Characteristics of Metro Fiber Deployments in the US

Sathiya Kumaran Mani*, Matthew Nance Hall†, Ramakrishnan Durairajan† and Paul Barford‡
*Microsoft, †University of Oregon, ‡University of Wisconsin—Madison

***Abstract*—Optical fiber deployments in metropolitan areas are critical for information distribution to businesses and large segments of the population. In this paper, we describe a characterization study of metropolitan area fiber networks in the US. The goal of our work is to elucidate the key aspects of these infrastructures and to assess how they can be enhanced to support growth in cloud-mobile via expanded connectivity to data centers. We collect maps of 204 metro fiber networks and transcribe these into a geographic information system for analysis and visualization. We report on characteristics including raw miles, geography, proximity to users, correspondence to other infrastructure and PoP/data center proximity. These characteristics indicate highly diverse deployments in different metro areas and suggest different strategies for future deployments. Next, we conduct a resource allocation analysis to assess how fiber infrastructure can be deployed in metro areas to reduce the physical distance to data centers over a range of cost scenarios. Our results show that a small number of new connections to data centers can significantly reduce physical distances to users.**

I. Introduction

The physical infrastructure that is the foundation for carrying traffic in the Internet is comprised of links (*e.g.,* fiber conduits, twisted pair copper, coax, etc.) and nodes (*e.g.,* points of presence (PoP), colocation centers, data centers, etc.) that are distributed around the globe. It is natural to think about physical Internet infrastructure simply as a large graph that carries traffic via distributed routing protocols. While this simple model may be sufficient for certain types of research studies, it belies the complexity of the details of the actual infrastructure that are critical for assessing issues such as performance, robustness, manageability, and opportunities for deployment of new technologies.

One way to taxonomize physical Internet infrastructure is by its geographic footprint and how that footprint relates to data sources (*e.g.,* data centers) and destinations (*e.g.,* user populations) connected by the infrastructure. Specifically, Internet infrastructure can be considered by *(i)* components that carry traffic over long distances between population centers (*e.g.,* submarine and long haul), *(ii)* components that carry traffic over moderate distances around population centers (*e.g.,* metropolitan) and *(iii)* components that carry traffic over short distances (last mile) to individual users (*e.g.,* broadband, cable and cellular). Prior work has focused on analyzing and characterizing certain aspects of the Internet's physical infrastructure including the long-haul fiber [1] and submarine cables [2], peering points [3], cross connects [4], cell towers [5], and interconnections [6], [7], [8]. These efforts provide important perspectives. However, to the best of our knowledge, the characteristics of the *metropolitan area (metro)* physical Internet infrastructure have not been investigated.

In this paper, we present a study of metro fiber deployments. The goal of our work is to elucidate the characteristics of these infrastructures and to utilize these insights to assess potential enhancements. Our particular focus is on enhancements for cloud-related technology. We seek to answer questions such as, what is the relative density of deployments in different metro areas? how do metro area deployments correspond to user populations? and how might cloud/mobile connectivity be improved by considering metro fiber deployments?

The primary challenge in studying physical Internet infrastructure is assembling a representative data corpus of operational networks since service providers are often hesitant to provide data for competitive and security reasons. To conduct our study, we follow the method described in [9]. Specifically, we use search to identify maps of 204 metro fiber infrastructures in the continental US.[1] We then transcribe these maps by hand into a standard representation that enables visualization and analysis via Geographic Information System (GIS). In particular, we use the GIS-based representation to relate the metro fiber maps to geographic characteristics and infrastructures (*e.g.,* datacenters, roads, etc.) in each area.

Our analysis of metro fiber networks reveals a variety of characteristics that have implications on serviceability, performance, robustness, future growth and competitiveness. We find a close correspondence between fiber deployments and roadways in all of the metro areas. We show that a handful of metro areas have a relatively large number of fiber providers (25 out of 204 areas have 10 or more different fiber providers), while others are much more limited. We also find that some metro areas with relatively low population density have a 3 to 4 times more deployed fiber than metro areas with similar population density. These characteristics and others included in our analysis highlight the diversity of current metro fiber deployments and suggest different strategies for future deployments in each area.

Using the maps and insights gained from our characterization study, we consider how cloud infrastructure *e.g.,* via third-party cloud connectivity providers [11], [12], [13] can be enhanced such that the physical distances between users and data centers are minimized. We consider this optimal node

*The work was completed while the author was at the University of Wisconsin—Madison.

[1]For example, a good source of metro fiber maps can be found at [10].

978-3-903176-27-0 © 2020 IFIP

placement problem over a range of scenarios in the New York, Newark, Jersey City metro area. Our results show that even a small number of new connections to data centers can significantly reduce physical distances to users, which has important implications for service performance and moving traffic closer to the end users.

To summarize, the research contributions of this paper include the first characterization study of metro fiber infrastructure deployments in the US. We also show how physical distances between users and data centers can be improved in metro areas via third-party cloud connectivity providers. While this paper represents an early milestone, we believe that the practical nature of our results can inform future deployments and future analyses of metro area networks. Finally, we embrace the importance of reproducibility in research. To that end and to facilitate future work, we will make a selection of the metro fiber maps that we assembled for this study available to the community in [14].

II. Datasets

We generated a list of metro fiber providers using Google search with terms such as "metro fiber maps", "metro fiber in manhattan", etc. We manually inspected provider sites to identify maps of their fiber deployments. We also utilize alternative data sources (*e.g.,* Telecom Ramblings [10]) to fill in details where available, however the vast majority of the maps come from service provider websites, which we consider to be highly credible sources. These efforts enabled us to identify 204 metro fiber infrastructures in the continental US. These maps show only the *physical link structure* of metro fiber deployments.[2] 368 metropolitan statistical areas (MSAs) are represented in the 204 maps since many maps span multiple MSAs and even states. MSAs are discrete geographic regions defined by the US Census Bureau that reflect core population areas of at least 50,000 people and the surrounding communities with strong economic and social integration [15].

Multiple providers have infrastructure in most MSAs. Sources for maps include [10], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Maps appear in a variety of formats from simple images (*e.g.,* png or jpg) to KML[3] files [10], [17]. If a KML file is available we import that map directly into ArcGIS (a widely-used platform for visualizing and analyzing spatial data). If a KML file is not available, such as in [20], then we transcribe the image into the GIS *shapefile* format by hand. After we collected all of the maps, we organized them by the provider and added them to a common GIS layer file. The final geocoded version of all of the 204 maps is shown in Figure 1.

It is important to be clear about how metro fiber fits into the larger physical network infrastructure that comprises the Internet. Durairajan et al. define a *long haul fiber 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" [1]. Our definition of *metro fiber* is primarily geographic *i.e.,* links that are confined to an MSA and that do not provide connectivity between MSAs. Long haul links connect with metro fiber links to enable distribution in metropolitan areas. Metro fiber also connects to broadband infrastructure to enable last mile connectivity to users who reside in metropolitan areas.

We argue that this dataset is of sufficient size to provide a valuable perspective on metro fiber deployments. Our 204 maps cover 368 out of the 381 MSAs in all of the lower 48 states in the US (which exclude Alaska and Hawaii). These 381 MSAs include over 80% of the US population. However, we do not argue that our maps show *all* deployed metro fiber in these MSAs (since there is no objective source of ground truth for total fiber deployments) nor can we assure the objective accuracy of our maps. We took great care in the transcription process — including a complete, after-the-fact auditing step to ensure exact shapefile correspondence to the original maps. During the audit step, we also updated the original maps with new data if available. Thus, the results reported in the following sections are accurate reflections of US metro fiber to the extent that the base maps from service providers and other sources are accurate.

We include a detailed survey of some of our largest data sources in Tables IV and V in the Appendix. All of the GIS shapefiles for the maps listed the Appendix plus a set of analysis scripts for ArcGIS will be available in [14] [4]. These enable reproduction of the key results in this paper for the top 10 most populated MSAs as listed in Table IV. We note that 48 out of the 49 maps indicated in Table V were gathered directly from service provider websites.

Augmenting fiber maps with node locations. To consider the issue of future data center (DC) infrastructure/connectivity deployments, we used search and data provided by Internet Atlas [9] to identify the *(i)* DCs (1,184), *(ii)* colocation facilities or colos (10) and *(iii)* PoPs (6,224) in the 368 MSAs represented in our maps. It is often the case that many PoPs are located in the same physical location/building. Similarly, micro data centers [5] can be colocated but actual data centers (*e.g.,* those operated by large cloud providers) cannot. However, some DC locations in our data do overlap because of the lack of specificity in the source data. Specifically, during the data collection process, we found that some service providers generically state that they have a PoP/colo/DC in a city without specifying the actual infrastructure type or its location. In such cases, we augmented our maps to note such features. If no specific location is given, we map the node to the city center but omit those nodes from further analysis. We also omit colos from analysis because so few are represented as compared to PoPs and data centers.

Basic features. The maps in our dataset represent ∼505K

[2] For convenience this paper we use the terms "fiber", "metro fiber" and "fiber link" when in reality what is represented in the maps is *metro fiber conduit i.e.,* conduit that contains multiple individual strands of optical fiber. To that end, we have identified and removed all duplicate fiber links in our datasets.

[3] KML, or Keyhole Markup Language, is a common encoding of geographic features into a file that can be used to render maps [25].

[4] Navigate to University of Wisconsin repository on the IMPACT website. We also include a file listing the URLs for all maps used in this study.

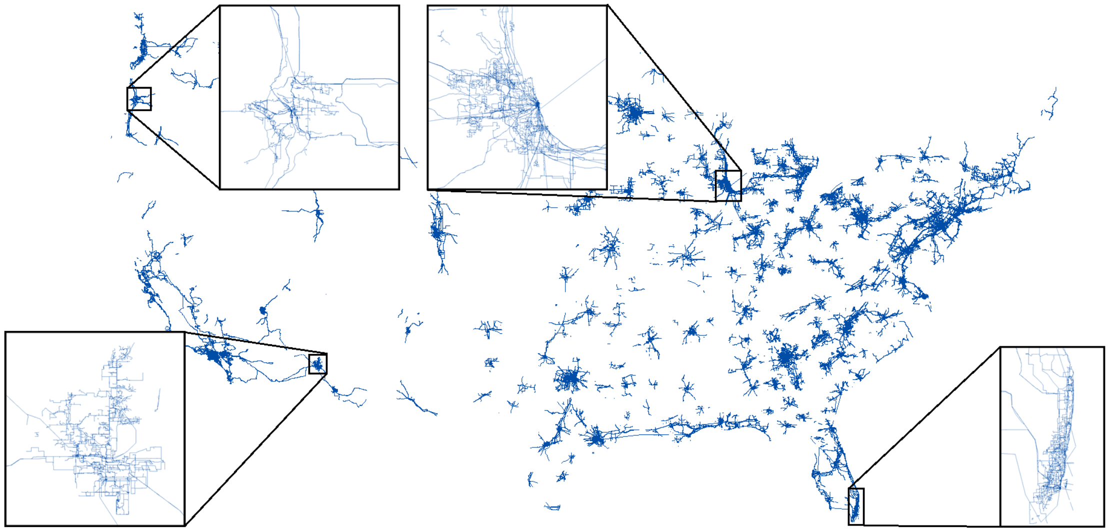

Fig. 1: Maps of the 204 metro fiber networks in the continental US used in our study. Magnified areas shown include Portland, Chicago, Phoenix, and Miami.

miles of fiber in US metropolitan areas. The most immediate takeaway from visualizing the maps is that metro fiber deployments are complex and diverse. As seen in Figure 1, there is rich connectivity among areas on the eastern seaboard and extending into the midwest. Deployments are more sparse on the west coast, with archipelago structures observable in key population areas such as Los Angeles, the San Francisco Bay area, and Seattle. There is a paucity of metro fiber maps in the central plains and rocky mountain region between the west coast and the midwest. These characteristics reflect the nature of population centers in the US, and also highlight areas where there may be future growth in metro connectivity.

Correspondence to roadways. One of the key benefits of the GIS-based approach that we use in this study is the ability to consider other data sets including transportation systems (*e.g.,* roadways, railways, etc.), population distributions and geographic features (lakes, rivers, mountains, etc.). As shown in [1], long haul infrastructure often follows road and rail infrastructure. We conducted an analysis for a similar correspondence in our metro fiber data. We found that nearly all of the metro fiber deployments in our data corpus indicate a very close correspondence to roadways in particular and almost never follow a path separated from other infrastructure. This characteristic can be explained by the fact that roads are a standard right of way (therefore do not require complex permitting) and facilitate maintenance and repair.

III. Deployment Analysis

In this section, we report on the basic characteristics of fiber deployment density in MSAs and how these relate to the populations that they serve in those areas. The goal of these analyses is to establish quantitative comparisons between MSAs as a means for assessing their ability to serve their populations and support future growth.

A. Population centers

Figure 2 shows the correlation between fiber coverage and population density for 368 MSAs. We can see that San Francisco-Oakland-Hayward, CA and Trenton, NJ have relatively low fiber coverage despite being among the MSAs with the highest population density. We also see that the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD MSA has the highest number of *fiber miles*—which we define as the total length of fiber links—while falling in the mid-range of all areas for population density.

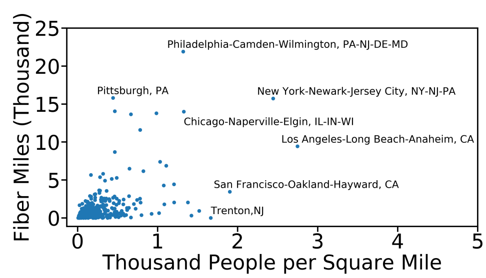

Fig. 2: Population density and fiber coverage for all US Metropolitan Statistical Areas in the lower 48 states.

B. Provider-specific characteristics

The online sources for our corpus of maps indicate ownership by 72 different service providers. These providers range from tier-1 and tier-2 providers (*e.g.,* Zayo [17] and Segra [22]) to local service providers such as Massachusetts Broadband [21] to regional dark fiber leasing companies such as Maine Fiber Company [23]. We make the following observations:

- Some metro areas have relatively low fiber representation, despite their relatively high population. For example, the Miami-Fort Lauderdale-West Palm Beach, FL area was estimated to have ∼6.1 million people as of 2016 [26], but only ∼4 thousand miles of metro fiber; one fifth the amount of fiber identified in Philadelphia-Camden-Wilmington, PA-NJ-DE-MD, an area with similar population and size (square miles of land). Interestingly, there are more service providers that own fiber in Miami (10) compared to Philadelphia (6), whose population is 3 times larger.

- Zayo has the largest presence across all areas in our study, with 290 different MSAs. Zayo is the sole fiber provider in 6 different MSAs, according to data we could procure. The provider with the second-highest presence is Windstream [24], with 266 different service areas across the US. Windstream was also the sole fiber presence in 5 distinct MSAs whose populations fell in the range of 90,000 to 200,000 people. These results may be an artifact of our sources although maps from all major service providers are included in our dataset.
- A handful of relatively small population centers have a high number of distinct providers present. For instance, the Youngstown-Warren-Boardman, OH-PA area has a population of roughly 500,000 people and has at least 14 distinct fiber providers.

Figure 3 shows the cumulative distribution of the number of fiber providers present among all 368 MSAs that we covered. We can see that the median number of providers is 5. Only 6.6% of MSAs have more than 10 different providers available. This has implications for competitiveness in metro areas.

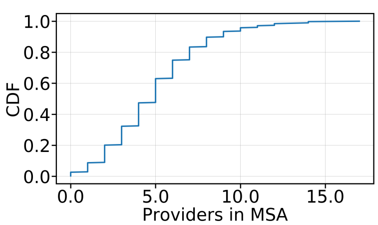

Fig. 3: CDF of the number of fiber providers in the US metropolitan areas.

C. *Fiber miles and population*

Using ArcGIS, we measure the fiber miles within each MSA. Then, we compare the availability/presence of fiber within these regions to the estimated number of people living there as of 2017 [26]. A summary of population and total fiber miles for the 10 most-populated MSAs is presented in Table I. The ratio of fiber miles per unit area within the MSAs is also shown. One of the interesting observations about these statistics is that the largest MSAs—for example, New York and Los Angeles—do not have as much fiber density as some of the smaller MSAs (0.0005 and 0.0003 miles per person respectively). Besides, we note that Philadelphia-Camden-Wilmington has the most fiber coverage of the MSAs while ranking 8^{th} in population, with 3.6 miles of fiber per thousand people (see Table I). While space constraints limit what we can report, we find diverse fiber density characteristics in MSAs beyond the top 10.

D. *Visualizing MSA fiber density*

While GIS is useful for certain types of analysis and visualization, we found that detailed assessment and characterization of street-level maps of metro fiber infrastructure was quite challenging. For instance, some fiber lines are collocated, (*e.g.,* two fiber lines might run along the same streets). Thus, we cannot easily see that the street has more fiber along with it from a glance at the raw map. We appealed to census block data from the US Census Bureau [27] to develop a means for quickly inferring characteristics of a fiber map. Census blocks partition MSAs into variable-sized *blocks* with size proportional to the population in the block. We measured the length of fiber (in miles) within each census block and produce a heat map of the MSA with colors corresponding to the length of fiber in a block. Unfortunately, we quickly realized that the variable size of census blocks also had undesirable impacts on the visualization: larger census blocks in the rural outskirts of an MSAs could contain long stretches of fiber which resulted in portraits that over-represented the coverage of fiber in large city blocks. Moreover, census blocks are densely packed within city centers, and thus we are not able to get a sense of the coverage of fiber coverage within these city centers.

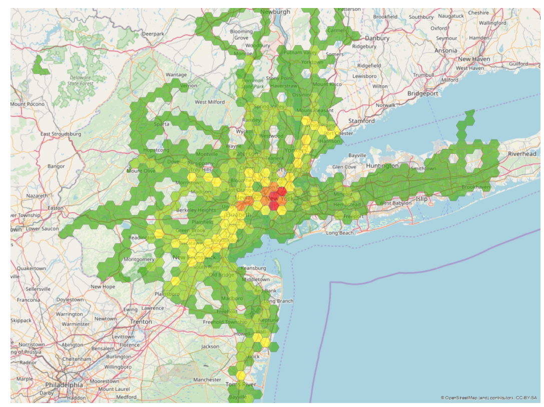

Fig. 4: Hexagonal tessellation-based heat map New York-Newark Jersey City MSA.

To visualize fiber coverage for MSAs effectively we need a technique that allows more coarse-grained representation than city blocks, with relatively similar size partitions. To this end, we create hexagonal tessellations over MSAs. We choose the size of hexagons based on the average census block size across the top ten MSAs, *i.e.,* 11 square miles. We posit that this view has several advantages. First, hexagonal tessellations allow quick inference of locations where fiber infrastructure deployments are dense. For example, in Figure 4, we see that the lower Manhattan area has the highest fiber density as inferred by the red hexagons. Second, hexagonal tessellations allow inference of paths or stretch with high fiber density. For instance, the area between Manhattan and New Brunswick, NJ appears to be a fiber artery (*i.e.,* a vital route) between the two districts as inferred by the yellow hexagons protruding from the lower Manhattan area.

Due to the visually appealing nature of the hexagonal tessellations, we use them as the basic unit from which we measure fine-grain statistics. For example, we use the center-point of hexagonal tessellations as a reference for measuring distance to DCs and PoPs across MSAs in § IV.

MSA	Population	Fiber Miles	Land Area ($miles^2$)	Fiber Density ($miles/miles^2$)
New York-Newark Jersey City, NY NJ PA	20,320,876	15,722	8,292	1.89
Los Angeles Long Beach Anaheim, CA	13,353,907	9,426	4,850	1.94
Chicago-Naperville-Elgin, IL IN WI	9,533,040	13,985	7,195	1.94
Dallas Fort Worth Arlington, TX	7,399,662	11,591	9,279	1.24
Houston The Woodlands Sugar Land, TX	6,892,427	6,171	8,265	0.74
Washington-Arlington-Alexandria, DC VA MD WV	6,216,589	13,773	6,246	2.20
Miami Fort Lauderdale West Palm Beach, FL	6,158,824	4,438	5,067	0.87
Philadelphia-Camden-Wilmington, PA NJ DE MD	6,096,120	21,895	4,603	4.75
Atlanta-Sandy Springs-Roswell, GA	5,884,736	13,708	8,685	1.57
Boston-Cambridge-Newton, MA NH	4,805,942	2,046	3,486	0.58

TABLE I: Metro fiber deployment summary statistics for top 10 most populated MSAs.

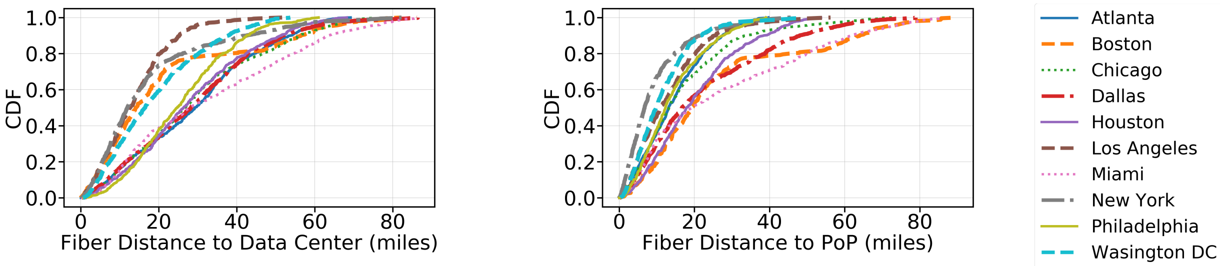

Fig. 5: Fiber distance to the nearest DC (left) and PoP (right) for the ten most populated MSAs.

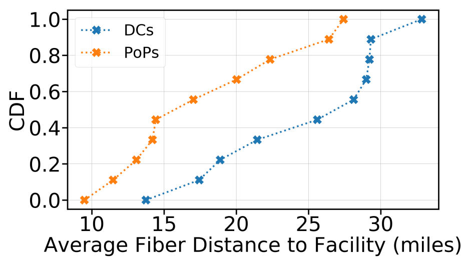

Fig. 6: Distribution of average fiber distance to the nearest DC and PoP for the ten most populated MSAs.

E. Distance metrics

Using our tessellation-based sampling, we quantify the fiber distance from any 11 sq. mile hexagon in an MSA to the nearest DC and PoP. We choose the top 10 most populated MSAs, as seen in table I, for this analysis. Figure 5 shows the distributions of these distances for the ten MSAs. We see that people in Los Angeles, New York, Boston, and Washington DC all have closer proximity to PoPs and DCs than their compatriots in Chicago, Dallas, Houston, Miami, Philadelphia, and Atlanta.[5]. The averages of these distributions are summarized in Figure 6, which indicates an average proximity of 10 to 27 miles for PoPs and 14 to 32 miles for DCs in these ten MSAs.

F. Ranking MSAs

We develop a ranking metric to compare metropolitan area fiber deployments, and to highlight areas where metro fiber connectivity might be improved. Our ranking metric considers three characteristics, *(i)* population (p), *(ii)* fiber miles (m), and *(iii)* land area (square miles, m^2). Having more fiber in the ground increases the score of an MSA, while more population and larger land areas decreases the score. Figure 7 shows the relationship between fiber availability (fiber miles per thousand persons, m/p) and fiber density (fiber miles per square mile, m/m^2) for all MSAs. We can see that most metro fiber deployments have 0 to 2 fiber miles per square mile, and also have between 0 and 6 miles of fiber per thousand people.

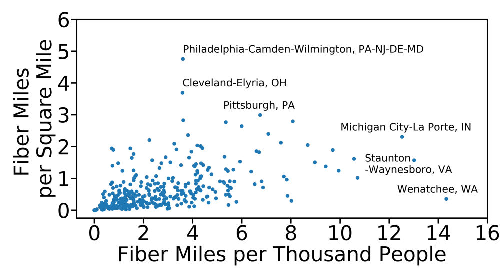

Fig. 7: Scatter plot of fiber availability and fiber density across all MSAs. Points in the upper-right corner have better fiber coverage.

The three MSAs with the highest fiber density or availability are annotated in the figure 7. Philadelphia-Camden-Willmington, PA-NJ-DE has the highest fiber density, with 4.75 m/m^2, followed by Cleveland-Elyria, OH, with 3.69 m/m^2, and Pittsburg, PA, with 2.99 m/m^2. The MSA with the highest fiber availability was Wenatchee, WA (a very small MSA, with a population of 99K), with 14.3 m/p. It was followed by Staunton-Waynesboro, VA (population of 118K), with 13.0 m/p. Finally, Michigan City-La Porte, IN (population of 110K) had 12.5 m/p.

Our ranking metric is calculated as follows. Let α be fiber availability. $\alpha = m/p$, where m is the total fiber miles of an MSA, and p is the population. Now, let ρ be the fiber density. $\rho = m/m^2$, where m^2 is the area of an MSA (Note: We only take into account land area. Area over water features such as lakes, rivers, or oceans do not affect our fiber density). Our metric M is the sum of fiber availability and fiber density. We reduce the range of M to [0,1) with the hyperbolic tangent function, thus $M = tanh(\alpha + \rho)$. Table II shows the statistics

[5]This data refers to the entire MSA, although only the first city-name of the MSA is given here and in the legend of figure 5 for brevity.

MSA	Rank	Fiber Miles	Population	Land Area	Total Providers
Philadelphia-Camden-Wilmington, PA NJ DE MD	0.999880	21,895	6,096,120	4603	6
Cleveland-Elyria, OH	0.999017	7,385	2,058,844	1999	11
Trenton, NJ	0.998920	819	374,733	224	6
Pittsburgh, PA	0.996033	15,803	2,333,367	5281	11
Akron, OH	0.994431	2,543	703,505	900	11
Erie, PA	0.994121	2,232	274,541	799	8
Ann Arbor, MI	0.993746	1,952	367,627	705	8
Elkhart-Goshen, IN	0.992163	1,224	85,557	463	9
South Bend-Mishawaka, IN MI	0.986987	2,272	440,933	947	12
Santa Cruz-Watsonville, CA	0.985706	1,050	3,337,685	445	4

TABLE II: Metro fiber deployment summary statistics for top 10 ranked MSAs.

for 10 highest-ranked MSAs. The rankings highlight the diverse deployment characteristics of metro fiber. The top 10 list contains 4 large MSAs with populations over 2M, but it also includes a very small MSA (Elkhart-Goshen, IN), with a population of about 85K. Similarly, land areas and fiber miles vary by a factor of 20 between largest and smallest.

G. Sparse metro fiber deployments

In this section, we focus on the Riverside-San Bernardino-Ontario metro in California and Beckley metro in West Virginia as examples of sparse metro fiber deployments in contrast to dense deployments such as the New York MSA. Based on our fiber rank metric from § III-F Riverside has a low fiber rank (0.2) considering population (4.5M), fiber miles (5.6K), and geographic area (27K square miles). The heatmap in Figure 8 shows that the low rank is primarily due to the large area the metro covers, much of which is uninhabited park land, or federal land (Mojave National Preserve/wilderness, and Joshua Tree National park, as well as multiple military bases). Much of the fiber is concentrated downtown, but the thin arms protruding to the north and west show potential vulnerability to physical damage/disruption for the areas connecting Riverside to Phoenix and Las Vegas.

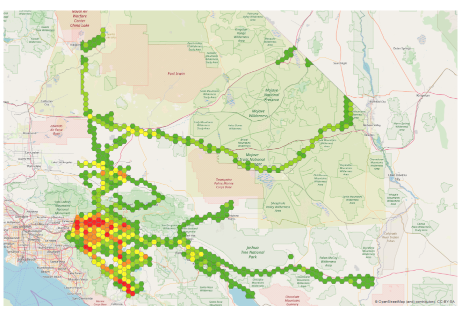

Fig. 8: Sparse Fiber Deployment of Riverside-San Bernardino-Ontario metropolitan statistical area.

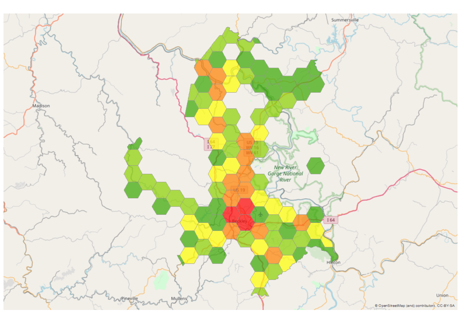

Fig. 9: Sparse Fiber Deployment of Beckely, West Virgina

In contrast to Riverside and the New York metro areas, Beckley, WV is more rural (Figure 9). Its population is 120K and its land area is about 1.2K square miles. While Beckley is smaller than many other MSAs, its fiber density per person is one of the highest in our study, with approximately 10 miles of fiber per thousand people person.

H. Spatial autocorrelation analysis

Next, we analyze the spatial autocorrelation of fiber deployments in MSAs using Moran's I statistics [37], which is a standard technique for assessing patterns in complex spatial data sets. Our goal is to compare and contrast different metro fiber deployments quantitatively. The first step in this analysis is to determine the appropriate distance threshold to be used for Moran's I. This begins with an initial distance threshold and tests incrementally higher values to determine a distance threshold that would provide peak Z-Scores. After determining the distance threshold, we use this threshold value to perform cluster and outlier analysis (Anselin Local Moran's I), to determine spatial auto-correlation of the fiber infrastructure for a given MSA. The goal of our Moran's I analysis is to categorize the fiber infrastructure into segments that show high autocorrelation vs. segments that show low autocorrelation. We posit that segments that show high autocorrelation indicate infrastructure that is likely to have been carefully designed over a larger area to serve the needs of an existing and/or projected population, while low autocorrelation is likely to indicate infrastructure that has been deployed incrementally in smaller areas or areas of unanticipated growth. Table III lists the percentage of the fiber infrastructure assigned to each category by Moran's I analysis for a few MSAs. These MSAs are selected randomly to represent a mixture of both dense and sparse MSA fiber deployments.

This analysis provides a practical and concise means for describing and comparing deployment styles within and between different metro areas. High spatial autocorrelation represents regions of interest where the infrastructure is highly clustered, which in turn could be used to identify shared risk when multiple providers are deploying fiber along the same path. For example, from Figure 10 (highlighting the downtown region), we see two different deployment styles between Miami and Portland MSAs. For Miami, we see that the infrastructure exhibiting high spatial autocorrelation is found on the outer edges of the MSA along the south-eastern coastline. This contrasts with Portland, where the infrastructure with high spatial autocorrelation is found in the center of the MSA in the

Fig. 10: A map of the downtown region for Miami (left) and Portland (right) MSAs showing fiber infrastructure labeled into various clustering regions by Moran's I analysis.

MSA (region)	High-High	Low-Low	High-Low	Low-High	Not Significant
Miami, FL	26.22	4.54	4.13	0.49	64.62
Portland, OR	34.4	8.23	9.59	0.07	47.7
Riverside, OR	25.6	0.68	0.77	0.16	72.79
Beckley, WV	21.42	2.21	1.35	0	75.02

TABLE III: Percentage of fiber infrastructure assigned to each Moran's I clustering category for various MSAs. High-High and Low-Low represent infrastructure with high and low spatial autocorrelation respectively, which helps identify areas of interest (*i.e.,* structured connectivity). High-Low and Low-High represent infrastructure categorized as outliers (*i.e.,* random connectivity) by Moran's I.

downtown area. Exploring methods to identify characteristics such as robustness, performance or accessibility using the spatial autocorrelation analysis is a focus of our on-going work.

IV. Enhanced Cloud Connectivity

In this section, we present a case study that considers how users can be connected to data centers (DC) via metro fiber. Our general objective is to demonstrate how a detailed understanding of metro fiber deployment can be applied to problems of interest. This particular case study is motivated by the emerging Micro-DC (MDC) offerings and cloud providers who are actively seeking partnerships within PoPs to bring their services and traffic closer to end users [11], [12], [13]. MDCs are growing in popularity as the line between cloud and transit networks has begun to blur [28]. We anticipate that reducing fiber miles to DCs will have a positive impact on how traffic is carried using metro fiber infrastructures.

We begin by characterizing the metro fiber distance (physical distance over the fiber network infrastructure) between users and existing DCs in an urban city center. Then, we posit that the physical distance between users and DCs can be reduced by introducing new MDCs that could be co-located within existing PoP locations identified in our maps. There are many resource allocation scenarios that could likewise be considered (*e.g.,* optimal placement of entirely new fiber or DC infrastructure), and we plan to investigate these in future work.

Methodology. We use the fiber map shown in Figure 4, and zoom into a 60 sq. mile region around Manhattan island, NYC to characterize user proximity to DCs. According to our data, this particular region houses at least 69 DCs, 189 PoPs, and 1,500 miles of optical fiber. We use the physical fiber routes, the DC locations, and 0.5 sq. mile hexagonal tessellations[6] over this region to calculate the fiber distance from the centroid of each of the hexagons in the tessellation (points of demand or PoDs) to its closest DC. Given the characterization of PoDs to DCs, we next consider an optimal node placement problem. The objective is to add new nodes (MDCs) to the network such that the sum of distances from each PoD to its nearest DC is minimized. We choose candidate MDCs from among the existing PoPs, reasoning that MDC deployment can benefits directly from a PoP's fiber connectivity.

Results. Figure 11 (left) characterizes the initial infrastructure with respect to DCs and PoDs. The lines extending from the DCs terminate at PoDs for whom fiber distance is minimized. Several PoDs within the region do not connect to any DC; this is due to the fact that fiber in the hexagons for those PoDs is disconnected from the other metro fiber in the area. The lines shown are line-of-sight. However, the actual distance measured is calculated based on the distance through the fiber network. We found that ∼15% of the Manhattan area is more than 4 miles away from the closest existing DC.

Figure 11 (left) shows fiber distances to DCs in the original map. We see that for some PoDs, the distance to the nearest DC is quite far (*e.g.,* points in the upper-right of Figure 11 (left)). Introducing one new MDC reduces the maximum distance from 9 miles to 7. This new MDC also improves distance for 189 PoDs in total, as seen in Figure 11 (center). Now, as more MDCs are added (five in Figure 11 (right)) the figures look less and less "cactus-like", which indicates that connectivity

[6]Manhattan city blocks are roughly 0.5 sq. miles.

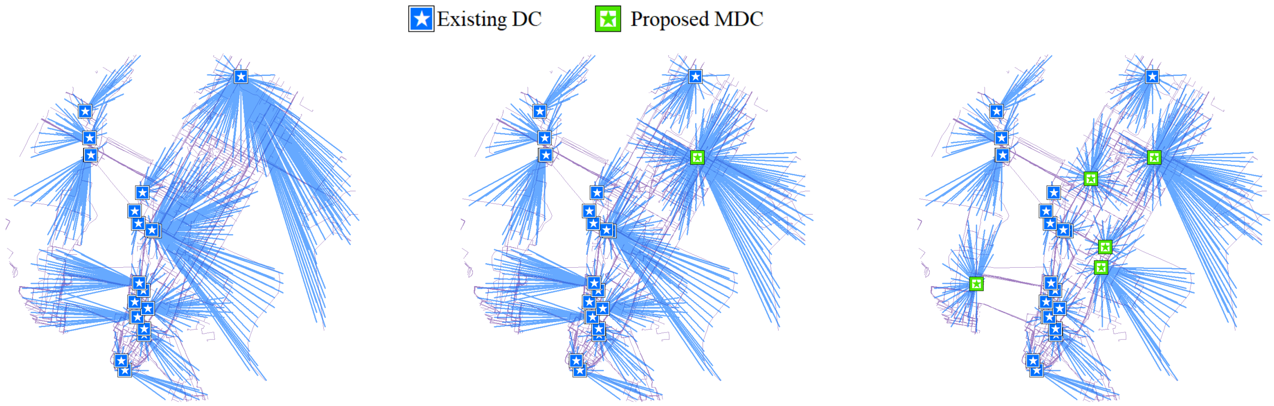

Fig. 11: Line-of-sight to the nearest DC, over the fiber network, for users in the Manhattan area (left) after the addition of one MDC (center) and five MDCs (right). Locations for new MDCs are restricted to being co-located with existing PoPs.

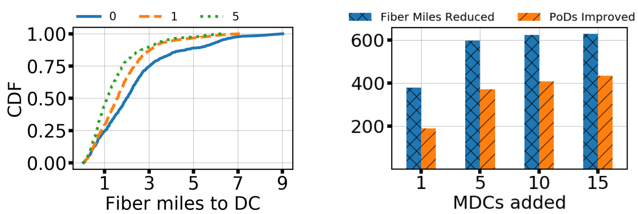

Fig. 12: (left) CDF of fiber distances of user locations in Manhattan to their closest cloud infrastructure node. Each line indicates the number of additional MDC connections in the area (thus, the baseline of no new MDCs is the blue line). (right) Total number of fiber miles reduced for all points of demand (PoDs) which benefit from the new cloud infrastructure node (dark bars with double hatching), and the total number of PoDs receiving a benefit (light bars with single hatching).

to DCs for users is improving. Figure 12 (left) shows that the introduction of just one MDC leads to a decrease in the 80th percentile fiber distance from 3.5 to 2.5 miles. Figure 12 (right) shows that the number of PoDs for whom fiber distance is reduced by the addition of new MDCs. It appears that a point of diminishing returns is met after introducing five MDCs, as the fiber miles reduced between five and ten new locations is negligible. The key result of this analysis is that cloud connectivity can be greatly improved by only adding a small number of new MDCs. Results for other MSAs vary and are omitted due to space constraints.

V. Related Work

Mapping and characterizing the Internet's physical infrastructure and connectivity offers the opportunity for insights on performance, robustness, manageability, and vulnerabilities, and has been of interest to the community for some time. Notable efforts include mapping connectivity at the router level [29], [30], [31], [32], [33], PoPs [34], [35], [36], and Autonomous Systems (ASes) [37], [38] connectivity. 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 [39], [40].

Mapping efforts from a physical deployment perspective include Internet Atlas [9] and Internet Topology Zoo [41]. Our work is related to and complements the study by Durairajan et al., which uses maps of physical deployments to characterize *long haul fiber infrastructure* in the US [1]. The key differences with that work are the fact that we focus on *metro fiber infrastructure* that has a limited geographic focus and that metro fiber networks are inherently more complex in terms of interconnection characteristics leading to our tessellation-based approach for analysis.

Prior studies have also examined interconnections between networks [6], [7], [8]. Our effort is also related to mobile and edge computing (MEC) efforts. The predominant focus of the efforts in MEC realm is on resource management and allocation [42], [43], load balancing [44], network architectures [45], virtualization [46], and deployment characterization [5], [42], [47], [48], [49]. To the best of our knowledge, *ours is the first study to elucidate metro-fiber connectivity in the US.*

VI. Summary

In this paper we present an analysis of metro fiber network infrastructures in the US. The goal of our work is to elucidate the characteristics of metro fiber networks and to demonstrate how more detailed knowledge of these infrastructures can serve as a foundation for future deployments and upgraded technologies. We collected 204 metro fiber maps using search and transformed these into a common representation that can be utilized in geographic information systems for visualization and analysis. We find that similar to long haul infrastructure, metro fiber follows standard rights of way, typically roadways, and is most densely deployed in or near the geographic center of urban areas. We find diverse characteristics in terms of fiber vs. population density, and that larger metro areas typically have lower density per person than smaller areas. We use the fiber map for the New York/New Jersey MSA to conduct a case study that focus on deployments that improve cloud connectivity. In on-going work, we are focused on assessing related issues including full end-to-end connectivity (long haul, metro and last mile infrastructures), risks to metro fiber and manageability of infrastructure in metro areas.

Acknowledgements

The authors thank our shepherd Ignacio Castro for his input. This work is supported by NSF CNS-1703592, DHS BAA 11-01, AFRL FA8750-12-2-0328, NSF CNS-1850297, UO faculty research award, and a Ripple fellowship. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF, DHS, AFRL, UO, Ripple, or the U.S. Government.

References

[1] R. Durairajan, P. Barford, J. Sommers, and W. Willinger, "InterTubes: A Study of the US Long-haul Fiber-optic Infrastructure," in *proceedings of ACM SIGCOMM*, 2015.

[2] Z. S. Bischof, R. Fontugne, and F. E. Bustamante, "Untangling the world-wide mesh of undersea cables," in *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*, 2018.

[3] V. Giotsas, G. Smaragdakis, B. Huffaker, M. Luckie *et al.*, "Mapping peering interconnections to a facility," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. ACM, 2015, p. 37.

[4] R. Motamedi, B. Chandrasekaran, B. Maggs, R. Rejaie, and W. Willinger, "On the geography of x-connects," Tech. Rep.

[5] M. Syamkumar, P. Barford, and R. Durairajan, "Deployment characteristics of the edge in mobile edge computing," in *Proceedings of the 2018 Workshop on Mobile Edge Communications*. ACM, 2018, pp. 43–49.

[6] A. Marder and J. M. Smith, "MAP-IT: Multipass Accurate Passive Inferences from Traceroute," in *IMC*. ACM, 2016.

[7] M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark *et al.*, "bdrmap: Inference of Borders Between IP Networks," in *IMC*. ACM, 2016.

[8] M. Alexander, M. Luckie, A. Dhamdhere, B. Huffaker, C. KC, and S. M. Jonathan, "Pushing the boundaries with bdrmapit: Mapping router ownership at internet scale," in *IMC*, ACM. ACM, 2018.

[9] R. Durairajan, S. Ghosh, X. Tang, P. Barford, and B. Eriksson, "Internet atlas: a geographic database of the internet," in *5th ACM workshop on HotPlanet*. ACM, 2013.

[10] "Telecom Ramblings: Metro Fiber Maps," https://www.telecomramblings.com/metro-fiber-maps/.

[11] "Expressroute partners and peering locations," https://docs.microsoft.com/en-us/azure/expressroute/expressroute-locations.

[12] "Google supported service providers," https://cloud.google.com/interconnect/docs/concepts/service-providers.

[13] "Aws direct connect partners," https://aws.amazon.com/directconnect/partners/.

[14] "DHS Impact Information Marketplace," https://www.impactcybertrust.org/.

[15] "Metropolitan and Micropolitan," https://www.census.gov/programs-surveys/metro-micro/about.html, 2018.

[16] "Metro Fiber Maps: Pacific Northwest," https://www.telecomramblings.com/metro-fiber-maps/pacific-northwest/.

[17] "Zayo building lists and kmz files," https://www.zayo.com/solutions/global-network/building-lists-kmz-files/.

[18] "Crown Castle Fiber Map," https://fiber.crowncastle.com/our-network.

[19] "Columbus FiberNet," http://columbusfiber.net/google-map.php.

[20] "Fiber Light: Atlanta Metro Area ," (Link).

[21] "Massachusetts broadband institute: Map gallery," https://broadband.masstech.org/news-and-resources/map-gallery.

[22] "Segra network," https://www.segra.com/network/.

[23] "Maine fiber company map," http://mainefiberco.com/the-network/maps/.

[24] "Windstream network map," https://www.windstreamenterprise.com/wholesale/interactive-map/.

[25] "Keyhole markup language," https://developers.google.com/kml/.

[26] "Annual Estimates of Resident Population," https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2018/GCTPEPANNR.US24PR, 2019.

[27] "Census Blocks and Block Groups," https://www2.census.gov/geo/pdfs/reference/GARM/Ch11GARM.pdf, 2018.

[28] B. Yeganeh, R. Durairajan, R. Rejaie, and W. Willinger, "How cloud traffic goes hiding: A study of amazon's peering fabric," in *Proceedings of the Internet Measurement Conference*, 2019, pp. 202–216.

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

[30] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane: An Information Plane for Distributed Services," in *USENIX OSDI*, 2006.

[31] R. Sherwood and N. Spring, "Touring the Internet in a TCP Sidecar," in *ACM SIGCOMM IMC*, 2006.

[32] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding Traceroute Anomalies with Paris Traceroute," in *ACM SIGCOMM IMC*, 2006.

[33] B. Augustin, B. Krishnamurthy, and W. Willinger, "IXPs: Mapped?" in *ACM SIGCOMM IMC*, 2009.

[34] Y. Shavitt and E. Shir, "DIMES: Let the Internet Measure Itself," in *ACM SIGCOMM CCR*, 2005.

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

[36] R. Sherwood, A. Bender, and N. Spring, "Discarte: A Disjunctive Internet Cartographer," in *ACM SIGCOMM*, 2008.

[37] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-Level Topology," in *ACM SIGCOMM CCR*, 2005.

[38] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an Accurate AS-Level Traceroute Tool," in *ACM SIGCOMM*, 2003.

[39] D. Meyer, "Routeviews project," 2002.

[40] "RIPE Atlas," https://atlas.ripe.net, 2015.

[41] S. Knight, H. X. Nguyen, N. Falkner, R. A. Bowden, and M. Roughan, "The Internet Topology Zoo," in *IEEE JSAC*, 2011.

[42] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," 2017.

[43] A. S. Prasad, M. Arumaithurai, D. Koll, and X. Fu, "Raera: A robust auctioning approach for edge resource allocation," in *Workshop on Mobile Edge Communications*, 2017.

[44] Y. Yu, X. Li, and C. Qian, "Sdlb: A scalable and dynamic software load balancer for fog and mobile edge computing," in *Workshop on Mobile Edge Communications*. ACM, 2017.

[45] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," 2017.

[46] K. V. Cardoso, M. J. Abdel-Rahman, A. B. MacKenzie, and L. A. DaSilva, "Virtualization and programmability in mobile wireless networks: Architecture and resource management," in *Workshop on Mobile Edge Communications*. ACM, 2017.

[47] R. Durairajan and P. Barford, "A techno-economic approach for broadband deployment in underserved areas," *ACM SIGCOMM Computer Communication Review*, vol. 47, no. 2, pp. 13–18, 2017.

[48] T. X. Tran, M.-P. Hosseini, and D. Pompili, "Mobile edge computing: Recent efforts and five key research directions," 2017.

[49] M. Bouet and V. Conan, "Geo-partitioning of mec resources," in *Workshop on Mobile Edge Communications*, 2017.

Appendix: Metro Fiber Provider and Source Information

In this section, we provide a detailed survey of metro fiber providers across the top 10 most populated MSAs in the US. Table IV lists metro fibers providers for each of the 10 MSAs by provider ID. And each provider ID is accompanied by a value in paranthesis that indicates the percentage of fiber miles contributed by that provider to the total fiber miles for that MSA. Table V lists information for each fiber provider including the URL used to access the fiber maps for that provider and the data type of the fiber maps. The data types could be KML, Image (IMG) or PDF. It should be noted that about 65% of the maps are of type KML and all but one fiber map (provider ID 25) are accessed directly from provider websites. Further, the table also notes additional information such as accessibility information and date the maps were last accessed. Unless specified otherwise, the last accessed date for the maps correspond to our last full audit date of Aug, 2017, when all maps used in this study were verified against the original source maps that we collected via Google search campaigns.

MSA	Provider IDs, (percentage fiber miles contributed)
New York-Newark-Jersey City, NY-NJ-PA	1 (23.45%), 2 (4.77%), 3 (2.04%), 4 (0.52%), 5 (1.96%), 6 (0.59%), 7 (6.59%), 8 (5.68%), 9 (45.67%), 10 (5.18%), 11 (0.01%), 12 (0.03%), 13 (0.86%), 14 (2.63%)
Los Angeles-Long Beach-Anaheim, CA	1 (37.72%), 7 (1.12%), 9 (13.78%), 12 (0.25%), 15 (4.15%), 16 (33.39%), 17 (4.59%), 18 (5.0%)
Chicago-Naperville-Elgin, IL-IN-WI	1 (8.53%), 7 (4.67%), 8 (17.88%), 9 (31.94%), 10 (8.99%), 12 (0.56%), 19 (0.5%), 20 (0.02%), 21 (3.09%), 22 (1.53%), 23 (1.33%), 24 (2.37%), 25 (0.63%), 26 (0.97%), 27 (12.71%), 28 (0.46%), 29 (3.8%)
Dallas-Fort Worth-Arlington, TX	1 (0.16%), 5 (10.44%), 7 (3.75%), 8 (13.91%), 9 (34.01%), 10 (5.98%), 12 (3.99%), 18 (5.77%), 26 (6.57%), 30 (5.25%), 31 (7.21%), 32 (1.03%), 33 (0.91%), 34 (1.0%)
Houston-The Woodlands-Sugar Land, TX	1 (0.17%), 5 (13.73%), 8 (17.5%), 9 (20.09%), 10 (9.73%), 18 (6.68%), 30 (7.46%), 31 (12.51%), 35 (12.13%)
Washington-Arlington-Alexandria, DC-VA-MD-WV	1 (6.86%), 5 (3.35%), 7 (5.24%), 8 (15.47%), 9 (37.37%), 10 (12.27%), 28 (3.19%), 31 (4.0%), 36 (3.61%), 37 (3.48%), 38 (4.95%), 39 (0.21%)
Miami-Fort Lauderdale-West Palm Beach, FL	1 (4.57%), 5 (49.42%), 7 (2.3%), 8 (5.27%), 9 (17.56%), 10 (4.33%), 13 (2.02%), 28 (1.62%), 31 (6.88%), 40 (6.04%)
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1 (33.85%), 5 (1.73%), 7 (5.5%), 8 (8.14%), 9 (43.06%), 10 (7.72%)
Atlanta-Sandy Springs-Roswell, GA	1 (8.64%), 5 (2.16%), 7 (12.08%), 8 (16.55%), 9 (44.73%), 10 (1.91%), 26 (0.96%), 28 (4.13%), 31 (1.46%), 36 (0.35%), 41 (6.68%), 42 (0.34%)
Boston-Cambridge-Newton, MA-NH	1 (12.56%), 8 (3.92%), 9 (31.08%), 10 (3.95%), 11 (0.42%), 12 (3.09%), 13 (1.06%), 28 (8.64%), 43 (5.8%), 44 (6.18%), 45 (0.54%), 46 (4.12%), 47 (5.3%), 48 (13.34%)

TABLE IV: Metro fiber providers for top 10 most populated MSAs. Provider ID refer to providers listed in Table 5. Values in parenthesis indicate the percentage of fiber miles contributed by each provider to the total fiber miles indicated in our data for an MSA.

ID	Provider name	Map source URL	Data type	Notes (if applicable)
1	Sunesys	http://www.sunesys.com/coverage/	KML	Now Crown Castle
2	Bestweb	http://bestweb.net/fiber-map/	KML	Aug, 2018
3	Cross River Fiber	http://crossriverfiber.com/assets/images/CRF_2016-Coverage-Area.pdf	PDF	Now Zenfi
4	Fiber Tech	www.fibertech.com/net_currentmaps.cfm?id=42	IMG	No longer accessible
5	FPL Fibernet	http://www.fplfibernet.com/network/map.html	KML	Now Crown Castle
6	Pangaea	https://pangaea.us/viewournetwork/	IMG	N/A
7	Uniti	https://uniti.com/network?map=leasing	KML	N/A
8	Windstream	http://carrier.windstreambusiness.com/interactive-map/	KML	N/A
9	Zayo	http://www.zayo.com/solutions/global-network/building-lists-kmz-files/	KML	N/A
10	Intelli Fiber	https://www.intellifiber.com/network/	KML	Now Windstream
11	First Light	https://www.firstlight.net/network/network-map/	IMG	N/A
12	Spectrum Business	https://business.spectrum.com/content/fibermap	PDF	N/A
13	Level3	http://www.level3.com/~/media/files/maps/en-network-services-level-3-network-map.ashx	KML	N/A
14	Optical Communications Group	http://ocgfiber.com/networkarea	IMG	No longer accessible
15	Abovenet	http://abovenet.com/maps/docs/Seattle%20Market%202011-12-12.pdf	PDF	No longer accessible
16	Edison Carrier Solutions	https://www.edisoncarriersolutions.com/explorenetworkmap	KML	Aug, 2018
17	Integra Network	http://www.integratelecom.com/pages/network-map.aspx	KML	Now Allstream
18	Santa Cruz Local Infrastructure	https://santacruzfiber.com/	KML	N/A
19	123Net	https://www.123.net/networkmap/	KML	N/A
20	ACD	https://acd.net/fibermap/	KML	N/A
21	Everstream	http://everstream.net/network/check-network/	KML	N/A
22	Ifiber	http://ifncom.co/network/statefibermap/	KML	Aug, 2018
23	IFN	https://www.intelligentfiber.com/network/statewidefibermap/	KML	Aug, 2018
24	Independent Fiber Network	http://ifnetwork.biz/network-maps/regional-map	KML	N/A
25	Lynx Routes	http://www.telecomramblings.com/media/maps/LYNX_Routes.kmz	KML	No longer accessible
26	UPN Networks	https://uniteprivatenetworks.com/interactive-map/	KML	N/A
27	Wow Business	https://www.wowforbusiness.com/company/partner-alliance/fiber-maps	KML	N/A
28	Crown Castle	http://www.crowncastle.com/Maps/Fiber_Maps/CentralFloridaFiber.pdf	KML	N/A
29	US Signal	https://ussignal.com/uploads/network_landing/Network_Map_-_online.pdf	PDF	N/A
30	Alpheus	http://alpheus.net/wp-content/uploads/2012/08/san-antonio-fiber-schematic.pdf	PDF	No longer accessible
31	Fiber Light	http://www.fiberlight.com/files/fiberlight/e7/e76c20ff-d3ca-4dd0-bde3-ab9502c15c95.pdf	PDF	N/A
32	Inner City Fibernet	https://www.innercityfiber.net/	KML	N/A
33	Light Link Networks	http://www.lightlinknetworks.com/pdfs/ep.pdf	PDF	N/A
34	Nextera Communications	https://nextera.net/fiber_coverage_map	KML	N/A
35	ICTX Wave Media	http://www.ictxwavemedia.net/images//mapbig.pdf	PDF	N/A
36	Lumos Fiber	https://www.segra.com/network/	KML	Now Segra, Aug 2018
37	Lumos Networks	https://www.segra.com/network/	KML	Now Segra, Aug 2018
38	Summitig	http://summitig.com/wp-content/uploads/2013/11/NoVA-LH.pdf	PDF	N/A
39	Ainet	http://www.ai.net/wp-content/uploads/2014/08/AiNET%20DC%20Detail%20Summer%202014.jpg	IMG	N/A
40	Tower Cloud	http://towercloud.com/wp-content/uploads/SouthGA-e1340636022388.png	IMG	N/A
41	GPW Network	https://gapublicweb.net/whatwedo/network/	KML	Aug, 2018
42	Spirit Fiber	https://www.segra.com/network/	KML	Now Segra, Aug 2018
43	Capenet	https://opencape.org/network-map	KML	Uses Open Cape
44	Maine Fiber Company	http://www.mainefiberco.com/	KML	N/A
45	Mass Broadband	https://broadband.masstech.org/news-and-updates/mapgallery/massbroadband123mapsdata	KML	N/A
46	Open Cape Network	https://opencape.org/about/network-map	KML	N/A
47	Towardex	http://towardex.com/network.html	IMG	N/A
48	186 Communications	http://186comm.com/interactive-network/	IMG	Now First Light

TABLE V: Details of metro fiber data sources for providers of various MSAs discussed in the paper.