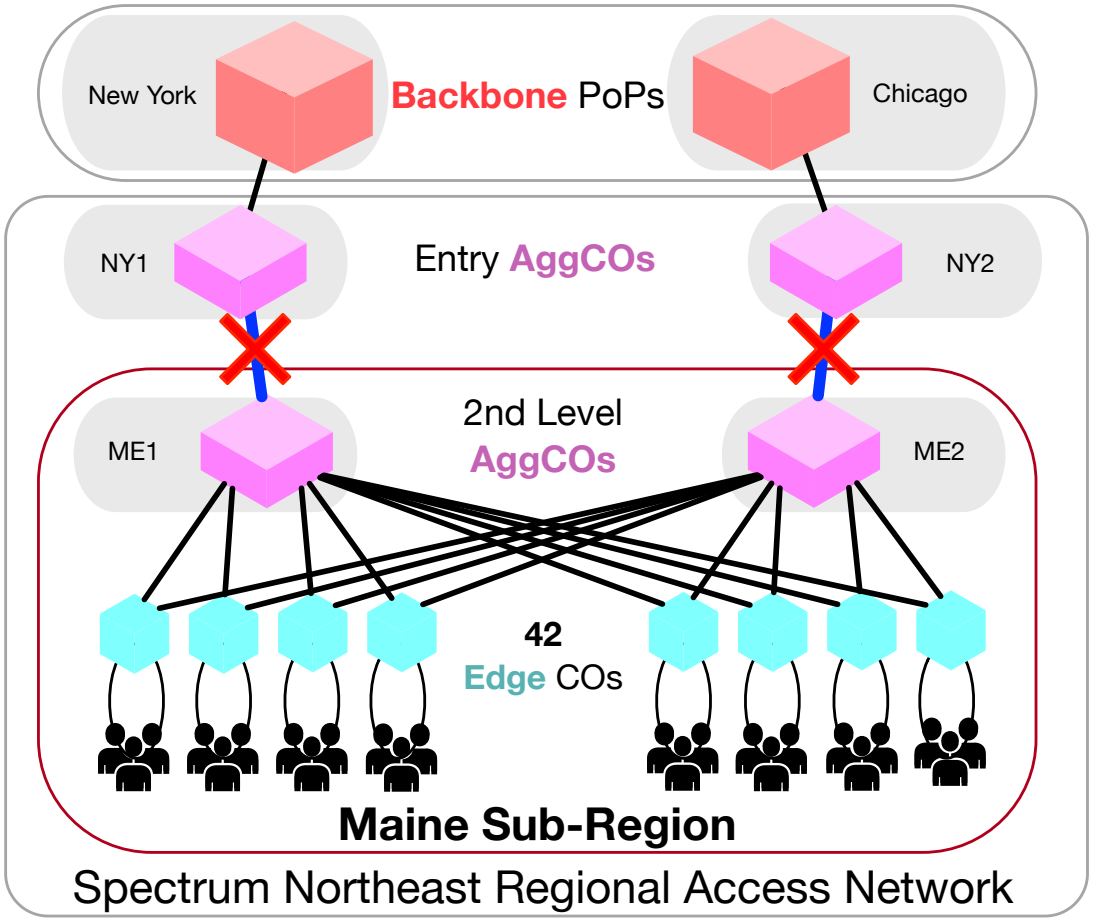


Figure 4.10. Spectrum's Maine sub-region includes two AggCOs leading to every EdgeCO. Two fiber cuts disconnected the AggCOs from the rest of the access network [117].

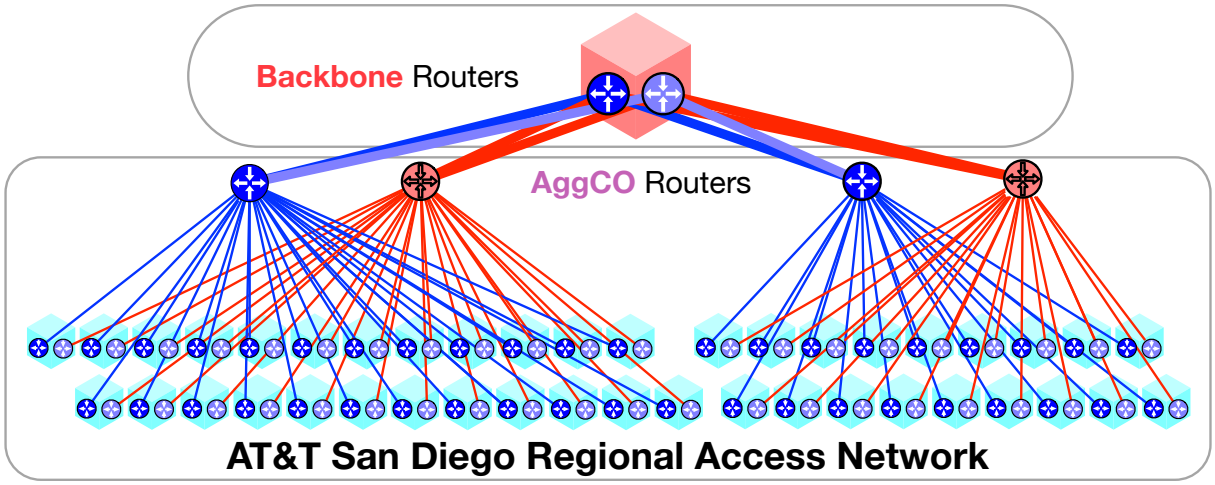


Figure 4.11. A partial outage appears to disconnect two AggCO routers (red), but customers remained connected.

customers in Maine could not connect to the rest of the access network or reach Spectrum's backbone. Furthermore, our pings included some Spectrum addresses connected to COs in upstate NY that did not depend on the Maine AggCOs and remained reachable throughout the outage.

Importantly, the outage confirms our hypothesis that the effects of AggCO outages cascade to their downstream EdgeCOs. It also suggests that an attacker might have hours to cut multiple fibers in different locations to cause large-scale outages. The Maine EdgeCOs required only one connection to the upstream AggCOs, but it took at least two hours to bring customers back online, indicating it took Spectrum at least two hours to fix one of the fiber cuts.

Degraded Service After AT&T AggCO Failure in San Diego.

We also examined a likely AggCO outage where the 2× redundancy maintained customer connectivity. According to our our map of the San Diego AT&T regional access network (Figure 4.11), all EdgeCOs connect to two of four AggCOs, which in turn connect to the two CRS routers in the San Diego backbone PoP. On December 18, 2020, two of the four AggCO routers disappeared from the Ark traceroutes for 16 hours (shown in red), leaving only half of the IP-level topology visible. The disappearance of these two AggCO routers suggests that they

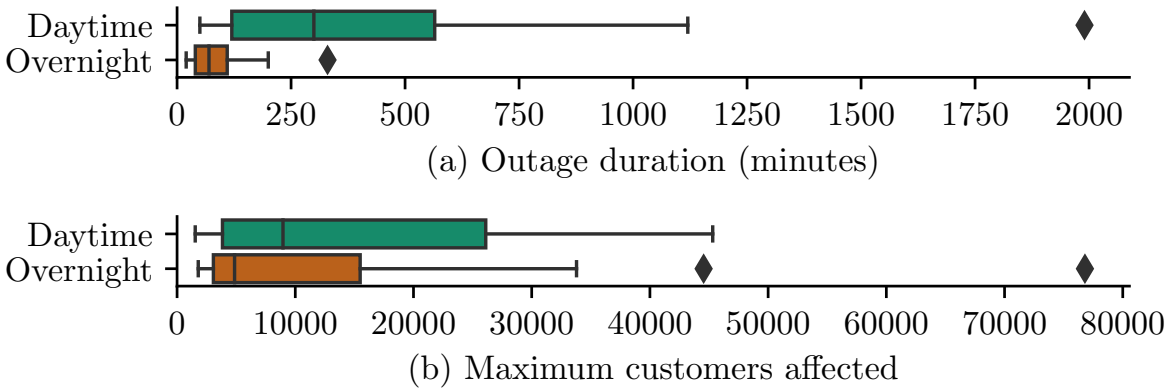


Figure 4.12. Single EdgeCO outages typically lasted 1–4.5 hours (a) and impacted 3–20K customers (b).

became disconnected, yet the Ark traceroutes continued to reach customers of the San Diego access network through the remaining AggCO routers.

Although the redundancy maintained AT&T customer connectivity, it appears that the remaining path could not handle peak traffic demand. Between 19:00 and 22:00 local time, Ark traceroutes inconsistently revealed the San Diego access network CO routers. The most likely explanation is that the increase in traffic during peak Internet usage hours congested the remaining CO interconnections, degrading customer connectivity. This explanation is consistent with DownDetector data [127] showing an increase in customer outage reports starting at the same time. This case shows shows that even when redundancy prevents a widespread access network outage, an attacker could still cause degraded service.

4.5.3 Case Studies: EdgeCOs Outages

In our study, EdgeCO outages without a corresponding AggCO outage occurred most commonly. Of those, 15 outages disconnected all customers connected to multiple EdgeCOs, and the remaining 49 outages affected a single EdgeCO. The single EdgeCO outages help indicate the expected fallout from an attack against an EdgeCO (Figure 4.12); they typically lasted 1–4.5 hours and affected 3–20K customers. We focus specifically on a multi-CO Spectrum outage in Los Angeles and a Comcast EdgeCO outage in Rio Vista.

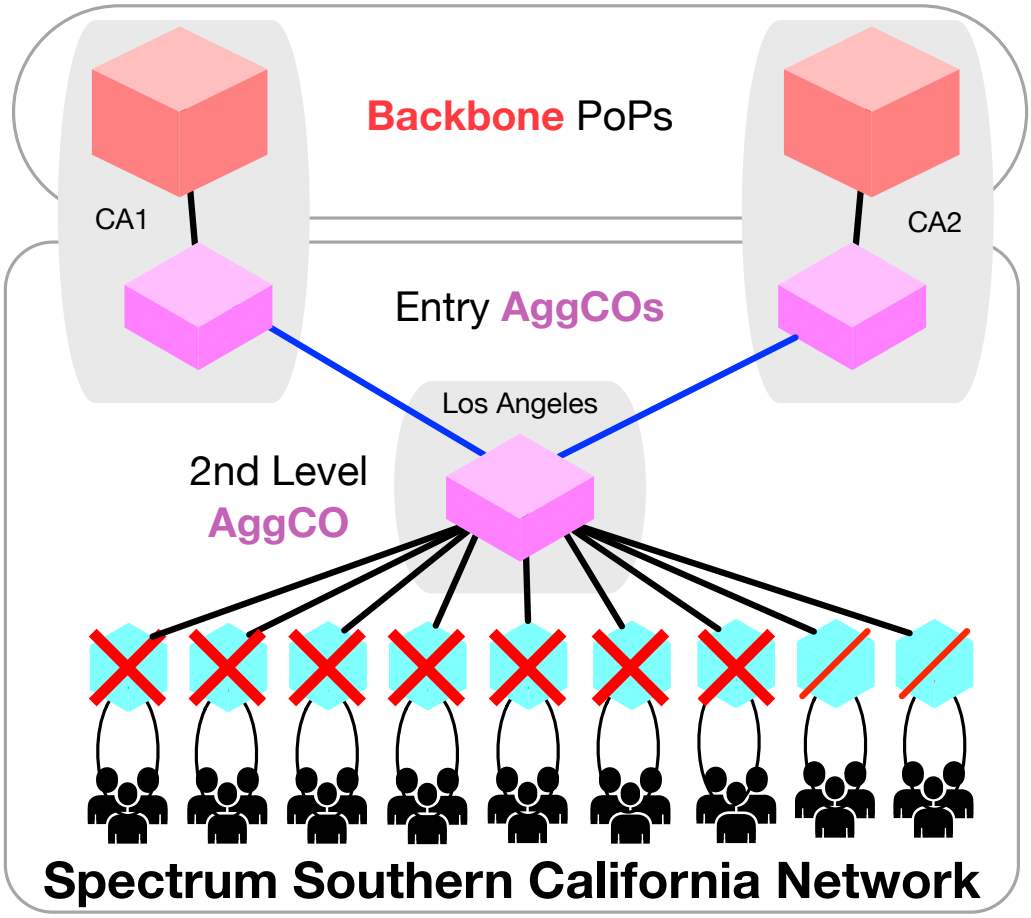


Figure 4.13. Multiple EdgeCO outage in Los Angeles, California affected up to 294,400 residential customers.

Nearby Spectrum EdgeCO Outage in Los Angeles Without AggCO Failure.

On February 2, 2021 we observed evidence that EdgeCO outages are not always independent. The outage spanned multiple Los Angeles EdgeCOs in Spectrum's Southern California regional access network (Figure 4.13), but the outage did not appear to originate at an AggCO. Starting at 18:00 and lasting 30 minutes, the outage disconnected 8 EdgeCOs from their single upstream AggCO in Los Angeles and degraded service to two other EdgeCOs. News reports confirmed the outage and its duration [8], but Spectrum did not publicly disclose the cause of the outage. This outage shows that even connecting to two EdgeCOs might be insufficient, since an attacker might be able to disconnect nearby COs simultaneously.

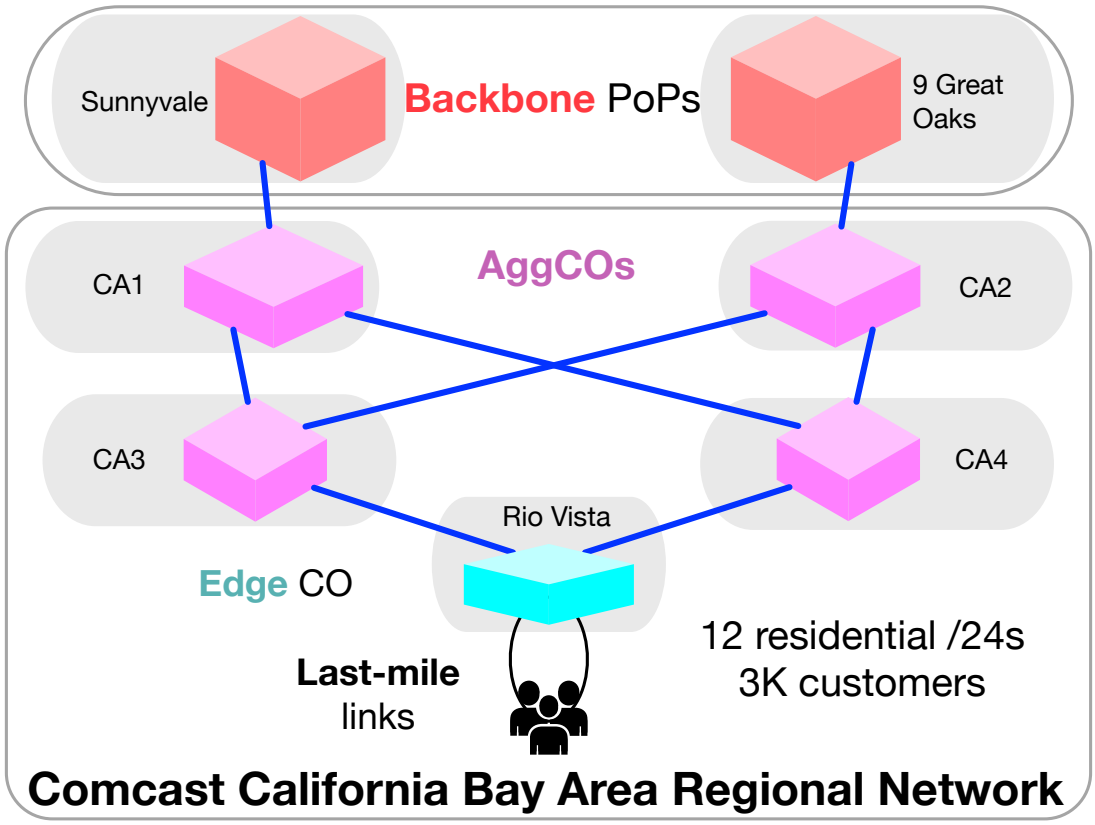


Figure 4.14. EdgeCO outage in Comcast's Bay Area regional network, affecting 3K residential customers for 40 minutes.

EdgeCO Outage Disconnected Customers From 911.

An outage in Rio Vista, California highlights that EdgeCO outages can affect customers in ways that customers might not expect. At 08:50 on March 23, 2021, we observed a 40-minute Comcast EdgeCO outage in Rio Vista, California that disconnected up to 3,000 customer devices (Figure 4.14). The Solano Country Sheriff's office reported the outage [162] to warn that during the outage Comcast-provided phone service could not reach 911 emergency services.

4.5.4 Security Takeaways

The outages and case studies illustrate three key access network properties that facilitate intentional attacks. (1) Combinations of power failures and fiber cuts frequently disconnect COs, despite their redundant design. If an attacker can disrupt power or fiber connectivity, they

will disconnect the CO. (2) Residential customers are typically connected to a single CO, and CO failures disconnect their connected customers. Attackers can target a single EdgeCO to target customers within the local geographic area. (3) Entire regional access networks can fail, as evidenced by the statewide outage in Maine and the complete access network failures in northern California and Nashville. Current access network design in the U.S. allows an attacker to disrupt Internet communication for millions of people by targeting specific COs, without hidden redundancy to maintain connections.

4.6 Feasibility of Targeted Attacks

Our synthesis of topology mapping with case studies of real outages demonstrates that attackers could disrupt Internet connectivity with physical attacks on COs or last-mile links. However, these case studies do not reveal if it is feasible to perform a targeted attack to disrupt a specific entity or geographic area. In this section, we show that attackers can precisely locate COs and predict the affected geographic area.

Hazardous Materials Records Can Locate COs.

Surprisingly, we find that safety regulations increase availability of CO street addresses. To satisfy FCC backup power requirements [58], COs typically use on-site diesel generators and battery cells as redundant power sources. These materials pose fire hazards, so local authorities require the networks to register the capacity and location of storage tanks and other hazardous materials (hazmat) with regulatory bodies. These records are often public [181, 179, 31], revealing CO street addresses in a geographic area (Fig 4.15). We implemented scripts to crawl hazmat records from four different regulators in the US, demonstrating the accessibility of the data.

Wardrive to Predict Service Areas.

Customers are not necessarily served by the closest EdgeCO due to regulatory, geographic, and financial constraints, but traceroutes in the target area can reveal the EdgeCO serving an

Figure 4.15. Three Florida CO locations from hazmat records.

area. Specifically, an attacker can cluster access points to the EdgeCO serving them using a "wardriving" approach to conduct traceroutes via public WiFi access points (APs) in fast-food restaurants and coffee shops, such as McDonald's and Starbucks. As a proof-of-concept, we conducted traceroutes from 114 public WiFi APs in southwest San Diego County to a server in our lab, and estimated the geographic service areas for each of an ISP's EdgeCOs (Figure 4.16). Notably, 89% of the WiFi APs connected to the geographically nearest EdgeCO.

Match CO Identifiers to Locations.

Synthesizing the hazmat records with DNS names can reveal even richer CO topology information. Some access networks include street or neighborhood names as CO identifiers in the DNS hostnames associated with access network router IP address. This allows an attacker to match CO locations in hazmat records to the IP addresses that traceroute reveals.

We matched the CO identifiers in a South Florida access network to the street and city names in public hazmat records (Figure 4.17). We validated the mappings with network operators, who asked us to anonymize the network for operational security reasons. This synthesis of physical and topological access network maps reveals the AggCO locations and the interconnections between the AggCO and EdgeCO locations. For example, the map indicates that an attack against AggCOs in Stuart and Pompano Beach could cause widespread outages

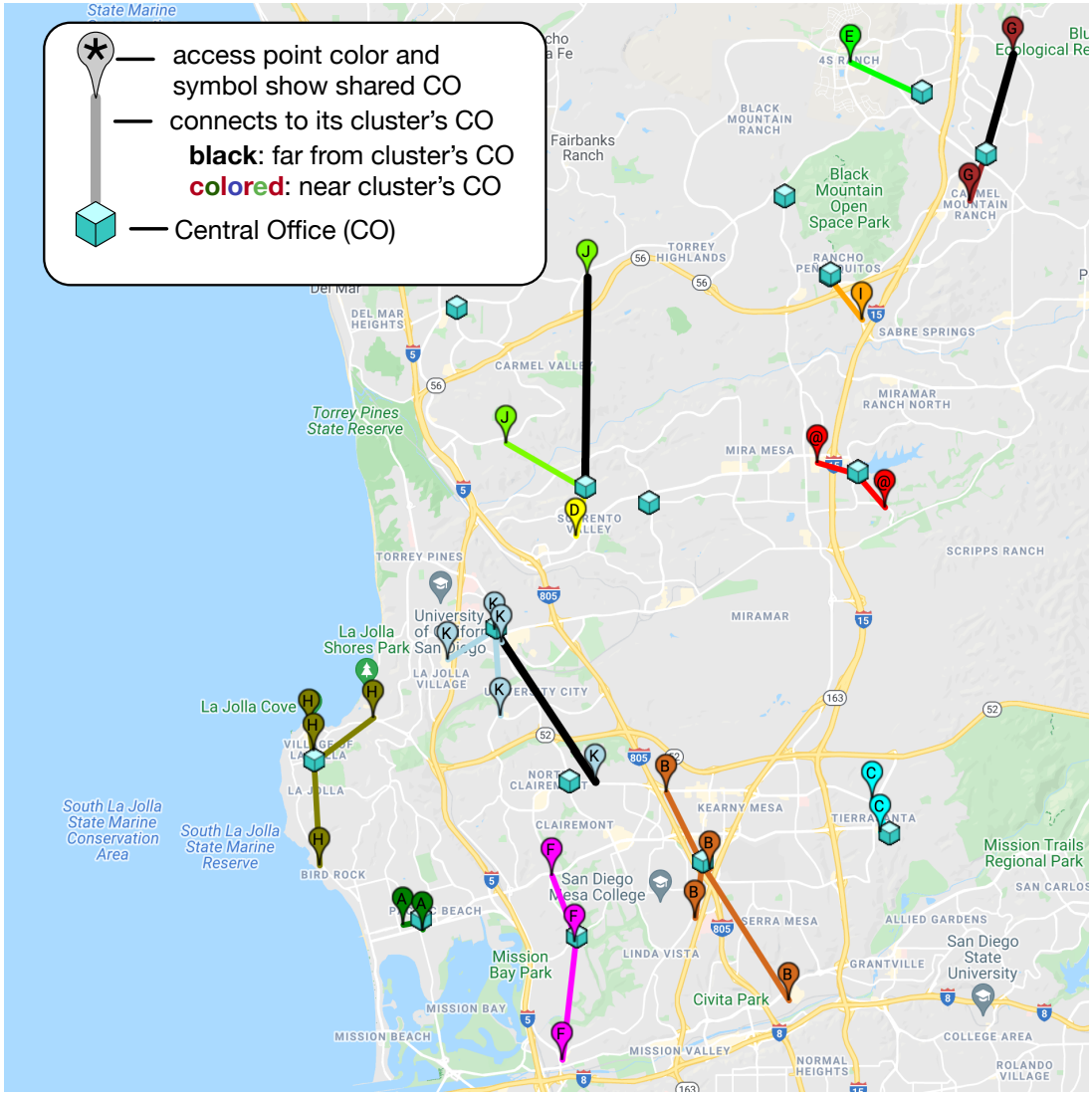


Figure 4.16. Inferred EdgeCOs for access points (APs) in a San Diego ISP. Marker color identifies APs connected to the same EdgeCO. Black lines indicate that the EdgeCO is not the closest CO.

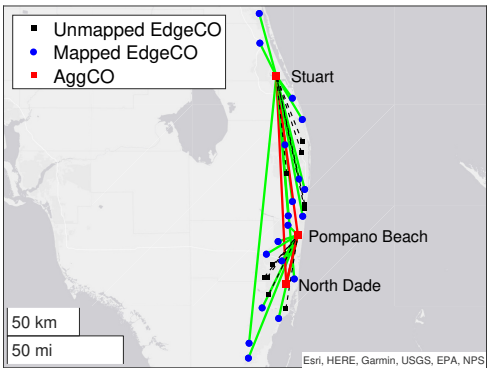


Figure 4.17. Map that combines hazmat records, DNS router names, and network topology measurements. Green lines map EdgeCOs (blue) to their corresponding AggCOs (red). Black squares are EdgeCOs that we could not map to DNS names.

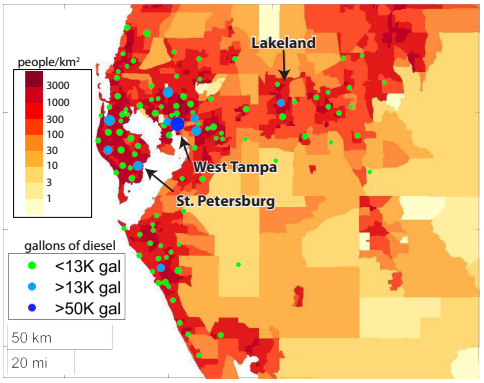


Figure 4.18. COs with large-capacity backup tanks in a Florida access network are located in highly populated areas.

extending to Palm Beach and Miami.

For access networks without useful CO identifiers in their DNS names, an attacker could also use the amount of registered fuel in the facilities to infer the aggregation level of the proximate CO. Compared to EdgeCOs, AggCOs often house equipment with greater power consumption that require more backup fuel. Figure 4.18 shows the locations and sizes of backup diesel tanks at COs in a West Florida access network overlaid on top of a population heat map. One facility in the West Tampa neighborhood (dark blue circle) stands out due to its exceptionally large tank size and the number of potential customers nearby.

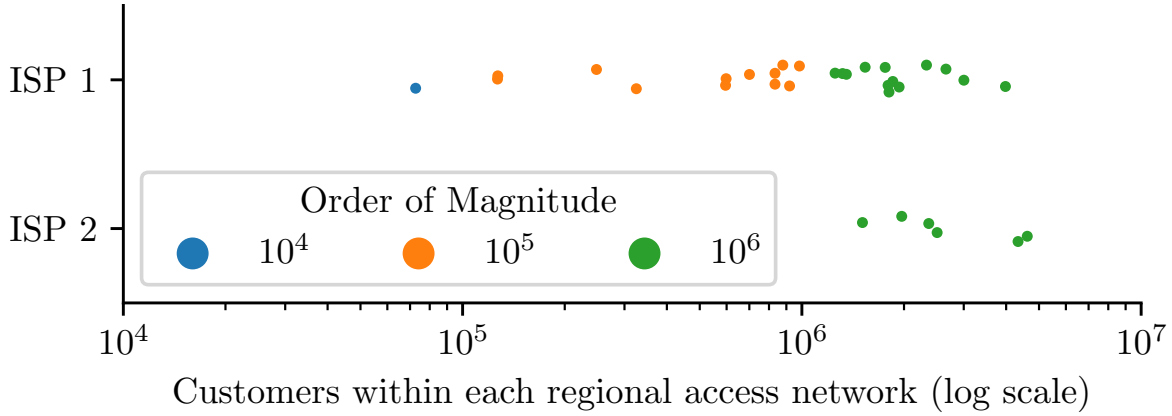


Figure 4.19. Causing both entry AggCOs to fail would disconnect over a million people in 59% of the regional access networks we study.

4.7 Assessing Outage Potential

After reviewing actual outages, we examine outage potential from intentional attacks based on the access network maps and customers connected to each CO.

Customers in Each Regional Network.

First, we examine the potential fallout from an attack that disconnects an entire access network, i.e., the entry AggCOs for the network. Nearly all regional access networks in our maps rely exclusively on two entry AggCOs to bridge customers to the Internet, and we can often precisely locate them remotely. If an attacker disables both entry AggCOs, it would disconnect more than 100K customers in all but a single region, and disconnect over 1M wireline customers in 59% of the regions (Figure 4.19). The Nashville outage also showed that entire access network outages can disconnect wireless customers that rely on the access network to reach the mobile packet core. The potential to disconnect millions of people, as well as other services that rely on the access networks, makes the regional network itself a compelling target for attack.

Customers Connected to EdgeCOs.

EdgeCOs present a softer target for intentional attack than AggCOs; operators indicated they are typically less fortified and might not have continual staff presence. According to

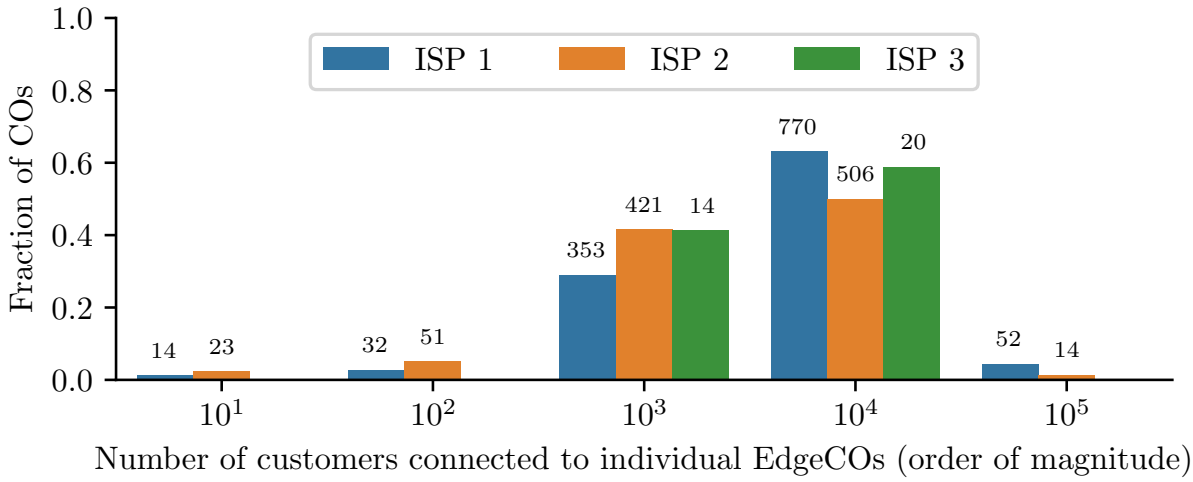


Figure 4.20. An EdgeCO outage would disconnect thousands to tens-of-thousands of customers for the EdgeCOs in our maps.

Table 4.3. Percentage of customers that ultimately rely on a single AggCO or backbone PoP. These customers are especially susceptible to natural and intentional disconnections.

	ISP 1	ISP 2	ISP 3
Single AggCO	11.3%	31.3%	100%

our inferred maps, an attack that disables an EdgeCO would disconnect thousands or tens-of-thousands of customers for 92% of the EdgeCOs (Figure 4.20), with a median of 12.7K customers. While EdgeCO failures disconnect their wireline residential and business customers, operators told us that wireless customers would often remain connected through nearby cell towers.

Customers Reliant on Single AggCO.

Finally, we analyze the fraction of customers that rely on a single AggCO in each network (Table 4.3), as these customers are especially susceptible to natural outages or intentional attack. In ISP 3, all EdgeCOs connect to multiple AggCOs, but the two regions we investigated rely on a single backbone PoP, so all customers ultimately rely on a single facility. For ISPs 1 and 2, each region relies on multiple backbone PoPs. In ISP 1, some of the smaller regions rely on a single entry AggCO that connects to multiple backbone PoPs, and the customers in these regions

lack redundant paths. All of the ISP 2 regions have multiple entry AggCOs, but many subregions connect EdgeCOs to only one AggCO. This topology leaves 31.3% of the customers reliant on a single AggCO, a nearly 3× increase compared to ISP 1.

4.8 Mitigations and Trade-offs

Our case studies and evaluation of targeted attacks reveal that ISPs are often not prepared for physical attacks on their regional infrastructure. We discussed the threat of intentional physical attack against COs with network operators, who were generally surprised at the level of detail we could reveal. The operators agreed that the threats exist but were unsure how to mitigate them cost effectively. In this section, we review potential mitigations that we discussed with access network operators, along with their perceived drawbacks, to inform future efforts to better secure these critical networks.

Operators consider the possibility of targeted attacks but face inherent tensions between the goals of decreasing the cost and complexity of network deployment, operation, repair, and defending against attacks. Our discussions revealed that the primary concerns for network operators are the cost and complexity of proposed mitigations, as well as retaining their ability to recover from common failure modes. Proposed mitigations that do not account for these concerns are unlikely to gain traction. Below we present the trade-offs operators identified in undertaking five potential mitigations to the attacks we consider.

Hide Locations of Central Offices.

The easiest way to cause widespread outages is to find a CO and disconnect either the power or fiber. There are two straightforward ways to precisely locate a CO: searching around a targeted area for the provider's signage on buildings, or search public databases to find records of buildings belonging to the targeted provider.

Providers can practice security through obscurity by hiding the location of COs. This is an inexpensive way to hide the infrastructure as the cost will primarily be labor to remove

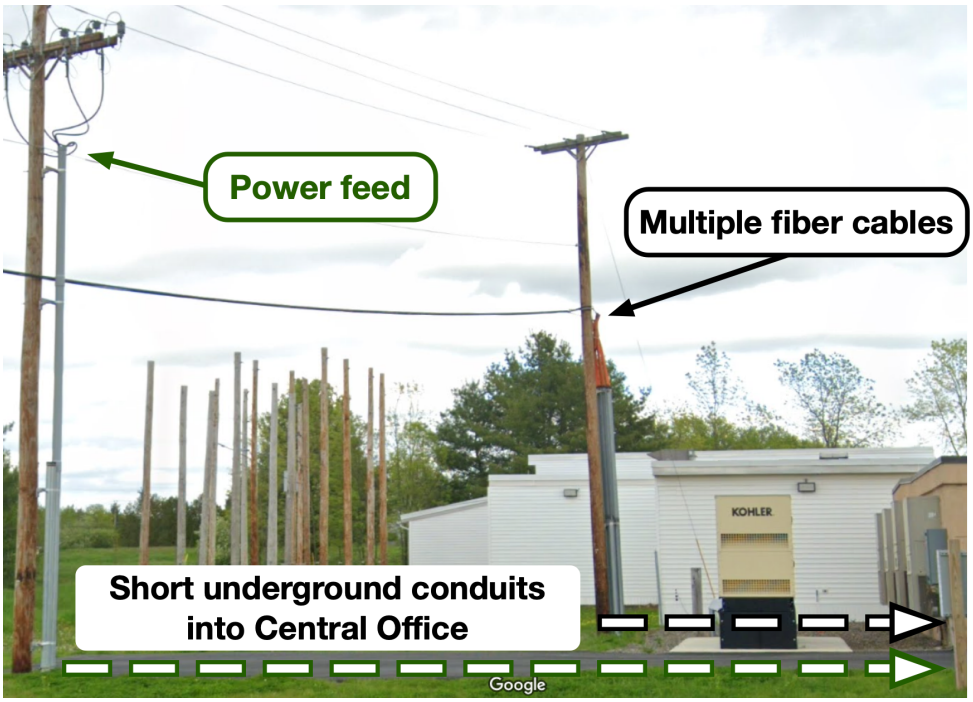


Figure 4.21. A CO's fiber and power are visible from the street.

signage from buildings. However, operators thought this could introduce many hidden costs. Operators told us they rely extensively on field technicians and contractors, and removing signs from CO buildings will make it harder for them to easily find the building in case of a problem. Operators also told us that COs are often unstaffed, so signage helps the public report problems to the ISP, such as when a building is on fire.

Similarly, providers can remove CO listings from public databases to prevent an attacker from remotely learning CO street addresses. However, the drawback is that public records of fuel-spill accidents are important for public health and environmental protection, leading governments to mandate them. There may be middle ground where some hazardous materials listings are obfuscated in public records so as to not reveal the purpose of the facility, or its owner. Costs would include paying administrative staff to both obfuscate and reveal the records when deemed necessary.

Hide CO Fiber/Power Lines.

Once an attacker finds a CO to target, it is easy to locate fiber and power serving that CO. Figure 4.21 shows an example of an EdgeCO that appears to have multiple fibers and power entering its premises on one pole just outside the CO entrance.

Burying the fiber and mains power into a CO—especially some distance away—could prevent an attacker from immediately finding the power and transport after locating a CO. Buried cables are also better protected from the elements than aerial cables. Unfortunately, operators told us that the costs of underground fiber ($25–75K per mile [46, 24]) and power (∼$500k per mile [45]) are significant. Additionally underground cables are considerably more expensive to repair.

ISPs also label their fiber infrastructure with tags, including underground cable runs, and removing labels from fiber runs would make it harder for attackers to identify fiber belonging to a particular ISP. Operators told us they label the fiber to prevent accidents and shorten repair time, so removing the labels would likely increase the number and duration of outages due to more common failure modes. This change also introduces the cost of removing labels on splice boxes placed at least every ∼1,000 ft along fiber runs [125].

Increase Last-mile Redundancy.

Some access networks do not include redundancy in their shared last-mile links, so a single fiber cut can take thousands of customers offline. Providers could add last-mile redundancy by adding a redundant connection back to the EdgeCO using a ring topology. The primary cost would be the extra network interfaces in the EdgeCOs(∼$24K per 20K customers for CMTS [38]) and redeploying last-mile fiber in a ring. ISPs could further improve redundancy by connecting customers to two EdgeCOs rather than one. Operators told us that some business customers pay to connect to multiple EdgeCOs, but that doing so for all customers is cost prohibitive.

Another approach is adding backup cellular connectivity to customer premises equipment. Costs include modem equipment and service plans. However, ISPs need to ensure the cellular backup link fails independently. This requires ISPs to provide more transparency about how

their wireline access network is used for backhaul in mobile networks.

Make Access Networks Passive.

COs depend on both power and fiber for connectivity. Removing the dependency on power would make networks more resilient, and remove an attack vector. Passive network equipment (e.g., optical splitters) are already used in the last mile. It may also be feasible to replace powered routers and CMTSes in EdgeCOs with entirely passive components driven by an AggCO. This technology has not yet been developed, and likely requires longer-term research to develop new passive network technologies. This solution would also incur the cost of upgrading network infrastructure across EdgeCOs.

Hide Access Networks in Measurements.

As we demonstrate, an attacker could learn physical topology from wardriving while performing ICMP traceroutes. It is possible to randomize IP address assignment within a given region making it more difficult for an attacker to geolocate infrastructure and users, but operators told us that doing so adds significant network management complexity. ISPs could also disable ICMP responses from their router infrastructure and remove reverse DNS, an inexpensive mitigation. However, this has a key drawback: operators told us that they and their customers rely on traceroute and reverse DNS to troubleshoot and diagnose problems. Also, we demonstrate that it remains possible to find COs with other methods (Section 4.6).

4.9 Related Work

The Internet is designed to be able to route around failures [41], yet large-scale failures are known to occur [137, 140, 9, 132]. Diverse factors cause failures including human error [108, 82], natural phenomena such as earthquakes [48], weather [132], solar activity [84], and equipment failure [185]. Our study focuses on vulnerabilities in access networks, since failures in these networks are challenging to route around.

Attempts to map topological diversity and understand physical network infrastructure vulnerabilities typically focused on backbone networks [166, 177, 89, 54, 65, 110] and submarine

cable networks [33, 207]. Analytic and probabilistic models were proposed to estimate the risk and survivability of physical attacks [10, 123, 68] and natural disasters [184, 50, 130]. Our study focuses on the topological diversity of regional access networks; we localized failures to specific EdgeCOs and AggCOs to inform a risk assessment of access network deployments.

This work builds on prior investigations into cyber attacks on related critical infrastructure: the electric grid. Internet access relies on power, and these prior threat assessments reveal how an attacker can force access networks to rely on backup power sources. Researchers found vulnerabilities in SCADA systems that manage electricity networks [178, 167, 20, 114], and real-world attacks that caused electricity outages for hundreds of thousands of endpoints [34, 37]. They also examined how an attacker can coordinate demand attacks over the Internet to cause cascading power grid failures [19, 12, 76]. Since these attacks require Internet connectivity to execute, this work provides some insight into how the power redundancy built into access networks may make it possible for an attacker to continue performing an attack even as it causes parts of the access network to lose power.

4.10 Conclusions

Although successful attacks on access networks require sophistication and planning, their impact on modern society—disconnecting critical infrastructure and economic activity—suggests that motivation for such attacks will increase. Given the increase in interdependence with other critical services, we believe our approach to considering resilience of this infrastructure must evolve. As with other critical ecosystems [36, 76], it would be better not to wait for high-profile attacks before undertaking this effort.

Our empirical approach combined new techniques for analyzing access network infrastructure deployments with measurements of weather-induced and accidental large-scale outages to quantify the potential cascading impact of targeted attacks. We discovered new insights into the physical attack surfaces and resiliency limit of regional access network infrastructure. We

also analyzed approaches to mitigating risks we identified, and associated tradeoffs in terms of cost and management complexity. Our results can inform risk assessments and reconsideration of approaches to safeguard this critical infrastructure on which our lives now depend.

4.11 Acknowledgement

Chapter 4, in part, is a reprint of the materials as it apears in *USENIX Security Symposium 2023*. Alexander Marder, Zesen Zhang, Ricky Mok, Ramakrishna Padmanabhan, Bradley Huffaker, Matthew Luckie, Alberto Dainotti, KC Claffy, Alex C. Snoeren, Aaron Schulman. The dissertation author was the co-primary investigator and author of this paper.

Chapter 5

How is LTE Base Stations' Scheduling Strategies Different Across Different Vendors

5.1 Introduction

The 3GPP standards [56] provide flexibility for base station vendors to provide their own proprietary downlink scheduling policies. For example, although the 3GPP suggests to use a certain data rate in a particular channel condition, it does not specify the exact relationship between these two variables [56]. The 3GPP also only recommends base station scheduler designers to balance between user demands and their equipment's capabilities, while simultaneously maintaining fairness. It does not define a strategy for how to achieve that, as with other mobile protocols [6]. Therefore, each of the four primary base station vendors in the world (i.e., Ericsson, Nokia, Samsung, and Huawei), can implement their own proprietary scheduling policies. These policies are the primary "secret sauce" that vendors can use to differentiate their base station's performance from other vendors.

Since all LTE base stations are 3GPP standards compliant, and compounded by the lack of visibility into what vendor base station a device is using, there has been an assumption that there homogeneous behavior across different vendors' schedulers. However, there are contradictory observations about scheduling behavior. Several prior studies found bursty patterns in how LTE

base stations allocated resources across time slots, [17, 18, 210], while other work found LTE base stations equally shared resources across UEs in every time slot [205, 206]. This apparent contradiction means cellular performance monitoring studies and tools may need to be evaluated on and tailored to a variety of base station scheduling policies.

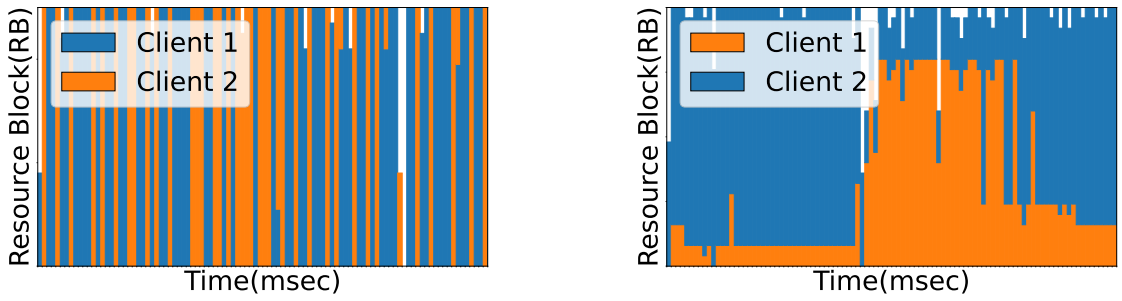
In this paper, we provide the first preliminary evidence of clear, but somewhat consistent, differences in base station scheduling policy across four of the top base station vendors: Ericsson, Samsung, Huawei and Nokia.

Unfortunately, it is challenging to determine what differences exist between downlink scheduling policies because: (1) Cellular modems have no built-in method to determine the vendor of a base station. (2) Observing the scheduling policy of a base station requires a controlled environment, namely an idle base station with two controlled users contending for resources. (3) Base stations may behave differently across multiple configurations, and deployment locations, but operators tend to deploy a particular vendor's equipment homogeneously within a particular region.

We determined base station vendor through a combination of eyeballing the logo on the base station enclosure, as well as using a base station vendor dataset from Revelare Networks [139]. We achieved a semi-controlled setting by performing all experiments overnight in non-residential areas and validated that the base stations were idle by checking if one use could use the entire resources of the base station. To evaluate a variety of vendors, and validate their consistency, we collected data from 20 different base stations, with different configurations, in three cities, across two countries, and across the deployments of the four major mobile carriers. We compared behavior of schedulers from all four of the most popular vendors, with two competing users, in a variety of conditions including differing buffer status, channel quality, and traffic sources (i.e., application and transport protocol). Our contributions are as follows:

1. We found differing, but consistent, *radio resource allocation policies* for competing users across base station vendors, providers, and channel conditions.

2. We found differing *rate adaptation* algorithms in use. We found that some base station



(a) All in TTI

(b) Split TTI

Figure 5.1. Two different possible base station schedulers.

vendors aggressively pick data rates, while others opt for a linear data rate mapping to channel quality. The aggressive strategy resulted in higher end-to-end throughput in high quality channel conditions.

3. We found that when competing users have different channel qualities, the different vendors in different deployments prioritize users differently. Some allocate resources unequally, while others evenly shared radio resources.

5.2 Background and Related Work

In this paper, we focused on evaluating LTE base station behavior as it is still the dominate standards in the world providing service. We put 5G results in the discussion section to open era for future work. This section introduces two major processes not directly defined by the 3GPP, which leaves their implementation up to the vendor, having a direct impact on network performance.

5.2.1 Radio resource scheduling

The scheduler is the process through which the base station distributes available data Physical Resource Blocks (PRBs) across the UEs. The base station scheduler manages the radio resource allocation for both, uplink and downlink directions. However, given the dominance of downlink traffic, we focus on downlink scheduling.

During our experiment, we found single UE will get all base station resources when it is the only available UE downloading traffic. Therefore, in this paper, we focused on discussing base station scheduling strategies in competing UE scenario. The scheduler takes a scheduling decision every TTI (1 ms). This means that the available PRBs of 1 TTI are distributed across the different UEs every millisecond.

In general, scheduling decisions are taken every Transmission Time Interval or TTI taking into account multiple sources of information from UEs (KPIs) as well as radio measurements taken by the base station combined with historic allocation data. Some of the most relevant KPIs generally used across schedulers are the number of UEs, UEs' CQI reports, buffer status reports and QoS rank [7, 176].

The literature defines several theoretical scheduling algorithms including Round Robin, Maximum CQI and Proportional Fair[149]. However, most commercial base stations implement a custom variant of proportional fair using the aforementioned KPIs among others, offering a balance between throughput and fairness by giving priority to the UEs that meet their custom criteria (e.g. haven't received resources in a while or they have more traffic to be sent) [21, 18].

By summarizing previous research efforts, we found design protocols based on a certain scheduling assumption observed in commercial base stations[17, 206, 18]. For example, some authors [17, 210] suggest that the base station scheduler assigns all the PRBs to one user in one TTI effectively generating a traffic burst. Base station scheduling schema in Figure 5.1a (the x-axis represents time in milliseconds and y-axis is the number of PRBs allocated) depicts that behavior. It shows how in each TTI, the base station tends to assign resources to only one UE. However, other trends in the literature [205, 206] observed a different base station's resource allocation behavior. As shown in Figure 5.1b, the base station tends to share the resources of one TTI between the two competing UEs.

This difference in the scheduling behavior limits the potential benefits offered by different optimizations proposed in the literature. For example, BurstTracker[17] will trigger false positive detection in the case shown in Fig 5.1b.

5.2.2 Link Adaptation policy

Link Adaptation is the process where the base station dynamically adapts the Modulation and Coding Scheme (MCS) to codify more or less information in each PRB depending on the channel quality with the objective of maximizing data rates and minimize loses. How to select the MCS according to the channel quality is decided by vendors and the 3GPP only provides recommendations. They generally rely on a combination of Channel Quality Indicator (CQI) periodically sent by the UE with other radio KPIs measured from the base station. In general, higher CQI values enable the use of more advanced modulation schemes, such as 64-QAM, whereas lower CQI values will trigger more robust coding and simpler modulation techniques, like QPSK.

As suggested in WiFi, rate control algorithm can greatly affect network throughput and power [133, 77, 71, 119]. How base station's rate link adaption work has been used in improving network performance has been discussed in previous research [98, 105, 13], but they were either based on simulation or based on single vendor. None of them notice the CQI related MCS selection strategy can be different across different vendors.

5.3 Methodology and Instrumentation

This section introduces the experimental setup and instrumentation utilized for examining scheduling algorithms of base stations. We demonstrate the efficacy of the methodology for data acquisition and its analysis on idle base stations.

5.3.1 Experimental Setup

Our goal is to delineate the differences between base station downlink schedulers implemented by various vendors.

All our scheduler measurements have been performed on identified idle base stations where we download traffic from a controlled server to our UEs. Figure 5.2 illustrates our

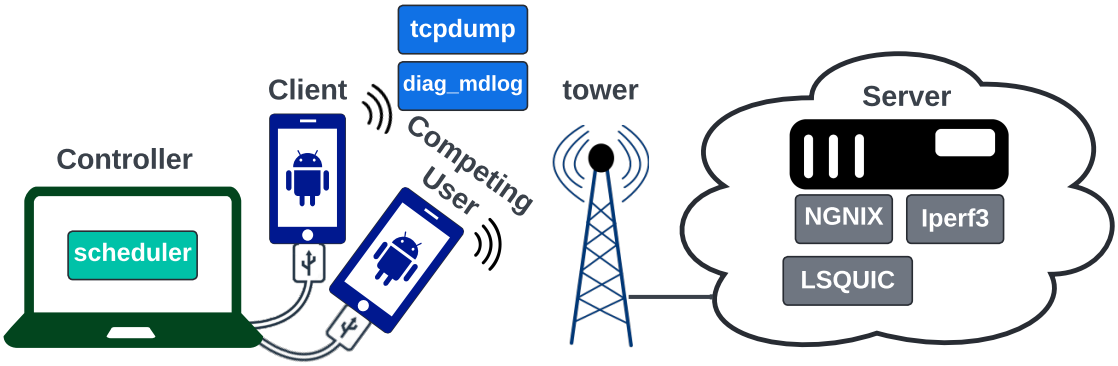


Figure 5.2. Experimental setup for data collection.

experimental setup, which is segmented into three distinct components.

Server Configuration. A server operating on Ubuntu 22.04 LTS with the Linux kernel 5.15 LTS was deployed within the same geographic region as the base stations to ensure a round-trip time (RTT) within 40 milliseconds, thereby reducing the impact of end-to-end latency. The server was configured with substantial egress bandwidth to prevent it from bottlenecking. Network traffic was generated using `iperf3 version 3.9` for TCP and UDP, `LSQUIC version 4.0.8` for QUIC, and `NGINX version 1.26.1` for HTTPS. We use BBR congestion control algorithm for both TCP and QUIC in our test.

Client Configuration. The client-side is comprised of two mobile phones (Oneplus Nord N30) operating on Android 14. Two phones competing for downlink resources by generating traffic using `iperf3`, `lsquic` and `curl`. The connection between the phones and a laptop was maintained using USB solely for synchronization purposes during code execution. To capture the base station's behavior at the Physical (PHY) layer, the Qualcomm Modem Diagnostic Log (QMDL) files were recorded using `diag_mdlog`. And we decoded them with QXDM software[136, 188]. Additionally, `tcpdump` was employed to monitor and later analyze network traffic at the transport layer on the mobile devices.

Controller Functionality. The synchronization of the two phones was managed via a

Table 5.1. Details about the 20 diverse base stations we observed from 4 vendors.

Downlink Config.	Ericsson										Samsung						Huawei			Nokia
	Macro						Micro				Macro				Micro		Macro			Macro
Provider	A	A	T	T	T	A	A	A	A	A	VZ	VZ	VZ	VZ	VZ	VZ	VO	VO	VO	A
BW (MHz)	50	50	35	40	35	20	40	40	35	40	60	50	60	50	20	20	30	30	25	40
MIMO	2	2	4	4	4	4	2	2	2	2	2	4	4	4	4	4	2	2	2	4
CA	4	4	2	3	3	1	3	3	2	3	4	4	4	4	2	3	2	2	2	2

laptop by using thread barrier to guarantee a competitive scenario.

5.3.2 Base Station Scheduler analysis

To understand the base station's scheduling behavior, we need to look at how the base station allocates its bandwidth resources. It requires us to find an idle base station so we can monitor all downlink traffic from base station. Idle base station also means we can reconstruct base station behavior by aggregating the PHY layer behavior from all UEs.

To find idle base station, we choose base stations located in non-residential areas during nocturnal hours to ensure a controlled environment. To reduce the potential for base station idleness disrupted by passing users, we perform an initial five-second bursty UDP download test at the beginning of each experimental phase as metadata. We can find the cell will allocate their 98.8% RBs to our single UE once the base station is idle. We assume the rest 1.2% of empty RBs are mostly reserved RBs for control message like MIB or SIB as they are in subframe 0 and 5[158]. We also perform experiment multiple times(more than 3) to reduce the likelihood of noise from nearby users.

During the experiments, we ensured that both phones were registered through the same Cell ID by using NetMonster [175] before proceeding. We force UEs to use LTE only to obtain LTE scheduling behavior and allow 5G to get 5G results. To observe how vendors schedule heavy downlink traffic differently, we filled the buffer at the base station for both User Equipments (UEs) by having two phones requesting 4 parallel thread each with 1 Gbps of downlink UDP traffic from the server. By having high packet loss rate with no congestion control for UDP, we filled the buffer in the base station for each UE throughout the experiment. Additionally, we

utilized prepaid SIM cards ensuring consistent QoS and traffic. In this case, we controlled as much factors we can have to observe base station scheduling differences.

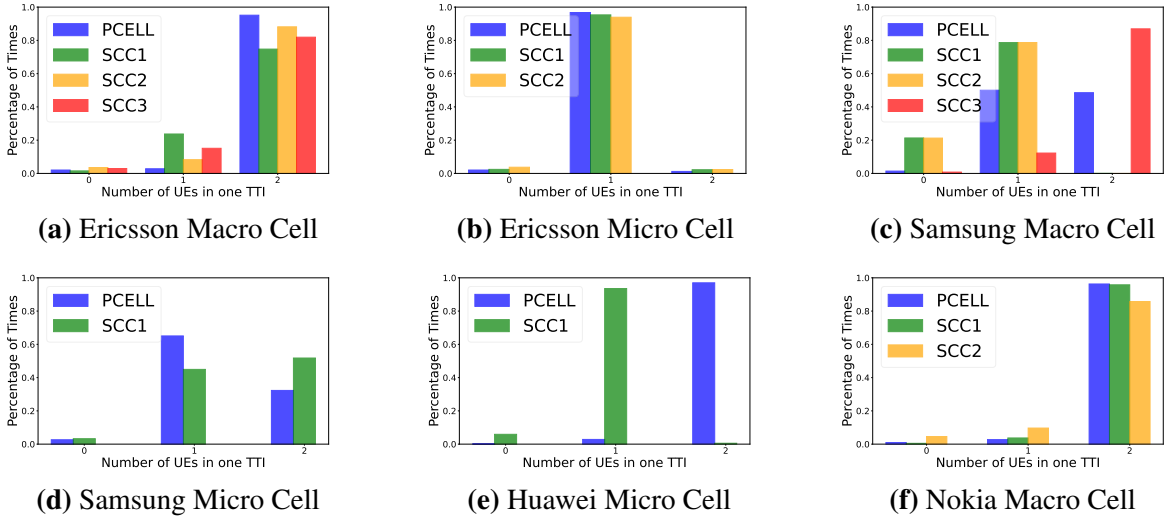
With high network traffic, carrier aggregation was triggered by the network. We believe they are carrier aggregations instead of dual connectivity by comparing through physical cell ID, RSSI and frame number synchronization. Hence, we perform analysis on all carriers as they are from the same eNB who shares the same scheduler.

5.3.3 Data Overview

We collected and analyzed downlink scheduler traces from 20 base stations across four of the most prominent cellular base station vendors: Ericsson, Samsung, Nokia, and Huawei (Table 5.1) used across four major carriers: AT&T, T-Mobile, Verizon Wireless and Vodafone. Micro Cells, 2 Samsung Micro Cells, 3 Huawei Macro Cells and 1 Nokia Macro Cells. Operators tend to homogeneously deploy a vendor's equipment in each deployment region (e.g., city and state), so to obtain a diversity of vendors we needed to collect data across four cities in the U.S., three on the west coast and one on the east coast, and one in Spain. This set likely represents many of the widely used base stations across the globe, however it is by no means a complete set. In this preliminary study, we use this set of diverse base stations to provide evidence of the differences that can exist between downlink schedulers, with the intention of showing the need to consider these differences.

5.4 Results

We found many differences in downlink scheduler behavior across different vendors, types of base stations. Specifically, we found there were significant differences across four dimensions of base station scheduling behavior: (1) number of radio Resource Blocks (RB) allocated to competing UEs per scheduling interval (TTI). (2) link adaption algorithm (i.e., how CQI is assigned to MCS) (3) UEs that have diverse channel quality to the base station. We, provide a preliminary view of scheduling differences on 5G NR base stations.



(a) Ericsson Macro Cell **(b)** Ericsson Micro Cell **(c)** Samsung Macro Cell

(d) Samsung Micro Cell **(e)** Huawei Micro Cell **(f)** Nokia Macro Cell

Figure 5.3. Base station schedulers across different vendors as well as macro and micro cell configurations.

5.4.1 Radio resource allocation policy

Figure 5.3 shows how often more than one of the two competing UEs are scheduled in one TTI. The resources scheduled to competing users varied across all vendors Ericsson, Samsung, Huawei and Nokia, and even varied within each vendor's macro and micro cells. However, the behavior was consistent within vendor and type of base station, regardless of carrier; therefore each plot in Figure 5.3 shows behavior of one representative base station of each vendor/type.

Ericsson and Nokia macro base stations (Figure 5.3a) from all three carriers show that the base station always allocates resources to both UEs in every TTI. Three aggregated cells are sending traffic to UEs simultaneously, including across one primary carrier and two secondary carriers simultaneously. Within each carrier, resources are distributed between the two competing phones consistently. Resources are allocated to UEs each TTI in varying proportions, and those ratios alternate every 5-6 TTIs. It is possible *Nokia*'s base stations are (Figure 5.3f) acting similar to Ericsson because they are used by the same providers in different regions [**?**]. In contrast, the Ericsson *micro* base stations (Figure 5.3b) generally allocate all the RBs to only one UE per TTI, resulting in a bursty resource distribution. However, across all of the aggregated carriers on the

base station, both UEs receive resources from at least one carrier, 80% of the time. This suggests that the Ericsson *micro* base stations divide up resources primarily by assigning them to different carriers.

Samsung macro base stations (Figure 5.3c) have a different resource allocation; 50% of the time assigning full RBs to one UE per TTI, while the other half the two UEs share resources in the same TTI in each carrier. Samsung's *micro* base station is a hybrid of Ericsson's macro and micro-cell behavior (Figure 5.3d). Frequently sharing resources among multiple UEs within the same TTI like the macro cell, but they occasionally allocate full RBs to one UE in each carrier like the micro cell.

Huawei micro base station (Figure 5.3e) behaves differently than Ericsson and Samsung mostly in how it uses carrier aggregation. The primary carrier always provides resources to both UEs at the same TTI, but the secondary carrier gives all its resources to one UE at one TTI. We observed its secondary carrier schedules its traffic alternatively to two UEs along each TTI which still ensures the fairness between UEs.

We also validated that our UDP probing method and found similar scheduling policies were displayed with iperf TCP files downloads, HTTPS, and QUIC protocols (Section 5.5).

5.4.2 Link adaption policy

Figure 5.4 shows the distribution we observed of MCS rate control across various CQI levels (i.e., link adaption). *Ericsson* is generally agressive, assigning high MCS levels for high CQI. *Samsung* and *Huawei* exhibit a more cautious approach, with MCS allocation showing a linear decline in median MCS as CQI decreases, but *Huawei* generally assigns higher than *Samsung*. *Nokia* is as aggressive as *Ericsson* with high CQI, but tends to be more conservative assigning CQI is low.

These observed differences distinct vendor-specific strategies. Ericsson appears to prioritize speed and throughput under favorable signal conditions, potentially improving throughput in scenarios with high CQI. While Samsung and Huawei's strategy seems to focus on maintaining

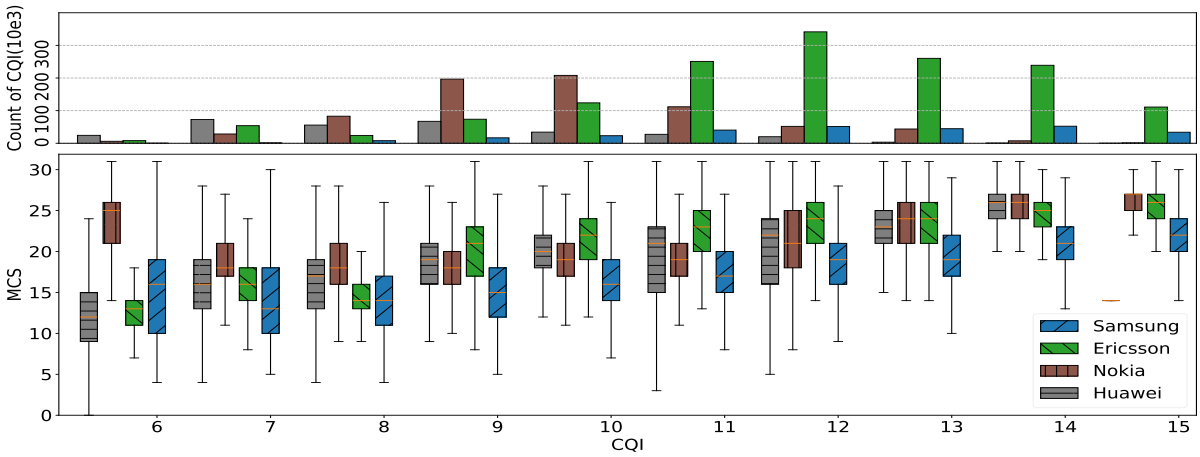


Figure 5.4. Box and histogram plot showing MCS allocation by vendors against different CQI. The box plot showed the range of MCS to each CQI. The histogram showed the number of data for each CQI.

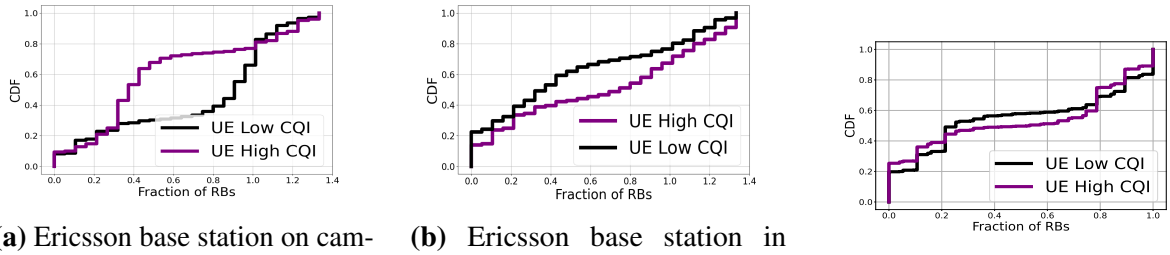
a balance, possibly to mitigate rapid throughput degradation under fluctuating signal strengths. Unfortunately we did not observe many low CQI scenarios (CQI 0–6), possibly because these would lead to a cell handoff if there is a better nearby cell.

These results indicate that when a UE has high CQI, they will get higher end-to-end throughput from an Ericsson base station than from a Samsung base station. We tested this by bringing the same UE to the location close to Samsung and Ericsson base station. We stay as close as possible to the base station and ensured UE's CQI is between 12-15 during the experiment (and during night when the base stations were idle). Using iperf UDP flood we measured the bandwidth-normalized throughput for Ericsson as 8.75 bits/sec/Hz and Samsung is 4.25 bits/sec/Hz (note both used 4x4 MIMO). This confirms that in similar radio conditions Ericsson base stations provider higher throughput. However, this performance gap is likely to narrow in the middle-range of CQI, as Ericsson's aggressive strategy may lead to more losses and retransmissions.

5.4.3 Policy for diverse channel quality

Next, we compare the resource allocation policy of macro base stations under varying network conditions. The focus was on observing how two vendors base station allocated

resources to User Equipments (UEs) competing for one base station's radio resources when one UE has a much higher CQI than the other (e.g., one is closer to the base station). We only study *Ericsson* and *Samsung* for this experiment to see if the divergent behavior they demonstrated in prior experiments also applied to this scenario.



(a) Ericsson base station on campus.

(b) Ericsson base station in crowd area.

(c) Samsung base station.

Figure 5.5. CDF of RB usage for UEs with different CQIs contending for the same base station's downlink resources

Figure 5.5 shows how different vendors' base stations allocate resource differently to UEs under different CQI. We compare the fraction of RBs allocated to the UE in each TTI divided by the total number of RBs the base station has (i.e., it's bandwidth). We discovered distinct resource allocation policies between vendors:

Ericsson: Ericsson presented a resource distribution based on signal quality differences between UEs. For UEs in Figure 5.5a, the CQI for the UE proximal to the base station is recorded at average 11.7, whereas the distant UE's average CQI is 6.9, with a higher allocation of resources directed towards the UE with superior connectivity. And we observed that the base station provides more resource to the UE which is away from it. In the meanwhile, for UEs in Fig 5.5b, the base station provides more resource to the UE which is under better channel quality condition. This could be because the network operator tuned their base station's behavior when they deploy their base station to meet the requirement of deployed location. The CQI for the closer UE to base station is with average 12.37 while the CQI which is further away from base station is 8.69.

Samsung: In contrast, Samsung's strategy for resource allocation appears more balanced

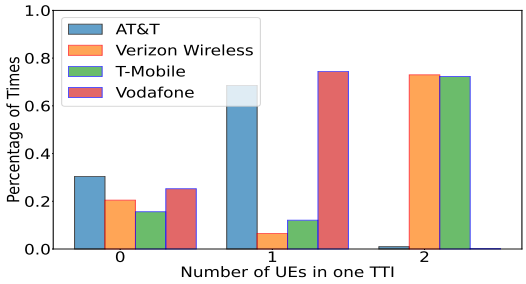


Figure 5.6. UEs per TTI for 5G base station schedulers

between two UEs with differing network qualities. This demonstrated a near-equal distribution of resource blocks throughout the experiment, as depicted in Figure 5.5c. For the UE near the base station, the average CQI was noted as 14.7, and for the more remote UE, it was 7.19.

These divergent strategies between Ericsson and Samsung base stations reflect varying vendor approaches to resource management. We will further collect data from other base station vendors in the future work.

5.5 Discussion

5.5.1 Scheduling policy diversity in 5G

This work primarily focuses on on 4G LTE downlink scheduler differences, as continues to be widely deployed. However, do these scheduling policy differences exist in 5G? We compared the radio resource scheduling on 5G base stations from Ericsson, Samsung, and Huawei. We tested on AT&T, Verizon, and Vodafone that use 5G NSA, and T-Mobile wich uses 5G SA. AT&T and T-Mobile use Ericsson, Verizon was Samsung, and Vodafone uses Huawei.

Figure 5.6 displays the frequency of the number of UEs appearing per TTI. AT&T, Verizon and Vodafone transmits packets through both 5G NR and LTE radios simultaneously as they use NSA core network. AT&T and Vodafone schedule one UE via the 5G radio and another via the LTE radio, making it look like 5G radio schedules only one UE per TTI. However, synchronization between 5G and LTE transmissions is challenging due to 5G using Time Division

Duplexing (TDD) and LTE using Frequency Division Duplexing (FDD) in the region we collected data from. Another problem is 5G packet transmission is sensitive to CQI as its radio frequency is high where it could be easily blocked by noise. Conversely, T-Mobile, operating on a 5G SA network, solely transmits packets through 5G NR. Its base station tends to share resources between UEs within the same TTI.

5.5.2 Policy effects on congestion control

Besides UDP test, we also conducted tests using iperf TCP and files downloads from a server via HTTPS and QUIC protocols with curl and lsquic to generate network traffic. The similar scheduling patterns were observed as when downloading packets with UDP in each Cell.

We analyzed throughput fairness between UEs by comparing 82 experiments conducted with two UEs under the same CQI. 85% time with less than 20% throughput difference between UEs indicates fairness is managed properly among all vendors despite differences in their scheduling processes.

We applied our findings into literature and found they lack in generalization of the methods they provide. For instance, BurstTracker [17] assumes that a base station schedules one UE in each TTI when managing traffic for contention users. Our research demonstrates BurstTracker's assumption is primarily used by Ericsson Micro base stations. Applying the same strategy to Ericsson Macro base stations or Samsung, Nokia base stations results in false positive detections of burst ends. On the other hand, PBE-CC [206] assumes that resources are shared equally among multiple UEs within the same subframe. Their model fits well with Ericsson Macro and Nokia base stations, where resources are shared among UEs for each carrier. However, Ericsson Micro base stations and occasionally Samsung base stations may allocate all RBs in a carrier to a single UE, which limits the effectiveness of PBE-CC when implemented with these vendors.

5.6 Conclusion

This study gave a first attempt analysis of the downlink scheduling algorithms employed by base stations from four major vendors, Ericsson, Samsung, Huawei and Nokia, across three cities. Our experiments highlighted the difference of vendor-specific strategies in resource allocation. Our findings also reveal significant differences in how these vendors allocate resources in LTE networks, particularly in competitive user scenarios. Ericsson's base stations tended to favor high MCS levels to users with better channel conditions, optimizing throughput in scenarios where signal quality was favorable. Conversely, Samsung's approach was more conservative, maintaining a balance across varying signal conditions, which might help in stabilizing user experience during fluctuating network quality. The detailed insights into the scheduling behavior and MCS allocation strategies provided by our analysis aim to enhance future research on cellular network performance improvements with considering vendor differences.

5.7 Acknowledgement

Chapter 5, is currently being prepared for submission for publication of material. Zesen Zhang, Jon Larrea, Jarrett Huddleston, Haoran Wan, Ricky Mok, Bradley Huffaker, KC Claffy, Kyle Jamieson, Alexander Marder, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

Chapter 6

MobileSDR: A Mobile Programmable Platform for Wireless Field Tests and Diagnostics

6.1 Introduction

Deploying and maintaining wireless networks is increasingly challenging due to the increasing rate of protocol development and scale of deployments. Diagnosing failures is often done by manually deploying specialized equipment; for instance, cellular providers send trucks with spectrum analyzers to diagnose if interference from unauthorized transmitters is causing poor performance [26]. Also, it is difficult to improve wireless networks that are built using unlicensed, or shared, spectrum as commodity devices do not offer visibility into the cause of poor performance, namely is it due to interference, high usage, incorrect settings (e.g., antenna gain), or even hardware failures [116].

This lack of visibility into deployments of wireless networks stems from the following key practical problems: (1) Field test equipment is specialized to specific protocols, so deployments of existing field test equipment can not be used for new protocols, leaving bleeding-edge protocols always behind in terms of the available field test equipment to find the problematic behavior. (2) Field test equipment is generally designed to operate locally on an attached screen, there is a so the cost of the equipment is extremely high making scale of field test deployments limited

to technicians that need it for in-person diagnosis. For instance, cellular tower installers often test newly deployed towers with an expensive real-time field test device the Agilent FieldFox which is on the order of $10,000–50,000 USD depending on options), and (3) Programmable test equipment (e.g., SDRs) is currently limited to stationary deployments, or deployments where an expert in signal processing software and hardware expert is physically present to manage the SDR. These result in extremely limited geographic coverage.

As a result of these practical issues, wireless field testing has yet to achieve large-scale wide-area coverage that is feasible from other crowd sourced measurement systems. Smartphones have the advantage of being an ideal hardware setup for mobile measurement. They have a built-in battery and built-in Internet connectivity to upload results. Indeed, smartphone-based wireless network sensing platforms have achieved worldwide coverage. OpenCellID [187]'s cell tower mapping has incorporated data from 49,000 sensors, and Wigle.net has an estimated half a million sensors. However, these platforms are extremely limited in what they can sense, mostly only capturing information about base stations from the wireless network scanning capability that is built-in to all smartphones.

Our goal is to build a mobile SDR platform that can achieve the same general field test capability as existing field test equipment with the wide-area coverage of crowdsourced wireless measurement platforms. Prior work demonstrated that smartphones have the I/O and CPU performance to be capable of being directly hosting Software Defined Radios [23, 94, 134, 161, 109]. However, this work only showed that there is sufficient I/O bandwidth and CPU performance to perform many typical diagnostic tasks from I/Q capture, to decoding WiFi packets [23]. However, no platform has yet demonstrated a smartphone can be used as a crowdsourced mobile SDR.

We present MobileSDR, the first platform to demonstrate the feasibility of a platform for crowdsourced mobile SDR-based field tests that can be remotely programmed to do custom, geo-tagged, signal measurements. These upload the results for remote users to access over the cloud. The principle that we apply to make this vision practical is to separate the operator of

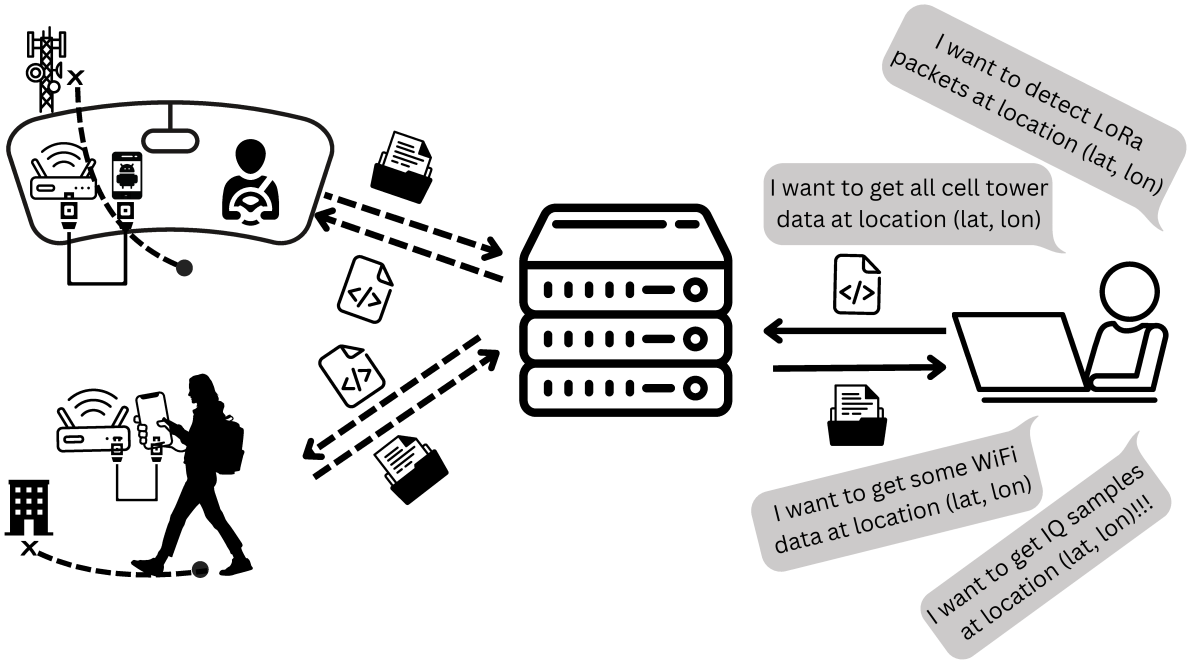


Figure 6.1. MobileSDR usage overview. Users can deploy signal processing programs on remote mobile SDRs in locations that they specify. Once MobileSDR collects data, it will backhual collected samples through wireless connection to broker to do further analysis on cloud and researchers can download those samples.

the crowdsourced mobile SDR from the user and programmer of the signal measurement and analysis coming from the SDR. To the best of our knowledge, this is the first demonstration that this separation is practically feasible on commodity low-cost smartphones and SDR platforms.

While there have been previous attempts at building platforms for remote SDR operation [62, 217, 72, 194], none have solved the following technical challenges that are needed to make a remotely-controlled programmable mobile wireless measurement platform:

First, the limited CPU cores in Android devices limit their ability to be used for wireless measurement. Our empirical study shows a smartphone struggles to channelize and decode four channels within a 4MHz band simultaneously, leading to overflow and decoding failures. Additionally, phone backhaul data bandwidth (e.g., WiFi and LTE) is is very limited so MobileSDR must filter the signal significantly before sending the samples to the remote user. Second, providing programmable signal processing capability on crowdsourced mobile devices could create

a new attack vector compromising the device owner's security. The current Software Defined Radio applications on Android are limited to using insecure pre-compiled native libraries that could be used by an attacker to gain control over the device, or leak sensitive information. Third, battery capacity on smartphones is extremely limited. Adding an SDR powered by the phone will reduce battery life energy to a phone, so a mobile SDR platform needs to operate as as efficiently possible.

Our work overcomes these challenges by making the following key contributions:

1. To overcome the limited CPU we demonstrate the feasibility of using a combination of existing approaches to packet detection using energy detection in an FPGA, and preamble detection on the phone's CPU to reduce the samples that need to be processed by the phone's CPU and uploaded to the remote user for further analysis.

2. To make mobile SDRs safe to program, we show that we can give users the ability to deploy GNURadio flowgraphs using the Domain-Specific Language (DSL) of GNURadio companion's xml/yaml flowgraphs. Achieving this required demonstrating that we can port the entire Python-based flowgraph processor of GNURadio into a Java-based Android app. We also show that this DSL can still achive the high-performance of pre-compiling all native C++ GNURadio blocks in a flowgraph and deploying them together as was done in prior work on smartphone-based SDRs [23].

3. We demonstrate that commodity SDR hardware platforms provide several opportunities to dynamically reduce power consumption to save energy between experiments. In particular, we show that several components of the SDR can be duty cycled (FPGA, oscillators, amplifiers). We found that there are tradeoffs in turning on/off these devices based on how much power is saved compared to the cost of turning a component back on once it has been powered off.

4. We evaluate the MobileSDR platform by conducting real-world case studies of its use for LoRa performance diagnosis and LTE base station information collection. For LoRa, We designed a close to COTS LoRa signal analysis GNU Radio block which can save IQ samples once it detects

LoRa packets. We observed LoRa signal will not only be interfered by other LoRa signals but also interfered by the presence of strong signals within the ISM bands, which prevent the LoRa signal from being decoded. Our findings demonstrate that with our mobileSDR, it is possible to move around buildings to identify areas with lower interference levels and successfully decode LoRa packets. For cellular, we showed the ability of collecting base station information from multiple frequency bands in a location to collect base station information and diagnose potential potential interference of base station signal.

6.2 Motivation

Measuring wireless networks in the field is crucial for optimizing infrastructure performance, ensuring connection reliability, and reducing costs. Field measurements provide empirical data that enhances the wireless network design and implementation, especially in complex environments.

Enhance Accuracy of Network Design: Field measurements validate and refine empirical models of signal propagation, which often fail in multi-floor or multi-building scenarios. Accurate path loss measurements can reduce the number of sensors needed by over 50%, leading to substantial cost savings[101]. Understanding actual conditions, such as access point (AP) density and channel usage, enables better load balancing and interference management, crucial in densely populated areas[214].

Provide Quality of Service (QoS) Assessment: Field test enables continuous monitoring of QoS parameters through both active and passive methods, which provides timely performance assessments, optimizing user experiences[11]. Field measurements can also pinpoint specific issues that affect network performance, such as connection setup times influenced by environmental factors[214].

Field measurements are indispensable for achieving reliable and efficient wireless network performance, particularly in dynamic environments. Some researchers argue that theoretical

models may suffice in less complex scenarios, suggesting that a balance between empirical and model-based approaches could be beneficial.

However, field measurements are resource-intensive. The lack of generic tools and expert human resources to collect data impedes researchers' ability to conduct field tests for their designs. This scenario has motivated us to develop a generic, portable tool with crowd-sourcing potential. This tool aims to offload the duty of data collection to the general public while still providing researchers with the flexibility to create programs tailored to their specific needs.

To do so, we choose leverage the popularity of smartphones to connect with SDR since SDR are typically more affordable than conventional spectrum analyzers, making them accessible for a wide range of applications, including research and public health assessments[218]. The ability to implement measurement functions through software minimizes the need for costly hardware upgrades[218].

Moreover, SDR is flexible enough for use in various applications, from monitoring soil nutrient sensors to assessing electric field strengths in urban settings[209][113]. Its ability to perform simultaneous measurements across multiple channels enhances the quality and reliability of data collected in diverse. technology enables advanced measurements that traditional methods cannot match, especially in dynamic environments.

6.3 System Model

The goal of MobileSDR is to provide a new and flexible platform that separate researchers' efforts of building wireless measurement program from collecting wireless data in the field. Therefore, we designed our system to achieve the following principles:

- **Portability.** MobileSDR turns mobile devices into convenient vantage points, which are easy to carry around and use for field testing.
- **Scalability.** MobileSDR decouples the program design from data collection process. It allows user without domain specific language collecting data for researchers. This

provides crowd sourcing potentials and reduces the bar of field test. The platform also allows researchers to task multiple wireless mobile vantage points simultaneously with a single signal analysis program.

- **Flexibility.** MobileSDR allows putting verified library into the next version of APP and uses domain specific language design, enables users to create their own wireless signal processing program by calling the pre-built block in the APP through python. It also leverages GNU Radio design to enable generating multiple flowgraphs with a few number of blocks.

- **Security.** MobileSDR permits only input non-executable XML/YAML files into mobile devices and reconstructs executable codes internally. The platform will also carefully verify the XML/YAML file before inserting it into endpoints. Moreover, whenever researchers want to add their own designed program, it requires developers do cross-compile of program as C++ dynamic library after verifying the legitimacy of the blocks before putting them into APP during version update.

- **Integrity.** MobileSDR maintains the necessary metadata in sigMF format[180] for data collected from mobile vantage points to ensure data integrity. It also allows researchers to upload their experiment configuration file and platform will automatically allocate the corresponding program to users who match the requirement. Moreover, the mobility of the vantage points in MobileSDR will allow users to get closer to the data source and further from interference, leading to higher data quality.

MobileSDR enables researchers to design their wireless measurement programs in the lab, while separate portable vantage points in the field will run and collect data based on the designed program. This separation of the program design and measurement execution processes offers greater opportunities for measuring wireless data in various locations and at different times. MobileSDR utilizes software-defined radio (SDR) technology to enable wireless measurements

with multiple protocols according to the researchers' program designs. It also provides the capability to backhaul raw IQ samples, offering extensive information for wireless measurement analysis.

MobileSDR adopts GNU Radio flowgraph design as define domain-specific language. It requires pre-compiling the analysis process as a library, which is then called upon when needed. In MobileSDR , a capable Android app has been designed for easy installation on mobile Android devices and straightforward version updates, complemented by a web GUI interface for researchers to easily interact with the platform.

6.3.1 System Architecture

For a wireless measurement project, we can divide it into two stages: the measurement program design stage and the measurement program execution stage. In the design stage, researchers determine the signals of interest and the methods needed for detection. They design programs to tune the hardware to specific signal bands and to detect and extract packets from the air. In the execution stage, they run and test their measurement code while collecting data from the real world.

MobileSDR provides strategies for separating the signal design and measurement execution processes, giving researchers flexibility in conducting experiments at measurement endpoints while maintaining security. This section outlines how we constructed our platform architecture to address the principles we listed in section 6.3 and achieve our goals. Section 6.4 describes the implementation of our platform on web servers and Android devices.

Overview

The MobileSDR architecture consists of three components: measurement endpoints, a broker server, and experiment controllers, as shown in Figure 6.2. Experiment controllers, who may be using a laptop or desktop, aim to measure wireless signals in the real world. Measurement endpoints are Android mobile devices equipped with commercially available SDRs. The broker

server acts as an intermediary, maintaining the message flow between controllers and endpoints.

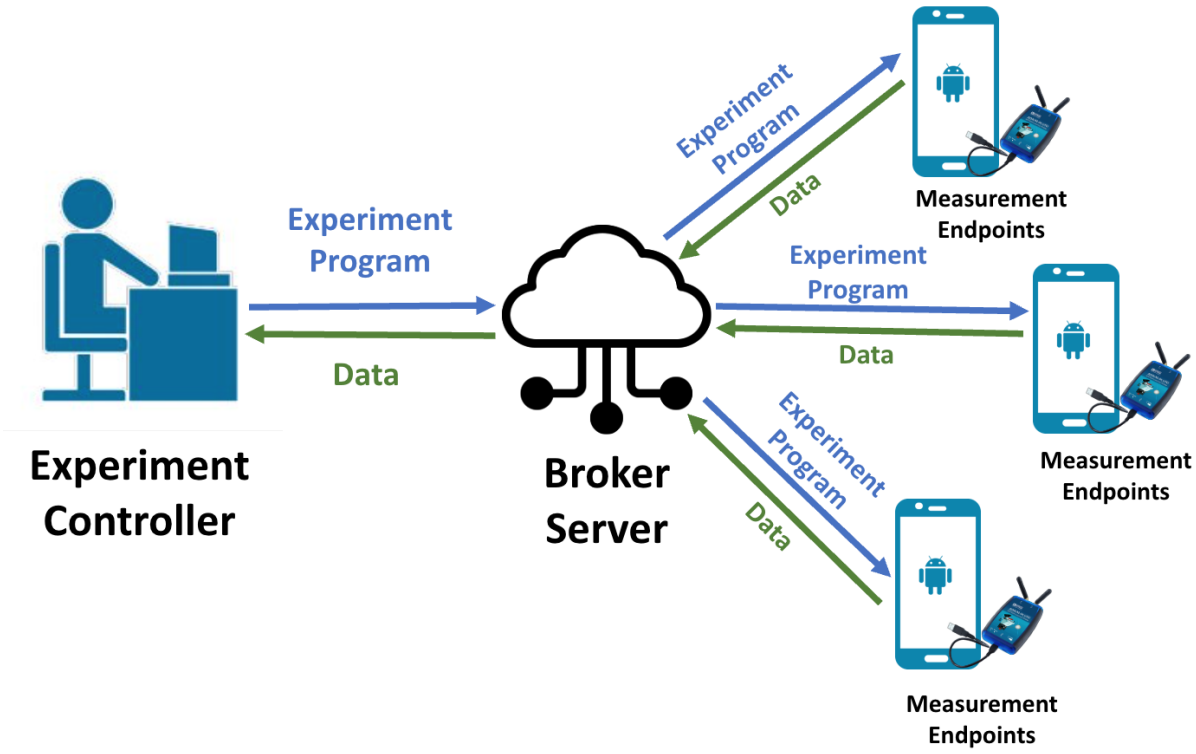


Figure 6.2. Overview of the platform architecture, including three main components: Experiment Controllers, Broker Server, and Measurement Endpoints. The Broker maintains a multi-to-multi relationship between Controllers and Endpoints to send the designed program and backhaul collected data.

In MobileSDR , experiment controllers design the signal process program. They upload this program to the Broker, which then assigns the program from the Controllers to the appropriate Endpoints. Subsequently, the Endpoints execute the program and backhaul the data and execution log to the Broker server. Controllers can actively monitor the log on the Broker and download the data files.

Separating Program Design and Measurement Execution

Separating the design and execution stages allows controllers to distribute tasks among multiple endpoints and collect data. To facilitate this separation, we have designed the measurement endpoints with generic ability to automatic accept and run programs. We also designed

broker to allocate programs according to configuration information provide by controllers and endpoints.

Measurement Endpoints

Measurement Endpoints fulfill the execution role in the platform. Our objective is to facilitate real-world signal measurement and learning. To this end, MobileSDR measurement endpoints combine Android mobile devices with portable SDRs, providing an interface for experiment controllers to load and execute measurement programs. The logic of the experiment resides with the controller, ensuring that the endpoint interface remains simple and universal.

Mobile Android device. In today's era of mobile internet, people commonly carry mobile devices, which have evolved into powerful computing platforms capable of performing complex computations traditionally done by laptops, albeit with some limitations. As demonstrated by bastibl[23], mobile devices can effectively execute various wireless measurement programs. Their portability and mobility offer researchers unique opportunities to gather real-world data.

Given its open nature and Linux-based foundation, Android facilitates the development and installation of custom applications, making it an ideal choice for hosting measurement endpoints due to its ease of integrating with Linux-based systems.

Adapting SDR. Mobile devices alone are insufficient for comprehensive wireless measurements due to their limited hardware capabilities for receiving diverse signal frequencies. Software Defined Radio (SDR) plays a crucial role in wireless measurements, converting traditional hardware implementations into software-based solutions. This approach enables endpoints to need only an Android device equipped with an SDR, handling various frequencies and bandwidths through software. Thus, the measurement endpoints use Android devices to run software programs designed by the Controllers and communicate with the SDR via USB.

Endpoint Information and Configuration. Certain experiments require endpoints to meet specific conditions, such as having a particular SDR, being located in a specific area, or supporting a certain data backhauling rate. MobileSDR endpoints accommodate these require-

ments by registering their configuration details with the Broker, as shown in Table 6.1. This process enables the Broker to assign tasks to suitable endpoints, ensuring efficient and accurate experiment execution.

Table 6.1. Endpoint Configuration Interface

Parameter	Description
Device name	Name of the endpoint, set during registration with the broker.
Geolocation	Current geographical location, automatically determined using the device's GPS.
SDR Config	Type of SDR device, identified automatically when connected to the Android device.
USB port	Type of USB port on the device, which influences the data backhauling rate from the SDR to the device. Automatically detected by the device.
System Info	Information about the device's processing capabilities, including CPU frequency and RAM size, which impacts the data processing capacity. Automatically detected by the device.

Broker Server

The Broker server plays a pivotal role in MobileSDR , facilitating the separation between experiment controllers and measurement endpoints. The Broker connects controllers and endpoints and provides as much information as it can to controllers to help controllers monitor data collection processes. As shown in Fig 6.3, we designed our Broker as a web service that offers several RESTful API interfaces for this communication. Controllers must register with the Broker and obtain an authorization key to access the platform. They can query the Broker for information on available endpoints and search for endpoints that meet their experiment's configuration requirements. Controllers can also send their experiment program to the Broker, along with the endpoint configuration requirements. The Broker then stores the program and assigns it to the qualifying endpoints automatically.

Endpoints must register with the Broker and provide their configuration information as mentioned in section 6.3.1. They are required to send an "alive check" ping to the Broker every second to confirm their activity. The Broker allocates experiments to different endpoints based on their configurations upon receiving the program configuration requirements from Controllers. Endpoints will execute the program automatically when they fetch one from the Broker and

backhual the required file, including runtime logs and collected data, automatically when the program finish and there is network connection.

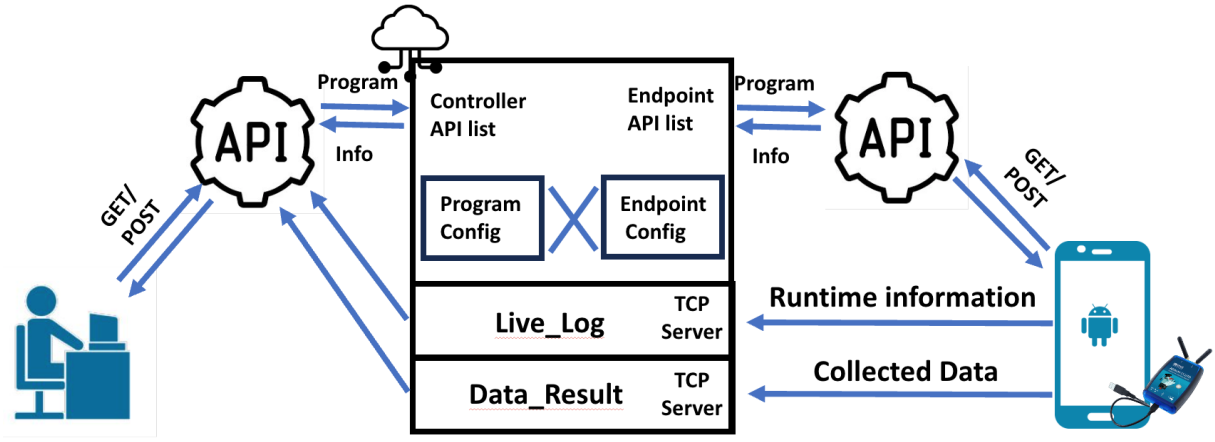


Figure 6.3. Interaction between Controllers and Endpoints with the Broker via API. Endpoints stream data and runtime logs back to the Broker during the experiment, while Controllers can directly download the data and view the logs through the Broker's GUI.

6.3.2 Providing Flexibility While Maintaining Security

In our platform, we enable researchers to conduct wireless signal measurements from multiple vantage points. This flexibility is essential, as researchers often need to monitor specific frequencies and bandwidths and apply unique methods for processing signals. However, deploying programs on arbitrary mobile vantage points poses challenges due to significant trust barriers. Platform operators must ensure that vantage point hosts are fully informed about the activities their devices will undertake, and they must trust controllers to adhere to these assurances.

Access control for measurement vantage points generally ranges from (1) free access, where trusted users can run any code, to (2) Restricted access, which allows any user to perform a limited set of predefined measurements and retrieve their results. These methods either compromise the security of the vantage points or limit researchers' flexibility to conduct experiments. The free access approach is particularly problematic on mobile Android devices, which restrict the insertion and execution of arbitrary binaries.

Domain-specific Language - GNU Radio

To address this challenge, we utilize a domain-specific language (DSL) for conducting wireless measurements. A DSL is tailored to a specific application domain, offering flexibility to researchers while alleviating security concerns. It enables platform operators to define allowable measurements and provides researchers with the tools to assemble their own experiments.

Flowgraph Design in GNU Radio is an example of a DSL. GNU Radio operates on a block-based system where each block serves a specific function, such as filtering, modulation, demodulation, or signal generation. Researchers can configure these blocks' parameters, like filter coefficients, modulation types, and frequencies, and connect them to form a flowgraph. In a GNU Radio flowgraph, data flows from source blocks (e.g., a file source or hardware device) through processing blocks to sink blocks (e.g., a file sink or audio output). Thus, providing access to these blocks allows researchers to construct and execute their own flowgraphs on the devices.

Python Wrapper Running programs on Android devices poses additional challenges. Directly shipping and executing compiled program on Android apps is impossible due to execution restrictions. Unlike server-based systems, it is also impractical to ship source code for compilation on Android devices. However, GNU Radio's Python wrapper enables interaction with the C++ implementations of signal processing blocks and functionalities. This wrapper acts as an intermediary between high-level Python scripts and low-level C++ code, allowing for the development and execution of GNU Radio applications in Python. A significant advantage of Python is its ability to run code directly without compilation. Rather than shipping Python code directly, we distribute XML/YAML files and convert them to python program by using template on Android devices, simplifying access control and program execution.

Access Control

Our primary goal is to develop a platform that enables the general public to participate, collect field data, and upload it to the Broker for sharing among Controllers. Participants with

measurement endpoints need only an Android device and an SDR to conduct tests, without delving into the complex data processing on the device. To address the risks associated with open interfaces and running flowgraphs from Controllers, we have designed several systems to restrict controller activities and safeguard endpoint security.

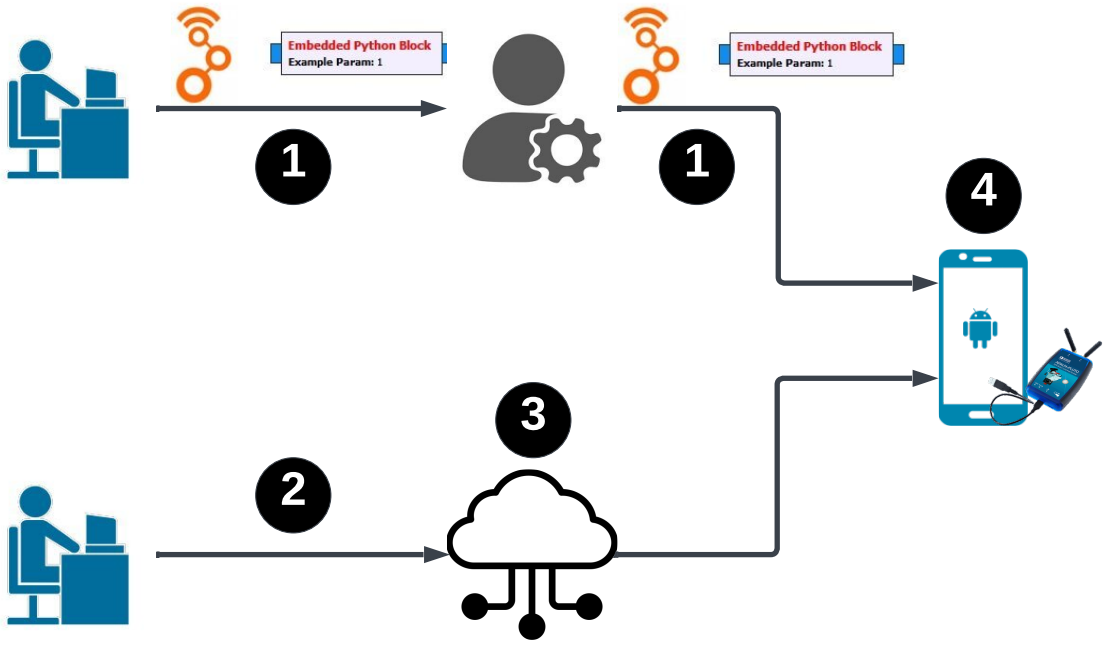


Figure 6.4. Access control flowgraph design. 1) Controllers submit new GNU Radio blocks to administrators for approval to prevent harmful processes. Approved blocks are added to the Android app in a version update. 2) Controllers must register with the Broker to send flowgraphs. 3) The Broker verifies the flowgraph's legality before distribution to endpoints. 4) The app restricts certain blocks and parameters to ensure security.

New Block Update Control. As depicted in Figure 6.4, we exercise complete control over app updates, including the permissible blocks. Controllers proposing new or updated blocks must submit their source code and cross-compiled dynamic libraries for administrative review. The administrator assesses these for vulnerabilities and potential harm to Android devices before integrating the new block into the platform. Following approval, the Android app is updated to support the new blocks, making them available to all platform controllers.

Controller Administration Control. Controller administration is enforced through

mandatory registration, during which basic information is collected. This strict access control enables us to authenticate Controller identities and prevent malicious platform use. Registered Controllers access the Broker's console with a secret key, enabling them to request endpoint lists and send permitted flowgraphs using their administration key.

Flowgraph Verification The Broker verifies the legality of flowgraphs before dispatching them to endpoints. This process includes checks on block permissibility, connection validity, and parameter safety to prevent malicious code insertion in XML/YAML files. Only flowgraphs that pass these checks can be allocated to endpoints.

Blocks Mask on Endpoints To avoid interference with existing signals, source block usage is heavily restricted in our app, which focuses on data collection. Endpoints are limited to capturing air signals using SDR, blocking signal source and network block usage within the platform. Data is temporarily stored locally on the device and then immediately uploaded to the Broker, from where Controllers can access it.

6.3.3 Ensuring Data Integrity

After collecting data, measurement endpoints must backhaul it to the Broker for Controllers to access. It is crucial to ensure the data's integrity, including the associated metadata, to prevent loss during transmission.

Metadata

In wireless data collection, metadata provides crucial context, detailing when, where, and how the data was collected. This information is vital for accurate data interpretation, particularly in dynamic environments. Our system adds metadata by using SigMF format to each backhauled dataset. We list major metadata will be included in the backhual metadata in Table 6.2.

Reliable Data Backhaul

Backhauling data over the Internet can be challenging, especially in environments with limited network bandwidth. Data integrity is paramount for wireless measurement, particularly

Table 6.2. Metadata for MobileSDR backhauled data.

- `Lat/Long` Geolocation where the data was collected.
- `Timestamp` Time when the endpoint completed data recording.
- `CF` Center Frequency used by the SDR to receive the signal.
- `BW` Bandwidth of the channel monitored by the SDR.
- `SR` Sampling Rate used by the SDR for data sampling.
- `Gain` Receiver gain set on the SDR.
- `User Identity` Additional metadata specified by Controllers, such as SNR.

for IQ sample data. MobileSDR employs two strategies to address these challenges.

Reliable Transmission: To ensure data integrity, the use of networking blocks like UDP Source, UDP Sink, and TCP Source blocks is restricted in our system. Instead, endpoints establish a TCP connection with the Broker for data backhaul during measurement execution. Controllers are advised to use file sink operations or save files directly in Android while subsequently use TCP backhaul to the Broker when a network connection is available.

Reducing Backhauled Data: Given the computational capabilities of mobile devices, we recommend that Controllers perform analyses, such as packet detection or demodulation/decoding on the phone, to backhaul only essential information rather than the entire frequency band. Section 6.6 will demonstrate that conducting packet detection and backhauling IQ samples in the field is feasible and yields more significant information.

6.4 Implementation

MobileSDR aims to provide a user-friendly platform that facilitates the study and analysis of real-world signals for researchers and developers. This goal is achieved through a straightforward Broker GUI and an easy-to-use Android app. MobileSDR functions on commercially available smartphones as a user-space service. Our app, designed for endpoints, simplifies

deployment and will be readily available for download from app stores like F-Droid for Android. Moreover, our APP can be adapted on non-rooted devices so that more people can use our platform. This section will detail the implementation of our measurement endpoints and the Broker.

6.4.1 Measurement Endpoints

MobileSDR measurement endpoints integrate Android mobile devices with portable SDRs, connecting to the Android devices via USB. We developed an app by leveraging GNU Radio, enabling signal processing blocks from GNU Radio to operate on Android devices. This setup manages signal reception from the SDR and conducts preprocessing tasks like channelizing, filtering, and packet detection on the mobile device before backhauling the remaining signal to the Broker for further analysis by Controllers.

6.4.2 Running GNU Radio on Android

To facilitate GNU Radio on Android, we cross-compiled the GNU Radio library for ARMv8-A (arm64-v8a, 64-bit), the dominant architecture in contemporary Android devices. This cross-compilation involved linking various dependent libraries, such as ZeroMQ, FFTW, GMP, USB, Boost, and Android-specific binaries like SWIG. Rather than utilizing the Android NDK, we opted for Termux, which provides a full Linux environment on Android without root access, featuring a package manager and a broad range of development tools and libraries. Termux enables direct compilation of C++ libraries and Python API dependencies on Android devices. Besides that, we leveraged some prebuilt dependent libraries like Boost from bastible's project [23].

Currently, our platform supports GNU Radio version 3.8 because versions above 3.9 depend on pybind11, which lacks support for packet compilation on Android.

Android apps are usually developed in Java or Kotlin, but GNU Radio is built with C/C++ and Python. To integrate GNU Radio with Android, we employed the Java Native Interface

(JNI) and Chaquopy. Chaquopy, a Python SDK for Android, allows incorporating Python code within Android projects, facilitating the use of Python libraries and scripts with Java or Kotlin. It supports a wide range of Python libraries and integrates well with the Gradle build system of Android Studio. In GNU Radio, Python serves as a wrapper for the C++ system library via SWIG, with Python code invoking underlying C++ libraries to process the data stream. To enable Python to call C++ dynamic libraries in the Android environment, these libraries must reside in the NativeLib path. Our use of JNI in the CMakeLists.txt file dynamically links these libraries, ensuring their upload to the NativeLib directory by Android.

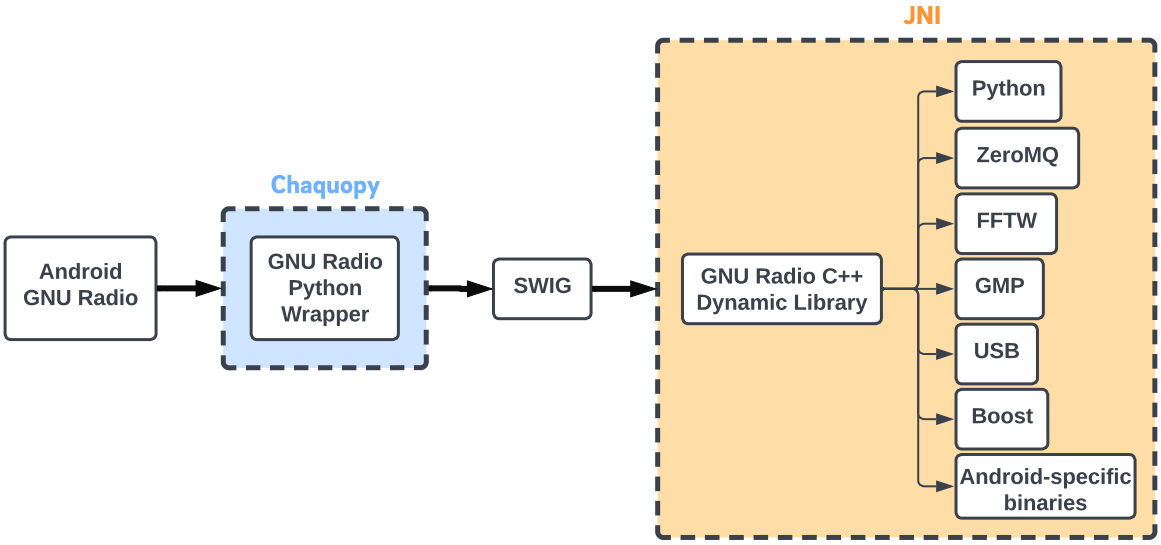


Figure 6.5. Library dependencies in Android GNU Radio. A -¿ B indicates that A depends on B.

6.4.3 Communicating with SDR through USB

Modern Android smartphones restrict hotspot interfaces for network sharing with SDR devices in master mode, necessitating USB interface communication with the SDR. Our system uses the libusb-1.0 dynamic library from Termux to enable phone-SDR communication. In USB host mode, the Android device acts as the USB host, powers the bus, and enumerates connected USB devices, allowing Android apps to interact with USB hardware via standard file I/O operations. Android generates an integer file descriptor for a low-level USB device

connection, facilitating read and write operations with the SDR.

Our platform supports popular SDRs like the USRP BXXX and PlutoSDR. For USRP BXXX SDRs, firmware compatible with Android is required to reattach the device to USB in a mode suitable for SDR applications. We use bastibl-developed firmware [23], loading it onto USRP devices at application startup, capturing the file descriptor and device name for Android-USRP communication. For PlutoSDR, we adapted the libiio and gr-iio codes for communication via the file descriptor, obtaining the descriptor and URI at process start to ensure PlutoSDR connectivity.

6.4.4 Converting Flowgraphs into Python Code

Our Broker distributes XML/YAML files, created by the GNU Radio Companion (GRC) GUI, to measurement endpoints. To utilize these XML/YAML formatted files, we convert them into Python scripts using GNU Radio's conversion tools, which employ Mako, a high-performance Python templating engine. We edit the template to make it suitable to run in our APP. This conversion checks the .grc file against YAML files in each GNU Radio block directory, tailoring Mako templates for No-GUI flowgraphs on Android, adjusting block settings for optimal Android performance. We've also customized conversion templates for various GNU Radio blocks to ensure smooth operation and integration in the Android environment.

6.4.5 App UI Interface

We developed an Android app providing endpoint interfaces on Android devices. Initially, the app makes an HTTP GET request to download task from the Broker and proceeds to convert the XML/YAML file and initiate GNU Radio signal processing within a Python module. The app's functionality was tested on Oneplus 10 pro, with a visual representation of the UI layout included in the accompanying figure.

6.4.6 Broker

The Broker in MobileSDR is a web server developed in Golang, orchestrating the interaction between controllers and endpoints. It manages registrations, oversees data transfers, and facilitates task assignment and reporting, ensuring secure and efficient communication. The Broker leverages various Go libraries such as `net/http`, `sync`, `sync.Map`, `dgrijalva/jwt-go`, and utility libraries like `strconv`, `strings`, and `encoding/json`.

Communication between the Broker, controllers, and endpoints follows a request-response model over HTTP, using RESTful APIs. This design ensures scalability and flexibility, supporting a range of operations crucial for real-world deployment.

For demonstration and testing, a GUI has been implemented in the Broker to improve user interaction and system visibility. This GUI dynamically displays available endpoints and their SDR information, allowing users to select and manage specific endpoints. It provides real-time results, enabling direct uploads of XML files for processing and monitoring responses. The GUI updates automatically, presenting the latest standard output data and IQ samples as tasks are accepted by endpoints.

6.5 Evaluation

6.5.1 Battery Usage and Power saving

In our platform, we use phones to power SDRs. However, power is a limited resource on Android devices when they are not connected to an external power source. The broadband SDRs we use, including PlutoSDR and USRP B200, have FPGAs that also require substantial power when functioning. As shown in Figure 6.6, when the program is running, it consumes nearly three times more power than the phone's idle power usage with the screen on. Moreover, even when no experiment is running, the power usage with just the SDR connected and firmware loaded is about 0.4mWh per second, which prevents the phone from lasting more than 12 hours in idle mode.

Android does not allow us to directly disconnect the USB via software. To conserve power, we place the SDR into ”low power” mode when not conducting experiments. The USRP B200 has two modes when connected to a USB port: an initial mode where the firmware is not loaded, which consumes only 0.1mWh per second—a rate four times less than when the firmware is loaded, as observed at the start of the red line in Figure 6.6. Therefore, we opt to reset the firmware when an experiment ended and reload B200's firmware before each experiment. For the PlutoSDR, we can turn off its RX and TX channels, saving 0.2mWh per second. As depicted in Figure 6.6, when an experiment is completed and no new experiment starts within the next 10 seconds, we switch the SDR to a ”lower power” mode to save battery.

However, resetting the B200's firmware causes its USB interface to reload, prompting Android to request user permission for the new USB device. For the PlutoSDR the FPGA is still running which consumes non-trivial number of powers even when the channel were turned off. We anticipate that shutting down the FPGA or running an idle FPGA in the SDRs will aid in power conservation, and we plan to develop an idle FPGA firmware and write it into the SDR as part of our future work.

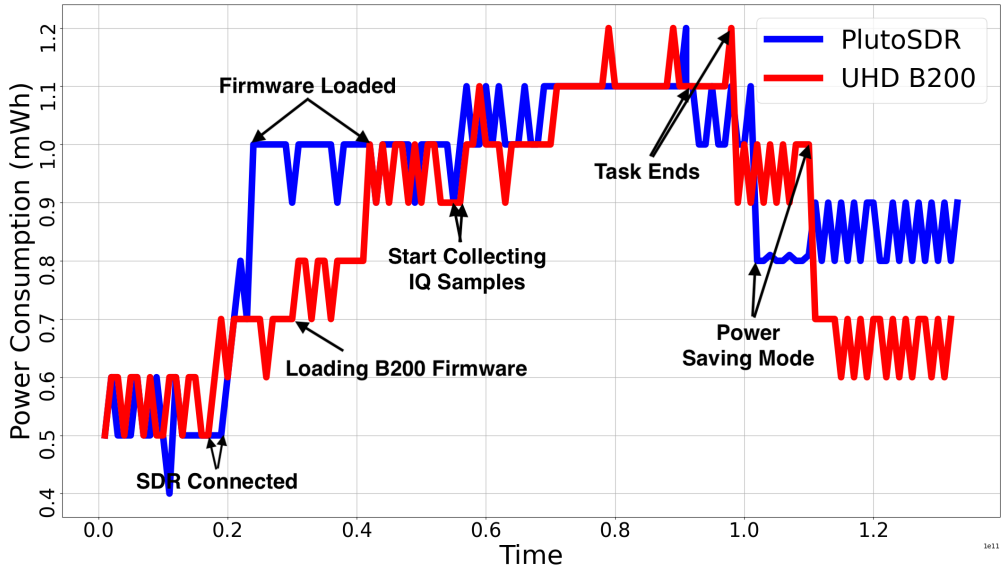


Figure 6.6. Battery usage for PlutoSDR and UHD B200 SDR for one round experiment.

6.5.2 SIMD and GPU benefits to the system

Given the limited CPU resources on Android devices, utilizing Single Instruction Multiple Data (SIMD) processors and GPUs to accelerate radio data processing is promising. As demonstrated in [23], using VOLK to implement SIMD-accelerated functions has proven to speed up data processing compared to traditional C implementations. Therefore, we plan to develop large data processing blocks using VOLK as part of our future work.

However, the results from GPU acceleration were not as promising for data processing. We tested data throughput on a OnePlus 10 Pro phone using an OpenCL benchmark tool developed by [23]. Contrary to [23], Figure 6.7 indicates that the GPU data processing rate is generally lower than that of the CPU when the buffer size is within 65,536 Bytes. The potential reason is that the time-consuming GPU loading process diminishes throughput gains more significantly than it enhances data processing speed. Therefore, GPUs may not be effective accelerators for signal processing in modern smartphones.

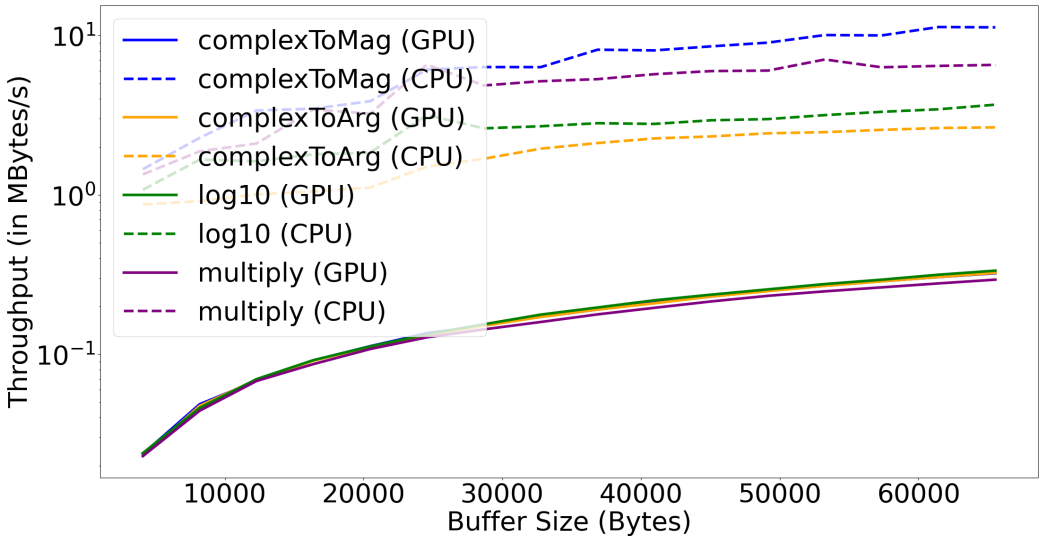
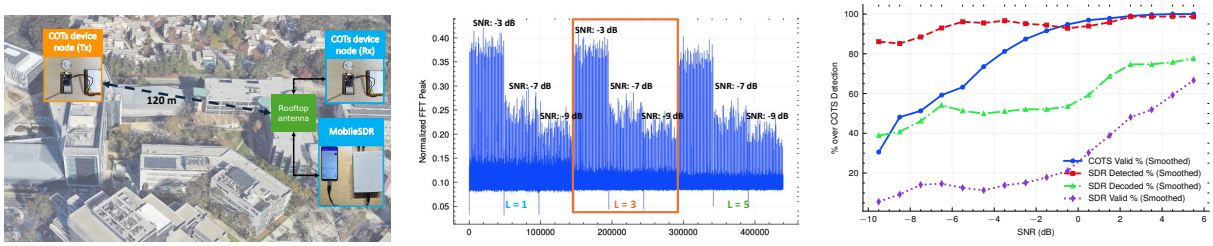


Figure 6.7. GPU data process throughput vs CPU data process throughput

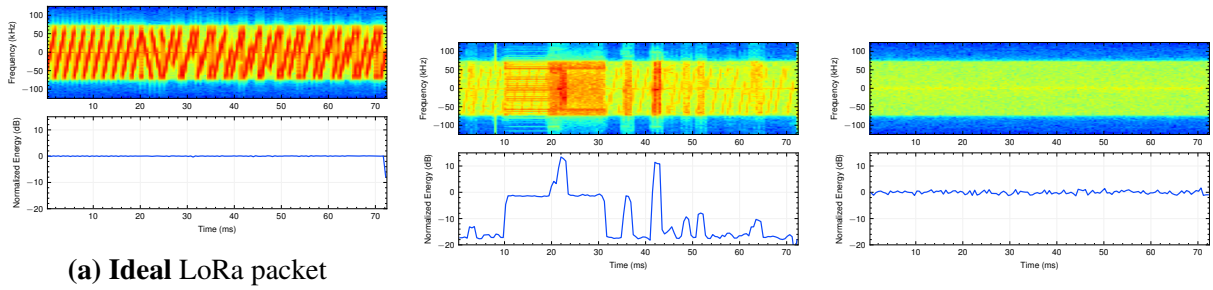


(a) Over-the-air experiments hardware setup with 2 COTs devices and MobileSDR.

(b) Normalized correlation for varying codeword (correlation seq) lengths and SNRs.

(c) MobileSDR detection, decoding and valid packets ratio with COTs detected packets.

Figure 6.8. Figure illustrating the over-the-air LoRa experiments with MobileSDR. Fig(a) shows the over-the-air experimental setup, and Fig (b) illustrates optimal codeword length (n) and threshold on normalized correlation. Fig (c) demonstrates that MobileSDR's detection rate is close to COTS detection (100%).



(a) Ideal LoRa packet spectrogram and normalized energy (instantaneous to avg ratio)

(b) LoRa decoding failed due to **interference**: spectrogram and normalized energy

(c) LoRa decoding failed due to **low SNR**: spectrogram and normalized energy

Figure 6.9. Figure demonstrating the lora packet spectrogram and normalized energy (ratio of instantaneous energy to the average energy across time). Illustrating idea packet and decoding failure scenarios with interference and low SNR.

6.6 Use case

Real-world outdoor measurements are pivotal in evaluating network performance and identifying deployment issues. To address this, we developed a setup characterized by its versatility, enabling it to function both as an independent unit and in conjunction with Commercial Off-The-Shelf (COTS) devices. This dual functionality enables the identification of issues and provides valuable insights into the causes of failure. Importantly, our platform is protocol-agnostic as long as the requirements for data transmission rates are not high. This characteristic ensures broad applicability across various protocols designed for low data rate transmissions,

such as LTE MIB and SIB measurements, Frequency Hopping Spread Spectrum (FHSS), and LoRa.

To demonstrate the practical advantages of our platform, we conducted two case studies, one focused on detecting LoRa interference in the ISM band while the other trying to decode LTE base station information by extracting MIB and SIB information.

6.6.1 LoRa Use case

LoRa-based IoT sensors are increasingly being deployed worldwide for various applications, with significant adoption in smart city initiatives such as smart parking, utility monitoring, waste management, and intelligent transportation systems [156]. This widespread usage is primarily driven by LoRa's long-range capabilities, energy efficiency, and low-complexity hardware. However, one challenge associated with LoRa is its susceptibility to interference, which can result in the failure to detect or correctly decode transmitted packets. Since LoRa operates in the ISM band alongside other protocols, interference from nearby LoRa transmissions or other protocol signals in the same frequency band can lead to a significant drop in performance, especially in high-interference environments.

To ensure optimal performance, LoRa gateways must be strategically placed to support the maximum number of IoT nodes while minimizing interference. Achieving this requires extensive data collection from various locations to identify potential performance drops and the underlying causes. While traditional COTS devices can estimate performance, they lack the capability to identify the specific reasons for failures. MobileSDR offers a more advanced solution by enabling data collection on the move and transmitting digitized (IQ) samples to the cloud whenever a LoRa packet is detected. This system allows data to be gathered across different locations, and the collected IQ samples help determine whether performance failures are due to low signal-to-noise ratio (SNR) or interference, providing a more comprehensive understanding of network's link and phy-layer issues.

Hardware Setup: The hardware setup for our over-the-air experiments is shown in

Figure-6.8a. We utilized two Adafruit RP2040 devices (COTs), one as the transmitter and the other as the receiver. A rooftop antenna was connected to a splitter, which was then linked to both the COTs receiver and the MobileSDR. On the transmission side, a COTs device was used to send known LoRa chirps. This setup was used for our evaluations, with normalized correlation, detection rate, decoding rate, and valid packets as performance metrics. A valid packet refers to one where the decoded bits match the transmitted bits; otherwise, the packet is considered invalid.

6.6.2 Improved packet detection

It is well known that Software-Defined Radio (SDR) detection and decoding performance is generally inferior to that of commercial off-the-shelf (COTS) devices. We observed that the detection accuracy of open-source gnu-radio implementations [142, 211] drops as SNR decreases. Typically for detecting the LoRa packets, at the receiver, they correlate with a known codeword (up-chirps), the peaks in correlated sequence help in determine whether there is a LoRa packet or not. However, at low SNRs these peaks burry into noise floor and receiver won't be able to detect packets [211]. To enhance detection rate, our approach used a normalized correlation technique coupled with thresholding and increasing the correlation sequence length.

$$r_{xy} = \frac{\sum_{i=1}^{n} y[i] * x^{H}[i]}{\sqrt{\sum_{i=1}^{n} \left\|y[i]\right\|^{2}} \sqrt{\sum_{i=1}^{n} \left\|x[i]\right\|^{2}}}$$

Where, n is correlation samples [146, 198], x is the known codeword or correlation sequence (L upchirps). y is the received digitized (IQ) samples after bandpass filtering.

As we increase the correlation length, we can see the correlation peaks even in low SNR, improving detection rate in low SNR conditions. However, the use of longer codeword increases the correlation gain for samples (noise) even without any preambles/packets, leading to false positives (Figure- (a)). To address this issue, we implemented a thresholding mechanism that filters out peaks below a defined threshold, thereby preventing false positive scenarios. As

illustrated in figure-6.8b, our detection rate depends on two hyperparameters - codeword length and threshold. In the Figure, we show vary three codeword lengths with L as 2, 4, and 6 chirps, for each scenario we captured -3, -7 and -9 SNR cases to show how normalized correlation changes with SNR and codeword lengths. At low SNRs, increasing the codeword lengths with thresholding enabled us differentiate actual correlation peaks from noise and improve detection rate.

We then compare MobileSDR detection rate with COTs devices. From over-the-air experiments outdoor setup, we captured COTS decoded packets, valid packets and SNR. With this ground truth we computed detection rate, decoded packets and valid packets ratio with repect to COTS detected packets. As illustrated in figure-6.8, we show that MobileSDR detection ration is close to COTS detected packets (100%). MobileSDR pushes the IQ samples of detected packets for further processing in server. We used demodulion block from open source git repository [211] to show decoding and valid packets ratio. However, we observed that SDR packet decoding and valid packet ratio is low compared to COTS and leading to many decoding failure packets.

6.6.3 Identifying reasons for decoding failure

MobileSDR can identify the causes of packet decoding failures, such as interference or low SNR, by calculating the normalized energy or the instantaneous energy-to-average energy ratio and visualizing the corresponding spectrogram plots for further analysis. As shown in Figure-6.9, the normalized energy for an ideal packet remains around 0 dB (Figure-6.9a). In the presence of interference, the normalized energy becomes distorted, with peaks deviating significantly from the average energy (Figure-6.9b). In contrast, under low SNR conditions, the normalized energy stays close to 0 dB, but the average energy over time is significantly lower than the expected level (Figure-6.9c).

Several novel academic approaches [73, 93, 202, 157] address challenges in low SNR, simultaneous packet decoding, and other LoRa issues by leveraging advanced signal processing

and machine learning models. Most of these methods require IQ samples as input. MobileSDR supports such approaches by providing IQ samples, enabling integration with the latest decoders that convert IQ samples into LoRa bytes.

6.6.4 Cellular Information Use Case

Base stations play a crucial role in the operation and maintenance of mobile networks. The information from base stations is a vital data source for location-based service (LBS) providers [192] and can help optimize networks by tracking extensive data transmission [61].

In cellular networks, the Master Information Block (MIB) and System Information Blocks (SIBs) are essential components that base stations use to communicate key operational parameters to mobile devices. Broadcasted periodically, the MIB contains critical information needed for mobile devices to access and communicate with the network, such as the downlink system bandwidth, system frame number, and Physical Cell ID. This information is essential for cell selection and synchronization. While the MIB provides foundational information, SIBs offer more detailed system information necessary for a device's network operation. SIBs are segmented by content type and importance; for example, SIB1 typically contains cell access options and SIB scheduling, SIB2 may include radio resource configurations, and other SIBs can provide details about neighboring cells, frequency information, and network-specific policies.

Collecting MIB and SIB information is crucial for base station vendors and ISPs to debug base stations in the field. However, these data are not directly retrievable from phones as they are encapsulated by Qualcomm modems. Additionally, phones usually attach to one frequency band, missing other frequencies in the same area while hard to have it switch to anther band. In this use case, we demonstrate how our platform collects MIB/SIB information.

6.6.5 IQ Sample Collection (at Phone)

Unlike laptops, a phone's data collection capability is limited not only by the USB port transmission rate but also by the CPU rate. To collect MIB/SIB information from a 20MHz

bandwidth base station, a 30.72MHz sample rate is required. We used a OnePlus 10 Pro Android device with a USB 3.0 OTG port. The device can continuously receive IQ samples without overflow at rates up to 10MHz. Beyond 10MHz, buffer and CPU usage quickly become saturated, leading to immediate overflow. However, we found the phone could still handle several megasamples in a short amount of time at high sampling rates, and since cellular base stations transmit MIB/SIB information every 80ms, we designed a GNU Radio block based on the UHD source block to immediately restart file overwriting when overflow occurs. It resets the sample count and stops immediately after receiving a user-defined number of samples.

To enable the collection of all cellular base station information in a specific area, the block allows users to select a list of frequencies for data collection. We designed a default flowgraph in our system for endpoints to run initially when no tasks are assigned. It collects 5M IQ samples for each cellular frequency band at a 30.72MHz sample rate with sigMF metadata and uploads the data to a broker once the experiment is complete.

6.6.6 MIB/SIB Information Decoding (at Server)

Once the collected IQ samples are uploaded to our broker, we extract MIB/SIB information using Matlab [79]. We referred to Matlab SIBRecovery [112] and used sigMF metadata to decode base station information. With the geolocation information, we post the collected base station data onto geographical maps. By cross-verifying the decoding results with the cell information obtained from the phone, we confirmed that the MIB information could be decoded correctly.

6.7 Conclusion

In this work, we introduced MobileSDR, a flexible and scalable platform for conducting wireless field tests using mobile devices equipped with Software Defined Radios (SDRs). By separating signal processing from data collection, MobileSDR enables widespread participation in wireless network diagnostics, allowing researchers to deploy experiments remotely while

relying on crowdsourced data collection.

The platform successfully overcomes the technical challenges associated with mobile-based SDRs, such as limited CPU resource, battery constraints, and security risks, by implementing innovative solutions like energy detection in FPGAs and secure, programmable flowgraph deployment through GNU Radio. Real-world case studies demonstrated the effectiveness of MobileSDR in diagnosing LoRa network interference and performing cellular base station measurements, highlighting its potential for broad applications in wireless diagnostics and optimization.

Future improvements, such as incorporating enhanced power-saving techniques and leveraging SIMD acceleration for more efficient signal processing, will further extend the capabilities of MobileSDR. Ultimately, this work establishes a foundation for large-scale, cost-effective, and flexible wireless network monitoring through crowdsourced efforts, making wireless field tests more accessible to researchers and industry professionals alike.

6.8 Acknowledgement

Chapter 6, in part, is currently being prepared for submission for publication of material. Zesen Zhang, Rohith Reddy Vennam, Maiyun Zhang, Yunxiang Chi, Dinesh Bharadia, Aaron Schulman. The dissertation author was the primary investigator and author of this material.

Chapter 7

Conclusion

Access networks are an essential components of the Internet infrastructure that connects millions of users to the Internet. Measuring and understanding access network designs and its infrastructure components can lead to successfully evaluate its performance, and resilience. In this dissertation, I identified key challenges that obstruct accurate analysis of access networks: lack of transparency of access network components, significant variability in network design, and deployment across vendors and regions. These issues have contributed to inaccurate assumptions and generalizations in prior research, limiting the applicability of their findings to the diverse real-world access network landscape. To overcome these challenges, I proposed a set of new methodologies that combine existing techniques with new access network-specific probing and analysis to improve the visibility of access network and revealing access networks.

By introducing McTraceroute and ShipTraceroute, I demonstrated the possibility of revealing regional access network topologies. In Chapter 3, we presented a comprehensive exploration of the topology of regional access networks. The methodologies and insights enable inference of these otherwise opaque access network topologies. Leveraging a diverse set of active measurement tools—ranging from traceroutes and alias resolution to innovative data collection via public WiFi hotspots and mobile devices in transit—I demonstrated it is feasible to map the hierarchical topology of both wireline and mobile ISPs' regional access networks. The finding reveals substantial variation in aggregation strategies and redundancy across providers

and regions. I highlighted their impact on resilience, latency, and end-user experience.

By revealing regional access networks' topologies, I assessed physical risks to reliability. In Chapter 4, we examined the physical vulnerabilities of U.S. regional access networks and demonstrated how these weaknesses pose critical threats to Internet availability and regional network resilience. Through a combination of large-scale empirical measurements, infrastructure mapping, and real-world outage case studies, we showed that access networks are susceptible to targeted physical attacks despite built-in redundancy. Our findings reveal how attackers can infer and locate critical network components using public records, wardriving techniques, and traceroute analysis, and how outages at even a single facility can impact millions of users. We also discussed feasible mitigation strategies and the operational trade-offs network operators must consider to protect this vital infrastructure.

By designing controlled mobile phone experiments, I uncovered LTE base station scheduler design variations across vendors and I evaluated the impact on the downlink throughput. In Chapter 5, I provided the first empirical comparison of proprietary downlink scheduling behaviors across four major cellular base station vendors—Ericsson, Samsung, Nokia, and Huawei. Through carefully controlled experiments in idle network environments, I demonstrated that vendors adopt distinct strategies for resource block distribution and modulation scheme selection based on channel quality, which can significantly affect throughput and user fairness under contention. I further observed that scheduling decisions vary not only across vendors but also between macro and micro deployments. These findings challenge the assumption of scheduling uniformity in prior LTE studies and underscore the need for network analysis and optimization tools to account for vendor-specific scheduling behaviors.

By presenting MobileSDR, I illustrate the feasibility of evaluating wireless signals under a variety of interference in the field. In Chapter 6, I introduced MobileSDR, a pioneering mobile SDR platform that enables flexible, crowdsourced wireless field testing using smartphones paired with SDR hardware. By decoupling signal processing from data collection and providing secure, programmable interfaces via GNU Radio, MobileSDR democratizes access to wireless

diagnostics traditionally limited to expensive, specialized equipment. I demonstrated that it is feasible to perform real-world spectrum analysis, interference diagnosis, and base station information extraction through portable endpoints. I performed detecting LoRa interference to decoding LTE MIB/SIB information as use cases for the platform. These results highlight MobileSDR's potential to transform wireless research and diagnostics by making scalable, region-flexible signal analysis widely accessible.

7.1 Future Work

While this thesis takes critical steps toward understanding regional access networks, I want to end my thesis with some open questions that immediately can be derived from my work.

Chapters 3 and 4 introduce several promising directions for enhancing our understanding and resilience of regional access network infrastructures. First, given the topology and detailed infrastructure we have observed, can we predict future risks? For instance, how might criminal threats or extreme weather events in a specific region impact network services? Additionally, can AI-driven anomaly detection be leveraged to identify risks early, enabling companies to respond proactively and mitigate potential attacks? Second, revealing access network topology presents new opportunities for more precise IP geolocation. Our topology maps illustrate how packets may be rerouted to distant aggregation centers before reaching a physically closer destination. Furthermore, my analysis of IPv6 addresses highlights the potential to extract building-level information, which could significantly improve mobile IP geolocation accuracy.

Chapter 5 provides a comprehensive comparison of downlink scheduling strategies across major LTE base station vendors, uncovering several opportunities for deeper exploration. Future research could enhance the intelligence of both user equipment (UEs) and base stations. For UEs, understanding network conditions—such as congestion in specific areas—could enable them to switch dynamically to base stations with more balanced traffic, improving data delivery and reducing webpage loading delays. For base stations, my experiments revealed that a single UE

demanding a large share of resources can degrade throughput for competing users. This suggests the need for more adaptive scheduling algorithms, potentially leveraging reinforcement learning or other AI-driven techniques, to optimize throughput and fairness in real time. Additionally, as 5G deployments expand, continuing this line of research with standalone 5G NR base stations will be essential. Similar experiments will help determine whether vendor-specific scheduling differences persist or if next-generation networks move toward greater standardization.

In Chapter 6, I present my wireless measurement platform, which significantly reduces the cost of measuring and collecting wireless signals in the field. This platform enables a range of potential use cases. For instance, we can easily deploy multiple endpoints on buses or provide them to truck drivers. These endpoints can automatically collect data during their regular routes and report back cellular signal conditions in various areas, helping ISPs refine their base station deployment strategies. Another promising direction is exploring how to intelligently distribute tasks across the platform's three components: the SDR, the smartphone, and the cloud. Effectively leveraging the computational and storage resources of each component to improve overall platform efficiency is an open problem that remains unexplored.

Appendix A

Additional Materials for Access Network Measurement

A.1 Ethics considerations

A.1.1 Transportation of Lithium Batteries

In the US, lithium batteries are considered hazardous materials, and shipment of them must comply with regulations set out by the US Department of Transportation (i.e. US 49 CFR §172.185 [126]). The relatively low capacity of batteries in smartphones, and the fact that they are contained within equipment (i.e., the smartphone), allows for them to be shipped by ground. Rules about shipment of a powered-on devices however, are not clearly specified. We are aware of a device with a similar operating mode that is widely in use today: shipment tracking devices that use Cellular radios and GPS to report package locations during shipment. One is even available directly from the US Postal Service [80].

Our institution's shipment coordinator—who routinely deals with shipments of hazardous materials such as medical supplies—contacted the US Department of Transportation for clarification on shipping powered-on smartphones. They confirmed that as long as the smartphones do not create a dangerous evolution of heat, or have the risk of catching fire while in transit, shipping powered on devices is permitted. To ensure there were no hazardous conditions our devices could enter while running this software, we thoroughly tested our smartphones in ex-

treme environmental conditions that could be experienced during shipment in trucks/railcars. We operated the at 44° C, and -2° C and for several hours while running our measurement, and the phone continued to operate properly. The smartphones we use also have an automatic thermal shutdown feature as an additional safeguard [147].

A.1.2 Characterizing Critical Infrastructure

Although this study does not involve experiments with human subjects, there are sensitivities with revealing information about critical infrastructure that may provide advantages to adversarial actors.

Although the Belmont report outlined principles relating to human subjects, the 2012 Menlo Report proposed a framework specifically targeting computer and information technology research [85]. Its companion report provided a set of case studies applying the framework [52]. The Menlo Report is a more appropriate framework for our analysis because it explicitly addresses stakeholders such as network/platform owners and providers but also acknowledges that they may warrent different consideration from that of individuals.

Our considered view is that the benefit of our research exceeds potential risk to infrastructure. We are now entirely dependent on this infrastructure but there has been little attention to independent objective understanding of its resilience and reliability. Given increasing attention to the need for regulatory oversight of the Internet as critical infrastructure, it is important to understand just how much a capable independent third party can accurately infer about various aspects of Internet infrastructure. We need to understand this capabilities in order to know what adversarial actors could likely achieve, as well as to know how benign actors might help to reduce the burdens of government by providing independent confirmation of claims of reliability and/or resilience of critical network infrastructure.

We also have long-standing cordial relationships with engineers at the providers we have studied, who are aware and supportive of our work. Specifically, we discussed our inferences with Comcast, Charter, AT&T, and T-Mobile engineers throughout our study, for the purposes of

Table A.1. To observing CO interconnections in traceroute, we map IP addresses to COs, and account for outdated and missing information.

	Comcast	Charter
Initial	204,744	54,079
Alias Resolution		
Changed	2.35%	1.10%
Added	2.76%	0.80%
Removed	0.86%	0.20%
	208,640	54,407
Point-to-Point Subnets		
Changed	0.04%	0.05%
Added	1.27%	0.48%
	211,295	54,670

validation of our findings.

A.2 Details about Comcast and Charter Mapping

A.2.1 Mapping IP Addresses to Hostnames

The traceroute probing yields IP address paths, and we attempt to map each individual addresses to a backbone or regional CO. We use both `dig` and the Rapid7 rDNS dataset to perform reverse lookups on the addresses, prioritizing the `dig` names to reduce potentially stale names in Rapid7. Comcast and Charter appear to connect both their backbone and regional routers with point-to-point links, so we also lookup names for all IP addresses in the same /30 subnet as a traceroute IP address. The /30 subnet includes all addresses possibly used in a point-to-point link with that address. Using regular expressions, we extract CO and region identifiers from the names, creating an initial mapping from IP addresses to COs. We perform two steps to improve the CO mappings to account for missing and outdated rDNS names (Table A.1): (1) resolve router aliases to map groups of addresses to COs, and (2) add additional constraints using point-to-point subnet addresses.

First, we use Mercator and Midar alias resolution to infer addresses that belong to the same router, since these addresses reside in the same CO. We included all of the traceroute

addresses, as well as the additional addresses in their /30 subnets. If more addresses in an inferred router map to one CO than any other CO, we remap all addresses in the group to that CO. We do not apply a minimum threshold for the number of router IP address hostnames containing a CO identifier. In the event of tie, we remove all CO mappings for the addresses to avoid potentially misleading information. The alias resolution modified or added more CO mappings for Comcast (5.1%) than Charter (1.9%).

Next, we use point-to-point subnets to further refine the CO mappings. Interconnected router interfaces must have IP addresses from the same IP subnet, and network operators usually assign these addresses from a point-to-point subnet; e.g., /30 or /31 subnets in IPv4, both of which include two usable interface addresses. Based on the IP addresses in our traceroutes, it appears that Comcast typically uses /30 subnets, while Charter uses /31 subnets to interconnect routers in different COs. Routers typically respond to traceroute with the inbound interface address, so the other address in the point-to-point subnet often belongs to the router at the prior traceroute hop. If that other address has a CO mapping provided by rDNS or alias resolution, we can use that information to refine the mapping for the prior hop.

Figure A.1 illustrates our approach to using point-to-point IP subnets to further refine the CO mappings with two traceroutes through a router in a regional access network, and initial CO mappings for each of the addresses. The initial mappings indicate that IP address x belongs to a router in CO1, but both paths reveal subsequent addresses where the other address in each subnet (y' and z') map to CO2. y' and z' most likely belong to the same router as x, so we use them as possible indications that we initially mapped x incorrectly. Here, more addresses map to CO2 than CO1, so we re-map x to CO2. If x lacked an initial CO mapping, then we would use the mappings for y' and z' to infer a mapping for x.

A.2.2 Removing CO Adjacencies

Initially, we collect all immediate IP address adjacencies where both addresses have a CO mapping (Table A.2). MPLS tunnels can cause false links to appear in traceroute, so we

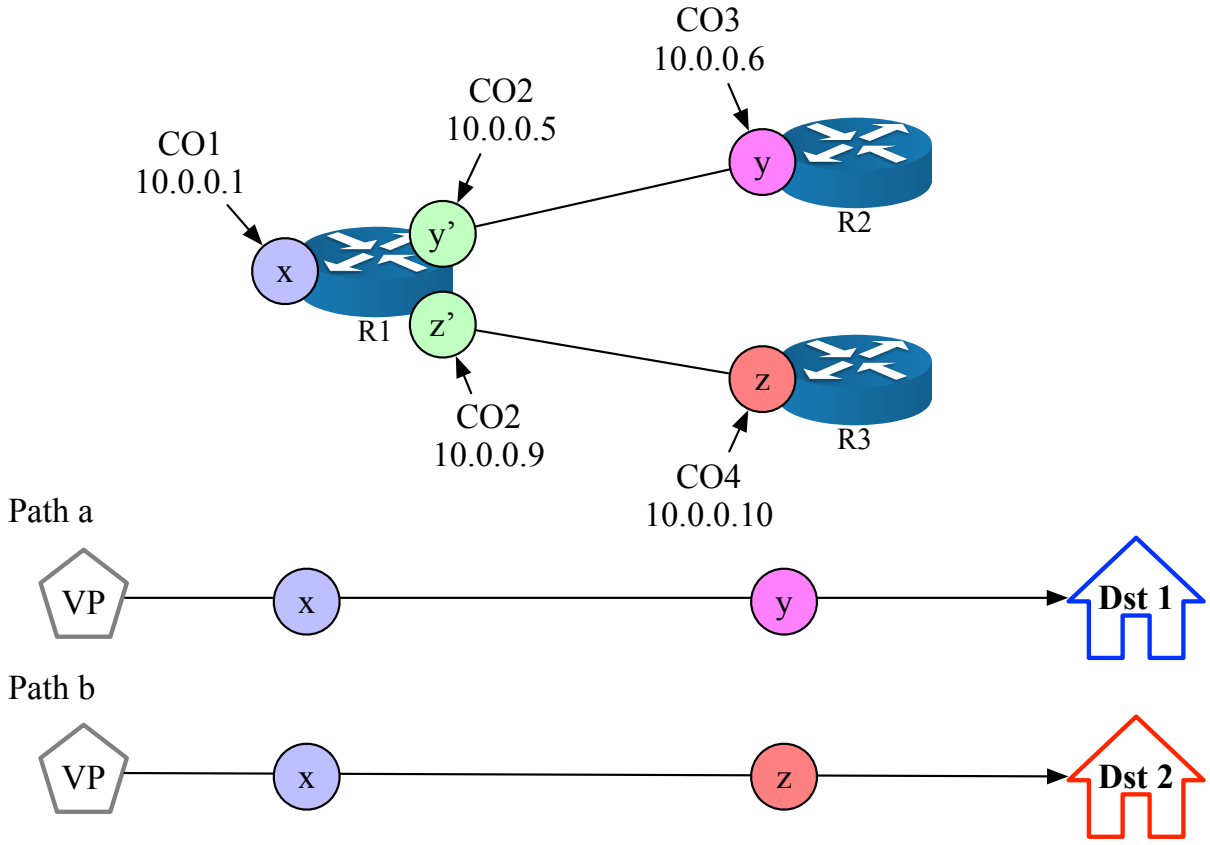


Figure A.1. The two paths reveal x followed by two different addresses, y and z. Presuming that y and z belong to /30 subnets, we use the other address in each subnet (y' and z') to correct the CO mapping for x.

Table A.2. The unique adjacent IP address adjacencies (IP Adjs) and unique CO adjacencies (CO Adjs) pruned to account for stale rDNS and traceroute path corruptions.

	Comcast		Charter	
	IP Adjs	CO Adjs	IP Adjs	CO Adjs
Initial	95,671	4777	64,667	3994
Backbone	26.07%	7.39%	11.67%	5.02%
Cross-Region	4.45%	18.78%	1.78%	2.37%
Single	0.06%	1.15%	0.03%	0.43%

use the approach by Vanaubel *et al*. [190] to reveal MPLS exits and the tunnel IP addresses by conducting follow-up traceroutes to all IP addresses mapped to COs in the original traceroute collection. If a pair of addresses appears adjacent in our initial probing, but are separated by one or more hops in the additional MPLS traceroutes, we remove the pair since it is likely the entry and exit of an MPLS tunnel. In our maps, we only observed MPLS tunnel behavior in one Charter region, although we observed this behavior throughout the region.

Although we attempt to adjust outdated rDNS CO references, outdated CO mappings remain problematic. To combat some of the stale CO mappings, we remove any adjacencies where each address maps to a CO in a different regional network. Prior knowledge of the extensive use of aggregation in each region, and conversations with network operators, indicated that a small number of entries into each region exist, so we remove likely invalid cross-region adjacencies. This removed far more of the unique CO adjacencies for Comcast than for Charter, likely due to more outdated rDNS in Comcast, although the cross-region CO adjacencies accounted for less than 5% of the IP adjacencies in both networks. We also remove adjacencies representing potential entries from the backbone into each region, where one of the IP addresses map to a backbone PoP, and we infer entries into each region in Section 3.5.2.

Finally, we remove any CO adjacencies that only appear once in the traceroute paths. Traceroute output occasionally contains anomalous output that results from network path changes during the probing. When COs appear interconnected in only one traceroute path, we conclude that the apparent interconnection might result from anomalous traceroute output, so we remove them. This removed 55 CO adjacencies for Comcast, and 9 for Charter.

A.2.3 Refining Region Graphs

After removing likely invalid adjacencies, we use the remaining adjacencies to create graphs of each regional network, with a directed edge from one CO to another corresponding to each CO adjacency. Access networks in the US generally use a star topology to connect EdgeCOs, so we attempt to conform our revealed CO topology to a star topology. The revealed topologies still contain noise, primarily in the form of misleading rDNS creating false adjacencies between EdgeCOs, as well as unrevealed CO interconnections. Our goal is to modify the graphs to conform to the likely physical star topology with as few modifications as possible.

First, we infer the cores of the stars; i.e., the AggCOs in each region. We expect that AggCOs should have more outgoing edges than other COs in the region, despite false CO adjacencies and some EdgeCOs actually connected only to another EdgeCO. To separate likely AggCOs from EdgeCOs we consider any CO with more than the mean outgoing edges plus one standard deviation a AggCO.

We then enforce the role of the AggCOs by removing any (x, y) edge from one EdgeCO to another EdgeCO, unless x has multiple outgoing edges to EdgeCOs that do not interconnect with AggCOs. In general, we expect that edges between EdgeCOs typically result from outdated rDNS, but when a CO appears to aggregate connectivity for multiple COs that otherwise lack connectivity, we conclude that the CO might function as a small AggCO. In total, we removed 26.9% of the unique CO edges in Comcast and 10.6% of the Charter CO edges. The higher fraction of removed Comcast edges reflects prior experience with stale Comcast rDNS.

Next, we infer related AggCOs that connect to the same set of EdgeCOs. Networks often connect an EdgeCO to two AggCOs to increase resiliency to AggCO failure, and we expect that two AggCOs that connect to the same EdgeCO typically connect to the same set of EdgeCOs, since access networks use bundled fiber rings to connect AggCOs to many EdgeCOs. We evaluate each combination of AggCO pairs in the same region, concluding a relationship between the two AggCOs AGG_x and AGG_y if at least 3/4 of the EdgeCOs connected to AGG_X

overlap with EdgeCOs connected to AGG_Y, and the overlap accounts for at least half of the EdgeCOs connected to AGG_Y. The overlap requirements help ensure that we only pair AggCOs with substantial downstream EdgeCO overlap. We also pair two AggCOs if one AggCO has 3/4 overlap with the other AggCO, and neither AggCO would otherwise have a relationship. To reflect the fact that EdgeCOs connect with fiber rings, we add edges to ensure that that all related AggCOs connect to the same EdgeCOs in the regional network graphs. This added 7.8% new edges to Comcast, and 6.1% new edges to Charter.

A.2.4 Redundant AggCO Connections

We inferred that 11.4% and 37.7% of the EdgeCOs in Comcast and Charter connected to a one other CO, respectively, but we never observed any CO-level redundancy for the Charter regional network in the southeastern US. This region is the only large regional network in Comcast or Charter where we did not observe any CO-level redundancy, suggesting we inferred an incomplete CO topology for the region. Excluding the southeast, 29.0% of the Charter EdgeCOs connect to a single upstream CO. Furthermore, of the EdgeCOs connected to one other CO, 33.7% of the Comcast COs and 42.2% of the Charter COs connect to another EdgeCO (not AggCO). Considering only the EdgeCOs connected to an AggCO, and excluding the Charter southeast region, 10.5% of the Comcast EdgeCOs and 18.4% of the Charter EdgeCOs connect to a single AggCO.

A.3 Details about AT&T Mapping

AT&T's regional network routers do not use rDNS names, so we cannot extend the DNS-based geolocation method (Section 3.5) to cluster AT&T's IP addresses into physical facilities. Additional visibility challenges arise from operational practices such as MPLS tunneling and ICMP filtering, which can both hide physical router topology from external traceroutes. These challenges make it critical to have a sufficiently large and strategically selected set of targets.

	Address	rDNS	reply-ttl
1	192.168.1.254		64
2	107.210.168.1	107-210-168-1.lightspeed .sndgca.sbcglobal.net	63
3	71.157.16.42		59
4	108.89.115.1	108-89-115-1.lightspeed .sndgca.sbcglobal.net	61

(a) Intra-region probing traceroute result. From a VP in San Diego, CA probe to a *lightspeed gateway (lspgw)* in the same city. The third hop is the IP of an EdgeCO router.

	Address	rDNS	reply-ttl
1	192.168.1.254		64
2	107.129.92.1	107-129-92-1.lightspeed .sntcca.sbcglobal.net	63
3	71.148.149.186		62
4	71.145.1.52		61
5	12.83.39.213		251
6	12.123.215.237		55
7	71.157.16.42		55
8	108.89.115.1	108-89-115-1.lightspeed .sndgca.sbcglobal.net	54

(b) Inter-region probing traceroute result. From a VP in Santa Cruz, CA to the same *lspgw* in San Diego, CA. The path first traversed COs in Santa Cruz region (hops 3-5), then AT&T's backbone network (hops 6-7), and finally San Diego region (hops 8-9).

Figure A.2. Traceroute examples of regional probing of AT&T.

Target selection.

To find responsive destinations with known geographic locations, we extracted location hints from rDNS names of the IP-DSLAMs connected to end-user modems (denoted as *lspgw*). From our pilot tests using Ark and RIPE Atlas, we found that AT&T encoded the rDNS names of *lspgws* with the regular expression *([\d-]+-1).lightspeed.([a-z]{6}).sbcglobal.net*, where the first part of the name is the dashed decimal notation of the corresponding IP address and the second part is a CLLI code-like 6-character string that represents the city and the state. For example, `sndgca` and `nsvltn` denoted San Diego, CA, and Nashville, TN, respectively. We denoted each unique combination as a *region*.[1] To obtain a comprehensive list of *lspgws*, we used Rapid7's rDNS dataset [138], which periodically resolves rDNS names of the entire IPv4 address space, to find hostnames (and IPs) that matched the regex. We found 95,821 IPs in 37 regions in the September 2020 dataset.

AT&T blocked traceroute measurements toward most of the *lspgws* from the public Internet, but allows traceroutes from within a region and from nearby regions. We used four CAIDA Ark VPs in and nearby San Diego in AT&T to conduct ICMP paris-traceroutes to *lspgws* IPs. This process partially revealed the topology that connected EdgeCOs and AggCOs in a region.

To observe the rest of the topology, we needed to expose MPLS tunnels between the BackboneCO and the *lspgws* that hide the AggCOs and many EdgeCOs. To expose these tunnels we needed to discover which IP prefixes are assigned to the EdgeCO routers in the region we are mapping. We used both *intra-* (McTraceroute) and *inter*-region (Ark) traceroutes to *lspgws* to discover these prefixes. Figure A.2a and Figure A.2b show samples of intra- and inter-region probing to a *lspgw* in San Diego from a RIPE Atlas VP in San Diego, CA and an Ark VP in Santa Cruz respectively. The San Diego VP reaches *lspgws* in the same region directly without crossing the backbone (Figure A.2a). The traceroute from the Santa Cruz VP traverses AT&T's

[1]Note that If the geolocation hint is stale we generally find some anomaly in the traceroute that reveals its staleness, e.g., a traceroute with backbone IP addresses in between nodes with the same geolocation hint likely involves a stale geolocation hint.

Table A.3. Targeted traceroutes to egress interfaces of MPLS tunnels reveals the paths hidden by the MPLS in intra-region probing (hop 4-5).

	Address	rDNS	reply-ttl
1	192.168.1.254		64
2	107.210.168.1	107-210-168-1.lightspeed .sndgca.sbcglobal.net	63
3	71.157.16.114		62
4	75.20.78.58		61
5	75.20.78.55		60
6	71.157.16.42		59

Table A.4. San Diego AT&T CO prefixes

Central Office type	prefix
Edge CO	71.157.6.0/24 71.148.118.0/24 71.148.71.0/24 71.148.104.0/24 71.148.70.0/24 71.157.16.0/24
Aggregation CO	75.20.78.0/24

backbone network, which uses prefix `12.0.0.0/8`, to reach other regions (Figure A.2b). We then extract a preliminary list router prefixes from hops between two *lspgws* in intra-region probing (i.e., hop 3 in Figure A.2a) and between the backbone and the destination *lspgws* in inter-region probing (i.e., hop 7 in Figure A.2b).

We applied the Direct Path Revelation (DPR) technique [191] to reveal the network paths in MPLS tunnels. We targeted inter- and intra- region traceroute measurements to all of the addresses in the EdgeCO router prefixes we discovered, which correspond to the egress interface of the tunnel (i.e., hop 3 in Figure A.2a and hop 7 in Figure A.2b), which allowed us to discover hidden links in the regional network. Table A.3 shows a sample traceroute within the San Diego region that revealed an additional link (hop 4 and 5 in Table A.3) that was hidden in traceroutes to *lspgws*. Table A.4 shows all the IP prefixes for routers we discovered in AT&T's San Diego region.

Table A.5. Inferred number of AT&T PGWs in each region.

Region	BTH	CNC	VNN	ALN	HST	CHC	AKR	ALP	NYC	ART	GSV
Region Bits	2030	2040	2090	2010	20a0	20b0	2000	2020	2050	2070	2080
MTSO Count	2	5	5	5	5	5	3	6	4	3	3

Table A.6. Inferred number of Verizon PGW in each region.

Backbone Region Name	SEA		SJC		LAX		
Wireless Region Name	RDMEWA	HLBOOR	SNVACA	RCKLCA	LSVKNV	AZUSCA	VISTCA
Region bits in IP addresses	100f:b0	100f:b1	1010:b0	1010:b1	1011:b0	1012:b0	1012:b1
PGW numbers	1	1	2	2	2	2	3
Backbone Region Name	CHI						PHIL
Wireless Region Name	HCHLIL	NWBLWI	SFLDMI	STLSMO	BLTNMN	OMALNE	ESYRNY
Region bits in IP addresses	1008:b0	1008:b1	1009:b1	100a:b0	1014:b1	1014:b1	1002:b1
PGW numbers	2	2	1	1	3	2	1
Backbone Region Name	DEN		DLLSTX			MIA	
Wireless Region Name	AURSCO	WJRDUT	ELSSTX	HSTWTX	BTRHLA	MIAMFL	ORLHFL
Region bits in IP addresses	100e:b0	100e:b1	100c:b2	100d:b0	100d:b1	100b:b0	100b:b1
PGW numbers	2	2	1	2	2	2	2

Backbone Region Name	ATL			IAD		NYC	BOS	
Wireless Region Name	CHR-XNC	WHC-KTN	ALP-SGA	CHN-TVA	JHT-WPA	WLTPNJ	WSB-OMA	BBT-PNJ
Region bits in IP addresses	1004:b0	1004:b1	1005:b0	1003:b0	1003:b1	1017:b0	1000:b0	1000:b1
PGW numbers	4	2	2	2	1	2	2	1

A.4 Details about Mobile Mapping

Target Selection.

We used the AS relationship dataset [27] to identify each mobile ISP's neighboring ASes. We found 266/406/213 neighboring ASes for AT&T/Verizon/T-Mobile, respectively. We then conducted a pilot test to compile lists of target IPs for each ISP. For each neighboring AS, we found one IPv4 and one IPv6 destination that were responsive to traceroute probes. We used the corresponding target list of the current mobile ISP to perform traceroute measurements.

The ShipTraceroute results showed that the network paths to all the targets shared the same paths within the mobile network until exiting the PGWs. Table A.5 and Table A.6 show the number of PGWs we inferred using region bits in AT&T and Verizon IPv6 addresses, respectively.

Bibliography

[1] Fcc's measure broadband america program.

[2] Netacuity. https://www.digitalelement.com/solutions/.

[3] Netalyzr. https://play.google.com/store/apps/details?id=edu.berkeley.icsi.netalyzr.android&hl=en_US.

[4] Speedof.me. https://speedof.me.

[5] Instabridge. https://instabridge.com/en/, 2020.

[6] 3GPP. 5G NR Medium Access Control (MAC) protocol specification. https://www.etsi.org/deliver/etsi_ts/138300_138399/138321/15.03.00_60/ts_138321v150300p.pdf.

[7] 3rd Generation Partnership Project (3GPP). Scheduling. https://www.3gpp.org/technologies/scheduling.

[8] ABC 7. Spectrum restores service to SoCal customers after brief outage. https://abc7.com/spectrum-outage-los-angeles-southern-california/10362980/, 2021.

[9] Giuseppe Aceto, Alessio Botta, Pietro Marchetta, Valerio Persico, and Antonio Pescapé. A comprehensive survey on internet outages. *Journal of Network and Computer Applications*, 2018.

[10] Pankaj K. Agarwal, Alon Efrat, Shashidhara K. Ganjugunte, David Hay, Swaminathan Sankararaman, and Gil Zussman. Network vulnerability to single, multiple, and probabilistic physical attacks. In *MILCOM*, 2010.

[11] Yazeed A Al-Sbou. Wireless networks performance monitoring based on passive-active quality of service measurements. *Int. J. Comput. Networks Commun*, 12(6):14–32, 2020.

[12] Sajjad Amini, Fabio Pasqualetti, and Hamed Mohsenian-Rad. Dynamic load altering attacks against power system stability: Attack models and protection schemes. *IEEE Transactions on Smart Grid*, 2018.

[13] Kamran Arshad. Lte system level performance in the presence of cqi feedback uplink delay and mobility. In *2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15)*, pages 1–5, 2015.

[14] ATT. West ca tandem and subtending end offices. https://tinyurl.com/7pyywwp7, 2008.

[15] AT&T. DRAFT at&t Iperf mobile application user guide. https://pdfslide.net/documents/draft-att-iperf-mobile-application-user-guide-aka-iperf-commands-205pdf.html, 2014.

[16] Vaibhav Bajpai, Steffie Jacob Eravuchira, and Jürgen Schönwölder. Dissecting last-mile latency characteristics. *ACM SIGCOMM Computer Communication Review (CCR)*, 2017.

[17] Arjun Balasingam, Manu Bansal, Rakesh Misra, Kanthi Nagaraj, Rahul Tandra, Sachin Katti, and Aaron Schulman. Detecting if lte is the bottleneck with bursttracker. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[18] Nimantha Baranasuriya, Vishnu Navda, Venkata N. Padmanabhan, and Seth Gilbert. Qprobe: locating the bottleneck in cellular communication. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '15, New York, NY, USA, 2015. Association for Computing Machinery.

[19] Carlos Barreto, Alvaro A. Cárdenas, Nicanor Quijano, and Eduardo Mojica-Nava. CPS: Market analysis of attacks against demand response in the smart grid. In *Annual Computer Security Applications Conference (ACSAC)*, 2014.

[20] Carlos Barreto, Jairo Giraldo, Alvaro A. Cardenas, Eduardo Mojica-Nava, and Nicanor Quijano. Control systems for the power grid and their resiliency to attacks. *IEEE Symposium on Security and Privacy (SP)*, 2014.

[21] Arkadiusz Biernacki and Kurt Tutschku. Comparative performance study of lte downlink schedulers. *Wireless personal communications*, 74:585–599, 2014.

[22] Zachary S. Bischof, John S. Otto, Mario A. Sánchez, John P. Rula, David R. Choffnes, and Fabián E. Bustamante. Crowdsourcing ISP characterization to the network edge. In *Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack - W-MUST*, 2011.

[23] Bastian Bloessl, Lars Baumgärtner, and Matthias Hollick. Hardware-Accelerated Real-Time Stream Data Processing on Android with GNU Radio. In *14th International Workshop on Wireless Network Testbeds, Experimental evaluation and Characterization (WiNTECH'20)*, London, UK, September 2020. ACM.

[24] Bloomer Telephone. Did you know? https://bloomer.net/did-you-know/.

[25] Ilker Nadi Bozkurt, Waqar Aqeel, Debopam Bhattacherjee, Balakrishnan Chandrasekaran, Philip Brighten Godfrey, Gregory Laughlin, Bruce M. Maggs, and Ankit Singla. Dissecting latency in the internet's fiber infrastructure, 2018.

[26] Clint Burgess. Employing spectrum analyzers to ensure reliable cellular coverage, May 2024. Accessed: 2024-09-06.

[27] CAIDA. As-relationships. https://publicdata.caida.org/datasets/as-relationships/.

[28] CAIDA. The CAIDA UCSD IPv4 routed /24 topology dataset. https://www.caida.org/catalog/datasets/ipv4_routed_24_topology_dataset/.

[29] CAIDA. DNS Decoded (DDec). http://ddec.caida.org.

[30] CAIDA. Archipelago (Ark) measurement infrastructure. https://www.caida.org/projects/ark/, 2007.

[31] CalEPA. Calepa regulated site portal. https://siteportal.calepa.ca.gov/nsite/.

[32] Igor Canadi, Paul Barford, and Joel Sommers. Revisiting broadband performance. In *ACM Internet Measurement Conference (IMC)*, 2012.

[33] Cong Cao, Moshe Zukerman, Weiwei Wu, Jonathan H. Manton, and Bill Moran. Survivable topology design of submarine networks. *Journal of Lightwave Technology*, 2013.

[34] Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016.

[35] Abhijnan Chakraborty, Vishnu Navda, Venkata N. Padmanabhan, and Ramachandran Ramjee. Coordinating cellular background transfers using LoadSense. In *ACM Conference on Mobile Computing and Networking (MobiCom)*, 2013.

[36] Stephen Checkoway, Damon McCoy, Brian Kantor, Danny Anderson, Hovav Shacham, Stefan Savage, Karl Koscher, Alexei Czeskis, Franziska Roesner, and Tadayoshi Kohno. Comprehensive experimental analyses of automotive attack surfaces. In *USENIX Security Symposium*, 2011.

[37] Anton Cherepanov. Win32/industroyer: A new threat for industrial control systems. *White paper, ESET*, 2017.

[38] Cisco. Cisco uBR-MC3GX60V broadband processing engine with full DOCSIS 3.0 support for the cisco uBR10012 universal broadband router. https://www.cisco.com/c/en/us/products/collateral/video/ubr10000-series-universal-broadband-routers/data_sheet_c78-642540.html.

[39] Cisco. Introduction to DWDM technology. https://www.cisco.com/c/dam/global/de_at/assets/docs/dwdm.pdf, 2000.

[40] Cisco. Lte design and deployment strategies. https://www.cisco.com/c/dam/global/en_ae/assets/expo2011/saudiarabia/pdfs/lte-design-and-deployment-strategies-zeljko-savic.pdf, 2011.

[41] David D. Clark. The design philosophy of the DARPA internet protocols. In *ACM SIGCOMM*, 1988.

[42] CloudFlare. Cloudflare speed test. https://speed.cloudflare.com.

[43] Comcast. Comcast maintenance notifications for non-service affecting maintenance activities associated with ethernet transport services and ethernet dedicated internet services. https://business.comcast.com/terms-conditions-ent/maintenance.

[44] Comcast. Xfinity speed test. http://speedtest.xfinity.com.

[45] Connecticut General Assembly. Undergrounding electric lines. https://www.cga.ct.gov/2011/rpt/2011-R-0338.htm.

[46] Critter Guard. Pros and cons of underground fiber optic cable. https://www.critterguard.org/blogs/articles/pros-and-cons-of-underground-fiber-optic-cable.

[47] Crown Infrastructure Partners. Ultra fast broadband. https://www.crowninfrastructure.govt.nz/ufb/what/, 2021.

[48] Alberto Dainotti, Roman Amman, Emile Aben, and Kimberly C. Claffy. Extracting benefit from harm: using malware pollution to analyze the impact of political and geophysical events on the internet. *ACM SIGCOMM Computer Communication Review (CCR)*, 2012.

[49] Mike Dano. An inside look at verizon's edge computing capabilities. https://www.lightreading.com/the-edge/an-inside-look-at-verizons-edge-computing-capabilities, 2019.

[50] Pankaz Das, Mahshid Rahnamay-Naeini, Nasir Ghani, , and Majeed M. Hayat. On the vulnerability of multi-level communication network under catastrophic events. In *IEEE International Conference on Computing, Networking and Communications*, 2017.

[51] Marcel Dischinger, Andreas Haeberlen, Krishna P. Gummadi, and Stefan Saroiu. Characterizing residential broadband networks. In *ACM Internet Measurement Conference (IMC)*, 2007.

[52] Dittrich, David and Kenneally, Erin and Bailey, Michael. Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report, 2013. http://ssrn.com/abstract=2342036.

[53] Wei Dong, Zihui Ge, and Seungjoon Lee. 3G meets the Internet: Understanding the performance of hierarchical routing in 3G networks. In *Proceedings of International Teletraffic Congress*, 2011.

[54] Ramakrishnan Durairajan, Paul Barford, Joel Sommers, and Walter Willinger. Intertubes: A study of the us long-haul fiber-optic infrastructure. In *ACM SIGCOMM*, 2015.

[55] Navid Ehsan, Mingyan Liu, and Roderick J. Ragland. Evaluation of performance enhancing proxies in internet over satellite. *International Journal of Communication Systems*, 16, May 2003.

[56] European Telecommunications Standards Institute. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 12.3.0 Release 12). Technical Specification TS 136 213 V12.3.0, ETSI, Sophia Antipolis, France, October 2014. Available: https://www.etsi.org/.

[57] Fast.com. Internet speed test. https://fast.com.

[58] FCC. Katrina panel. https://www.fcc.gov/katrina-panel, 2007.

[59] FCC. June 15, 2020 t-mobile network outage report, 2020.

[60] Romain Fontugne, Anant Shah, and Kenjiro Cho. Persistent last-mile congestion: Not so uncommon. In *ACM Internet Measurement Conference (IMC)*, 2020.

[61] Galooli. What is a base station?, 2024. Accessed: 2024-09-02.

[62] John A Gamon, AF Rahman, JL Dungan, M Schildhauer, and KF Huemmrich. Spectral network (specnet)—what is it and why do we need it? *Remote Sensing of Environment*, 103(3):227–235, 2006.

[63] SF Gate. $250,000 reward in phone cable vandalism. https://www.sfgate.com/bayarea/article/250-000-reward-in-phone-cable-vandalism-3245341.php, 2009.

[64] GeoTel Communications, LLC. GeoTel Communications: Advanced Fiber Maps and Telecom Data, 2025. Accessed: 2025-03-13.

[65] Monia Ghobadi and Ratul Mahajan. Optical layer failures in a large backbone. In *ACM Internet Measurement Conference (IMC)*, 2016.

[66] Joey Gill and Sebastian Posey. AT&T outages across Tennessee, Kentucky affecting multiple 911 services. https://www.wkrn.com/news/local-news/att-outages-across-tennessee-kentucky-affecting-multiple-911-services/, 2020.

[67] Utkarsh Goel, Ajay Miyyapuram, Mike P Wittie, and Qing Yang. Mitate: Mobile internet testbed for application traffic experimentation. In *Mobile and Ubiquitous Systems: Computing, Networking, and Services: 10th International Conference, MOBIQUITOUS 2013, Tokyo, Japan, December 2-4, 2013, Revised Selected Papers 10*, pages 224–236. Springer, 2014.

[68] Omer Gold and Reuven Cohen. Coping with physical attacks on random network structures. In *IEEE International Conference on Communications (ICC)*, 2014.

[69] Ramesh Govindan and Hongsuda Tangmunarunkit. Heuristics for Internet map discovery. In *IEEE Conference on Computer Communications (INFOCOM)*, 2000.

[70] Matthias Gunkel, Malte Schneiders, Sascha Vorbeck, Werner Weiershausen, Ralph Leppla, Frank Rumpf, Ralf Herber, Volker Furst, and Markus Rodenfels. Aggregation networks: Cost comparison of WDM ring vs. double star topology. In *ONDM*, 2008.

[71] Varun Gupta, Craig Gutterman, Yigal Bejerano, and Gil Zussman. Experimental evaluation of large scale wifi multicast rate control. *IEEE Transactions on Wireless Communications*, 17(4):2319–2332, 2018.

[72] Mehrdad Hessar, Ali Najafi, Vikram Iyer, and Shyamnath Gollakota. {TinySDR}:{Low-Power}{SDR} platform for {Over-the-Air} programmable {IoT} testbeds. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pages 1031–1046, 2020.

[73] Ningning Hou, Xianjin Xia, and Yuanqing Zheng. Don't miss weak packets: Boosting lora reception with antenna diversities. *ACM Transactions on Sensor Networks*, 19(2):1–25, 2023.

[74] Jiyao Hu, Zhenyu Zhou, Xiaowei Yang, Jacob Malone, and Jonathan W Williams. CableMon: Improving the reliabilityof cable broadband networksvia proactive network maintenance. In *Symposium on Networked Systems Design and Implementation (NSDI)*, 2020.

[75] Wenlu Hu, Ying Gao, Kiryong Ha, Junjue Wang, Brandon Amos, Zhuo Chen, Padmanabhan Pillai, and Mahadev Satyanarayanan. Quantifying the impact of edge computing on mobile applications. In *Proceedings of the 7th ACM SIGOPS Asia-Pacific Workshop on Systems*, 2016.

[76] Bing Huang, Alvaro A. Cardenas, and Ross Baldick. Not everything is dark and gloomy: Power grid protections against IoT demand attacks. In *USENIX Security Symposium*, 2019.

[77] Thomas Huehn and Cigdem Sengul. Practical power and rate control for wifi. In *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, pages 1–7. IEEE, 2012.

[78] B. Huffaker, M. Fomenkov, and k. claffy. DRoP:DNS-based Router Positioning. *ACM SIGCOMM Computer Communication Review (CCR)*, 44:6–13, 2014.

[79] The MathWorks Inc. Matlab version: 9.13.0 (r2022b), 2022.

[80] Tive Inc. Tive releases return-by-mail supply chain tracker. https://blog.tive.co/tive-releases-return-by-mail-supply-chain-tracker, 2019.

[81] Internet Censorship Lab. Internet Censorship Lab, 2024. Accessed: 2024-02-23.

[82] Santosh Janardhan. More details about the october 4 2021 outage, 2021.

[83] Juniper Networks. Metro ethernet design guide. https://www.juniper.net/documentation/en_US/release-independent/solutions/information-products/pathway-pages/solutions/metro-ethernet-dg.pdf, 2016.

[84] Sangeetha Abdu Jyothi. Solar superstorms: Planning for an internet apocalypse. In *ACM SIGCOMM*, 2021.

[85] Kenneally, Erin and Dittrich, David. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research, 2012. http://ssrn.com/abstract=2445102.

[86] Ken Keys. Periscope Looking Glass API, 2019. https://www.caida.org/tools/utilities/looking-glass-api/.

[87] Ken Keys, Young Hyun, Matthew Luckie, and k claffy. Internet-scale IPv4 alias resolution with MIDAR. *IEEE/ACM Transactions on Networking*, 21, 2013.

[88] Keysight Technologies. Fieldfox handheld rf and microwave analyzers. https://www.keysight.com/us/en/products/network-analyzers/fieldfox-handheld-rf-microwave-analyzers.html, 2024. Accessed: 2024-09-05.

[89] Simon Knight, Hung X. Nguyen, Nickolas Falkner, Rhys Bowden, and Matthew Roughan. The internet topology zoo. *IEEE Journal on Selected Areas in Communications (J-SAC)*, 2011.

[90] Christian Kreibich, Nicholas Weaver, Boris Nechaev, and Vern Paxson. Netalyzr. In *ACM Internet Measurement Conference (IMC)*, 2010.

[91] Kimberlee Kruesi, Michael Balsamo, and Eric Tucker. FBI at home of possible person of interest in Nashville bomb. https://apnews.com/article/us-news-nashville-coronavirus-pandemic-tennessee-dc6eb653053967a4187f0ca8276d20, 2020.

[92] Eric Kuhnke. Nashville. https://mailman.nanog.org/pipermail/nanog/2020-December/211081.html, 2020.

[93] Chenning Li, Hanqing Guo, Shuai Tong, Xiao Zeng, Zhichao Cao, Mi Zhang, Qiben Yan, Li Xiao, Jiliang Wang, and Yunhao Liu. Nelora: Towards ultra-low snr lora communication with neural-enhanced demodulation. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 56–68, 2021.

[94] Yilong Li, Yijing Zeng, and Suman Banerjee. Enabling wideband, mobile spectrum sensing through onboard heterogeneous computing. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 85–91, 2021.

[95] Aemen Lodhi, Natalie Larson, Amogh Dhamdhere, Constantine Dovrolis, and kc claffy. Using peeringdb to understand the peering ecosystem. *SIGCOMM Comput. Commun. Rev.*, 2014.

[96] Laura Lollman. Power pole arson causes major internet outage for cox customers in north phoenix. https://web.archive.org/web/20210119053739/https://www.azfamily.com/news/power-pole-arson-causes-major-internet-outage-for-cox-customers-in-north-phoenix/article_9b699a52-fe73-11ea-b6b6-a74c606e5175.html, 2020.

[97] Feng Lu, Hao Du, Ankur Jain, Geoffrey M. Voelker, Alex C. Snoeren, and Andreas Terzis. CQIC: Revisiting cross-layer congestion control for cellular networks. 2015.

[98] Feng Lu, Hao Du, Ankur Jain, Geoffrey M. Voelker, Alex C. Snoeren, and Andreas Terzis. Cqic: Revisiting cross-layer congestion control for cellular networks. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, page 45–50, New York, NY, USA, 2015. Association for Computing Machinery.

[99] M. Luckie. Scamper: a scalable and extensible packet prober for active measurement of the internet. In *Proc. ACM IMC*, 2010.

[100] Matthew Luckie, Bradley Huffaker, and k claffy. Learning regexes to extract router names from hostnames. In *ACM Internet Measurement Conference (IMC)*, 2019.

[101] Jenry Luis and Cesar A Santivanez. Assessing the impact of field-measurement on the design of spectrum sensing wsn. In *2023 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6. IEEE, 2023.

[102] M-Lab. BigQuery QuickStart. https://www.measurementlab.net/data/docs/bq/quickstart/.

[103] M-Lab. M-Lab naming service. http://mlab-ns.appspot.com/.

[104] Xiaoqiang Ma, Yuan Zhao, Lei Zhang, Haiyang Wang, and Limei Peng. When mobile terminals meet the cloud: computation offloading as the bridge. *IEEE Network*, pages 28–33, 2013.

[105] Helka-Liina Maattanen, Toni Huovinen, Tommi Koivisto, Mihai Enescu, Olav Tirkkonen, and Mikko Valkama. Performance evaluations for multiuser cqi enhancements for lte-advanced. In *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2011.

[106] MagicSDR. Magicsdr, 2024. Accessed: 2024-09-02.

[107] Ratul Mahajan, Neil Spring, David Wetherall, and Tom Anderson. Inferring link weights using end-to-end measurements. In *ACM Internet Measurement Workshop (IMW)*, 2002.

[108] Ratul Mahajan, David Wetherall, and Thomas Anderson. Understanding BGP misconfiguration. In *ACM SIGCOMM*, 2002.

[109] MAIA-SDR. Maia-sdr, 2024. Accessed: 2024-09-02.

[110] Sathiya Kumaran Mani, Matthew Nance Hall, Ramakrishnan Durairajan, and Paul Barford. Characteristics of metro fiber deployments in the US. In *Network Traffic Measurement and Analysis Conference (TMA)*, 2020.

[111] Yun Mao, Hani Jamjoom, Shu Tao, and Jonathan M. Smith. NetworkMD: Topology inference and failure diagnosis in the last mile. In *ACM Internet Measurement Conference (IMC)*, 2007.

[112] MathWorks. Cell search, mib and sib1 recovery, 2024. Accessed: 2024-09-02.

[113] Franco Minucci, Dieter Verbruggen, Hazem Sallouha, Vladimir Volski, Guy Vandenbosch, Gérôme Bovet, and Sofie Pollin. Measuring 5g electric fields strength with software defined radios. *IEEE Open Journal of the Communications Society*, 3:2258–2271, 2022.

[114] Saleh Soltan Prateek Mittal and H. Vincent Poor. BlackIoT: IoT botnet of high wattage devicescan disrupt the power grid. In *USENIX Security Symposium*, 2018.

[115] Nitinder Mohan, Lorenzo Corneo, Aleksandr Zavodovski, Suzan Bayhan, Walter Wong, and Jussi Kangasharju. Pruning edge research with latency shears. pages 182–189, 2020.

[116] Susanna Mosleh, Yao Ma, Jacob D Rezac, and Jason B Coder. Dynamic spectrum access with reinforcement learning for unlicensed access in 5g and beyond. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–7. IEEE, 2020.

[117] Edward D. Murphy and Dennis Hoey. Spectrum internet outage hits thousands of customers in Maine and New Hampshire. https://www.pressherald.com/2021/04/05/spectrum-restores-internet-service-after-outage-affects-thousands-across-maine/, 2021.

[118] Kate Murphy. The cyberthreat under the street. https://www.nytimes.com/2015/11/08/sunday-review/the-cyberthreat-under-the-street.html, 2015.

[119] David Murray, Terry Koziniec, Michael Dixon, and Kevin Lee. Measuring the reliability of 802.11 wifi networks. In *2015 Internet Technologies and Applications (ITA)*, pages 233–238. IEEE, 2015.

[120] N. Spring and R. Mahajan and D. Wetherall. Measuring ISP topologies with Rocketfuel. In *ACM SIGCOMM*, 2002.

[121] NBN Co. Nbn multi technology mix. https://www.nbnco.com.au/learn/network-technology, 2021.

[122] GSMA Future Network. Cloud ar/vr whitepaper, 2019.

[123] Sebastian Neumayer and Eytan Modiano. Network reliability under random circular cuts. In *IEEE Global Communications Conference (GLOBECOM)*, 2011.

[124] Ashkan Nikravesh, Hongyi Yao, Shichang Xu, David Choffnes, and Z. Morley Mao. Mobilyzer: An open platform for controllable mobile network measurements. In *ACM Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2015.

[125] City of Port St Lucie. Fiber optic network. https://utility.cityofpsl.com/media/1097/fiber-protocol.pdf.

[126] US Department of Transportation. 49 cfr 172.185 – lithium cells and batteries. https://www.law.cornell.edu/cfr/text/49/173.185.

[127] Ookla. At&t network outage in dec 18th, 2020. https://downdetector.com/status/att/news/356027-problems-at-att/.

[128] Ookla. Speedtest. http://www.speedtest.net.

[129] Ookla. How to install & submit server. https://support.ookla.com/hc/en-us/articles/234578568-How-To-Install-Submit-Server, June 2022.

[130] Jorik Oostenbrink and Fernando Kuipers. The risk of successive disasters: A blow-by-blow network vulnerability analysis. In *IFIP Networking*, 2019.

[131] Ramakrishna Padmanabhan, Aaron Schulman, Alberto Dainotti, Dave Levin, and Neil Spring. How to find correlated internet failures. In *Passive and Active Network Measurement Conference (PAM)*, 2019.

[132] Ramakrishna Padmanabhan, Aaron Schulman, Dave Levin, and Neil Spring. Residential links under the weather. In *ACM SIGCOMM*, 2019.

[133] Sourav Pal, Sumantra R Kundu, Kalyan Basu, and Sajal K Das. Ieee 802.11 rate control algorithms: Experimentation and performance evaluation in infrastructure mode. In *Passive and Active Measurement Conference*. Citeseer, 2006.

[134] Carla Parra, Edison Tatayo, Alejandro Paccha, Christian Tipantuna, and Jorge Carvajal. Sdr-based portable open-source gsm/gprs network for emergency scenarios. In *2019 Sixth International Conference on EDemocracy & EGovernment (ICEDEG)*, pages 268–273. IEEE, 2019.

[135] Larry Peterson, Tom Anderson, Sachin Katti, Nick McKeown, Guru Parulkar, Jennifer Rexford, Mahadev Satyanarayanan, Oguz Sunay, and Amin Vahdat. Democratizing the network edge. *ACM SIGCOMM Computer Communication Review (CCR)*, 2019.

[136] Qualcomm. *eXtensible Diagnostic Monitor*.

[137] Lin Quan, John Heidemann, and Yuri Pradkin. Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *ACM SIGCOMM*, 2013.

[138] Rapid7 Labs. Reverse DNS (RDNS). https://opendata.rapid7.com/sonar.rdns_v2/, 2021.

[139] RevelareNet. Home page. https://www.revelarenet.com/. Accessed: [insert date of access].

[140] Philipp Richter, Ramakrishna Padmanabhan, Neil Spring, Arthur Berger, and David Clark. Advancing the art of internet edge outage detection. In *ACM Internet Measurement Conference (IMC)*, 2018.

[141] RIPE. Ripe atlas. https://atlas.ripe.net/landing/measurements-and-tools/, 2019.

[142] Pieter Robyns, Peter Quax, Wim Lamotte, and William Thenaers. gr-lora: An efficient lora decoder for gnu radio. *Zenodo Ed*, 10:5281, 2017.

[143] Rick Rojas, Jamie McGee, Edmund Lee, and Steve Cavendish. When Nashville bombing hit a telecom hub, the ripples reached far beyond. https://www.nytimes.com/2020/12/29/us/nashville-bombing-telecommunications.html, 2020.

[144] John P. Rula, Zachary S. Bischof, and Fabian E. Bustamante. Second chance: Understanding diversity in broadband access network performance. In *Proc. of the ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, 2015.

[145] John P Rula and Fabian E Bustamante. Behind the curtain: Cellular DNS and content replica selection. In *ACM Internet Measurement Conference (IMC)*, 2014.

[146] Rutgers. *Correlation of Discrete-Time Signals*. https://eceweb1.rutgers.edu/~gajic/solmanual/slides/chapter9_CORR.pdf, September 2024.

[147] Samsung. Keep your galaxy device at its normal operating temperature, 2021.

[148] Mario A. Sánchez, John S. Otto, Zachary S. Bischof, and David R. Choffnes. Dasu: Pushing experiments to the Internet's edge. In *Symposium on Networked Systems Design and Implementation (NSDI)*, 2013.

[149] Sindhura Sarepalli. *LTE Downlink Scheduling Algorithms*. PhD thesis, 2016. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-07-21.

[150] Mahadev Satyanarayanan. The emergence of edge computing. *Computer*, 50:30–39, 2017.

[151] Mahadev Satyanarayanan. The emergence of edge computing. *Computer*, 50:30–39, 2017.

[152] Southwestern Bell Company (SBC). Rates and tariffs quote sheet—OC3c purchased under the OC-n point to point service offering. https://tinyurl.com/uwh3puzj, 2005.

[153] Brandon Schlinker, Kyriakos Zarifis, Italo Cunha, Nick Feamster, and Ethan Katz-Bassett. Peering: An as for us. In *Proceedings of the 13th ACM Workshop on Hot Topics in Networks*, pages 1–7, 2014.

[154] Paul Schmitt, Morgan Vigil, and Elizabeth Belding. A study of MVNO data paths and performance. In *Passive and Active Network Measurement Conference (PAM)*, 2016.

[155] Aaron Schulman and Neil Spring. Pingin' in the rain. In *ACM Internet Measurement Conference (IMC)*, 2011.

[156] Semtech. *WHY LoRaWAN IS THE CONNECTIVITY PLATFORM FOR SMART CITY APPLICATIONS*. https://lora-alliance.org/wp-content/uploads/2020/11/LA_WhitePaper_SmartCities_0520_v1.pdf, September 2024.

[157] Muhammad Osama Shahid, Millan Philipose, Krishna Chintalapudi, Suman Banerjee, and Bhuvana Krishnaswamy. Concurrent interference cancellation: Decoding multi-packet collisions in lora. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, pages 503–515, 2021.

[158] ShareTechnote. Sib scheduling. https://www.sharetechnote.com/html/BasicProcedure_LTE_SIB_Scheduling.html, 2024.

[159] Rob Sherwood, Adam Bender, and Neil Spring. DisCarte: A disjunctive Internet cartographer. In *ACM SIGCOMM*, 2008.

[160] Tony Shin and Monica Garske. AT&T cables vandalized, $250,000 reward offered for information. https://www.nbcsandiego.com/news/local/att-cables-cut-vandalized-250000-reward-offered/1924959/, 2012.

[161] Phillip Smith, Anh Luong, Shamik Sarkar, Harsimran Singh, Aarti Singh, Neal Patwari, Sneha Kasera, and Kurt Derr. A novel software defined radio for practical, mobile crowdsourced spectrum sensing. *IEEE Transactions on Mobile Computing*, 22(3):1289–1300, 2021.

[162] Solano County Sheriff's Office. Comcast outage. https://www.facebook.com/SolanoSheriff/posts/comcast-has-reported-an-outage-affecting-1016-users-in-fairfield-94533-and-rio-v/2131795846957800/, 2021.

[163] Spectrum. Spectrum maintenance update. https://www.spectrum.net/support/tv/spectrum-tv-maintenance-update.

[164] Neil Spring. Scriptroute: A public internet measurement facility. In *4th USENIX Symposium on Internet Technologies and Systems (USITS 03)*, 2003.

[165] Neil Spring, Ratul Mahajan, and Thomas Anderson. The causes of path inflation. In *ACM SIGCOMM*, 2003.

[166] Neil Spring, Ratul Mahajan, David Wetherall, and Thomas Anderson. Measuring ISP topologies with rocketfuel. In *ACM SIGCOMM*, 2002.

[167] Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyber-physical system security for the electric power grid. *Proceedings of the IEEE*, 2012.

[168] Brian Stelter, Kay Jones, and Hollie Silverman. AT&T working to restore outages after Nashville explosion. https://www.cnn.com/2020/12/25/us/nashville-explosion-service-disruptions/index.html, 2020.

[169] Srikanth Sundaresan, Sam Burnett, Nick Feamster, and Walter de Donato. Bismark: A testbed for deploying measurements and applications in broadband access networks. In *USENIX*, 2014.

[170] Srikanth Sundaresan, Walter de Donato, Nick Feamster, Renata Teixeira, Sam Crawford, and Antonio Pescapè. Measuring home broadband performance. *Commun. ACM*, 2012.

[171] Srikanth Sundaresan, Nick Feamster, and Renata Teixeira. Home network or access link? Locating last-mile downstream throughput bottlenecks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[172] Srikanth Sundaresan, Renata Teixeira, Georgia Tech, Nick Feamster, Antonio Pescapè, and Sam Crawford. Broadband Internet Performance : A View From the Gateway. *ACM SIGCOMM Computer Communication Review (CCR)*, 2011.

[173] Hirotaka Suzuki, Yuusuke Kawakita, and Haruhisa Ichikawa. Remote implementation of gnu radio-based sdr development environment. In *2016 22nd Asia-Pacific Conference on Communications (APCC)*, pages 355–360. IEEE, 2016.

[174] Dean Takahashi. Comcast can now pinpoint fiber optic cable breaks within minutes. https://venturebeat.com/2021/07/22/comcast-can-now-pinpoint-fiber-optic-cable-breaks-within-minutes/, 2021.

[175] NetMonster Team. Netmonster – advanced signal discovery. https://netmonster.app/, 2024.

[176] Tech LTE World. Lte mac scheduler. https://techlteworld.com/lte-mac-scheduler.

[177] Renata Teixeira, Keith Marzullo, Stefan Savage, and Geoffrey M. Voelker. In search of path diversity in ISP networks. In *ACM Internet Measurement Conference (IMC)*, 2003.

[178] Chee-Wooi Ten, Chen-Ching Liu, and Govindarasu Manimaran. Vulnerability assessment of cybersecurity for SCADA systems. *IEEE Transactions on Power Systems*, 2008.

[179] The County of San Diego. Citizen access portal. https://publicservices.sandiegocounty.gov/CitizenAccess/Default.aspx.

[180] The GNU Radio Foundation, Inc. The signal metadata format (sigmf), 2018. License: CC-BY-SA-4.0.

[181] The State of Florida, Department of Environmental Protection. Storage tanks and contamination monitoring. https://prodlamp.dep.state.fl.us/www_stcm/reports/DorFacilities, 2021.

[182] Ye Tian, Ratan Dey, Yong Liu, and Keith W Ross. Topology mapping and geolocating for china's internet. *IEEE Transactions on Parallel and Distributed Systems*, 24:1908–1917, 2012.

[183] Hoang Tran. Case study: Ethernet cell site backhaul request for quotation process, 2012.

[184] Phuong Nga Tran and Hiroshi Saito. Enhancing physical network robustness against earthquake disasters with additional links. *Journal of Lightwave Technology*, 2016.

[185] Daniel Turner, Kirill Levchenko, Alex C. Snoeren, and Stefan Savage. California fault lines: Understanding the causes and impact of network failures. In *ACM SIGCOMM*, 2010.

[186] Dalibor Uhlir, Dominik Kovac, and Jiri Hosek. Multi service proxy: Mobile web traffic entitlement point in 4g core network. *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, 4, 2015.

[187] Unwired Labs. Opencellid - largest open database of cell towers & geolocation. https://opencellid.org/, 2024. Accessed: 2024-09-06.

[188] Narseo Vallina-Rodriguez, Andrius Auçinas, Mario Almeida, Yan Grunenberger, Konstantina Papagiannaki, and Jon Crowcroft. Rilanalyzer: a comprehensive 3g monitor on your phone. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, page 257–264, New York, NY, USA, 2013. Association for Computing Machinery.

[189] Bertold Van den Bergh, Domenico Giustiniano, Héctor Cordobés, Markus Fuchs, Roberto Calvo-Palomino, Sofie Pollin, Sreeraj Rajendran, and Vincent Lenders. Electrosense: Crowdsourcing spectrum monitoring. In *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pages 1–2, 2017.

[190] Yves Vanaubel, Pascal Mérindol, Jean-Jacques Pansiot, and Benoit Donnet. Mpls under the microscope: Revealing actual transit path diversity. In *ACM Internet Measurement Conference (IMC)*, 2015.

[191] Yves Vanaubel, Pascal Mérindol, Jean-Jacques Pansiot, and Benoit Donnet. Through the wormhole: Tracking invisible mpls tunnels. In *ACM Internet Measurement Conference (IMC)*, pages 29–42, 2017.

[192] Hai Wang, Su Xie, Ke Li, and M Omair Ahmad. Big data-driven cellular information detection and coverage identification. *Sensors*, 19(4):937, 2019.

[193] Elizabeth Clement Webb. FBI releases report on Nashville bombing. https://www.fbi.gov/contact-us/field-offices/memphis/news/press-releases/fbi-releases-report-on-nashville-bombing, 2021.

[194] WebSDR. Websdr, 2024. Accessed: 2024-09-02.

[195] David Wells. Vandalism blamed for 13-hour Comcast outage in SLC area. https://www.fox13now.com/news/local-news/vandalism-blamed-for-13-hour-comcast-outage-in-slc-area, 2021.

[196] Chris Whitaker. The Comcast enterprise network story. https://www.slideshare.net/cwhita002/the-comcast-enterprise-network-story, 2011.

[197] Wikipedia. Arcgis, 2024. Accessed: 2024-09-02.

[198] Wikipedia. *Pearson correlation coefficient*. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient, September 2024.

[199] Wikipedia contributors. WiGLE, 2025. Accessed: 2025-03-13.

[200] Walter Willinger, David Alderson, and John C. Doyle. Mathematics and the Internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society*, 2009.

[201] Daoyuan Wu, Rocky Chang, Weichao Li, Eric Cheng, and Debin Gao. Mopeye: Opportunistic monitoring of per-app mobile network performance. In *Proceedings of USENIX ATC*, 2017.

[202] Xianjin Xia, Qianwu Chen, Ningning Hou, Yuanqing Zheng, and Mo Li. Xcopy: Boosting weak links for reliable lora communication. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[203] Xiufeng Xie, Xinyu Zhang, Swarun Kumar, and Li Erran Li. piStream: Physical layer informed adaptive video streaming over lte. In *ACM Conference on Mobile Computing and Networking (MobiCom)*, 2015.

[204] Xiufeng Xie, Xinyu Zhang, and Shilin Zhu. Accelerating mobile web loading using cellular link information. In *ACM Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2017.

[205] Yaxiong Xie and Kyle Jamieson. Ng-scope: Fine-grained telemetry for nextg cellular networks. *Proc. ACM Meas. Anal. Comput. Syst.*, 6(1), feb 2022.

[206] Yaxiong Xie, Fan Yi, and Kyle Jamieson. Pbe-cc: Congestion control via endpoint-centric, physical-layer bandwidth measurements. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '20, page 451–464, New York, NY, USA, 2020. Association for Computing Machinery.

[207] Yongshun Xie and Chengjin Wang. Vulnerability of submarine cable network of mainland China: Comparison of vulnerability between before and after construction of trans-arctic cable system. *Complexity*, 2021.

[208] Qiang Xu, Junxian Huang, Zhaoguang Wang, Feng Qian, Alexandre Gerber, and Zhuoqing Morley Mao. Cellular data network infrastructure characterization and implication on mobile content placement. In *ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2011.

[209] Yihan Xu, Reza K Amineh, Ziqian Dong, Fang Li, and Michael Kohler. Wireless sensing with software defined radio. In *2023 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (USNC-URSI)*, pages 1747–1748. IEEE, 2023.

[210] Yin Xu, Zixiao Wang, Wai Kay Leong, and Ben Leong. An end-to-end measurement study of modern cellular data networks. In *Proceedings of the 15th International Conference on Passive and Active Measurement - Volume 8362*, PAM 2014, page 34–45, Berlin, Heidelberg, 2014. Springer-Verlag.

[211] Zhenqiang Xu, Pengjin Xie, and Jiliang Wang. Pyramid: Real-time lora collision decoding with peak tracking. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2021.

[212] Hyo-Sik Yang, M. Herzog, M. Maier, and M. Reisslein. Metro wdm networks: performance comparison of slotted ring and awg star networks. *IEEE Journal on Selected Areas in Communications (J-SAC)*, 2004.

[213] Kyriakos Zarifis, Tobias Flach, Srikanth Nori, David Choffnes, Ramesh Govindan, Ethan Katz-Bassett, Z Morley Mao, and Matt Welsh. Diagnosing path inflation of mobile client traffic. In *Passive and Active Network Measurement Conference (PAM)*, 2014.

[214] Chengwei Zhang, Xiaojun Hei, and Brahim Bensaou. A measurement study of campus wifi networks using wifitracer. *Cyber-Physical Systems: Architecture, Security and Application*, pages 19–42, 2019.

[215] Shiwei Zhang, Weichao Li, Daoyuan Wu, Bo Jin, Rocky Chang, Debin Gao, Yi Wang, and Ricky Mok. An empirical study of mobile network behavior and application performance in the wild. In *Proceedings of IEEE IWQoS*, 2019.

[216] Zesen Zhang, Alexander Marder, Ricky Mok, Bradley Huffaker, Matthew Luckie, kc claffy, and Aaron Schulman. Inferring regional access network topologies: Methods and applications. *IMC*, November 2021.

[217] Mariya Zhivkova Zheleva, Ranveer Chandra, Aakanksha Chowdhery, Paul Garnett, Anoop Gupta, Ashish Kapoor, and Matt Valerio. Enabling a nationwide radio frequency inventory using the spectrum observatory. *IEEE Transactions on Mobile Computing*, 17(2):362–375, 2018.

[218] Mirela Șorecău, Emil Șorecău, Annamaria Sarbu, and Paul Bechet. Real-time statistical measurement of wideband signals based on software defined radio technology. *Electronics*, 12(13):2920, 2023.