Tracking Submarine Cables in the Wild

Ioana Livadariu[a,*], Ahmed Elmokashfi[a] and Georgios Smaragdakis[b]

[a]Simula Metropolitan, Oslo, Norway
[b]Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Internet infrastructure
Internet measurement
Internet physical topology
Fiber cables
Submarine cables

ABSTRACT

During the last ten years, thousands of kilometers of submarine cables have been rolled out to connect regions around the globe and improve intercontinental connectivity. However, while it is relatively easy to get information about the frequent roll-outs of these cables, it is challenging to translate these developments into network information to facilitate networking research. For example, announcements for new submarine cables typically mention landing points and not router IP addresses. With this network information, it is easier to assess the impact of a new submarine cable on end-to-end delays in the connecting regions. In this paper, we investigate the necessary and sufficient conditions to translate public announcements for submarine cables to network information that enables networking research on this topic. We also develop and evaluate a methodology to automatically extract IP-level information for deployed submarine cables and assess their impact on end-to-end performance.

1. Introduction

Investments in the Internet's infrastructure are at an all-time high [1] to connect the growing population of Internet users, estimated to be 5.3 Billion in 2023 [2]. Both governments, as well as private corporations, invest heavily in datacenters [3, 4, 5, 6], broadband and mobile infrastructure [7], and fiber cables [8] to connect users and to improve Web experience and application performance. Today, a significant fraction of the Internet population is only a few milliseconds away from servers of popular applications thanks to the deployment of content delivery networks and datacenters [9, 10, 11].

However, the deployment of datacenters around the globe and remote person-to-person communication require fast highways between continents and countries. This is realized to a large extent with submarine cables. Submarine cables are also a lifeline when it comes to connectivity of remote, developed and developing regions alike. Thus, the global submarine cable network is a critical part of the core Internet infrastructure carrying a significant part of the Internet traffic [12, 13].

Submarine cables were laid as early as 1850 when the first submarine cable connected Europe to North America. Connecting the same continents and deployed significantly later in 1988, TAT-8 was the first all-optic transatlantic undersea cable [14]. Figure 1 shows the yearly number of deployed cables and cable capacity. As of 2021, more than 400 submarine cables are deployed worldwide. Over half of them were deployed after 2004, and a further 36 new cables are scheduled to be ready for service in the following three years [15].

Cable capacity has also increased over time. Using cable capacity of 246 submarines cables deployed in recent

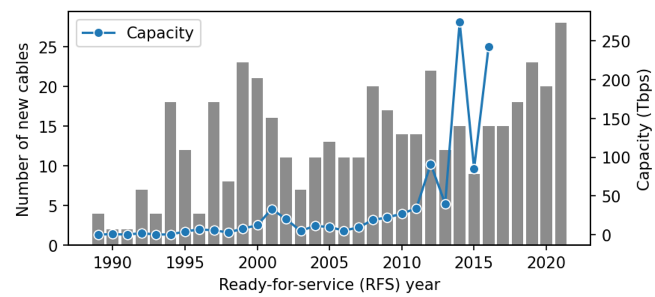


Figure 1: Evolution of the deployed submarine cables [source: Telegeography [15]]. Half of the submarine cables were deployed after 2004.

years [16], we notice that the newer cables have brought a significant increase in capacity. Cables deployed in 2000 have a median capacity of 480 Gbps, whereas those deployed in 2016 have a median capacity of 23.9 Tbps. We note that the majority of high-capacity cables have been deployed between 2014 and 2016. One such example is the MAREA submarine cable deployed in 2018. Connecting Spain to the United States, this cable possesses a design capacity of 208 Tbps capacity [17].

In terms of connectivity, three quarters of submarine cable connect countries located within the same continent. Countries in Europe and Asia are the most connected ones. Europe and Africa are the most connected pair of continents, followed by Asia and Africa. In contrast, as of 2023, there is no submarine cable between North America and Africa. Similarly, there is no submarine cable between South America and Oceania.

Submarine cables used to be either owned by a consortium or just one organization, which are governments or national carriers. However, this model has been changing in the last few years. Currently, content providers and hypergiants [18], like Google, Microsoft or Facebook are becoming cable owners or co-investors [19]. Approximately

*Corresponding author
ioana@simula.no (I. Livadariu); ahmed@simula.no (A. Elmokashfi); G.Smaragdakis@tudelft.nl (G. Smaragdakis)
ORCID(s): 0000-0003-2263-6934 (I. Livadariu); 0000-0001-9964-214X (A. Elmokashfi); 0000-0002-4127-3617 (G. Smaragdakis)

two thirds of the deployed cables are single-owned and span short distances (less than 200 kilometers).

While announcements about the roll-out of new submarine cables are publicly available, they only contain information about the landing location. This information alone is not useful for networking research as it does not provide IP-level information. Without these insights it is difficult to assess what is the return of submarine cable investments that can inform public policy debate as well as future investments and operation decisions.

The first work that investigated the impact of a new submarine cable was the paper authored by Fanou *et al.* [20]. The authors analyze the effect of the South Atlantic Cable System (SACS) [21, 22] cable, that connected Angola and Brazil in 2018, on the end-to-end performance and routing system. In the process, the authors identify the IP-level link that maps to the submarine cable. Our study takes the next step by proposing a generalized methodology for inferring submarine cable IP links in the wild. Specifically, we introduce and evaluate a number of techniques, namely, (i) owner and cable named based identification, (ii) cable IP link visibility, (iii) cable IP link centrality, and (iv) path directionality tests, to accurately map landing points to IPs for arbitrary submarine cables. These contributions help us in detecting the submarine cable IP links when we apply our method to already available public IP-level path data.

Our contributions can be summarized as follows:

- We develop a generalized methodology to translate public announcements about submarine cables to information relevant for networking research, i.e., IP addresses that are involved.
- We describe the necessary and sufficient conditions to translate public announcements into helpful information for networking research, and we propose a methodology to validate our results.
- We use our methodology to assess the impact of submarine cables on end-to-end delay at various regions around the globe, and we show that the roll-out of a new cable may only be beneficial for some of the networks and, thus, users of a connecting region.
- The active measurements collected for this work will be made publicly available.

The rest of the paper is organized as follows. Section 2 presents the challenges faced when inferring submarine cable IP links in the wild. In Section 3 we describe in detail the datasets used in this paper. In Section 4 we propose a generalized method for tracking submarine cables in traceroute data, and in Section 5, we validate our results. Using the validated results, in Section 6, we present the impact of the cable deployment on the end-to-end performance. Section 8 details the related work and in Section 9, we discuss our findings. Section 10 lists our conclusions and avenues for future research.

2. Challenges in Inferring Submarine Cable IP links

Identifying submarine cables that are crossed by a traffic flow as well as the general impact of laying a new submarine cable have largely remained unstudied. This is partially due to the challenges that come when developing a method that aims to associate traceroutes with submarine cables. However, even if such a method exists, available traceroutes highly impact its success. In this study, we develop a method that tracks submarine cables in available traceroutes. Consequently, we process publicly available IP-level paths collected from different vantage points located across the world. However, using this data to infer IP links that map to submarine cables comes with a series of challenges, which we discuss below.

2.1. IP path data challenges

Our approach maps IP links to a submarine cable by searching for cable-related hints within the IP-level paths collected between the countries that are inter-connected by the cable. Examples of such hints include the IPs owned by the cable owners or advertised by them. We also hypothesize that a submarine IP link will only become visible after the respective cable is deployed. Hence, for a given cable, we consider relevant IP paths collected before and after the cable *Ready-for-Service* (RFS) date. Our starting point is always a set of traceroutes between vantage points that are located in a pair of countries that are interconnected by a cable of interest. However, this approach comes with challenges that are linked both to the traceroute data itself and the available measurement platform.

Network measurements collected using traceroute between pairs of vantage points provide the IP address and delay for each hop between these locations. However, such measurements are influenced by different network conditions, resulting in delay variability and missing values in the traceroute data [23]. Our method uses both the IP addresses and the delay difference between consecutive hops. We account for any variations in the IP link delay by using the median values of delay difference over six months. At the same time, we factor in the missing values by considering an IP link as the link between two consecutive responsive hops.

Another data-related challenge is *coverage* in terms of available public measurements between the pairs of countries that we are interested in. Our work leverages traceroute data collected by RIPE Atlas probes. These vantage points are located in 173 countries [24]. The probe's geographic diversity, however, does not necessarily guarantee coverage in terms of landing point countries. Also, it does not guarantee that we have traceroutes traversing cable owners' networks. One such example is the Tui-Samoa cable [25] which connects Wallis and Futuna, Fiji and Samoa. However, there is no RIPE Atlas probe in Wallis and Futuna. Hence, we could only just analyze one of the three cable segments, i.e., the segment that connects Fiji and Samoa. Moreover, installation and operation of probes at a certain location does not guarantee that useful traceroutes are available. For example, from the 13 Angolan and 252 Brazilian probes, only one in Angola and 71 in Brazil are usable[1]. Hence, we could potentially rely only on these probes to study the South Atlantic Cable System (SACS) cable connecting Angola and Brazil [22]. However, we did not find any traceroutes between these probes in the publicly available measurements, which means that no user has scheduled measurements between these probes. Hence, studying the SACS cable requires dedicated measurements. In their study, Fanou *et al.* [20] focused on this particular cable, and used custom measurements to identify the submarine IP link. The authors relied on looking glass (LG) servers and CAIDA's Ark infrastructure [26] to design their measurements and collect data. Specifically, they ran traceroute campaigns for two days in March 2019 and collected traces between eight Ark vantage points located in Brazil and one LG located in Angola. The authors published this dataset [27] and we use it to cross-check our methodology (see Section 5).

2.2. Cable characteristics challenges

Tracking the deployment submarine cable depends on whether we are able to map the IP-level path to the specific cable by relying on the cable characteristics like cable name or cable owners. To this end, our inference approach relies on multiple datasets from which we extract information related to the submarine cable owners and geographic location. We provide an overview of our datasets in section 3.

In the ideal scenario, IP addresses that map to the cable link would be owned by one of the cable owners and located close to the landing points in the cable owner network. Moreover, the DNS pointer record (*PTR* for short) that provides the domain name associated with these IPs would comprise both geo-hints and hints about the cable. For example, the Monet cable [28] satisfies all these criteria. The cable is owned by a consortium composed of Algar Telecom, Angola Cables, Antel Uruguay, and Google. It connects Brazil to the Southeast of the United States. The cable landing points are Fortaleza and Santos in Brazil, and Boca Raton in the United States. These are located either in large cities, like Fortaleza, or next to a large city, i.e., Santos is close to Sao Paulo and Boca Raton is close to Miami. We find traces within the collected data before and after the cable is deployed, between Brazil and the United States. Moreover, for a subset of the traces the vantage points are in close proximity to the landing points. For the IP hops along these traces, we extract the PTR records and routing information. PTR records for some IP addresses include the cable name (*monet*), the owner name (*Algar Telecom*) and the location (*spo-piaf* which maps to Sao Paulo). In addition, we find IPs advertised by one of the owners.

Deployed in the same region, the South America-1 (SAm-1) cable connects eight countries located in South and Central America [29]. Owned by Telxius, the cable has been in use since 2001. However, this specific cable is currently not analyzable, since there is no public dataset that covers the deployment period. Moreover, the cable owner does not have any registered autonomous system number (ASN). Thus, we are unable to identify any IP space that is advertised by the owner. In this case, the owner most likely rents out to third parties, which most likely appear in the IP path data. Cables similar to SAm-1 are not analyzable without dedicated measurements and extra information about companies renting capacity on it.

2.3. Distinguishing cables

Another challenge is to distinguish between multiple cables that connect popular landing points. For example, we observe multiple cables that connect Great Britain with the US East Coast, and some of them use the same or close landing points, e.g., city ports close to London and New York metropolitan regions. The granularity of the traceroute measurements may not be sufficient to accurately locate the landing points. Thus, multiple hints have to be utilized (as described above) to assign an IP link to a physical cable. Such an example is the FLAG Atlantic-1 [30] and Atlantic Crossing-1 [31] cables. The landing points in the New York metropolitan area are located within 150 km of each other: Island Park and Northport for the former cable, and Brookhaven for the latter one. Both cables also have landing points on the west coast of Great Britain: Skewjack and Whitesands Bay.

3. Datasets

We describe in this section the public[2] datasets used in our work.

Submarine cable data: TeleGeography collects information on the deployment of submarine cables. Using this data, TeleGeography generates and maintains a submarine cable map that summarizes key aspects of submarine cables [15]. For each cable, this data comprises the *cable name*, *ready-for-service (RFS) date*, *landing points* and *cable owners*. However, as of 2021, TeleGeography's dataset is no longer available[3] and only the submarine cable map is publicly available. Thus, we conduct our study with the data collected earlier that comprises information on cables deployed until 2020. For most submarine cables, the RFS date provides both the year and month when submarine cable was deployed. However, for a few cables the RFS date is only the deployment year. In such cases, we consider January and December of the deployment year as the start and end of the deployment period, respectively.

Traceroute data: The RIPE Atlas project provides active vantage points that can be used to measure Internet performance and availability [24]. As of the 1st of July 2021, the measurement platform comprises 11,939 probes in 173 countries and 3,686 networks. We use traceroutes between

[1] We consider as usable vantage points, the RIPE Atlas probes that have the status value set to *Connected*.

[2] Both TeleGeography and Rapid7 datasets were available when our method was developed.

[3] TeleGeography does not make available their repository on GitHub for their data.

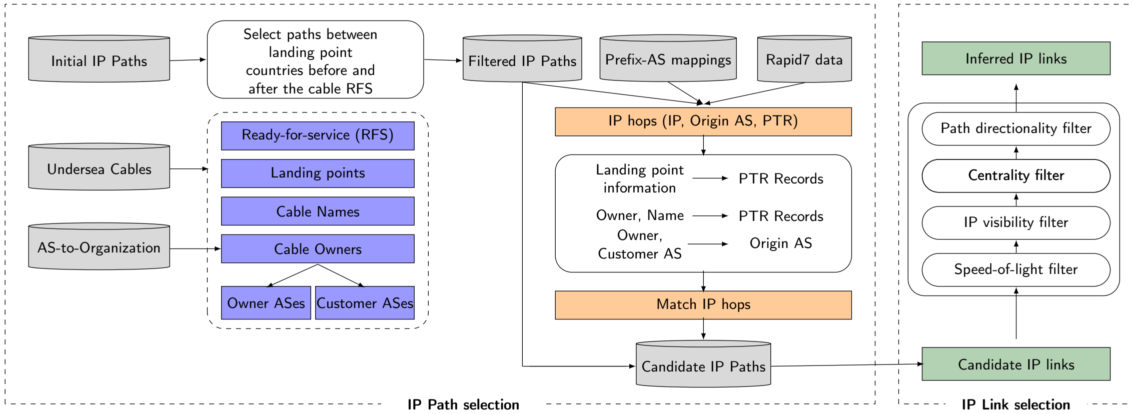


Figure 2: Overview of the submarine cable IP link inferring methodology. Our two-step method uses a multi-datasets approach to identify submarine cable links in the IP path data.

RIPE Atlas probes to infer and analyze the impact of submarine cable links. These traces include both periodic and user measurements. We collect data on the first day of each month from 2015 to 2019. Our choice for the start of the measurement period is determined by the significant number of RIPE Atlas probes deployed in 2015 [32]. At the same time, the collection periods between the traceroute data and the submarine cable data must align. Hence, we consider only traceroute measurements until the end of 2019. The collected data includes traces between 318,566 RIPE probe pairs and 1,429 country pairs.

AS-to-Organization (AS2ORG) dataset: CAIDA regularly publishes data that maps ASes to organizations [33]. We use these mappings to determine the cable owners ASes at the time of data collection.

BGP data: We use prefix-to-AS mappings derived from RouteViews data [34] and published by CAIDA [35] to determine the IP address blocks originated by submarine cable owners.

Rapid7 Open data: Rapid7's Project Sonar [36] is a security research project that regularly scans and analyzes the public Internet, and generates a set of datasets. We use the datasets that include the IPv4 PTR lookups for the advertised IPv4 space [37] to find cable related hints in PTR records. Specifically, we search in the collected PTR records for the name of owners and/or submarine cables. In the same records, we also search for geo-hints that match information extracted from the cable's landing points. Starting from 2021, Rapid7 restricts access to their datasets [38]. Note that we discuss in section 9 the usage of other publicly available DNS data sources. However, our collected data covers the measurement period.

4. Inferring cable links

In this section, we describe our two-step methodology for inferring cable links in traceroutes. Figure 2 illustrates the main blocks of our approach.

4.1. IP-level path selection

The first step is to identify the IP-level paths that potentially cross submarine cables. As shown in Figure 2, we start from a set of *initial paths* and select paths between countries where the cable landing points are located. We also rely on the cable ready-for-service (RFS) date to filter paths that are collected within six months of the RFS date, which we refer to as *filtered paths* in the following. We choose a time window of six months as we believe that this offers sufficient time to observe changes in the IP path data. Recall that for cables that have listed only the year as the RFS date, we consider January and December as the start and end deployment periods, respectively. For such cable, we collect traces six months before (after) the start (end) of the RFS period.

We started our analysis with an initial set of 64 submarine cables that have RFS dates between June 2015 and June 2019. Our choice of the measurement period is motivated by the significant number of RIPE Atlas probes deployed in 2015 [32], i.e., 7,490 connected probes were deployed as of 2015. Filtering out cables that are limited to a single country left us with 39 cables. Amongst the cables that connect different countries, we can only consider cables where we have traceroutes between country pairs with landing points. Due to the *IP path data challenges*, only 26 cables satisfy this criteria.

We further apply a set of filters to identify traceroute hops that can be linked to submarine cables. Here, we identify IP hops that are related to cable owners as well as hops with PTR records that contain either the cable name, owner name or cable landing points geo-hints. For each cable, we first extract the cable name and cable owners from the Telegeography data and use the AS2ORG dataset [33] to identify the owners' ASes as well as the customer ASes of the owners. Second, for each hop along the filtered traces we extract the AS origin and PTR records. We use airport codes [39] and the international codes for cities and country administrative divisions [40] to generate a set of geo-hints for each landing point. We then search the PTR records for the *cable name*, *owner name*, *landing point country* and

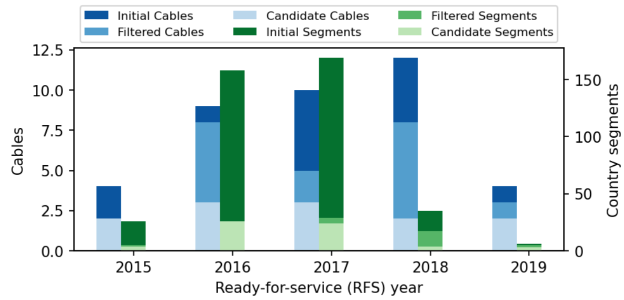


Figure 3: Annual number of cables and cable segments. Our method find candidate paths for 64 submarine cables and 93 submarine segments.

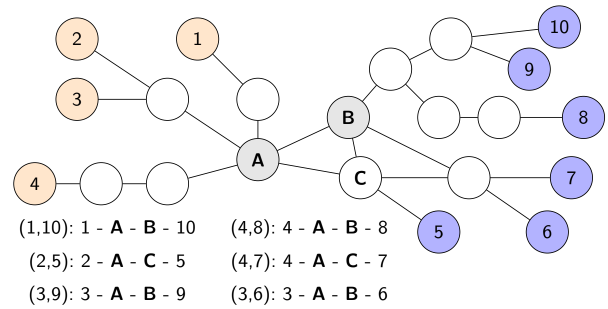


Figure 4: Vantage point centrality. Color-coded nodes indicate vantage points located in different countries. From the listed paths we infer that the nodes with the highest centrality are A (6) and B (4). Consequently, the (A,B) link is selected as the cable IP link.

landing point geolocation hints. Recall that we detail in section 2 the *cable characteristics challenges* in mapping IPs to submarine cables. We label each IP hop with the matching criteria, and filter out traces that comprise only unlabelled IP hops. We refer to IP paths that have at least one labeled hop as *candidate paths*. Figure 3 breaks down the process above per year. Note that we show here both cables and cable segments. A cable may comprise several segments and span several countries. We find candidate paths for 11 of the 64 and 93 of the 433 initial submarine cables and segments, respectively.

4.2. IP link selection

We next devise a set of four filters that aim to strip out all candidate links that are unlikely to be a cable hop.

Speed-of-light (SoL) filter: Our first filter removes links that are traversed faster than the submarine cable links. For each cable segment, we use the landing points geographic coordinates to compute the cable segment length (l). Specifically, we compute the length as the haversine distance between these two cable landing point locations. We further use this value and the speed-of-light to compute the theoretical delay $D_t(l)$ between landing points as:

$$D_t(l) = 2 \cdot \frac{l}{\frac{2}{3} \cdot c} \tag{1}$$

where c is the speed-of-light. For each candidate IP link, we select the delay difference between the source and destination hop and extract the median delay over the six months period after the cable RFS date. The delay values used when computing the delay difference are collected from the same traceroute measurement. Finally, we filter out IP links for which this median value is smaller than the theoretical delay. Note that the method proposed by Fanou *et al.* [20] also uses the speed-of-light test to filter IP links.

IP visibility filter: Submarine cable owners are likely to renumber/assign IP address blocks to IP links that map the submarine cable segments. Hence, these links would become visible after the cable's deployment date. Thus, we select IP links for which both the source and destination IPs are visible only after the cable is laid.

Vantage point centrality filter: We hypothesize that paths collected from multiple vantage point pairs that are located in two landing point countries are likely to traverse the submarine cable. Thus, the source and/or destination IPs of the IP link that maps to the submarine cable segment becomes visible on these paths. In other words, in a connectivity graph built from the selected candidate traces, these hops would have a high centrality value. The nodes in this graph represent the IP hops on the traces, and the edges correspond to the IP links. We assign to each IP hop in the graph a visibility metric computed as the number of vantage point pairs between which we find the specific hop on the IP path. We order the IP hops based on this value and select the IP links with high visibility, i.e., the destination and/or source is in the top 10% in terms of vantage point visibility value. Figure 4 shows the traces between vantage points located in two countries. Note that we color-code the nodes in these countries. In this example, hop A, B, and C have the highest centrality: 6, 4, and 2, respectively, and thus, we select the (A, B) link as the cable IP link.

Path directionality filter: Our final filter aims to reduce the number of candidate IP links by taking into account the path prefix sequence as follows. We consider the top two most visible IP links, and select traces that traverse the selected links to construct the prefix sequence for each link. Specifically, we build a sequence composed from prefixes that appear before and after the IP links on the path. We further apply the constructed pattern to the least visible edges, and filter out the links that do not match the constructed pattern.

4.3. Inferred submarine cable IP links

When applying the devised filters to the candidate IP level paths we infer 717 IP links that map to 11 cables and 67 cable segments. Analyzing the number of IP links per segment, we find that our method infers on average 21 IP links for each cable segment. Table 1 lists the number and percentage of IP links that are removed by each filter. We observe that the speed-of-light filter removes most of the IP links, followed by the IP visibility filter.

Table 1
Number (Percentage) of IP links removed by each filter: Speed-of-light (SoL), IP visibility (Visibility), Vantage point centrality (Centrality) and Path directionality (Path).

Filter	IP Links
SoL	4619 (76.30%)
Visibility	401 (6.62%)
Centrality	148 (2.44%)
Path	169 (2.80%)

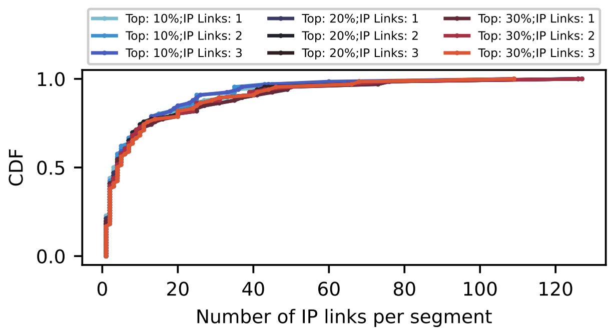


Figure 5: Distribution of the number of IP links per cable segment when varying the centrality and path directionality parameters. The number of links is comparable across different value of the parameters indicating the robustness of our method.

4.4. Sensitivity analysis

We compute the number of inferred links per cable segment when varying the vantage point centrality and path directionality filter parameters. Specifically, we vary the centrality threshold to consider the connectivity graph's top 10%, 20%, and 30% central IP addresses. We also consider one, two and three as the number of most visible IP links when applying the path directionality filter. Figure 5 shows the inferred link distribution for the considered parameters. We find the number of links to be comparable across the considered values, thus confirming the robustness of our method. Moreover, we compare the increase in the number of links when varying just one parameter. Thus, varying only the centrality threshold and setting the number of most visible links to two results in an increase of at most 17% in the number of inferred links. Similarly, varying the number of most visible links and considering only 10% as the threshold for the centrality filter generates at most 5% more links.

5. Validation

We follow a two-step process to validate the inferred cable links. For a given inferred link, we first validate the geographical location of the link's end points, and further use these results to validate the IP link, i.e., it is an operational submarine cable. But before embarking on this process, we take a closer look at the inferred links and narrow them further down by imposing an upper limit on measured round trip time (RTT). In other words, we consider links with delays that are relatively close to the theoretical link delay. For an inferred cable segment, we compute the minimum theoretical link delay as follows. We identify the nearest largest city to each landing point [41]. Then, we estimate the theoretical RTT as the sum of delays from these cities to respective landing points and the delay between the landing points. RTTs for approximately half of the inferred links are within 50% of their theoretical RTT. We further focus on validating these links.

5.1. Landing points

We rely on active measurements to confirm the physical location of source and destination IP addresses for each inferred link. Our goal is to find vantage points in the same or close to the landing point location in order to probe the inferred IP addresses resulting in small delays in ping measurements. To this end, we use publicly accessible looking glasses (LG) to probe the IP addresses. We choose LGs that are both located near the landing points and hosted within the owner network. For example, we consider the Tasman Global Access (TGA) cable [42] that is owned by Spark New Zealand, Telstra and Vodafone, and connects Oxford Falls, Australia to Raglan, New Zealand. For IP addresses mapped to the Oxford Falls in Australia, we search for LGs within the Telstra network located in Sydney, Australia, which is approximately 20 kilometers away from the landing point. Although done manually for this work, this process can be automated. In fact, AliceLG [43] and CAIDA's Periscope Looking Glass API [44] offer an automated interface to run such measurements from different IXPs. AliceLG provides such an interface for 23 IXPs located in seven countries. Cross-checking these LGs against the geographic location (city and country) of the valid submarine cable landing points results in five matches. Periscope has a higher coverage as it provides an interface to LGs located in 75 countries and 329 ASes. However, we find that only 12 ASes are registered to the organization that own any of the valid submarine cables. Thus, relying only the two automated tool to validated possible inferred landing points is not feasible. As future work we plan to develop an automated tool for validating the inferred submarine cable links.

Depending on the LG location, and the probing result we classify each IP address as *valid*, *misclassified*, or *unclassifiable*. We consider the IP address mapped to the cable landing point, *valid*, if the delay between the chosen LG and the IP addresses is within 5 milliseconds. Given that the IP addresses are within one or two hops from the LGs, we hypothesize that this value reflects the landing point validation. In fact, our collected validation data shows that for most cases the delay value is at most five milliseconds. For misclassified IPs the delay is greater than this value. We label an IP address as unclassifiable when there is no valid LG (*no-LG*), the IP address is unresponsive (*unresponsive*)

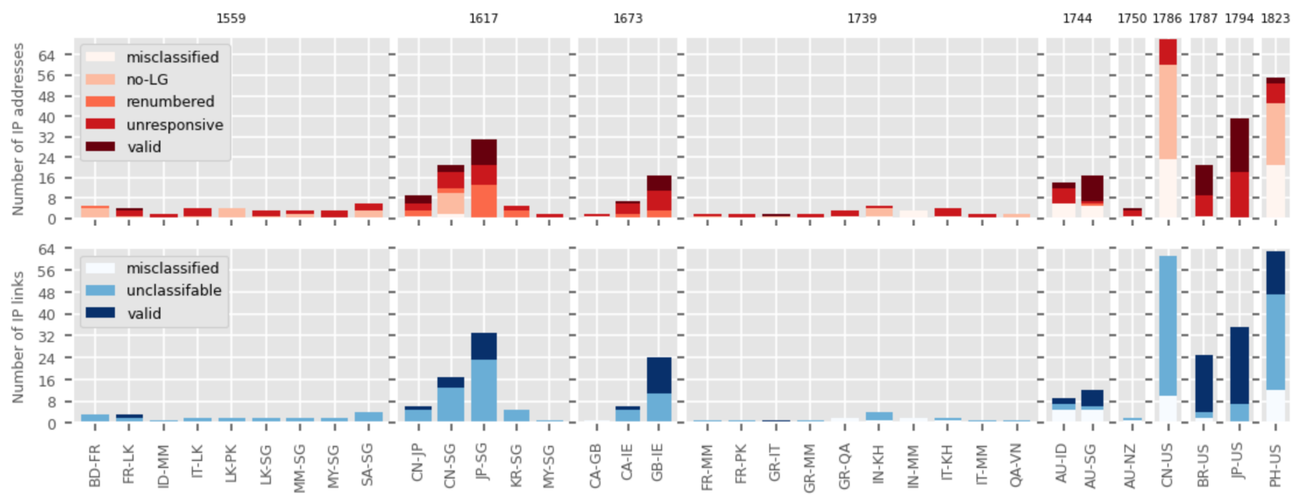


Figure 6: Submarine cables validation per cable segment: IP addresses (top) and IP links (bottom). We indicate the identifier for each cable on top of the figures. For each cable we mark the segment on the x-axis. We validate 71% of the 146 classifiable cables.

or the PTR record indicates that the address space is renumbered (*renumbered*) [4]. The top plot in Figure 6 shows the number of IPs within each class per each cable segment. We group the cable segments for each cable (marked as the x-axis), and indicate on top of each figure the cable identifier. Our validation shows that the inferred IPs that map to most segments are either valid or unclassifiable.

The ability to validate a segment varies widely depending on landing point location and cable owners. Most of the misclassified IPs are mapped to two submarine cables segments: New Cross Pacific (NCP) Cable System cable from China to United States (ID: 1786, CN-US) and the SEA-US segment between Philippines and the United States (ID: 1823, PH-US). A closer investigation reveals the root causes for misclassification. For the CN-US segment, lack of validation vantage points is the main cause. We are unable to use the LGs located in China as the Chinese landing points are located more than 1,000 kilometers and 3,000 kilometers from the LG located near Hong Kong and Urumqi, respectively. For the SEA-US cable segment, the presence of multiple cables on parts of the path appears to be the root cause of misclassification.

5.2. Cable links

To validate the inferred IP links we account for both the IP source and destination classification. A link is considered valid if both ends are validated or one is validated and the other is unresponsive. A link is considered misclassified (unclassifiable) if any of its end points is misclassified (unclassifiable).

The bottom part of Figure 6 shows the IP link classification. We label 146 and 191 links as classifiable and unclassifiable, respectively. About 71% of the 146 classifiable links are valid. For eight cable segments, we validate most of the classifiable links. We validate all such links for the GTT Express cable's (ID: 1673) [45] segments between Great Britain and Ireland (GB-IE) and Canada and Ireland (CA-IE), and Japan to United States (JP-US) segment for the FASTER cable (ID: 1739). Similarly, we validate 21 of the 25 inferred IP links for the Monet cable (ID: 1787) [28].

We note that the all inferred links that map to 14 of the 34 cable segments are unclassifiable. For the SeaMeWe-5 cable (ID: 1559) [46], we validate just one link on the segment between France and Sri Lanka (FR-LK). For most of the remaining links we are unable to probe one IP address of the link, and the other IP is unresponsive or renumbered. Hence, we label these IP links as unclassifiable. Similarly, for half of the ten segments of the AAE-1 cable (ID: 1739) [47] and two segments for the Asia Pacific Gateway cable (ID: 1617) [48] we label all the links are unclassifiable.

5.3. Cross-check with other methods

Fanou *et al.* [20] analyzed the impact of a new cable on end-to-end paths and performance, which they apply on the South Atlantic Cable System (SACS). Our proposal is a generalized approach to identifying IP-level path links that map to submarine cables. Thus, our method takes as input a list of submarine cables and public traceroute data collected during the deployment period of the submarine cables and infers the IP links in the traces that traverse the submarine cables. To this end, we leverage information extracted from multiple sources and propose a set of filters to identify the potential submarine IP links. Compared to the previous study, our work's focus is on inferring submarine cable links in the wild, while Fanou *et al.* focus on analyzing the impact of the South Atlantic Cable System. Moreover, Fanou *et al.* ran custom measurements from LGs and CAIDA

[4] PTRs for renumbered IPs include geo-hints pointing to locations other than the area of the landing point.

Ark nodes to collect a few traces mapped to the SACS cable. Our method, however, was applied to a significantly higher number of passive traceroutes and inferred IP links for 11 submarine cables.

We compare their study with our inference approach as follows. First, we apply our proposed methodology on the data published by Fanou *et al.*, and compare the results from our methodology with the results published by the authors [27]. Next, we apply the above mentioned method on public traceroute data and compare the results. Note that we rely in part on our code and our metadata as Fanou *et al.* made available only the code corresponding to the SACS deployment analysis.

In their study, Fanou *et al.* collect traceroutes using custom measurements between Angolan LG located in the cable owner network and CAIDA Ark's vantage points in Brazil [20]. They further use the speed-of-light test, IP geolocation and hostname-based geolocation to infer two IP addresses as the SACS IP link. The study also reports 47 IP-aliases for the two IPs (29 IPs for one end of the IP link, and 18 IPs for the other end). We used the collected traceroute data as input to our method, and identified IP links for which at least one IP hop is either advertised by the owner AS or/and the corresponding PTR record contains the cable or owner name. Our methods infers 30 IP links that correspond to the SACS cable. We find both the IPs inferred by Fanou *et al.* [20] as part of three IP links for which the cable owner AS advertised the IP addresses. Our method also infers that at least one IP address from 17 links either contains the owner name within the corresponding PTR record or is advertised by the AS owner. For the remaining links only the latter criteria is satisfied.

We also apply the method proposed by Fanou *et al.* [20] on traceroute data since our proposed method tracks submarine cable IP links in such data. To this end, we consider from Fanou's study the first two steps: *collecting the IP paths* and *identifying the router IP interfaces of the submarine cables*. To complete the first step, we select traceroutes used to build our methodology. In their second step, the authors identify the IP cable link by relying on the speed-of-light test, IP geolocation and hostname-geolocation. We rely on our extracted metadata to run the proposed approach on the traces we collected from the RIPE Atlas platform. Specifically, we first select the links that satisfy the speed-of-light criteria. Next, we consider links for which the PTR record for at least one hop contains either the landing point country and landing point geolocation hints. These steps yield an overall 6,841 IP links for 32 submarine cables and 118 cable segments with an average of 57 IP links per cable. Imposing that both ends match the geolocation criteria reduces the number of inferred IP links to 1,844 from 31 cables and 71 cable segments.

We further investigate whether this method infers the same results as the method proposed in this study. Hence we search within the set of links generated by Fanou's method, i.e., 6,841 and 1,844 IP links, for the set valid IP links. We find 64 and 33 of the valid IP links inferred by the analyzable method. Thus, we conclude that relying on Fanou *et al.* approach yields a high number of false positive.

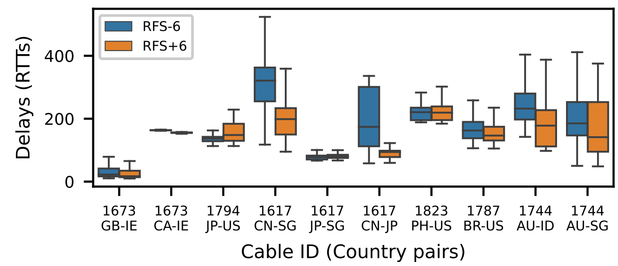


Figure 7: Delay values per cable segment before (RFS-6) and after (RFS+6) the submarine cable are deployed. We observe improvement in performance for eight of the ten cable segments.

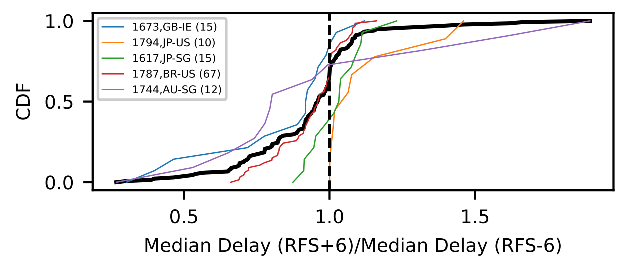


Figure 8: Median delay distribution per vantage point pairs. We observe improvement in the delay value for seven of the cable segments for at least 75% of the pairs.

6. Impact on Performance

Having validated the inferred submarine links, we further investigate the submarine cable impact on performance.

6.1. Overall analysis

We use the list of validated IP links to evaluate the impact of the submarine cable on end-to-end delay. To this end, we first select the source and destination vantage points of the traces that traverse the inferred links. Next, we extract the delays between the selected pairs before and after the ready for service date of the cable. We group these delays for each cable segment. In Figure 7, we plot the delay values before (RFS - 6 months) and after (RFS + 6 months) the RFS date. We mark on the x-axis the submarine cable identifier and the countries between which the cable segment is rolled out. We find an improvement in performance for eight of the ten cable segments. Moreover, for four of the nine cable segments the median delay value decreases by more than 20%. These results show a clear impact of the submarine cable deployment on end-to-end performance.

6.2. Submarine cable country segment impact

Having seen an overall impact on the performance, we further break down our analysis per vantage point pairs.

Thus, we compute the delay fraction per vantage point pairs and show in Figure 8 the distribution of these values. Note that we compare pairs with measurements before and after the RFS date. We marked with a dashed line the limit between the vantage point pairs that experience a decrease (left) and an increase (right) in delay. The black line corresponds to the distribution across all the vantage point pairs. For 21 of the 136 pairs the RTTs vary with 1% after the cable is deployed. However, for 80 and 35 pairs we find a decrease and an increase in the median delay after the cable is laid, respectively.

Taking this one step further, we break the distribution per cable segment, and plot in the same figure the CDF of the delay fraction for these segments. We color-code each cable segment and indicate in parenthesis the number of vantage point pairs used in the analysis. For example, we use 15 pairs to plot the distribution for the Asia Pacific Gateway (APG) cable segment between Japan and Singapore. Our analysis shows that more than 75% of the pairs experience an improvement in the delay value for seven of the cable segments. Not surprisingly, these segments also experience an overall improvement in the delay value. These findings reinforce our observation that the submarine cable deployment impacts the end-to-end performance.

6.3. Causes of changes in delay

We analyze the AS-path level between the vantage point pairs to understand the difference (increase/decrease) in delay values. Recall that we use BGP data to map the IP addresses to ASes. We hypothesize that owner organizations are going to utilize their cables. Hence, we would observe AS-path changes with the owner AS present on these paths. We select the most frequent AS paths before and after the deployment date and group the pairs based on whether we observe changes between these sets and whether the owner AS is present on the paths before and/or after the cable is deployed. Table 2 lists the number of pairs for each category when the median delay value decreases/increases after the cable is deployed. We list in parentheses the number of pairs for which the delay remains relatively stable, i.e., the value varies within 5 and 3 milliseconds for cable segments longer and shorter than 1,000 kilometers, respectively. Only the cable segment from Great Britain to Ireland is shorter than 1,000 kilometers. For example we consider the corresponding AS paths for the Monet cable (ID: 1787) between Brazil and the United State. We observe an AS path change between most of the vantage point pairs as the path remains the same and already is traversing the AS cable owner for only between seven pairs. Note that for three (four) of these pairs we find that the delay increases (decreases). Overall, the AS-path changes for 110 of the 136 pairs. Moreover, we find that the owner AS appears on the paths for 82 pairs with two thirds experiencing improvement in performance. Our results suggest that cable deployments result in topological changes and routing optimization. Thus, we consider that the delay change is linked to the cable deployment.

Table 2
Number of pairs per undersea cable segment with no change/change in the AS Path without and with the owner AS (w/oOwner and wOwner) present after the cable is deployed. We observe an AS-level path change between 80% of the vantage point pairs.

Cable ID,CCs	AS Path Change		No AS Path Change	
	w/oOwner	wOwner	w/oOwner	wOwner
Increase delay				
1673, GB-IE	2 (2)	0	0	0
1794, JP-US	0	4 (3)	0	4 (0)
1617, CN-SG	0	0	0	1 (0)
1617, JP-SG	0	6 (1)	0	3 (3)
1823, PH-US	0	2 (2)	0	0
1787, BR-US	3 (2)	10 (2)	0	3 (1)
1744, AU-SG	0	2 (0)	0	0
Total	**5 (4)**	**24 (8)**	**0**	**11 (4)**
Decrease delay				
1673, GB-IE	4 (2)	7 (3)	2 (1)	0
1673, CA-IE	0	1 (0)	0	0
1794, JP-US	0	0	0	2 (2)
1617, CN-SG	0	2 (0)	0	2 (0)
1617, JP-SG	0	2 (2)	0	4 (1)
1617, CN-JP	0	0	0	1 (0)
1823, PH-US	0	2 (1)	0	0
1787, BR-US	16 (4)	31 (10)	0	4 (0)
1744, AU-ID	0	6 (0)	0	0
1744, AU-SG	3 (1)	7 (1)	0	0
Total	**23 (7)**	**58 (17)**	**2 (1)**	**13 (3)**

7. Control plane

In this section, we study the deployed submarine cables impact on the routing system. Intuitively, we expect to observe changes on routing paths where the cable owner's network is present as well as in the group of networks that use these paths. Hence, we take the following steps in assessing the control-level plane impact. First, we collect routing tables from the RouteViews project [34] during the first seven days of each month over our five year measurement period. Second, for each inferred submarine cable owner in our dataset we use the AS-to-org dataset [33] to identify the ASes owned by these organizations. Third, we parse the routing data and filter the routes that either originate or traverse the cable owners. Lastly, for each submarine cable we compare three months of filter data before and after the RFS date. With the exception of the GTA TeleGuam, we collected paths that traverse at least one cable owner AS. Specifically, we collect data for 44 cable owners and 370 owner ASes.

7.1. AS pairs

Using the filtered AS paths, we further investigate the prevalence of the cable owners on these paths before and after each cable is deployed. Thus, for each of these paths we extract the source and destination AS and group these AS pairs per cable owner AS. We find that cable owners appear on paths that connect on average 161,555 AS pairs.

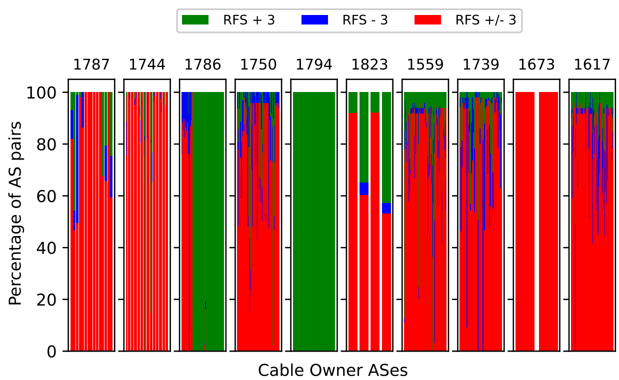


Figure 9: Percentage of AS pairs per cable owner before and/or after the RFS date. Each line corresponds to one cable owner and is color-coded to map the percentage of AS pairs.

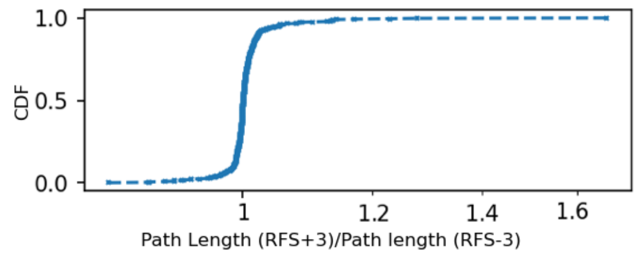


Figure 10: Distribution of fraction of average path length per submarine cable owner's AS. Half of the cable owner networks are traversed by slightly longer AS paths after the cable is deployed.

We further analyze whether we observe any change in the connecting pairs. To this end, we compute the number of AS pairs seen only three months before or after the RFS date of each cable as well as throughout the six months period. Half of the pairs are seen before and after the submarine cable is deployed, while 21% and 26% are observed only before and only after the RFS, respectively. Figure 9 shows the percentage of AS pairs seen before (blue color), after (green color), and during (red color) the roll-out for each cable owner. The numbers above the figure correspond to the cable identifiers, and each line corresponds to a cable owner. The Faster cable (cable ID: 1794) appears to be traversed by paths that connect only new AS pairs across all its owners.

Half of the owners appear before and after on the AS paths that connect 60% of the AS pairs. We notice that paths traversing the cable are used, on average, only before the roll-out of the cable by 11.21% AS pairs. For an average of 31.78% AS pairs, pairs appear to be used after the roll-out of the cables. We further focus on cable owners' ASes that are traversed by AS paths which connect new AS pairs. In Figure 9, we mark such pairs with green. These pairs account for at least 20% of the overall observed pairs for half of these cable owners. Our analysis shows variability in the owners' network connectivity as part of the owners change their peering relationships while for others we observe that they do not change their connectivity.

7.2. AS Paths

Our next step is to evaluate whether any changes occurred on the paths that traverse the submarine cables. To this end, we focus our analysis on AS pairs that appear to utilize the cable before and after the RFS date. For each such pair we first extract all the AS paths that appear to traverse cable network owners. Next, we compute the frequency of the AS paths for each pair and select only the ones with the highest frequency. By analyzing how these paths change for each cable owner, we find that 35% remain the same. Our analysis also shows that 43% (21%) of these paths appear only after (only before) the cable is deployed.

We further turn our attention to the length of the AS paths, and find that paths that appear throughout the cable deployment period have an average length of 3.39 hops. These paths are shorter compared to the ones seen only before and only after the RFS date. A close inspection of only these two latter categories shows that the AS paths that appear only before the cable deployment are on average slightly shorter than the ones that we observe only after the deployment. Breaking this analysis for each cable owner AS, we compute the average AS path length after the RFS date over the average length before the same date. Figure 10 shows the distribution of these values across all the owner ASes. Half of the cable owner networks appear to be traversed by longer paths after the cable deployment. However, the path increases at most with 20% for most of the networks. We also find that path length is reduced for approximately 40% of the owner networks.

8. Related work

Despite the major role that the the global submarine cable network plays in supporting reliable connectivity across the world, there is surprisingly only a few research studies that focus on this network [13, 20, 49, 50]. Bischof *et al.* [13] highlighted the importance of studying the submarine cable network for inter-continental end-to-end connection. In their follow-up work [50] the authors used traceroute data collected between RIPE Atlas probes and the most popular web resources within different regions to analyze the reliance of submarine cable for the specific regions. Their analysis reported that on average 28% of the resources were accessed via submarine cables. Our work has a different focus. In our study, we focused on tracking down submarine cables' IP links by analyzing publicly available data, and we presented a first-order analysis on the impact of deploying such cables at different parts of the world.

Closely related to our work is the study published by Fanou *et al.* [20]. In their work, the authors use a method to identify submarine cables using active measurements and use it to identify the SACS cable IP-level links. The focus of the study is to investigate the impact on end-to-end delay from different geographic regions and on routing. Surprisingly, the authors find that there are unintended consequences: the end-to-end latency to or from regions in Africa towards Angola increase after utilizing the newly deployed SACS cable. Our method extends this study by providing the necessary and sufficient conditions to map landing locations to network-level information (router IPs) by publicly available traceroute data that are not collected for this purpose. It also applies additional heuristics to map the landing locations to IPs for arbitrary submarine cables rolled out during the last year.

9. Discussion

Data availability. In this study we proposed a data-driven method that relies on a set of datasets that were public at the time of the study. However, as noted throughout our work at least two of these data (Telegeograph and Rapid7 datasets) are no longer publicly available. Information on the deployment of these cables is usually made public by the cable owners [21, 51]. We aim as future work to build a collection system for gathering information on submarine cables, and make the data available to the research and network operational community. At the same time, we plan to integrate in our inference method another publicly available data source that can substitute the Rapid7 dataset [52]. OpenIntel [53] is an active DNS measurements project that collects DNS data daily and makes the data available upon request [54]. As of 2020, the project also measures the IPv4 reverse DNS name space. ZDNS [55] is an fast open source measurement toolkit for fast DNS lookups. In their recent paper [52], the authors show that ZDNS can run PTR queries for the entire IPv4 address space in half of day. We thus believe that using DNS PTR records retried from these two active measurement project could replace the Rapid7 data.

Inferring submarine cable IP-level links in the wild. Our two-step proposed method relied on a series of datasets and heuristics to identify submarine cables links that map in the IP-level path data. Starting from a large set of IP paths we leverage PTR records, geo-hints and routing data to identify IP hops along the paths that could possibly match the end points of the submarine cable segments. After identifying such links, we further apply a set of four filters to remove links that are unlikely to match the submarine cable segments. While our inference method leverages existing approaches, it offers an automated method of finding IP links mapped to cables. Moreover, we analyze the impact on the data and control plane of the laid cables, showing that it can both increase and decrease performance. Also, it highlights cases when the cable owner does not appear on the path, which implies the need for developing further the inference method to include such cases.

Utility of our research findings. Computer scientists can utilize our research findings to understand better the expansion of the physical Internet infrastructure and better study the Internet traffic flow. The techniques we developed are limited to submarine cables and can be used for studies of long-haul terrestrial fiber cables. We plan to investigate this as part of our future work. Our findings are also helpful for scientists outside computer science alike. Engineers and mathematicians can use our results to model better the risk level of attacks, outages, and misconfigurations that will affect the smooth operation of submarine cables that are part of the Internet's critical infrastructure. Economists and policymakers can use our results to assess the “return on investment” for the roll-out of submarine cables. Our results show that the roll-out may significantly impact end-to-end delay; however, to our surprise, this is only sometimes the case due to the complex routing and business relationships on the Internet. Our results also provide unique insights on the state of the submarine cables today that can seed a debate on the best practices for sustainable and impactful deployment of submarine cables in the future.

10. Conclusions

We present our approach for translating public announcements for submarine roll-outs to useful information for networking research. Using multiple datasets, we develop a methodology and sketch the necessary and sufficient conditions to infer IP-level information from publicly available traceroute data archives. We showcase the efficiency of our approach by focusing on the submarine cables that were rolled out during the last five years. Furthermore, we use active measurements to validate approximately half of the inferred links. Our validation finds that up to 71% of the classifiable links can be mapped to a single submarine cable. We also comment on the impact that newly deployed submarine cables have on the end-to-end delay of users in connecting regions. Deploying some of the submarine cables appears to have an impact on the routing system. Specifically, for some cables we observe a slight increase in the number of networks that appear to connect through the cables.

As part of our future research agenda, we plan to expand our study to study new submarine cables as well as terrestrial cables. We also plan to incorporate additional measurements and meta-information in our methodology to address the challenges of inferring the submarine cable IP anchors, links, paths, and cable characteristics. Moreover, we plan to automate this study's validation process and incorporate it into our inferring submarine cable IP link approach.

Acknowledgments

Ioana Livadariu was partially funded by Simula Metropolitan internal funding, the Norwegian Research Council (grant #288744) and the RIPE NCC Community Projects Fund. The work of Georgios Smaragdakis was partially funded by the European Research Council under starting grant ResolutioNet (679158), the Dutch 6G flagship project "Future Network Services" with funding from the National

Growth Fund, and the European Union research and innovations programme Horizon Europe under grants SEPTON (101094901), MLSysOps (101092912), and TANGO (101070052).

References

[1] Organisation for Economic Co-operation and Development, Key ICT Indicators, https://www.oecd.org/digital/oecdkeyictindicators.htm (2021).

[2] International Telecommunication Union, Statistics - ITU, https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx (2023).

[3] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Holzle, S. Stuart, A. Vahdat, B4: Experience with a Globally-Deployed Software Defined WAN, in: ACM SIGCOMM, 2013.

[4] Google, Google Datacenters, https://www.google.com/about/datacenters/ (2021).

[5] R. Singh and M. Ghobadi and K.-T. Foerster and M. Filer and P. Gill, RADWAN: Rate Adaptive Wide Area Network, in: ACM SIGCOMM, 2018.

[6] Facebook, Facebook Datacenters, https://datacenters.fb.com/ (2021).

[7] ITU/UNESCO Broadband Commission for Sustainable Development, The State of Broadband 2021: People-Centred Approaches for Universal Broadband, https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.23-2021-PDF-E.pdf (2021).

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

[9] M. Trevisan, D. Giordano, I. Drago, M. Mellia, M. Munafo, Five Years at the Edge: Watching Internet from the ISP Network, in: CoNEXT, 2018.

[10] E. Pujol, P. Richter, B. Chandrasekaran, G. Smaragdakis, A. Feldmann, B. Maggs, K. Ng, Back-Office Web Traffic on The Internet, in: ACM IMC, 2014.

[11] T. Leighton, Improving performance on the internet, Commun. ACM 52 (2) (2009) 44–51. doi:10.1145/1461928.1461944.
URL https://doi.org/10.1145/1461928.1461944

[12] David-Brown, 10 Facts About the Internet's Undersea Cables, https://www.mentalfloss.com/article/60150/10-facts-about-internets-undersea-cables (2015).

[13] Z. S. Bischof, R. Fontugne, F. E. Bustamante, Untangling the world-wide mesh of undersea cables, HotNets '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 78–84.

[14] History of the Atlantic Cable & Undersea Communications, https://atlantic-cable.com/Cables/CableTimeLine/atlantic.htm, [Online; accessed 1-June-2021] (2021).

[15] Telegeography, Submarine Cable Map, https://www.submarinecablemap.com/, [Online; accessed 1-November-2021] (2021).

[16] Greg's Cable Map, https://cablemap.info/_default.aspx, [Online; accessed 1-June-2021] (2016).

[17] MAREA Cable System Reaches 200 Tbps of Capacity), https://www.submarinenetworks.com/en/systems/trans-atlantic/marea/marea-cable-system-reaches-200-tbps-of-capacity, [Online; accessed 1-November-2021] (2021).

[18] P. Gigis, M. Calder, L. Manassakis, G. Nomikos, V. Kotronis, X. Dimitropoulos, E. Katz-Bassett, G. Smaragdakis, Seven years in the life of hypergiants' off-nets, in: Proceedings of the 2021 ACM SIGCOMM 2021 Conference, Association for Computing Machinery, New York, NY, USA, 2021, p. 516–533.

[19] Telegeography Blog, This is What Our 2019 Submarine Cable Map Shows Us About Content Provider Cables, https://blog.telegeography.com/this-is-what-our-2019-submarine-cable-map-shows-us-about-content-provider-cables, [Online; accessed 1-June-2021] (2019).

[20] R. Fanou, B. Huffaker, R. Mok, K. C. Claffy, Unintended Consequences: Effects of Submarine Cable Deployment on Internet Routing, in: Passive and Active Measurement, Springer International Publishing, 2020.

[21] Submarine Networks, SACS submarine cable, https://www.submarinenetworks.com/systems/brazil-africa/sacs (2022).

[22] Telegeography: SACS, South Atlantic Cable System (SACS), https://www.submarinecablemap.com/submarine-cable/south-atlantic-cable-system-sacs, [Online; accessed 1-November-2021] (2021).

[23] R. Fontugne, C. Pelsser, E. Aben, R. Bush, Pinpointing delay and forwarding anomalies using large-scale traceroute measurements, in: Proceedings of the 2017 Internet Measurement Conference, IMC '17, Association for Computing Machinery, New York, NY, USA, 2017.

[24] RIPE Atlas, https://atlas.ripe.net/, [Online; accessed 1-June-2021] (2021).

[25] Telegeography, Tui-Samoa, https://www.submarinecablemap.com/submarine-cable/tui-samoa, [Online; accessed 1-November-2021] (2021).

[26] Archipelago (Ark) Measurement Infrastructure, http://www.caida.org/projects/ark/.

[27] CAIDA, Submarine cable impact analysis, https://github.com/CAIDA/submarine-cable-impact-analysis-public#readme.

[28] Telegeography, Monet, https://www.submarinecablemap.com/#/submarine-cable/monet, [Online; accessed 1-November-2021] (2021).

[29] Telegeography: SAm-1, South America-1 (SAm-1), https://www.submarinecablemap.com/submarine-cable/south-america-1-sam-1, [Online; accessed 1-November-2021] (2021).

[30] Telegeography: FA-1, FLAG Atlantic-1 (FA-1), https://www.submarinecablemap.com/submarine-cable/flag-atlantic-1-fa-1, [Online; accessed 1-November-2021] (2021).

[31] Telegeography: AC-1, Atlantic Crossing-1 (AC-1), https://www.submarinecablemap.com/submarine-cable/atlantic-crossing-1-ac-1, [Online; accessed 1-November-2021] (2021).

[32] Global RIPE Atlas Network Coverage, https://atlas.ripe.net/results/maps/network-coverage/, [Online; accessed 1-June-2021] (2021).

[33] CAIDA, Mapping Autonomous Systems to Organizations: CAIDA's Inference Methodology, https://www.caida.org/archive/as2org/, [Online; accessed 1-June-2021] (2021).

[34] Routeviews Project – University of Oregon, http://www.routeviews.org/, [Online; accessed 1-June-2021].

[35] RouteViews IPv4 Prefix to AS mappings - coalesced, https://catalog.caida.org/details/dataset/routeviews_ipv4_prefix2as_coalesced, [Online; accessed 1-June-2021].

[36] Rapid7: Open data, https://opendata.rapid7.com/, [Online; accessed 1-June-2021] (2021).

[37] Rapid7: Reverse DNS (RDNS), https://opendata.rapid7.com/sonar.rdns_v2/, [Online; accessed 1-June-2021] (2021).

[38] Rapid7: About Open Data, https://opendata.rapid7.com/about/ (2021).

[39] IATA, Airline and Location Code Search, https://www.iata.org/en/publications/directories/code-search/, [Online; accessed 1-June-2021] (2021).

[40] International Organization for Standardization (ISO, The International Standard for country codes and codes for their subdivisions, https://www.iso.org/iso-3166-country-codes.html, [Online; accessed 1-June-2021].

[41] GeoNames, Large cities, https://www.geonames.org/countries/, [Online; accessed 1-June-2021].

[42] Telegeography, Tasman Global Access (TGA) Cable, https://www.submarinecablemap.com/submarine-cable/tasman-global-access-tga-cable, [Online; accessed 1-November-2021] (2021).

[43] Alice-LG - Your friendly looking glass, https://github.com/alice-lg/alice-lg.

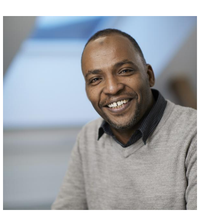
[44] V. Giotsas, A. Dhamdhere, k. claffy, Periscope: Unifying Looking Glass Querying, in: PAM, 2016.

[45] Telegeography: GTT, GTT express), https://www.submarinecablemap.com/#/submarine-cable/gtt-express, [Online; accessed 1-November-2021] (2021).

[46] Telegeography: , South East Asia–Middle East–Western Europe 5 Submarine Cable System, http://www.seamewe5.com, [Online; accessed 1-November-2021] (2021).

[47] Telegeography: AAE-1, Asia-Africa-Europe-1 (AAE-1), https://www.aaeone.com/aaeportal/, [Online; accessed 1-November-2021] (2021).
[48] Telegeography: APG, Asia Pacific Gateway (APG), https://www.submarinecablemap.com/#/submarine-cable/asia-pacific-gateway-apg, [Online; accessed 1-November-2021] (2021).
[49] S. Anderson, L. Salamatian, Z. S. Bischof, A. Dainotti, P. Barford, IGDB: Connecting the Physical and Logical Layers of the Internet, IMC '22, 2022.
[50] S. Liu, Z. S. Bischof, I. Madan, P. K. Chan, F. Bustamante, Out of Sight, Not Out of Mind: A User-View on the Criticality of the Submarine Cable Network, in: ACM IMC, 2020.
[51] Zayo, Explore Our Global Network, https://www.zayo.com/solutions/global-network/ (2018).
[52] L. Izhikevich, G. Akiwate, B. Berger, S. Drakontaidis, A. Ascheman, P. Pearce, D. Adrian, Z. Durumeric, ZDNS: A Fast DNS Toolkit for Internet Measurement, IMC '22, 2022.
[53] OpenINTEL, OpenINTEL: Active DNS Measurement Project, https://openintel.nl (2024).
[54] OpenINTEL, OpenINTEL: Data Access, https://openintel.nl/data-access/ (2024).
[55] The ZMap Project, zdns, 2024.

Ioana Livadariu is a Research Scientist at SimulaMet in Oslo, Norway since 2022. Prior, she worked towards completing her Ph.D on monitoring and understanding the IPv6 adoption from the University of Oslo. Her project was conducted in close collaboration with researchers from CAIDA (Center of Advanced Internet Data Analysis) at the University of California San Diego where she was a visiting researcher in 2013 and 2015. As part of her Ph.D project she conducted an in-depth empirical characterization of the IPv4 Transfer Markets in which she analyzed the reported IPv4 transfers from different perspectives, i.e., overall evolution, type, size and utilization of the exchanged IPv4 blocks, and the impact on the routing table and IPv6 adoption. Also, she proposed a method for inferring IP transfers from publicly available data. Her research also made significant contributions in estimating the deployment of Carrier-Grade NAT (NAT444) and on measuring the adoption of IPv6 in Africa. Her work has been published in top academic venues, including IEEE INFOCOM and ACM CoNEXT. She has produced an array of publicly available datasets as outcome of her research, which are being used by research groups around the world.



Ahmed Elmokashfi received the Ph.D. degree from the University of Oslo in 2011. He is a Research Professor with the Simula Metropolitan Center for Digital Engineering in Norway. Until 2022, he was the Head of the Center for Resilient Networks and Applications, which is part of the Simula Metropolitan Center, which is funded by the Norwegian Ministry of Transport and Communication. In particular, he focused on studying resilience, scalability, and evolution of the Internet infrastructure; the measurement and quantification of robustness in mobile broadband networks; and the understanding of dynamical complex systems. Over the past few years, he has been leading and contributing to the development, operation and management of the NorNet testbed infrastructure, which is a countrywide measurement setup for monitoring the performance of mobile broadband networks in Norway. His research interests lie in network measurements and performance. Here goes the biography details.

Georgios Smaragdakis is currently a Full Professor, the Chair, and the Section Head of Cybersecurity with the Faculty of Electrical Engineering, Mathematics, and Computer Science at Delft University of Technology (TU Delft). He is also a Researcher with the Max Planck Institute for Informatics (MPI-INF) and a Principal Investigator and a Fellow of the Berlin Institute for the Foundations of Learning and Data (BIFOLD). Since 2021, he has been a Scientific Coordinator of AI for peace, justice, and security initiative at TU Delft. His research interests include data- and measurement-driven approach to the study of the Internet security, resilience, state, and performance, and the enhancement of web privacy and security. Professor Smaragdakis is an ACM Distinguished Member and an IEEE Senior Member. His research was recognized with a European Research Council (ERC) Starting Grant Award, in 2015, a Marie Curie International Outgoing Fellowship, in 2013, Best Paper Awards at ACM SIGCOMM, in 2021, ACM IMC, in 2018, 2016, and 2011, ACM CoNEXT, in 2019 and 2015, IEEE INFOCOM, in 2017, three IETF/IRTF Applied Networking Research Prizes, in 2022, 2020, and 2019, "Best of ACM SIGCOMM Computer Communication Review," in 2019, and was selected for Communications of the ACM (CACM) Research Highlights, in 2021. He served as the Technical Program Chair for TMA, in 2020, ACM CoNEXT, in 2019, and PAM, in 2018. He has also been involved in the Organization and Technical Program Committees of many conferences, including, ACM SIGCOMM, ACM CCS, ACM IMC, ACM SIGMETRICS, ACM ASPLOS, ACM CoNEXT, ACM Web Conference, ACM Web Science, ACM HotNets, IEEE EuroS&P, IEEE Infocom, IEEE HotWeb, USENIX NSDI, USENIX ATC, PETS, ACM EuroSys, and ESORICS. George received the Diploma degree in electronic and computer engineering from the Technical University of Crete and the Ph.D. degree in computer science from Boston University in 2009. In 2008, he was a Research Intern at Telefonica Research. From 2008 to 2014, he acted as a Senior Researcher at the Deutsche Telekom Laboratories (T-Labs). From 2014 to 2017, he was a Marie Curie Fellow at the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), and a Research Affiliate with the MIT Internet Policy Research Initiative (IPRI), from 2015 to 2018. He was a Professor at TU Berlin, from 2017 to 2021, and a Research Collaborator with Akamai Technologies, from 2014 to 2021.