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

Ramakrishnan Durairajan[†], Paul Barford[†*], Joel Sommers[+], Walter Willinger[‡]
{rkrish,pb}@cs.wisc.edu, jsommers@colgate.edu, wwillinger@niksun.com
[†]University of Wisconsin - Madison [*]comScore, Inc. [+]Colgate University [‡]NIKSUN, Inc.

ABSTRACT

The complexity and enormous costs of installing new long-haul fiber-optic infrastructure has led to a significant amount of infrastructure sharing in previously installed conduits. In this paper, we study the characteristics and implications of infrastructure sharing by analyzing the long-haul fiber-optic network in the US.

We start by using fiber maps provided by tier-1 ISPs and major cable providers 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[1] connectivity characteristics and confirm a clear correspondence between long-haul fiber-optic, roadway, and railway infrastructures. Next, we examine the prevalence of high-risk links by mapping end-to-end paths resulting from large-scale traceroute campaigns onto our fiber-optic infrastructure map. 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. In particular, focusing on a subset of high-risk links is sufficient to improve the overall robustness of the network to failures. Finally, we discuss the implications of our findings on issues related to performance, net neutrality, and policy decision-making.

CCS Concepts

•Networks → Physical links; Physical topologies;

Keywords

Long-haul fiber map; shared risk; risk mitigation

[1]The constructed long-haul map along with datasets are openly available to the community through the U.S. DHS PREDICT portal (www.predict.org).

SIGCOMM '15, August 17–21, 2015, London, United Kingdom
© 2015 ACM. ISBN 978-1-4503-3542-3/15/08…$15.00
DOI: http://dx.doi.org/10.1145/2785956.2787499

1 Introduction

The desire to tackle the many challenges posed by novel designs, technologies and applications such as data centers, cloud services, software-defined networking (SDN), network functions virtualization (NFV), mobile communication and the Internet-of-Things (IoT) has fueled many of the recent research efforts in networking. The excitement surrounding the future envisioned by such new architectural 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. 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 today's *physical* Internet where individual components such as cell towers, routers or switches, and fiber-optic cables are concrete entities with well-defined geographic locations (see, *e.g.*, [2, 36, 83]). This general lack of a basic understanding of the physical Internet is exemplified by the much-ridiculed metaphor used in 2006 by the late U.S. Senator Ted Stevens (R-Alaska) who referred to the Internet as "a series of tubes" [65].[2]

The focus of this paper 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),

[2]Ironically, this infamous metaphor turns out to be not all that far-fetched when it comes to describing the portion of the physical Internet considered in this paper.

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 paper 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 [34], 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 paper. 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 [99] 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 [100]). 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 [7]. 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).

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*[3], and it is this fixed infrastructure which we seek to identify.

Our high-level definition of a long-haul link[4] 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

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

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

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 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.

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 [83]. 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 [83]. 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[5] from 5 tier-1 and 4 major cable providers: AT&T [6], Comcast [16], Cogent [14], EarthLink [29], Integra [43], Level3 [48], Suddenlink [63], Verizon [72] and Zayo [73]. For example, the map we used for Comcast's network [16] lists all the node information along with the exact geography of long-haul fiber links. Table 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.

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.*, [13, 18, 26]), environmental impact statements (*e.g.*, [71]), documentation released by third-party fiber services (*e.g.*, [3–5, 10]), indefeasible rights of use (IRU) agreements (*e.g.*, [44, 45]), press releases (*e.g.*, [49, 50, 52, 53]), and other related resources (*e.g.*, [8, 11, 23, 27, 28, 59, 67]).

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 [31]), and which local organization has jurisdiction varies state-by-state [1]. 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 [12], *(2)* a franchise agreement [20,21] 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 [24] 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 [68] shows pictures [69] of Comcast and AT&T's fiber deployed in the Urbana, IL area, and *(5)* documents from the CASF project [70] in Nevada county, CA show that Comcast has deployed fiber along with AT&T and Suddenlink.

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 con-

[5] 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 [58], 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 [35] and validation/inference (step 2) during the process.

Table 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
Number of nodes	25	26	69	248	27	240	39	116	98
Number of links	57	71	84	370	36	336	42	151	111

duits without considering the 9 ISPs above). For example, for Sprint's network [60], 102 links were added and for CenturyLink's network [9], 134 links were added.

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 [60] 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 [13] 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 [22] 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 [22]). 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 [19]) 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 [47] 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 [61] and their settlement details are publicly available [62], *(2)* a whitepaper related to a research network initiative in Virginia identifies link location and sharing details regarding Sprint fiber [27], *(3)* the "coastal route" [13] 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 [33] 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.

2.5 The US Long-haul Fiber Map

The final map constructed through the process described in this section is shown in Figure 1, 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).

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.

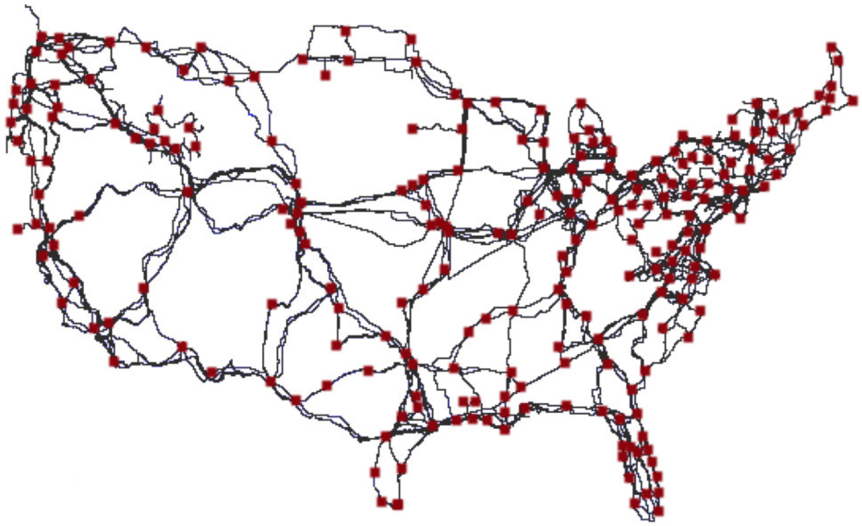

Figure 1: Location of physical conduits for networks considered in the continental United States.

Interestingly, recommendation 6.4 made by the FCC in chapter 6 of the National Broadband Plan [7] 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.*, [34]. 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 [41][6]), 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 information that is available from publicly available records, albeit in a convenient but non-free form.

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 [36]. 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 [51]. The geographic layout of our roadway and railway data sets can be seen in Figure 2 and Figure 3, respectively. In comparison, the physical link geographic information for the networks under consideration can be seen in the Figure 1.

Figure 2: NationalAtlas roadway infrastructure locations.

Figure 3: NationalAtlas railway infrastructure locations.

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

We use the *polygon overlap* analysis capability in the ArcGIS [30] to quantify the correspondence between physical links and transportation infrastructure. In Figure 4, aggregating across all networks under consideration, we compare the fraction of each path that is co-located with 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.

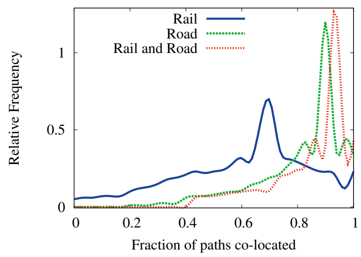

Figure 4: Fraction of physical links co-located with transportation infrastructure.

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 we show the Level 3-provided physical link locations outside Laurel, MS, and in the right-hand plot we show Google Maps [37] 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.

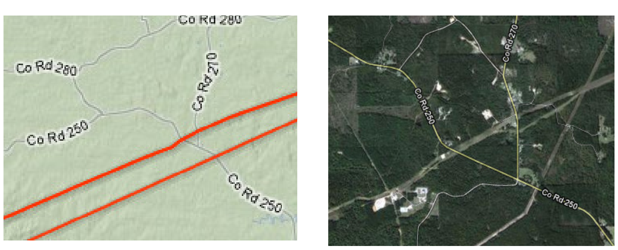

Figure 5: Satellite image validated right-of-way outside of Laurel, MS. (Left) - Level 3 Provided fiber map. (Right) - Google Maps satellite view.

A few examples can be shown in Level3's network [48], 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 [56], 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 pipeline. Similarly, the link from Houston, TX to Atlanta, GA is deployed along with NGL pipelines.

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 [2]. We intend to analyze different dimensions of network resilience in future work.

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 §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 §2.

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 6 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 1), 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.

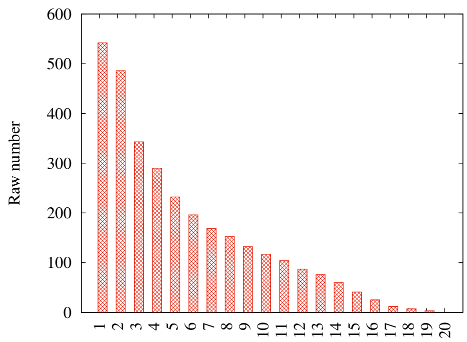

Figure 6: ISP Ranking.

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 [77, 83, 103]).
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 6. 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.

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* [25] and *open trench* [55], 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.

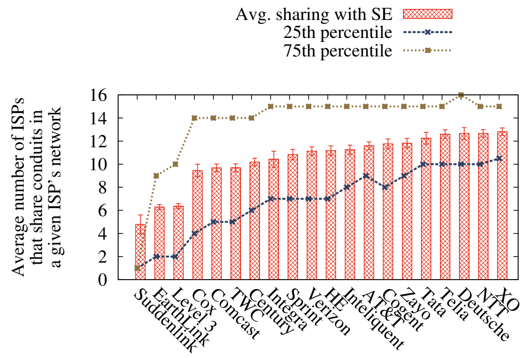

Figure 7: The raw number of shared conduits by ISPs.

How similar are ISP risk profiles? Using the risk matrix we calculate the Hamming [39] 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 8 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.

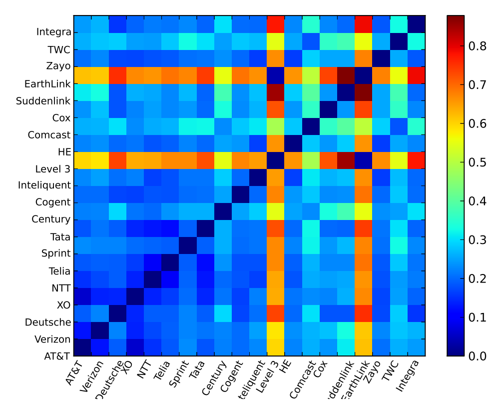

Figure 8: Similarity of risk profiles of ISPs calculated using Hamming distance.

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 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.

4.3 Risk Metric: Connectivity + Traffic

In this section, we follow the method of [99] 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 [80] 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 [78, 92], 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 [101] 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 2 and Table 3 show the top 20 conduits for west-origin east-bound and east-origin west-bound probes[7] 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 9 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 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.

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

Table 2: 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

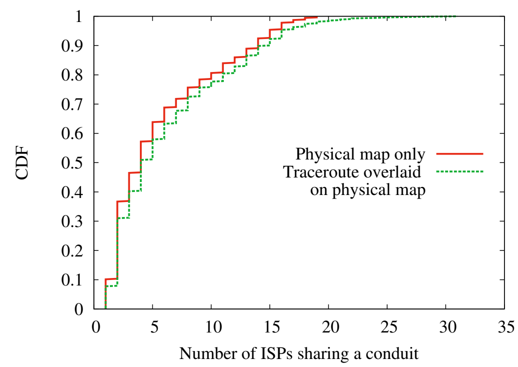

Figure 9: CDF of number of ISPs sharing a conduit before and after considering the traceroute data as a proxy for traffic volumes.

Table 3: 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

Distribution of traffic. We also ranked the ISPs based on the number of conduits used to carry traffic. Table 4 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. 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.

Table 4: 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

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.

5.1 Increasing Network Robustness (I)

We first examine the possibility of improving the long-haul infrastructure's robustness (*i.e.*, to reduce the impact of fiber cuts by reducing the level of conduit sharing among ISPs[8]) 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 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,

$$OP_{i,j}^{robust} = \min_{P_{i,j} \in \mathscr{E}^A} SR\left(\mathbf{P_{i,j}}\right) \qquad (1)$$

where $\mathscr{E}^A$ 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 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.

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

Figure 10 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 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.

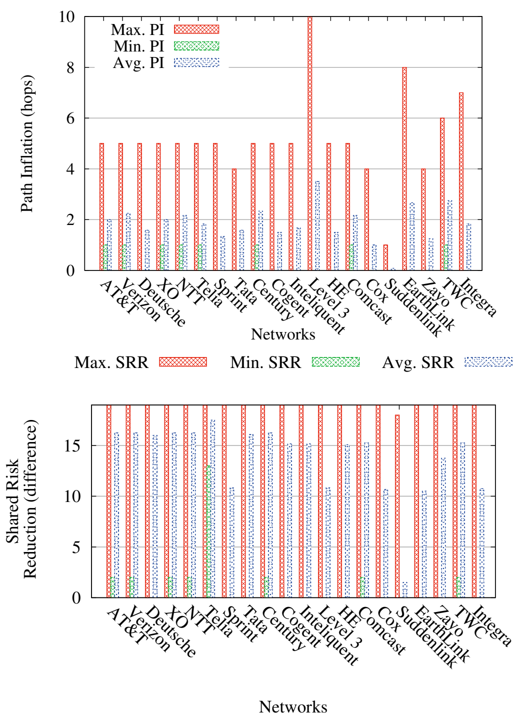

Figure 10: Path Inflation (top) and Shared Risk Reduction (bottom) based on the robustness suggestion framework for the twelve heavily shared links.

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,

Table 5: 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

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.

5.2 Increasing Network Robustness (II)

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 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 ∈ V and {u,v} ∉ 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 ∪ S) for some set S ∈ $\hat{E}$. We want to find S ∈ $\hat{E}$ of size k such that

$$S = arg\max(\lambda_A - \lambda_{\hat{A}}) \quad (2)$$

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.

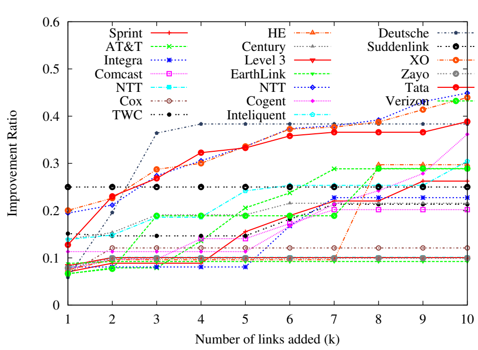

Figure 11: Potential improvements to ISP

Figure 11 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.

5.3 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.

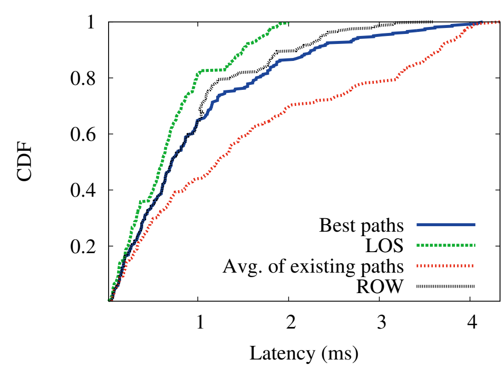

Figure 12: Comparison of best links against avg. latencies of links, ROW links and LOS links.

Figure 12 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 [32]). 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.

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.

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 [54].

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 [17] 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 [38]. 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.

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 [79]. 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.[9] 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 [15].

7 Related Work

The Internet's basic design [81] makes it robust against failures of physical components such as routers and links. While IP routing allows the network to dynamically detect and route around failures, events such as natural or technological disasters (*e.g.*, [42,57]), malicious attacks (*e.g.*, [66]) and benign incidents (*e.g.*, [64]) can have localized effects, including the loss of connectivity for varying numbers of Internet users for certain amounts of time. The main reasons for such localized and temporal Internet outages are typically a lack

[9]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.

of geographic diversity in connectivity [2,40] and a tendency for significant physical infrastructure sharing among the affected providers—the very focus of this paper. In particular, this paper is not about the Internet's vulnerability to non-physical cyber attacks (*e.g.*, [46]) that rely on the existence and full functionality of the Internet's physical infrastructure to achieve their goals and do maximal damage [82].

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 [82, 104], vulnerability [86, 87, 106], survivability [90, 91], resilience analysis [74,84,105], reachability [76,94], security and robustness of components [93], fault detection/localization [85, 95, 98], and the development of resilient routing protocols [75,88,89, 102, 107]. 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.

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 [2, 36]), 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 [97]). 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 [30], and is unprecedented both in terms of accuracy and ability for validation. In contrast to the work by Lakhina *et al.* [96] 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 [36].

8 Summary and Future Work

In this paper 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. In future work, we plan 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. We also plan to generate annotated versions of our map, focusing in particular on traffic and propagation delay.

Acknowledgements

We thank our shepherd, David Wetherall, and the anonymous reviewers for their invaluable feedback. This material is based upon work supported by the NSF under grant CNS-1054985, DHS grant BAA 11-01 and AFRL grant FA8750-12-2-0328. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF, DHS or AFRL.

9 References

[1] 50-State survey of rights-of-way statutes. http://www.ntia.doc.gov/legacy/ntiahome/staterow/rowtable.pdf.

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

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

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

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

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

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

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

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

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

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

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

[13] Coastal Route - SFO. to LA. for several ISPs. http://www.ustaxcourt.gov/InOpHistoric/anschutz.TCM.WPD.pdf.
[14] Cogent Network Map. www.internetatlas.org.
[15] Colorado Department of Tranportation document. http://www.wmxsystems.com/EndUserFiles/44489.pdf.
[16] Comcast Network Map. http://business.comcast.com/about-us/our-network.
[17] Communications Act of 1934. http://transition.fcc.gov/Reports/1934new.pdf.
[18] County document for Zayo and ATT. http://www2.ntia.doc.gov/files/grantees/nt10bix5570098_california_broadband_cooperative_inc_ppr2013_q3.pdf.
[19] Cox and Level3 lease agreement - Ocala, FL. http://ocalafl.iqm2.com/Citizens/FileOpen.aspx?Type=30&ID=2988.
[20] Cox Franchise Agreement in Fairfax county, Virginia. http://www.fairfaxcounty.gov/cable/regulation/franchise/cox/franchise_agreement_cox_fairfax_cty.pdf.
[21] Cox Franchise Agreement in Fairfax county, Virginia. https://www.natoa.org/events/RickEllrod.pdf.
[22] Cox Network Map. http://www.cox.com/wcm/en/business/datasheet/brc-backbone-map-q4-2013.pdf?campcode=brc_un_07_101513.
[23] Dascom Systems Report. http://dascom-systems.com/new/wp-content/uploads/eCare-Overview-of-Nevada-project-for-ICBN-Constituency.pdf.
[24] 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.
[25] Dig Once ordinance: Muni Networks. http://www.muninetworks.org/tags/tags/dig-once.
[26] Douglas County, CO document for CenturyLink and Level3. http://www.crgov.com/DocumentCenter/Home/View/945.
[27] E-Corridors. http://www.ecorridors.vt.edu/research/papers/stircne/vol02-connecting.pdf.
[28] EAGLENET. https://www.co-eaglenet.net/wp-content/uploads/2013/05/EAGLE-NET-001-Addenda-4-Responses-to-Questions-2013-05-29.pdf.
[29] EarthLink Network Map. http://www.earthlinkbusiness.com/support/network-map.xea.
[30] ESRI ArcGIS. http://www.arcgis.com/features/.
[31] 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.
[32] Fiber-optic latency. http://www.m2optics.com/blog/bid/70587/Calculating-Optical-Fiber-Latency.
[33] Fiber Optic Settlement. https://fiberopticsettlements.com.
[34] FiberLocator Online. http://www.fiberlocator.com/product/fiberlocator-online.
[35] Georeferencing in ArcGIS. http://resources.arcgis.com/en/help/main/10.1/index.html#/Fundamentals_of_georeferencing_a_raster_dataset/009t000000mn000000.
[36] GMU Mapping Project.. http://gembinski.com/interactive/GMU/research.html.
[37] Google Maps. http://maps.google.com/.
[38] Google to FCC. http://www.engadget.com/2015/01/01/google-letter-fcc-title-ii/.
[39] Hamming Distance. http://en.wikipedia.org/wiki/Hamming_distance.
[40] How to Destroy the Internet. http://gizmodo.com/5912383/how-to-destroy-the-internet.
[41] Image from Geo-tel. http://www.techrepublic.com/article/the-google-fiber-lottery/.
[42] Impact of the 2003 blackouts on Internet communications (Preliminary report), Nov. 2003. http://research.dyn.com/content/uploads/2013/05/Renesys_BlackoutReport.pdf.
[43] Integra Telecom Network Map. http://www.integratelecom.com/resources/Assets/long-haul-fiber-network-map-integra.pdf.
[44] IRU between Level3 and Comcast. http://investors.level3.com/investor-relations/press-releases/press-release-details/2004/Comcast-Extends-National-Fiber-Infrastructure/default.aspx.
[45] IRU Swaps. http://chogendorn.web.wesleyan.edu/excessive.pdf.
[46] Is Internet backbone vulnerable to cyber attack? http://www.sciencedaily.com/releases/2010/12/101214085539.htm.
[47] Level 3 and Comcast. http://corporate.comcast.com/news-information/news-feed/comcast-extends-national-fiber-infrastructure.
[48] Level3 Network Map. http://maps.level3.com/default/.
[49] Media Advertisement. http://www.wikinvest.com/stock/XO_Holdings_Inc_(XOHO)/Cox_Communications_Las_Vegas_Indefeasible_Right_Iru_Agreement.
[50] Media Advertisement. http://www.wikinvest.com/stock/XO_Holdings_Inc_(XOHO)/Level.
[51] National Atlas. http://www.nationalatlas.gov/.
[52] News article from SMW3. http://www.smw3.com/smw3/signin/download/iru/Standard%20Agreement/AG-IRU-Final%20(Mar03).pdf.
[53] News article from Telecom Ramblings. http://www.telecomramblings.com/2014/09/tw-telecom-pushes-north-new-york/.
[54] NSFNet. http://en.wikipedia.org/wiki/National_Science_Foundation_Network.
[55] Open Trench: Muni Networks. http://muninetworks.org/tags/tags/open-trench.
[56] Pipeline 101. http://www.pipeline101.org/where-are-pipelines-located.
[57] Quake shakes up the net, Dec. 2006. http://www.thestar.com.my/story/?file=%2f2006%2f12%2f28%2fnation%2f16426778&sec=nation.
[58] 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.
[59] SCAC. http://securities.stanford.edu/filings-documents/1023/GX02/2004322_r01c_02910.pdf.
[60] Sprint Network Map. https://www.sprint.net/images/network_maps/full/NorthAmerica-Global-IP.png.
[61] Sprint using Level3's fiber in Detroit. http://apps.fcc.gov/ecfs/document/view;jsessionid=yzRnRk4NJnlqDtlg0shlW2QDTSd3J2x0nz9Ryg5TJpMyX0rnYXxG!-1694890999!-477673473?id=6516791680.
[62] Sprint using Level3's fiber in Detroit settlement. https://fiberopticsettlements.com/michigan/LinkClick.aspx?fileticket=912Y0QkAV2E%3d&tabid=62&mid=420.
[63] SuddenLink Network Map. http://www.suddenlinkcarrier.com/index.php?page=our-network.

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

[65] The internet is not a big truck. It's a series of tubes. http://en.wikipedia.org/wiki/Series_of_tubes.

[66] The National Research Council of the National Academies, 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.

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

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

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

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

[71] US 36 Corridor Final Environmental Impact Statement: Chapter 4 âAffected Environment and Environmental Consequences. https://www.codot.gov/projects/us36eis/documents/us-36-final-eis-volume-i/section-4-18_utilities.pdf.

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

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

[74] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman. The Resilience of WDM Networks to Probabilistic Geographical Failures. In *IEEE INFOCOM*, 2011.

[75] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris. Resilient Overlay Networks. In *ACM SOSP*, 2001.

[76] R. Bush, O. Maennel, M. Roughan, and S. Uhlig. Internet Optometry: Assessing the Broken Glasses in Internet Reachability. In *ACM IMC*, 2009.

[77] K. L. Calvert, M. B. Doar, and E. W. Zegura. Modeling Internet Topology. *IEEE Communications*, 1997.

[78] J. Chabarek and P. Barford. What's in a Name? Decoding Router Interface Names. In *ACM HotPlanet*, 2013.

[79] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There is More to IXPs Than Meets the Eye. *SIGCOMM CCR*, 2013.

[80] D. R. Choffnes and F. E. Bustamante. Taming the Torrent: A Practical Approach to Reducing Cross-ISP Traffic in Peer-to-peer Systems. In *ACM SIGCOMM*, 2008.

[81] D. Clark. The design philosophy of the darpa internet protocols. *SIGCOMM CCR*, 1988.

[82] J. C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, R. Tanaka, and W. Willinger. The "robust yet fragile" nature of the Internet. In *PNAS*, 2005.

[83] R. Durairajan, S. Ghosh, X. Tang, P. Barford, and B. Eriksson. Internet atlas: A geographic database of the internet. In *ACM HotPlanet*, 2013.

[84] B. Eriksson, R. Durairajan, and P. Barford. RiskRoute: A Framework for Mitigating Network Outage Threats. In *ACM CoNEXT*, 2013.

[85] E. Glatz and X. Dimitropoulos. Classifying Internet One-way Traffic. In *ACM IMC*, 2012.

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

[87] S. P. Gorman, L. Schintler, R. Kulkarni, and R. Stough. The Revenge of Distance: Vulnerability Analysis of Critical Information Infrastructure. *JCCM*, 2004.

[88] K. P. Gummadi, H. V. Madhyastha, S. D. Gribble, H. M. Levy, and D. Wetherall. Improving the Reliability of Internet Paths with One-hop Source Routing. In *USENIX OSDI*, 2004.

[89] A. F. Hansen, A. Kvalbein, T. Cicic, and S. Gjessing. Resilient Routing Layers for Network Disaster Planning. In *IEEE ICN*, 2005.

[90] P. E. Heegaard and K. S. Trivedi. Network Survivability Modeling. 2009.

[91] 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.

[92] B. Huffaker, M. Fomenkov, and k. claffy. DRoP:DNS-based Router Positioning. *SIGCOMM CCR*, 2014.

[93] K. Kant and C. Deccio. Security and Robustness in the Internet Infrastructure.

[94] E. Katz-Bassett, H. V. Madhyastha, J. P. John, A. Krishnamurthy, D. Wetherall, and T. Anderson. Studying Black Holes in the Internet with Hubble. In *USENIX NSDI*, 2008.

[95] E. Katz-Bassett, C. Scott, D. R. Choffnes, I. Cunha, V. Valancius, N. Feamster, H. V. Madhyastha, T. E. Anderson, and A. Krishnamurthy. LIFEGUARD: Practical Repair of Persistent Route Failures. In *ACM SIGCOMM*, 2012.

[96] A. Lakhina, J. Byers, M. Crovella, and I. Matta. On the Geographic Location of Internet Resources. In *IEEE JSAC*, 2003.

[97] E. J. Malecki. The economic geography of the internet's infrastructure. *Economic geography*, 2002.

[98] L. Quan, J. Heidemann, and Y. Pradkin. Detecting Internet Outages with Precise Active Probing (extended). In *USC Technical Report*, 2012.

[99] M. A. Sanchez, F. E. Bustamante, B. Krishnamurthy, W. Willinger, G. Smaragdakis, and J. Erman. Inter-domain traffic estimation for the outsider. In *ACM IMC*, 2014.

[100] A. Singla, B. Chandrasekaran, P. B. Godfrey, and B. Maggs. The internet at the speed of light. In *ACM Hotnets*, 2014.

[101] J. Sommers, P. Barford, and B. Eriksson. On the Prevalence and Characteristics of MPLS Deployments in the Open Internet. In *ACM IMC*, 2011.

[102] H. Wang, Y. R. Yang, P. H. Liu, J. Wang, A. Gerber, and A. Greenberg. Reliability as an Interdomain Service. In *ACM SIGCOMM*, 2007.

[103] B. M. Waxman. Routing of multipoint connections. *IEEE JSAC*, 1988.

[104] W. Willinger and J. Doyle. Robustness and the internet: Design and evolution. *Robust-Design: A Repertoire of Biological, Ecological, and Engineering Case Studies*, 2002.

[105] J. Wu, Y. Zhang, Z. M. Mao, and K. G. Shin. Internet Routing Resilience to Failures: Analysis and Implications. In *ACM CoNEXT*, 2007.

[106] L. Zhou. Vulnerability Analysis of the Physical Part of the Internet. In *IJCI*, 2010.

[107] Y. Zhu, A. Bavier, N. Feamster, S. Rangarajan, and J. Rexford. UFO: A Resilient Layered Routing Architecture. *SIGCOMM CCR*, 2008.