Unsteady Underwater - On the Constancy of Submarine Path Properties

Mia Weaver[†], Paul Barford[†], Fabián E. Bustamante[*], Esteban Carisimo[*], Weili Wu[‡], Lynne Stokes[°]
[†]University of Wisconsin-Madison [*]Northwestern University [‡]Georgia Institute of Technology [°]Southern Methodist University
mweaver6@wisc.edu, pb@cs.wisc.edu, fabianb@northwestern.edu, esteban.carisimo@northwestern.edu,
weili.wu@gatech.edu, slstokes@mail.smu.edu

***Abstract*—The world-wide submarine cable network (SCN) is one of the most costly, complex, and critical components in the Internet. In this paper, we describe a measurement study of packet dynamics on SCN paths. To conduct this study we built an active probe-based measurement system that harnesses looking glass nodes in strategic locations in service provider networks to gather data from 134 SCN paths. During the 18 month period of our study, we also tracked outage reports from online sources on a weekly basis. Our main finding is that key dynamic characteristics of the SCN including latency, loss, path stability, and outages vary widely. To provide perspective, we assess the operational constancy (*i.e.,* remaining within bounds considered operationally equivalent) of SCN path properties and find a wide range of behavior. To further clarify these characteristics, we report the results of a cluster analysis and find that paths separate into distinct groups where behavior is either relatively stable or highly variable. Our findings have implications for operational practices, applications that utilize these paths, and future measurement and monitoring of the SCN.**

I. Introduction

The worldwide submarine cable network (SCN) is one of the most costly, complex, and critical components of the Internet. Today's SCN consists of over 500 cable systems that vary widely in their physical deployments. It includes short-haul links spanning tens to hundreds of kilometers that connect communities in remote areas, to ultra-long-haul, intercontinental links that span thousands of kilometers and reach every major region in the world [1]. The SCN is at once vitally important to the global economy and essential for enabling sustainable growth in developing regions [2], [3].

Despite its importance, key behavioral characteristics of the SCN remain relatively opaque. While prior work has shown that web resources rely heavily on the SCN [4], to the best of our knowledge, the general aspects of packet dynamics over the SCN (*e.g.,* latency and loss) and path dynamics (*e.g.,* traffic engineering) have not been examined in detail. Understanding these characteristics has important implications for application performance, measurement-based protocols, operation of SCN infrastructure, and measurement-based study of the SCN.

Motivated by these observations, the objective of our work is to deepen our understanding of SCN path properties through empirical study. Our specific objective is to characterize the dynamic properties of a diverse set of SCN paths on timescales that are relevant to applications and operations[1] (*e.g.,* bulk data transfers, which may not achieve optimal performance over intercontinental links [5]). An important focus of our analysis is identifying the *operational constancy* of SCN paths where "the key criterion in operational constancy is whether an application (or other operational entity) would care about the changes in the dataset" [6].

To conduct our study we built an active probe-based measurement system for SCN infrastructure. Developing such a system has several significant technical challenges including identifying vantage points (VPs) to launch measurements and targets for traceroute probes such that the measurements traverse pairs of IP addresses spanning submarine infrastructure. Throughout this paper we refer to these as *SCN paths*, since the path between these IP address pairs could include infrastructure hidden from our probes (*e.g.,* layer 2 hops). Our measurement system is composed of looking glass nodes (LGNs) in service provider networks that enabled us to conduct active probing campaigns on 134 distinct SCN paths.

We report our findings on dynamic properties of SCN paths based on data collected with our LGN-based measurement system over a period of 18 months from July '22 through January '24. A high-level examination of measurements of packet loss, latency, and path stability shows that behavior varies widely across SCN paths. Specific to operational constancy, we find that change-free regions (CFRs - periods of time over which a path property remains steady) vary widely in duration from one day to nearly a month. Furthermore, by monitoring on-line sources during our study, we identified 31 outage events that impacted 59 submarine cables. Through careful analysis we show that the observed variability in our measurements can sometimes be the result of outage events, though such events on SCN paths are relatively uncommon.

To understand the characteristics of SCN path properties in greater detail, we perform a cluster analysis on the time series data collected from the monitored paths. We find that SCN paths generally fall into 6 behavioral clusters (§IV). The largest of the clusters includes 54 of the monitored paths, which exhibit stable behavior. Applications traversing these paths should experience consistent conditions. In contrast, several of the clusters include paths that exhibit frequent

[1]Fine-grained measurements on timescales relevant to *e.g.,* individual TCP connections are not an objective of our study.

ISBN 978-3-903176-72-0

changes in conditions (*i.e.,* low operational constancy), high latency variability and high packet loss rates, which would all adversely affect application performance.

In summary, this paper makes several contributions to the body of work on understanding the SCN. First, we describe the key components of an LGN-based system that we use for active probe-based measurement of 134 SCN paths. Next, we report findings on the dynamic properties of SCN paths using data collected over a period of 18 months. Finally, we show that (*i*) dynamic properties vary widely on SCN paths, (*ii*) outages, though uncommon, can explain some of the observed variability in our data, and (*iii*) a noteworthy percentage of paths are only operationally constant over a period of days. Our findings have implications for applications that traverse the SCN, for operational practices, and for the design of future SCN monitoring systems.

II. Measuring the SCN

A. Vantage Point Selection

Our goal in VP selection is to measure a set of SCN paths that are a representative sample of the diversity of submarine deployments across continents, countries, regions, and networks. Network elements invisible to layer 3 probes can introduce additive random noise to measurements. We aim to minimize this noise by selecting VPs geographically and topologically close to target SCN paths. Ideally, a VP is hosted by the owner of the submarine cable of interest or in one of its immediate AS-level neighbors.

Public looking glass *infrastructure* (LGI) has traditionally been deployed by large service providers to allow diagnostic visibility on their networks. LGIs are composed of *nodes* that are located at major network Points of Presence (PoPs) (such as demonstrated by Twelve99's LGI [7]). We expect LGNs to be geographically and topologically close to the ingress points of SCN paths due to their importance in world-wide communication. We also expect that traceroute measurements from an LGN located near an SCN path to an IP address target in the same network *on the opposite side of the SCN path* will never exit that network and will utilize the SCN path directly, making LGNs suitable VPs.

To identify the specific LGNs to use as VPs in our study, we (*i*) select networks with access to submarine infrastructure that also host LGI (or peer with LGI hosts) and (*ii*) determine the best VP locations for direct measurement of the SCN paths using publicly available network maps. Since intercontinental carriers differentiate themselves from competitors with distinctive network footprints and their exclusive access to a resource (*e.g.,* right to use a submarine cable), many carriers make their network maps publicly available to promote their services. We identify maps of networks hosting LGI through online search and reference other sources such as Infrapedia [8] and Telegeography's SCN map [1] which we use to identify candidate LGI. Then, we overlap submarine cables in these network maps with their offered LGNs to identify VPs nearest to each cable's landing points. If there are no accessible LGNs for a given network, we use LGNs from a peering AS, chosen to minimize potential noise in the collected measurements.

In total, we identified 90 LGNs from 6 LGIs for use as VPs in our study. 85 (63%) of the SCN paths monitored had VPs at the SCN path ingress and 106 (88%) had VPs within 5ms round trip time (RTT) to the SCN path ingress.

B. Selecting Measurement Targets

To create a list of IP address targets for our measurements, we first launched traceroutes from the selected VPs to destinations that must be connected via underwater infrastructure. We meet this condition by applying the drivability concept from [4] that uses the absence of a contiguous terrestrial path between two points as a first step to detecting paths traversing the SCN (*e.g.,* a VP in Japan to a target in the US). These traceroutes reveal the next hop from the selected VPs. If a next hop is in the proximity of the VP and is owned by the same network, its IP address can be used as a target destination for that network and landing point location. We confirm proximity to VPs by observing the geolocation hints embedded in DNS pointer records and verify that the RTT to the hop is in a range expected of co-located nodes (*i.e.,* $\leq$1ms). It is not uncommon that an LGN is situated at the ingress point of an SCN path such that the next hop is not in the proximity of the VP but rather is the egress point of an SCN path. In this case, we either use exploratory traceroutes directed towards the VP location to identify a potential IP address match or exclude that location as a measurement destination for that network. The SCN egress points identified via these exploratory traceroutes are also collected and used as measurement targets.

C. Implementing the Measurement System

VP – target IP pairs are used to directly measure SCN paths in 6 ASes and indirectly measure SCN paths in 3 peer ASes. A breakdown of paths, networks, and countries covered by our system is summarized in Table I. Our system monitors 134 geographically diverse SCN paths of varying lengths, connects 48 countries across 6 continents, and includes infrastructure operated by major cable providers. A map of our coverage is shown in Figure 1. We posit that the scope, scale, and diversity of our measurement system is sufficient to provide broadly applicable insights on the SCN.

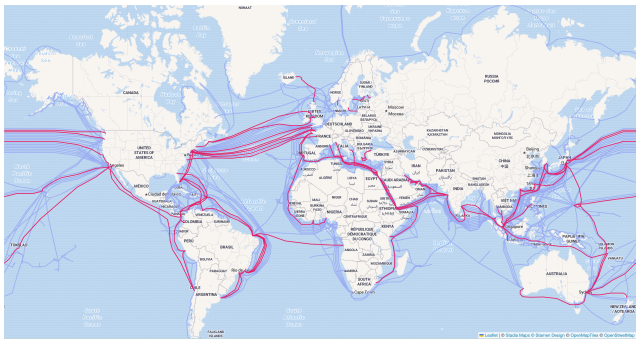

Figure 1: SCN coverage of our measurement system. Red lines indicate SCN paths monitored by our infrastructure.

The use of LGNs for SCN measurement presents several challenges. First, LGNs are typically implemented on routers

and may introduce additional packet delay to measurements. We expect that this noise will be relatively constant on top of the dynamic characteristics we measure and do not attempt to remove it from our measurements. Next, LGIs do not provide users with a standard interface for launching measurements [9] or displaying traceroute output, so samples from different networks must be collected and parsed on a case-by-case basis.

We wrote software to programmatically initiate traceroutes between VP – target IP pairs using LGI web-based interfaces.[2] We configured our system with a period of 10 minutes between traceroute measurements on an SCN path, consistent with prior active probe-based studies (*e.g.,* [10]). This probe rate is examined in more detail in (§III-C).

AS	# Countries Connected	# SCN Paths	# Traceroute Samples	Longest Path (ms)
NTT	8	19	669,665	142.0
GCX	13	26	1,293,623	143.1
HE	13	19	257,049	81.0
Telecom Italia	6	6	327,906	106.6
GlobeNet	4	10	589,531	71.5
Arelion	10	9	434,968	72.8
Orange	5	4	171,074	101.0
Level3	9	38	505,213	76.8
Telstra	2	2	15,421	39.3
AngolaCables	2	1	18,210	61.1

Table I: Coverage of our measurement system where longest path is the minimum RTT of the longest SCN path monitored per AS.

Our measurement campaign was initiated on *July 1, 2022*. We discarded erroneous traceroutes (*e.g.,* include bad characters, all packets in a hop are dropped such that the presence of an SCN path cannot be inferred, target IP address is not reached). In the case of a single hop measurement, the VP is also the entry point of the SCN path, and the ingress hop is considered as having an RTT of 0. We report results on over 4.2 million successful traceroutes collected through *January 1, 2024*. Our measurement system uses the egress points of SCN paths as traceroute targets, and we map SCN paths to the last two hops in a traceroute measurement. From these hops, we extract the IP addresses, geolocation data, and RTT values of the SCN ingress and egress points.

D. Dynamic Properties in the SCN

Using our data corpus, we focus on characterizing packet loss, latency, and path change behaviors in the SCN. We define key terms relating to our analyses as follows:

SCN path. The last hop(s) of a traceroute collected using a unique VP – target IP pair, where the final hop is associated with the egress point of the SCN path and the second-to-last hop (or VP) is associated with the ingress point.

Latency. For each traceroute, we compute latency by taking the minimum of the egress hop RTT trio and subtracting from it the minimum RTT value associated with the ingress router across *all* measurements of the path over the full 18-month measurement campaign.

Packet loss. For each traceroute sample on a path, we extract the number of packets dropped at the egress and construct a raw loss time series. On some paths, at least one packet was dropped from every traceroute probe trio. We attribute this to ICMP rate limiting by the router [11], [12]. To account for this, we convert the raw loss measurements into a binary time series that only registers a loss when the number of packets dropped is greater than the minimum over the entire time series. The ratio of data points observing packet loss is called the *packet loss rate* (PLR).

Path change. It is possible for there to be oscillation at the network layer with multiple IP addresses responding at the SCN path ingress. We process all measurements of an SCN path and collect the set of ingress IP addresses encountered. We identify samples utilizing the most prevalent ingress IP address and mark all others as path changes. The result is a binary time series where a value of '1' represents the presence of a path change. The ratio of data points during a path change, called the *path change rate* (PCR), is a measure of stability.

E. Operational Constancy of SCN Paths

A key reference for our study is the seminal paper by Zhang *et al.* [6], which explores the constancy of Internet path properties and introduces the notion of *operational constancy*. A network measurement dataset exhibits *operational constancy* if the quantities of interest remain within bounds considered operationally equivalent (*e.g.,* loss rates between 5-10%). Zhang *et al.* analyzed operational constancy on end-to-end paths for packet loss, latency, and throughput measurements from distributed end-hosts[3].

We follow these methods for evaluating operational constancy in the SCN, identifying the CFRs of each SCN path's time series representations of packet loss and latency. A CFR is the number of consecutive interval periods during which the PLR or average latency remains within bounds considered "operationally equivalent." We refrain from comparing our results directly to the findings of Zhang *et al.* since their measurements were made at significantly higher rates (10Hz/20Hz). In what follows, we use the terms *constancy*, *steadiness*, and *operational constancy* interchangeably.

During our study, there were occasional LGN web-page errors, script errors, and network outages that disrupted our measurements. Thus, to conduct our constancy analysis, we select a month-long snapshot during which the largest number of SCN paths have sufficient, continuous data. We exclude any datasets without sufficient data during this time frame. The CFRs reported in this section were calculated using measurements collected on 117 SCN paths from November 13, 2022 to December 14, 2022. When there is an interval during which no measurements are collected on a path, we split the current CFR. Using 24-hour intervals, the average number of CFR splits due to lack of data is 0.31 per path.

[2]We initially attempted to use CAIDA's Periscope [9] to schedule our measurements, but found that the platform did not support the interfaces of many of our selected VPs.

[3]A recent study by [13] provides a reassessment of constancy in end-to-end latency utilizing RIPE Atlas measurements, though our study differs in our focus on the SCN using LGNs.

To calculate packet loss CFRs (PLR CFRs) on an SCN path, we compute the PLRs of traceroute samples over 12- and 24-hour intervals. The PLR of an interval period falls into one of these bins: [0 - 0.5%), [0.5 - 2%), [2 - 5%), [5 - 10%), [10 - 20%), [20 - 100%]. Consecutive intervals with PLRs in the same bin are aggregated into a single CFR (*e.g.,* a CFR of 3 days). Due to our coarse sampling rate, our analysis of PLR CFRs differs slightly from that of [6] in that we do not aggregate consecutive packet losses into single *loss events*. This choice is guided by the assumption that "back-to-back losses" separated by 10 minutes do not necessarily correspond to the same burst.

Next, we extract latency CFRs (LAT CFRs) by averaging latencies over 12- and 24-hour intervals before sorting each interval into one of the following bins: [0 - 100ms), [100 - 200ms), [200 - 300ms), [300 - 400ms), [400 - 1000ms+], where consecutive intervals in the same bin are aggregated into a single LAT CFR.

To highlight the variation in constancy observed in the SCN, we will present the ratio of PLR and LAT CFRs that persist for $\approx$ 90%, 60%, and 30% of the analysis time frame, or 28 days, 19 days, and 10 days. We report also the proportion of minimal duration CFRs (considered as 1 day when using 24-hour intervals or $\leq$ 1 day when using 12-hour intervals).

F. Identifying Anomalous Behaviors in the SCN

While it is possible that any observed unsteadiness is a result of outages and other noteworthy events, we expect these to be relatively uncommon. To validate this, we identify the real-world events observable in the features captured by our measurement system by collecting cable failure reports and announcements of *specific* network outages as identified by sub-tel forums and articles, ISP social media posts, ISP outage reports, and Cloudflare reports. We also reference reports of real-world events (*i.e.,* natural disasters and extreme weather) accompanied by location-specific (but not ISP-specific) outage reports. We then cross-reference our time series plots to visually identify these events in our data, which appear as disruptions to standard path behavior.

We also identify events using SIFT [14], a tool which parses results from Google trends queries to identify spikes in searches for specific key words such as "Outage" during a specified time frame. Time series events identified via SIFT cannot necessarily be tied to a widely-reported event, but can still be correlated with outages.

III. The Operational Constancy of SCN properties

A. Feature Constancy

Packet loss, at a glance, appears unsteady. When using 24-hour intervals, 44.5% of CFRs are of the minimal duration, 1 day, and only 9.6%, 16.5%, and 25.7% persist for $\geq$ 28 days, 19 days, and 10 days respectively. A deeper look, however, reveals that 40.2% of the sampled paths contain a PLR CFR of at least 28 days, while only 12% of paths contain a maximum length CFR of $<$ 10 days, suggesting that a subset of paths are responsible for the apparent unsteadiness of packet loss.

We repeated the analysis using 12-hour intervals to uncover any additional characteristics of packet loss constancy not apparent over longer periods. Only 6.3%, 9.6%, and 17.0% of these CFRs persist for $\geq$ 28 days, 19 days, and 10 days respectively. We conclude that a shorter interval reveals a larger number of SCN paths as persistently unsteady. However, 37.6% of the sampled SCN paths still contain a CFR of at least 28 days, further suggesting that the steadiness of packet loss in the SCN varies considerably.

Latency in the SCN achieves a higher degree of steadiness than packet loss. As many as 42.7%, 60.1%, and 81.5% of latency CFRs are $\geq$ 28 days, 19 days, and 10 days respectively, with 65.0% of the sampled paths containing a CFR of at least 28 days. However, minimal duration CFRs still make up 7.3% of CFRs calculated.

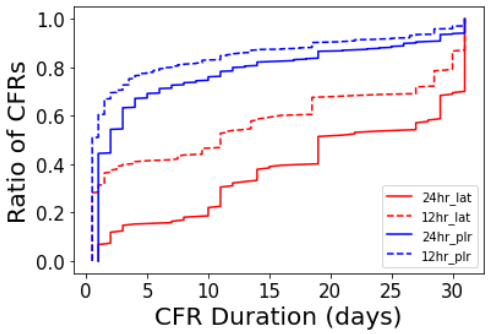

Figure 2: CDFs of CFRs for packet loss (plr, blue) and latency (lat, red) broken down by interval lengths: 12 (dashed), 24 (solid) hours.

Decreasing the interval to 12 hours predictably results in reduced steadiness. Still, as many as 28.0%, 32.3%, and 53.3% of SCN paths contain a CFR of $\geq$ 28 days, 19 days, and 10 days respectively. Though, 31.5% of latency CFRs are 1 day *or less*. We investigate whether this unsteadiness is consistent throughout the sampled paths in (§IV) and display all CFR distributions in Figure 2.

B. Assessing Anomalous Behaviors

We manually identified 59 significant anomalous behaviors in our time series plots that correspond to 31 real-world events, several of which span multiple cable systems. Many of these events correspond to reports of cable failures such as the example in Figure 3, which shows a segment of SEA-ME-WE 4 crossing the Suez Canal during two failures on nearby cable SEA-ME-WE 5. During the first failure on August 18, 2022, there is a level shift in latency and a transient path change. During the second event on November 30, 2022, there is another transient path change followed by latency spikes on the prevalent path. This could be the result of increased traffic during the severance of SEA-ME-WE 5 [15], [16].

A portion of FNAL/RNAL connecting Japan to South Korea exhibited latency spikes from Aug 9, 2022 until Aug 20, 2022 during a period of extreme weather that resulted in widespread outages in Japan [17]. This is shown in Figure 3.

We make no claim that this is the complete set of anomalous events that occurred on the SCN paths we monitored. However,

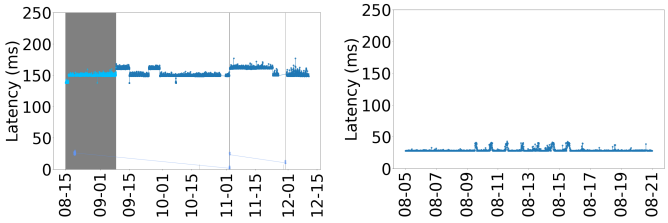

Figure 3: (Left) SEA-ME-WE 4 during damage of nearby path on 8/18/2022, 11/30/2022. Path changes (gray vertical lines, with path change latencies plotted in lighter color) and increased latency likely due to increased traffic. (Right) Latency spikes on SCN path connecting JP and KR during Aug. 2022 storms in JP.

the identification of these events is useful for understanding the effects of outages on dynamic properties of SCN paths, though the infrequency of these events indicates that operational practices and changing user demands are more likely the cause of the observed unsteadiness.

C. Higher Frequency Sampling

We assess the extent to which our coarse-grained sampling explains the observed unsteadiness, selecting two SCN paths for additional, higher-frequency probing: one with low latency variation and one with comparatively high latency variation. We examine the behaviors captured at sampling rates of once every minute and once every two and a half minutes for one hour daily over a week and compute the autocorrelation associated with each sampling rate.

The low variation exemplar, Skagen Fiber West Cable, connecting Norway to Denmark (170km in length), had estimated first-order autocorrelations of 0.005, 0.04, and -0.002 from the 1, 2.5, and 10 minute interval datasets respectively. The p-values for tests of both positive and negative autocorrelation indicate no evidence of autocorrelation; distributions calculated using high frequency sampling were similar to those calculated with low frequency sampling. We removed a single outlier value $>$ 500ms from the coarse-grained data; with this value removed, the largest latency is 11ms.

The autocorrelation statistics for our high variation exemplar, BCS North – Phase 2, connecting Finland and Russia (280km in length), were consistent with this finding, with estimated first-order autocorrelations of -0.023, -0.003, -0.009 from the 1, 2.5, and 10 minute interval datasets respectively. Again, the p-values for tests of both positive and negative autocorrelation based on these statistics exceed 0.20 for all three series, indicating no evidence of autocorrelation. We conclude that higher frequency sampling does not reveal any additional qualities of standard behavior[4].

IV. Clusters of Constancy

To gain a broader understanding of operational constancy in the SCN, we applied unsupervised learning to categorize the SCN paths into clusters based on their dynamic characteristics and assess the constancy within each cluster. Through this analysis, we (*i*) confirm that steadiness is not consistent throughout the set of monitored paths and that only a subset of SCN paths are responsible for the unsteadiness we observed and (*ii*) identify key predictors of unsteady behavior.

A. Dynamic Behaviors in the SCN

To identify sources of unsteadiness in the SCN, we first examine the dynamic behaviors of the 134 sampled paths, focusing on latency standard deviation (LAT DEV), PCRs, and PLRs calculated during the full measurement campaign. We seek to identify how SCN paths differ on these axes so that we can develop metrics to differentiate paths. The observations reported here guided our development of the clustering process.

Latency. Latency variation is a key characteristic in determining a baseline behavior since it establishes the expected latency bounds that each incoming measurement is likely to fall within. Just over half of the 134 SCN paths monitored exhibited a LAT DEV of $<$ 5ms, and 67.7% were under 10ms. These paths experienced few or no bursts or long-term changes in latency. However, a noteworthy subset of SCN paths, 12.8%, exhibited a LAT DEV of $\geq$ 20ms. These paths are potential culprits for unsteadiness in the SCN and should be isolated in the clustering analysis.

Packet loss. The majority of sampled paths reliably carried packets to their egress routers over longer time periods, with 86.5% of the paths exhibiting a PLR of $\leq$ 1%. However, 5.3% of SCN paths exhibited a PLR of $\geq$ 50%, suggesting that a small but noteworthy subset of SCN paths are subject to routing policies and/or outside factors that result in the routine dropping of packets. Such a behavior is indicative of high traffic demands or operational practices that obscure SCN path properties from our VPs. Thus, we expect that clustering analysis should also isolate these paths.

Path changes. As many as 60.2% of SCN paths have a PCR of $\leq$ 0.5%, suggesting that many of the SCN paths are highly stable in regards to path changes. However, 15% of SCN paths exhibit a PCR of $\geq$ 30%. A high frequency of path changes on SCN paths does not necessarily indicate a change in infrastructure but rather may be caused by traffic engineering. In cases where a path regularly and evenly alternates between ingress points in the same AS and location, and that are equidistant from the VP, we assume that these alternating ingress points are interfaces of the same router. We also identify long-term path changes that do not correlate to a change in baseline latency or packet loss behavior and infer that such instances are similarly indicative of interface changes on an ingress router. This finding raises the question of whether high path change rates correspond to reduced steadiness in latency or packet loss.

Assumptions must be made about whether or not a transient path change traverses the same infrastructure based on other behavioral characteristics. Short term path changes corresponding to a change in baseline behavior are likely due to a change in infrastructure. Such behaviors should not

[4]Much higher probe rates would be required to observe detailed features of congestion events [18].

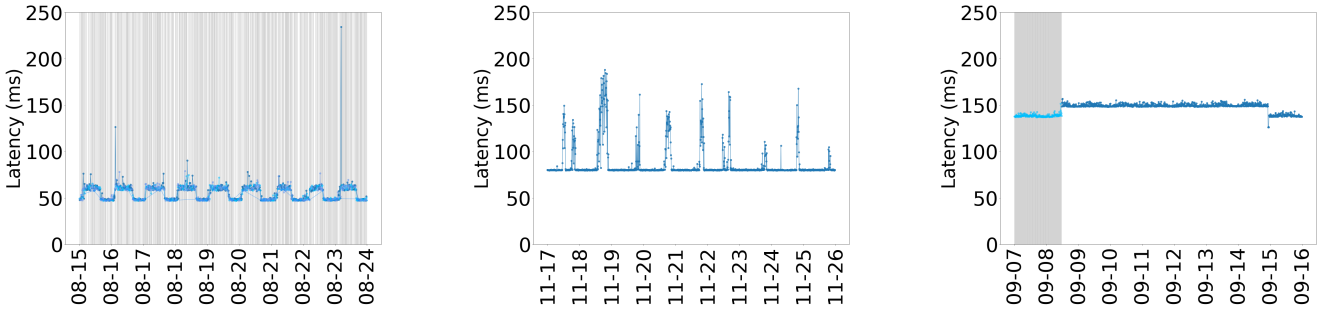

Figure 4: SCN behaviors; prevalent path latency plotted in darker color; vertical shading indicates path changes; path change latencies plotted in lighter color. (Left) Diurnal routing pattern in latency and alternating ingress interfaces on segment of FLAG Europe Asia [HK → JP]. (Center) Latency bursts on Yellow AC-2 [US → GB]. (Right) Path changes and latency level shifts on SEA-ME-WE 4 [SG → FR].

be included in the cluster of paths exhibiting long-term path changes or traffic engineering, and will likely exhibit PCRs in between the two extremes described in the previous paragraph.

Figure 4 shows time series highlighting common behaviors on SCN paths such as traffic engineering, congestion events, and a potential change in infrastructure.

B. Time Series Clustering

To gain deeper insights into sources of unsteadiness, we conducted a cluster analysis using measurements collected from the beginning of our measurement campaign through October 6, 2023 on the SCN paths used in §III. We use the standard *K-means* technique, implemented with the `sklearn.cluster` library. This algorithm was selected for clustering due to its simple implementation and wide use in empirical studies.

Disadvantages of K-means are the requirement for the user to specify the number of clusters as an input parameter, k, and its lack of robustness to outliers. We use a principled approach to select the number of clusters called the *elbow method*, which uses the decrease in cluster spread yielded by increments of k to determine the optimal value [19], [20]. The cluster spread associated with a value k is calculated as the *within-cluster sum of squares error (SSE)* summed across the resulting clusters.

For each SCN path, we coalesce the standard dynamic properties of network paths that have been studied for decades [21]: (*i*) PCR, (*ii*) normalized LAT DEV, and (*iii*) PLR. We also include a measure of (*iv*) PLR constancy and (*v*) latency constancy to ensure that clusters will be based in part on steadiness. We measure constancy by dividing the number of CFRs (using 24-hour intervals) by the number of days in the clustering time frame, giving us CFR *split rates* where mild split rates correspond to higher steadiness.

Each value in the feature set is normalized such that the minimum and maximum values are 0 and 100. We call a value in the range [0, 10) "mild" (*e.g.,* mild PCR, mild LAT CFR split rate, etc.), and values in ranges [10, 30), [30, 50), and [50, 100] "moderate", "high", and "extreme" respectively.

We identified an outlier SCN path with a LAT DEV of $\geq$ 60ms. The inclusion of this SCN path in the normalization process has a compressing effect on the remainder SCN paths' LAT DEVs, so we excluded the outlier during clustering. From the remaining 116 SCN paths, we identify 6 clusters, providing a succinct representation of the diversity of dynamic characteristics observed. When discussing LAT DEV within the clusters, we will report on the values *before* normalization. A summary of each cluster is provided in Table II.

ID	Paths	PCR	LAT	PLR	PLR CFR	LAT CFR
0	54	○ - ◖	○	○	○ - ◖	○
1	7	○ - ◖	○ - ◖	●	○	○ - ◗
2	16	◗ - ●	○ - ◖	○	○ - ◖	○ - ◖
3	8	○	◖ - ◗	○	○ - ◗	○ - ◖
4	18	○ - ◖	○ - ◖	○	○ - ◖	○
5	13	○ - ◖	○ - ◖	○ - ◖	◖ - ●	○ - ◖

Table II: Number of paths and characteristics of each cluster. PCR, LAT DEV, PLR, PLR CFR split rate, and LAT CFR split rate. Categories: mild (○): [0, 10), moderate (◖): [10, 30), high (◗): [30, 50), Extreme (●): [50, 100].

Cluster 0. This cluster contains approximately 46% of the SCN paths included in the clustering analysis and exhibits mild/moderate PCRs as well as mild LAT DEVs and PLRs. LAT and PLR CFR split rates are relatively low in this cluster.

Cluster 1. SCN paths in this cluster exhibit PLRs of $\geq$ 90% and mild PLR CFR split rates, suggesting that PLRs are consistently high. There is a wide spread of LAT DEV values, and LAT CFR split rates range from mild to high. This is the smallest cluster, comprising of only 7 SCN paths.

Cluster 2. SCN paths in this cluster have a PCR of $\geq$ 30% and exhibit one of two behaviors: oscillations between $\geq$ 2 ingress router export interfaces or long-term path changes. These paths exhibit mild to moderate LAT DEVs and mild to moderate LAT and PLR CFR split rates. PLRs within this cluster are mild. This cluster includes 16 SCN paths.

Cluster 3. The 8 SCN paths in this cluster have the highest variation in latency, but minimal PLRs and PCRs. LAT and PLR CFR split rates here are higher than in cluster 0, indicating low operational constancy on both axes.

Cluster 4. This cluster exhibits higher LAT DEVs than cluster 0, but lower values than cluster 3, as well as mild PLRs and PCRs. The LAT and PLR split rates are comparable to those in cluster 0. This cluster includes 18 SCN paths.

Cluster 5. The 13 paths in this cluster are characterized by short-lived PLR CFRs, as well as mild to moderate LAT DEVs, PCRs, and PLRs. Despite the observed unsteadiness in packet loss, LAT CFR splits in this cluster are rare.

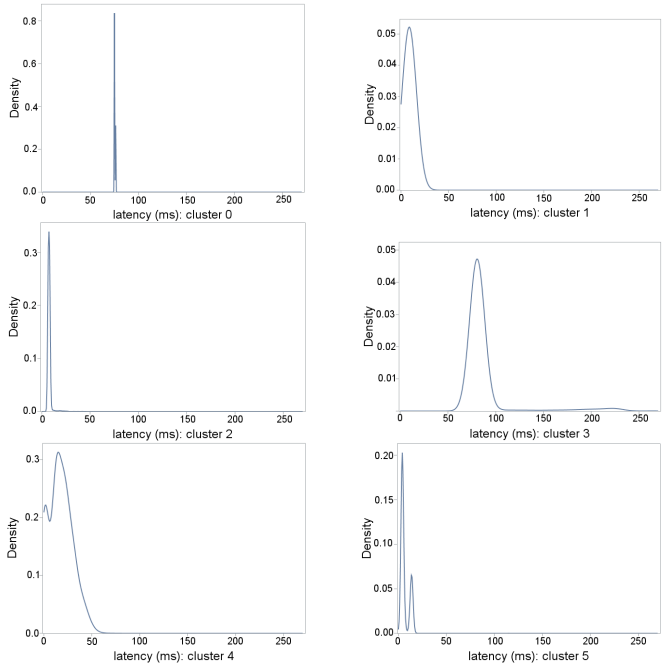

Figure 5: Latency densities of exemplar SCN paths from cluster 0 (Top Left), 1 (Top Right), 2 (Middle Left), 3 (Middle Right), 4 (Bottom Left), and 5 (Bottom Right).

C. Latency Density of SCN Path Clusters

For further perspective on cluster characteristics, we present estimates of latency density of cluster exemplars that were selected based on proximity to cluster centroid and number of samples collected. Our estimations were calculated using the following equation: $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\phi_h(x - X_i)$, where X_i denotes the n sample values and $\phi_h(...)$ is the density of a normal random variable with a mean of 0 and a LAT DEV value of h. The value of h is chosen using the formula in [22]. Our results are shown in Figure 5.

D. Constancy in SCN path Clusters

To confirm that the unsteadiness identified in §III is indeed due to a subset of the SCN, we repeat our constancy analysis, this time organizing CFRs by cluster. We find that the duration of LAT and PLR CFRs vary significantly across clusters and are not consistent throughout the SCN. The CFRs reported in this section were calculated over 24-hour intervals to filter out less severe disturbances. We display and summarize our results in Figure 6 and Table III respectively.

Cluster 0 is relatively steady on both axes, particularly latency. Cluster 1 achieves an even higher degree of packet loss steadiness than cluster 0, though it also exhibits the highest PLRs and the *lowest* degree of latency constancy. From this, we conclude that packet loss steadiness does not necessarily correlate to latency steadiness, and that high PLRs may correlate to reduced latency steadiness.

This is further indicated in cluster 5 which achieves a high degree of latency steadiness despite the majority of its PLR CFRs being of the minimal duration. This indicates that short-lived disturbances on these paths are typical. This unsteadiness presents challenges for monitoring, measurement-based protocols, and operators that seek to provide a consistent service. The density plot of the cluster 5 exemplar reveals the bimodal latencies observed on the path, as shown in Figure 5. We infer the slightly higher latency values to be the result of mild congestion, occurring in tandem with packet loss bursts that disrupt standard behavior.

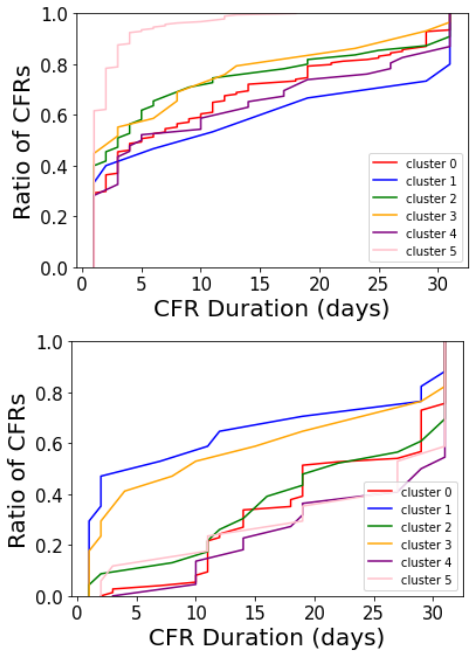

Figure 6: (Top) PLR, (Bottom) LAT CFRs by cluster identified in IV-B.

Cluster 3 illustrates the link between high latency variation and reduced steadiness on both axes. High LAT DEVs in this cluster are spurred by one of two phenomena: (*i*) infrequent, unstable, and extreme latency bursts disrupting otherwise stable behaviors, and (*ii*) sustained high variation in latency that still remains within stable bounds, exhibiting a relatively steady moving average. The high latency variation and moderate LAT CFR split rates in this cluster are indicative of longer-lived disturbances than observed in other clusters.

Packet loss in cluster 4 is less steady than in cluster 0 but more steady than in cluster 3. To understand cluster 4 packet loss in more detail, we consider its LAT DEV values, which fall between the maximum value observed in cluster 0 and the minimum value observed in cluster 3. This cluster encompasses a set of "mid-tier" LAT DEV values, further strengthening the conclusion that higher variation in latency correlates to less steady latency *and* packet loss.

Cluster 2 is characterized by high PCRs. It exhibits relative steadiness on both axes, with LAT and PLR CFRs slightly

shorter than those in cluster 0. Thus, a higher PCR is not necessarily a source of unsteadiness. This suggests that, in the face of congestion, paths in cluster 2 achieve low latency variation likely due to traffic engineering, an operational practice meant to accommodate high demands on a router.

Unsteadiness is most notable in clusters 1, 3, and 5, which exhibit high PLRs and/or latency variation, comprising approximately 24% of the 116 monitored paths used. Due to their relatively low latency variation, PCRs, and PLRs, as well as their high degree of steadiness, paths in cluster 0 could be considered as exhibiting the ideal behavior, providing reliable, steady service. Given that paths in cluster 0 comprise nearly 50% of the monitored paths used in the clustering process, and that paths from clusters 2 and 4 also achieve a high degree of steadiness, we confirm that the observed unsteadiness is indeed due to a subset of the SCN exhibiting high latency variation and PLRs, which we have isolated with our clustering process.

ID	28 days		19 days		10 days	
	PLR	LAT	PLR	LAT	PLR	LAT
0	14.3	47.3	27.9	66.2	40.9	94.6
1	31.3	27.8	37.5	33.3	50.0	44.4
2	14.3	41.7	21.4	58.3	28.6	83.3
3	10.0	27.8	16.7	38.9	30.0	50.0
4	17.0	56.5	29.8	73.9	46.8	95.6
5	0.0	44.4	0.0	72.2	3.8	83.3

Table III: % of cluster paths with CFRs of $\geq$ 28, 19, and 10 days.

V. Discussion

The biggest hurdle in deploying a comprehensive SCN monitoring system – which is our long-term goal – is the lack of available and easily accessible measurement infrastructure. Although some LGNs used in our study are quite near submarine landing points, many ISPs have LGNs deployed in only a subset of their geographic locations. For ISPs that do not host LGNs, in some instances we were able to use LGNs operated by peer ISPs to monitor SCN paths that they utilize. However, using LGNs to monitor SCN paths outside of a host network can introduce unwanted variability into measurements. Furthermore, SCN paths monitored from external networks are subject to operational practices reducing visibility (*e.g.,* packets from external networks are routinely dropped). This points to the long-term need for more comprehensive deployment or access to infrastructure that can be used for the express purpose of SCN monitoring.

Insights on dynamic path properties are inherently limited by sampling rate. While we show that sampling at a rate of one measurement per minute would not have a material impact on our findings, some phenomena of interest (*e.g.,* congestion events and short-lived TCP transactions) could be revealed in greater detail with higher frequency sampling. This capability would likely exceed what is available through standard LGNs and require dedicated infrastructure.

Our findings show that a significant subset of SCN paths exhibit a relatively high degree of constancy. This bodes well for applications that traverse these paths and is indicative of the effectiveness of the operational practices used to maintain them. This suggests that it may not be difficult to find providers that can deliver service over SCN paths that support good application performance, and that there may be opportunities to improve configurations and management practices on unsteady SCN paths.

Our findings point toward an opportunity for detecting anomalous events on SCN paths using cluster *signatures* (*i.e.*, a vector of the five path features defined in §IV-B averaged across the cluster's paths). Baselines and thresholds could be established for each feature based on the range of values observed within the cluster. When conditions on a path push its measured properties outside of these thresholds, we would consider this as an *anomalous event*. We leave a detailed study of anomaly detection based on these signatures to future work.

Finally, there are ethical implications of utilizing LGIs that are made available to service providers to enhance network operations. We alerted the owners of the LGIs used in this study to our interest in utilizing their infrastructure for research. We either received a positive response or no response from those providers. We maintain low sampling rates so as to not flood the LGNs with traffic. For the same purpose, higher frequency sampling is conducted only over short intervals for only a handful of datasets.

VI. Related Work

Our work is informed by previous studies focused on monitoring network paths and characterizing dynamic properties. Several studies have utilized LGNs to address the limited coverage of measurement platforms. [23] utilized LGN traceroute measurements to reveal peering relationships within Internet Exchange Points (IXPs). Similarly, [24] leveraged LGNs to augment the visibility of multilateral peering. In a follow-up work, [9] developed Periscope, a platform automating LGN measurements by aggregating LGNs from various providers in a programming interface. The SCN has also received attention in recent years. A position paper by [25] emphasized the criticality of the SCN, providing examples of SCN failures. [26] investigated the impact of a recent cable deployment in the South Atlantic on African countries. In [4], it is revealed that popular websites include objects only reachable via SCN infrastructure. More recently, [27] studied the structure of *long-haul links* of the Internet, finding that network-layer intercontinental links often terminate at far inland locations, thousands of kilometers away from submarine landing points. In another effort to map SCN infrastructure, [28] proposes Nautilus, which uses topological information collected from probes in the proximity of landing points.

The notion of operational constancy was first introduced in [6], which utilizes a purpose-built measurement infrastructure to measure packet loss, latency, and throughput. Our analysis is based in part on these methods. Many platforms such as [29] and [30] are used to constantly probe the Internet to assess performance characteristics and network outages. There is a large body of work on the characteristics of network path properties starting with [31]. Other studies developed techniques to improve measurement accuracy (*e.g.,* [18]). Finally, measurement-based analysis of network events have been

reported in several studies, including [32], [33]. Recently, [14] tracks user-affecting Internet outages using aggregated Google queries collected via Google trends.

VII. Summary and Conclusion

In this paper, we report the results of our study of the dynamic properties of SCN paths. We developed an active probe-based monitoring system that utilizes 90 LGNs distributed in different networks around the world to send traceroute probes across 134 SCN paths during a period of 18 months. We analyze characteristics of packet loss, latency and path stability, focusing specifically on the notion of *operational constancy* for these metrics. Our main finding is that these characteristics vary widely from path to path. To organize our findings and to provide further detail, we conduct a cluster analysis and find that a relatively small set of behaviorally defined clusters capture the wide range of performance characteristics observed across the SCN paths. We also find that several of the clusters include paths that exhibit unsteady behaviors, high latency variability, and high packet loss rates, which would all adversely affect application performance. Our findings have implications for applications that traverse the SCN, for operational practices, and for the design of future SCN monitoring systems. In future work, we plan to undertake further expansion of our monitoring system and to enhance its capabilities for identifying events that have an impact on operational constancy.

Acknowledgments

This work is supported by the National Science Foundation grants CNS-2107392, CNS-1703592, CNS-2039146, and CNS-2106517. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] TeleGeography, "Submarine cable map," https://www.submarinecablemap.com/.

[2] C. Wall, "Invisible and vital: Undersea cables and transatlantic security," Center for Strategic and International Studies, Tech. Rep., June 2021.

[3] M. Hirosawa and R. Yasoshima, "G-7 to support deep-sea cable network for emerging nations," Nikkey Asia, April 2023, https://s.nikkei.com/43rrQJY.

[4] S. Liu, Z. S. Bischof, I. Madan, P. K. Chan, and F. E. Bustamante, "Out of sight, not out of mind - a user-view on the criticality of the submarine cable network conference," in *Proc. of IMC*, 2020.

[5] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, "Bbr: Congestion-based congestion control: Measuring bottleneck bandwidth and round-trip propagation time," *Queue*, vol. 14, no. 5, pp. 20–53, 2016.

[6] Y. Zhang and N. Duffield, "On the constancy of internet path properties," in *Proc. ACM IMW*, 2001.

[7] Twelve99, "Twelve99 looking glass," https://lg.twelve99.net/.

[8] Infrapedia, "The infrastructure map."

[9] V. Giotsas, A. Dhamdhere, and K. C. Claffy, "Periscope: Unifying looking glass querying," in *Proc. of PAM*, T. Karagiannis and X. Dimitropoulos, Eds., 2016.

[10] L. Quan, J. Heidemann, and Y. Pradkin, "Trinocular: understanding internet reliability through adaptive probing," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, ser. SIGCOMM '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 255–266. [Online]. Available: https://doi.org/10.1145/2486001.2486017

[11] H. Guo and J. Heidemann, "Detecting icmp rate limiting in the internet," in *Passive and Active Measurement: 19th International Conference, PAM 2018, Berlin, Germany, March 26–27, 2018, Proceedings 19*. Springer, 2018, pp. 3–17.

[12] R. Ravaioli, G. Urvoy-Keller, and C. Barakat, "Characterizing icmp rate limitation on routers," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 6043–6049.

[13] L. Davisson, J. Jakovleski, N. Ngo, C. Pham, and J. Sommers, "Re-assessing the Constancy of End-to-End Internet Latency," in *In Proceedings of the IFIP Internet Traffic Measurement and Analysis Conference*, 2021.

[14] E. C. Kirci, M. Vahlensieck, and L. Vanbever, ""is my internet down?": Sifting through user-affecting outages with google trends," in *Proceedings of the 22nd ACM Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2022, p. 290–297. [Online]. Available: https://doi.org/10.1145/3517745.3561428

[15] K. Ali, "Sea-me-we-5 fault causing disruptions," https://subtelforum.com/sea-me-we-5-fault-causing-disruptions/, 2022, accessed: May 25, 2023.

[16] P. Lipscombe, "Damage to sea-me-we-5 cable caused internet disruption to asia and africa for several hours," https://www.datacenterdynamics.com/en/news/damage-to-sea-me-we-5-cable-caused-internet-disruption-to-asia-and-africa-for-several-hours/, 2022, accessed: May 25, 2023.

[17] S. Nussey, "Tropical storm meari lashes japan with heavy rain," https://www.reuters.com/business/environment/tropical-storm-meari-lashes-japan-with-heavy-rain-2022-08-13/, 2022, accessed: May 25, 2023.

[18] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Improving accuracy in end-to-end packet loss measurement," in *In Proceedings of ACM SIGCOMM*, 2005.

[19] R. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, p. 267–276, 1953.

[20] M. Syakur, B. Khusnul Khotimah, E. Rohman, and B. Dwi Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," *OP Conference Series: Materials Science and Engineering*, vol. 336, 2018.

[21] V. E. Paxson, "Measurements and analysis of end-to-end internet dynamics," Ph.D. dissertation, USA, 1998, uMI Order No. GAX98-03325.

[22] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.

[23] B. Augustin, B. Krishnamurthy, and W. Willinger, "Ixps: Mapped?" in *Proc. of IMC*, 2009.

[24] V. Giotsas, S. Zhou, M. Luckie, and k. claffy, "Inferring multilateral peering," in *Proc. of CoNEXT*, 2013.

[25] 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, pp. 78–84.

[26] R. Fanou, B. Huffaker, R. Mok, and K. Claffy, "Unintended consequences: Effects of submarine cable deployment on internet routing," in *Proc. of PAM*, 2020.

[27] E. Carisimo, C. Wang, M. Weaver, F. E. Bustamante, and P. Barford, "A hop away from everywhere: A view of the intercontinental long-haul infrastructure," in *Proc. of ACM SIGMETRICS*, 2024.

[28] A. Ramanathan and S. Abdu Jyothi, "Nautilus: A framework for cross-layer cartography of submarine cables and ip links," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2023.

[29] CAIDA, "Archipelago (ark) measurement infrastructure," https://www.caida.org/projects/ark/.

[30] R. NCC, "What is ripe atlas?" https://atlas.ripe.net/about/.

[31] J.-C. Bolot, "End-to-end packet delay and loss behavior in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 23, no. 4, 1993.

[32] L. Quan, J. Heidemann, and Y. Pradkin, "Understanding Internet Reliability Through Adaptive Probing," in *In Proceedings of ACM SIGCOMM*, 2013.

[33] A. Dainotti, C. Squarcella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescape, "Analysis of country-wide internet outages caused by censorship," *EEE/ACM Transactions on Networking*, vol. 22, no. 6, 2014.