TOPOLOGY-AWARE PERSPECTIVE FOR ENHANCING ROBUSTNESS IN THE INTERNET

By

Ramakrishnan Durairajan

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: August 7, 2017

The dissertation is approved by the following members of the Final Oral Committee:

Paul Barford, Professor, Computer Sciences
Aditya Akella, Professor, Computer Sciences
Remzi Arpaci-Dusseau, Professor, Computer Sciences and ECE
Suman Banerjee, Professor, Computer Sciences
Joel Sommers, Associate Professor, Colgate University

TOPOLOGY-AWARE PERSPECTIVE FOR ENHANCING ROBUSTNESS IN THE INTERNET

Ramakrishnan Durairajan

Under the supervision of
Professor Paul Barford
At the University of Wisconsin–Madison

Outages and loss of connectivity in the Internet can have a significant impact on businesses and users. While the Internet was designed to recover from simple failures, there are numerous examples of accidents and attacks over the past two decades that have resulted in large-scale loss of connectivity. This suggests that standard mechanisms to ensure Internet robustness require renewed consideration. Toward the goal of improving Internet robustness, this dissertation has two themes: *(i)* understanding the structural complexity and inherent risks in the Internet—two factors which makes the Internet failure-prone, and *(ii)* building scalable, robust and easy-to-deploy systems to enhance network robustness.

In this dissertation, we first investigate how a comprehensive understanding of structural complexity and inherent risk are critical for robust Internet design and operation. A key challenge in this regard is that Internet is in a constant state of flux. To overcome this challenge, we take a bottom-up approach and build Internet Atlas, which is a comprehensive repository of the physical Internet. The main goal of Internet Atlas is to develop a geographically accurate representation of the Internet's physical interconnection infrastructure map to understand the structural complexity. We also develop a new probing heuristic called POPsicle to broadly identify Internet infrastructure that has a fixed geographic location such as Point-of-presences (POPs), Internet exchange points (IXPs), datacenters and other kinds of hosting facilities.

One of the most striking characteristics of the constructed map is a significant amount of observed infrastructure sharing, which in turn aggravates the risks inherent on topology. Such infrastructure sharing is the result of a common practice among many of the existing Internet service providers (ISPs) to deploy their fiber in jointly-used and previously installed conduits and is dictated by simple economics—substantial cost savings, among other objectives, as compared to deploying fiber in newly constructed conduits. By considering different metrics for measuring the risks associated with infrastructure sharing, we examine the presence of high-risk links in the existing infrastructure, both from a connectivity and usage perspective.

Given the understanding of structural complexity, we next develop systems that take advantage of emerging technology to satisfy ISP objectives and to minimize shared risks. First, we create a decision support framework that uses geo-based multi-objective optimization to identify target areas with the highest concentration of un/underserved users at the the lowest cost to service providers for network infrastructure deployment. Second, we propose a system called GreyFiber, which provides a means to offer easy and cost-effective access to unused fiber-optic paths between participating endpoints on demand based on market economics, for arbitrary durations, and possibly with industry-specific performance guarantees.

Dedicated with special gratitude and warmest affection to my parents, my family, my friends, and my mentors, for their unconditional love and support.

Acknowledgments

Right after my Ph.D. defense, I took a long walk around UW-Madison campus, reflecting on successes, failures, rejections, and people who played an important part in the last 6 years. The following is a brief excerpt from that reflection.

First, my family. Thank you Amma and Appa for your unconditional love and emotional support. Amma: I am deeply indebted to you for teaching me your never-give-up attitude. Appa: you are the hardest working person I've ever seen in my life; I am very grateful to you for teaching me that. You two always provided me with more than what I ever wanted. Thank you for being a rock behind me and for instilling strong work ethic and punctuality in me. Next, many thanks to my sister for being there for me and for putting up with all my craziness. Thanks for setting the bar very high, every time. I still remember my childhood days, watching and learning from her. Thanks to my brother-in-law for his support throughout this journey. Specials thanks to the kids for reminding me "after all, there is life". I dedicate this thesis to my parents and my family!

Next, I am very grateful to my amazing colleagues without whom none of the work presented in this thesis would have been possible. In particular, many thanks to my advisor, Paul Barford, for his professional support, mentorship and friendship. In his guidance I have never had second thoughts about returning to graduate school, leaving a well-paid job in India behind. Paul is definitely one of the best teachers in the

department. The way he sets the context before discussing a topic and his knowledge about the topic itself are spectacular. In fact, I am choosing an academic career because of the inspiration drawn from his teaching style and research rigor. Paul: thank you. I have immensely benefited from your mentoring over the years and have always found you to be *(1)* a great advisor, whenever I cannot geolocate (no pun intended!) myself in a research problem; *(2)* an intelligent colleague, who is always available to brainstorm new ideas and thoughts; and *(3)* a warm friend during difficult personal situations. Looking forward to the future collaborations.

Thank you Joel Sommers for being my “unofficial" co-advisor, for your patience, and for your friendship. Special thanks to you for your attention to details, writing tips, friendly how-is-the-paper-looking emails before deadlines, and your company/support, especially in conferences. Joel: thank you. I could not ask for a better collaborator. Thank you Walter Willinger for being the “Dumbledore" of Internet measurements research, for asking all the right questions during our meetings and for being a wonderful collaborator. Walter: thank you. One of the first papers that I read when I started my Ph.D. was yours and you are my constant source of inspiration. I am very thankful to you for that one meeting (at SIGCOMM 2014) that shaped many aspects of this thesis. Special thanks to my prelim and defense committee members—specifically, Aditya Akella, Suman Banerjee, Remzi Arpaci-Dusseau and Robert Nowak—for all their valuable comments and suggestions that helped shape this thesis. I am also grateful to my internship mentors. Special thanks to Dave Plonka and Arthur Berger (Akamai, Inc.); Lee Breslau, Zihui Ge, and Vijay Gopalakrishnan (AT&T Research); and Stefan M. Petters (CISTER research). Thanks to my undergraduate advisor, Uma Maheshwari, for her guidance and for helping me get that proverbial foot on the door.

Thanks to the present and past members of WAIL, WISDOM, WiNGS and Akella group including Scott Alfeld, Subhadip Ghosh, Dave Plonka,

Joe Chabarek, Igor Canadi, Brian Eriksson, Meena Syamkumar, Sathiya Kumaran Mani, Ashok Anand, Theo Benson, Aaron Gember-Jacobson, Wenfei Wu, Ashish Patro, and Tan Zhang for their professional support, feedback and encouragement. Thanks to Mike Blodgett for helping me with many of my experimental needs and setups. Thank you Anton Kapela for patiently answering all my questions about real ISP deployments. Thanks to Bruce Maggs, Somesh Jha, Mark Crovella, Vyas Sekar, Dave Choffnes, and Dongsu Han for their suggestions and pointers on my job application.

Finally, it goes without saying that my friends (a.k.a my other family) created the perfect “home away from home". They restored my sanity from time-to-time and I am immensely grateful for that. What started as a simple invite to Sankaralingam Panneerselvam’s birthday party in the hallway (in 2011) has grown into a gang of great friends. Special thanks to Uthra Srinath, Srinath Srinath, Sankaralingam Paneerselvam, Venkatanathan Varadarajan, Vijay Chidambaram, Sajala Rajendran, Theva Chandereng, Thanumalayan Pillai, Sandeep Vishwanathan, Shriram Sridharan, Sreenivaas Sudhakar, Surya Narayanan, Sibin Philip, Madhav Venkateswaran, Ragunanth Venkatesan, Chetan Rao, Gautam Prakriya and many others for weekend potlucks, games, hangout sessions, roadtrips, useless discussions, and all the “good" times. Thanks to Andrea Keapproth, Venkatesh Srinivasan, Adam Everspaugh, Robert Grandl, Junaid Khalid, Raajay Viswanathan, Vinitha Raajay, and Aparna Subramanian for their friendship; I cannot imagine the last couple of years of Ph.D without their company. Thanks to my friends in India—Rahul Shankar and Ananth Raghav—for their lifelong friendship.

THANK YOU to every single one of you, especially for being there during my failures and rejections, for helping me bear the cold Wisconsin winters with warm friendship, and for standing by my side throughout this journey without any expectation.

This work was supported in part by NSF grants CNS-0831427, CNS-

0905186, CNS-1054985; ARL grant W911NF1110227; AFRL grant FA8750-12-2-0328; and DHS grant BAA 11-01. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF, ARL, AFRL or DHS.

Contents

List of Figures and Tables

1 Introduction

Internet is the most gigantic world-wide communications infrastructure ever built. The rise of new novel designs, technologies and applications such as datacenters, cloud services, software-defined networking (SDN), network functions virtualization (NFV), mobile communication and the Internet-of-Things (IoT), has fueled the recent evolution of the Internet. The excitement surrounding the future envisioned by such new designs, services, and applications is understandable, both from a research and industry perspective. At the same time, it is either taken for granted or implicitly assumed that the physical infrastructure of tomorrow's Internet will have the capacity, performance, and resilience required to develop and support ever more bandwidth-hungry, delay-intolerant, or QoS-sensitive services and applications. Furthermore, the basic design [195] to adapt itself almost instantaneously to damage or outages should make the Internet robust against failures of physical infrastructure components such as routers and links.

While IP routing allows the network to dynamically detect and route around localized failures, events such as natural or technological disasters (*e.g.*, [80, 122]), benign incidents (*e.g.*, [138]), accidents (*e.g.*, the Baltimore Howard Street Tunnel fire [266] or Mediterranean Cable Cuts [318]), misconfigurations (*e.g.*, Pakistani YouTube routing [235]), terrorism (*e.g.*, the World Trade Center attack [142] or a potential Electromagnetic Pulse Attack [294]), or censorship (*e.g.*, response to the 2011 Egyptian uprising [158]) can have significant effects, including the loss of connectivity

for large sections of Internet users for extended periods of time. This suggests that, while in principle the Internet is designed to be robust, the underlying robustness mechanisms are often insufficient. In addition, while many other dynamic aspects (like routing) of the Internet have been examined in prior work [176, 228, 230, 299, 317], the underlying *physical infrastructure*—specifically, the geographic locations of nodes (*e.g.*, POPs, IXPs, datacenters, etc.) and links (*e.g.*, fiber strands housed in physical conduits)—that make up the Internet are, by definition, static[1] and are completely ignored, which leads to a *robustness gap* (*e.g.*, unpredictability in routing convergence, cascading failures, etc.). Bridging this robustness gap is an open problem.

1.1 Approach

In this dissertation, we are motivated by the questions, "why is today's Internet far from being robust with numerous loss-of-connectivity episodes and how do we transcend the robustness gap to create a robust Internet?" These questions leads us to hypothesize that the *(i)* ignored physical infrastructural complexity (*i.e.,* lack of understanding of nodes, links and connectivity) and inherent risks on topology are what makes the Internet failure-prone and *(ii)* creating methods and frameworks to unravel the complexity and risks in the Internet is the first logical step towards a truly robust Internet. To this end, we take a bottom-up approach to understand the structural complexity and its associated risks and show how the understanding is critical for enabling robust Internet operation.

As actionable items, using the lessons learned from measuring and understanding the Internet's infrastructural complexity, the risks and their associated root causes, we build new systems to mitigate the infrastructural

[1]More precisely, installed conduits rarely become defunct and deploying new conduits takes time.

risks and to further enhance the robustness of the Internet. In short, we take a “measure-and-then-build" approach to create a more robust Internet—the main vision of this research.

1.1.1 Thesis Statement

The Internet is continuously evolving and the scope of Internet usage is beyond its original design. By understanding the structural complexity and risks inherent on physical infrastructure, we build topology-aware systems to create a more robust Internet.

1.1.2 Research Questions

The specific research questions we address in this dissertation are:

- How do we unravel the complexity of the Internet infrastructure, and develop scalable techniques to better reveal the physical infrastructure?
- What are the problems and risks in the infrastructure? What are the root causes for infrastructural problems and risks?
- Given the understanding of complexity and risks, how do we build systems to create a more robust Internet?

1.1.3 Unravelling Structural Complexity

Among the factors impeding the comprehensive unravelling of complexity, one stand out: *physical topology*. Why? Studies that aim to map the Internet’s topological structure have been motivated for many years by a number of compelling applications including the possibilities of improving performance, security and robustness. While these motivations remain as compelling as ever, the ability to accurately and comprehensively map the Internet has, for the most part, remained beyond our grasp. In fact, despite some 20 years of research efforts that have focused on understanding aspects of the Internet’s infrastructure such as its router-level topology

or the graph structure resulting from its inter-connected Autonomous Systems (AS), very little is known about the complexity of today's *physical* Internet.

The primary challenges to thoroughly mapping the Internet stem from its enormous size, distributed ownership, and constantly changing characteristics. Faced with these challenges, the most widely used approach to Internet mapping, to produce "network-layer maps," has been based on gathering data from network-layer measurements using TTL-limited probes[2]. Ideally, network-layer maps reflect a timely representation of network topology as well as the dynamic aspects of management and configurations. Great progress has been made on solving some of the specific problems related to using these network-layer maps for understanding aspects of Internet topological characteristics. However, the fact remains that layer 3 data are inherently tied to the management policies and operational objectives of Internet service providers (ISPs), which may be at odds with comprehensive and accurate mapping of the Internet.

We posit that the starting point towards understanding the complexity of Internet is to understand the physical topology. To investigate this, we developed Internet Atlas (or Atlas) [206], which is a geographically anchored representation of the *physical Internet* including *(i)* nodes (*e.g.,* hosting facilities and data centers), *(ii)* conduits/links that connect these nodes, and *(iii)* relevant meta data (*e.g.,* source provenance). The objective of this effort is to build a comprehensive map of the physical Internet and is built by using search to identify primary source data such as maps and other repositories of service provider network information. This data is then carefully entered into the database using a combination of manual and automated processes including consistency checks and methods for geocoding both node and link data. Atlas currently contains over 25K Point-of-Presence (POP) locations and nearly 27K links for over 1200 networks

[2]Maps can also be created using BGP updates or application-layer data, however those are not the focus of this work.

around the world. Customized interfaces were built to import a variety of dynamic (*e.g.,* BGP updates, Twitter feeds and weather updates) and static (*e.g.,* highway, rail and census) data into Atlas, and to layer it on top of the physical representation. We refer to the collection of these data and their geographic locations as “physical maps" of the Internet. These maps are valuable because they reflect a ground truth perspective of service provider infrastructure.

Next, we investigate the hypothesis that physical maps of service provider infrastructure can be used to effectively guide topology discovery based on network layer TTL-limited measurement. The goal is to focus layer 3-based probing on broadly identifying *Internet infrastructure that has a fixed geographic location* such as POPs, IXPs and other kinds of hosting facilities. We begin by comparing more than 1.5 years of TTL-limited probe data from the Ark [143] project with maps of service provider infrastructure from the Internet Atlas [206] project. We find that there are substantially more nodes and links identified in the service provider map data versus the probe data. Next, we describe a new method for probe-based measurement of physical infrastructure called *POPsicle* [207] that is based on careful selection of probe source-destination pairs. We demonstrate the capability of our method through an extensive measurement study using existing “looking glass” vantage points distributed throughout the Internet and show that it reveals 2.4 times more physical node locations versus standard probing methods. To demonstrate the deployability of POPsicle we also conduct tests at an IXP. Our results again show that POPsicle can identify more physical node locations compared with standard layer 3 probes, and through this deployment approach it can be used to measure thousands of networks world wide.

Finally, we study the characteristics of the long-haul fiber-optic network in the US [204]. We start by using fiber maps of tier-1 ISPs and major cable providers from Internet Atlas to construct a map of the long-haul US

fiber-optic infrastructure. We also rely on previously under-utilized data sources in the form of public records from federal, state, and municipal agencies to improve the fidelity of our map. We quantify the resulting map's connectivity characteristics and confirm a clear correspondence between long-haul fiber-optic, roadway, and railway infrastructures.

1.1.4 Understanding Topological Risks

A striking characteristic of the constructed maps is a significant amount of observed infrastructure sharing [204] which leads to the problem called *Shared Risk*: physical conduits shared by many service providers are at an inherently risky situation since damage to those conduits will affect many several providers. Such infrastructure sharing is the result of a common practice among many of the existing ISPs to deploy their fiber in jointly-used and previously installed conduits and is dictated by simple economics—substantial cost savings, among other ISP objectives, as compared to deploying fiber in newly constructed conduits. A qualitative assessment of the risk inherent in this observed sharing and their corresponding root causes forms the second contribution of this thesis.

By considering different metrics for measuring the risks associated with infrastructure sharing, we examine the presence of high-risk links in the existing long-haul infrastructure, both from a connectivity and usage perspective. In the process, we also do a detailed analysis of how to improve the existing long-haul fiber-optic infrastructure so as to increase its resilience to failures of individual links or entire shared conduits, or to achieve better performance in terms of reduced propagation delay along deployed fiber routes. By framing the issues as appropriately formulated optimization problems, we show that both robustness and performance can be improved by deploying new fiber routes in just a few strategically-chosen areas along previously unused transportation corridors and right-of-ways (ROWs), and we quantify the achievable improvements in terms of

reduced risk (*i.e.*, less infrastructure sharing) and decreased propagation delay (*i.e.*, faster Internet [289]).

These technical solutions often conflict with currently-discussed legislation that favors policies such as “dig once", “joint trenching" or “shadow conduits" due to the substantial savings that result when fiber builds involve multiple prospective providers or are coordinated with other infrastructure projects (*i.e.*, utilities) targeting the same ROW [18]. In particular, we discuss our technical solutions in view of the current net neutrality debate concerning the treatment of broadband Internet providers as telecommunications services under Title II. We argue that the current debate would benefit from a quantitative assessment of the unavoidable trade-offs that have to be made between the substantial cost savings enjoyed by future Title II regulated service providers (due to their ensuing rights to gain access to existing essential infrastructure owned primarily by utilities) and an increasingly vulnerable national long-haul fiber-optic infrastructure (due to legislation that implicitly reduced overall resilience by explicitly enabling increased infrastructure sharing).

1.1.5 Building Systems for Robust Internet

Given the understanding of complexity and risks in the Internet, our next step is to investigate and build systems that are easily deployable making the path towards more robust Internet a plausible vision. While prior efforts focused on routing mechanisms which are the traditional focus for robustness including the development of resilient routing protocols [176, 228, 230, 299, 317], they often preclude issues such as deployability and manageability.

In this dissertation, we design and build two systems with deployability and manageability in mind. First, we build Deployment-as-a-Service (DaaS), a framework whose objective is to provide flexible decision support on opportunities for broadband deployment that enables economic

and technical issues to be considered simultaneously. Specifically, our framework considers *(i)* infrastructure proximity, *(ii)* demographics, and *(iii)* deployment costs. We employ geographically-based, multi-objective optimization to identify the *highest* concentrations of un/underserved users and that can be upgraded to the broadband threshold at the *lowest* cost. Our work takes advantage of our maps of long-haul infrastructure in the US that are critical for accurate cost modeling. We demonstrate the efficacy of our approach by considering US demographic data and two different deployment models: upgrading existing infrastructure and deploying new infrastructure. Our results highlight the tradeoffs of the different deployment models and identify a list of US counties that would be attractive targets for broadband deployment from both cost and impact perspectives and that correspond closely with areas identified by Connect America map [34]. While our analysis focuses on the US, our method is generic and can be applied in other regions where similar data is available.

Second, we propose and build GreyFiber [205]. The main idea for GreyFiber is to provide a means to offer easy and cost-effective access to unused fiber-optic paths between participating endpoints (*e.g.*, colocation facilities) on demand, for arbitrary durations, and possibly with industry-specific performance guarantees (*e.g.*, ultra-low delay for high-frequency trading applications or gaming services; fully diverse physical paths for mission-critical business applications). In this sense, GreyFiber can be thought of as offering *Wide Area Connectivity-as-a-Service*. However, GreyFiber differs from standard cloud computing services (*e.g.*, SaaS, PaaS and IaaS) in that it is fundamentally concerned with connectivity, not computation. To demonstrate the feasibility of our approach and examine its efficacy, we describe an implementation of our design and deploy it in the GENI testbed [161]. This prototype system addresses the technical challenges associated with circuit provisioning and enables performance evaluation over a range of use scenarios. First, we show that as many as 50

paths can be provisioned between endpoints in less than a minute, which demonstrates the scalable and rapid provisioning capabilities of GreyFiber. Next, to enable higher infrastructure resilience during network outages and/or planned maintenance events, we show how GreyFiber can be used to create an effective backup solution. Specifically, GreyFiber can reactively detect path failures and provision a new path within 1.25s, which outperforms the traditional OSPF-based backup solution by 28x. This agility of GreyFiber benefits many applications by allowing them to be oblivious to underlying network failures. Finally, we dynamically provision paths between endpoints to create on-demand high-capacity connectivity and demonstrate the resulting performance benefits of GreyFiber.

We believe that DaaS is paramount for satisfying ISP objectives (*e.g.*, minimizing costs), while GreyFiber is a way to solve the problem of shared risk in the Internet.

1.2 Summary of Major Contributions

The following are the contributions made by this dissertation.

- We built Internet Atlas—a new visualization and analysis portal for diverse Internet measurement data, which is by far the largest repository of Physical Internet maps in the world. Internet Atlas is used by over 150 researchers in the community world-wide.
- We perform a first-of-its-kind comparison of large repositories of physical and network maps and find that physical maps typically reveal a much larger number of nodes (*e.g.*, POPs and hosting infrastructure).
- We consider the targeting problem and find that using sources and destinations within the same autonomous system for probing reveals the most physical infrastructure.
- We develop a layer 1-informed heuristic algorithm for probe source-destination selection called POPsicle that identifies 2.4 times as many

nodes as standard probing methods.

- We identify the fact that sources co-located as IXPs can be used to amplify POPsicle-based probing broadly throughout the Internet resulting in layer 3 maps that can be more effectively applied to problems of interest. To that end, we deployed our method at a real IXP and found that our method finds almost all POPs compared to Atlas and additional POPs compared to Ark for the ISPs studied.
- We solve what had been a *two-decade-long open problem* by constructing an accurate, first-of-its-kind reproducible map of the physical Internet infrastructure. To improve the fidelity of our map, we validate the constructed map using a surprisingly new dataset: previously underutilized data sources in the form of public records from federal, state, and municipal agencies.
- We develop qualitative assessment of the risk inherent in the constructed representation and their corresponding root causes.
- We show how both risk and latency (*i.e.*, propagation delay) can be reduced by deploying new links along previously unused transportation corridors and rights-of-way using simple optimization models. In particular, we show that focusing on a subset of high-risk links is sufficient to improve the overall robustness of the network to failures. We discuss the implications of our findings on issues related to performance, net neutrality, and policy decision-making.
- We develop a framework for satisfying ISP objectives that considers economic and technical issues simultaneously.
- We built a new system for offering *Wide Area Connectivity-as-a-Service*, which, apart from mitigating shared risk, is key enabler for diverse applications: *(1)* ultra-low delay paths for high-frequency trading or gaming services, *(2)* fully diverse physical paths for mission-critical business applications, *(3)* capacity-scalable paths for Big Data transfers, and *(4)* flexible means for wide-area NFV chaining.

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows. Chapter 2 discusses relevant prior efforts and studies in this space. We describe the Internet Atlas repository in Chapter 3. Chapter 4 describes the study that compares and contrasts the physical and network layer maps. It also describes POPsicle, a new layer 3-based probing to broadly identify physical infrastructure assets. Chapter 5 studies the characteristics and implications of observed infrastructure sharing in the long-haul fiber-optic network in the US. Chapter 6 outlines a new framework for providing flexible decision support for ISP deployments. The idea of GreyFiber and the realization of GreyFiber in GENI testbed, along with performance guarantees and applications are discussed in Chapter 7. We summarize and describe future work in Chapter 8.

2 Related Work

2.1 Mapping Infrastructure and Connectivity

Analyzing the interconnection structure of the Internet has been the subject of a large number of studies over the past decade. The network scope of these studies range from the router-level (*e.g.,* [188, 263]), to POP-level (*e.g.,* [284, 292]), to the autonomous system-level (*e.g.,* [265, 310]). Most of these studies are based on layer 3 measurements from traceroute-like tools at the router-level, and BGP announcements at the AS level (*e.g.,* [126]). The dynamic nature of the data used in these studies presents significant challenges in recovering details of the underlying physical infrastructure.

Layer 3-based Mapping Efforts. There has been a great deal of effort made to harness layer 3 TTL-limited probes for network mapping since the introduction of the traceroute tool [236]. Some efforts (*e.g.,* [287, 292]) have focused on the goal of developing a comprehensive network-layer view of the Internet *i.e.,* unique identification of nodes and links. Other efforts have focused on developing new probing techniques that expand the ability to collect data and thereby improve accuracy and mapping coverage, *e.g.,* [179, 180, 288]. More recent efforts have focused on analyzing and addressing various inaccuracies inherent in probe-based network mapping [291, 313]. For example, Roughan, *et al.* and Eriksson, *et al.* develop inferential techniques to quantify the nodes and links that are missed through network-layer mapping [211, 280]. Other researchers have looked

closely at the rise of Internet Exchange Points (IXPs) and the effects of IXPs on inaccuracies of network-layer mapping, *e.g.*, [175, 180]. Concurrent with the rise of IXPs has come a "flattening" of the Internet's peering structure [201, 220, 254], which affects the very nature of end-to-end paths through the Internet. Still other researchers have observed that increased use of network virtualization techniques such as MPLS have led to additional inaccuracies in layer 3 mapping, and which are likely to continue to thwart probe-based mapping efforts [202, 287, 290]. We posit that layer 3 mapping efforts will continue to be important sources of Internet topology information and that complementary efforts to build repositories of physical Internet maps (*e.g.,* [206, 251]) will result in representations of Internet topology that are more accurate and applicable to problems of interest than either representation in isolation.

The study that bears the strongest resemblance to ours is the Internet Topology Zoo [251], which offers representations of service provider maps that are discovered from search (Gorman used a similar approach in his Ph.D. thesis [222] although no maps from that work are available). Indeed, we used the Topology Zoo as a source of data for our repository. However, Internet Atlas goes well beyond Topology Zoo by including: *(1)* Dynamic, GIS-based visualization including search, filtering and overlaying of multiple networks, *(2)* A larger and more diverse repository of network node locations, *(3)* Real-time data feeds (*e.g.,* BGPmon and NOAA weather reports) layered on top of maps of physical infrastructure, *(4)* Static geo-coded data (*e.g.,* census and infrastructure), *(5)* Geo-location of physical links (when available from maps) and *(6)* Geo-spatial analyses (*e.g.,* Kriging estimation). To the best of our knowledge, Internet Atlas is the first academic effort to establish a GIS-based web portal that includes diverse measurement data layers on top of a map of the physical Internet.

2.2 Targeted TTL-limited Probing

The *targeting problem* that is a focus of POPsicle is informed by prior studies that analyze the intrinsic importance of measurement infrastructure in Internet topology mapping. Barford *et al.* were among the first to quantify the value of vantage points in discovery of nodes and links in core and edges of the Internet [182]. More recently, Shavitt and Weinsberg consider the problem of bias in measurements based on vantage point distributions and show that a broad distribution of vantage points reduces bias in resulting maps [285]. Our work differs from these studies in that we are focused on using layer 3 probes to identify specific infrastructure targets.

2.3 IP Geolocation

Identifying the geographic location of nodes that have been assigned specific IP addresses (*i.e., IP geolocation*) is a challenging problem that is highly relevant to our study. Some of the earliest work on this problem was done by Paxson, who developed the idea of using DNS hints to identify the geographic locations of nodes that were responding to TTL-limited probes [269]. We use similar methods in our study. Since then, many studies have addressed the problem of IP and POP geolocation using a variety of measurement techniques (*e.g.,* [210, 214, 227, 245, 268, 276, 300, 304]). The fact that POP locations in physical maps are often given at the street address level offers the possibility to improve IP geolocation estimates using standard measurement-based methods. We plan to investigate another possibility of leveraging state-of-the-art geolocation techniques (*e.g.,* [234, 300]) to enhance the accuracy of our location extraction approach in future work.

2.4 Network Robustness and Outage Studies

Analyzing the robustness of the physical Internet has been the focus of many prior research efforts. These include studies on its robust yet fragile nature [203, 303], vulnerability [222, 223, 315], survivability [231, 232], resilience analysis [174, 212, 305], reachability [186, 246], security and robustness of components [242], fault detection/localization [221, 247, 272], and the development of resilient routing protocols [176, 228, 230, 299, 317]. In contrast to these and similar prior efforts, our study is the first to consider the extensive levels of physical infrastructure sharing in today's Internet, use various metrics to quantify the resulting shared risk and offer viable suggestions for improving the overall robustness of the physical Internet to link and/or router failures.

The main reasons for localized and temporal Internet outages are typically a lack of geographic diversity in connectivity [5, 76] and a tendency for significant physical infrastructure sharing among the affected providers—the very focus of our work. In particular, our work is not about the Internet's vulnerability to non-physical cyber attacks (*e.g.*, [89]) that rely on the existence and full functionality of the Internet's physical infrastructure to achieve their goals and do maximal damage [203].

Our study centers around the construction of a high-fidelity map of the long-haul fiber-optic routes in the US Internet and relies critically on a first-of-its-kind analysis of the detailed geography of these routes. On the one hand, there exists prior work on mapping the US long-haul fiber-optic network (see for example [5, 68]), but the resulting maps are of uncertain quality, lack important details, and are not reproducible. There have also been prior studies that examine different aspects of the Internet infrastructure and various spatial patterns that have emerged (see for example [264]). On the other hand, the basic map constructed as part of our work is based on rich information from publicly available resources and can be reproduced by anybody who has the time and energy to gather

the available but not necessarily easy-to-locate information.

The detailed analysis of our long-haul fiber-optic network map is made possible by using geocoded network maps and the ArcGIS framework [50], and is unprecedented both in terms of accuracy and ability for validation. In contrast to the work by Lakhina *et al.* [256] who use geolocation databases to obtain the approximate link lengths between geolocated routers, our study avoids the issues related to router-level granularity (*e.g.*, errors in geolocating routers, use of line-of-sight for estimating link distances) by exploiting the detailed geography of the long-haul fiber-optic routes between major city pairs and computing their actual lengths. In the process, we compare our long-haul fiber-optic map to existing transportation infrastructure (*e.g.*, railway, roadways) and quantify previously made qualitative observations that place much of the long-haul fiber-optic infrastructure along railways and roadways [68].

2.5 Connecting Un-/Under-served Communities and Internet Penetration Studies

Understanding the Internet penetration rate and its economic impact has been a subject of inquiry for the last two decades [197, 198, 244]. These studies consistently conclude that Internet connectivity at broadband speeds is essential for growth and economic prosperity. Since the dot-com bubble, several efforts studied the Internet adoption rate in un/underserved areas, both empirically [178] and qualitatively [99], and found several interesting rate determining factors, including gender [273, 277], age [181] and race [279]. Even though these factors influence Internet penetration to some extent, key determinants like availability of telecom infrastructure, federal regulations and economic affordability play a signif-

icant role in closing the digital divide in un/underserved areas [192, 302]. Finally, several research projects have proposed paradigms [282], technologies (both traditional [278] and alternative [11]) and approaches [295] for improving Internet penetration in un/unserved communities.

Determining target areas for infrastructure deployment and optimizing deployment costs are two key components of our framework. While we take a GIS-based approach similar to prior efforts [187, 283] for the former, we use insights from Ranaweera *et al.* [275] for various cost optimizations (*e.g.* upgrading existing infrastructure) to address the latter. We argue that our framework offers the ability to assess technological and economic tradeoffs in deploying or upgrading infrastructure in a way that has not been considered in these prior studies.

2.6 Infrastructure Provisioning

The notion of provisioning dedicated circuits between endpoint pairs, *i.e.*, *circuit switching* and its cousin *virtual circuit switching*, has been well studied in the community for decades. Our approach of establishing dynamic paths between two endpoints is informed by the seminal work on circuit switching by Erlang *et al.* [213] and later by Kelly *et al.* [248].

In the context of datacenter and WAN settings, infrastructure provisioning has been of interest to both industrial and academic communities [176, 233, 237, 239, 258, 299, 317]. SDN-based provisioning approaches include B4 [237], SWAN [233] and OWAN [239], each of which are aimed at improving the utilization of inter-datacenter and wide area networks. We posit that deployment of such efforts along with acquiring access to physical paths (via IRU or GreyFiber) between DCs has the potential to produce better performance results than considering either of these solutions in isolation. In particular, we argue that such an environment, which considers provisioning and access to physical paths in tandem, can facilitate improve-

ments at the physical layer [183, 307], network layer optimizations [253], and cross-layer enhancements [185, 193, 226, 274]. From a traffic engineering perspective, efforts like Tempus [241] and Amoeba [311] have the goal of guaranteeing timely transfer of bulk data across datacenters.

Provisioning of infrastructure to enhance the robustness of networked systems has been a key focus of many prior works. Notable efforts include backup routing [219], preventive routing via risk analysis [212], management system for provisioning [184], and enabling backup paths using either optimization of IGP link weights [218] or RSVP-TE's *fast reroute* mechanism [157].

2.7 Internet Economics

Incorporating pricing models for networks has been of interest to researchers since the Internet's infancy [249, 260, 262, 286, 296]. Recently, many efforts have focused on increasing revenues for service providers and customer satisfaction via flexible pricing models. For instance, Jalaparthi *et al.* [238] accommodates both deadlines and demands into a time-dependent pricing model to create Pretium, a framework which considers economics and traffic engineering issues in tandem. Similarly, a pricing model for transit ISPs based on tiers and traffic demand is proposed in [297].

The auction model in GreyFiber is motivated by online auction research in the theory literature. Specifically, we use the classical results on Generalized Second Price (GSP) [208] or Vickrey-Clarke-Groves (VCG) [196, 225, 298] in our framework. Economics on infrastructure—specifically, auction models and bidding strategies in cloud settings—are studied in [314]. Furthermore, several industrial efforts on infrastructure economics include bandwidth markets (*e.g.*, Enron [49]), spot pricing markets (*e.g.*, Invisible Hand Networks [86]), and fiber arbiters (*e.g.*, IXReach [91], Pack-

etFabric [115]). In particular, IXReach (which was acquired in 2015 by IIX, Inc. [78] which in turn was renamed as Console [90]) provides the ability to expand network footprint at locations that are of interest to service providers à la GreyFiber.

3

Internet Atlas: A Geographic Database of the Physical Internet

3.1 Introduction

Accurate and timely maps of the Internet would provide a starting point for diverse research topics such as assessing infrastructure vulnerabilities, understanding routing behavior, developing new protocols, analyzing application performance, and large scale network simulation studies. Such maps would also be highly valuable in network operations since they could provide insights to fault diagnosis, opportunities for peering and transit, and planning for future growth. However, despite the many and varied efforts over the years, there remains no central repository of accurate Internet maps.

The lack of a central repository of Internet maps can be attributed to several significant challenges. First, it is well known that the Internet is a gigantic and complex world-wide infrastructure that is in a constant state of flux. Second, the fundamental “network-of-networks" organization of the Internet means that no single provider can offer an authoritative perspective on structure. Third, the de facto use of IP addresses gathered from TTL-limited probing campaigns as the basis for inferring structure has inherent difficulties. These include the well known interface disambiguation problem [292], widely varying policies on probe blocking

among providers, and difficulties in managing large scale measurement infrastructures [270]. We believe that a different approach to building and maintaining a repository of Internet maps is required.

In this chaper, we describe the *Internet Atlas* project. Our objective is to build a comprehensive, geographically accurate map of the physical Internet *i.e.,* nodes (*e.g.,* hosting facilities and data centers) and links (*e.g.,* optical fiber conduits), extend this map with relevant, related data (*e.g.,* BGP updates, etc.), and to make the data available through a web portal for visualization and analysis. We posit that such a map or repository would lend itself directly to a wide range of research and operational applications that have motivated prior Internet mapping efforts.

Our first focus is on developing a comprehensive and geographically accurate representation of the Internet's *physical interconnection structure*. Specifically, we seek to identify the locations of the buildings that house switching and routing equipment as well as the paths of physical conduits that connect them. We argue that this physical perspective is *(i)* likely to be a much smaller graph than those produced from layer 3 measurements, thus potentially making the mapping process more tractable, *(ii)* likely to change over much longer time scales, again making the mapping process more tractable, and *(iii)* a foundation for understanding other Internet-related measurements.

To produce a comprehensive map of the physical Internet we use search to identify infrastructure maps and other repositories that are published online by service providers. Unfortunately, the maps that are identified have no consistent format. For example, some are images, while others are embedded in Flash applications; all are given with a variety of geographic details. Geographic accuracy is aided by the fact that many service providers list street addresses of the locations of their POPs, and listings of the same street addresses from multiple service providers (indicating a third party hosting facility) increases confidence in the overall map. Some

of our data collection and entry process are automated, however others must still be done by hand (*e.g.*, verification). The current Atlas repository includes over 25K PoP locations and 27K links for over 1200 networks (including all tier-1 providers) around the world. A snapshot of the Atlas portal showing all the identified POPs can be seen in Figure 3.1. Published maps of service provider networks are periodically checked for updates, and new maps are being added to the repository on an on-going basis.

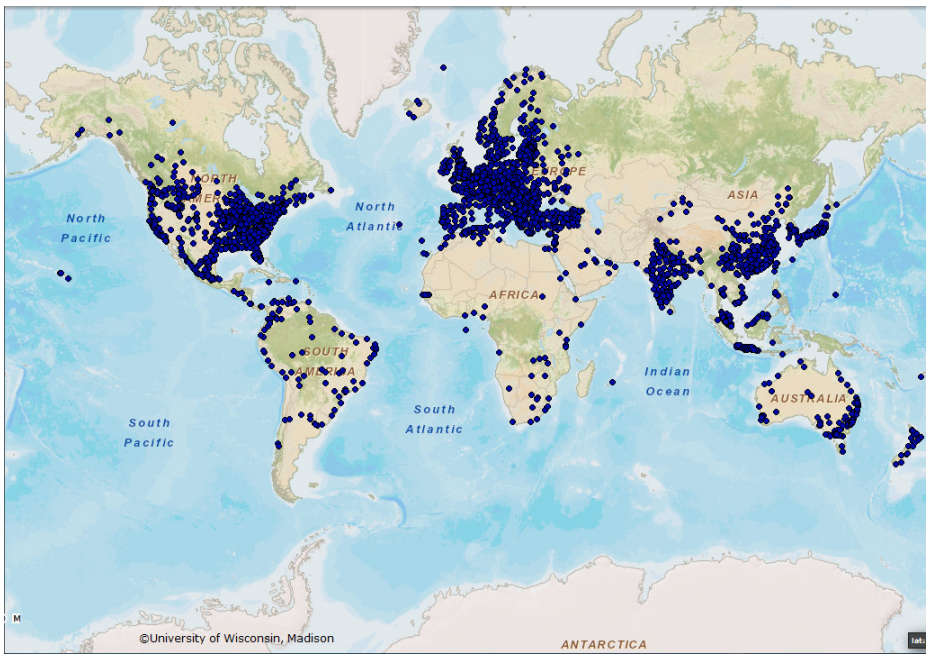

Figure 3.1: Map of the locations of Points of Presence that are currently available in Internet Atlas.

Our second focus is to use the representation of the physical Internet as the starting point for assembling and associating additional data types. Specifically, we seek to use geographic anchors to tie together diverse data types including Internet measurements, social media, public infrastructure, weather reports, *etc.* We argue that this combined representation offers a unique perspective and opportunity to address a broad set of

research and operational problems. We have implemented a variety of interface mechanisms that enable diverse data to be automatically pulled into the Atlas repository. The current Atlas repository includes BGP updates, weather service, hurricane and earthquake data, Twitter feeds, US census data, Dshield attack data, US highway and rail maps. Additional data types are being added on an on-going basis.

The Atlas repository is implemented on top of the widely-used Geographic Information System, ArcGIS [50], which includes an object-relational database that is purpose-built for data that is geographically organized. This design choice was motivated by the robust visualization and geo-analytic capabilities that are available in ArcGIS, the availability of a large number of third party geo-coded databases (*e.g.,* census and infrastructure data), and our desire to be able to flexibly overlay and analyze multiple data types on top of our physical Internet representation.

Our third focus is to enhance the utility and flexibility of the repository through a public web portal. Beyond the interface to ArcGIS, requirements for the portal include broad compatibility, high performance, ease-of-use, extensibility, security, auditing, and user based access control. To satisfy these requirements the portal is implemented in Java using the Spring Framework [131]. We also use Adobe's Flex SDK [173] to build rich, dynamic applications on top of the ArcGIS server. The portal enables users to visualize all of the data in a layered fashion, browse the repository, search the repository, flexibly organize and display aggregates of networks over base maps of the earth. Similar to the repository, the new capabilities are being added to the portal on an on-going basis. The current version of the portal is openly available to the community at [141].

3.2 Internet Atlas Data Repository

In this section, we describe the processes and tools that were developed to build and maintain the Atlas repository.

3.2.1 Identifying Network Data Sources

Internet Atlas is predicated on the assumption that detailed information on network infrastructure can be found on publicly available webpages.[1] This implies that Internet search can be used as the primary data gathering tool. In addition to the major search engines, we used search aggregators (*e.g.*, Soovle [130] and SidePad [129]) to enhance the ability to find network maps such as the ones shown in Figure 3.2.

Our search objective is to find any and all maps/listings of Internet infrastructure. Several important lessons were learned through extensive trial-and-error exploration of relevant search terms. For example, simple one-word terms, such as "co-location" or "datacenter", resulted in discovery of very few previously unseen networks/locations, while multiple word phrases, such as "co-location facility" or "telecom hotels" were more productive.[2] The most important lesson learned is that geographic specificity in search terms is extremely important in revealing regional and local providers. While this may seem obvious, it is complicated by the vast number of local service providers that are only concerned with last mile connectivity.[3]

[1] Such information is often available from ISPs since it aids in their sales and marketing efforts.

[2] The current search term library is entirely in English. Moving beyond English is an objective in future work.

[3] Mapping last mile connectivity is a future goal for Internet Atlas.

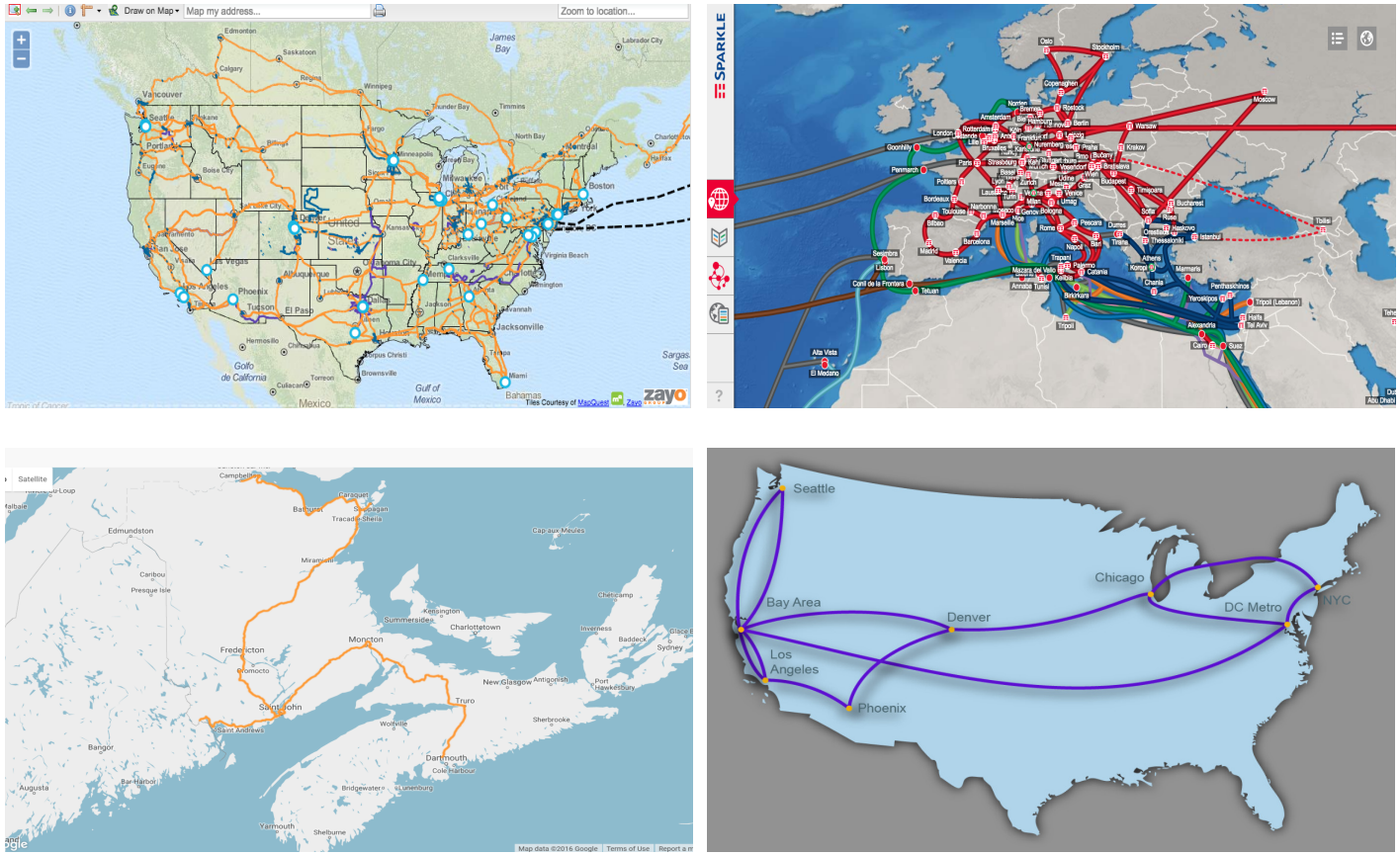

Figure 3.2: Example network maps found using web search. Maps shown belong to Zayo (upper left), Telecom Italia (upper right), F6 networks (lower left), and Layer 42 (lower right). Images courtesy of individual ISPs.

In addition to Internet search, we appeal to the large number of existing Internet systems and publicly available data they provide. This includes PeeringDB [146], Network Time Protocol (NTP) servers, Domain Name System servers (DNS), listings of Internet Exchange Points (IXPs), Looking Glass servers, traceroute servers, Network Access Points, *etc.* Beyond their intrinsic interest, it is important to recognize that NTP servers [105] often publish their Lat/Lon coordinates and are typically co-located with other networking/computing equipment. Similarly, DNS servers routinely publish their location via the LOC record [6]. In total, over 4,700 network resources of various types are annotated in the Internet Atlas database.

3.2.2 Transcription of Network Data

Once a target network has been discovered via search, we transcribe the information to Atlas' GIS database. This is complicated by the varying data formats used by each provider. Network maps can range from images (such as the Sprint Network Map [132]), to interactive maps (such as, the Flash-based AT&T Map [16] and the Google Maps-based Level3 Map [95]).

Visualization-centric representations often reveal no information about link paths other than connectivity (*e.g.,* line-of-sight abstractions are common). For these we enter the network adjacency graph by hand into Atlas. However, some maps provide highly detailed geographic layouts of fiber conduit connectivity (*e.g.,* Level3 [95]). We transcribe these, maintaining geographic accuracy, into the Atlas using a process and scripts that *(i)* capture high resolution sub-images, *(ii)* patch sub-images into a composite image, *(iii)* extract a network link image using color masking techniques, *(iv)* project the link-only image into ArcGIS using geographic reference points (*e.g.,* cities), and *(v)* use link vectorization in ArcGIS to enable analysis (*e.g.,* distance estimation) of the links.

Node locations in primary source data are provided in four forms: Lat/Lon, street address, city or state. If none of these location types is pro-

vided, then the node is not entered into the repository. All node locations are geo-coded into the Atlas repository as a Lat/Lon, while maintaining the source information as meta data. If a Lat/Lon for a network resource is provided, that is transcribed directly into the repository. If a street address for a resource is provided, that address is translated into a Lat/Lon using ArcGIS' inherent capabilities. If only a city/state location is provided, then that is translated into the Lat/Lon of the city/state center if no other more specific addresses for network infrastructure are available in that city/state. Otherwise, the Lat/Lon of the location in that city/state that has the most references from other networks is used. While clearly this could be inaccurate, we believe it is likely to be more accurate than simply leaving the location in the city/state center.

Provider maps often contain additional information about network node resources. This information can range from location (potentially down to Lat/Lon coordinates), to IP addresses, to resource or service types. Our ability to extract network node information from the discovered resources is dependent on an assembly of scripts that include Flash-based extraction and parsing tools [60], optical character recognition parsing tools [110], PDF-based parsing tools [116], in addition to standard text manipulation tools. This library of parsing scripts can extract information and enter it into database automatically. For instances where none of the tools or scripts are successful on the provider data, we manually parse and enter the data.

As an example, consider the map of the Layer42 [92] network, which is show in Figure 3.3 (upper left). To extract the node, link and text information from the image, we convert it to CMYK color model and consider the *yellow* (third) and *magenta* (second) channels to identify nodes and edges respectively. If we then consider nodes and edges as connected components, we can construct a graph simply by identifying edges that touch a given node. We can then convert the input image to grayscale to

find the text embedded in the image.

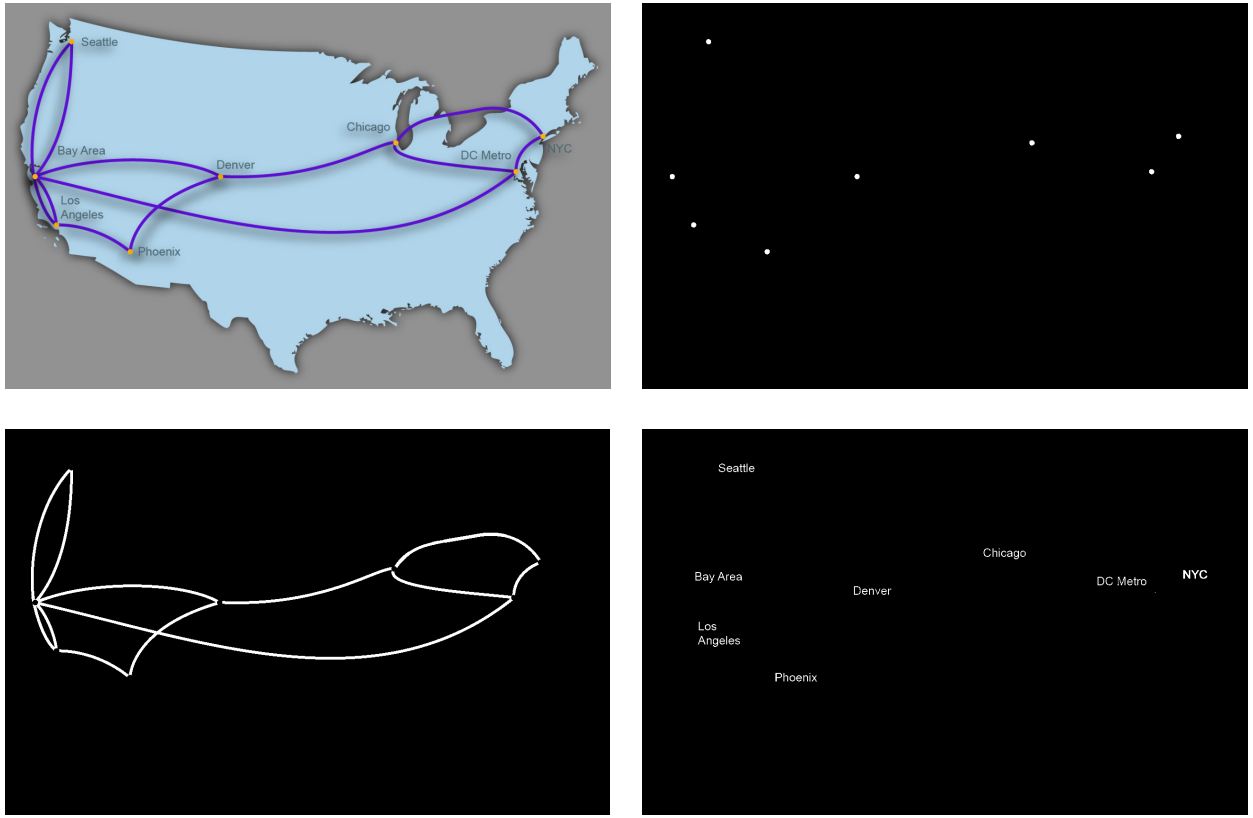

Figure 3.3: Original map of Layer42's network [92] (upper left). Nodes (upper right), edges (lower left) and text (lower right) extracted by the Atlas automated parsing framework.

3.2.3 Verification and Updating of Network Data

Despite our efforts to automate the data transcription process, we still have to enter data manually from time to time. Thus, we must account for human error in this process. We employ several completeness and consistency checks to verify that that manually entered data accurately reflects the primary source data. These checks include having someone other than the person who entered the data manually compare it to the source data.

It is also important to recognize that network footprints will change periodically *e.g.,* a POP will move from one hosting facility to another. Our goal is that the network maps in Atlas reflect current versions of primary source data. To that end, we periodically refer to previously identified primary sources and compare the current online information with the current representation in Atlas. Simple hashes of images enable potential changes to be identified. For network maps that have been automatically transcribed, we can use their vectorized images in Atlas as the basis for automated comparison.

Let the graph generated from the data in the Internet Atlas be $G_{atlas} : \{nodes_{atlas}, edges_{atlas}\}$. Let $G_{source} : \{nodes_{source}, edges_{source}\}$ be the new graph extracted from the primary source. Given the two graphs, verification is checking if G_{source} and G_{atlas} are isomorphic. To facilitate this analysis, we use the proximity analysis capability available in ArcGIS. While this method is not infallible, it is useful for identifying changes.

Finally, it is important to distinguish between *verification* and *validation*. The former is meant to ensure the accuracy of the transcription process from source materials and that Atlas reflects the latest maps available from a primary source. We have taken significant steps to ensure the data in Atlas is verified. The latter is meant to ensure that the data reflects physical reality. It is entirely beyond the scope of our work to actually visit

all locations, and we believe that this is an unreasonable standard. It is also clear that the current published maps may not reflect the most recent changes in a network. However, since we are using source information provided by ISPs (typically providing street or at least city addresses), we argue that there is already a high level of confidence in the data. Further, by virtue of the fact that *many* locations for PoPs appear in *multiple* networks, this is an further and important measure of self-consistency that increases confidence in the repository. We also avail ourselves of other data sources such as the CAIDA [188] and PeeringDB [146] repositories as means for additional consistency checks.

3.2.4 Precision of Network Data

The *precision* of the data in the Atlas repository is defined by the specificity of the locations in the primary source data. As noted above, nodes are entered into the repository based on the availability of at least one of four location types — Lat/Lon is the most precise and state name is the least. We use a separate column in the Atlas database to capture the precision of the data source. We follow a simple heuristic to assign values to each node. If the source data for a node has a Lat/Lon or street level address, then we assign a Source Score of 1.0. If the source data has only a city name, then we assign a Source Score of 0.75. If the source data has only a state name, then we assign a Source Score of 0.5.

We use a similar designation for the precision of paths. While we cannot verify that maps that provide highly detailed geographic layouts of fiber conduit connectivity are accurate, we posit that such detail would not be provided if it were not accurate. Furthermore, visual inspection of many of these maps indicates that the paths typically follow major road/rail infrastructure, which is know to be common practice. This further enhances our confidence in the accuracy of the data.

3.2.5 Completeness of Network Data

While it is likely to be impossible to identify and include the locations of *all* buildings that house Internet infrastructure and the links that connect them, we try our best and add new networks to the repository in an on-going basis. Furthermore, we frequently conduct audits and update the maps to reflect ISP mergers, acquisitions, etc.

3.2.6 Real-time Data Feeds

A key feature of Internet Atlas is the ability to include and visualize real-time data feeds in the portal. The current set of real-time data feeds include Twitter, BGPmon [308], and weather and disaster reports from NOAA and FEMA. Our intention in including this selection of data feeds is to demonstrate the flexibility and extensibility of our platform and to enable Internet Atlas to be used, for example, to analyze the risks of Internet infrastructure to natural disasters.

We use ArcGIS tracking server [51] as a tracking engine for the real-time data. The key capability of the tracking engine is to integrate real-time data feeds with the other geographic data in the repository. With the tracking server, one can receive data in any format and distribute it to clients, perform filtering and generate alerts based on attributes of data or spatial positions, and log the data for later use.

Our primary task in extending the repository to include real-time data was to develop scripts to connect the tracking server to the various remote APIs. This is typically straightforward, but requires special capabilities for each data feed. For example, our interface to Twitter captures all tweets that include some mention of a network-related event. To that end, we have implemented a targeted dictionary of terms and phrases that we look before we archive a tweet (*e.g.,* hurricane, disaster, network failure, poor connectivity, topology change). We rely on Twitter's reported

coordinates for geo-coding understanding that these may not be accurate. Our approach is similar for the other data feeds.

3.2.7 Static Data

One of the important benefits of developing Atlas on top of ArcGIS is that we can immediately take advantage of a wide variety of geo-coded data sets that are available for that system. To demonstrate this capability, we currently include road and rail infrastructure in the US along with population data from the US census. This data can be used, for example, to assess the density of Internet infrastructure in different metropolitan areas or rights-of-way for future fiber conduits.

We also include daily downloads of Internet attack data from DSHIELD.org. DSHIELD data includes the source IP addresses of attacking nodes from over 1700 networks world-wide [309]. We currently use MaxMind to geo-code the locations of attacking nodes. While the accuracy of IP geo-coding is a subject of on-going work, our intention is simply to demonstrate the capability of Atlas in the area of security.

3.3 Internet Atlas Portal Implementation

In this section, we describe the implementation of Internet Atlas, including the web portal, the ArcGIS-based visualization and analysis engine, and the network information database.

3.3.1 Internet Atlas Web Portal

Our overall goal in developing a web portal for Atlas was to provide a visualization and analysis environment that would make the underlying repository useful for both research and operations. Specific technical goals

included making the portal highly responsive, scalable, extensible and secure.

To accomplish these goals, we implemented the portal in Java and JSP using the Spring Framework [131], which enables access from any web browser and enables declarative, annotation-driven support for transactions and caching. Similarly, Spring Security provides a highly customizable authentication and user access-control framework. We also use Adobe's Flex SDK [173] to build dynamic applications on top of the ArcGIS server.

The primary challenge that we faced in developing the web portal was to balance the ease-of-use and flexibility, with security for the underlying data. We addressed that challenge by leveraging the powerful collection of performance and authentication APIs provided by Spring Framework. Every critical aspect of the portal has been modeled using interface-driven design principles to ensure flexibility and security.

Internet Atlas runs on an Apache Tomcat 6.0 web server installed on a Intel-based Red Hat Enterprise Linux machine.

3.3.2 Visualization via ArcGIS

We use ArcGIS Version 10.4 [50] as visualization engine for the data in our repository. ArcGIS enables base maps of the earth or other infrastructures (*e.g.,* roads or rails) to be displayed in layers below individual network maps or aggregates of network maps. The system also provides basic spatial navigation, which is critical for examining maps at multiple scales. In addition, the use of ArcGIS allows for fine-grained control over the data access (*e.g.,* in contrast to Google Maps) and the ability to use openly available data processing and spatial analysis tools that have been written to the ArcGIS standard.

ArcGIS is accessed by the web portal through Adobe's Flex (see Figure 3.4 for example). The challenges in developing this capability were

primarily in creating and deploying a custom GIS web-mapping application that supports data display, interactive querying, and geocoding. We addressed these challenges by leveraging the extensible widget programming model supported by Adobe's Flex.

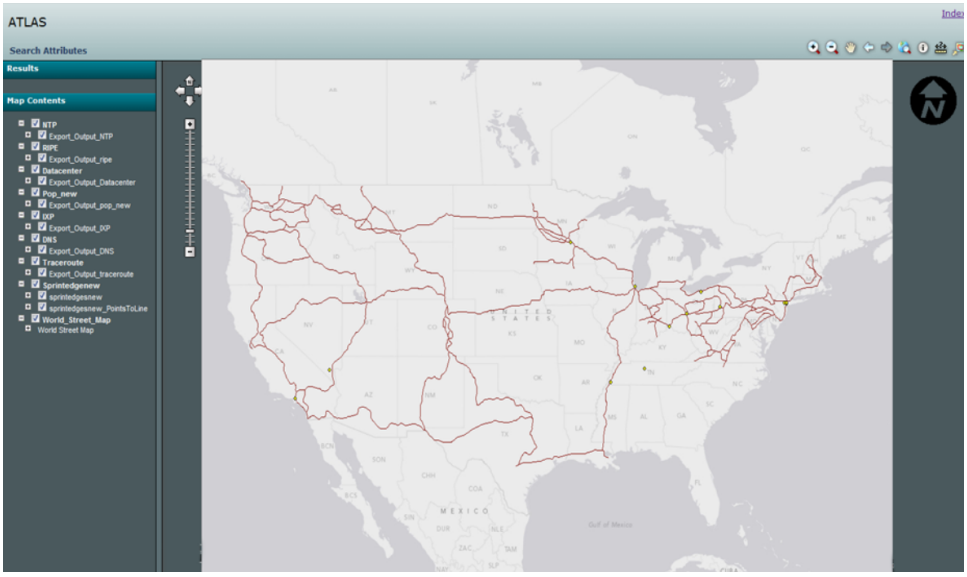

Figure 3.4: Example visualization of network map using ArcGIS.

3.3.3 Internet Atlas Database Structure

The ArcGIS system references our database repository, which contains the physical network data, real-time feeds and static data. The repository also contains meta-data, including source provenance (*e.g.*, where and when data was acquired), IP address ranges associated with ISPs, the perceived confidence in the observed data, and other relevant information on the service provider's network. We also archive snapshots of all network maps and data.

The backend data repository is based on MySQL. The repository contains 16 tables that captures the various source information described

above. For each network resource, we hand-enter the information found in Table 3.5, and for the network connectivity information we hand-enter the information in Table 3.6. The current size of the repository (excluding the real-time data) is ~2 GB.

3.4 Summary

Internet Atlas is a visualization and analysis environment that is based on geographically accurate maps of the physical Internet. We assemble this physical representation of the Internet using search to identify maps and other repositories of the locations of buildings that house networking equipment and the conduits that connect them. We have developed a substantial collection of scripted tools that automate many aspects of identifying, verifying, collecting and transcribing maps of physical Internet infrastructure into the Atlas database repository. The Atlas data repository is available through an openly available web portal. This system is based on the widely-used ArcGIS, which includes a large set of built-in tools that enable a broad set of spatial and statistical analyses. To further increase the utility of Atlas, we extend the repository with relevant, related real-time feeds (*e.g.*, BGP updates, Twitter feeds, and weather data) and static data (*e.g.*, DSHIELD logs, road/rail infrastructure).

Table 3.5: Summary of Internet Atlas node information.

Property	Description	Example
Name	Human readable name	*Deutsche_PoP_Palo_Alto* to "Deutsche Telekom, Palo Alto, California"
IP Address	IP Address of node	192.168.0.1
Location	Best acquired geographic location	*25 Broadway, New York, NY* or *Chicago, IL*
Lat/Lon	Geographic coordinates location (if available)	43.4N 89.24W
Confidence	Indicates source of network data	"1" complete confidence, "0.5" questionable
Accuracy	Resolution of network geolocation information	"1.00" street level, "0.75" city/region, "0.5" country
Node Type	Classification of network resource type	NTP, traceroute server, Looking Glass server
Last Update	Most previous change to ATLAS data	11/10/2011
Metadata	Source of network information	Network provider URL
Network Name	Provider network name	AT&T, Level3, Deutsche Telekom
Tier	Provider network classification	Tier-1, Regional, Metro

Table 3.6: Summary of Internet Atlas edge information.

Property	Description	Example
Node A	Human readable name	*Deutsche_PoP_Palo_Alto* to "Deutsche Telekom, Palo Alto, California"
Node B	Human readable name	*Deutsche_PoP_San_Jose* to "Deutsche Telekom, San Jose, California"

4

Layer 1-Informed Internet Topology Measurement

4.1 Introduction

In this chapter we investigate the hypothesis that physical maps can be used to guide and reinforce the process of collecting layer 3 probe data toward the goal of expanding the scope of physical infrastructure captured in network-layer maps. This conjecture leads directly to two key research questions: *(i)* how do physical layer maps compare and contrast with network-layer maps? and *(ii)* how can probe methods used by projects like Ark [143] be improved to reveal a larger portion of physical infrastructure? We contend that some of the challenges inherent in generating maps from layer 3 probes can be overcome by using the constructive approach of first identifying key infrastructure (POPs, etc.) and then identifying nodes (identified by disambiguating IP addresses or using DNS names) that reside in those locations.

Our study begins by considering physical map data from Internet Atlas project and network-layer map data from Ark project. We focus specifically on infrastructure in North America. At the time of this study, we used physical map data from 78 Internet service providers with over 2600 nodes and over 3580 links. Nodes in the Atlas data refer to hosting centers or points of presence (POPs), with links referring to physical connections between those locations. We use Ark measurements collected

from September 2011 to March 2013 (approximately the same period over which the Atlas repository was assembled). We resolve the IP addresses from this corpus to DNS names and then use location hints to associate these with physical locations (*e.g.*, cities), which becomes the basis for our comparisons.

Several characteristics are immediately evident in the data. Most prominent is the fact that among the 50 networks that are the focus of our comparison study, we observe many more nodes and links in the physical maps. There can be a number of explanations for this observation, including *(i)* the limitations of exploiting DNS naming conventions, *(ii)* the use of tunneling protocols (*e.g.*, MPLS) or the lack of layer 3 services which can render nodes invisible to probes, *(iii)* the limited perspective of the network mapping infrastructure and *(iv)* the fact that layer 3 routing configurations may simply obviate the ability to observe all networks, nodes and links. This supposition is supported by the observation that all Ark probes are confined to a minority subset of networks, with the majority of probes traversing an even smaller subset of networks. Despite this, there are still some nodes/locations/links that appear in the network-layer map but are not indicated in the physical map. This can be explained by physical maps that are out of date or are either intentionally or erroneously incomplete.

The differences between the physical and network-layer maps suggests opportunities for *reinforcement* between the data sets. First, networks observed in Ark that do not appear in Atlas offer clues for searching for new maps that would expand the repository. Second, nodes or links in Atlas that do not appear in Ark can become targets for additional probing that could expand the scope of resulting network-layer maps, thereby making them more useful in target applications. We focus specifically on the possibility of identifying new nodes in layer 3 measurements through targeted probing in the second component of our study.

We define the *targeting problem* as identifying source-destination pairs

for layer 3 probes that reveal nodes indicated in the physical maps[1]. Probing sources (or Vantage Points—VPs) are publicly available infrastructure such as looking glass and traceroute servers and PlanetLab nodes from which probes can be sent. Destinations are simply IP addresses that may respond to probes. We began our targeting analysis by identifying a subset of 596 POPs from the physical maps across 25 networks as our target set. We then conducted extensive probe-based measurements using 266 unique sources and 742 destination addresses in the target networks using two core ideas: *(i)* source-destination pairs should be proximal to the target geographically and in address space, and *(ii)* verification of measurements using multiple sources is required. We verify the identification of infrastructure using location hints in DNS names and using records available in PeeringDB [146]. Our analysis shows that probing between sources and destinations that are both *within the same autonomous system as the target(s)* reveals the most physical infrastructure.

The results of our targeting experiments motivate a new heuristic algorithm for probe targeting that we call *POPsicle*. We show that POPsicle finds 2.4 times as many nodes as are identified by Ark. We compare the number of POPs found by POPsicle with POPs found using Rocketfuel [292] and in all cases POPsicle performs better. We also found that IXPs play a critical role in the way probes traverse a given network. Specifically, sources that are co-located with IXPs have the advantage of appearing—from a layer 3 perspective—as being internal to any/all of the networks that are connected at that location. Thus, a single source that is co-located within an IXP may enhance the identification of infrastructure across all networks that connect to the IXP. This has the effect of significantly broadening the scope of the infrastructure that can be identified using our approach. To validate this idea, we deployed POPsicle at the

[1] *Efficient targeting* is a related problem that seeks to identify infrastructure with a minimal number of probes. We do not directly consider minimizing probe budget in this study.

Equinix IXP in Chicago, USA, and measured the number of POPs for 10 ISPs and found that POPsicle reveals almost all POPs compared to Atlas and extra POPs (in certain cases) compared to Ark. We also find through a case study of Cogent network that POPsicle identifies over 90% of the nodes identified in Atlas or by the reverse DNS technique of [215], compared with about 65% of the POPs identified through Ark, and only 25% identified in the most recently available Rocketfuel data.

4.2 Datasets

In this section we describe the datasets used in our study. One of the key contributions of our work is the comparison of physical topology data from primary sources and network-layer topology data extracted from layer 3 TTL-limited probes, as described below. In the case of physical infrastructure data, we use the latest maps from service providers collected as part of Internet Atlas project [206]. For network-layer topology data, we rely on traceroute data collected as part of CAIDA's Archipelago (Ark) project [143].

4.2.1 Physical Topology Data

In this study, we rely on the publicly available physical topology data from the Internet Atlas project. From Atlas, we obtain detailed geographic information of 7 Tier-1 networks and 71 non-Tier-1/regional networks with a presence in North America consisting of 2611 POPs and 3588 links. Though there is no guarantee as to their timeliness or completeness, we use this data as ground truth of service provider infrastructure in this study.

4.2.2 Network-layer Topology Data

We seek to improve the state-of-the-art in Internet topology mapping by investigating structural characteristics revealed by layer 3 probes. Our goal is to broaden the understanding of Internet topology by investigating how topological characteristics as revealed by layer 3 traceroute probes compare to and contrast with physical structure derived from service provider maps.

The network-layer probe data that we use are collected as part of the Ark project, and include traceroute measurements from a set of 77 monitoring systems distributed around the globe to all routed /24 prefixes in the IPv4 Internet. We used the traceroute data gathered by the Ark project since it represents a canonical system for large-scale Internet topology measurement. We note, however, that the measurements collected in Ark are subject to a variety of network management policies, including blocking or limiting responses to TTL-limited probes, routing configurations and MPLS tunnels, each of which can limit the scope of the measurement data.

Ark is the canonical example of what we might call a *generalized topology probing system*. POPsicle, on the other hand, has a specific goal, which is to discover unique nodes based on guidance from physical maps. Given the difference in goals, the comparisons of the number of unique nodes identified by Ark vs. POPsicle should be interpreted as a comparison between a generalized and a purpose-built system, *i.e.*, POPsicle can be implemented as an extension to Ark or as the basis for designing an entirely new *coordinated large-scale traceroute-based topology measurement system*.

4.2.3 DNS Data

The DNS data we use are also collected as part of the IPv4 Routed /24 DNS Names Dataset [139], and provide fully-qualified domain names

for IP addresses seen in the Ark traces. In this work, our consideration of network-layer topology data is limited by the scope and placement of Ark monitors. However, that project has taken pains to include a broad spectrum of network types (*e.g.*, research, commercial, and educational networks) as vantage points for their monitoring systems, and it provides a widely-used view of the Internet's topology.

Leveraging *location hints* present in domain names to classify IP addresses into POPs is fraught with challenges as described in Section 4.3. We believe that the accuracy of our results could be improved further either with better techniques of handling DNS naming hints, *e.g.,* using the techniques of Huffaker *et al.* [234] or Chabarek *et al.* [190], or by using non-DNS-based techniques to classify IP addresses to their corresponding POPs [291].

4.2.4 Scope of Comparison Study

In this study, we restrict our analysis of Ark data to a period of 19 months, from September 2011 to March 2013, which is contemporaneous with data collection in Atlas. Our focus is on understanding the composite views of networks offered by both data sets over this period. From each individual traceroute in the source data, we extract all the *internal* network IP addresses and links. That is, after processing each traceroute, there is a corresponding interface list (*e.g.*, IP1, IP2, IP3, IP4, *etc.*) and link list (*e.g.*, IP1-IP2, IP3-IP4, *etc.*). For instance, if the traceroute contains a probe of the form A-B-C-D-E (where A, B, C, D and E are IP addresses), we ignore the end point IP addresses (A and E) and extract only the network IP addresses (B, C and D). The interface list thus contains IP addresses B, C and D, and the link list contains B-C and C-D. We merge all the interface/link lists after removing all the duplicate entries to produce a final list of interface IP addresses and links. We then use the corresponding DNS dataset and join the list of interface IPs to their corresponding DNS entries.

4.3 Data Analysis

In this section, we describe the methods we use to analyze the network-layer topology data. We begin with discussing results from processing the network-layer data followed by associating geographic locations to the network-layer data.

4.3.1 Network-layer Data Analysis

In this section, we describe the two-step mapping algorithm that we use to associate a physical location to the IP address interface list obtained from processing network-layer traceroute data, as described above.

Key Idea. One of the aspects of this algorithm is to translate location-based patterns in DNS names that refer to router interfaces to physical (geographic) locations. The influential topology mapping work of Spring *et al.* [292] used such "hints" in their undns tool as part of the Rocketfuel project in order to infer locations of network POPs. Many network service providers employ naming conventions that include geographically relevant information such as airport codes, city names, or other location information. By exploiting these conventions and developing rules to infer geographic locations from them, we can build a network-layer topology map.

Challenges. Leveraging naming conventions in DNS entries has two important challenges. The first is that these names may be out-of-date or misconfigured, which would lead to invalid geographic inferences. The work of Zhang *et al.* [312] quantified the prevalence such problems and found them to occur infrequently, but to have potentially large impact on topology mapping studies that rely on DNS information. They developed a set of heuristics to avoid such problems, including the detection of POP-level loops within a single provider (which should not occur, assuming that the ISP's intra-domain routing protocols are configured properly). We

also use such techniques in our work to avoid problems with exploiting DNS naming conventions. A second challenge with using DNS entries is that there are inherent ambiguities associated with them, *e.g.*, a single string may be used by two different ISPs to refer to two *different* physical locations. To cope with these problems, a set of regular expression patterns can encode different rules to disambiguate location hints from different providers. This approach was also taken in the earlier undns tool [292]. Table 4.1 shows several example patterns and how they are used to resolve ambiguities in naming.

Algorithm. The algorithm for developing a network-layer map from raw traceroute data takes four inputs:

- *regular expression patterns* to extract the location code from DNS entries. The location code is that part of the hostname that contains location data. For example, for A.B.C.LAX2.D.NET, the location code is LAX, which is the airport code for Los Angeles, CA, USA;
- *mapping codes* [190] to translate location code obtained from DNS entries to physical location (a latitude/longitude pair);
- the *list of nodes* (along with each corresponding DNS entry) obtained by parsing the traceroute data from Ark as described above;
- and the *list of links* obtained by parsing the traceroute data from Ark, also as described above.

Using these inputs, we associate physical locations to the IP addresses in the interface list using the following steps:

- First, we match the domain names against the regular expression location patterns and extract a location code from every entry.
- Next, we translate the location code to an actual physical location using the mapping codes. The result of this second step is that we have location information associated with every interface IP address that has a DNS entry with location hints embedded in it. We also use Team Cymru's IP-to-ASN mapping service [137] to classify the list of nodes and links

into different ISPs based on the Autonomous System (AS) Numbers.

Table 4.1: Examples of regular expressions used for extracting location hints from DNS entries.

Regular expression	**Explanation**
/\.(birmingham)\d*\.(level3)\.net$/i	birmingham could refer to a city like Birmingham, UK, but means Birmingham, AL, USA to Level3.
/\.(manchester)\d*\.(level3)\.net$/i	manchester could refer to a city like Manchester, NH, USA, but it means Manchester, UK to Level3.
/\.(mad)\.(verizon\-gni)\.net$/i	mad could refer to a city like Madrid, Spain, but it means Madison, NJ, USA to Verizon.
/\.(ham)\d*\.(alter)\.net$/i	ham could refer to a city like Hamburg, Germany, but it means Hamilton, Canada to alter.net
/\.(cam)\-bar\d*\.(ja)\.net$/i	cam could refer to a city like Cambridge WI, USA, but it means Cambridge, MA, USA to bbnplanet, and Cambridge, UK to ja.net

At the end of applying this algorithm we have network-layer maps for different autonomous systems in which the nodes refer to geographic locations of POPs, and links refer to the fact that packets can be forwarded between a pair of POPs. Note that we do not consider intra-POP links, or individual routers in POPs. The result is that we have a network-layer map that can be equitably compared with the physical map available from Internet Atlas.

4.3.2 How are POPs in the Same City Identified?

To identify POPs located within the same city, we leverage three types of information: *(1)* Personal email communication with network operators and administrators who run the ISPs, *(2)* IP address allocation information from publicly available databases like PeeringDB, and *(3)* naming conventions recorded from ISP websites. In what follows, we give a list of examples for all three cases.

- Tinet (now Intelliquent) has multiple POPs at multiple cities. To identify those POP locations, we contacted one of the network operators [209] from Tinet and identified the naming convention followed by them — the first three letters are city code, and next digit is location code. For instance, ams10 and ams20 are two different POPs in Amsterdam.
- Another reliable source of information that is frequently updated and maintained by network operators is PeeringDB. Apart from providing the list of peers at a particular facility (or an IXP), PeeringDB also provides information like address space allocation, network operator contacts, etc. For instance, GTT has multiple POP locations in New York. One of them peers at NYIIX and has 198.32.160.0/24 as its address space, and one another POP peers at Coresite NY with 206.51.45.0/24 as its address space.
- ISPs routinely publish their naming conventions in their websites along with inter-city POP details. For instance, Lumos Networks and Atlantic

Metro Communications publicly list all inter-city POP naming conventions [170, 171].

4.3.3 Associating Geographic Locations with Traceroute Data

We first provide details on results from processing the traceroute data used for building network-layer topologies. Over the 19 months of Ark data considered in our study, we identified 14,593,457 unique interface IP addresses, comprising 31,055 unique ASes. On these traceroute measurements, we applied the algorithm described above to construct network-layer topologies for comparison with the physical networks chosen for our study. Table 4.2 shows several statistics resulting from applying our algorithm.

Table 4.2: Basic results from processing 19 months of Ark traceroute data using the algorithm described in Section 4.2

Total traceroutes processed	2,674,959,041
Number of unique interface IP addresses	14,593,457
Number of unique ASes	31,055
Valid DNS entries found	6,936,146
No associated DNS name found	7,657,311
DNS entries with location hints	704,935
Number of ASes with at least one geographically identifiable interface address	4,135

As shown in Table 4.2, there were a number of situations in which we could not reliably use the traceroute data for building network-layer topologies. In particular, over 13M IP addresses did not have an associated DNS name with any (obvious) location information embedded in

it[2], which represents 95.16% of all IP addresses observed in our data. Of these, over 6M were unusable because of DNS resolution failures, *e.g.*, fail.non-authoritative.in-addr.arpa, which represented 40.31% of all IP addresses observed in our data. While these results certainly limit our ability to compare physical and network-layer topologies for *all* networks, the remaining "usable" trace information represents 4,135 separate autonomous systems, which we argue still represents a significant slice of the Internet.

An issue we encountered when applying the algorithm of Section 4.3.1 was that, in some cases, there were no associated AS numbers indicated by the Team Cymru IP-to-AS mapping service or available in other whois databases. For such networks, we used a manual keyword search (*e.g.*, layer42.net refers to the Layer42 ISP), which was effective for subnets with at least one associated DNS entry.

4.4 Comparing Layer 1 Maps with Layer 3 Probe Data

In this section, we analyze the physical and network-layer topology data. We begin with comparing the two views of Internet topology by considering how each view intersects and differs from one another, and also how the two views of network topology reinforce each other. We focus our discussion on 50 regional and national ISPs with footprints in North America. We focus on these particular networks because there is significant detail within the Internet Atlas data regarding POPs and inter-POP links for these ISPs.

[2]For example, the DNS naming conventions may not be oriented around physical node location and thus be unusable for our purposes, *e.g.*, entries such as 216-19-195-15.getnet.net and 173-244-236-242.unassigned.ntelos.net.

Table 4.3: Summary comparison of nodes and links observed in physical and network-layer topologies for networks with a footprint in North America.

ISP	Physical		Network-layer		Nodes			Links			N_{Index}
	Nodes	Links	Nodes	Links	∩	Only in P	Only in N	∩	Only in P	Only in N	
AT&T	25	57	39	72	25	0	14	51	6	21	100
Cogent	186	245	122	172	122	64	0	171	74	1	63
NTT	47	216	65	229	47	0	18	189	27	40	57
Tinet	122	132	64	79	57	65	7	79	53	0	37
Sprint	63	102	67	108	63	0	4	98	4	10	54
Level3	240	336	129	237	129	111	0	237	99	0	63
Tata	69	111	0	0	0	69	0	0	111	0	40
Abiline	11	14	8	13	8	3	0	13	1	0	100
Ans	18	25	0	0	0	18	0	0	25	0	94
ATMnet	21	22	0	0	0	21	0	0	22	0	100
Bandcon	22	28	14	22	14	8	0	22	6	0	100
BBNPlanet	27	28	0	0	0	27	0	0	28	0	100
BellCanada	48	65	22	0	22	26	0	0	65	0	56
BellSouth	50	66	0	0	0	50	0	0	66	0	76
BTNorthAmerica	33	76	0	0	0	33	0	0	76	0	85
CompuServe	11	17	0	0	0	11	0	0	17	0	100
DarkStrand	28	31	0	0	0	28	0	0	31	0	96
DataXchange	6	11	0	0	0	6	0	0	11	0	100
Digex	31	38	0	0	0	31	0	0	38	0	97
Epoch	6	7	0	0	0	6	0	0	7	0	100
Getnet	7	8	0	0	0	7	0	0	8	0	100
Globalcenter	9	36	0	0	0	9	0	0	36	0	89
Gridnet	9	20	0	0	0	9	0	0	20	0	100
HiberniaCanada	10	14	0	0	0	10	0	0	14	0	60
HiberniaUS	20	29	0	0	0	20	0	0	29	0	100
Highwinds	18	53	0	0	0	18	0	0	53	0	80
HostwayIntl.	16	21	0	0	0	16	0	0	21	0	94
HE	24	37	23	41	23	1	0	34	3	7	100
Integra	27	36	0	0	0	27	0	0	36	0	74
Intellifiber	70	97	0	0	0	70	0	0	97	0	77
Iris	51	64	0	0	0	51	0	0	64	0	27
Istar	19	23	0	0	0	19	0	0	23	0	84
Layer42	9	12	10	6	9	0	1	4	8	2	100
Napnet	6	7	0	0	0	6	0	0	7	0	100
Navigata	13	17	0	0	0	13	0	0	17	0	100
Netrail	7	10	0	0	0	7	0	0	10	0	100
NetworkUSA	35	39	0	0	0	35	0	0	39	0	34
Noel	19	25	2	0	2	17	0	0	25	0	16
NSFnet	13	15	0	0	0	13	0	0	15	0	92
Ntelos	48	61	0	0	0	48	0	0	61	0	48
Oxford	20	26	0	0	0	20	0	0	26	0	50
PacketExchange	21	27	0	0	0	21	0	0	27	0	100
Palmetto	45	70	0	0	0	45	0	0	70	0	49
Peer1	16	20	0	0	0	16	0	0	20	0	100
RedBestel	82	101	0	0	0	82	0	0	101	0	9
Syringa	66	74	0	0	0	66	0	0	74	0	9
USSignal	61	79	0	0	0	61	0	0	79	0	46
VisionNet	22	23	0	0	0	22	0	0	23	0	23
Xeex	24	34	4	3	4	20	0	3	31	0	96
Xspedius	34	49	0	0	0	34	0	0	49	0	100

4.4.1 Comparison of Physical and Network-layer Nodes and Links

We now compare the physical and network-layer topologies obtained from the Atlas data and the Ark data, respectively. Again, the basic entities we compare are *nodes*, which represent city-level points of presence or data centers, and *links*, which represent physical and/or logical connectivity between two city-level POPs.

Table 4.3 shows the number of nodes and links observed in each topology type, for each of the 50 networks under study. We first see that while all physical networks have non-zero nodes and links, there are some network-layer topologies for which there are zero nodes and/or links observed. There are two reasons for this. First, an interface IP address for a given network may have no clear location information embedded in its associated DNS entry. For example, for 21 out of 50 networks, there were no location hints observable in the related DNS records. This result may be because of non-obvious naming conventions, or simply that there are no name records available. We note that although some ISPs in our list of 50 have been acquired by other companies, the AS number and address blocks assigned to these companies still refer to the original ISP.[3]

The second reason we may observe zero nodes and/or links for a given network is that we may simply not have observed *any* interface addresses for a given network in 19 months of traceroute data. This observation was true for 16 out of the 50 networks included in our study. Considering the fact that the Ark project targets *every* routable /24 in the IPv4 Internet, this is a surprising result. Still, there may be a variety of reasons for this observation. First, some ISPs may configure their routers not to respond to hop-limited probes with ICMP time exceeded messages (resulting in

[3]For example, although BellSouth was acquired by AT&T in 2006, the name BellSouth is still referred to in whois databases and appears in recent address block usage reports (http://www.cidr-report.org/as2.0/).

"asterisks" in the traceroute output). Second, some networks may use tunneling protocols such as MPLS, and configure these tunnels to be completely hidden. Third, there may be interfaces controlled by an ISP under study that are configured with IP addresses from a third party, *e.g.*, an IXP. In the end, we were left with 13 networks that had DNS entries for which we could identify a physical location.

To assess how the physical and network-layer views of a network compare, we consider node and link *intersection*, as well as the number of nodes and links *only* observed in one or other other topology. To determine the intersection, we consider a node to intersect each topology if we identify the same POP location in each one. We consider a link to intersect each topology if there are POPs identified in the same two locations in each topology and there is a link identified between them. For example, if we observe nodes in Chicago and Kansas City in both the physical and network-layer topologies for a given ISP, and a link between those two cities, we say the link and two nodes intersect.

Table 4.3 shows results from the intersection analysis. We also show in the table nodes or links that *only* appear in one or the other topology. We see that, in general, there are more nodes and links observed in the physical topologies than are seen in the network-layer topologies. For the networks for which this observation holds, the number of nodes and links observed is, in some cases, *significantly* larger than those seen using the traceroute data. These results strongly suggest that sole reliance on layer 3 probes to generate physical network maps is likely to result in an incomplete view of Internet topology. On the other hand, the table shows that there are a small number of networks in which we observe *more* nodes and links in the network-layer topology. In particular, we see this for AT&T, Tinet, NTT, Sprint, Layer42, and Hurricane Electric (abbreviated as HE in the table). This observation suggests that while published physical maps *usually* offer an authoritative view of physical infrastructure, the published

maps may lag recent deployments which can be observed through layer 3 probing.

More broadly, analysis of the Ark traces shows that there are at least 448 distinct networks in North America (that are not part of Atlas). This number is identified by first searching for all North American location DNS hints and then identifying unique service providers in the DNS names. This compares to the 320 distinct networks in the Atlas repository, which have been identified through extensive search-based methods. An implication for this difference is that measurements from Ark can be used as guidance for identifying service provider networks that could be included in Atlas. For example, many small/regional networks like Adera Networks (CA), Grande Communications (TX) and Atala T (NY) were found in the traces of Ark and such networks could be incorporated into future search-based campaigns.

Of the 448 distinct networks identified through Ark measurements, the vast majority of probes pass through tier-1 and major ISPs, as shown in Figure 4.4. Thus, while it is likely that the POP-level topology of well-connected ISPs can be largely identified through general probing techniques, smaller ISPs are unlikely to be well-mapped. This observation is supported by prior studies on sampling bias in network topology measurements (*e.g.*, [285]). An implication for this observation is that *targeted probing methods* may be necessary to obtain a more comprehensive topological picture of physical Internet infrastructure.

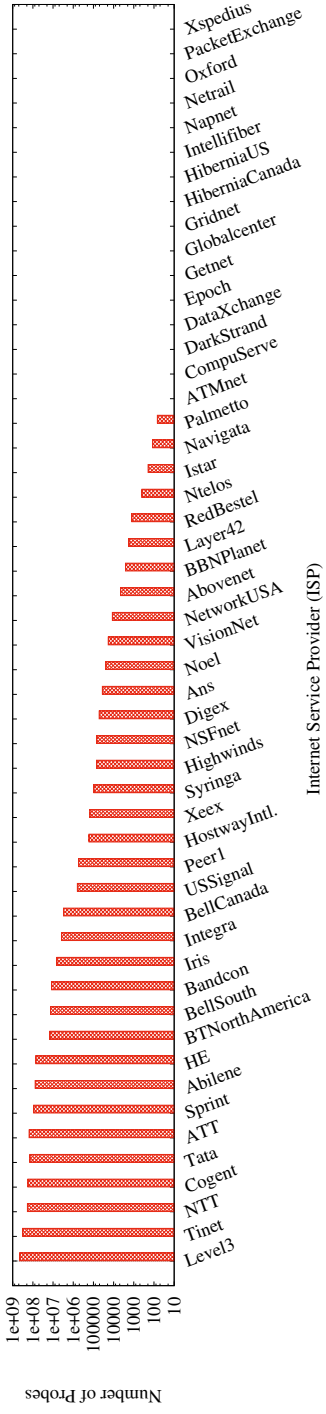

Figure 4.4: Number of probes sent out by Ark across Internet Service Providers

Lastly, we consider one form of *validation* of physical node locations when we observe the same location for nodes in more than one network. We define the metric N_{Index} as the percentage of nodes identified for a given network that have the same physical location as a node in another network. The right-most column of Table 4.3 shows the N_{Index} for each network. The intuition for why this metric provides some level of validation of the physical location has to do with common industry practices of using co-location facilities and telecom hotels. While this observation may not hold universally, we believe that co-location practices are generally observed for small regional networks in geographically isolated areas since the costs associated with setting up new facilities is high. For instance, larger national ISPs like Layer42, Napnet, Navigata and Netrail show an N_{Index} of 100 (complete overlap with nodes in other networks), whereas smaller regional cariers such as NetworkUSA, RedBestel and Syringa[4] show an N_{Index} less than 20 (mostly their own locations). The combination of a high N_{Index} and overlap with traceroute probes provides perhaps the best validation of node locations.

4.4.2 Case study: Tinet

Tinet (now Intelliquent, Inc.), shown in Figure 4.5 represented an interesting special case: the physical topology contained nodes not present in the network-layer topology, and the network-layer topology also contained nodes not present in the physical topology. In particular, there were 65 nodes only present in the physical topology, and 7 nodes that were only observed in the Ark data and network-layer topology. For example, the Tinet physical network map shows four nodes for Amsterdam, Netherlands, one node in San Jose, CA, two nodes in Milan, Italy and two nodes in Washington, DC. However, the network-layer topology revealed additional

[4]NetworkUSA is a regional carrier serving Louisiana, RedBestel operates in the Guadalajara region of Mexico, and Syringa is a regional carrier in Idaho.

nodes for these locations. The missing nodes from the physical network may be due to Tinet's network maps not reflecting the most up-to-date deployments. Missing nodes and links in the network-layer view may be due to a variety of reasons, including the inability to gain a broad perspective on Tinet's network from Ark vantage points. What these results indicate is that to gain a *complete* view of a network's topology, multiple data sources must be considered.

4.4.3 Main Findings and Implications

The main findings of our comparison of physical and network-layer topologies are as follows.

- We observe many more nodes and links in the physical maps, which may be due to a variety of reasons, but is most critically due to the fact that layer 3 routing configurations simply eliminate the possibility to observe all networks, nodes, and links through end-to-end probing. This likelihood is supported by the fact that all Ark probes are limited to a relatively small subset of networks, with the majority of probes passing through an even smaller set of networks.
- There are still some nodes, locations, and links that appear in the network-layer map but are not observed in physical maps. The likely reason is that the physical maps are out of date or incomplete.
- The observed differences between the physical and network-layer maps suggest opportunities for using one to *reinforce* the other. In particular, networks observed in Ark that do not appear in Atlas offer clues for searching for new maps to expand Atlas. Similarly, nodes or links in Atlas that do not appear in Ark can become targets for additional probing in order to broaden the scope of the resulting network-layer maps.

Indeed, in the next section we focus specifically on how to emit targeted layer 3 probes in order to confirm the existence of nodes identified in physical maps, as well as to identify additional physical nodes.

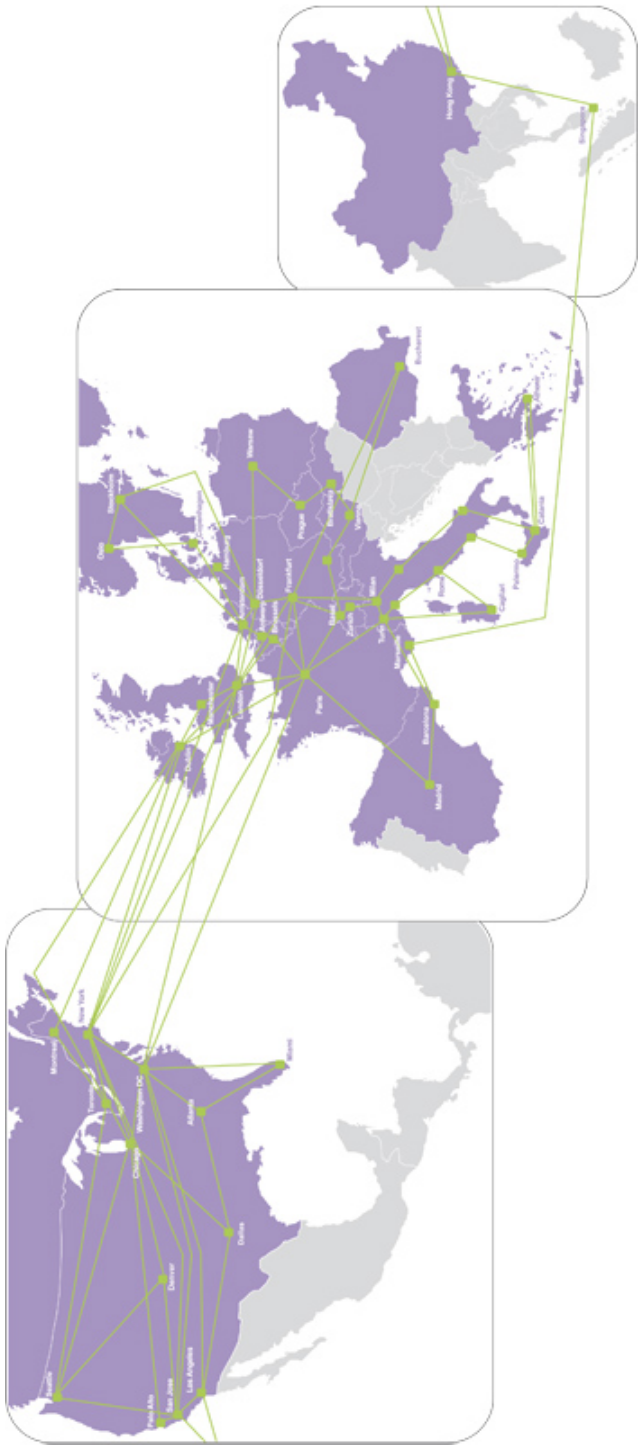

Figure 4.5: Network map of Tinet. Image courtesy of Intelliquent, Inc.

4.5 Effects of Vantage Points on Node Identification

In this section, we examine the effects of source-destination selection on the ability to identify POPs within a service provider using targeted layer 3 probes. Specifically, we examine the differences between using vantage points (probing sources) internal or external to an ISP containing target POP(s), and destinations either internal or external to the ISP. Furthermore, we examine the effects that IXPs may have on probe-based POP identification and how IXP placement may be exploited to aid in node identification by providing a larger set of internal vantage points.

4.5.1 Effects of Vantage Point and Destination Selection

To examine the impact of vantage point and destination IP address selection for identifying all target POPs in an ISP, we leverage publicly available traceroute servers, looking glass servers and Planetlab nodes as VPs[5], and select different combinations of them located within or external to different service providers. In particular, we use three combinations: probing from VPs outside an ISP to destinations inside (denoted VP_{out} to t_{in}), from VPs inside an ISP to destinations outside (denoted VP_{in} to t_{out}) and from VPs inside an ISP and destinations inside (denoted VP_{in} to t_{in}). For each directional modality ($VP_{out} \rightarrow t_{in}$, $VP_{in} \rightarrow t_{out}$, $VP_{in} \rightarrow t_{in}$), we use a greedy approach to identify probe source-destination pairs based on geographic proximity. We choose the VP geographically closest to a target POP, then successively choose from the set of destinations that are also geographically proximal to the target until the target is identified. For

[5]We followed principles established in prior work, *e.g.*, [292], to avoid burdening these public servers with excessive load.

instance, a probe from planetlab4.wail.wisc.edu to 184.105.184.158[6] with the aim to identify Hurricane Electric's POP in Los Angeles identified two additional POPs (in Chicago and Denver) in addition to identifying the Los Angeles POP. If we can not identify the POP from a given vantage point, we choose the next closest VP, and so forth (specific details of this method are provided in Section 4.6).

Using a subset of 25 ISP networks that assign DNS names with location hints and that contain 596 target POPs, we analyze the source-destination combinations. Figure 4.6 shows the fraction of target POPs discovered by these three probing modalities relative to the number of POPs identified in Atlas. The figure shows clearly that the most effective strategy is to send probes from vantage points located *within* an ISP to destinations that are *also within the ISP* (VP_{in} to t_{in}). We further observe that using a VP located within an ISP is more effective than choosing one external to the ISP. We hypothesize that these differences are due to the effects of interdomain versus intradomain routing on probes. In the case of both VP and destination located within an ISP, there is a greater chance for a diversity of paths to be observed due to ECMP, the fact that more information about shortest paths is available, and the greater degree of flexibility that a service provider has in routing packets within its own infrastructure. In the case of either VP or destination being external to the ISP that contains a target POP, interdomain routing protocol effects come into play, such as hot-potato routing and the forced choice of a single best path.

Lastly, we note that in absolute numbers, we observed a total of 188 POPs using VP_{in} to t_{in}, 157 POPs via VP_{in} to t_{out}, and 93 with VP_{out} to t_{in}. For 11 networks we observed zero POPs. Similar to our earlier observations in which we do not see POPs identified in physical maps, this may be due to MPLS deployments, traffic management/routing policies.

[6] lightower-fiber-networks.gigabitethernet4-10.core1.lax2.he.net

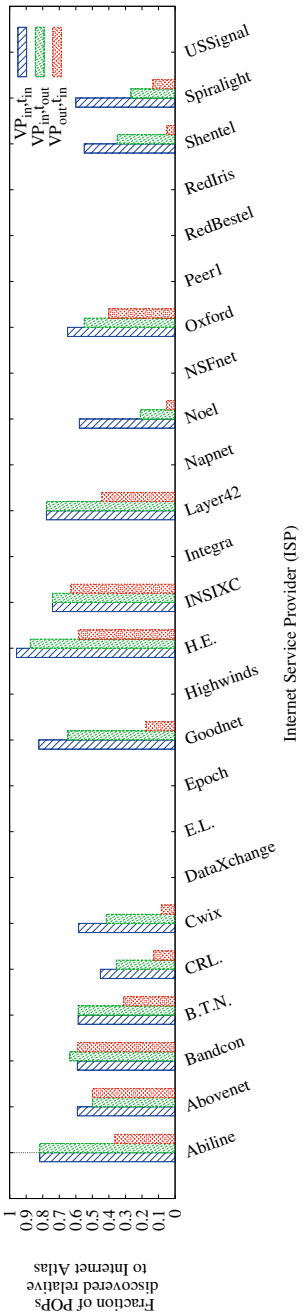

Figure 4.6: Number of POPs discovered by the probing modalities.

4.5.2 Using IXPs to Expand Perspective

Given the result that the most effective probing strategy for identifying physical infrastructure is to choose source-destination pairs that are within an ISP, it is important to recognize that broad deployment of such targeted measurements is inherently limited by the availability of VPs within provider networks. Indeed, the 266 VPs used in this paper are restricted to 248 separate networks, which is substantially less than the total number of networks identified in North America by Ark in Section 4.4.

Recent work in [180] has highlighted the enormous amount of layer 2 peering that is taking place at IXPs. This leads us to posit that VPs co-located with IXPs might be leveraged to dramatically expand our ability to identify physical infrastructure. Indeed, there is anecdotal evidence that much of the rapid growth in peering at IXPs is being driven by local and regional ISPs and that Tier-1 ISPs have been slower to connect [243]. This offers a tantalizing opportunity since it is generally the smaller networks that are more difficult to map and those networks often do not deploy looking glass servers that are necessary for mapping physical infrastructure.

To consider this possibility, we begin by looking for VPs that are co-located with IXPs in North America. We find that 14 out of 65 IXPs have co-located VPs. Using PeeringDB [146] we find that the total number of unique ISPs that peer at these 14 IXPs is 642. A comparison between these ISPs and those in with VPs used in our study shows that an additional 625 unique networks could be measured from these 14 IXPs alone. This suggests that deployment of VPs in other IXPs could be the starting point for comprehensive mapping of physical Internet infrastructure.

4.5.3 Main Findings and Implications

In summary, we consider how to choose sources and destinations for probing in order to identify POPs within a service provider, as well as to discover new POPs. Specifically, we examine whether it is better to use vantage points (probing sources) internal or external to an ISP containing the target POP(s) and destinations either internal or external to the ISP. Our results show that it is best to choose both source and destination to be *within* the ISP that contains the targeted POP(s), which we hypothesize is largely due to intradomain versus interdomain route selection. Furthermore, we observe that co-locating a probing vantage point at an IXP may be particularly useful in that the VP can effectively appear as being internal to all ISPs that peer at the IXP.

4.6 Enhancing Node Identification

In this section, we build on the observations and experiments of Section 4.5 to describe a new targeted probing algorithm called *POPsicle*. We evaluate POPsicle's effectiveness for reinforcing and confirming information available in physical maps. We deploy POPsicle at an IXP in Chicago, and describe results of experiments carried out at the IXP.

4.6.1 POPsicle Algorithm

POPsicle is designed to send traceroute-like probes toward a target with a known geographic location based on information from a physical map. The objective is to detect the target *at the network layer*. POPsicle is based on the insight that vantage points co-located with IXPs can be used to launch probes in many different networks, and that probe-based detection of target physical infrastructure is most effective when both VP and destination are located within the same service provider network.

Algorithm 1 POPsicle algorithm

input: targetNet = target network
input: L^T = list of targets to be identified
input: L^S_{vp} = list of source VPs with known coordinates

```
// Scan target network to find reachable hosts
1 scanResults = scan(targetNet);
  L^D_vp = inferLocations(scanResults);
  foreach t in L^T do
      // Choose destination VPs that are closest to reachable hosts
2     S^t_vp = geographicallyNearest(t, L^S_vp);
      foreach vp in S^t_vp do
          // Greedily choose probing destinations within a cone extending from vp to t
3         D^t_vp = searchCone(vp, t);
          foreach dst in D^t_vp do
4             send probe from vp to dst;
              if t found then
5                 record success for t;
                  goto step 3;
```

Algorithm 1 shows the key steps of POPsicle. The inputs to the algorithm are (1) the name and address prefix(s) of the ISP within which physical targets are to be identified, (2) the specific list of targets (*e.g.*, POPs) to be identified, including their geographic locations according to physical mapping information, and (3) a list of VPs and their known geographic coordinates. The algorithm proceeds by first scanning the target network to identify which hosts are accessible[7]. This step is performed to collect a set of hosts that can be used as probe destinations. The geographic locations of these hosts are then inferred using DNS location hints. Another option at this step would be to use IP geolocation algorithms or tools. However, the accuracy of these techniques is a subject of ongoing

[7]We employ the nmap tool for this step with the command line nmap -Pn -sn prefix. Even though nmap is considered bad, we only did a passive scan without causing any trouble to ISPs

research (*e.g.,* [210]) so we do not use them in POPsicle, but they could be easily incorporated.

Next, POPsicle iterates through the list of target nodes to be identified. For each target, we obtain a list of VPs for initiating probes in step 2, ordered by proximity in Euclidean space (using the Haversine formula [74]) to the target. For each VP, we then select a set of destinations that are also ordered by proximity to the target. Destinations, compiled from a variety of sources like Internet Atlas portal and PeeringDB, are IP addresses of infrastructure, like looking glass servers, traceroute servers, telecom hotels, and other entities that may simply respond to probes. From this set, we sub-select the destinations such that the square of the Euclidean distance between the VP and destination is greater than the sum of the squares of the distance between VP and target and VP and destination. This has the effect of creating a “measurement cone" centered at the VP and directed toward the target node (step 3). These destinations are then iteratively probed using traceroute. For each completed trace we determine whether the target has been found using location hints. If it has, the algorithm completes. If not, we continue until we have exhausted all VPs and their corresponding destination sets. Figure 4.7 depicts the targeting process of POPsicle.

POPsicle is based on the notion that target POPs will be part of routes that connect sources and destinations located on either side (from a Euclidean perspective) of the target. We argue that this is likely due to shortest path intra-domain routing. POPsicle is also currently dependent on location hints from DNS for both destination identification and to identify when a target has been discovered. IP geolocation could be used to address the former, while the latter could be addressed through publicly available data (*e.g.,* PeeringDB) by associating IP address ranges with POP locations.

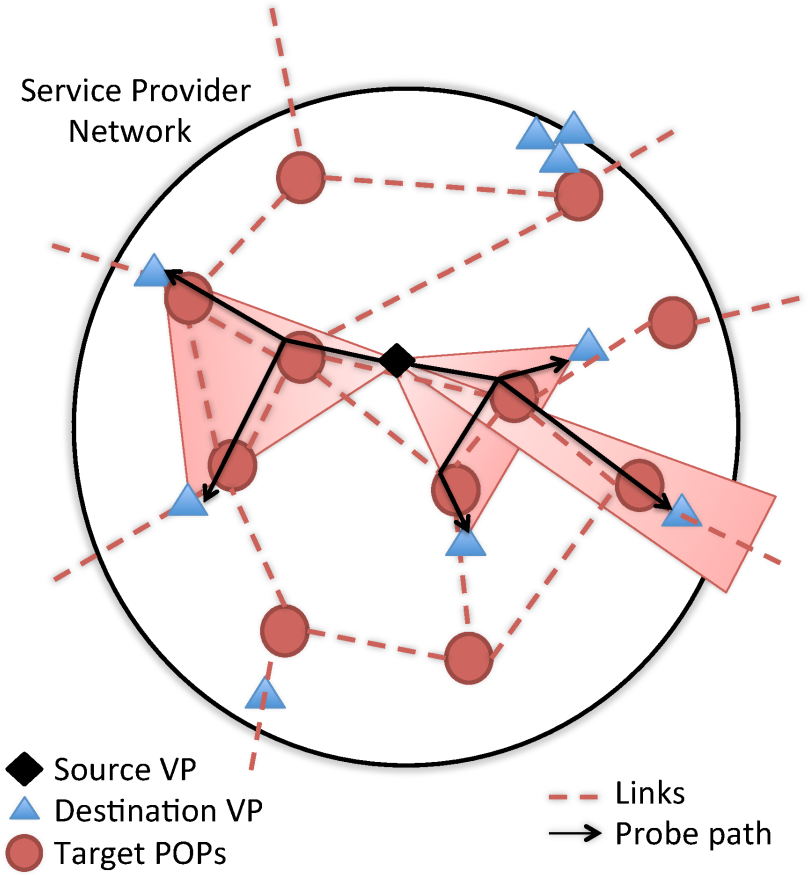

Figure 4.7: POPsicle targeting process. VPs within the ISP that are geographically closest to the target are selected along with destinations that are geographically closest to the target and "on the other side" of the VP.

4.6.2 POPsicle Evaluation

We selected 30 looking glass servers from the Atlas database that satisfied the following criteria: (1) the server is co-located with an IXP in North America, and (2) the ground truth information of the POPs is available either from Internet Atlas or in PeeringDB [146]. The vast majority of providers we selected for analysis are regional providers since we found them to be poorly represented in Ark probing results and thus prime

Table 4.8: Summary results of network POPs identified with POPsicle, Atlas, Ark, and Rocketfuel for POPsicle deployed at publicly accessible looking glass servers.

	POPsicle	Atlas	Ark	Rocketfuel
Abovenet	13	22	13	13
BellCanada	34	48	30	29
Centauri	7	14	3	—
Cyberverse	2	2	2	—
Data102	2	2	2	—
HopOne	4	4	4	—
HE	23	24	23	8
Inerail	3	25	3	—
Internet2	10	10	10	10
Interserver.net	2	2	1	—
Steadfast.net	3	3	3	—
Towardex	7	8	6	—
XO	42	80	42	39

candidates for detailed study. In terms of the number of networks used in this study, the coverage of our technique can appear limited. We had to remove several networks from our study due either to the incompleteness of the physical or network maps, or due to the lack of DNS locations hints.

The selection of these 30 looking glass servers resulted in 13 service provider networks that were the focus for our evaluation. We began by examining Internet2, which we consider a special case since complete ground truth for all the layer1, layer2, and layer3 devices is available [85]. POPsicle-directed probing found 10 out of the 10 POPs in Internet2 that house layer 3 infrastructure.

We initiated probing on the remaining set of 12 ISPs using POPsicle-directed probing to verify and map the POPs for each of those networks. Table 4.8 shows the results from all of our probing experiments. Overall, for 8 out of 13 ISPs, we see all or almost all of the POPs identified in physical maps. These 8 ISPs include Cyberverse, data102, HopOne, Hurricane

Electric (HE), Inerail, Interserver.net, Steadfast.net, and Towardex. For several ISPs, we also observed additional POP locations which we verified using PeeringDB. We also compare with the most recently available measurements from Rocketfuel. Although the Rocketfuel measurements are not especially recent, we note that it is likely that POP deployments are fairly stable. We observe, for example, that POPsicle and Rocketfuel identify the same number of POPs for 3 out of 5 ISPs. Lastly, we note that Rocketfuel data were unavailable for 8 ISPs.

In the following we discuss various special cases and observations related to results for each ISP:

- For BellCanada, POPsicle identified significantly more POPs than were revealed in Ark data. Additional locations identified were in New York, Palo Alto, Seattle, and Woodbridge. We confirmed these locations with Equinix Palo Alto, NYIIX, and SIX exchange points in PeeringDB. The Woodbridge location could not be confirmed in PeeringDB.
- For Centauri Communications, POPsicle identified four additional POP locations in comparison with Ark, including Palo Alto, San Francisco, San Jose, and Sunnyvale. These locations were all confirmed by SFIX and SFMIX in PeeringDB.
- For cyberverse, data102, Steadfast.net, Inerail, Internet2, Hurricane Electric and XO Communications POPsicle identified the same POPs as were observed using the Ark data.
- For HopOne, POPsicle found one extra POP location in Palo Alto (which is not seen in either Ark or physical topology maps), which was confirmed in PeeringDB. POPsicle did *not* observe a node in Mclean, VA, which was seen in the Ark data.
- For Interserver.net, POPsicle identified one additional POP location in New Jersey which is confirmed by Equinix New York IX.
- For Towardex, POPsicle found an extra POP in Boston which is confirmed in PeeringDB (Boston IX).

In addition to mapping POPs of the 13 ISPs described above, we evaluated POPsicle's effectiveness for mapping and confirming additional *infrastructural* nodes that have known/published physical locations. This test set included data centers, DNS root servers, NTP servers (both stratum 1 and stratum 2), and IXPs. Table 4.9 shows results of these experiments, as well as summary results of the POP-identification experiments. We can see from the table that POPsicle is able to identify network-layer locations for this larger and much more diverse set of devices. In total, it finds 1.04 times more POPs, 1.54 times more data centers, 9 times more DNS servers, over 11 times more NTP servers, and 1.48 times more IXPs[8] (in North America) compared to nodes found by standard end-to-end layer 3 probing campaigns. Overall, POPsicle reveals and confirms 2.4 times more physical node locations versus standard probe-based topology measurement methods.

[8]We expect our result to coincide with [180] if we have access to more vantage points.

Table 4.9: Summary of results from mapping infrastructural nodes.

	POPs (13 ISPs)	Datacenters	DNS Servers	NTP Servers	IXPs	Total Locations
POPsicle	149	487	9	627	37	1309
Ark	143	315	1	55	25	539
Atlas	244	641	13	827	65	1790
POPsicle vs. Atlas	61.07%	75.98%	69.23%	75.82%	56.92%	73.13%
Ark vs. Atlas	54.60%	49.14%	7.69%	6.65%	38.46%	30.11%
Improvement	1.04x	1.54x	9x	11.40x	1.48x	2.42x

4.6.3 IXP Deployment of POPsicle

We observe in Section 4.5 that a VP co-located with an IXP can provide what appears to be an *internal* probing source for *any* ISP that peers at the IXP as depicted in Figure 4.10. From such a vantage point, a tool implementing the POPsicle algorithm could be employed to map and identify POPs and other nodes of interest in any one of the adjacent ISPs.

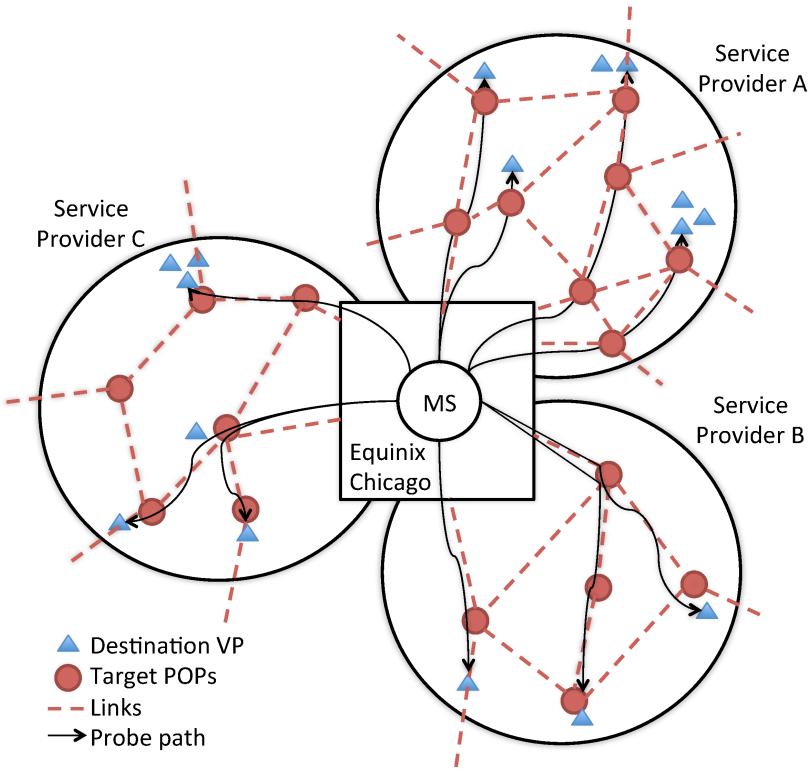

Figure 4.10: Multiplexing an IXP-based measurement server across multiple ISPs using POPsicle.

To substantiate this idea, we deployed a tool implementing POPsicle on a server at the Equinix Chicago Internet Exchange with the help of network operators, and we conducted a week-long measurement study. We chose 10 ISPs that peer at Equinix Chicago for targeted probing. These ISPs were

chosen because (1) there was information available in PeeringDB, or we had operator contacts who could verify our inferences, and (2) location hints were available in DNS for IP addresses within the ISP. Unfortunately, the vast majority of ISPs that peer at Equinix Chicago do not have publicly available ground truth information and/or location hints available via DNS thus we could not include them in this initial study. Also, we note that the ISPs we considered in our POPsicle deployment have, unfortunately, little overlap with the ISPs we consider in Section 7.4 (and which appear in Table 4.8) due to our requirement that we have ground truth and location hint information available — only Hurricane Electric and HopOne are in common.

Table 4.11 shows the results of our IXP-based POPsicle deployment. We observe from the table that POPsicle finds all nodes for 8 out of the 10 ISPs (as compared with the Atlas physical topology data). In the Ark measurements, 6 out of 10 ISPs are fully mapped. For the two ISPs the POPsicle is not able to fully map, a very likely possibility is that the unobserved POPs are invisible to layer3 probes due to configured router policies [255], thus the results we show may be the best that can be achieved through active probing. Overall, our results suggest that POPsicle could be deployed more broadly to accurately map (to the extent possible) ISPs for which we do not have ground truth.

Special Cases. *(1)* The number of POPs found for HE in Table 4.8 is 23 but in Table 4.11 the number of POPs found for HE is 24. That is, POPsicle deployed at Equinix Chicago saw an extra node in Calgary, Canada (YYC) which is verified with Datahive IX. One possible implication of this result is that such probe-based measurements are biased towards the vantage points selected. *(2)* For 2 ISPs (PaeTec and Atlantic Metro) some POPs were not visible to our probes, which we intend to investigate further in future work. There is anecdotal evidence that ISPs typically do not expose certain locations to traceroute probes (or any access methods from outside)

even when layer 3 services are available at that particular location due to security reasons [255].

Case Study: Cogent. In [215], Ferguson *et al.* present an analysis of Cogent Communication's network based on using reverse DNS records, as well as location-based naming hints. We used the dataset made public by these authors to evaluate, compare and validate POPsicle's probe-based measurement of Cogent's network. We processed the DNS names from Ferguson *et al.*'s dataset using the modified version of location inference technique developed by Chabarek *et al.* [190] and identified 187 POP locations. We then used POPsicle deployed at the Equinix Chicago IXP to target routers within Cogent's network, and it identified 173 POPs. In Table 4.3, we see that there are 186 POP locations identified in the Atlas physical topology; it is likely that the additional POP identified in the Ferguson *et al.* dataset is a more recent deployment than was found in Atlas. Also in the Table 4.3, we see that there are 122 POPs identified through the Ark probes. Lastly, we note that in the most recent Rocketfuel data, there are only 45 POP locations identified. Altogether, these results show that POPsicle's probing technique is very effective for discovering locations of physical infrastructure like POPs, is much better than existing probe-based techniques, and nearly as good as exhaustive use of reverse DNS records.

4.6.4 Main Findings and Implications

We describe a new targeted probing technique called POPsicle that is designed to reveal and confirm the presence and location of physical infrastructure such as POPs. To evaluate our method, we used publicly accessible looking glass servers deployed at IXPs, and made a custom deployment of POPsicle at the Equinix Chicago IXP. POPsicle finds 2.4x more physical nodes than Ark probes, and in our custom deployment in Chicago, POPsicle finds nearly all POPs identified in the Atlas physical

Table 4.11: Summary results of network POPs identified with POPsicle deployed at the Equinix Chicago IXP.

ISP Name	POPsicle	Atlas	Ark
BTN	29	29	28
HE	24	24	23
Internet2	10	10	10
PaeTec	54	61	54
Nexicom	9	9	9
HopOne	3	3	3
Indiana Gigapop	2	2	2
MOREnet	4	4	4
Atlantic Metro	9	12	8
Steadfast.net	3	3	3

topologies. In a case study of Cogent's network, POPsicle identified more than 90% of the POPs known through Atlas as well as through the recently described technique based on using reverse DNS records [215]. Moreover, it found many more POPs than Ark probes, or the most recent Rocketfuel measurements.

Overall, our results show that an IXP deployment provides a prime location from which to launch targeted topology discovery probes. Since Rocketfuel maps are commonly used in networking studies that require realistic and representative network topologies, we view this deployment paradigm as having significant potential for generating machine-readable topological information on an on-going basis. We plan to investigate the possibility for additional IXP deployments and a full-fledged system for generating up-to-date network topology data in future work.

The peering model in which IXPs operate is different across different continents. For instance, an IXP in Europe is completely different from an IXP in North America. On one hand the peering model in North America typically involves a commercial colo-operator who also operates the peering equipment. On the other hand, the exchange points in Europe

tend to be non-profit, community-based organizations, and the colocation and peering equipment operators are different [175]. We believe that such a peering model will lead to differences on the results that we observe for networks in our study compared to networks in Europe.

4.7 Summary

To summarize, the key contributions of this chapter are as follows. First, we perform a first-of-its-kind comparison of large repositories of physical and network maps and find that physical maps typically reveal a much larger number of nodes (*e.g.,* POPs and hosting infrastructure). Next, we consider the targeting problem and find that using sources and destinations within the same autonomous system for probing reveals the most physical infrastructure. We develop a layer 1-informed heuristic algorithm for probe source-destination selection called POPsicle that identifies 2.4 times as many nodes as standard probing methods. Finally, we identify the fact that sources co-located as IXPs can be used to amplify POPsicle-based probing broadly throughout the Internet resulting in layer 3 maps that can be more effectively applied to problems of interest. To that end, we deployed our method at a real IXP and found that our method finds almost all POPs compared to Atlas and additional POPs compared to Ark for the ISPs studied.

5

InterTubes: A Study of the US Long-haul Fiber-optic Infrastructure

5.1 Introduction

The focus of this chapter is the physical Internet. In particular, we are concerned with the physical aspects of the wired Internet, ignoring entirely the wireless access portion of the Internet as well as satellite or any other form of wireless communication. Moreover, we are exclusively interested in the long-haul fiber-optic portion of the wired Internet in the US. The detailed metro-level fiber maps (with corresponding colocation and data center facilities) and international undersea cable maps (with corresponding landing stations) are only accounted for to the extent necessary. In contrast to short-haul fiber routes that are specifically built for short distance use and purpose (*e.g.*, to add or drop off network services in many different places within metro-sized areas), long-haul fiber routes (including ultra long-haul routes) typically run between major city pairs and allow for minimal use of repeaters.

With the US long-haul fiber-optic network being the main focal point of our work, the first contribution of this chapter consists of constructing a reproducible map of this basic component of the physical Internet infrastructure. To that end, we rely on publicly available fiber maps provided by many of the tier-1 ISPs and major cable providers. While some of these

maps include the precise geographic locations of all the long-haul routes deployed or used by the corresponding networks, other maps lack such detailed information. For the latter, we make extensive use of previously neglected or under-utilized data sources in the form of public records from federal, state, or municipal agencies or documentation generated by commercial entities (*e.g.*, commercial fiber map providers [58], utility rights-of-way (ROW) information, environmental impact statements, fiber sharing arrangements by the different states' DOTs). When combined, the information available in these records is often sufficient to reverse-engineer the geography of the actual long-haul fiber routes of those networks that have decided against publishing their fiber maps. We study the resulting map's diverse connectivity characteristics and quantify the ways in which the observed long-haul fiber-optic connectivity is consistent with existing transportation (*e.g.*, roadway and railway) infrastructure. We note that our work can be repeated by anyone for every other region of the world assuming similar source materials.

A striking characteristic of the constructed US long-haul fiber-optic network is a significant amount of observed infrastructure sharing. A qualitative assessment of the risk inherent in this observed sharing of the US long-haul fiber-optic infrastructure forms the second contribution of this chapter. Such infrastructure sharing is the result of a common practice among many of the existing service providers to deploy their fiber in jointly-used and previously installed conduits and is dictated by simple economics—substantial cost savings as compared to deploying fiber in newly constructed conduits. By considering different metrics for measuring the risks associated with infrastructure sharing, we examine the presence of high-risk links in the existing long-haul infrastructure, both from a connectivity and usage perspective. In the process, we follow prior work [281] and use the popularity of a route on the Internet as an informative proxy for the volume of traffic that route carries. End-to-end

paths derived from large-scale traceroute campaigns are overlaid on the actual long-haul fiber-optic routes traversed by the corresponding traceroute probes. The resulting first-of-its-kind map enables the identification of those components of the long-haul fiber-optic infrastructure which experience high levels of infrastructure sharing as well as high volumes of traffic.

The third and final contribution of our work is a detailed analysis of how to improve the existing long-haul fiber-optic infrastructure in the US so as to increase its resilience to failures of individual links or entire shared conduits, or to achieve better performance in terms of reduced propagation delay along deployed fiber routes. By framing the issues as appropriately formulated optimization problems, we show that both robustness and performance can be improved by deploying new fiber routes in just a few strategically-chosen areas along previously unused transportation corridors and ROW, and we quantify the achievable improvements in terms of reduced risk (*i.e.*, less infrastructure sharing) and decreased propagation delay (*i.e.*, faster Internet [289]). As actionable items, these technical solutions often conflict with currently-discussed legislation that favors policies such as “dig once", “joint trenching" or “shadow conduits" due to the substantial savings that result when fiber builds involve multiple prospective providers or are coordinated with other infrastructure projects (*i.e.*, utilities) targeting the same ROW [18]. In particular, we discuss our technical solutions in view of the current net neutrality debate concerning the treatment of broadband Internet providers as telecommunications services under Title II. We argue that the current debate would benefit from a quantitative assessment of the unavoidable trade-offs that have to be made between the substantial cost savings enjoyed by future Title II regulated service providers (due to their ensuing rights to gain access to existing essential infrastructure owned primarily by utilities) and an increasingly vulnerable national long-haul fiber-optic infrastructure (due to legislation

that implicitly reduced overall resilience by explicitly enabling increased infrastructure sharing).

5.2 Mapping Core Long-haul Infrastructure

In this section we describe the process by which we construct a map of the Internet's long-haul fiber infrastructure in the continental United States. While many *dynamic* aspects of the Internet's topology have been examined in prior work, the underlying long-haul fiber paths that make up the Internet are, by definition, *static*[1], and it is this fixed infrastructure which we seek to identify.

Our high-level definition of a long-haul link[2] is one that connects major city-pairs. In order to be consistent when processing existing map data, however, we use the following concrete definition. We define a long-haul link as one that spans at least 30 miles, *or* that connects population centers of at least 100,000 people, *or* that is shared by at least 2 providers. These numbers are not proscriptive, rather they emerged through an iterative process of refining our base map (details below).

The steps we take in the mapping process are as follows: *(1)* we create an initial map by using publicly available fiber maps from tier-1 ISPs and major cable providers which contain explicit geocoded information about long-haul link locations; *(2)* we validate these link locations and infer whether fiber conduits are shared by using a variety of public records documents such as utility right-of-way information; *(3)* we add links from publicly available ISP fiber maps (both tier-1 and major providers) which have geographic information about link *endpoints*, but which do not have explicit information about geographic pathways of fiber links; and *(4)* we

[1]More precisely, installed conduits rarely become defunct, and deploying new conduits takes considerable time.

[2]In the rest of the chapter, we will use the terms "link" and "conduit" interchangeably—a "tube" or trench specially built to house the fiber of potentially multiple providers.

again employ a variety of public records to infer the geographic locations of this latter set of links added to the map. Below, we describe this process in detail, providing examples to illustrate how we employ different information sources.

5.2.1 Step 1: Build an Initial Map

The first step in our fiber map-building process is to leverage maps of ISP fiber infrastructure with explicit geocoding of links from Internet Atlas project [206]. Internet Atlas is a measurement portal created to investigate and unravel the structural complexity of the *physical* Internet. Detailed geography of fiber maps are captured using the procedure described, esp. §3.2, in [206]. We start with these maps because of their potential to provide a significant and reliable portion of the overall map.

Specifically, we used detailed fiber deployment maps[3] from 5 tier-1 and 4 major cable providers: AT&T [17], Comcast [31], Cogent [29], EarthLink [47], Integra [84], Level3 [95], Suddenlink [136], Verizon [152] and Zayo [156]. For example, the map we used for Comcast's network [31] lists all the node information along with the exact geography of long-haul fiber links. Table 5.1 shows the number of nodes and links we include in the map for each of the 9 providers we considered. These ISPs contributed 267 unique nodes, 1258 links, and a total of 512 conduits to the map. Note that some of these links may follow exactly the same physical pathway (*i.e.*, using the same conduit). We infer such conduit sharing in step 2.

[3] Although some of the maps date back a number of years, due to the static nature of fiber deployments and especially due to the reuse of existing conduits for new fiber deployments [125], these maps remain very valuable and provide detailed information about the physical location of conduits in current use. Also, due to varying accuracy of the sources, some maps required manual annotation, georeferencing [67] and validation/inference (step 2) during the process.

Table 5.1: Number of nodes and long-haul fiber links included in the initial map for each ISP considered in step 1.

ISP	AT&T	Comcast	Cogent	EarthLink	Integra	Level 3	Suddenlink	Verizon	Zayo
# nodes	25	26	69	248	27	240	39	116	98
# links	57	71	84	370	36	336	42	151	111

5.2.2 Step 2: Checking the Initial Map

While the link location data gathered as part of the first step are usually reliable due to the stability and static nature of the underlying fiber infrastructure, the second step in the mapping process is to collect additional information sources to validate these data. We also use these additional information sources to infer whether some links follow the same physical ROW, which indicates that the fiber links either reside in the same fiber bundle, or in an adjacent conduit.

In this step of the process, we use a variety of public records to geolocate and validate link endpoints and conduits. These records tend to be rich with detail, but have been under-utilized in prior work that has sought to identify the physical components that make up the Internet. Our working assumption is that ISPs, government agencies, and other relevant parties often archive documents on public-facing websites, and that these documents can be used to validate and identify link/conduit locations. Specifically, we seek information that can be extracted from government agency filings (*e.g.*, [28, 36, 44]), environmental impact statements (*e.g.*, [1]), documentation released by third-party fiber services (*e.g.*, [13–15, 24]), indefeasible rights of use (IRU) agreements (*e.g.*, [87, 88]), press releases (*e.g.*, [100, 101, 106, 107]), and other related resources (*e.g.*, [20, 25, 41, 45, 46, 128, 147]).

Public records concerning rights-of-way are of particular importance to our work since highly-detailed location and conduit sharing information can be gleaned from these resources. Laws governing rights of way are established on a state-by-state basis (*e.g.*, see [56]), and which local organization has jurisdiction varies state-by-state [4]. As a result, care must be taken when validating or inferring the ROW used for a particular fiber link. Since these state-specific laws are public, however, they establish a number of key parameters to drive a systematic search for government-related public filings.

In addition to public records, the fact that a fiber-optic link's location aligns with a known ROW serves as a type of validation. Moreover, if link locations for multiple service providers align along the same geographic path, we consider those links to be validated.

To continue the example of Comcast's network, we used, in part, the following documents to validate the locations of links and to determine which links run along shared paths with other networks: *(1)* a broadband environment study by the FCC details several conduits shared by Comcast and other providers in Colorado [26], *(2)* a franchise agreement [38, 39] made by Cox with Fairfax county, VA suggests the presence of a link running along the ROW with Comcast and Verizon, *(3)* page 4 (utilities section) of a project document [42] to design services for Wekiva Parkway from Lake County to the east of Round Lake Road (Orlando, FL) demonstrates the presence of Comcast's infrastructure along a ROW with other entities like CenturyLink, Progress Energy and TECO/People's Gas, *(4)* an Urbana city council project update [148] shows pictures [149] of Comcast and AT&T's fiber deployed in the Urbana, IL area, and *(5)* documents from the CASF project [150] in Nevada county, CA show that Comcast has deployed fiber along with AT&T and Suddenlink.

5.2.3 Step 3: Build an Augmented Map

The third step of our long-haul fiber map construction process is to use published maps of tier-1 and large regional ISPs which do not contain explicit geocoded information. We tentatively add the fiber links from these ISPs to the map by aligning the logical links indicated in their published maps along the closest known right-of-way (*e.g.*, road or rail). We validate and/or correct these tentative placements in the next step.

In this step, we used published maps from 7 tier-1 and 4 regional providers: CenturyLink, Cox, Deutsche Telekom, HE, Inteliquent, NTT, Sprint, Tata, TeliaSonera, TWC, XO. Adding these ISPs resulted in an

addition of 6 nodes, 41 links, and 30 conduits (196 nodes, 1153 links, and 347 conduits without considering the 9 ISPs above). For example, for Sprint's network [133], 102 links were added and for CenturyLink's network [21], 134 links were added.

5.2.4 Step 4: Validate the Augmented Map

The fourth and last step of the mapping process is nearly identical to step 2. In particular, we use public filings with state and local governments regarding ROW access, environmental impact statements, publicly available IRU agreements and the like to validate locations of links that are inferred in step 3. We also identify which links share the same ROW. Specifically with respect to inferring whether conduits or ROWs are shared, we are helped by the fact that the number of possible rights-of-way between the endpoints of a fiber link are limited. As a result, it may be that we simply need to *rule out* one or more ROWs in order to establish sufficient evidence for the path that a fiber link follows.

Individual Link Illustration: Many ISPs list only POP-level connectivity. For such maps, we leverage the corpus of search terms that we capture in Internet Atlas and search for public evidence. For example, Sprint's network [133] is extracted from the Internet Atlas repository. The map contains detailed node information, but the geography of long-haul links is not provided in detail. To infer the conduit information, for instance, from Los Angeles, CA to San Francisco, CA, we start by searching "los angeles to san francisco fiber iru at&t sprint" to obtain an agency filing [28] which shows that AT&T and Sprint share that particular route, along with other ISPs like CenturyLink, Level 3 and Verizon. The same document also shows conduit sharing between CenturyLink and Verizon at multiple locations like Houston, TX to Dallas, TX; Dallas, TX to Houston, TX; Denver, CO to El Paso, TX; Santa Clara, CA to Salt Lake City, UT; and Wells, NV to Salt Lake City, UT.

As another example, the IP backbone map of Cox's network [40] shows that there is a link between Gainesville, FL and Ocala, FL. But the geography of the fiber deployment is absent (*i.e.*, shown as a simple point with two names in [40]). We start the search using other ISP names (*e.g.*,"level 3 and cox fiber iru ocala") and obtain publicly available evidence (*e.g.*, lease agreement [37]) indicating that Cox uses Level3's fiber optic lines from Ocala, FL to Gainesville, FL. Next, we repeat the search with different combinations for other ISPs (*e.g.*, news article [93] shows that Comcast uses 19,000 miles of fiber from Level3; see map at bottom of that page which highlights the Ocala to Gainesville route, among others) and infer that Comcast is also present in that particular conduit. Given that we know the detailed fiber maps of ISPs (*e.g.*, Level 3) and the inferred conduit information for other ISPs (*e.g.*, Cox), we systematically infer conduit sharing across ISPs.

Resource Illustration: To illustrate some of the resources used to validate the locations of Sprint's network links, publicly available documents reveal that *(1)* Sprint uses Level 3's fiber in Detroit [134] and their settlement details are publicly available [135], *(2)* a whitepaper related to a research network initiative in Virginia identifies link location and sharing details regarding Sprint fiber [45], *(3)* the "coastal route" [28] conduit installation project started by Qwest (now CenturyLink) from Los Angeles, CA to San Francisco, CA shows that, along with Sprint, fiber-optic cables of several other ISPs like AT&T, MCI (now Verizon) and WilTel (now Level 3) were pulled through the portions of the conduit purchased/leased by those ISPs, and *(4)* the fiber-optic settlements website [57] has been established to provide information regarding class action settlements involving land next to or under railroad rights-of-way where ISPs like Sprint, Qwest (now CenturyLink), Level 3 and WilTel (now Level 3) have installed telecommunications facilities, such as fiber-optic cables.

5.2.5 The US Long-haul Fiber Map

The final map constructed through the process described in this section is shown in Figure 5.2, and contains 273 nodes/cities, 2411 links, and 542 conduits (with multiple tenants). Prominent features of the map include *(i)* dense deployments (*e.g.*, the northeast and coastal areas), *(ii)* long-haul hubs (*e.g.*, Denver and Salt Lake City) *(iii)* pronounced absence of infrastructure (*e.g.*, the upper plains and four corners regions), *(iv)* parallel deployments (*e.g.*, Kansas City to Denver) and *(v)* spurs (*e.g.*, along northern routes).

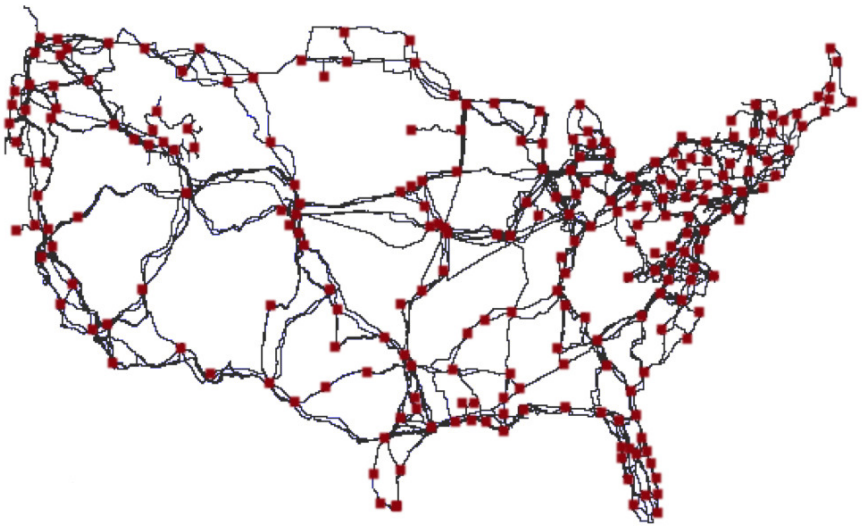

Figure 5.2: Location of physical conduits for networks considered in the continental United States.

While mapping efforts like the one described in this section invariably raise the question of the quality of the constructed map (*i.e.*, completeness), it is safe to state that despite our efforts to sift through hundreds of relevant documents, the constructed map is not complete. At the same time, we are confident that to the extent that the process detailed in this section reveals long-haul infrastructure for the sources considered, the constructed map

is of sufficient quality for studying issues that do not require local details typically found in metro-level fiber maps. Moreover, as with other Internet-related mapping efforts (*e.g.*, AS-level maps), we hope this work will spark a community effort aimed at gradually improving the overall fidelity of our basic map by contributing to a growing database of information about geocoded conduits and their tenants.

The methodological blueprint we give in this section shows that constructing such a detailed map of the US's long-haul fiber infrastructure is feasible, and since all data sources we use are publicly available, the effort is reproducible. The fact that our work can be replicated is not only important from a scientific perspective, it suggests that the same effort can be applied more broadly to construct similar maps of the long-haul fiber infrastructure in other countries and on other continents.

Interestingly, recommendation 6.4 made by the FCC in chapter 6 of the National Broadband Plan [18] states that "the FCC should improve the collection and availability regarding the location and availability of poles, ducts, conduits, and rights-of-way.". It also mentions the example of Germany, where such information is being systematically mapped. Clearly, such data would obviate the need to expend significant effort to search for and identify the relevant public records and other documents.

Lastly, it is also important to note that there are commercial (fee-based) services that supply location information for long-haul and metro fiber segments, *e.g.*, [58]. We investigated these services as part of our study and found that they typically offer maps of some small number (5–7) of national ISPs, and that, similar to the map we create (see map in [79][4]), many of these ISPs have substantial overlap in their locations of fiber deployments. Unfortunately, it is not clear how these services obtain their source information and/or how reliable these data are. Although it is not possible to confirm, in the best case these services offer much of the same

[4]Visually, all the commercially-produced maps agree with our basic map, hinting at the common use of supporting evidence.

information that is available from publicly available records, albeit in a convenient but non-free form.

5.3 Geography of Fiber Deployments

In this section, we analyze the constructed map of long-haul fiber-optic infrastructure in the US in terms of its alignment with existing transportation networks. In particular, we examine the relationship between the geography of physical Internet links and road and rail infrastructure.

While the conduits through which the long-haul fiber-optic links that form the physical infrastructure of the Internet are widely assumed to follow a combination of transportation infrastructure locations (*i.e.,* railways and roadways) along with public/private right-of-ways, we are aware of very few prior studies that have attempted to confirm or quantify this assumption [68]. Understanding the relationship between the physical links that make up the Internet and the physical pathways that form transportation corridors helps to elucidate the prevalence of conduit sharing by multiple service providers and informs decisions on where future conduits might be deployed.

Our analysis is performed by comparing the physical link locations identified in our constructed map to geocoded information for both roadways and railways from the United States National Atlas website [145]. The geographic layout of our roadway and railway data sets can be seen in Figure 5.3 and Figure 5.4, respectively. In comparison, the physical link geographic information for the networks under consideration can be seen in the Figure 5.2.

We use the *polygon overlap* analysis capability in the ArcGIS [50] to quantify the correspondence between physical links and transportation infrastructure. In Figure 5.5, aggregating across all networks under consideration, we compare the fraction of each path that is co-located with

Figure 5.3: NationalAtlas roadway infrastructure locations.

Figure 5.4: NationalAtlas railway infrastructure locations.

roadways, railways, or a combination of the two using histogram distributions. These plots show that a significant fraction of all the physical links are co-located with roadway infrastructure. The plots also show that it is more common for fiber conduits to run alongside roadways than railways, and an even higher percentage are co-located with some combination of roadways *and* railway infrastructure. Furthermore, for a vast majority of

the paths, we find that physical link paths more often follow roadway infrastructure compared with rail infrastructure.

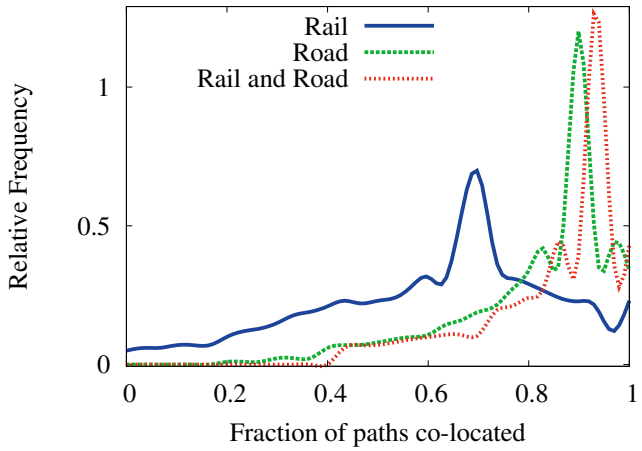

Figure 5.5: Fraction of physical links co-located with transportation infrastructure.

Despite the results reported above there remain conduits in our infrastructure map that are *not* co-located with transportation ROWs. For example, in the left-hand plot of Figure 5.6 we show the Level 3-provided physical link locations outside Laurel, MS, and in the right-hand plot we show Google Maps [70] satellite imagery for the same location. These images shows the presence of network links, but no known transportation infrastructure is co-located. In what follows, we list examples by considering other types of rights-of-way, such as natural gas and/or petroleum pipelines, but leave details to future work.

A few examples can be shown in Level3's network [95], where the map shows the existence of link from *(1)* Anaheim, CA to Las Vegas, NV, and *(2)* Houston, TX to Atlanta, GA, but no known transportation infrastructure is co-located. By considering other types of rights-of-way [118], many of these situations could be explained. Visually, we can verify that the link from Anaheim, CA to Las Vegas, NV is co-located with refined-products

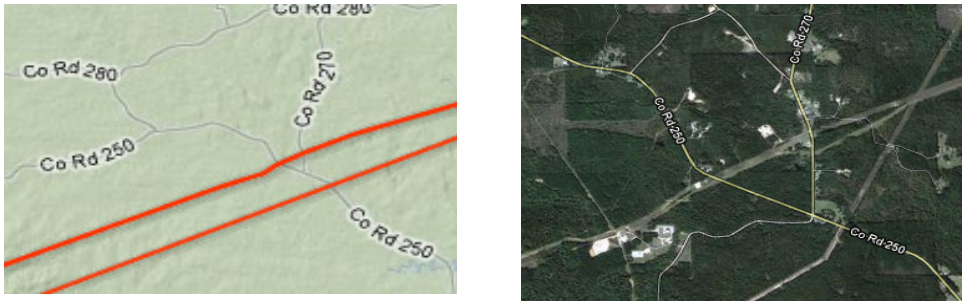

Figure 5.6: Satellite image validated right-of-way outside of Laurel, MS. (Left) - Level 3 Provided fiber map. (Right) - Google Maps satellite view.

pipeline. Similarly, the link from Houston, TX to Atlanta, GA is deployed along with NGL pipelines.

5.4 Assessing Shared Risk

In this section, we describe and analyze two notions of *risk* associated with sharing fiber-optic conduits in the Internet. At a high level, we consider conduits that are shared by many service providers as an inherently risky situation since damage to that conduit will affect several providers. Our choice of such a risk model that considers the degree of link sharing and not the overall physical topology as a means to analyze robustness is based on the fact that our map is highly incomplete compared to the 40K plus ASes and certain metrics (*e.g.*, number of fiber cuts to partition the US long-haul infrastructure) have associated security implications [5]. We intend to analyze different dimensions of network resilience in future work.

5.4.1 Risk Matrix

Our analysis begins by creating a *risk matrix* based on a simple counting-based approach. The goal of this matrix is to capture the level of infrastructure sharing and establish a measure of shared risk due to lack of diversity in physical connectivity. The risk matrix is populated as follows: we start with a tier-1 ISP that has vast infrastructure in the US and subsequently add other tier-1 and major cable Internet providers to the matrix. The rows are ISPs and columns are physical conduits carrying long-haul fiber-optic links for those ISPs. Integer entries in the matrix refer to the number of ISPs that share a particular conduit. As a result, values in the matrix increase as the level of conduit-sharing increases.

As an illustrative example, we choose Level 3 as a "base" network due to its very rich connectivity in the US. We use our constructed physical network map (*i.e.*, the map we describe in §5.2) and extract all conduit endpoints across city pairs, such as "SLC-Denver" (c1 below), SLC-Sacramento (c2 below), and Sacramento-Palo Alto (c3 below), etc., and assign 1 for all conduits that are part of Level 3's physical network footprint. A partial matrix is then:

	c1	c2	c3
Level 3	1	1	1

Next, say we include another provider, *e.g.*, Sprint. We add a new row for Sprint to the matrix, then for any conduit used in Sprint's physical network, we increment *all* entries in each corresponding column. For this example, Sprint's network shares the SLC-Denver and SLC-Sacramento conduits with other providers (including Level 3), but not the Sacramento-Palo Alto conduit. Thus, the matrix becomes:

	c1	c2	c3
Level 3	2	2	1
Sprint	2	2	0

We repeat this process for all the twelve tier-1 and eight major Internet service providers, *i.e.*, the same ISPs used as part of constructing our physical map of long-haul fiber-optic infrastructure in the US in §5.2.

5.4.2 Risk Metric: Connectivity-only

How many ISPs share a link? Using the risk matrix, we count the number of ISPs sharing a particular conduit. Figure 5.7 shows the number of conduits (y axis) for which at least k ISPs (x axis) share the conduit. For example, there are 542 distinct conduits in our physical map (Figure 5.2), thus the bar at x=1 is 542, and 486 conduits are shared by at least 2 ISPs, thus the bar at x=2 is 486. This plot highlights the fact that it is relatively uncommon for conduits *not* to be shared by more than two providers. Overall, we observe that 89.67%, 63.28% and 53.50% of the conduits are shared by at least two, three and four major ISPs, respectively.

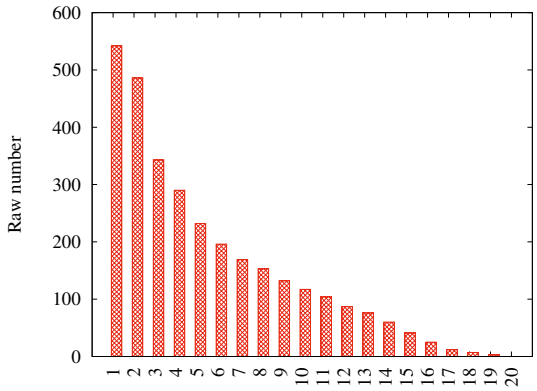

Figure 5.7: Raw number of conduits for which at least k ISPs (x axis) share the conduit.

In some of the more extreme cases, we observe that 12 out of 542 conduits are shared by *more than 17 ISPs*. These situations may arise where such conduits run between major population centers, or between cities separated by imposing geographic constraints (*e.g.*, the Rocky Mountains).

For example, conduits that are shared by 19 ISPs include *1)* Phoenix, AZ to Tucson, AZ, *(2)* Salt Lake City, UT to Denver, CO, and *(3)* Philadelphia, PA to New York, NY.

Implication: When it comes to physically deployed connectivity, the US long-haul infrastructure lacks much of the diversity that is a hallmark of all the commonly-known models and maps of the more logical Internet topologies (e.g., router- or AS-level graphs [189, 206, 301]).

Which ISPs do the most infrastructure sharing? To better understand the infrastructural sharing risks to which individual ISPs are exposed, we leverage the risk matrix and rank the ISPs based on increasing average shared risk. The average of the values across a row in the risk matrix (*i.e.*, values for an individual ISP) with standard error bars, 25^{th} and 75^{th} percentile are shown in Figure 5.7. The average values are plotted in a sorted fashion, resulting in an increasing level of infrastructure sharing when reading the plot from left to right.

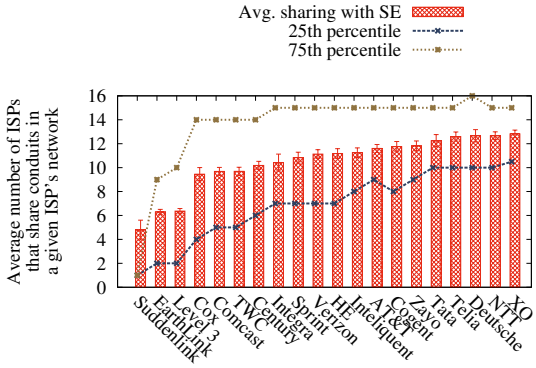

Figure 5.8: The raw number of shared conduits by ISPs.

From this plot we observe that Suddenlink has the smallest average number of ISPs that share the conduits used in its network, which can be explained by its diverse geographical deployments. It is followed by EarthLink and Level 3. Deutsche Telekom, NTT and XO, on the other

hand, use conduits that are, on average, shared by a large number of other ISPs.

Implication: Non-US service providers (*e.g.*, Deutsche Telekom, NTT, Tata, etc.) use policies like *dig once* [43] and *open trench* [111], and/or lease dark fibers to expand their presence in the US. Such policies may save deployment costs, but appear to be counter-productive as far as overall network resilience is concerned.

How similar are ISP risk profiles? Using the risk matrix we calculate the Hamming [73] distance similarity metric among ISPs, *i.e.*, by comparing every row in the risk matrix to every other row to assess their similarity. Our intuition for using such a metric is that if two ISPs are physically similar (in terms of fiber deployments and the level of infrastructure sharing), their risk profiles are also similar.

Figure 5.9 shows a heat map generated by computing the Hamming distance metric for every pair of ISPs considered in the construction of our physical map. For this metric, the smaller the number, the greater the shared risk between the corresponding (two) ISPs. We observe in the plot that EarthLink and Level 3 exhibit fairly low risk profiles among the ISPs we considered, similar to results described above when we consider the average number of ISPs sharing conduits used in these networks. These two ISPs are followed by Cox, Comcast and Time Warner Cable, which likely exhibit lower risk according to the Hamming distance metric due to their rich fiber connectivity in the US.

Somewhat surprisingly, although the average number of ISPs that share conduits in Suddenlink's network is, on average, low, the Hamming distance metric suggests that it is exposed to risks due to its geographically diverse deployments. While Level 3 and EarthLink also have geographically diverse deployments, they also have diverse paths that can be used to reach various destinations without using highly-shared conduits. On the other hand, Suddenlink has few alternate physical paths, thus they

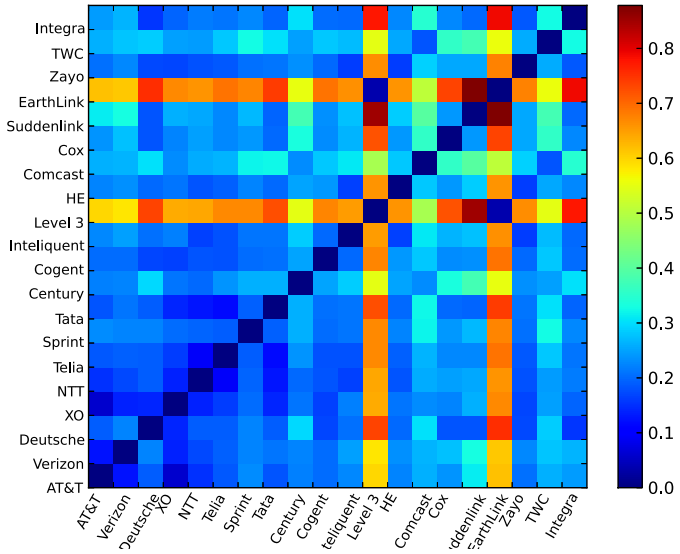

Figure 5.9: Similarity of risk profiles of ISPs calculated using Hamming distance. Risk profiles are similar for ISPs exhibiting the same color.

must depend on certain highly-shared conduits to reach certain locations. TATA, TeliaSonera, Deutsche Telekom, NTT and XO each use conduits that are very highly shared, thus they have similar risk profiles according to the Hamming distance metric.

Implication: Multiple metrics are required to precisely characterize and capture the level of infrastructure sharing by service providers. Geographically diverse deployment may reduce the risk only when the ISP has diverse paths to avoid the critical choke points to reach different destinations.

5.4.3 Risk Metric: Connectivity + Traffic

In this section, we follow the method of [281] and use the popularity of different routes on the Internet as measured through traceroute probes as a way to infer relative volumes of traffic on those routes. We use traceroute

data from the Edgescope [194] project and restrict our analysis to a period of 3 months, from January 1, 2014 to March 31, 2014. These data consisted of 4,908,223 individual traceroutes by clients in diverse locations. By using geolocation information and naming hints in the traceroute data [190, 234], we are able to overlay individual layer 3 links onto our underlying physical map of Internet infrastructure. As a result, we are able to identify those components of the long-haul fiber-optic infrastructure which experience high levels of infrastructure sharing as well as high volumes of traffic.

The prevalent use of MPLS tunnels in the Internet [290] poses one potential pitfall with overlaying observed layer 3 routes onto our physical map. While we certainly do see segments along individual traceroutes that likely pass through MPLS tunnels, we observe the frequency of these segments to be relatively low. Thus, we believe that their impact on the results we describe below is limited.

Ranking by frequency. Table 5.10 and Table 5.12 show the top 20 conduits for west-origin east-bound and east-origin west-bound probes[5] ranked based on frequency. Interestingly, for these tables we observe high volumes of traffic flowing through certain cities (*e.g., Dallas, TX, Salt Lake City, UT*) in either direction, and that while many of the conduit endpoints are major population centers, there are a number of endpoint cities that are simply popular waypoints (*e.g.*, Casper, WY and Billings, MT in the East to West direction).

Additional ISPs. Figure 5.11 compares the CDF of the number of ISPs sharing a conduit with a CDF of conduit frequencies observed through the traceroute data. In the plot, we observe that the conduits identified in our physical map appear on large numbers of paths in the traceroute data, and that when we consider traffic characteristics, the shared risk of certain conduits is *only greater*. Through analysis of naming conventions in the traceroute data, we infer that there are even larger numbers of ISPs

[5]Classified based on geolocation information for source/destination hops in the traceroute data.

Table 5.10: Top 20 base long-haul conduits and their corresponding frequencies of west-origin to east-bound traceroute probes.

Location	Location	# Probes
Trenton, NJ	Edison, NJ	78402
Kalamazoo, MI	Battle Creek, MI	78384
Dallas, TX	Fort Worth, TX	56233
Baltimore, MD	Towson, MD	46336
Baton Rouge, LA	New Orleans, LA	46328
Livonia, MI	Southfield, MI	46287
Topeka, KS	Lincoln, NE	46275
Spokane, WA	Boise, ID	44461
Dallas, TX	Atlanta, GA	41008
Dallas, TX	Bryan, TX	39232
Shreveport, LA	Dallas, TX	39210
Wichita Falls, TX	Dallas, TX	39180
San Luis Obispo, CA	Lompoc, CA	32381
San Francisco, CA	Las Vegas, NV	22986
Wichita, KS	Las Vegas, NV	22169
Las Vegas, NV	Salt Lake City, UT	22094
Battle Creek, MI	Lansing, MI	15027
South Bend, IN	Battle Creek, MI	14795
Philadelphia, PA	Allentown, PA	12905
Philadelphia, PA	Edison, NJ	12901

that share the conduits identified in our physical map, thus the potential risks due to infrastructure sharing are magnified when considering traffic characteristics. For example, our physical map establishes that the conduit between Portland, OR and Seattle, WA is shared by 18 ISPs. Upon analysis of the traceroute data, we inferred the presence of an additional 13 ISPs that also share that conduit.

Distribution of traffic. We also ranked the ISPs based on the number of conduits used to carry traffic. Table 5.13 lists the top 10 ISPs in terms of number of conduits observed to carry traceroute traffic. We see that Level 3's infrastructure is the most widely used. Using the traceroute frequencies as a proxy, we also infer that Level 3 carries the most traffic.

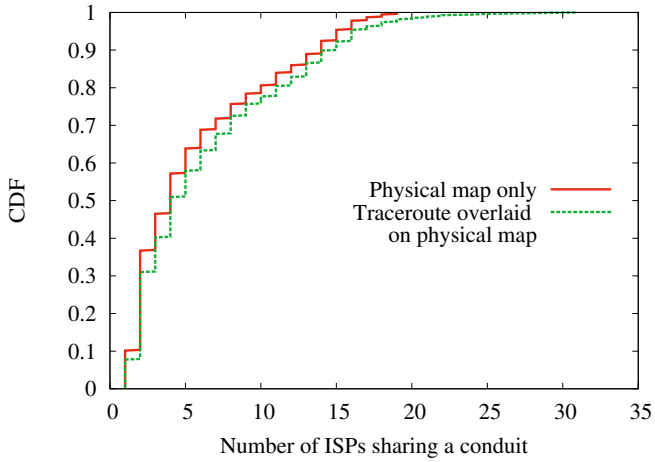

Figure 5.11: CDF of number of ISPs sharing a conduit before and after considering the traceroute data as a proxy for traffic volumes.

In fact, it has a significantly higher number of conduits used compared to the next few "top" ISPs. Interestingly, although XO is also considered to be a Tier-1 provider, it carries approximately 25% of the volume that Level 3 carries, at least inferred through these data.

5.5 Mitigating Risks

In this section we describe two optimization analyses in which we examine how best to improve the existing physical infrastructure to either increase the robustness of long-haul infrastructure to fiber cuts, or to minimize propagation delay between pairs of cities.

5.5.1 Increasing Network Robustness without Adding New Conduits

We first examine the possibility of improving the long-haul infrastructure's robustness (*i.e.*, to reduce the impact of fiber cuts by reducing

Table 5.12: Top 20 base long-haul graph conduits and their corresponding frequencies of east-origin to west-bound traceroute probes.

Location	Location	# Probes
West Palm Beach, FL	Boca Raton, FL	155774
Lynchburg, VA	Charlottesville, VA	155079
Sedona, AZ	Camp Verde, AZ	54067
Bozeman, MT	Billings, MT	50879
Billings, MT	Casper, WY	50818
Casper, WY	Cheyenne, WY	50817
White Plains, NY	Stamford, CT	25784
Amarillo, TX	Wichita Falls, TX	16354
Eugene, OR	Chico, CA	12234
Phoenix, AZ	Dallas, TX	9725
Salt Lake City, UT	Provo, UT	9433
Salt Lake City, UT	Los Angeles, CA	8921
Dallas, TX	Oklahoma City, OK	8242
Wichita Falls, TX	Dallas, TX	8150
Seattle, WA	Portland, OR	8094
Eau Claire, WI	Madison, WI	7476
Salt Lake City, UT	Cheyenne, WY	7380
Bakersfield, CA	Los Angeles, CA	6874
Seattle, WA	Hillsboro, OR	6854
Santa Barbara, CA	Los Angeles, CA	6641

the level of conduit sharing among ISPs[6]) by either *(1)* utilizing existing conduits that are not currently part of that ISPs physical footprint, or *(2)* carefully choosing ISPs to peer with at particular locations such that the addition of the peer adds diversity in terms of physical conduits utilized. In either case, we rely on the existing physical infrastructure and the careful choice of conduits rather than introduce any new links.

We call the optimization framework used in this first analysis a *robustness suggestion*, as it is designed to find a set of links or set of ISPs to peer with at different points in the network such that global shared risk (*i.e.*, shared risk across all ISPs) is minimized. We refer to this set of additional

[6]When accounting for alternate routes via undersea cables, network partitioning for the US Internet is a very unlikely scenario.

Table 5.13: Top 10 ISPs in terms of number of conduits carrying probe traffic measured in the traceroute data.

ISP	# conduits
Level 3	62
Comcast	48
AT&T	41
Cogent	37
SoftLayer	30
MFN	21
Verizon	21
Cox	18
CenturyLink	16
XO	15

links or peering as the *robust backup infrastructure*. We define the optimized path between two city-level nodes *i* and *j*, $OP_{i,j}^{robust}$, as,

$$OP_{i,j}^{robust} = \min_{P_{i,j} \in \mathcal{E}^A} SR\left(\mathbf{P_{i,j}}\right) \tag{5.1}$$

where $\mathcal{E}^A$ is the set of all possible paths obtained from the risk matrix. The difference between the original set of existing network hops and the hops seen in the optimized paths produced from equation 5.1 forms the additional peering points. Depending on operational needs and robustness requirements, the framework can used to optimize specific paths or the entire network, thereby improving the robustness of the network at different granularities.

In our analysis of the constructed physical map of the fiber-optic infrastructure in the US, we found that there are 12 out of 542 conduits that are shared by more than 17 out of the 20 ISPs we considered in our study. We begin by analyzing these twelve links and how network robustness could be improved through our robustness suggestion framework. We use two specific metrics to evaluate the effectiveness of the robustness suggestion: *(1)* path inflation (PI) *i.e.*, the difference between the number of hops in

the original path and the optimized path, and *(2)* shared risk reduction (SRR), *i.e.*, the difference in the number of ISPs sharing the conduit on the original path versus the optimized path.

Figure 5.14 shows the PI and SRR results for optimizing the 12 highly-shared links, for all ISPs considered in our study. Overall, these plots show that, on average, an addition of between one and two conduits that were not previously used by a particular ISP results in a significant reduction in shared risk across all networks. We observe that nearly all the benefit of shared risk reduction is obtained through these modest additions.

Apart from finding optimal paths with minimum shared risk, the robustness suggestion optimization framework can also be used to infer additional peering (hops) that can improve the overall robustness of the network. Table 5.15 shows the top three beneficial peering additions based on minimizing shared risk in the network for the twelve most highly-shared links. Level 3 is predominantly the best peer that any ISP could add to improve robustness, largely due to their already-robust infrastructure. AT&T and CenturyLink are also prominent peers to add, mainly due to the diversity in geographic paths that border on the 12 highly-shared links.

Besides focusing on optimizing network robustness by addressing the 12 heavily-shared links, we also considered how to optimize ISP networks considering *all* 542 conduits with lit fiber identified in our map of the physical infrastructure in the US. We do not show detailed results due to space constraints, but what we found in our analysis was that many of the existing paths used by ISPs were already the best paths, and that the potential gains were minimal compared to the gains obtained when just considering the 12 conduits.

Overall, these results are encouraging, because they imply that it is sufficient to optimize the network around a targeted set of highly-shared links. They also suggest that modest additions of city-to-city backup links would be enough to get most of the potential robustness gains.

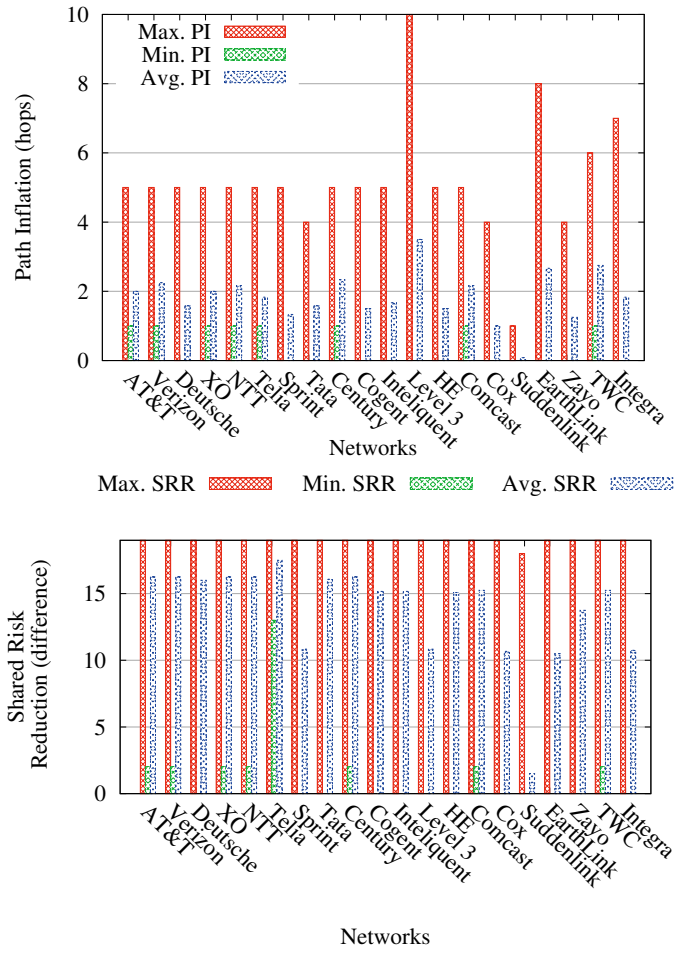

Figure 5.14: Path Inflation (top) and Shared Risk Reduction (bottom) based on the robustness suggestion framework for the twelve heavily shared links.

5.5.2 Increasing Network Robustness through Targeted Infrastructure Additions

In this section we consider how to improve network robustness by adding up to k new city-to-city fiber conduits. We consider the existing physical map as a graph $G = V, E$ along with the risk matrix A. Our goal

Table 5.15: Top 3 best peering suggested by the optimization framework for optimizing the twelve shared links.

ISP	Suggested Peering
AT&T	Level 3 | Century | Verizon
Verizon	Level 3 | Century | AT&T
Deutsche	Level 3 | AT&T | Century
XO	Level 3 | AT&T | Century
NTT	Level 3 | AT&T | Century
Telia	Level 3 | Century | AT&T
Sprint	Level 3 | AT&T | Century
Tata	Level 3 | AT&T | Century
Century	Level 3 | AT&T | Verizon
Cogent	Level 3 | AT&T | CenturyLink
Inteliquent	Level 3 | Century | AT&T
Level 3	Century | Integra | EarthLink
HE	Level 3 | AT&T | Century
Comcast	Level 3 | AT&T | Verizon
Cox	AT&T | Level 3 | Century
Suddenlink	Level 3 | AT&T | Sprint
EarthLink	Tata | Integra | AT&T
Zayo	Level 3 | AT&T | Century
TWC	Level 3 | AT&T | Verizon
Integra	Level 3 | Sprint | Century

is to identify a new set of edges along with E such that the addition *(1)* causes the largest increase in overall robustness, *i.e.*, greatest reduction in shared risk, and *(2)* while imposing the smallest deployment cost (DC), *i.e.*, the cost per fiber conduit mile, compared with alternate shortest paths between two city pairs.

Formally, let $\hat{E} = \{\{u,v\} : u,v \in V \text{ and } \{u,v\} \notin E\}$ be the set of edges not in G and let $\hat{A}$ be the reduced shared risk matrix of network $\hat{G} = (V, E \cup S)$ for some set $S \in \hat{E}$. We want to find $S \in \hat{E}$ of size k such that

$$S = \arg\max(\lambda_A - \lambda_{\hat{A}}) \tag{5.2}$$

where $\lambda = \sum_{i=1,j=1}^{m,n} SRR_{i,j} + \sum_{i=1,j=1}^{m,n} DC_{i,j}$ and $DC_{i,j}$ is the alternate

shortest path with reduced cost and physically shortest, different, redundant path between i and j.

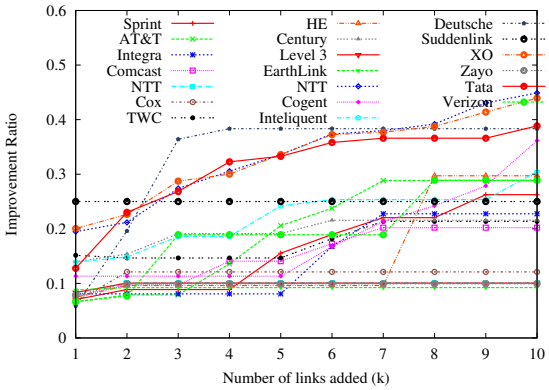

Figure 5.16: Potential improvements to ISP

Figure 5.16 shows the *improvement ratio* (avg. shared risk after adding link(s) divided by avg. shared risk before adding link(s)) for the 20 ISPs considered in our study. The objective function is to deploy new fiber at geographically diverse locations such that the deployment cost (*i.e.*, the length of fiber) is minimized and global shared risk is reduced. As expected, we see good improvement for ISPs with smaller infrastructural footprints in the US, *e.g.*, for Telia, Tata, etc. and very little improvement for large US-based ISPs such as Level 3, CenturyLink, and Cogent, since their networks already have fairly rich connectivity. An interesting case is Suddenlink, which shows no improvement even after adding multiple links. We attribute this result to the dependency on the other ISPs to reach destinations because of its geographically diverse conduit paths.

5.5.3 Reducing Propagation Delay

In this section we examine propagation delays between individual city pairs in our map of the physical fiber infrastructure in the US. Since there may be multiple existing physical conduit paths between two cities,

we consider the average delay across all physical paths versus the best (lowest) delay along one of the existing physical paths. We also consider how delay may be reduced by adding new physical conduit paths that follow existing roads or railways (*i.e.*, existing rights-of-way) between a pair of cities. Lastly, we consider the possibility of adding new physical conduit that ignores rights-of-way and simply follows the line-of-sight (LOS). Although following the LOS is in most cases practically infeasible, it represents the minimum achievable delay between two cities and thus provides a lower bound on performance.

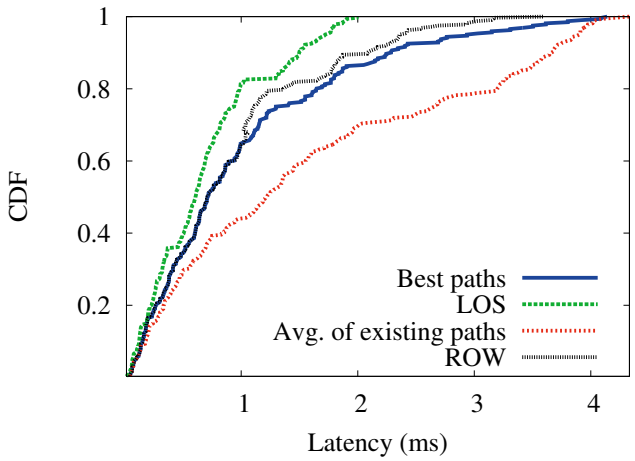

Figure 5.17: Comparison of best links against avg. latencies of links, ROW links and LOS links.

Figure 5.17 plots the cumulative distribution function of delays across all city pairs that have existing conduits between them. We first observe in the figure that the average delays of existing links between city pairs are often substantially higher than the best existing link. This result suggests that there are some long-haul fiber links that traverse much longer distances than necessary between two cities, perhaps due to ease of deployment or lower costs in certain conduits. We also observe that even the best existing paths do not follow the shortest rights-of-way between two cities,

but that the difference in many cases is fairly small. In particular, about 65% of the best paths are also the best ROW paths. Lastly, we observe that the LOS distance between two cities versus the best ROW path (or best existing path) varies. For 50% of the paths, the difference is under 100 microseconds (*i.e.*, approximately 20 km), but for 25% of the paths the difference is more than 500 microseconds (*i.e.*, more than 100 km), with some differences exceeding 2 milliseconds (*i.e.*, more than 400 km; see [19]). These results indicate that it is important to consider rights-of-way when evaluating possible improvements to propagation delays in the Internet, since line-of-sight distances may differ significantly and may not be practically achievable.

5.6 Discussion

In this section, we discuss the broader implications of our findings and offer ideas on how the additional infrastructure indicated by our analysis might be practically deployed.

5.6.1 Implications for Service Providers

Our base map of the US long-haul fiber infrastructure highlights the fiber conduits used to transmit data between large population centers. While infrastructure such as content delivery networks and data centers complicate the details of data flows, this map can support and inform decisions by service providers on provisioning and management of their infrastructures. Beyond performance and robustness analysis, the base map can inform decisions on local/regional broadband deployment, peering, and route selection, as well as provide competitive insights. Further, the fact that there is widespread and sometimes significant conduit sharing complicates the task of identifying and configuring backup paths since these critical details are often opaque to higher layers. Enrichment of this

map through the addition of long-haul links in other regions around the world, undersea cable maps for inter-continental connectivity, and metro-level fiber maps will improve our global view of the physical Internet and will provide valuable insights for all involved players (*e.g.*, regional, national, or global-scale providers). Finally, the map also informs regulatory and oversight activities that focus on ensuring a safe and accessible physical communications infrastructure.

While much prior work on aspects of (logical) Internet connectivity at layer 3 and above points to the dynamic nature of the corresponding graph structures as an invariant, it is important to recognize that the (physical) long-haul infrastructure is comparably static by definition (*i.e.*, deploying new fiber takes time). In that sense, the links reflected in our map can also be considered an Internet invariant, and it is instructive to compare the basic structure of our map to the NSFNET backbone circa 1995 [108].

5.6.2 The FCC and Title II

Over the past several years, there have been many discussions about the topic of network neutrality. The US Communications Act of 1934 [32] is mentioned frequently in those discussions since Title II of that Act enables the FCC to specify communications providers as “common carriers". One implication of the recent FCC decision to reclassify broadband Internet providers as common carriers is that parts of a provider’s infrastructure, including utility poles and conduits, will need to be made available to third parties. If this decision is upheld, it will likely lead to third party providers taking advantage of expensive already-existing long-haul infrastructure to facilitate the build out of their own infrastructure at considerably lower cost. Indeed, this is exactly the issue that has been raised by Google in their current fiber deployment efforts [71]. Furthermore, an important consequence of the additional sharing of long-haul infrastructure that will likely take place if the Title II classification is upheld is a significant *increase*

in shared risk. We argue that this tradeoff between broader metro-area fiber deployments (*e.g.*, Google) and the increased risks in shared long-haul infrastructure requires more careful consideration in the broader Title II debate.

5.6.3 Enriching US Long-Haul Infrastructure

On the one hand, our study shows that the addition of a small number of conduits can lead to significant reductions in shared risk and propagation delays. At the same time, our examination of public records also shows that new conduit infrastructure is being deployed at a steady rate. Assuming that the locations for these actual deployments are based on a combination of business-related factors and are not necessarily aligned with the links that our techniques identify, the question that arises is how the conduits identified in our analysis might actually be deployed.

We believe that a version of the Internet exchange point (IXP) model could be adapted for conduits. IXPs largely grew out of efforts by consortia of service providers as means for keeping local traffic local [191]. We argue that the deployment of key long-haul links such as those identified in our study would be compelling for a potentially large number of service providers, especially if the cost for participating providers would be competitive. At the same time, given the implications for shared risk and the critical nature of communications infrastructure, government support may be warranted.[7] In fact, the involvement of some states' DOTs in the build-out and leasing of new conduits can be viewed as an early form of the proposed "link exchange" model [30].

[7]Similar arguments are being made for hardening the electrical power grid, *e.g.,* http://www.wsj.com/articles/grid-terror-attacks-u-s-government-is-urged-to-takes-steps-for-protection-1404672802.

5.7 Summary

In this chapter, we study the Internet's long-haul fiber-optic infrastructure in the US. Our first contribution is in building a first-of-its-kind map of long-haul infrastructure using openly available maps from tier-1 ISPs and cable providers. We validate the map rigorously by appealing to public information sources such as government agency filings, environmental impact statements, press releases, and others. Examination of the map confirms the close correspondence of fiber deployments and road/rail infrastructure and reveals significant link sharing among providers. Our second contribution is to apply different metrics to examine the issue of shared risk in the long-haul map. Our results point to high-risk links where there are significant levels of sharing among service providers. Our final contribution is to identify public ROWs that could be targets for new link conduits that would reduce shared risk and improve path performance. We discuss implications of our findings in general and point out how they expand the current discussion on how Title II and net neutrality.

6

A Techno-Economic Approach for Broadband Deployment in Underserved Areas

6.1 Introduction

The focus of this chapter is to develop a decision support framework to satisfy ISP objectives (e.g., identify target areas with the highest concentration of un/underserved users at the the lowest cost to service providers for network infrastructure deployment), with an emphasis on broadband deployments.

The importance of broadband connectivity in the US is highlighted by the following quote from the FCC's National Broadband Plan, "Like electricity a century ago, broadband is a foundation for economic growth, job creation, global competitiveness and a better way of life" [54]. Despite the compelling case for broadband access and significant efforts by the FCC over the past six years, 6% of the Americans still lack access to broadband service (threshold defined to be 25 Mbps download/3 Mbps upload for fixed services) and the percentages are much higher in rural and tribal areas [55].

Expansion of broadband access in the US, as it is in other states, is a complex matter. First, the FCC does not build, own or operate Internet infrastructure. Instead it works with municipalities, private service providers and other sponsors by providing guidance and economic in-

centives to deploy broadband infrastructure in un/underserved areas (*e.g.*, via the Connect America Fund [33]). Second, there are legal and policy concerns such as laws that limit or prohibit non-telecom companies from deploying communication infrastructure [62]. Third, defining and identifying underserved areas that are the *best targets* for new or upgraded infrastructure deployment requires consideration of a variety of geographic, economic and demographic factors—the main focus of this work.

In this chapter, we describe a techno-economic framework and system for identifying targets for future broadband expansion. The objective of our work is to provide flexible decision support on opportunities for broadband deployment that enables economic and technical issues to be considered simultaneously. Specifically, our framework considers *(i)* infrastructure proximity, *(ii)* demographics, and *(iii)* deployment costs. We employ geographically-based, multi-objective optimization to identify the *highest* concentrations of un/underserved users and that can be upgraded to the broadband threshold at the *lowest* cost. Our work takes advantage of new maps of long-haul infrastructure in the US (§5) that are critical for accurate cost modeling.

We demonstrate the efficacy of our approach by considering US demographic data and two different deployment models: upgrading existing infrastructure and deploying new infrastructure. Our results highlight the tradeoffs of the different deployment models and identify a list of US counties that would be attractive targets for broadband deployment from both cost and impact perspectives and that correspond closely with areas identified by Connect America map [34]. While our analysis focuses on the US, our method is generic and can be applied in other regions where similar data is available.

6.2 Connectivity Analysis

In this section, we assess connectivity and need in US counties (and county equivalents) using *provider data* [2] from broadbandmap.gov, *census data* [151] from census.gov and *infrastructure data* from Internet Atlas [204, 206]. Our analysis considers the presence of Internet Service Providers (ISP) and the characteristics of user populations in counties. We spatially integrate the infrastructure datasets from Internet Atlas and census.gov to highlight the presence of “digitally divided" regions across US.

6.2.1 Service Provider Prevalence

Similar to [54], our analysis of connectivity begins by counting the number of providers with presence in US counties. First, we extract population information and FIPS codes of 3,142 US counties using census data. Next, we look up FIPS code in provider data and count the unique number of service providers present in each county in the form of a broadband/fiber provider or a reseller.

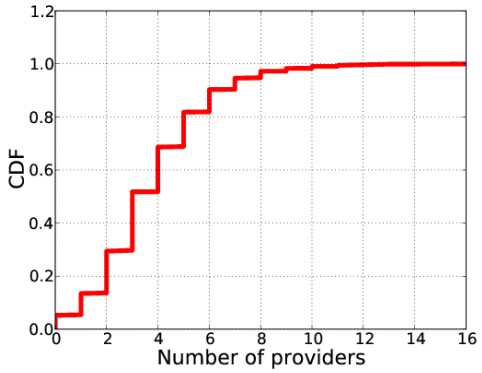

Figure 6.1: CDF of number of providers with presence in 3,142 US counties (and county equivalents).

Figure 6.1 shows the distribution of service providers in US counties.

For 50% of the counties, the number of service providers present is less than or equal to 3. Surprisingly, 170 counties do not have *any* provider presence. These counties are spread across 30 states leaving 38,464,508 users—or 12% of the US population[1]—disconnected from the Internet, which is consistent with observation from others [3]. Finally we observe that less than 1% of the counties (across 17 states) have provider presence greater than 10. Manual comparison with physical infrastructure repository and fiber assets [204, 206] showed that the increased presence of providers in these locations corresponds with the presence of either *(1)* a co-location facility, an Internet Exchange Point (IXP) and/or a submarine cable landing station, or *(2)* high availability of fiber resources to meet large user demand (*e.g.*, a major metropolitan area).

6.2.2 Infrastructure vs. Population

We compare the availability of infrastructure versus population to assess the prevalence of underserved communities. Similar to [54], we use the unique number of service providers with a presence in a county as a proxy for the infrastructure availability. Our intuition for this analysis is that the trend in population should be proportional to number of unique providers to completely connect *all* communities in a region.

Figure 6.2 depicts the normalized population versus the normalized infrastructure availability in US counties. The expected and the actual deployments are also shown. The plot highlights the fact that there are a sizable number of population centers in the US that have infrastructure provided by a small number of ISPs.

A natural question is can a region with only one service provider effectively serve and provide broadband access to every community in that region? Even though such a scenario is possible, we argue that the geographical diversity of infrastructure deployments will suffer as a con-

[1]Based on projected US population of 320,090,857 on Jan. 01, 2015 [119].

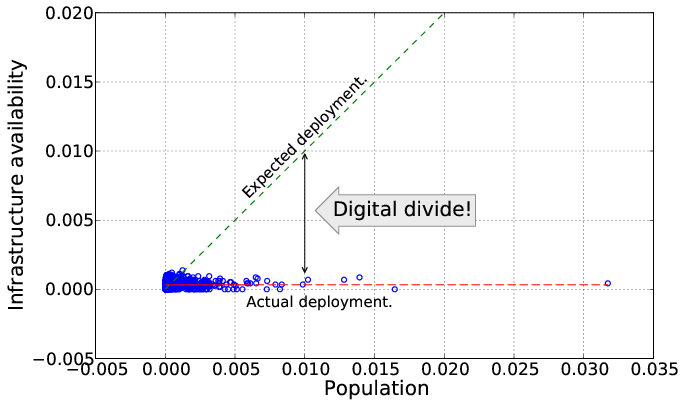

Figure 6.2: Normalized population vs. normalized infrastructure availability in the US counties along with expected and actual deployments.

sequence of one-provider-services-all model since business imperatives may lead to delays in broadband deployments to all communities. It may also lead to choke points and single points of failure in the Internet [204] that may otherwise be obviated in more competitive areas.

6.2.3 Availability of Infrastructure

Finally, we consider the issue of level of service in an area by using a Geographic Information System (GIS) to spatially integrate areas of counties from census.gov and physical infrastructure assets from *(1)* the Internet Atlas and *(2)* the long-haul infrastructure information from the InterTubes projects (§5). Our objective is to analyze the proximity of population centers to infrastructure for network connectivity. To facilitate this analysis, we use the *spatial query* and *overlap* capabilities in ESRI ArcGIS [50].

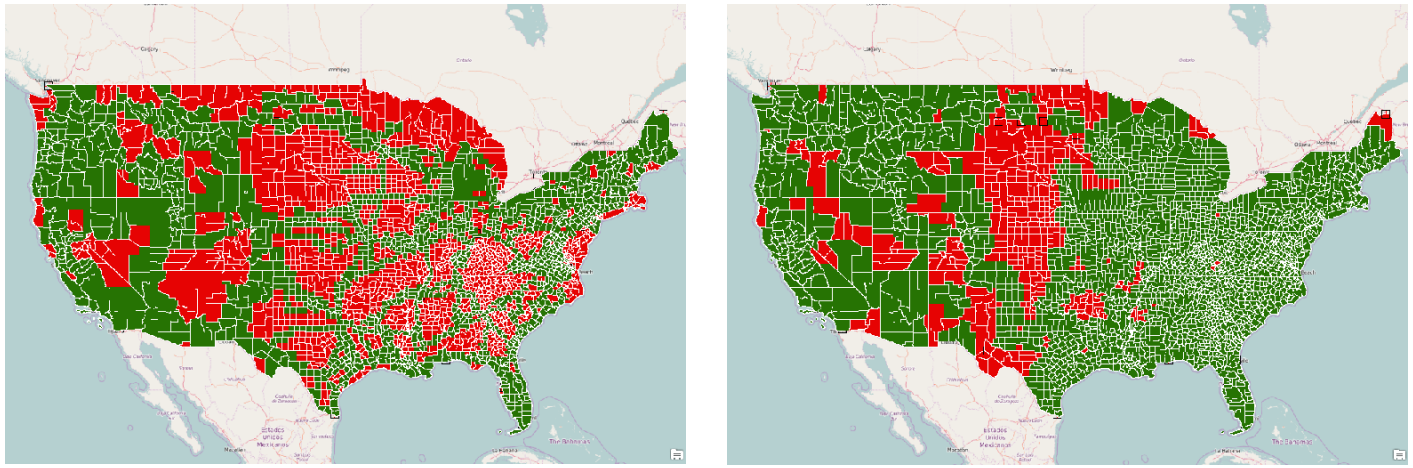

Figure 6.3: Spatial selection of counties using D1 [204] (left) and D2 [206] (right) datasets. Counties with and without infrastructure are shown in green and red respectively.

We start by layering the infrastructure shape files from Internet Atlas and InterTubes atop the counties. We invoke *spatial overlap* and *select by location* queries on these spatially integrated datasets. Figures 6.3-(left) and -(right) distinguish the digitally disconnected regions (in red) from those that are well connected (in green) to physical long-haul fiber data from InterTubes (dataset D1) and 100 US-based networks from Internet Atlas repository (dataset D2) respectively. We call these the *infrastructure availability map*. These maps form the basis of our targeting assessments described below.

6.3 Deployment Objectives

In this section, we state the broad objectives for an ISP's operational success, which are important in understanding how to create incentives for broadband deployment in un/underserved areas.

Maximize value. The primary objective of any company is to maximize shareholder value. The question is how ISPs go about doing this? While large ISPs have complex business models that are beyond the scope of this paper, several key factors including revenue growth, cost management, customer satisfaction, and maintaining technological and operational capabilities.

Growing the user base. Revenue growth can be directly tied to expansion of an ISPs user base. This can be done in a variety of ways including expanding infrastructure into previously unserved or underserved areas or by upgrading capabilities that allow for higher service charges. Expanding the user based is one of the primary motivations for expanding to un/underserved areas.

Minimize CAPEX. Expanding or upgrading infrastructure is capital expense (*i.e.,* an investment that depreciates over time) for ISPs. Many factors must be considered before making capital expenditures including

(1) proximity/type of current infrastructure, *(2)* geographical feasibility (mountains vs. existing right of ways) and *(3)* market economics and competition. CAPEX is one of the primary deterrents to expanding to un/underserved areas.

Minimize OPEX. Operational expense (OPEX) refer to costs associated with operating and maintaining an infrastructure. A variety of factors contribute to OPEX including environmental factors (*e.g.*, power, cooling, etc.), miscellaneous factors (*e.g.*, taxes, repairs, etc.) and personnel costs. Economies of scale for OPEX argue for expanding to un/underserved areas.

Minimize risk. Any infrastructure or service expansion implies CAPEX and OPEX commitment. Any analysis of the opportunities for increased revenue through new user service adoption must be complemented by an analysis of the risks associated with deployment and operating costs. The more accurate these analyses, the more likely service providers are to commit to expansion. This is one of the goals of our work.

6.4 Techno-Economic Framework

In this section, we describe our geo-based optimization framework that guides infrastructure deployment in new geographic locations. We identify two deployment scenarios that are affordable for the end users and that are practical and cost-effective for the ISPs. We conclude this section with an evaluation of the identified scenarios using our framework.

6.4.1 Techno-Economic Model

We consider the problem of assigning a list of *nodes* to a list of *locations*, where our objective is to assign each node (*i.e.*, network infrastructure) to a location such that the total cost is *minimized* and the number of users[2]

[2]For simplicity, we simply consider the total population in a target area.

is *maximized*. This is an extension of Koopmans-Beckmann version of the Quadratic Assignment Problem (QAP) [252] where, apart from the objective of minimizing costs associated with a node assignment to a location, we also consider maximizing number of end users who could benefit from the *new* deployments.

Note that the objective of maximizing the number of users is in conflict with the objective of minimizing total costs. For example, more users implies a larger infrastructure and thus higher the total costs for CAPEX and OPEX (unless one further assumes a per user revenue model, which we argue is not of intrinsic importance to this step in the analysis—revenue modeling including incentives can be done post-facto). Because of the conflicting nature of these two objectives, we model the assignment problem as a multi-objective optimization problem, subject to various technical, economical and ISP-centric constraints. Specifically, given a list N of k nodes, where N is defined as,

$$N = \{n_1, n_2, n_3, \ldots, n_k\}$$

the multi-objective problem can be formulated as,

$$\textbf{max.} \sum_{i=n_1}^{i=n_k} B_{i\gamma(i)} + \textbf{min.} \sum_{i=n_1}^{i=n_k} C_{i\gamma(i)} \tag{6.1}$$

subject to the following constraints,

$$Budget_{min} \leqslant C_{i\gamma(i)} \leqslant Budget_{max}, \quad \forall i = n_1, \ldots, n_k \tag{6.2}$$

$$k \leqslant K \tag{6.3}$$

where, $B_{i\gamma(i)}$ is the benefit factor to users at location $\gamma(i)$ for deploying a node i, $C_{i\gamma(i)}$ is the total cost of deploying node i at location $\gamma(i)$, $Budget_{min}$ and $Budget_{max}$ are the minimum and maximum budgets allocated for deployments, and K is the maximum number of deployments

planned by the ISP.

Implementation. The optimization model described above is implemented in approximately 450 lines of python code using the DEAP evolutionary computation framework [216]. DEAP enables rapid prototyping of *any* evolutionary algorithm with minimal developer efforts.

Advantages. Our optimization framework has the following advantages: (1) *flexibility*, where equation 6.1 can be extended to accommodate other objectives such as considering only a subset of user population (*e.g.*, based on economics) and the ones described in §6.3 instead of maximizing the number of users; (2) *simplicity*, where the cost and benefit factors can be varied as per service provider's requirement; and (3) *modularity*, where different evolutionary algorithms can be plugged in to perform a wide spectrum of analyses.[3]

6.4.2 The Solutions

To facilitate our deployment analysis, we studied solutions proposed by researchers and consider both the practicality and cost-effectiveness of each. First, we study solutions including *(a)* WiMax [77]; *(b)* radio wave mesh-based networking [52]; *(c)* li-fi technology [144]; and *(d)* satellite-, balloon- and aircraft-based networking [53, 72]. Our conclusion is that these technologies are quite costly for deployments that cover broad geographic areas, which are common in underserved areas. For example, a typical satellite deployment costs about $500M and includes high equipment costs ($150-200M), high maintenance and operational costs ($120M for launch, $20M for launch insurance, $20M for in-orbit insurance, $15M for operations, and special manpower at about $10M a year per specialist) [127]. It is somewhat surprising that industrial projects [53, 72] continue to push at these solutions despite the challenges and practicality issues.

[3] In our evaluation, we use NGSA-II evolutionary algorithm [200] with Ant heuristics.

Next, we investigated a set of technologies that are more cost-effective and practical. To that end, we consider the following two options: (1) connect existing transmission infrastructure (*e.g.*, public switched telephone network (PSTN) or cable television network) to IP infrastructure using Multi-service Access Node (MSAN) at strategic locations and use cable or DSL modems at the user end; or (2) leverage power line infrastructure to enable connectivity. Even at locations where PSTN is not installed, there are almost always power lines installed, which enables broadband over power line (BPL) or distribution line carrier (DLC). Since the latter is proven successful and is already the goal of many companies [97], in our evaluation we only explore the former (scenario 1 below). Finally, to add perspective, we also consider the scenario where a service provider is willing to invest on building new fiber infrastructure to connect a region.

6.4.3 Evaluation

Scenario1: Upgrading existing infrastructure. We first examine the possibility of leveraging existing infrastructure (*e.g.*, PSTN and cable network) to connect un/underserved counties. We augment the GIS-based approach described in §6.2.3 with other analysis capabilities in ArcGIS and QGIS to identify new deployment locations. Specifically, for this scenario, we leverage the *hub distance* tool in MMQGIS [103] to identify a number of locations that do not have any connectivity and that could be cost-effectively connected to other areas with connectivity in Figures 6.3. By using the infrastructure availability map as input to the hub distance tool, we create hubs in green polygons which serve as the deployment location for MSANs. These MSAN locations are connected to the nearest red polygon, which indicates the absence of connectivity. Figures 6.4-(left) and -(right) depicts the hub-based deployments for datasets D1 and D2 respectively.

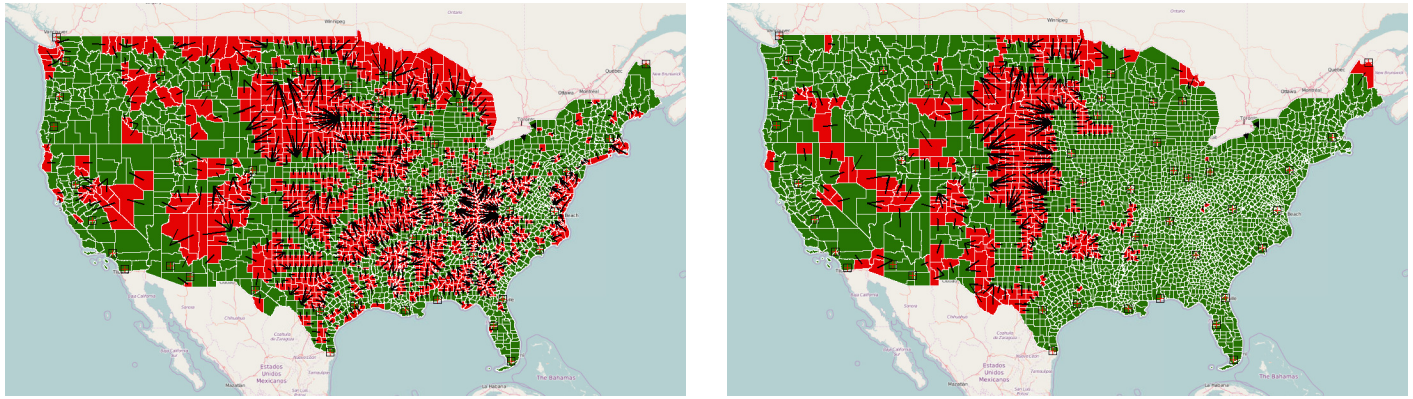

Figure 6.4: Spatial selection of counties using D1 (left) and D2 (right). New hub-based MSAN upgrades are shown in black.

Since all the identified hub-locations cannot be connected, as it is impractical in terms of cost, we apply our techno-economic framework to maximize connectivity with minimum deployment costs for a given deployment budget. For this scenario, we assume that the cost[4] of an MSAN is $100K and that the telephone and cable networks are available in all the un/underserved areas. We also assume that the cost to connect a household with a modem is $25 and that the cost to connect network access points in underserved regions to the households in every region is negligible. So, the cost to connect a region using this scenario is simply the sum of MSAN costs at hubs divided by the number of counties sharing that hub plus the cost to install modems in every household in a region. We set the maximum deployment budget per location to be $100K.

Figure 6.5 shows both the Pareto-optimal or non-dominated solutions (in red) and the evolution of these solutions (in blue) for this scenario. For example, based on our cost model for hub-based deployment, a little over than 4.2M users in all (red) counties in D1 can be connected at a cost of $2.2M. Note that all the Pareto-optimal solutions are also globally optimal solutions. By analyzing the tradeoff between the multiple objectives and depending on the deployment budget, the network operator can choose a particular solution to make an appropriate deployment decision.

[4] All costs in our study are based on personal communication with network operators [104].

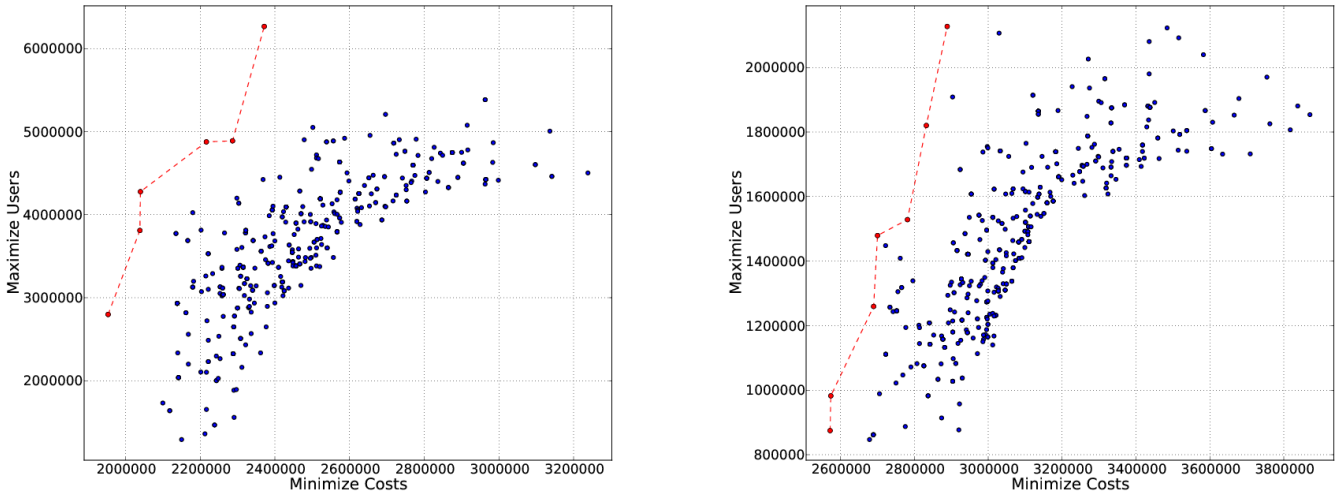

Figure 6.5: Deployment solutions produced by our framework for D1 (left) and D2 (right). The evolution (blue) and pareto front (red) of the solutions are also shown.

Scenario2: Deploying new infrastructure. In this scenario, we assume that the service provider is considering building their own infrastructure in un/underserved areas by deploying fiber assets along the existing ROWs (*e.g.*, road and rail). We begin by layering the ROW shapefiles [206] on top of the infrastructure availability map and use spatial overlap capability to select only those ROW features that intersect with the regions that do not have any connectivity. Next, we use the *cost distance* capability in ArcGIS and create a low-cost minimum spanning tree of the ROW features to identify the new fiber ROW deployments. Figures 6.6-(left) and -(right) shows the ROW-based fiber deployments for datasets D1 and D2 respectively. As one might expect based on results in [206], the resulting infrastructure bears a striking resemblance to current fiber deployments.

Next, we apply our techno-economic framework to create a more optimized deployment scenario. For this scenario, we assume CAPEX cost of fiber per mile is $1500 and OPEX per mile per year is $300. So, the cost to connect a region is the sum of fiber miles multiplied by these costs. Our objective for this scenario is to minimize these costs. Figures 6.7 plots both the Pareto-optimal solutions (in red) and the evolution of these solutions (in blue) for the ROW-based fiber deployment scenario. Based on our cost model for this scenario, a little less than 5M users in all (red) counties in D1 can be connected at a cost of about $14.5M.

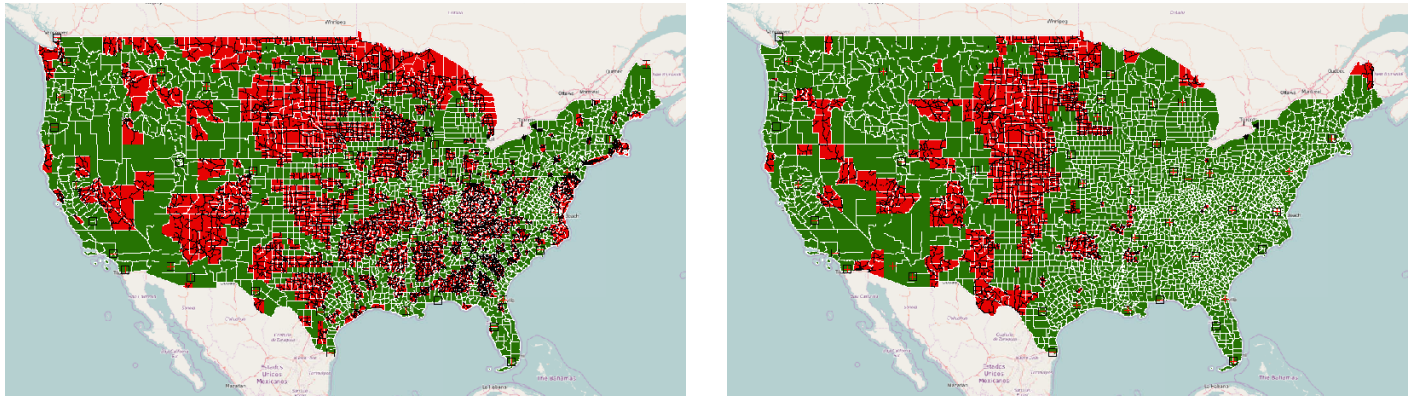

Figure 6.6: Spatial selection of counties using D1 (left) and D2 (right). New fiber-based ROW deployments are shown in black.

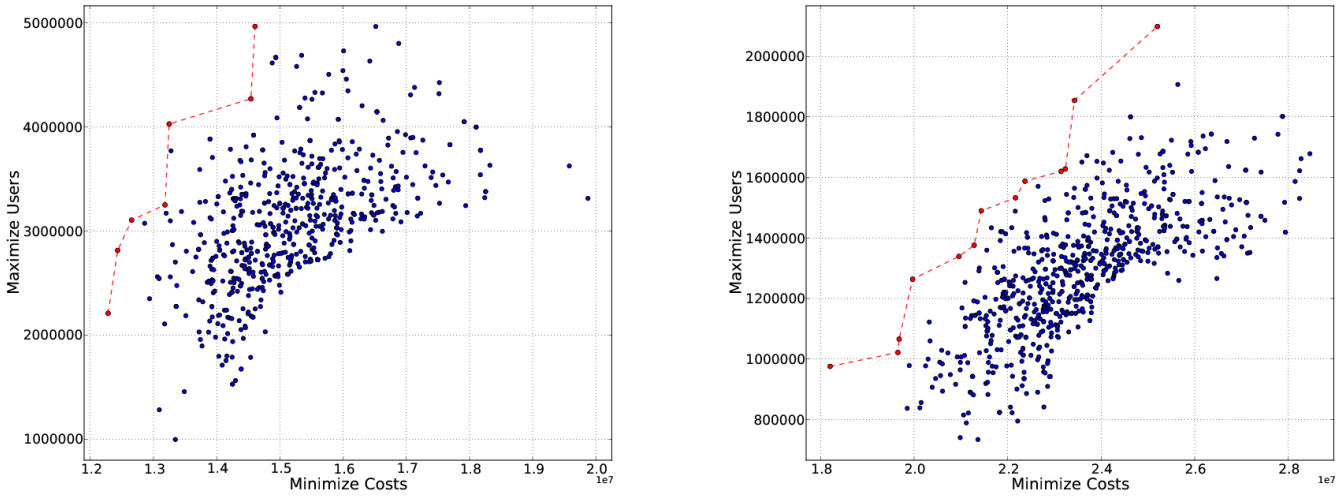

Figure 6.7: Deployment solutions produced by our framework for D1 (left) and D2 (right). The evolution (blue) and pareto front (red) of the solutions are also shown.

Top 20 deployment targets. Based on the above two scenarios, we identified the following 20 counties based on its occurrence across both the scenarios. The counties include Niobrara, Lamar, Weston, Hot Springs, Foard, Crook, Coahoma, Washakie, Val Verde, Slope, Schleicher, San Juan, Roosevelt, Panola, Newton, Mono, Mercer, McDonough, Los Alamos, and La Paz. Unsurprisingly, these counties are rural areas that are predominantly located in states like Texas, Wyoming, North- and South-Dakotas—an observation consistent with prior work [99].

Validation with Connect America map. To validate our methodology for selecting deployment targets, we compare the regions identified by the FCC's Connect America Fund for phase II funding [34] and the ones identified by our framework. Specifically, we calculate the percentage of agreement between the FCC's accepted areas dataset and the counties identified by our analysis above. For D2 dataset, our framework has 86.62% agreement with that of the accepted areas (395 out of 456 counties). Similarly, for D1 dataset, 1405 out of 1521 counties (*i.e.*, 92.38%) identified by Connect America agrees with our analysis.

In short, these results, apart from validating our framework, shows that the funding attempt by Connect America is progressing in a way that is balancing deployment in areas with a large number of users with costs. Note that we see a higher percentage of agreement for D1 because counties listed by Connect America are based on long-haul providers, which is the main focus of D1 dataset. More broadly, we believe that this comparison highlights the utility of our framework and the potential for its application in other areas and under a wide variety of cost/impact assumptions.

6.5 Summary

In this chapter, we consider the problem of identifying target areas for network infrastructure deployment in un/underserved areas. Our

techno-economic approach applies geo-based multi-objective optimization to find the areas with the highest concentration of un/underserved users at the the lowest cost to service providers. We demonstrate the efficacy of our methodology by considering physical infrastructure and demographic data for US counties along with deployment cost models that include upgrading existing infrastructure and deploying new infrastructure. While we do not argue that the quantitative aspects of our cost models are representative of any specific service provider, our results identify a list of counties that would be attractive targets for broadband deployment and that correspond closely with those already identified for future deployments in the US.

7

GreyFiber: A System for Providing Flexible Access to Wide-Area Connectivity

7.1 Introduction

The focus of this chapter is to develop a system to minimize shared risk in the Internet by borrowing concepts from cloud computing. The premise of cloud computing is that instead of building and maintaining their own in-house computing and storage infrastructures, users (*e.g.*, companies, organizations, individuals) can consider and consume compute resources as a utility. "The network is the computer" is a much-used phrase that was coined to succinctly describe this utilitarian approach that has been the driving force for much of the ongoing "cloudification" of today's Internet. For users, the benefits of relying on the network to perform tasks that traditionally ran in local compute environments are all too obvious. For one, the ability to spin up compute resources on demand for pretty much any type of task or workload offers enormous flexibility for self-service provisioning by relieving the users from having to make decisions about the physical placement of the desired compute resources and their interconnectivity. Second, being able to scale up or down as computing needs increase or decrease provides a degree of elasticity that conventional, statically-provisioned in-house compute facilities cannot

match unless prohibitively expensive over-provisioning of local compute resources is acceptable. Lastly, the commonly-adopted pay-per-use billing model for cloud computing (*i.e.*, users only pay for the resources they actually consume) has proven to be a key economic incentive and responsible for much of the recent emergence of an entire cloud-related ecosystems (*i.e.*, cloud providers, cloud services).

However, these benefits also have a direct impact on the type of traffic that is generated in a cloud-centric Internet and on how that traffic is routed over the existing physical Internet infrastructure (see for example [250, 257]). Consider for example the simple case of different users spinning up virtual machines (VMs) for running big data analytics applications that require the transfer of large datasets from a geographically-dispersed set of database servers, possibly with additional performance- or security-related requirements (*e.g.*, low-delay, resilience to outages, avoiding certain networks or regions). Such transfers can potentially consume significant portions of the available bandwidth along their routes, but the onus is squarely on the user's cloud provider or on that cloud provider's transit provider to ensure that the user's application gets the necessary data as required. Traditionally, traffic engineering and routing have been used to address such issues (*e.g.*, see [177, 199, 217, 240, 241, 258, 261, 311]), but what if the nature of the generated traffic is such that it periodically exceeds the available capacity on the primary and backup paths and no alternative paths are available?

There have been recent efforts to study the problem of dealing with highly variable and unpredictable workloads in inter-datacenter (inter-DC) WANs such as those operated by Google or Microsoft. In particular, B4 [237] and SWAN [233] leverage SDN technology and rely on a wide area network view to dynamically change routing and rate allocations to ensure high network utilization while meeting the deadlines of the data transfers. However, by assuming a fixed network- or router-level WAN

topology, these efforts ignore the opportunities that arise from reconfiguring equipment in the underlying optical or physical layer to dynamically change the router topology. Such a joint (and central) control of both the physical and network layer has recently been considered in [239] where the authors describe Owan, a new SDN-based system for orchestrating bulk transfers that computes and implements the optical circuit configuration (*i.e.*, the optical circuits that implement the network-layer topology) and the routing configuration (*i.e.*, the paths and rate allocation for each transfer) to ensure high network utilization and optimize bulk transfers.

In this chapter, we move beyond [239] and borrow a page from cloud computing. In particular, we describe the design and implemention of *GreyFiber*, a new platform for establishing fiber-optic connectivity in the Internet. Similar to how the cloud enables arbitrary users to spin up VMs as needed, GreyFiber makes it possible for infrastructure providers to spin up optical circuits on demand to handle the highly variable and unpredictable workloads that a cloud-centric Internet entails. In a sense, GreyFiber is to the wide-area Internet as 3D beamforming is to DCs [316]. While the technologies, economics, and operations underlying these two approaches differ drastically, their objectives are the same. That is, to alleviate traffic hotspots as they occur as the result of highly unpredictable traffic, the original (fixed) means of data communication is complemented by unused communication channels that are made available as needed—idle optical circuits in the case of GreyFiber in the wide-area Internet, and idle wireless links in the 60 GHz band for 3D beamforming in DCs. In fact, where available, GreyFiber could include the sort of microwave communication that is used for high-frequency trading applications between New York and Chicago [61, 75] (see also [289])[1].

The main idea for GreyFiber is to provide a means to offer easy and cost-effective access to unused fiber-optic paths between participating

[1] In this chapter, our focus is on utilizing unused optical circuits.

endpoints (*e.g.*, colocation facilities) on demand, for arbitrary durations, and possibly with industry-specific performance guarantees (*e.g.*, ultra-low delay for high-frequency trading applications or gaming services; fully diverse physical paths for mission-critical business applications). In this sense, GreyFiber can be thought of as offering *wide area connectivity as a service*. However, GreyFiber differs from standard cloud computing services (*e.g.*, SaaS, PaaS and IaaS) in that it is fundamentally concerned with connectivity, not computation. In the rest of the chapter, we use the following terminology. The unit of connectivity in GreyFiber is a *link* which refers to a single strand of fiber. A link may contain one or more *circuits*, which are defined as logical connections across endpoints with unique wavelengths and which are configurable sub-units in GreyFiber. Multiple links are bundled in a *path* (also known as a *conduit*) and each path/conduit is physically installed between endpoints at distinct geographic locations.

The design of GreyFiber requires the careful integration of three critical components. First, to ensure that GreyFiber is an economically viable option, we monetize the current over-supply of buried fiber in existing conduits in today's physical Internet infrastructure [140, 155][2] by proposing an auction-based Fiber Exchange that attracts potential buyers and sellers of GreyFiber. This idea is based on the insights gained from Internet Atlas (§3) and InterTubes (§5) projects. Second, we leverage the fact that fiber-optic technology has advanced to the point where today's fiber-optic gear allows fast remote reconfigurations. For example, provisioning of an idle circuit can be done on the order of milliseconds to seconds [82, 83, 109, 162, 193] which suggests that spinning up an optical circuit between two participating endpoints can be achieved at time scales that are commensurate with those required for launching a cloud service. Finally, the operation of our GreyFiber platform is inspired by prior work [239] and relies on a central controller that allows for direct

[2]Our focus here is strictly US-centered.

and end-to-end control of all GreyFiber-affected devices and simplifies overall network management.

To demonstrate the feasibility of our approach and examine its efficacy, we describe an implementation of our GreyFiber design and deploy it in the GENI testbed. This prototype system addresses the technical challenges associated with circuit provisioning and enables performance evaluation over a range of use scenarios. First, we show that as many as 50 paths can be provisioned between endpoints in less than a minute, which demonstrates the scalable and rapid provisioning capabilities of GreyFiber. Next, to enable higher infrastructure resilience during network outages and/or planned maintenance events, we show how GreyFiber can be used to create an effective backup solution. Specifically, GreyFiber can reactively detect path failures and provision a new path within 1.25s, which outperforms the traditional OSPF-based backup solution by 28x. This agility of GreyFiber benefits many applications by allowing them to be oblivious to underlying network failures. Finally, we dynamically provision paths between endpoints to create on-demand high-capacity connectivity and demonstrate the resulting performance benefits of GreyFiber.

We quantify the overhead of our system versus the underlying infrastructure and highlight the critical path performance of GreyFiber. By examining the log files produced during circuit provisioning, our analysis shows that GreyFiber has minimal system overheads. We find that the latency overhead for on-demand path provisioning is completely dependent on the underlying network substrate (*e.g.*, hardware), which highlights avenues for improvement and expansion of the range of use scenarios of GreyFiber in the future.

7.2 The Case for GreyFiber

Over the past several years, there have been significant changes among network service and infrastructure providers that motivate the timeliness of *wide-area connectivity as a service* embodied in GreyFiber.

Consolidation of dark fiber providers. There has been a trend toward consolidation among dark fiber providers. Examples include CenturyLink's acquisition of Qwest in '11 (resulting in a combined 190k mile fiber network [159]), Zayo's acquisition of Abovenet in '12 (resulting in a combined 6.7M fiber mile network connecting some 800 datacenters [160]), Level 3's acquisition of tw telecom in '14 [94], Lighttower merging with Fibertech in '15 [96], CenturyLink's acquisition of Level 3 in '16 [22], and Verizon's recent announcement to acquire XO communications' fiber-optic network business [153]. A clear consequence of these mergers is that *there are fewer fiber-optic network providers, but the remaining ones have larger fiber footprints*.

Evolution in the datacenter market. There has been consolidation as well as expansion within the datacenter market. Among the tier-1 datacenter providers (*i.e.*, serving major metro areas and large cities), examples of consolidation include Equinix' acquisition of Telecity Group (EU/UK) [165] and Bit-Isle (JP) [164] in '15, Digital Reality Trust's acquisition of Telx [163] in '15, AT&T announcement to sell datacenter assets [167] in '15 and Windstream announcement to sell its datacenter business to TierPoint [168] in '15. At the same time, the growing demand for cloud services has put pressure on the largest cloud providers to have presence in more locations and also closer to their customers, which has led to the emergence of an increasing number of new 2nd-tier datacenter providers (*e.g.*, EdgeConneX [48]) that are focused on medium-sized markets such as Portland, OR and Pittsburgh, PA. The combined effects of this cloud-driven, broader user-base and higher volatility of workloads could be mollified via GreyFiber. These trends indicate an *expanding geographic*

distribution of datacenter capacity that could benefit from GreyFiber connectivity.

Dark fiber providers acquiring datacenters. There are recent examples of dark fiber providers acquiring datacenters, which presents an opportunity for one provider to supply high-bandwidth connectivity between datacenter co-location endpoints to customers who need it. One example of a provider with this nascent capability is Lightower [35], which acquired ColocationZone in '15 [166] and Datacenter101 [169] in '16. Similarly, Allied Fiber[3] aimed to be a network-neutral and dark fiber "superstructure" with a footprint across the United States and offered traditional 20-year and non-traditional 12, 24, and 36-month Indefeasible Rights of Use (IRU) options [8, 9]. *These developments indicate that there exist business opportunities for companies that offer integrated (network-neutral) colocation/dark fiber services and that could benefit from available GreyFiber connectivity to boost their existing but maybe constrained dark fiber infrastructure.*

Implementation challenges. Our framework for GreyFiber includes three high level aspects: a fiber exchange, a circuit provisioning system and a central controller. Each component has its own technical challenges to enable scalable use across diverse physical infrastructures. In most respects, the fiber exchange has the same requirements as other auction-based systems (*e.g.,* Amazon EC2 spot pricing system [12]), and indeed those provide a blueprint for our GreyFiber prototype described in §7.3. Next, driven by demands in datacenters, new optical switching equipment is being designed to speed and simplify configuration and management of optical connections [162]. For example Infinera's Open Transport Switch [82] is a software layer that runs on top of of their optical cross connect hardware to enable fiber-optic wavelengths to be put into service on demand. We believe that this trend in switch technology will continue in the future and this sort of equipment and capability is a key enabler for GreyFiber. Finally, the global controller must coordinate between user

[3] Allied Fiber is now defunct primarily because they were not able to build an adequate customer base quickly enough [10].

requirements and the underlying physical infrastructure to ensure that service commitments are satisfied. These requirements are akin to SDN controllers, which serve as a model for our GreyFiber prototype (§7.3).

Incentives for GreyFiber. While corporate and technical trends indicate the opportunity for GreyFiber, practical incentives motivate broader deployment and use. We consider the incentives for GreyFiber versus IP transit (*i.e.*, lit fiber) and dark fiber, which are the standard fiber options in the Internet today. In particular, we compare and contrast the three market options using five different metrics: economic incentives; potential market size; control over routing, physical route diversity and control over performance.

Table 7.1: Incentives of GreyFiber-based fiber market in comparison with the IP transit and dark fiber options (L = low, M = medium, H = high).

	Dark Fiber	IP Transit	GreyFiber
Economic Incentives	L	M	H
Potential market size	L	M	H
Control over routing	H	L	M
Physical route diversity	M	L	H
Control over performance	M	M	M

Table 7.1 shows a relative comparison between the GreyFiber and other fiber markets. Based on the IP-transit and dark fiber price sheets compiled from three different US service providers, we posit that dark fiber has the lowest economic incentive if one considers a broad set of customers. First, there is the required 20+ year commitment for an IRU, which locks in capital expenditures (CAPEX) and operational expenditures (OPEX) over that duration. The standard pricing model for dark fiber includes an upfront payment for the IRU along with substantial CAPEX to light fiber. Reoccurring costs include CAPEX at ~$1000 to 3000 per mile per year and OPEX at ~$250 per mile per year. These costs and the duration of the commitment tend to reduce market size. Benefits of dark fiber include

control over routing, physical route diversity and low latency due to direct interconnection to peers at the colocation facilities.

Fiber pricing in the IP transit market is ~$500 to $600 per Gbps per month. Benefits include medium-term commitments (3–5 years) for fully managed services, no OPEX or CAPEX. The one-stop shopping, fully managed service aspect of IP transit leads to a medium sized market. The drawbacks include (*1*) no physical route diversity (unless explicitly specified at additional cost), (*2*) no routing control (*3*) latency determined by SLA, which may not be sufficient due to indirect routing and lack of direct interconnection at peering points.

In this chapter, we assume GreyFiber will initially be offered in auction-based exchanges as managed layer-3 services. Thus, the benefits of GreyFiber include *(1)* a flexible pay-as-you-go model and no upfront costs, which opens the fiber market to a potentially large customer base; *(2)* the ability to choose diverse routes; and *(3)* control over performance (*i.e.*, low latency) due to direct interconnection with peers. As a consequence, the only drawback is that the customers will likely have limited control over routing.

Use cases. We envision three use cases for GreyFiber: *(1)* improving network resilience through redundant connections, *(2)* providing (ultra) low latency paths, and *(3)* providing on-demand high-capacity paths over arbitrary durations. Internet outages are common and occur due to a variety of reasons including accidents, misconfigurations and censorship (*e.g.*, [158, 235, 266, 318]). Outage can be mitigated by temporary paths that reconnect points within a network. Addition of a long-haul path might also be considered as a preemptive measure in the case of a planned maintenance outage, or knowledge of an impending weather event that may affect the network. Next, a reduction of milliseconds or even microseconds in latency can yield competitive advantages in the financial sector or in gaming. The addition of new fiber links through GreyFiber may

be used to provide more direct paths and thereby reduce end-to-end latency. Finally, the need to transfer (large) data sets across the wide-area Internet or between datacenters is likely to continue to grow. Improving throughput and scheduling of large inter-datacenter transfers has been the subject of recent research [238, 258, 306], and could benefit from additional high-capacity paths via GreyFiber.

For each of these motivating use-cases, there may be quite different requirements in terms of capacity and the time duration over which the additional capacity is needed. For example, (*i*) short lifetime capacity to address an unexpected outage, (*ii*) short lifetime capacity to address unexpected demand, (*iii*) short lifetime capacity to enable better performance between two points, (*iv*) medium lifetime capacity to service expected demand that has no specific deadlines, (*v*) short-to-medium lifetime capacity for transit, backhaul, etc. We believe that these scenarios create a compelling case for the utility of on-demand connectivity offered by GreyFiber.

7.3 GreyFiber System Design

In this section, we describe the system requirements and design, auction model used, and events that take place when provisioning new paths through GreyFiber.

7.3.1 System Requirements

A GreyFiber system must satisfy the following requirements:

Scalability and extensibility. A GreyFiber system must scale to meet the demands of envisioned sellers and buyers. From sellers' perspectives, this could mean providing access to many thousands of circuits driven by diverse hardware across a broad geographic region. From buyers' perspectives, this means having access to potentially thousands of end-to-

end paths that are available in many/most colocation facilities in a broad geographic region. Further, the fiber exchange must scale to meet diverse demands of buyers in a timely fashion.

High availability. Service and content providers typically seek to guarantee five-nines availability to their customers [224] (*i.e.*, available 99.999% of the time). Likewise, the GreyFiber system must be highly available in order to function as a flexible provider of wide-area connectivity. Additionally, the resources (*i.e.*, endpoints and links) provisioned by the system should also enable/support five-nines availability. Two positive consequences of such a highly-available system are that failures can be treated as a normal situation to be handled [259], and that a high level of service can be guaranteed through service level agreements (SLA), with low risk to the provider.

Rapid provisioning. Hardware resources must be able to be provisioned over short timescales (ideally on the order of millisecond or submillisecond). This capability enables GreyFiber paths to be available over very short timescales (*e.g.*, in response to workload bursts) and to put paths into service quickly when needed by a customer to recover from an unexpected failure. Naturally, for service providers, a fast infrastructure provisioning capability simplifies the process of activating backup resources during network maintenance or outage events. Rapid provisioning also implies the need for a system that is easy-to-use after it has been initially configured.

Flexible access. Current dark fiber leases (based on 20+ year IRUs) and IP transit commitments (typically 3-5 years) inherently limit access to connectivity. To overcome this impediment GreyFiber requires access to infrastructure over a wide range of timescales (sub-second to years). This enables many opportunities for buyers and sellers including economic benefits, reselling unused resources, and ease of expansion at diverse geographical regions.

7.3.2 GreyFiber Overview

GreyFiber is a three-tiered system whose goal is to provide wide area connectivity as a service over a range of timescales. GreyFiber consists of three components: *(1)* Global Control, *(2)* Local Site Control, and *(3)* Physical Infrastructure. The overall architecture of GreyFiber is depicted in Figure 7.2, which is inspired by the hybrid control proposed by Mukerjee *et al.* [267].

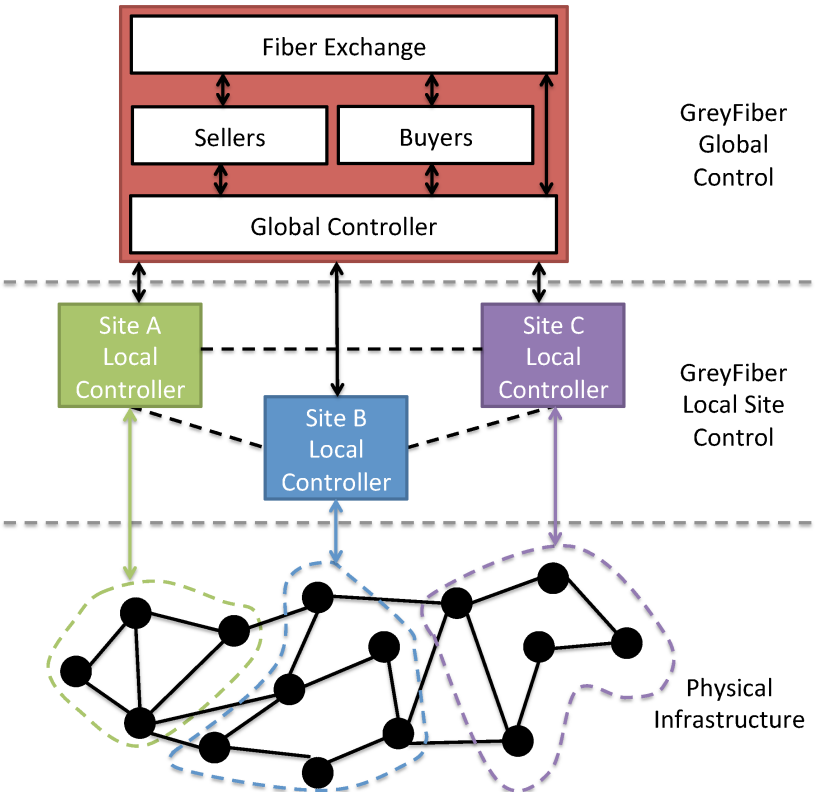

Figure 7.2: GreyFiber Architecture Overview.

GreyFiber Global Control. The highest level of the system is the GreyFiber Global Control (GGC), which serves as a command center for the entire system by providing a common interface for all the entities involved. To meet scalability, extensibility and availability requirements, the GGC

resides either in the cloud or in a datacenter and consists of the following four sub-components/entities:

- *Fiber Exchange.* An auction management system (explained in §7.3.4) that is similar to an ad auction [69, 112, 172] or cloud resource auction system [314]. This subcomponent can either be co-located with the Global Controller or can reside in a different location (*e.g.*, the cloud).
- *Buyers.* The entities (*e.g.*, ISPs, CDNs, enterprise networks, etc.) or the customers of GreyFiber who specify their connectivity needs—also known as *resource requests*—including geography, performance, timescales, deadlines (if any) and bandwidth requirements, along with their bids are called *Buyers*. Support for these options allows planning over longer time scales where buyers can manage costs (*i.e.*, leasing vs. digging new conduits) or over short time scales when there is a specific need (*i.e.*, during specific Internet events like high-traffic streaming events, planned outages due to maintenance, etc.).
- *Sellers.* The entities (*e.g.*, service, cable and fiber providers) who own or have the ability to provide a link or set of links to the GreyFiber ecosystem are called *Sellers*. To support evolution in the physical Internet and to enable a GreyFiber-based connectivity service, an entity has to meet the following constraints: *(i)* provide access to all (layer 1) hardware such as endpoints and links, *(ii)* provide access to the routing substrate in order to direct packet traffic to the lit fiber, and *(iii)* support for a wrapper API to get circuit provision/tear down decisions from the auction-based decision process. We call these three constraints *seller requirements*. Similar to any market with competing entities, we hypothesize that different sellers compete based on factors including fiber costs, geographical diversity and robustness of their paths, and simplicity in establishing/tearing down connectivity.
- *Global Control.* A centralized controller (similar to an SDN controller) that has a global view of all site controllers, also known as GLSCs (ex-

plained below), at different geographic locations. Various applications including traffic engineering, time-based circuit provisioning, network management and backup restoration are implemented within this entity.

In the GreyFiber system, access to connectivity is based on winning auctions for available resources (*e.g.*, either via generalized second price auction (GSP) [208] or Vickrey-Clarke-Groves (VCG) [196, 225, 298] auctions). Winning a bid results in a configuration that is pushed instantaneously across the sites for a specified customer. Circuit creation is similar to the flow installation using a circuit pusher [23] application in FloodLight. A wide variety of time-based circuit provisioning capabilities, as explained in §7.3.3, are also supported in GreyFiber.

GreyFiber Local Site Control. Below the GGC is the GreyFiber Local Site Control (GLSC) which mimics minimal functionalities from the GGC in a local context (*e.g.*, local decisions on failures, provisioning next available resource in case of failure, etc.) and provides local control for individual sites at marked geographic locations. With the rise of Internet Exchange Points (IXP) researchers have observed a "flattening" of the peering structure in the Internet [201, 220, 254], affecting the structure of end-to-end paths; these facilities are natural locations for GLSCs. Accordingly, we assume that GLSCs are available in every colocation facility. A GLSC has the following capabilities:

- *register* with the Fiber Exchange where the registration includes information about the set of links, capacity required, geographic reach and the potential buyers that are directly connected to a particular GLSC;
- *configure* links, that is, when a buyer wins an auction, connectivity is established for the specified period of time over the specified link(s) and then to tear down these connections when the time expires;
- *report status* information to the exchange since the link may not always be available or buyers might be interested in real time status, especially on links that are used by multiple buyers;

- *control* a set of physical infrastructure (explained below) during connection setup and tear down; and
- *monitor* links connected to them and maintain different performance indicators like packet loss, latency, and connection stability.

In the future, we envision replacing these GLSC units with either SDN-enabled IXPs or simply Software-Defined Exchanges (SDX) [229], where inside an SDX, the route servers are local SDN controllers and SDN-enabled switches where multiple ASes participate, connect and exchange traffic.

Physical Infrastructure. The final layer in the GreyFiber ecosystem is the Physical Internet Infrastructure, which is composed of traditional nodes (fiber connection points) and links (fiber paths) (from §3 and §5). The physical Internet layer encompasses both long-haul and metro fibers, which provide intra- and inter-GLSC connectivity. Although we conceive of this layer as *physical* infrastructure, any network substrate for which the required GLSC functions can be implemented can fulfill this role, *e.g.*, overlay or virtual network topologies created using Planetlab [271], Mininet [293], or GENI [161]. Moreover, infrastructure from *any* real service provider that meets the seller requirements (above) can be seamlessly connected to GreyFiber.

It is important to note that there are many *technical* and *engineering* challenges that must be overcome at the physical layer to realize rapid connection setup/teardown. Technical issues include signaling across various endpoints, hardware limits such as transmission power, and fiber-specific challenges such as attenuation and chromatic and polarization mode dispersion [113]. In this paper, we assume that these factors are already addressed and that the Sellers expose the configurable wavelengths of fiber strands (as part of Seller requirements) to GGC to ensure that the wavelengths are unique for each created circuit. Moreover, since the signal-to-noise ratio of other wavelengths is affected when a new wave-

length is added dynamically, the optical power needs adjustment every time a new circuit is added. We plan to consider such power adjustments in future work. In addition, our future efforts will investigate the efficacy of CDC ROADM-based wavelength reconfigurability ([113, slide 39]) in GreyFiber.

Some of the engineering challenges include determining locations for infrastructure build outs, deploying endpoint-specific capabilities (*e.g.*, amplifier, multiplexer, signal regeneration equipment, etc.), patching endpoints to fiber strands, and electricity needed to power the deployments. Since the speed at which the GreyFiber system can put new paths into service is dependent on many factors, including the engineering challenges mentioned above, our requirement is that they not add any significant overhead to the provisioning times of the underlying paths and/or links under its control. Furthermore, we assume that these factors are taken care of at the Sellers' end before using GreyFiber. That is, *the fiber path is already lit between endpoints and every seller controls, manages and maintains their own portion of the physical infrastructure.*

7.3.3 Supported Circuit Provision Scenarios

To overcome the inflexibilities in standard infrastructure leasing (§7.3.1) and to address the need for quick, dynamic and on-demand network parallelization and/or circuit provisioning, the GGC in GreyFiber supports a wide taxonomy of time-based provisioning scenarios. At the highest level, the provisioning module that implements the time-based provisioning logic classifies the *resource requests* from buyers into either a *realtime* or a *non-realtime* request. Once the immediateness of a given request is identified by the provisioning logic, it is further sub-classified based on *(i)* timescales during which the path is needed, *(ii)* backup requirements, and *(iii)* scalability/performance constraints.

GreyFiber supports a variety of circuit provisioning scenarios at vary-

ing timescales including *small* (from seconds to minutes), *medium* (hours), *large* (from days to months) and *extra-large* (years similar to a standard fiber lease or IRU). In addition, circuits could be dynamically provisioned to serve as backups during (or quickly after) either an outage event or a scheduled maintenance operation. Furthermore, in order to meet performance constraints in the SLA at peak times, links could be elastically spun up and/or down using GreyFiber.

7.3.4 Auction Model

To enable flexibility in infrastructure pricing, the GGC—in particular, the Fiber Exchange subcomponent—uses an auction model to lease seller's infrastructure to interested buyers/customers. Making GreyFiber resources available via auction recognizes that the value in wide area connectivity as a service is in the excess capacity available over a variety of time scales (similar to the motivation for spot markets in cloud infrastructures). Should customers wish a longer term IRU, traditional dark fiber and IRU-based leasing model are assumed to be available.

Fiber Exchange offers auctions from a list of k links[4]. Specifically, there is list L of k links, where L is defined as,

$$L = \{l_1, l_2, l_3, \ldots, l_k\}$$

There are N ($> k$) customers, each of whom submits one bid per link, a non-negative value b_i, independently and simultaneously with other bidders. Note that a customer can bid for multiple links (*e.g.*, l_1, l_2 and l_3) separately and a path (p) can be a composition of either multiple links (say l_1-l_2-l_3) that is laid sequentially in different conduits or three strands of

[4]In this work, our key focus is to enable leasing the fiber/link resources. However, there is nothing limiting in GreyFiber to support leasing of other types of resources (*e.g.*, routers).

fiber laid in parallel within the same conduit. In what follows, we explain the generalized second price (GSP) auction [208], which is the default resource auction model in GreyFiber.

The auction format is GSP with perfect information, and the selection rule is such that the highest k bidders are ranked by their bid values. The payment that the winner makes is the second-highest bid among those submitted by the players who do not win for a particular link. In such a setting, the payoff function, which also denotes the preference of customer i for a link l_i, is given by:

$$u_i = \begin{cases} v_i - \hat{b} & \text{if } b_i \geqslant \hat{b} \text{ and } v_i > v_j \text{ if } b_j = \hat{b} \\ 0 & \text{if } b_i < \hat{b} \end{cases}$$

subject to the following (seller) constraint,

$$v_i \leqslant b_i \quad \forall i = l_1, \ldots, l_k$$

where, each bidder/customer submits a (sealed) bid b_i, and $\hat{b}$ is the highest bid submitted by a customer other than i. v_i is the value that seller attaches to every link l_i to maintain revenue. In short, if the customer obtains a link, they receive a payoff $v_i - b_i$. Otherwise, their payoff is zero. Furthermore, the benefits of GSP including enabling a more user-friendly market that is less prone to gaming by other bidders is shown by Edelman *et al.* [208].

Note that our auction mechanism does not preclude a traditional lease, since a contract could be offered on an exchange with the reserve price set at the standard lease rate. Therefore, GreyFiber is backwards compatible. It is possible for a new entrant to use GreyFiber with short-term leasing option while others use a legacy model with long-term IRU-based leasing. Furthermore, while the idea of applying auction-based methods for leasing a service provider's infrastructure in GreyFiber is new, the auction

mechanisms are well known[5].

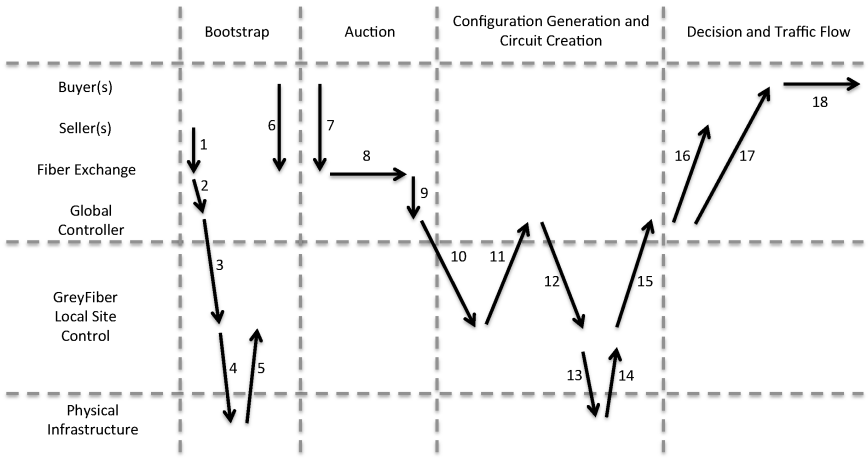

Figure 7.3: Timeline of events to enable end-to-end connectivity in GreyFiber.

7.3.5 End-to-end Events in GreyFiber

Assumptions. To establish an end-to-end circuit between endpoints (A and B) of a customer, we assume that the customer has one (or more) of the following options between their endpoints and a colocation facility that is GreyFiber-enabled: *(1)* metro-fiber or broadband or wireless connectivity (and access) in the last mile (*e.g.*, Verizon's Interconnection services [154]), or *(2)* a dedicated private connection (*e.g.*, Microsoft's ExpressRoute [102]), or *(3)* Fibre to the Premises (FTTP) on Demand [59] with dedicated infrastructure [120]. Furthermore, we assume that the connectivity between customer endpoints and GLSC units are already lit and tested.

Given the aforementioned components and assumptions, we now describe the sequence of events that take place to establish end-to-end connectivity in GreyFiber as depicted in Figure 7.3.

[5]Other forms of auction mechanisms such as GSP with reserve pricing could also be used.

- Every seller registers with the GreyFiber system with information that includes the geography of nodes and links, peering and link properties (*e.g.*, capacity, performance indicators). This information is communicated to the Fiber Exchange and is also advertised to a list of buyers in the ecosystem (step 1). Every buyer must also register with the GreyFiber system prior to entering bids (step 6).
- Once the registration is complete, the GGC forwards the information to the appropriate GLSCs (steps 2 and 3), which monitor the requested set of links for various performance indicators including latency perceived, loss and link utilization (steps 4 and 5).
- If there is a demand from a buyer in the form of resource requests, their bids along with other relevant information are accepted (step 7) and the Fiber Exchange runs an auction to determine the winner (step 8).
- Once a winner for a link or a set of fiber links is determined, the Fiber Exchange communicates the winner information and their corresponding fiber requirements to the Global Controller (step 9).
- At the global controller, creation of a circuit between endpoints of a buyer occurs in two stages. First, the physical topology graph G is queried (step 10). G is composed of fiber circuits from multiple sellers. Each edge in the graph is annotated with maximum and available bandwidth, and total number of fiber strands. Only if the available bandwidth and the number of fiber strands in G are greater than a buyer's resource request does GreyFiber proceed to the second stage (step 11); otherwise, circuit creation is aborted and the buyer is informed that the resource is unavailable.
- The actual establishment of an end-to-end circuit happens in the second stage and is composed of multiple events (steps 12 to 15). The logical end-to-end circuit, with a unique identifier, is stitched from individual one-to-one circuits in G. Buyer requirements are translated into a set of configurations that get pushed into the corresponding GLSCs to

create individual circuits (steps 12 and 13). Now the connections across endpoints are set up for the duration requested by the buyer in her bid (step 14 and 15). Subsequently, available bandwidth and the number of fiber strand counters are updated (step 16).

- The buyer is notified about the decision, along with the connectivity information to access the circuit (step 17). On receipt of this message, end-to-end traffic flow can be initiated by the buyers (step 18). The circuits are continuously monitored by the GLSC to create instant backups in case of failure events.
- Finally, connection tear down simply causes the established circuit to be revoked between the endpoints. When the lease time of buyers end, this process is triggered automatically.

7.4 GreyFiber Implementation and Evaluation

In this section, we describe an implementation of GreyFiber, which was developed to provide insights on feasibility and performance. We also describe results of our evaluation of the implementation in the GENI testbed.

Implementation. The GreyFiber, along with GGC, GLSC, Fiber Exchange, interfaces for buyers and sellers, and monitoring and measurement subcomponents described in §7.3 were all implemented in Python. Our implementation includes broad functionality for each GreyFiber component[6]. This enables all aspects of the GreyFiber event sequence and important aspects of performance to be evaluated.

The GGC is designed to efficiently serve simultaneous requests from multiple buyers in a multi-threaded fashion and has communication interfaces to different entities including buyers, sellers and Fiber Exchange

[6]Commercial GreyFiber deployments will be more scalable and robust, and will reflect details of both business and operational requirements.

(via GGC). Resource requests from buyers are sent via the buyer interface as <Endpoint_A, Endpoint_B, #OfStrandsNeeded, BidAmount, Time, CapacityNeeded, ClientName> tuples in a json format. Next, the physical infrastructure information from the sellers are encoded as topology graphs using the networkx library and are sent via the seller interface. Provisions are available in GreyFiber for both fiber providers and customers to update seller and buyer information respectively. Finally, bid amount and client information extracted from the resource requests are sent to the Fiber Exchange. We note that all the data as well as messages communicated using the aforementioned interfaces are both compressed and encrypted.

Auctions are run at the Fiber Exchange, which implements both GSP- and VCG-based models, and the winner is determined. The winner information from a given auction is communicated to the GGC using the interface specific to Fiber Exchange. The GGC further communicates the winner information to individual GLSC locations. A GLSC, as noted in §7.3, is similar to the GGC albeit with a restricted set of functions and is multi-threaded to improve efficiency. Specifically, it monitors resources using the ping tool, transmits resource information to the GGC through an interface to the Fiber Exchange, and uses infrastructure-specific libraries for creating and pushing configurations to physical infrastructure (as explained below). For our experiments, both GGC and GLSC reside on a Macbook Pro laptop equipped with Intel's i5 processor and 4GB RAM.

Experimental testbed. We demonstrate and evaluate the GreyFiber system through deployment in the Global Environment for Network Innovations (GENI) testbed [161]. GENI enables relatively controlled testing across a homogeneous infrastructure. GENI also offers access to network-based devices that are useful for GreyFiber tests. We also developed a GLSC that interfaces with Mininet as the underlying network substrate. We measured the total time taken to bring a circuit into service using each of these systems and while latencies for setting up GreyFiber-

internal components were consistent between GENI and Mininet, circuit creation times in Mininet were very small (on the order of microseconds). Although the GENI-imposed circuit creation latencies are fairly large (as we discuss in detail below), we use it as the basis for our evaluation due to the feasibility of experiments and the realism inherent in its wide-area reach. In particular, we do not consider Mininet any further in our subsequent experiments. Moreover, while some aspects of GENI are idiosyncratic, the availability of configurable devices along end-to-end paths make it an attractive target for our GreyFiber demonstration.

In our experiments, the resource requests are randomly generated based on the *resource pool* information populated by the sellers in the system. Since we use GENI to evaluate GreyFiber, infrastructure information [63] from the GENI resource center is used to populate the resource pool and bootstrap our system. Similarly, we use GENI's *stitcher* service [66] to create/tear down circuits across GENI endpoints.

Next, in all our experiments, we use a GSP-based auction to elect a winner. If the buyer wins the auction for fiber resource(s), the global controller in GGC issues a new GENI Resource Specification (RSpec) [64]—an XML-formatted configuration file used to reserve network resources in GENI—generation request to GLSC and a new RSpec for circuit creation/revocation is created at the GLSC. This generated configuration file is pushed into the GENI infrastructure *only* if the GLSC monitoring a specific set of resources in a geographic location has determined that those resources are continuously available (stable). In our experiments, we use simple active probes using ping to determine availability, and set the monitoring interval to 1s. We note that our approach of monitoring resources is based on ideas borrowed from prior efforts [7, 176]. Furthermore, the monitoring interval is tunable and can be changed by any entity deploying GreyFiber.

The GLSC assigns an available resource in a particular location to satisfy

a provisioning request. If the requested resource is either unavailable (as determined using the monitors at GLSC locations) or if the request failed due to unavoidable errors (*e.g.*, hardware failure), the next resource at the location is assigned to satisfy the request. For example, if a node in the California is unavailable, GLSC assigns the request from GGC to the next available node in the resource pool in California. We note that network resources across different locations can be dynamically added/removed from the resource pool by the GGC.

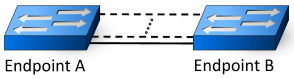

Figure 7.4: Dumbbell topology used for scaling experiments in GreyFiber.

Evaluation methodology. In our evaluation, we start by focusing on the feasibility and scalability of GreyFiber system. Next, to demonstrate the ability of GreyFiber to adapt to network dynamics (*e.g.*, failures), we run our experiments in an end-to-end, wide-area setting. Tests consider both the performance and responsiveness of the system in the presence of background traffic. Specifically, our evaluation is organized around four main questions:

Q1. Can GreyFiber effectively scale if multiple links are required on demand?

Q2. What are the performance overheads in the GreyFiber system?

Q3. How performant and responsive is GreyFiber during network outage(s)?

Q4. How does the performance of GreyFiber for provisioning an alternate path in reaction to a failure compare with rerouting overheads, *e.g.*, using OSPF?

7.4.1 Scalability of GreyFiber

To assess the scalability of GreyFiber, we increase the number of links in a simple dumbbell topology depicted in Figure 7.4. In this experiment, the two endpoints (or node pairs) are located at two different geographic locations. We repeated the scaling experiments 5 times with different node pairs that are selected randomly from GENI nodes [63], at different locations and at different times of the day.

Table 7.5: Configuration generation and provision times on scaling the number of links in GreyFiber system.

Number of Links	Configuration Generation (s)	Circuit Provision (s)
1	0.124	19
2	0.116	22
3	0.107	21
4	0.148	25
5	0.126	24
10	0.112	33
20	0.119	35
30	0.120	37
40	0.112	47
50	0.121	54

Table 7.5 shows the averages of time taken (in seconds) to generate configuration files and provisioning of the circuits on increasing the number of links between the endpoints A and B for 5 runs of the scaling experiment. The time taken to generate the configuration is about 120ms on average, independent of the number of links. The time taken to provision circuits from scratch ranges from 19s for one circuit to within a minute (54s) for 50 circuits. We note that these provisioning times are entirely dependent on characteristics of the underlying physical infrastructure (in this case, GENI) which are outside the control of the GreyFiber system. For a differ-

ent infrastructure (*e.g.*, controlled through modern optical transport gear), these circuit provisioning times would likely differ significantly.

While GreyFiber requirements indicate scaling to thousands of circuits, the GENI infrastructure limits our ability to experiment at that scale. Thus, we consider these results as “proof of concept" and intend to continue to investigate scaling in future work. Our expectation is that future cloud-based or distributed versions of the GGC will satisfy the outlined scalability requirements. Apart from improving scalability, such distributed versions of the GGC would also enable the consideration of regional differences between various sellers, buyers, market economies, and geographic considerations. In particular, such distributed GGC units, and in turn the fiber exchanges, could account for regional differences in prices. For example, the north-eastern region may be dictated by the prevailing business needs of customers requiring low-latency paths for financial transactions. Similarly, the west region may be defined by the need for physical diversity of routes across the Rockies.

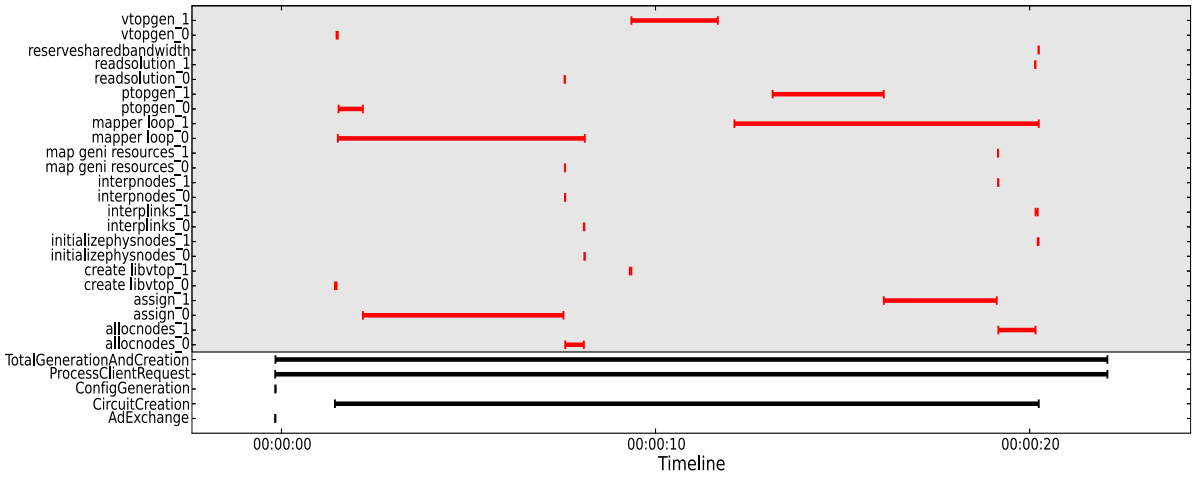

Figure 7.6: Time taken by different components in GreyFiber and GENI. Time Timestamps extracted from the *spew log* file for GENI endpoints A and B are marked with "_0" and "_1" respectively.

7.4.2 Overheads in GreyFiber

We drill down on the time taken by different components in GreyFiber to provision a circuit between two endpoints and quantify the overhead in the GreyFiber system versus the underlying network substrate. Specifically, we measure the time spent to generate the configuration files, provision the actual circuits between node pairs, determine the winners of the auction at Fiber Exchange, and total response time to process a buyer's resource request.

Since our measurement framework is opaque to the underlying network gear in GENI, measuring the time taken by individual components (*e.g.*, hardware, configuration software, etc.) that are used for circuit provisioning/tear down is beyond the control of GreyFiber. This calls for integration of intuitive measurement methods into our system to effectively measure the GreyFiber overhead. To that end, our measurement framework utilizes information from GENI *spew log* files that are emitted during circuit provisioning to quantify the overheads in the underlying network substrate. Specifically, we extract information such as timestamps and debug messages from the log files to tease out the overheads in GENI versus the overheads in GreyFiber.

Figure 7.6 depicts the time taken by different modules as reported by our measurement framework, which is available as part of the GreyFiber system. Timestamps extracted from spew log files that correspond to GENI infrastructure are shown in red and are marked with a grey background. Processing time taken by individual GreyFiber-specific components including Fiber Exchange (177ms), configuration generation (124ms), circuit creation (18.813s) and client requests (22.245s) for provisioning one circuit between endpoints A and B in Figure 7.4 are also shown. Next, we map the circuit creation process into individual GENI-specific functions using the spew log file in the measurement component to account for testbed—in particular, GENI-specific—overheads. GENI-specific func-

tions include *vtopgen* which takes 22ms and 2.310s for endpoints A and B respectively. Next, *create libvtop* accounted for about 40ms for both A and B. We note that the predominant overhead is caused by *mapper loop* function which encompasses other functions including *ptopgen*, *assign* and *interpnodes*, *interplinks*, and *allocnodes*. Endpoint A and B spent 6.602s and 8.134s, respectively, in the mapper loop function. Overall, we observe from Figure 7.6 that the circuit creation process is responsible for the bulk of the total time required, and that the GreyFiber system itself introduces little latency (just over 300ms). Again, we observe that this inherent latency is completely dependent on the underlying network substrate—an observation consistent with anecdotal evidence from a service provider [117].

7.4.3 Performance of GreyFiber

In this experiment, we demonstrate the performance gains—specifically, improvements in throughput—that can be achieved when incrementally adding physical capacity using GreyFiber. For this analysis, we reused the dumbbell topology from earlier experiments, adding an an iPerf server and client at each end point. Next, we bootstrapped the experiment with five hosts on either side of the bottleneck link, creating five different TCP flows.

To show the performance benefits of GreyFiber, we scale the number of links between the dumbbell topology endpoints by dynamically provisioning a new circuit every 30s. Figure 7.7-(left) shows the improvements in performance on scaling the number of links. At the start of the experiment, *i.e.*, during the initial 30s, all five flows contended heavily for the bottleneck link and the average effective throughput, as observed from H1, is ~4Mbps. Upon provisioning two additional links at 30s and 60s, the throughput increases to ~8Mbps and ~12Mbps respectively. On further addition of a link at 90s, an average throughput of ~16Mbps is achieved. Finally, on yet another addition of a link at 120s (leading to a total of 5

links between the dumbbell endpoints), an average effective throughput of ~20Mbps is achieved by all the five competing flows.

We repeated the experiment (above) on CloudLab [27] using the same GENI RSpec, by changing the capacity to 10Gbps. The results are depicted in Figure 7.7-(right). Similar to Figure 7.7-(left), all five flows contended heavily for the bottleneck link initially and throughput across is ~1.7Gbps. At 30s and 60s two additional links were provisioned, which increased the throughput to ~3.7Gbps and ~5.3Gbps respectively. On further addition of a link at 90s, an average throughput of ~7.6Gbps is achieved. Lastly, an average effective throughput of ~9.55Gbps is achieved by all the five competing flows on provisioning the fifth link at 120s. From this result, we make two key observations: *(1)* GreyFiber scales effectively on links with larger bandwidths *without* any performance degradation and *(2)* GreyFiber is generic and adaptable to different networking substrates. These results, apart from showing the efficacy of GreyFiber, also demonstrate the kinds of performance gains that could be achieved using GreyFiber.

7.4.4 Effectiveness in the Face of Outages

Finally, we show how GreyFiber could be effectively used to provide backup physical connectivity during network maintenance and/or outage events. We start with one link between node pairs and run an iPerf server at A and an iPerf client at B (in Figure 7.4) for 90 seconds. The first 30s is the warmup phase to account for TCP artifacts like congestion control. Next, we manually introduce a *link failure event* at the 60th second between A and B by using the tc (traffic control) command on interface A and disrupt connectivity in different ways. For each experiment, we measure and show the throughput (in bytes per second).

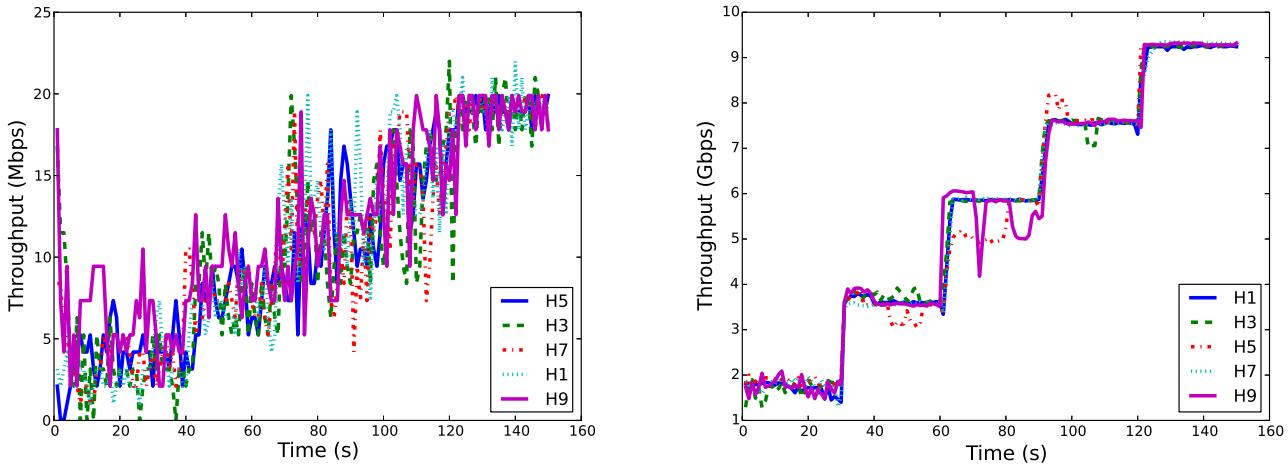

Figure 7.7: Performance improvements achieved using GreyFiber on GENI (left) and CloudLab (right) testbeds.

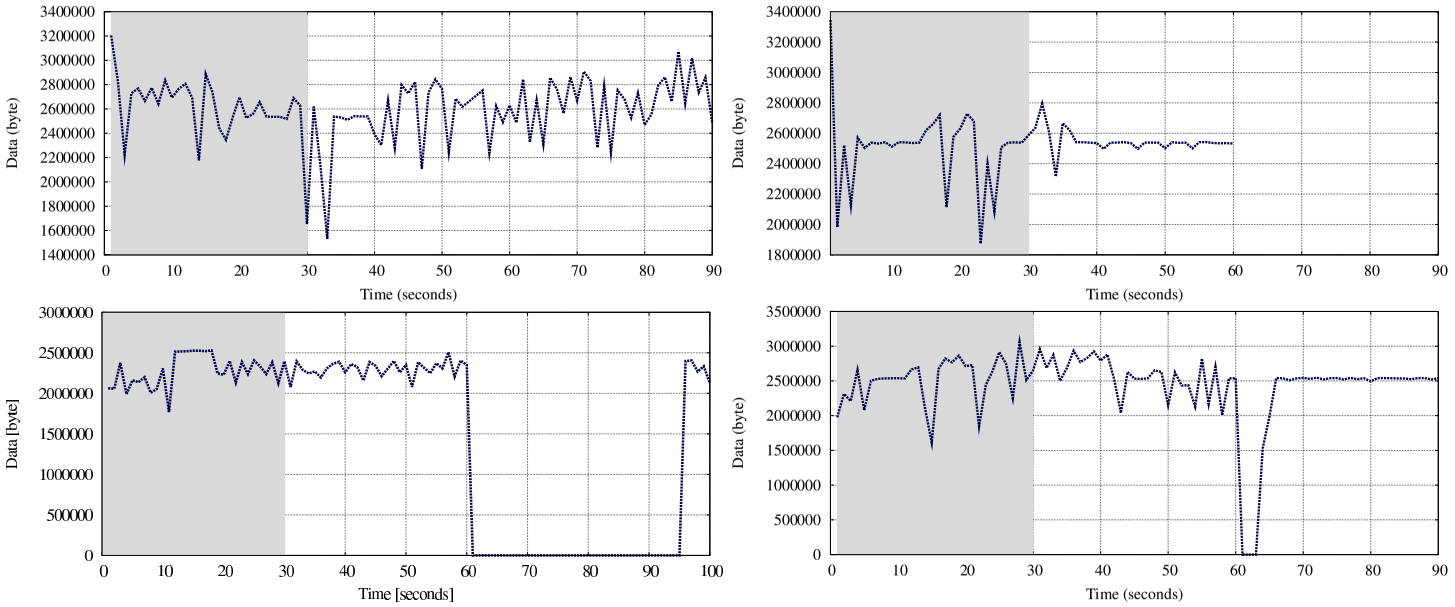

Figure 7.8: Throughput (bytes per second) results from dynamic outage detection and recovery experiments. Warmup phase of each experiment is shown with grey background. Plots shown for no failures (top left), with failures but no backups (top right), with failures and backup using OSPF (bottom left) and GreyFiber (bottom right).

Scenario 1: No failures. We begin our evaluation by showing the best case scenario where there is no failure event introduced between endpoints A and B. The top-left plot of Figure 7.8 shows throughput as observed from endpoint A, respectively. In this scenario, a total data of 1.88Gb is transferred from A and B and the throughput observed is 20.78 Mbps.

Scenario 2: No backup solution. Next, we show the effect of a link failure event *without* any instantaneous and reactive backup solution in this scenario. This is the worst case scenario. The top-right plot of Figure 7.8 depicts the throughput for this situation The connection between the endpoints stalled at the 60th second. Furthermore, the total data transferred dropped to 1.23Gb, with an average throughput of 20.19Mbps from endpoint A, up to the time of failure.

Scenario 3: Using OSPF-based backup. Third, we evaluate an OSPF-based backup solution to reroute traffic during the link failure event. In this scenario, hello and dead intervals are set to 10s and 40s respectively, which are off-the-shelf default values for OSPF. In this scenario, we used one additional link as a backup to reroute traffic. Also, the experiments were run for 100s to illustrate OSPF's recovery.

The bottom-left plot of Figure 7.8 shows the throughput using OSPF routing to reroute traffic. We observe a lag of 36s to re-establish connectivity using OSPF-based backup[7] with a total data transfer of 1.26Gb at 13.04Mbps. These results are as bad as the no-backup scenario 2.

For the experiment in this scenario, as noted above, we use the default values for time intervals. These values are not proscriptive but are used by service providers in traditional OSPF settings. An alternative way is to reduce the timer hello and dead timer values. However, anecdotal evidence shows that the configurations generated from reduced timer values can be sub-optimal and can result in route flaps [123]. In addition, since we use *quagga*-based routers [121] at endpoints A and B, to the best of

[7]During the time at which the measurement was taken, the wait time interval was 4s. Hence an OSPF-backup was initiated at the $dead - wait^{th}$ second.

our knowledge, there are no known implementations for mechanism like *fast reroute* [157] and *fast hello* [114]. We intend to evaluate these solutions against GreyFiber-based backup solution as part of future work.

Scenario 4: Using GreyFiber-based backup. Finally, in this scenario, we outline the efficacy of a GreyFiber-based backup solution. Specifically, we show how the link failure event introduced at 60th second is rapidly detected by the GLSC, which monitors every network provisioned resource associated to it (by default, every second). In short, as soon as the failure event is detected, a new link between A and B is provisioned by the GLSC thereby initiating a backup.

The bottom-right plot of Figure 7.8 shows the throughput as the circuit is provisioned using GreyFiber on detecting a link outage (at around 60s). During this scenario, a total data of 1.76Gb is transferred from A and B at rate of 19.48Mbps.

The GLSC took 1s to detect the link failure event and another 240ms to provision/activate a link in the existing *shared vlan* [65] configured through the GENI infrastructure, and reroute flows via the newly created path. This results in a 28x faster recovery than the OSPF-based scenario. Since, for this experiment, we used a TEQL-based load-sharing technique [98] while provisioning circuits between A and B, links are effectively aggregated and backup creation is rapid. While the latency of activating the backup link (240ms) is GENI-infrastructure-specific, it is similar to switching times found in published specifications from commercial optical networking gear, *e.g.*, [83].

While the monitoring interval employed by a GLSC is tunable and the physical infrastructure imposes unavoidable latency in the provisioning process, our results illustrate how GreyFiber could be used to quickly recover from network outages with minimal impact on user traffic. For example, a video streaming application with modest buffering would not perceive any glitch, and for chat, interactive shell, and other realtime

applications, the impact would be short-lived. Lastly, for web traffic, the waiting time to lose a user has been observed to be ~4s [124]. Even with a more stringent two-second rule for webpage load times [81], the GreyFiber system can sufficiently provision a backup path.

7.5 Summary

Our work is motivated by the fact that market forces and technology trends have evolved to the point where alternatives to the decades-old methods for gaining access to physical network infrastructure (dark vs. lit fiber) are now feasible. In this chapter, we describe GreyFiber, which is designed to enable wide area connectivity as a service, similar to the way that cloud computing has enabled computation-based service offerings that have had a transformative impact. The objective of GreyFiber is to offer flexible access to fiber-optic paths between end points (*e.g.*, colocation facilities) over a range of timescales, and through a Fiber Exchange, which makes this connectivity available to the highest bidders. We describe details of the design of a GreyFiber system and deploy a fully functional instance of it in the GENI testbed for demonstration and assessment. Our experiments consider scalability, performance overhead, and responsiveness to outages. Our results show that the system generates configurations in less than 150ms and that circuit provisioning scales roughly linearly in the number of links. Detailed examination shows that provisioning overheads are tightly coupled with the GENI infrastructure. Finally, our results show how path performance and reaction time to outages (vs. OSPF) can be improved with GreyFiber.

8

Conclusion and Future Work

Standard mechanisms to ensure robustness in the Internet require ongoing consideration. Toward enhancing the robustness in the Internet, we posit that two requirements are paramount: *(1)* understanding physical Internet's infrastructural complexities/risks and incorporating them in the solutions developed; and *(2)* improving mechanisms in the evolving Internet ecosystem. While the first requirement is looking backward at what is *missing* in this space so far, the second requirement is looking forward at what is *needed* to ensure robust operation, given the Internet's continuous evolution.

To address the first requirement, my dissertation research has two themes: *(i)* understanding the structural complexity and risks in the Internet and *(ii)* building scalable, robust and easy-to-deploy systems to enhance network robustness by mitigating outages. The goal is to design and evolve each of these two themes independently while being aware of the other, so that they function synergistically towards the tackling the first requirement. In this closing chapter, we first summarize our key findings, contributions and solutions (§8.1). We then present directions for future research in §8.2, where we outline ideas to address the second requirement of improving mechanisms to further enhance Internet's robustness in the face of continuous evolution. We conclude in §8.3.

8.1 Summary of Contributions

8.1.1 Unravelling Internet's Structure and Risks

We start by investigating how a comprehensive understanding of structural complexity and inherent risk are critical for robust Internet design and operation, where we take a bottom-up approach and build Internet Atlas, which is a comprehensive repository of the physical Internet. The main goal of Internet Atlas is geographically accurate representation of the Internet's physical interconnection infrastructure map. We assemble this physical representation of the Internet using search to identify maps and other repositories of the locations of buildings that house networking equipment and the conduits that connect them. We have developed a substantial collection of scripted tools that automate many aspects of identifying, verifying, collecting and transcribing maps of physical Internet infrastructure into the Atlas database repository.

The Atlas data repository is available through an openly available web portal. Internet Atlas is based on the widely-used ArcGIS geographic information system, which includes a spatial database that enables flexible aggregation and visualization of diverse, geo-coded data sets. ArcGIS also includes a large set of built-in tools that enable a broad set of spatial and statistical analyses. To further increase the utility of Atlas, we extend the repository with relevant, related real-time feeds (*e.g.,* BGP updates, Twitter feeds, and weather data) and static data (*e.g.,* DSHIELD logs, road/rail infrastructure).

We develop a new probing heuristic to broadly identify Internet infrastructure that has a fixed geographic location such as Point-of-presences (POPs), Internet exchange points (IXPs), datacenters and other kinds of hosting facilities. The starting point of our study is to understand how physical and network-layer maps differ. To that end, we compare large repositories of physical and network maps and find that physical maps

typically reveal a much larger number of nodes (*e.g.,* POPs and hosting infrastructure). For the selected networks, we find that: *(i)* the physical maps typically show many more nodes/links than the network-layer maps, *(ii)* there is often a high amount of overlap in nodes/links that appear in both data sets, and *(iii)* network-layer maps sometimes include some nodes/links that are not in physical maps due to incomplete or out-of-date published topologies.

These results motivate the development of probing techniques for targeting the identification of nodes with known or suspected physical locations. We develop a layer 1-informed heuristic algorithm for probe source-destination selection called POPsicle that identifies 2.4 times as many nodes as standard probing methods. Finally, we identify the fact that sources co-located as IXPs can be used to amplify POPsicle-based probing since an IXP-based vantage point can be considered to reside within all of the service providers that peer at the IXP. To that end, we deployed POPsicle at a real IXP and found that it finds almost all POPs compared to Atlas, and additional POPs compared with Ark.

Next, we study the Internet's long-haul fiber-optic infrastructure in the US. Specifically, we build a first-of-its-kind map of long-haul infrastructure using openly available maps from tier-1 ISPs and cable providers. We validate the map rigorously by appealing to public information sources such as government agency filings, environmental impact statements, press releases, and others. Examination of the map confirms the close correspondence of fiber deployments and road/rail infrastructure and reveals significant link sharing among providers, which in turn aggravates the risks inherent on topology.

We apply different metrics to examine the issue of shared risk in the long-haul map. Our results point to high-risk links where there are significant levels of sharing among service providers. Such infrastructure sharing is the result of a common practice among many of the existing Internet

service providers (ISPs) to deploy their fiber in jointly-used and previously installed conduits and is dictated by simple economics—substantial cost savings, among other objectives, as compared to deploying fiber in newly constructed conduits.

Finally, we identify public ROWs that could be targets for new link conduits that would reduce shared risk and improve path performance. We discuss implications of our findings in general and point out how they expand the current discussion on how Title II and net neutrality.

8.1.2 Systems for Robust Internet

Given the understanding of structural complexity, we next develop systems that take advantage of emerging technology to satisfy ISP objectives and to minimize shared risks. First, we create a decision support framework, called Deployment as a Service (DaaS), that uses geo-based multi-objective optimization to identify target areas with the highest concentration of un/underserved users at the the lowest cost to service providers for network infrastructure deployment. Second, we propose and build a system called GreyFiber, which provides a means to offer easy and cost-effective access to unused fiber-optic paths between participating endpoints on demand based on market economics, for arbitrary durations, and possibly with industry-specific performance guarantees.

DaaS. We consider the problem of identifying target areas for network infrastructure deployment in un/underserved areas. Our techno-economic approach applies geo-based multi-objective optimization to find the areas with the highest concentration of un/underserved users at the the lowest cost to service providers. We demonstrate the efficacy of our methodology by considering physical infrastructure and demographic data for US counties along with deployment cost models that include upgrading existing infrastructure and deploying new infrastructure. While we do not argue that the quantitative aspects of our cost models are rep-

resentative of any specific service provider, our results identify a list of counties that would be attractive targets for broadband deployment and that correspond closely with those already identified for future deployments in the US.

GreyFiber. Our GreyFiber work is motivated by the fact that market forces and technology trends have evolved to the point where alternatives to the decades-old methods for gaining access to physical network infrastructure (dark vs. lit fiber) are now feasible. We describe GreyFiber, which is designed to enable wide area connectivity as a service, similar to the way that cloud computing has enabled computation-based service offerings that have had a transformative impact. The objective of GreyFiber is to offer flexible access to fiber-optic paths between end points (*e.g.*, colocation facilities) over a range of timescales, and through a Fiber Exchange, which makes this connectivity available to the highest bidders. We describe details of the design of a GreyFiber system and deploy a fully functional instance of it in the GENI testbed for demonstration and assessment. Our experiments consider scalability, performance overhead, and responsiveness to outages. Our results show that the system generates configurations in less than 150ms and that circuit provisioning scales roughly linearly in the number of links. Detailed examination shows that provisioning overheads are tightly coupled with the GENI infrastructure. Finally, our results show how path performance and reaction time to outages (vs. OSPF) can be improved with GreyFiber.

8.2 Future Work

8.2.1 Bridging the Robustness Gap

In future work, we will continue to expand the Internet Atlas (in chapter 3) by adding additional maps as they are discovered in on-going search

campaigns. We will also begin further expansion and diversification of the repository *e.g.,* by adding real time measurement capability.

We plan to use POPsicle, explained in Chapter 4, to more broadly confirm and map network-layer nodes by exploiting additional available IXP-based VPs and by deploying it to new IXPs. We also intend to examine potential efficiency gains in POPsicle's algorithm by more aggressively pruning the search space of destination VPs. Future efforts will include benchmarking versus simple probing methods and more targeted approaches (like iPlane [263]) that will enable us to reason about and quantify the efficiency and effectiveness of the tool in a broader deployment. We are considering how to fully automate and integrate POPsicle with the Internet Atlas in order to accurately and quickly assemble multi-layer maps of network service providers.

In addition, we intend to appeal to regional and metro fiber maps to improve the coverage of the long-haul map and to continue the process of link validation in Chapter 5. We also plan to generate annotated versions of our map, focusing in particular on traffic and propagation delay.

We will also consider how to enrich our model discussed in Chapter 6 to provide details that can catalyze infrastructure deployments on multiple geographic levels. This would accommodate underserved users in areas that may not otherwise be overlooked. More broadly, we believe that our framework can be applied in areas beyond the US that have limited or different types of data that could provide insights on deployment opportunities. We also argue that while our framework is currently focused on un/underserved areas, it could also be used to consider other business needs of service providers including identifying new market opportunities.

While our results in Chapter 7 demonstrate the efficacy of our Grey-Fiber design, there is much to be done in future work to develop the core concepts into reliable, high performance systems that deliver wide area

connectivity as a service. In on-going work we will develop partnerships with service and equipment providers toward the goal of deploying GreyFiber in a live environment. One of the key aspects of this work is to push functionality as close to the physical layer as possible in order to reduce provisioning latency. At the same time, we plan to address scaling and distributing the GGC. We also plan to expand our cost, pricing and deployment analyses in order to assess the feasibility of wide area connectivity as a service in a range of markets.

8.2.2 Improving Internet Mechanisms

To tackle the second requirement, this dissertation opens up a new problem that we call “Internet Measurement Data Science" (IMDS) based on the following key insight. The scale and diversity of the Internet measurement data continues to grow and has the opportunity to take advantage of the algorithms, systems and methods from the “Big Data" community. For example, a distributed system to secure ISP deployments has the ability to generate vast amounts of measurement data on a daily basis. However, there are no suitable tools to organize and extract meaningful insights from the generated data. At the core of IMDS is the ability to employ techniques/theories drawn from broad areas of Big Data and seek feedback from network operators and service providers to create new insights that are often overlooked by state-of-the-art measurement and inference techniques. We believe IMDS to be a promising direction for interesting and impactful research in the Internet. To realize IMDS, four key components are paramount:

- **Rethinking Internet Measurements.** In the era of IPv6 and IoT, along with the revenue model introduced by ad-driven economies, measuring many billions of next-generation devices (*e.g.*, mobile phones, smart sensors) and users (*e.g.*, retention rate, click) in the Internet ecosystem will be fraught with challenges. At a first glance, it might appear that

many of the existing measurement techniques could be directly applied for studying the entities in the Internet. However, we argue that such an application is infeasible because of the size of address space, network management policies, bandwidth, storage and power requirements, deployment characteristics, and heterogeneity of devices. For example, while scanning the entire IPv4 address space could be performed in less than five minutes, such techniques are not directly applicable for an Internet with IPv6 addressing and billions of IoT devices connected. In short, we are at a crossroads for creating new measurement techniques and/or repurposing the existing measurement techniques to effectively measure the evolving Internet ecosystem at scale.

- **Internet Data Management.** Faced with the challenge of measuring all entities is the fact that the measurement efforts will produce voluminous amounts of data with differing format and processing requirements. This calls for a new data management framework which considers: (i) analysis objective of a measurement study (in terms of when an analysis is required), (ii) computational complexity of the analysis method used, (iii) treatment of the actual data (what is collected and quantity), (iv) network topology, (v) streaming vs. storage requirements, and (vi) resource constraints (*e.g.*, power, bandwidth, etc.). The framework would allow one to reason about what kinds of analysis and transformations need to be considered at the point of measurement, what data needs to be moved and when, and how can different datasets compare to, contrast with and inform each other. Borrowing concepts from other areas (*e.g.*, Compressed Sensing), we envision building a generic data-driven framework to collect optimal amount of data across various entities and curate them efficiently, with a focus on databases and systems research.
- **Large-scale Internet Analytics.** Extracting meaningful information from the collected and curated data calls for rethinking the inference and visualization techniques, algorithms and systems used. Specifically,

we plan to develop new algorithms and deep learning-based systems to develop insights from the data. We are also interested in the problem of big-data visualization where the goal is to create a decision support framework and present such large-scale data to researchers, network operators and service providers in an efficient, lucid and actionable way that they can quickly transform into network practices.

8.3 Closing Remarks

The high level objective of this thesis is to move closer to the goal of having comprehensive and accurate maps of the Internet's topology that can be applied to a wide range of problems. In particular, we apply the maps to improve Internet's robustness by taking a bottom-up approach: that is, by understanding the structural complexity and risks inherent on physical infrastructure, we build topology-aware systems to close the robustness gap. Moving forward, keeping up with the technological evolution in the Internet ecosystem will be key to its continued robust operation. This calls for improving and rethinking various measurement, inference and data management mechanisms/techniques used in the Internet. This thesis takes important steps in this direction by creating a better understanding of the structure, risks, and systems needed to address the former; while simultaneously proposing new future research directions in the space for the latter.

Bibliography

[1] https://www.codot.gov/projects/us36eis/documents/us-36-final-eis-volume-i/section-4-18_utilities.pdf.

[2] 2014 Provider Dataset from broadbandmap.gov. http://www.broadbandmap.gov/data-download.

[3] 2016 Broadband Progress Report by FCC. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2016-broadband-progress-report.

[4] 50-State survey of rights-of-way statutes. http://www.ntia.doc.gov/legacy/ntiahome/staterow/rowtable.pdf.

[5] A Dissertation So Good It Might Be Classified. http://archive.wired.com/wired/archive/12.01/start.html?pg=10.

[6] A Means for Expressing Location Information in the Domain Name System. http://tools.ietf.org/rfc/rfc1876.txt.

[7] Akamai's SureRoute. https://developer.akamai.com/stuff/Optimization/SureRoute.html.

[8] Allied fiber. http://www.alliedfiber.com.

[9] Allied Fiber: Long Haul Dark Fiber. http://www.alliedfiber.com/products/long-haul-dark-fiber/.

[10] Allied Fiber's units file bankruptcy protection. http://www.globaltelecomsbusiness.com/Article/3549199/Allied-Fibers-units-file-bankruptcy-protection.html.

[11] Alternative Network Deployments. Taxonomy, Characterization, Technologies and Architectures. https://datatracker.ietf.org/doc/rfc7962/.

[12] Amazon EC2 Spot Pricing. https://aws.amazon.com/ec2/spot/pricing/.

[13] American Fiber Services - Response1. http://www.ntia.doc.gov/legacy/broadbandgrants/applications/responses/774PNR.pdf.

[14] American Fiber Services - Response2. http://www.ntia.doc.gov/legacy/broadbandgrants/applications/responses/804PNR.pdf.

[15] American Fiber Services - Zayo Release. http://www.zayo.com/images/uploads/resources/Earnings_Releases/FY2011Q1_10-Q.pdf.

[16] AT&T Flash-based Network Map. http://www.corp.att.com/globalnetworking/.

[17] AT&T Network Map. http://www.sura.org/images/programs/att_sura_map_big.gif.

[18] Broadband plan. http://www.broadband.gov/plan/6-infrastructure/.

[19] Calculating fiber optic latency. http://www.m2optics.com/blog/bid/70587/Calculating-Optical-Fiber-Latency.

[20] CapeNet. http://www.muni.ri.net/middletown/documents/technology/RFIresponses/CapeNet.pdf.

[21] CenturyLink Network Map. http://www.centurylink.com/business/asset/network-map/fiber-network-nm090928.pdf.

[22] CenturyLink to Buy Level 3 for $34 Billion in Cash, Stock. https://www.bloomberg.com/news/articles/2016-10-31/centurylink-agrees-to-buy-level-3-for-34-billion-in-cash-stock.

[23] Circuit Pusher. FloodLight Controller. https://github.com/wallnerryan/floodlight/blob/master/apps/qos/circuitpusher.py.

[24] City of Boulder, CO. http://www.branfiber.net/Conduit%20Lease%20Agreement%20between%20Zayo%20and%20the%20City%20of%20Boulder.pdf.

[25] City of Santa Clara, CA. http://vantagedatacenters.com/wp-content/uploads/2014/06/Vantage-Network-Connectivity-Report.pdf.

[26] Clear Creek and Gilpin Counties Broadband Assessment. http://apps.fcc.gov/ecfs/document/view?id=7521088690.

[27] Cloudlab. http://www.cloudlab.us/.

[28] Coastal Route - SFO. to LA. for several ISPs. http://www.ustaxcourt.gov/InOpHistoric/anschutz.TCM.WPD.pdf.

[29] Cogent Network Map. www.internetatlas.org.

[30] Colorado Department of Tranportation document. http://www.wmxsystems.com/EndUserFiles/44489.pdf.

[31] Comcast Network Map. http://business.comcast.com/about-us/our-network.

[32] Communications Act of 1934. http://transition.fcc.gov/Reports/1934new.pdf.

[33] Connect America. https://www.fcc.gov/general/connect-america-fund-caf.

[34] Connect America Accepted Areas. https://www.fcc.gov/reports-research/maps/caf-2-accepted-map.

[35] Control and Flexibility with Unlimited Scale Through Lightower Dark Fiber Network. http://www.lightower.com/network-services/dark-fiber-service/#overview.

[36] County document for Zayo and ATT. http://www2.ntia.doc.gov/files/grantees/nt10bix5570098_california_broadband_cooperative_inc_ppr2013_q3.pdf.

[37] Cox and Level3 lease agreement - Ocala, FL. http://ocalafl.iqm2.com/Citizens/FileOpen.aspx?Type=30&ID=2988.

[38] Cox Franchise Agreement in Fairfax county, Virginia. http://www.fairfaxcounty.gov/cable/regulation/franchise/cox/franchise_agreement_cox_fairfax_cty.pdf.

[39] Cox Franchise Agreement in Fairfax county, Virginia. https://www.natoa.org/events/RickEllrod.pdf.

[40] Cox Network Map. http://www.cox.com/wcm/en/business/datasheet/brc-backbone-map-q4-2013.pdf?campcode=brc_un_07_101513.

[41] Dascom Systems Report. http://dascom-systems.com/new/wp-content/uploads/eCare-Overview-of-Nevada-project-for-ICBN-Constituency.pdf.

[42] Design service demonstrating Comcast's presence with other entities along ROW. https://www.cfxway.com/portals/0/procurement_pdf/892ec203-d259-46d7-a6c7-b67b8c01dddd.pdf.

[43] Dig Once ordinance: Muni Networks. http://www.muninetworks.org/tags/tags/dig-once.

[44] Douglas County, CO document for CenturyLink and Level3. http://www.crgov.com/DocumentCenter/Home/View/945.

[45] E-Corridors. http://www.ecorridors.vt.edu/research/papers/stircne/vol02-connecting.pdf.

[46] EAGLENET. https://www.co-eaglenet.net/wp-content/uploads/2013/05/EAGLE-NET-001-Addenda-4-Responses-to-Questions-2013-05-29.pdf.

[47] EarthLink Network Map. http://www.earthlinkbusiness.com/support/network-map.xea.

[48] EdgeConneX. http://www.edgeconnex.com/.

[49] Enron. https://en.wikipedia.org/wiki/Enron.

[50] ESRI ArcGIS. http://www.arcgis.com/features/.

[51] ESRI ArcGIS Tracking Server. http://www.esri.com/software/arcgis/tracking-server.

[52] Facebook Developing Radio Wave Mesh to Connect Offline Areas. https://thestack.com/cloud/2016/02/10/facebook-developing-radio-wave-mesh-to-connect-offline-areas/.

[53] Facebook's Connectivity Lab. https://info.internet.org/en/story/connectivity-lab/.

[54] FCC BroadBand Plan. https://www.fcc.gov/general/national-broadband-plan.

[55] FCC BroadBand Progress. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/eighth-broadband-progress-report.

[56] Federal Highway Administration: State by state status report — utility rights of way. http://www.fhwa.dot.gov/real_estate/right-of-way/utility_rights-of-way/utilsr.cfm.

[57] Fiber Optic Settlement. https://fiberopticsettlements.com.

[58] FiberLocator Online. http://www.fiberlocator.com/product/fiberlocator-online.

[59] Fibre to the Premises (FTTP) on Demand. https://www.openreach.co.uk/orpg/home/products/ultrafastfibreaccess/fttpondemand/fttpod.do.

[60] FlashGot. http://www.flashgot.net/.

[61] From Chicago To New York And Back In 8.5 Milliseconds. http://www.zerohedge.com/news/chicago-new-york-and-back-85-milliseconds.

[62] GAO Report. http://www.gao.gov/assets/670/662711.pdf.

[63] GENI Infrastructure Newcomers Welcome. http://groups.geni.net/geni/wiki/GeniNewcomersWelcome.

[64] GENI Resource Specification. http://groups.geni.net/geni/wiki/GENIExperimenter/RSpecs.

[65] GENI Shared VLAN. http://groups.geni.net/geni/wiki/HowTo/ShareALan.

[66] GENI Stitcher. http://trac.gpolab.bbn.com/gcf/wiki/Stitcher.

[67] Georeferencing in ArcGIS. http://resources.arcgis.com/en/help/main/10.1/index.html#/Fundamentals_of_georeferencing_a_raster_dataset/009t000000mn000000.

[68] GMU Mapping Project. http://gembinski.com/interactive/GMU/research.html.

[69] Google Adwords. https://www.google.com/adwords/.

[70] Google Maps. http://maps.google.com/.

[71] Google to FCC. http://www.engadget.com/2015/01/01/google-letter-fcc-title-ii/.

[72] Google's Project Loon. https://www.google.com/loon/where/.

[73] Hamming Distance. http://en.wikipedia.org/wiki/Hamming_distance.

[74] Haversine formula. http://en.wikipedia.org/wiki/Haversine_formula.

[75] High-Frequency Traders Find Microwaves Suit Their Need for Speed. http://www.bloomberg.com/news/articles/2014-07-24/high-frequency-traders-find-microwaves-suit-their-need-for-speed.

[76] How to Destroy the Internet. http://gizmodo.com/5912383/how-to-destroy-the-internet.

[77] IEEE 802.16 standard. https://standards.ieee.org/about/get/802/802.16.html.

[78] IIX Acquires Leading Interconnection Provider IX Reach. http://www.iix.net/news/iix-acquires-leading-interconnection-provider-ix-reach/.

[79] Image from Geo-tel. http://www.techrepublic.com/article/the-google-fiber-lottery/.

[80] Impact of the 2003 blackouts on Internet communications (Preliminary report), Nov. 2003. http://research.dyn.com/content/uploads/2013/05/Renesys_BlackoutReport.pdf.

[81] Impatient Web Users Flee Slow Loading Sites. http://www.nytimes.com/2012/03/01/technology/impatient-web-users-flee-slow-loading-sites.html?_r=0.

[82] infinera: Transport SDN. https://www.infinera.com/technology/transport-sdn/.

[83] infinera XTM Series Datasheet. https://www.infinera.com/wp-content/uploads/2015/09/infinera-ds-1x8-roadm_50-ghz-optical-networks.pdf.

[84] Integra Telecom Network Map. http://www.integratelecom.com/resources/Assets/long-haul-fiber-network-map-integra.pdf.

[85] Internet2 Network Map. http://www.internet2.edu/media/medialibrary/2013/07/31/Internet2-Network-Infrastructure-Topology.pdf.

[86] Invisible Hand Networks. http://www.invisiblehand.net/.

[87] IRU between Level3 and Comcast. http://investors.level3.com/investor-relations/press-releases/press-release-details/2004/Comcast-Extends-National-Fiber-Infrastructure/default.aspx.

[88] IRU Swaps. http://chogendorn.web.wesleyan.edu/excessive.pdf.

[89] Is Internet backbone vulnerable to cyber attack? http://www.sciencedaily.com/releases/2010/12/101214085539.htm.

[90] IX Reach Changes Name To Console Network Solutions. https://www.console.to/news/ix-reach-changes-name-console-network-solutions/.

[91] IXReach. http://www.ixreach.com/.

[92] Layer42 Map. http://www.layer42.net/network/national.html.

[93] Level 3 and Comcast. http://corporate.comcast.com/news-information/news-feed/comcast-extends-national-fiber-infrastructure.

[94] Level3 Completes Acquisition OF tw Telecom. http://investors.level3.com/investor-relations/press-releases/press-release-details/2014/Level-3-Completes-Acquisition-of-tw-telecom/default.aspx.

[95] Level3 Network Map. http://maps.level3.com/default/.

[96] Lightower Fiber Networks to Merge with Fibertech Networks. http://www.lightower.com/company/news/press-releases/lightower-fiber-networks-to-merge-with-fibertech-networks/#.V92nIz4rKL0.

[97] List of PLC manufacturers. https://en.wikipedia.org/wiki/List_of_PLC_manufacturers.

[98] Load-sharing across multiple interfaces. http://lartc.org/howto/lartc.loadshare.html.

[99] Mapping the Digital Divide. https://www.whitehouse.gov/sites/default/files/wh_digital_divide_issue_brief.pdf.

[100] Media Advertisement. http://www.wikinvest.com/stock/XO_Holdings_Inc_(XOHO)/Cox_Communications_Las_Vegas_Indefeasible_Right_Iru_Agreement.

[101] Media Advertisement. http://www.wikinvest.com/stock/XO_Holdings_Inc_(XOHO)/Level.

[102] Microsoft Azure ExpressRoute. https://azure.microsoft.com/en-us/services/expressroute/.

[103] MMQGIS. Hub-distance capability. http://michaelminn.com/linux/mmqgis/.

[104] Network Operators. Personal Communication, 2016.

[105] Network Time Protocol Project Website. http://www.ntp.org/.

[106] News article from SMW3. http://www.smw3.com/smw3/signin/download/iru/Standard%20Agreement/AG-IRU-Final%20(Mar03).pdf.

[107] News article from Telecom Ramblings. http://www.telecomramblings.com/2014/09/tw-telecom-pushes-north-new-york/.

[108] NSFNet. http://en.wikipedia.org/wiki/National_Science_Foundation_Network.

[109] Oclaro Delivers Industry-First 1x23 WSS Featuring 10X Faster Switching Speeds. http://investor.oclaro.com/releasedetail.cfm?releaseid=652644.

[110] OnlineOCR. http://www.onlineocr.net/.

[111] Open Trench: Muni Networks. http://muninetworks.org/tags/tags/open-trench.

[112] OpenX Ad Exchange. http://openx.com/.

[113] Optical Networking Tutorial. https://www.nanog.org/sites/default/files/2_Steenbergen_Tutorial_New_And_v2.pdf.

[114] OSPF Fast Hello. http://www.cisco.com/c/en/us/td/docs/ios/12_0s/feature/guide/fasthelo.html.

[115] PacketFabric. http://www.packetfabric.com/.

[116] PDF2XL. http://www.cogniview.com.

[117] Personal Communication. Service provider not disclosed to preserve anonymity, 2016.

[118] Pipeline 101. http://www.pipeline101.org/where-are-pipelines-located.

[119] Population Estimates from census.gov. http://www.census.gov/newsroom/press-releases/2014/cb14-tps90.html.

[120] Profitable Network Services via Dedicated Network Infrastructure. https://www.juniper.net/assets/us/en/local/pdf/solutionbriefs/3510539-en.pdf.

[121] Quagga. http://www.nongnu.org/quagga/index.html.

[122] Quake shakes up the net, Dec. 2006. http://www.thestar.com.my/story/?file=%2f2006%2f12%2f28%2fnation%2f16426778&sec=nation.

[123] Reducing timer values in Quagga. http://bird.network.cz/pipermail/bird-users/2012-June/003066.html.

[124] Response Times: Three Important Limits. https://www.nngroup.com/articles/response-times-3-important-limits/.

[125] Reuse of old fiber conduits. https://books.google.co.in/books?id=ynvMx7mMgJAC&pg=PA87&lpg=PA87&dq=reuse+of+old+fiber+conduits&source=bl&ots=_QHokk-JPX&sig=5fIzF6Jc5w6TmsewGS8ITX79VKU&hl=en&sa=X&ei=LTF3VfrDD8Xt8gXd2oGoCg&ved=0CCMQ6AEwAQ#v=onepage&q=reuse%20of%20old%20fiber%20conduits&f=false.

[126] Route Views Project. http://www.routeviews.org/.

[127] Satellite Deployment Costs. http://www.globalcomsatphone.com/hughesnet/satellite/costs.html.

[128] SCAC. http://securities.stanford.edu/filings-documents/1023/GX02/2004322_r01c_02910.pdf.

[129] SidePad. http://www.sidepad.gov/.

[130] Soovle. http://www.soovle.gov/.

[131] Spring Framework. http://www.springsource.org.

[132] Sprint Maps. https://www.sprint.net/network_maps.php.

[133] Sprint US Network Map. https://www.sprint.net/images/network_maps/full/NorthAmerica-Global-IP.png.

[134] Sprint using Level3's fiber in Detroit. http://apps.fcc.gov/ecfs/document/view;jsessionid=yzRnRk4NJnlqDtlg0shlW2QDTSd3J2x0nz9Ryg5TJpMyX0rnYXxG!-1694890999!-477673473?id=6516791680.

[135] Sprint using Level3's fiber in Detroit settlement. https://fiberopticsettlements.com/michigan/LinkClick.aspx?fileticket=912Y0QkAV2E%3d&tabid=62&mid=420.

[136] SuddenLink Network Map. http://www.suddenlinkcarrier.com/index.php?page=our-network.

[137] Team Cymru IP-to-ASN service. http://www.team-cymru.org/Services/ip-to-asn.html.

[138] The Backhoe: A Real Cyberthreat. http://archive.wired.com/science/discoveries/news/2006/01/70040?currentPage=all.

[139] The CAIDA UCSD IPv4 Routed /24 DNS Names Dataset - September 2011–March 2013. http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml.

[140] The Fiber-Optic “Glut" – in a New Light. http://www.bloomberg.com/news/articles/2001-08-30/the-fiber-optic-glut-in-a-new-light.

[141] The Internet Atlas. http://internetatlas.org/.

[142] The Internet Under Crisis Conditions: Learning from September 11, 2003. http://www.nap.edu/catalog/10569/the-internet-under-crisis-conditions-learning-from-september-11.

[143] The IPv4 Routed /24 AS Links Dataset: September 2011–March 2013. http://www.caida.org/data/active/ipv4_routed_topology_aslinks_dataset.xml.

[144] The Li-fi technology. http://www.ted.com/talks/harald_haas_a_breakthrough_new_kind_of_wireless_internet.

[145] The National Atlas. http://nationalatlas.gov/.

[146] The PeeringDB. https://www.peeringdb.com/.

[147] UEN. http://home.chpc.utah.edu/~corbato/montana/montana-friends-25-jul-2013.pdf.

[148] Urbana city council project update. http://media01.atlas.uiuc.edu/gslis/gslis-v-2010-3/Digital_Divide_smeltzer_slidesandnotes.pdf.

[149] Urbana city council project update - pictures. http://urbanaillinois.us/sites/default/files/attachments/uc2b-urbana-council-update-10-25-10s.pdf.

[150] Urbana city council project update - pictures. http://www.tellusventure.com/downloads/casf/feb_2013_round/casf_project_bright_fiber_1feb2013.pdf.

[151] U.S. Census Bureau.

[152] Verizon Network Map. https://www22.verizon.com/wholesale/images/networkMap.png.

[153] Verizon to Buy XO’s Fiber Unit From Carl Icahn for $1.8 Billion. http://www.bloomberg.com/news/articles/2016-02-22/verizon-to-acquire-xo-communications-fiber-unit-for-1-8-billion.

[154] Verizon's Data Center Interconnection Services. http://www.verizonenterprise.com/external/service_guide/reg/cp_dcis_plus_data_center_interconnection_services.pdf.

[155] Why the Glut In Fiber Lines Remains Huge. http://www.wsj.com/articles/SB111584986236831034.

[156] Zayo Network Map. http://www.zayo.com/network/interactive-map.

[157] Fast Reroute Extensions to RSVP-TE for LSP Tunnels. http://www.ietf.org/rfc/rfc4090.txt, 2005.

[158] Analysis of Country-wide Internet Outages Caused by Censorship. In *Proceedings of ACM IMC*, 2011.

[159] Centurylink and qwest complete merger. http://news.centurylink.com/news/centurylink-and-qwest-complete-merger, 2011.

[160] Zayo group completes acquisition of Abovenet. http://www.zayo.com/news/zayo-group-completes-acquisition-of-abovenet-3/, 2012.

[161] GENI: A Federated Testbed for Innovative Network Experiments. *Computer Networks*, 2014.

[162] Data center interconnect sales growth driver for fiber-optic network gear says Ovum. http://www.lightwaveonline.com/articles/2015/01/data-center-interconnect-sales-growth-driver-for-fiber-optic-network-gear-says-ovum.html, 2015.

[163] Digital Realty Closes $1.9B Telx Acquisition. http://www.datacenterknowledge.com/archives/2015/10/12/digital-realty-closes-1-9b-telx-acquisition/, 2015.

[164] Equinix Closes Bit-isle Deal, Expands Japan Data Center Footprint. http://www.datacenterknowledge.com/archives/2015/11/04/equinix-closes-bit-isle-deal-expands-japan-data-center-footprint/, 2015.

[165] Equinix closes its blockbuster $3.8B acquisition TelecityGroup acquisition. http://www.datacenterknowledge.com/archives/2016/01/15/equinix-closes-blockbuster-3-8b-telecitygroup-acquisition/, 2015.

[166] Lightower Acquires ColocationZone, Enterprise-Class Data Center in Chicago. http://www.lightower.com/company/news/press-releases/lightower-acquires-colocationzone-enterprise-class-data-center-in-chicago/#.V9yYazsgkdc, 2015.

[167] Report: AT&T to Sell $2B Worth of Data Center Assets. http://www.datacenterknowledge.com/archives/2015/02/03/report-2b-worth-of-att-data-centers-may-be-up-for-sale/, 2015.

[168] Windstream to Sell Data Center Business for $575M. http://www.datacenterknowledge.com/archives/2015/10/19/windstream-to-sell-data-center-business-for-575m/, 2015.

[169] Lightower Fiber Networks Acquires Datacenter101, Leading Data Center in Columbus, Ohio. http://www.lightower.com/company/news/press-releases/lightower-fiber-networks-acquires-datacenter101-leading-data-center-in-columbus-ohio/#.V9yX5Tsgkdc, 2016.

[170] Atlantic Metro Communications. http://www.atlanticmetro.net/resources/maps.php, Accessed February 2013.

[171] Lumos Networks. https://www.lumosnetworks.com/sites/default/files/POP-Colocation-List-Feb2013.xlsx, Accessed February 2013.

[172] Zoë Abrams and Michael Schwarz. Ad Auction Design and User Experience. In *Internet and Network Economics*. 2007.

[173] Adobe. The Flex SDK. http://www.adobe.com/devnet/flex.html.

[174] Pankaj K Agarwal, Alon Efrat, Shashidhara K Ganjugunte, David Hay, Swaminathan Sankararaman, and Gil Zussman. The Resilience of WDM Networks to Probabilistic Geographical Failures. In *IEEE INFOCOM*, 2011.

[175] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a Large European IXP. In *ACM SIGCOMM*, 2012.

[176] David Andersen, Hari Balakrishnan, Frans Kaashoek, and Robert Morris. Resilient Overlay Networks. In *ACM SOSP*, 2001.

[177] David Applegate and Edith Cohen. Making Intra-domain Routing Robust to Changing and Uncertain Traffic Demands: Understanding Fundamental Tradeoffs. In *ACM SIGCOMM*, 2003.

[178] Debra J Aron and David E Burnstein. Broadband Adoption in the United States: An empirical analysis. *Down to the Wire: Studies in the Diffusion and Regulation of Telecommunications Technologies, Allan L. Shampine, ed*, 2003.

[179] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Clémence Magnien, and Renata Teixeira. Avoiding Traceroute Anomalies with Paris Traceroute. In *ACM SIGCOMM IMC*, 2006.

[180] Brice Augustin, Balachander Krishnamurthy, and Walter Willinger. IXPs: Mapped? In *ACM SIGCOMM IMC*, 2009.

[181] Ian Austen. Studies Reveal a Rush of Older Women to the Web. *The New York Times.*, 2000.

[182] Paul Barford, Azer Bestavros, John Byers, and Mark Crovella. On the Marginal Utility of Network Topology Measurements. In *ACM IMW*, 2001.

[183] Balagangadhar G Bathula, Rakesh K Sinha, Angela L Chiu, Mark D Feuer, Guangzhi Li, Sheryl L Woodward, Weiyi Zhang, Robert Doverspike, Peter Magill, and Keren Bergman. Constraint Routing and Regenerator Site Concentration in ROADM Networks. *JOCN*, 2013.

[184] Raouf Boutaba, Wojciech Golab, and Youssef Iraqi. Lightpaths on Demand: A Web-services-based Management System. *IEEE Communications magazine*, 2004.

[185] Andrew Brzezinski and Eytan Modiano. Dynamic Reconfiguration and Routing Algorithms for IP-over-WDM Networks with Stochastic Traffic. *IEEE Journal of Lightwave Technology*, 2005.

[186] Randy Bush, Olaf Maennel, Matthew Roughan, and Steve Uhlig. Internet Optometry: Assessing the Broken Glasses in Internet Reachability. In *ACM IMC*, 2009.

[187] Guoray Cai. A GIS Approach to the Spatial Assessment of Telecommunications Infrastructure. *NSE*, 2002.

[188] CAIDA. The Skitter Project. http://www.caida.org/tools/measurement/skitter/, 2007.

[189] Kenneth L Calvert, Matthew B Doar, and Ellen W Zegura. Modeling internet topology. *IEEE Communications Magazine*, 1997.

[190] Joseph Chabarek and Paul Barford. What's in a Name? Decoding Router Interface Names. In *ACM SIGCOMM HotPlanet*, 2013.

[191] Nikolaos Chatzis, Georgios Smaragdakis, Anja Feldmann, and Walter Willinger. There is More to IXPs Than Meets the Eye. *SIGCOMM CCR*, 2013.

[192] Menzie D Chinn and Robert W Fairlie. The Determinants of the Global Digital Divide: A Cross-country Analysis of Computer and Internet Penetration. *Oxford Economic Papers*, 2007.

[193] Angela L Chiu, Gagan Choudhury, George Clapp, Robert Doverspike, Mark Feuer, Joel W Gannett, Janet Jackel, Gi Tae Kim, John G Klincewicz, Taek Jin Kwon, et al. Architectures and Protocols for Capacity Efficient, Highly Dynamic and Highly Resilient Core Networks. In *JOCN*, 2012.

[194] David R. Choffnes and Fabián E. Bustamante. Taming the Torrent: A Practical Approach to Reducing Cross-isp Traffic in Peer-to-peer Systems. In *ACM SIGCOMM*, August 2008.

[195] David Clark. The Design Philosophy of the DARPA Internet Protocols. In *ACM SIGCOMM CCR*, 1988.

[196] Edward H. Clarke. Multipart Pricing of Public Goods. In *Public Choice*, 1971.

[197] Robert W Crandall, William Lehr, and Robert E Litan. *The Effects of Broadband Deployment on Output and Employment: A Cross-sectional Analysis of US Data*. Brookings Institution, 2007.

[198] Nina Czernich, Oliver Falck, Tobias Kretschmer, and Ludger Woessmann. Broadband Infrastructure and Economic Growth. *The Economic Journal*, 2011.

[199] Emilie Danna, Subhasree Mandal, and Arjun Singh. A Practical Algorithm for Balancing the Max-Min Fairness and Throughput Objectives in Traffic Engineering. In *IEEE INFOCOM*, 2012.

[200] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. In *IEEE TEC*, 2002.

[201] Amogh Dhamdhere and Constantine Dovrolis. The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh. In *ACM CoNEXT*, 2010.

[202] Benoit Donnet, Matthew Luckie, Pascal Mérindol, and Jean-Jacques Pansiot. Revealing MPLS Tunnels Obscured from Traceroute. In *ACM SIGCOMM CCR*, 2012.

[203] John C. Doyle, David Alderson, Lun Li, Steven Low, Matthew Roughan, Reiko Tanaka, and Walter Willinger. The "Robust Yet Fragile" Nature of the Internet. In *National Academy of Sciences*, 2005.

[204] Ramakrishnan Durairajan, Paul Barford, Joel Sommers, and Walter Willinger. InterTubes: A Study of the US Long-haul Fiber-optic Infrastructure. In *ACM SIGCOMM*, 2015.

[205] Ramakrishnan Durairajan, Paul Barford, Joel Sommers, and Walter Willinger. GreyFiber: A System for Providing Flexible Fiber-optic Connectivity. In *submission*, 2017.

[206] Ramakrishnan Durairajan, Subhadip Ghosh, Xin Tang, Paul Barford, and Brian Eriksson. Internet Atlas: A Geographic Database of the Internet. In *ACM SIGCOMM HotPlanet*, 2013.

[207] Ramakrishnan Durairajan, Joel Sommers, and Paul Barford. Layer 1-Informed Internet Topology Measurement. In *ACM SIGCOMM IMC*, 2014.

[208] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *The American economic review*, 2007.

[209] Sven Engelhardt. Personal communication, 2014.

[210] Brian Eriksson, Paul Barford, Bruce Maggs, and Robert Nowak. Posit: A Lightweight Approach for IP Geolocation. In *ACM SIGMETRICS Performance Evaluation Review*, 2012.

[211] Brian Eriksson, Paul Barford, Joel Sommers, and Robert Nowak. Inferring Unseen Components of the Internet Core. In *IEEE JSAC*, 2011.

[212] Brian Eriksson, Ramakrishnan Durairajan, and Paul Barford. RiskRoute: A Framework for Mitigating Network Outage Threats. In *ACM CoNEXT*, 2013.

[213] Agner Krarup Erlang. Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. *Elektrotkeknikeren*, 1917.

[214] Dima Feldman, Yuval Shavitt, and Noa Zilberman. A Structural Approach for POP Geo-location. In *Computer Networks*, 2012.

[215] Andrew D. Ferguson, Jordan Place, and Rodrigo Fonseca. Growth Analysis of a Large ISP. In *ACM SIGCOMM IMC*, 2013.

[216] Félix-Antoine Fortin, De Rainville, Marc-André Gardner Gardner, Marc Parizeau, Christian Gagné, et al. DEAP: Evolutionary Algorithms Made Easy. In *JMLR*, 2012.

[217] Bernard Fortz and Mikkel Thorup. Internet Traffic Engineering by Optimizing OSPF Weights. In *IEEE INFOCOM*, 2000.

[218] Bernard Fortz and Mikkel Thorup. Optimizing OSPF/IS-IS Weights in a Changing World. In *IEEE JSAC*, 2002.

[219] Lixin Gao, Tim Griffin, and Jennifer Rexford. Inherently Safe Backup Routing with BGP. In *IEEE INFOCOM*, 2001.

[220] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse? In *PAM*. 2008.

[221] Eduard Glatz and Xenofontas Dimitropoulos. Classifying Internet One-way Traffic. In *ACM IMC*, 2012.

[222] Sean P. Gorman. *Networks, Security And Complexity: The Role of Public Policy in Critical Infrastructure Protection*. Edward Elgar, 2005.

[223] Sean P. Gorman, Laurie Schintler, Raj Kulkarni, and Roger Stough. The Revenge of Distance: Vulnerability Analysis of Critical Information Infrastructure. In *JCCM*, 2004.

[224] Ramesh Govindan, Ina Minei, Mahesh Kallahalla, Bikash Koley, and Amin Vahdat. Evolve or Die: High-Availability Design Principles Drawn from Googles Network Infrastructure. In *ACM SIGCOMM*, 2016.

[225] Theodore Groves. Incentives in Teams. *Econometrica*, 1973.

[226] Kyle Chi Guan. *Cost-effective Optical Network Architecture: A Joint Optimization of Topology, Switching, Routing and Wavelength Assignment.* PhD thesis, Massachusetts Institute of Technology, 2007.

[227] Bamba Gueye, Artur Ziviani, Mark Crovella, and Serge Fdida. Constraint-Based Geolocation of Internet Hosts. In *IEEE/ACM TON*, 2006.

[228] Krishna P. Gummadi, Harsha V. Madhyastha, Steven D. Gribble, Henry M. Levy, and David Wetherall. Improving the Reliability of Internet Paths with One-hop Source Routing. In *USENIX OSDI*, 2004.

[229] Arpit Gupta, Laurent Vanbever, Muhammad Shahbaz, Sean P Donovan, Brandon Schlinker, Nick Feamster, Jennifer Rexford, Scott Shenker, Russ Clark, and Ethan Katz-Bassett. SDX: A Software-Defined Internet Exchange. In *ACM SIGCOMM*, 2014.

[230] Audun Fosselie Hansen, Amund Kvalbein, Tarik Cicic, and Stein Gjessing. Resilient Routing Layers for Network Disaster Planning. In *ICN*, 2005.

[231] Poul E. Heegaard and Kishor S. Trivedi. Network Survivability Modeling. 2009.

[232] P-H Ho, J. Tapolcai, and H. Mouftah. On Achieving Optimal Survivable Routing for Shared Protection in Survivable Next-Generation Internet. *IEEE Transactions on Reliability*, 2004.

[233] Chi-Yao Hong, Srikanth Kandula, Ratul Mahajan, Ming Zhang, Vijay Gill, Mohan Nanduri, and Roger Wattenhofer. Achieving High Utilization with Software-driven WAN. In *ACM SIGCOMM*, 2013.

[234] Bradley Huffaker, Marina Fomenkov, and kc Claffy. DRoP:DNS-based Router Positioning. In *ACM SIGCOMM CCR*, 2014.

[235] Philip Hunter. Pakistan YouTube Block Exposes Fundamental Internet Security Weakness. *Computer Fraud and Security*, 2008.

[236] Van Jacobson and Steve Deering. Traceroute, 1989.

[237] Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, et al. B4: Experience with a Globally-deployed Software Defined WAN. In *ACM SIGCOMM*, 2013.

[238] Virajith Jalaparti, Ivan Bliznets, Srikanth Kandula, Brendan Lucier, and Ishai Menache. Dynamic Pricing and Traffic Engineering for Timely Inter-Datacenter Transfers. In *ACM SIGCOMM*, 2016.

[239] Xin Jin, Yiran Li, Da Wei, Siming Li, Jie Gao, Lei Xu, Guangzhi Li, Wei Xu, and Jennifer Rexford. Optimizing Bulk Transfers with Software-Defined Optical WAN. In *ACM SIGCOMM*, 2016.

[240] Srikanth Kandula, Dina Katabi, Bruce Davie, and Anna Charny. Walking the Tightrope: Responsive yet Stable Traffic Engineering. In *ACM SIGCOMM*, 2005.

[241] Srikanth Kandula, Ishai Menache, Roy Schwartz, and Spandana Raj Babbula. Calendaring for Wide Area Networks. In *ACM SIGCOMM*, 2014.

[242] Krishna Kant and Casey Deccio. Security and Robustness in the Internet Infrastructure.

[243] Anton Kapela. Personal communication, 2014.

[244] Raul Katz. The Impact of Broadband on the Economy: Research to Date and Policy Issues. *Broadband Series*, 2012.

[245] Ethan Katz-Bassett, John P John, Arvind Krishnamurthy, David Wetherall, Thomas Anderson, and Yatin Chawathe. Towards IP Geolocation Using Delay and Topology Measurements. In *ACM SIGCOMM IMC*, 2006.

[246] Ethan Katz-Bassett, Harsha V. Madhyastha, John P. John, Arvind Krishnamurthy, David Wetherall, and Thomas Anderson. Studying Black Holes in the Internet with Hubble. In *USENIX NSDI*, 2008.

[247] Ethan Katz-Bassett, Colin Scott, David R. Choffnes, Italo Cunha, Vytautas Valancius, Nick Feamster, Harsha V. Madhyastha, Thomas E. Anderson, and Arvind Krishnamurthy. LIFEGUARD: Practical Repair of Persistent Route Failures. In *ACM SIGCOMM*, 2012.

[248] Frank P Kelly. Blocking Probabilities in Large Circuit-Switched Networks. *Advances in applied probability*, 1986.

[249] Frank P Kelly, Aman K Maulloo, and David KH Tan. Rate Control for Communication Networks: Shadow prices, Proportional fairness and Stability. *Journal of the Operational Research society*, 1998.

[250] Mirosl Aw Klinkowski and Krzysztof Walkowiak. On the Advantages of Elastic Optical Networks for Provisioning of Cloud Computing Traffic. In *IEEE Network*, 2013.

[251] Simon Knight, Hung X. Nguyen, Nick Falkner, Rhys Alistair Bowden, and Matthew Roughan. The Internet Topology Zoo. In *IEEE JSAC*, 2011.

[252] Tjalling Koopmans and Martin Beckmann. Assignment Problems and the Location of Economic Activities. *Econometrica*, 1957.

[253] Alok Kumar, Sushant Jain, Uday Naik, Anand Raghuraman, Nikhil Kasinadhuni, Enrique Cauich Zermeno, C Stephen Gunn, Jing Ai, Björn Carlin, Mihai Amarandei-Stavila, et al. BwE: Flexible, Hierarchical Bandwidth Allocation for WAN Distributed Computing. In *ACM SIGCOMM*, 2015.

[254] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. Internet Inter-domain Traffic. In *ACM SIGCOMM*, 2010.

[255] Bruce LaBuda. Personal communication, 2014.

[256] Anukool Lakhina, John Byers, Mark Crovella, and Ibrahim Matta. On the geographic location of internet resources. In *IEEE JSAC*, 2003.

[257] Cedric F. Lam, Hong Liu, Bikash Koley, Xiaoxue Zhao, Valey Kamalov, and Vijay Gill. Fiber Optic Communication Technologies: What's Needed for Datacenter Network Operations. *IEEE Communications Magazine*, 2010.

[258] Nikolaos Laoutaris, Michael Sirivianos, Xiaoyuan Yang, and Pablo Rodriguez. Inter-datacenter Bulk Transfers with NetStitcher. In *ACM SIGCOMM*, 2011.

[259] Tom Limoncelli, Jesse Robbins, Kripa Krishnan, and John Allspaw. Resilience Engineering: Learning to Embrace Failure. *Communications of the ACM*, 2012.

[260] Hongqiang Harry Liu, Srikanth Kandula, Ratul Mahajan, Ming Zhang, and David Gelernter. Traffic Engineering with Forward Fault Correction. In *ACM SIGCOMM*, 2014.

[261] Hongqiang Harry Liu, Srikanth Kandula, Ratul Mahajan, Ming Zhang, and David Gelernter. Traffic Engineering with Forward Fault Correction. In *ACM SIGCOMM*, 2015.

[262] Richard TB Ma, Dah Ming Chiu, John Lui, Vishal Misra, and Dan Rubenstein. Internet Economics: The Use of Shapley Value for ISP Settlement. In *IEEE/ACM TON*, 2010.

[263] Harsha V Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani. iPlane: An Information Plane for Distributed Services. In *USENIX OSDI*, 2006.

[264] Edward J Malecki. The Economic Geography of the Internet's Infrastructure. In *Economic geography*, 2002.

[265] Zhuoqing Morley Mao, Jennifer Rexford, Jia Wang, and Randy H Katz. Towards an Accurate AS-Level Traceroute Tool. In *ACM SIGCOMM*, 2003.

[266] Kevin McGrattan and Anthony Hamins. Numerical Simulation of the Howard Street Tunnel Fire. *Fire Technology*, 2006.

[267] Matthew K Mukerjee, David Naylor, Junchen Jiang, Dongsu Han, Srinivasan Seshan, and Hui Zhang. Practical, Real-time Centralized Control for CDN-based Live Video Delivery. In *ACM SIGCOMM*, 2015.

[268] Venkat Padmanabhan and Lakshmi Subramanian. An Investigation of Geographic Mapping Techniques for Internet Hosts. In *ACM SIGCOMM*, 2001.

[269] Vern Paxson. *Measurement and Analysis of End-to-end Internet Dynamics*. PhD thesis, University of California at Berkeley, 1997.

[270] Vern Paxson, Andrew K Adams, and Matt Mathis. Experiences with NIMI. In *IEEE SAINT*, 2002.

[271] Larry Peterson and Timothy Roscoe. The Design Principles of PlanetLab. *ACM SIGOPS OS Review*, 2006.

[272] Lin Quan, John Heidemann, and Yuri Pradkin. Detecting Internet Outages with Precise Active Probing (extended). In *USC Technical Report*, February 2012.

[273] Lee Rainie, Susannah Fox, John Horrigan, Amanda Lenhart, and Tom Spooner. Tracking Online Life: How Women Use the Internet to Cultivate Relationships with Family and Friends. *Washington, DC: The Pew Internet and American Life Project*, 2000.

[274] Byrav Ramamurthy and Ashok Ramakrishnan. Virtual Topology Reconfiguration of Wavelength-routed Optical WDM Networks. In *IEEE GLOBECOM*, 2000.

[275] Chathurika Ranaweera, Pat Iannone, Kostas Oikonomou, Ken C Reichmann, and Rakesh Sinha. Cost Optimization of Fiber Deployment for Small Cell Backhaul. In *NFOEC*, 2013.

[276] Amir H Rasti, Nazanin Magharei, Reza Rejaie, and Walter Willinger. Eyeball ASes: From Geography to Connectivity. In *ACM IMC*, 2010.

[277] Anne Rickert, Anya Sacharow, et al. It's a Woman's World Wide Web. *Media Metrix and Jupiter Communications*, 2000.

[278] C Osvaldo Rodriguez. Affordable Wireless Connectivity Linking Poor Latin American Communities Binding Their Schools by Sharing ICT Training for "Maestros" of Primary Schools. In *Internationalization, Design and Global Development*. 2009.

[279] Gregory L Rohde, Robert Shapiro, et al. Falling Through the Net: Toward Digital Inclusion. *US Department of Commerce*, 2000.

[280] Matthew Roughan, Simon Jonathan Tuke, and Olaf Maennel. Bigfoot, sasquatch, the yeti and other missing links: what we don't know about the AS graph. In *Proceedings of ACM Internet measurement conference*, 2008.

[281] Mario A. Sanchez, Fabian E. Bustamante, Balachander Krishnamurthy, Walter Willinger, Georgios Smaragdakis, and Jeffrey Erman. Inter-domain traffic estimation for the outsider. In *ACM IMC*, 2014.

[282] Arjuna Sathiaseelan and Jon Crowcroft. LCD-Net: Lowest Cost Denominator Networking. *ACM SIGCOMM CCR*, 2013.

[283] M Sawada, Daniel Cossette, Barry Wellar, and Tolga Kurt. Analysis of the Urban/Rural Broadband Divide in Canada: Using GIS in Planning Terrestrial Wireless Deployment. *GIQ*, 2006.

[284] Yuval Shavitt and Eran Shir. DIMES: Let the Internet Measure Itself. In *ACM SIGCOMM CCR*, 2005.

[285] Yuval Shavitt and Udi Weinsberg. Quantifying the Importance of Vantage Points Distribution in Internet Topology Measurements. In *IEEE INFOCOM*, 2009.

[286] Scott Shenker, David Clark, Deborah Estrin, and Shai Herzog. Pricing in Computer Networks: Reshaping the Research Agenda. In *ACM SIGCOMM CCR*, 1996.

[287] Rob Sherwood, Adam Bender, and Neil Spring. Discarte: A Disjunctive Internet Cartographer. In *ACM SIGCOMM*, 2008.

[288] Rob Sherwood and Neil Spring. Touring the Internet in a TCP Sidecar. In *ACM SIGCOMM IMC*, 2006.

[289] Ankit Singla, Balakrishnan Chandrasekaran, P. Brighten Godfrey, and Bruce Maggs. The Internet at the Speed of Light. In *ACM Hotnets*, 2014.

[290] Joel Sommers, Paul Barford, and Brian Eriksson. On the Prevalence and Characteristics of MPLS Deployments in the Open Internet. In *ACM SIGCOMM IMC*, 2011.

[291] Larissa Spinelli, Mark Crovella, and Brian Eriksson. AliasCluster: A lightweight approach to interface disambiguation. In *Global Internet Symposium*, 2013.

[292] Neil Spring, Ratul Mahajan, and David Wetherall. Measuring ISP topologies with Rocketfuel. *ACM SIGCOMM*, 2002.

[293] Mininet Team. Mininet: An Instant Virtual Network on Your Laptop (or Other PC), 2012.

[294] Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack. Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack: : Critical National Infrastructures. In *Critical National Infrastructures Report*, 2004.

[295] Leanne Townsend, Arjuna Sathiaseelan, Gorry Fairhurst, and Claire Wallace. Enhanced Broadband Access as a Solution to the Social and Economic Problems of the Rural Digital Divide. *LE*, 2013.

[296] Vytautas Valancius, Nick Feamster, Ramesh Johari, and Vijay Vazirani. Mint: A Market for Internet Transit. In *ACM CoNEXT*, 2008.

[297] Vytautas Valancius, Cristian Lumezanu, Nick Feamster, Ramesh Johari, and Vijay V Vazirani. How Many Tiers? Pricing in the Internet Transit Market. In *ACM SIGCOMM*, 2011.

[298] William Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 1961.

[299] Hao Wang, Yang Richard Yang, Paul H. Liu, Jia Wang, Alexandre Gerber, and Albert Greenberg. Reliability as an Interdomain Service. In *ACM SIGCOMM*, 2007.

[300] Yong Wang, Daniel Burgener, Marcel Flores, Aleksandar Kuzmanovic, and Cheng Huang. Towards Street-Level Client-Independent IP Geolocation. In *USENIX NSDI*, 2011.

[301] Bernard M Waxman. Routing of multipoint connections. *IEEE JSAC*, 1988.

[302] Darrell M West. Digital Divide: Improving Internet Access in the Developing World through Affordable Services and Diverse Content. *Brookings Institution*, 2015.

[303] Walter Willinger and John Doyle. Robustness and the internet: Design and evolution. *Robust-Design: A Repertoire of Biological, Ecological, and Engineering Case Studies*, 2002.

[304] Bernard Wong, Ivan Stoyanov, and Emin Gün Sirer. Octant: A Comprehensive Framework for the Geolocation of Internet Hosts. In *USENIX NSDI*, 2007.

[305] Jian Wu, Ying Zhang, Z. Morley Mao, and Kang G. Shin. Internet Routing Resilience to Failures: Analysis and Implications. In *ACM CoNEXT*, 2007.

[306] Yu Wu, Zhizhong Zhang, Chuan Wu, Chuanxiong Guo, Zongpeng Li, and Francis CM Lau. Orchestrating Bulk Data Transfers Across Geo-distributed Datacenters. In *IEEE TCC*, 2015.

[307] Dahai Xu, Guangzhi Li, Byrav Ramamurthy, Angela Chiu, Dongmei Wang, and Robert Doverspike. On Provisioning Diverse Circuits in Heterogeneous Multi-layer Optical Networks. In *Computer Communications*, 2013.

[308] He Yan, Ricardo Oliveira, Kevin Burnett, Dave Matthews, Lixia Zhang, and Dan Massey. BGPmon: A Real-Time, Scalable, Extensible Monitoring System. In *CATCH*, 2009.

[309] Vinod Yegneswaran, Paul Barford, and Johannes Ullrich. Internet Intrusions: Global Characteristics and Prevalence. In *ACM SIGMETRICS*, 2003.

[310] Beichuan Zhang, Raymond Liu, Daniel Massey, and Lixia Zhang. Collecting the Internet AS-Level Topology. In *ACM SIGCOMM CCR*, 2005.

[311] Hong Zhang, Kai Chen, Wei Bai, Dongsu Han, Chen Tian, Hao Wang, Haibing Guan, and Ming Zhang. Guaranteeing Deadlines for Inter-datacenter Transfers. In *ACM Eurosys*, 2015.

[312] Ming Zhang, Yaoping Ruan, Vivek S Pai, and Jennifer Rexford. How DNS Misnaming Distorts Internet Topology Mapping. In *USENIX ATC*, 2006.

[313] Yu Zhang, Ricardo Oliveira, Yangyang Wang, Shen Su, Baobao Zhang, Jun Bi, Hongli Zhang, and Lixia Zhang. A Framework to Quantify the Pitfalls of Using Traceroute in AS-level Topology Measurement. *IEEE JSAC*, 2011.

[314] Liang Zheng, Carlee Joe-Wong, Chee Wei Tan, Mung Chiang, and Xinyu Wang. How to Bid the Cloud. In *ACM SIGCOMM*, 2015.

[315] Ling Zhou. Vulnerability Analysis of the Physical Part of the Internet. In *International Journal of Critical Infrastructures*, 2010.

[316] Xia Zhou, Zengbin Zhang, Yibo Zhu, Yubo Li, Saipriya Kumar, Amin Vahdat, Ben Y. Zhao, and Haitao Zheng. Mirror Mirror on the Ceiling: Flexible Wireless Links for Data Centers. In *ACM SIGCOMM*, 2012.

[317] Yaping Zhu, Andy Bavier, Nick Feamster, Sampath Rangarajan, and Jennifer Rexford. UFO: A Resilient Layered Routing Architecture. In *ACM SIGCOMM CCR*, 2008.

[318] Earl Zmijewski. Mediterranean cable break. *Renesys Blog*, 2008.