Towards Global Outage Detection for LEO Networks

Manda Tran
University of California, Los Angeles
mandat@ucla.edu

Khiet Huynh
University of California, Los Angeles
khietlhuynh@ucla.edu

Dravya Jain
University of California, Los Angeles
dravyajain@ucla.edu

Dylan Truong
University of California, Los Angeles
dylantruong5@ucla.edu

Sirapop Theeranantachai
University of California, Los Angeles
stheera@ucla.edu

Beichuan Zhang
University of Arizona
bzhang@cs.arizona.edu

Lixia Zhang
University of California, Los Angeles
lixia@cs.ucla.edu

Liz Izhikevich
University of California, Los Angeles
lizhikev@ucla.edu

ABSTRACT

Low Earth Orbit (LEO) satellite networks are increasingly deployed, yet users continue to experience frequent, short-lived outages. We present Roman-HitchHiking, a system for measuring LEO satellite outages globally and in near-real time. Roman-HitchHiking significantly reduces the measurement overhead by leveraging path redundancy to eliminate duplicate probes to shared pre-satellite routers, thereby reducing overall network traffic and increasing coverage. With Roman-HitchHiking, we identify large clusters of simultaneous outages across geographically diverse regions, pointing to potential centralized failures that traditional outage detection systems overlook. Roman-HitchHiking is open-sourced to enable reproducibility and foster further research on LEO outages.

CCS CONCEPTS

• **Networks → Network monitoring**; *Network measurement*; *Network dynamics*.

KEYWORDS

LEO, Satellite, Network, Measurement, Starlink, Outages

ACM Reference Format:

Manda Tran, Khiet Huynh, Dravya Jain, Dylan Truong, Sirapop Theeranantachai, Beichuan Zhang, Lixia Zhang, and Liz Izhikevich. 2025. Towards Global Outage Detection for LEO Networks. In *3rd Workshop on LEO Networking and Communication (LEO-Net '25), September 8–11, 2025, Coimbra, Portugal.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3748749.3749088

1 INTRODUCTION

Despite the growing deployment of Low Earth Orbit (LEO) satellite networks, users continue to experience frequent outages that differ in nature from those observed in terrestrial networks [19]. Disruptions in LEO connectivity stem from satellite mobility, the dish's susceptibility to obstructions, and sporadic geomagnetic storms [11, 22]—all of which can trigger outages at both the individual and regional levels [16, 19]. Understanding whether particular regions and customers are predisposed to outages would allow researchers to determine user quality of experience, identify disparities in service reliability, and prioritize improvements for the most affected populations. However, to date, there has been no systematic global way to study outages in LEO satellite networks because such analysis requires high-resolution data captured in near-real time across a wide geographic scale.

Conventional outage detection systems designed for terrestrial networks are inadequate for analyzing outages affecting LEO satellite networks [12, 23]. These systems focus on identifying large-scale, long-duration outages, whereas LEO satellite disruptions often occur at a much finer timescale—on the order of seconds. While Izhikevich et al. [13] introduced HitchHiking, a methodology for measuring global Starlink customer latency, we demonstrate this approach does not scale to the demands of collecting global, simultaneous outage data. Rather, in practice, naively running HitchHiking leads to massive packet loss.

In this work, we introduce a methodology to measure LEO satellite outages at a global scale: Roman-HitchHiking. Roman-HitchHiking solves HitchHiking's packet loss problem by capitalizing on a key insight: widespread path redundancy in the network allows us to reduce the number of paths we need to probe. Roman-HitchHiking is inspired by the idiom "all roads lead to Rome": at any given moment, if a destination just before the satellite hop is reachable (or unreachable) via one network path, it is highly likely to be reachable (or unreachable) via many others.

Roman-HitchHiking reduces measurement-induced packet loss from 73.4%—as seen in prior methods—to under 0.01%, enabling over four orders of magnitude greater coverage for simultaneously measuring customer outages. We apply Roman-HitchHiking to collect measurements from Starlink [26], the largest LEO satellite provider, over a three day period. Our analysis reveals large clusters of simultaneous outages account for most disruptions, with Australia consistently among the most impacted areas. We open-source Roman-HitchHiking at https://github.com/UCLA-SCaN/roman_hitchhiking and hope it serves as a foundation for future research on LEO network outages.

LEO-Net '25, September 8–11, 2025, Coimbra, Portugal

ACM ISBN 979-8-4007-2090-1/25/09
https://doi.org/10.1145/3748749.3749088

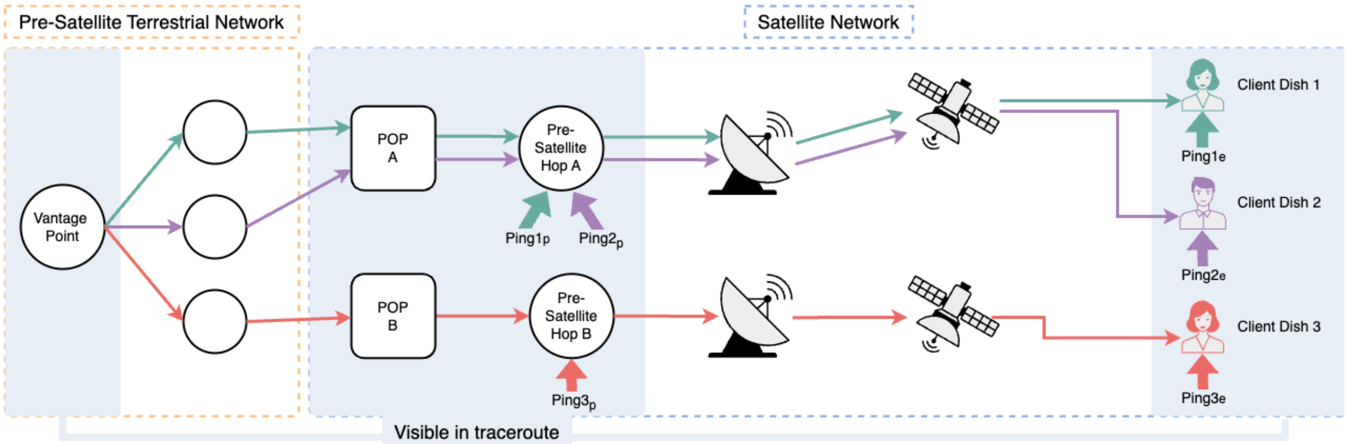


Figure 1: Identifying the Satellite Link from a Simplified Traceroute Using HitchHiking— HitchHiking first identifies the satellite link by tracerouting from a Vantage Point to Client Dish 1, 2, and 3. The traceroute reveals the Point of Presence (POP) and pre-satellite hop router. An outage occurring over the satellite link can be identified by simultaneously pinging the pre-satellite hop and the client dish. An outage has occurred if the pre-satellite hop responds to the probe but the client dish does not respond. Pings are denoted as $PingN_{\{p,e\}}$, where N is the Client Dish number, **p** refers to a **p**re-satellite hop ping, and **e** refers to a customer **e**ndpoint ping. For example, $Ping3_p$ is the probe sent to the *p*re-satellite hop along Client Dish *3*'s path.

2 BACKGROUND

To enable global, scalable outage detection in LEO satellite networks, a low-barrier, widely deployable technique is essential. We begin by reviewing an existing methodology, HitchHiking, and exploring how it might be naively adapted for outage detection. In Section 3, we experimentally demonstrate the scalability limitations of the HitchHiking approach, which motivates the development of Roman-HitchHiking.

Defining Outages. Starlink's Service Level Agreement defines an "outage" as "…a period where the Starlink is unable to send/receive pings to/from servers at a Starlink Point of Presence" [27]. In this paper, we employ a narrower definition and focus on outages that occur within the satellite link path rather than any outage that occurs within the Starlink network path. In particular, we define an *outage* as a period of time in which the Starlink client is unresponsive to the Starlink pre-satellite hop router.

HitchHiking Overview. LEO HitchHiking, introduced by Izhikevich et al. [13], infers network characteristics by probing publicly accessible satellite-routed devices. As illustrated in Figure 1, HitchHiking first identifies the pre-satellite hop using ICMP Paris traceroute. It maps the path from the last visible pre-satellite router to the exposed Starlink client IP (e.g., Client Dish 1, 2, 3). A TTL-limited ping is then used to isolate latency at specific hops: one ping targets the pre-satellite hop, and another the client dish. The difference in latency estimates the satellite link delay. We adopt this technique to identify pre-satellite hops and isolate the satellite segment in our methodology.

HitchHiking Outages. HitchHiking can be naively applied to identify outages. For instance, in Figure 1, Client Dish 1 experiences an outage if the probe $Ping1_p$ is responsive but $Ping1_e$ is not responsive. Note that since we are only conducting one-way measurements from the "outside-in," an unresponsive client can indicate that they are unable to send *or* receive pings from our server, but we cannot distinguish between the two. Even if the endpoint receives the ping, the absence of a response is still an outage by our definition. While filtering policies could in theory explain missing responses, we observe that clients respond to probes both before and after the outage presented in Section 4. This transient behavior makes persistent filtering rules an unlikely explanation.

3 SIMULTANEOUS LATENCY MEASUREMENTS

To understand outages in LEO satellite networks at a global scale, we need a measurement system capable of probing customer services simultaneously—or as close to the same point in time as possible—to capture a consistent view of network conditions. In this section, we show that naively adopting prior work presents scalability challenges. Thus, we propose Roman-HitchHiking and evaluate it against the naive approach.

3.1 Naive HitchHiking

To measure global outages, we naively modify the HitchHiking methodology to run simultaneous TTL-limited pings for all exposed service routers and their corresponding terrestrial router. For example, in Figure 1, to measure a network with three publicly exposed IPs at a single point in time, we would simultaneously send 6 TTL-limited pings (illustrated as *Ping*-labeled arrows). To measure outages across all customers who expose services over a 60-second window at a one second interval, a naïve implementation of HitchHiking would simultaneously send a probe to each customer service and their corresponding pre-satellite hop once per second for 60 seconds. We conduct naive measurements from Stanford University on May 20, 2025. Figure 2 shows that running naive HitchHiking leads to over 50% packet loss at the pre-satellite hop,

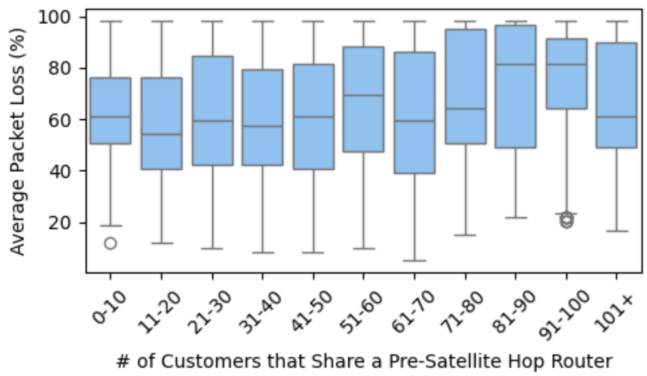


Figure 2: Substantial Packet Loss Due to HitchHiking Overload of Pre-Satellite Hop—The Naive HitchHiking implementation results over 50% of packet loss to each pre-satellite hop router over a 60 second measurement period, probing once per second.

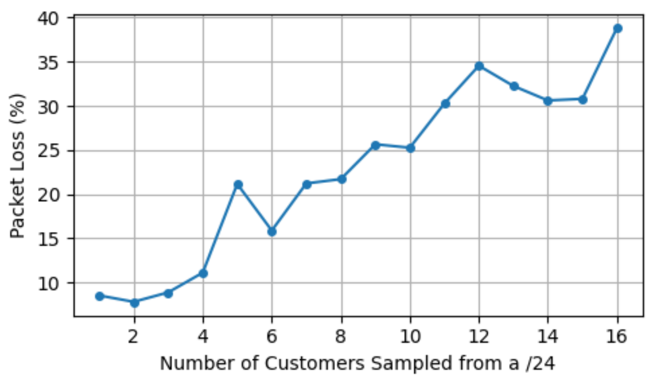


Figure 3: Packet Loss Increases as Sample Size Increases in Naive HitchHiking—Packet loss substantially increases after subsampling more than 4 customers per /24 subnet.

regardless of the number of endpoints probed. Packet loss is not due to customer packet drop. Rather, packet loss is likely due to routers rate-limiting ICMP responses to prevent denial-of-service attacks. This rate-limiting hampers our ability to probe multiple endpoints simultaneously, limiting visibility and reducing the effectiveness of outage detection.

To better understand and mitigate packet loss during measurements, we explore how packet loss varies with the number of endpoints probed. We partition our dataset into /24 subnets and subsample customers within each subnet. For a given sample size n, we probe n endpoints from each subnet. Figure 3 shows that packet loss significantly increases when probing more than 4 customers per /24 [1]. Thus, we require a new outage measurement solution that reduces packet loss while maintaining coverage.

[1]In Appendix A.1, we examine the packet loss using an alternative partition method: by shared pre-satellite hop. However we do not observe a difference in packet loss compared to the pre-satellite partition.

3.2 Simultaneous Measurements Using Roman-HitchHiking

Roman-HitchHiking capitalizes on a key insight: widespread path redundancy in the network allows us to reduce the number of paths we need to probe. Many exposed customer services converge on the same pre-satellite hop router—in the worst case, up to 288 customers share a router in our experiment from Section 3.1.

To lower the overall probing load and mitigate packet loss at scale, Roman-HitchHiking sends TTL pings to every exposed endpoint, while issuing only a single TTL ping to each unique pre-satellite hop router. For example, in Figure 1, to collect one measurement of satellite link latency, Roman-HitchHiking sends TTL pings to the endpoints, Client Dish 1-3 ($Ping1_e$, $Ping2_e$, $Ping2_e$), and sends *one* TTL ping to Pre-Satellite Hop A (either $Ping1_p$ *or* $Ping2_p$) and one TTL ping to Pre-Satellite Hop B ($Ping3_p$).

To improve coverage and reduce filtering—often triggered at the source IP level [29]—Roman-HitchHiking distributes probes across multiple source IP addresses. As we later show, this design significantly improves response rates and measurement reliability.

3.3 Evaluation

We evaluate Roman-HitchHiking across three dimensions to demonstrate its effectiveness and scalability in detecting outages across LEO satellite networks: (1) How well does Roman-HitchHiking reduce measurement-induced packet loss at large scales? (2) How does increasing the number of source IPs improve measurement coverage? (3) Does optimizing for scale (e.g., eliminating redundant probes, varying vantage points) preserve outage detection accuracy? Our results show that Roman-HitchHiking reduces packet loss and thereby increases coverage—by over four orders of magnitude compared to naive methods—while preserving measurement accuracy.

To perform our data collection we use Censys [5, 8] to find exposed Starlink services. We filter exposed services for customer endpoints and exclude endpoints using performance enhancing proxies (PEP) [13] [2]. We use Scamper [4, 18] to run Paris traceroutes to determine the pre-satellite hop.

Scalability and Packet Loss. We begin by measuring how well Roman-HitchHiking scales to large numbers of customers. We use Naive HitchHiking as a comparison, which suffers from severe packet loss when probing many customers simultaneously. Figure 4 shows that packet loss rates in the naive setup can reach 79.7%. By removing redundant probes to the same pre-satellite hop and distributing probes across multiple source IPs, Roman-HitchHiking reduces packet loss to just 0.01%—a four-order-of-magnitude improvement. This enables significantly higher measurement volume and coverage without sacrificing data quality.

Impact of Source IP Diversity. While eliminating redundant probes reduces loss by a factor of 1.93, Roman-HitchHiking still experiences a packet loss rate of 37.9% when using a single source IP (Figure 4). This loss is largely due to complete packet drops—where

[2]We only include only hosts whose DNS PTR record follow the Starlink customer format `customer.[location].pop.starlinkisp.net` and exclude hosts with Peplink TLS certificates, the most common PEP.

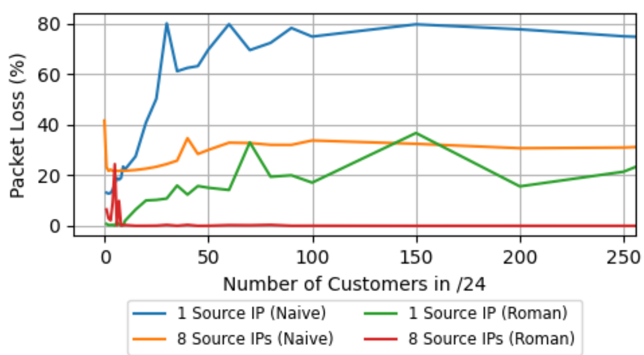


Figure 4: Roman-HitchHiking Experiences Significantly Less Packet Loss Compared to Other Methodologies—Roman-HitchHiking with 8 source IPs consistently report close to 0% packet loss. In the worst case, Naive HitchHiking with 1 source IP experiences up to 79.7% packet loss when sampling 150 customers in a /24.

neither the pre-satellite hop nor the endpoint responds (see Figures 3 and 4). As detailed in Appendix A.3, such losses suggest filtering behavior at the source IP level.

Motivated by prior findings on scanning-induced filtering [29], we hypothesize that using multiple source IPs can mitigate this issue. To test this, we compare performance using a single source IP versus all 8 available IPs. The results validate our hypothesis: packet loss drops dramatically to just 0.01% with full source IP diversity.

Measurement Validity. We evaluate whether removing redundant pre-satellite probes and using multiple source IPs impacts outage detection accuracy. We run three concurrent configurations: (1) *Naive*: Naive HitchHiking, probing 4 customer endpoints per /24 (maximum scale before packet loss spikes) and (2) *Roman*: Roman-HitchHiking, probing 4 endpoints per /24 with deduplication. (3) *Roman Large*: Roman-HitchHiking, probing 8 endpoints per /24 with deduplication. In each configuration, we probe each endpoint once every second for 5 minutes.

In our evaluation (and Section 4 analysis) we focus on *sustained* outages over 5 seconds and over 15 seconds long. We choose durations over 5 seconds because it is the maximum outage length Starlink's mobile app categorizes (0.1s+, 2s+, and 5s+). We also evaluate outages over 15 seconds because prior work has shown that Starlink has a 15 second reconfiguration period [20]. If an outage persists for more than 15 seconds, it suggests that even satellite handovers failed to resolve the issue. An outage is considered detected by all configurations if any temporal overlap exists between corresponding outage periods. Outages that occur exclusively in one method and last longer than the evaluated outage length (5s+ or 15s+) are treated as method-specific detections. For all measurements, we exclude endpoints with packet loss rates exceeding two standard deviations from the mean.

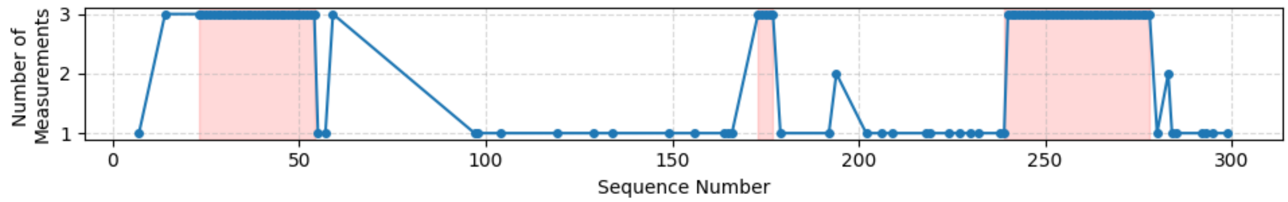
We first perform a case study to cross-validate measurement consistency across the configurations targeting the same customer. Figure 5 shows all configurations detect the same sustained outages, anecdotally confirming that Roman-HitchHiking produces consistent results regardless of measurement strategy.

Beyond the single case study, Roman-HitchHiking scales without sacrificing fidelity: among the 1,092 overlapping customers between experimental configurations, the Jaccard Similarity between the *Naive* and *Roman Large* configurations is 0.85 for 5-second outages and 0.88 for 15-second outages, indicating high agreement. We also compare outage detection across two vantage points by running *Roman* and *Roman Large* configurations simultaneously from two different source IPs. With 1,265 overlapping customers, we again observe a Jaccard Similarity of 0.88 for both 5 and 15-second outages, confirming that vantage point diversity does not compromise consistency.

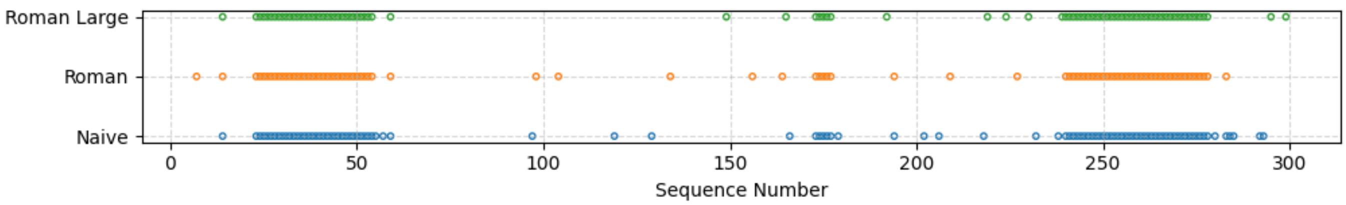
We further verify that reducing redundant scanning does not introduce false positives or negatives, by verifying that within a given second, a pre-satellite hop is reachable regardless of the terrestrial path taken. For example, suppose Client A and B share the same Pre-Satellite Hop A but traverse different terrestrial paths 1 and 2, respectively (Figure 1). If Pre-Satellite Hop A is reachable, probes that traverse through either terrestrial path will report reachability. If Pre-Satellite Hop A is unreachable, both paths will report unreachable. We analyze the concurrently sent pre-satellite probes from the *Naive* configuration and do not find a single instance in which the pre-satellite hop was responsive to one terrestrial path and unresponsive to different terrestrial path.

Obtaining ground truth for Starlink outages is challenging due to the scarcity of public measurement data, limited access to community-operated vantage points, and the absence of our own hardware deployment. We explored several options to validate our measurements. First, we examined RIPE Atlas [1], but found that its measurement granularity was too coarse to capture the transient, sub-minute outages that Roman-HitchHiking targets. Second, we considered using LEOScope [17] nodes; however, at the time of our measurements, only three nodes were active, none of which exposed public IPs, preventing us from validating "inside-out" outages from an "outside-in" perspective. Existing outage detection platforms [6, 7, 12] only report sustained, macroscopic outages and do not reflect the micro-outages we observe. As a result, direct validation remains an open challenge. In the meantime, we increase confidence in our results by showing that different sources observe the same outage events, suggesting that these are not artifacts of local misconfiguration or filtering. In future work, we plan to deploy our own Starlink terminal to enable ground-truth comparisons as well as more rigorously validate observed outages from geographically distributed sources.

Ethics. We follow the best practices described by Izhikevich et al. [13] and Durumeric et al. [9]. Our scanner's IPs re-direct to a page that informs end-users about the usage for research and how to opt out. We do not receive any requests to opt out. It is common practice to use multiple vantage points to increase coverage in measurement studies [2, 29]. Additionally, ICMP echo requests are small, stateless packets widely used for diagnostic purposes, and are considered low-impact in the context of network measurements. Additionally, we consciously limit the number of probes we send by leveraging existing scans from Censys [8] to find publicly exposed satellite



(a) All Measurement Configurations Detect the Same Outages for 129.222.5.64—The y-axis indicates the number of measurement configurations that identify an outage at a particular second. The highlighted ranges are the sequence numbers in which *Roman Large* (Roman-HitchHiking sampling 8 customers per /24) identifies a >5s outage. The outages it observes is consistent with all other methodologies.



(b) Outages at each Sequence Number Across Measurement Types for 129.222.5.64—Each circle indicates an outage instance and identifies at which point in time each measurement method detects and outage. There are three distinct intervals in which all three methods observe outages over 5 seconds: [23,54], [173, 177], [240-274].

Figure 5: Roman-HitchHiking Accurately Determines Sustained Outages (>5s) —All three measurement configurations as described in Section 3.3 (*Naive* with $n = 4$, *Roman* with $n = 4$, and *Roman Large* with $n = 8$) are run simultaneously to the same exposed endpoint. Each configuration independently detects the same three distinct outages, each lasting over 5 seconds and occurring within overlapping temporal ranges.

services. Critically, Roman-HitchHiking methodology reduces the number of packets sent to pre-satellite hops therefore leaving a smaller footprint than prior methodologies.

4 OUTAGES

In this section, we analyze global outage measurements and show that large clusters of simultaneous outages account for most disruptions. These events span geographically diverse regions, with Australia consistently emerging as one of the most impacted areas.

Dataset. Using the same setup described in Section 3.3, we collect data using Roman-HitchHiking with 8 source IPs for a 5 minute sample, probing at a one-second interval on three different days (May 27-29, 2025). We geolocate the customer endpoints using Starlink's published IP Geolocation feed [25]. We only include endpoints in our data that have successful paris-traceroutes and filter out endpoints that have packet loss over 2 standard deviations of the mean. Across the three datasets, this leaves us with 14,436 to 15,068 exposed (and measurable) endpoints which are distributed across over 98-101 countries (63.64–65.58% of countries with Starlink customers).

Analysis. Our global measurements show that LEO satellite outages are frequent, short-lived, and significant enough to disrupt user experience. Across our datasets, we observe 812 outage events longer than 5 seconds on May 27, 344 on May 28, and 574 on May 29. Most outages last under 60 seconds, but many fall in the 50–75 second range—long enough to affect applications like video calls or gaming. Notably, these outages would be missed by traditional detection systems that probe at multi-minute intervals,

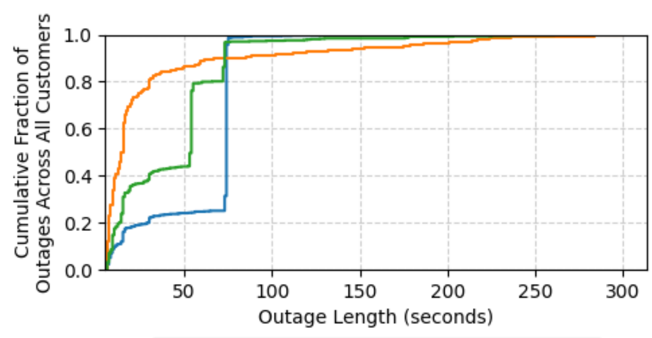


Figure 6: Outages Lasting Longer than 5 seconds are Typically Less Than a Minute Long —For each day of measurement, the cumulative fraction of outage is computed for outages longer than 5 seconds. On May 27 and May 29, there are large increases around 55 seconds and 75 seconds, respectively. These increases are attributed to country-specific outage events which we describe in Section 4.

Large-scale clustered outage events occur on a daily basis. On May 27, 597 outages occurred simultaneously, each lasting between 70 and 75 seconds—accounting for 73.5% of all outages that day. On May 29, we observe two similar clusters of outages lasting 50–55 seconds and 70–75 seconds, respectively, with both groups overlapping in time. These events suggest centralized failures, likely at the satellite link level, that affect many users at once.

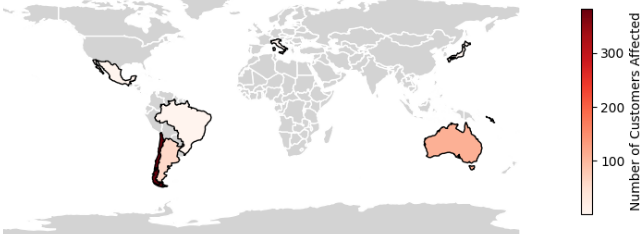


Figure 7: Customers in Chile and Argentina are Largely Affected by a 70-75 Second Outage—On May 27th, several regions including Argentina, Australia, Chile, Brazil, Italy, Mexico, and Japan, simultaneously experienced an outage lasting over 70 seconds.

Figure 8: Customers in Mexico and Australia are Largely Affected by a 50-75 Second Outage—On May 29th, Australia, Mexico, United States, France, Guatemala, and El Salvador, simultaneously experienced an outage lasting between 50 - 70 seconds.

These clustered outages affect users in widely separated geographic regions, potentially suggesting shared satellite infrastructure as the source. For example, the May 27 event impacted 64% of affected users in Chile, 10% in Argentina, and 18% in Australia. On May 29, the 50–55 second outages affected 38% of users in Mexico, while 34% of the 70–75 second outages occurred in Virginia, United States. This global spread makes it unlikely that local ground conditions are the cause.

Australia experiences more frequent and longer outages than any other region in our dataset. In every measurement sample, Australia ranks among the most affected countries, both in number and duration of outages. Other impacted regions include Chile, Argentina, and parts of Oceania such as the Solomon Islands and Tonga, pointing to possible coverage gaps or handover inefficiencies in the southern hemisphere. Country-level statistical analysis confirms that outage rates vary significantly by region. Using a z-test with a 95% confidence level, we find that Chile, Argentina, and Australia had significantly higher outage rates on May 27; Italy and Peru on May 28; and Australia and Mexico on May 29. These findings indicate that outage risk is not evenly distributed across the network.

In summary, LEO outages are shaped by large, simultaneous events and persistent regional disparities. Our findings show the need for high-resolution outage detection to accurately assess user experience and guide the development of more reliable satellite systems.

5 RELATED WORK

While prior work has laid the groundwork for network measurement and macroscopic outage detection, none address the need for a scalable, fine-grained system capable of capturing second-scale LEO satellite outages across a global footprint: the gap our methodology aims to fill.

LEO satellite measurement systems have primarily relied on community-driven platforms (e.g., RIPE Atlas [1]), research testbeds and simulations (e.g., LEOScope [17], Hypatia [24], and StarryNet [15]), and industry tools (e.g., Ookla SpeedTest [21], M-Lab [10]). While valuable for characterizing performance metrics like latency and throughput, these systems were not built to detect outages and are limited by sparse vantage points and coarse temporal resolution.

In contrast, outage detection systems developed by both academia and industry have focused on long-duration disruptions. For example, IODA [12] uses BGP updates, darknet traffic, and active probing to detect macroscopic Internet outages, probing /24 blocks every 10 minutes. Similarly, Hubble [14] targets persistent Internet reachability issues lasting over 15 minutes. USC's ANT Lab [28] has built large-scale infrastructure (e.g., Trinocular [23]) to monitor IPv4 space through extensive active probing at 11-minute granularity.

Industry platforms like Cloudflare Radar [7] and Cisco ThousandEyes [6] also monitor broad outages, but with limited temporal granularity (typically 10–15 minutes) and often proprietary data. Downdetector [3] aggregates user-reported issues but provides only coarse, daily time-series data, limiting its utility for fine-grained analysis. Moreover, few of these tools support public access or offer visibility into customer-level outages.

6 CONCLUSION

We present Roman-HitchHiking, a scalable methodology for measuring customer-level outages in LEO satellite networks at a global scale and in near-real time. By leveraging path redundancy, Roman-HitchHiking reduces packet loss by four orders of magnitude compared to prior approaches. We analyze Starlink customer outage patterns over three days and find large clusters of simultaneous outages across geographically diverse regions. We open-source Roman-HitchHiking to support the future study of global LEO outages.

REFERENCES

[1] RIPE Atlas. [n. d.]. RIPE Atlas. https://atlas.ripe.net/. Accessed: 2025-04-15.

[2] Guillermo Baltra, Tarang Saluja, Yuri Pradkin, and John Heidemann. 2024. Measuring Partial Reachability in the Public Internet. *arXiv preprint arXiv:2412.09711* (2024).

[3] Downdetector by Ookla. [n. d.]. Starlink ooutage or service down?). https://downdetector.com/status/starlink/map/ Accessed: 2025-04-16.

[4] CAIDA. [n. d.]. Scamper. https://www.caida.org/catalog/software/scamper/. Accessed: 2025-04-16.

[5] Censys. [n. d.]. Research Access to Censys Data. https://docs.censys.com/docs/research-access-to-censys-data#censys-datasets. Accessed: 2025-04-16.

[6] Cisco. 2025. ThousandEyes. https://www.thousandeyes.com/. Accessed: 2025-05-13.

[7] Cloudflare. [n. d.]. Outage Center | Cloudflare Radar). https://radar.cloudflare.com/outage-center Accessed: 2025-04-16.

[8] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *22nd ACM Conference on Computer and Communications Security.*

[9] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. {ZMap}: Fast internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*. 605–620.

[10] Phillipa Gill, Christophe Diot, Lai Yi Ohlsen, Matt Mathis, and Stephen Soltesz. 2022. M-Lab: User Initiated Internet Data for the Research Community. *SIGCOMM Computer Communication Review* 52, 1 (2022), 34–37. https://doi.org/10.1145/3523230.3523236 Accessed: 2025-05-03.

[11] Alex DP Hands, Keith A Ryden, Nigel P Meredith, Sarah A Glauert, and Richard B Horne. 2018. Radiation effects on satellites during extreme space weather events. *Space Weather* 16, 9 (2018), 1216–1226.

[12] Georgia Tech Internet Intelligence Lab. [n. d.]. IODA: Internet Outage Detection & Analysis. https://ioda.inetintel.cc.gatech.edu/ Accessed: 2025-04-16.

[13] Liz Izhikevich, Manda Tran, Katherine Izhikevich, Gautam Akiwate, and Zakir Durumeric. 2024. Democratizing leo satellite network measurement. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8, 1 (2024), 1–26.

[14] Ethan Katz-Bassett, Harsha V Madhyastha, John P John, Arvind Krishnamurthy, David Wetherall, and Thomas E Anderson. 2008. Studying Black Holes in the Internet with Hubble.. In *NSDI*, Vol. 8. 247–262.

[15] Zeqi Lai, Hewu Li, Yangtao Deng, Qian Wu, Jun Liu, Yuanjie Li, Jihao Li, Lixin Liu, Weisen Liu, and Jianping Wu. 2023. StarryNet: Empowering Researchers to Evaluate Futuristic Integrated Space and Terrestrial Networks. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. USENIX Association, Boston, MA, 1309–1324. https://www.usenix.org/conference/nsdi23/presentation/lai-zeqi

[16] Dominic Laniewski, Eric Lanter, Simon Beainn, Jan Dunker, Michael Dückers, and Nils Aschenbruck. 2024. Starlink on the Road: A First Look at Mobile Starlink Performance in Central Europe. In *2024 8th Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 1–8.

[17] LEOScope. [n. d.]. LEOScope: A Measurement Testbed for Low-Earth Orbit (LEO) Satellite Networks. https://leoscope.surrey.ac.uk/. Accessed: 2025-04-15.

[18] Matthew Luckie. 2010. Scamper: a scalable and extensible packet prober for active measurement of the internet. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 239–245.

[19] Sami Ma, Yi Ching Chou, Haoyuan Zhao, Long Chen, Xiaoqiang Ma, and Jiangchuan Liu. 2023. Network characteristics of leo satellite constellations: A starlink-based measurement from end users. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 1–10.

[20] Nitinder Mohan, Andrew E. Ferguson, Hendrik Cech, Rohan Bose, Prakita Rayyan Renatin, Mahesh K. Marina, and Jörg Ott. 2024. A Multifaceted Look at Starlink Performance. In *Proceedings of the ACM Web Conference 2024*. Association for Computing Machinery, 2723–2734. Accessed: 2025-05-03.

[21] Ookla. 2023. Ookla's Speedtest Methodology. https://www.ookla.com/resources/guides/speedtest-methodology. Accessed: 2025-05-03.

[22] William E Parker and Richard Linares. 2024. Satellite Drag Analysis During the May 2024 Geomagnetic Storm. *arXiv preprint arXiv:2406.08617* (2024).

[23] Lin Quan, John Heidemann, and Yuri Pradkin. 2013. Trinocular: Understanding internet reliability through adaptive probing. *ACM SIGCOMM Computer Communication Review* 43, 4 (2013), 255–266.

[24] André Baptista Águas Jens Eirik Saethre Simon Kassing, Debopam Bhattacherjee and Ankit Singla. 2020. Exploring the "Internet from space" with Hypatia. In *Proceedings of the ACM Internet Measurement Conference*. Association for Computing Machinery, 214–229. Accessed: 2025-05-03.

[25] Starlink. [n. d.]. geoip.starlinkisp.net. https://geoip.starlinkisp.net/. Accessed: 2025-05-02.

[26] Starlink. [n. d.]. Starlink. https://www.starlink.com/ Accessed: 2025-05-07.

[27] Starlink. [n. d.]. What is the Starlink Service Level Agreement? https://www.starlink.com/support/article/b7142f8e-9762-6fcf-03f5-0d53c2895b3f Accessed: 2025-05-07.

[28] USC. 2025. The ANT Lab: Analysis of Network Traffic. https://ant.isi.edu/about/index.html. Accessed: 2025-05-13.

[29] Gerry Wan, Liz Izhikevich, David Adrian, Katsunari Yoshioka, Ralph Holz, Christian Rossow, and Zakir Durumeric. 2020. On the Origin of Scanning: The Impact of Location on Internet-Wide Scans. In *ACM Internet Measurement Conference*.

A APPENDIX

A.1 Subsampling Partitions

In addition to partitioning customers by /24 subnets, we also investigate if partitioning customers by shared pre-satellite hop impacts packet loss. In both cases, we find that packet loss begins to increase after sampling 4 or 5 (respectively) customers per partition (Figure 9). Additionally we find that packet loss scales with the number of customers sampled rather than the method of partition. Because we do not observe a difference between parition method,

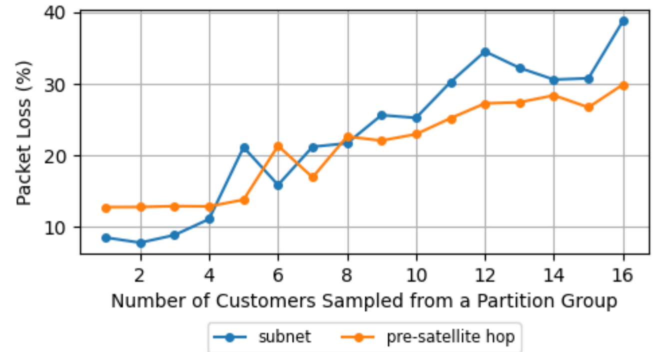


Figure 9: Packet Loss Increases as Sample Size Increases in Naive HitchHiking—Both sample partitions, by /24 subnet and by shared pre-satellite hop, experience increasing packet loss rates as the sample size for each partition grows.

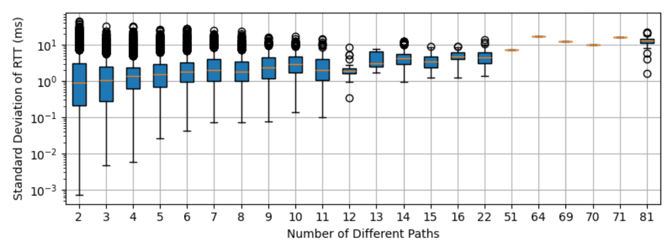
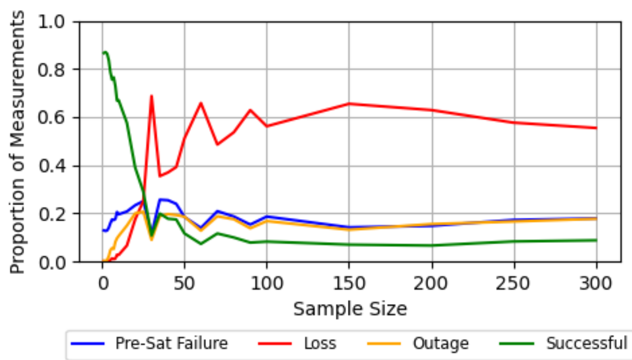


Figure 10: The Standard Deviation of Pre-Satellite Hop RTTs is Comparable Across Different Paths—For the largest sample of Naive HitchHiking with successful measurements (subnet partition with $n = 4$), we find that the RTTs to pre-satellite hops are relatively stable. The majority of the measurements have <1ms standard deviation from the mean RTT.

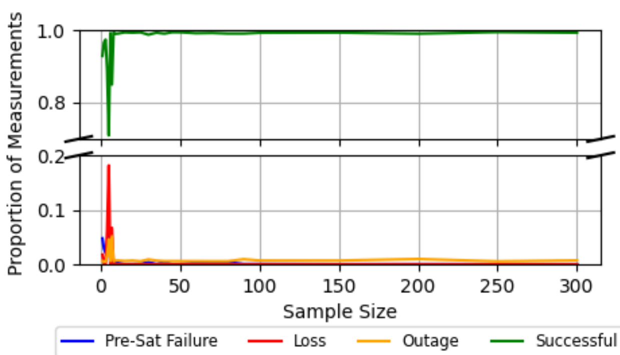
we choose to partition our data by /24 subnet, which is consistent with existing outage detection systems [14, 23].

A.2 Methodology Comparison

Figure 10 shows that the RTT measurements are relatively the same across all paths to the same pre-satellite hop. The RTT variance does increase with the number of paths taken, however, the majority of routes to the pre-satellite hop experience close to 1 ms standard deviation across all paths. These results can be leveraged in future work to understand global latencies in near real-time. Additionally, we find that the decline in valid measurements does not impact the RTT latencies, even with high packet loss, but the entire visibility to an endpoint (e.g., comparing RTTs of same the endpoints in /24 subnet samples $n = 1$ and $n = 14$, the mean difference is only -0.71 ms).



(a) Naive Concurrent with a Single Source IP



(b) Roman-HitchHiking with Multiple Source IPs

Figure 11: Roman-HitchHiking Has a Comparable Distribution of Measurements for All Sample Sizes—Roman-HitchHiking has stable distributions across all data types for all sample sizes. The distribution of data types across all Roman-HitchHiking samples comparable to the distribution of data types in Naive Concurrent samples before they experience high loss rates ($n = 1$ to $n = 4$). Contrastingly, Naive Concurrent measurements begin experiencing high loss rates after sample sizes greater than $n = 4$. The lack of coverage in the data can be attributed to complete *loss* measurements.

A.3 Data Type Distributions over Sampling Sizes

To evaluate how different measurement strategies degrade under scale, we define the following measurement outcomes:

- **Success**: The endpoint and its corresponding pre-satellite hop are both reachable, allowing satellite path reachability to be inferred.
- **Outage**: The pre-satellite hop are reachable, but the endpoint is not. This indicates a potential disruption in the satellite segment (see Section 2).
- **Loss**: Neither the pre-satellite hop nor the endpoint is reachable. This represents a complete failure in the measurement path.
- **Pre-satellite failure**: The endpoint is reachable, but the pre-satellite hop is not. This limits our ability to isolate satellite-specific issues but still indicates partial path availability.

We compare the best performing and worst performing measurement setups from Figure 4. Naive Concurrent HitchHiking experiences a large decline in the fraction of success measurements when increasing the sample size. The large decline is due to the amount of *loss* experienced (we do not receive responses from both the pre-satellite hop and the endpoint). We speculate that this decline is due to routers rate-limiting our source IP (Figure 11a). On the other hand, Roman-HitchHiking with multiple IPs have consistent proportions in data types across all sample sizes. The stable proportions are comparable to the proportions in the Naive Concurrent measurements at small samples ($n <= 4$). We note that the spike in loss at the beginning of Figure 11b is impacted by the small sample size: if a pre-satellite hop shared by several endpoints doesn't respond, the loss rate is disproportionately high.

Received May 2025; revised July 2025; accepted July 2025