universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„Turning Failure into Knowledge: Extending the Observable Internet by Leveraging Delivery Failures“

verfasst von / submitted by

Dipl.-Ing. Florian Holzbauer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktor der technischen Wissenschaften (Dr.techn.)

Wien, 2026 / Vienna, 2026

Studienkennzahl lt. Studienblatt / degree programme code:	UA 786 880
Studienrichtung lt. Studienblatt / degree programme:	Informatik
Betreut von / Supervisor:	Univ.-Prof. DI Dr.techn. Johanna Ullrich Privatdoz. BSc. Univ.-Prof. Dipl.-Ing. Mag. Dr.techn. Edgar Weippl

Acknowledgements

> **Manifesto:** *Yes, We Scan! I dedicate this thesis to the principle that responsible Internet measurements serve the public good. Our scans are conducted for transparency and research, and always with scientific rigor.*

This manifesto is dedicated to the many colleagues, collaborators, and contacts I have made in the Internet measurement community throughout my PhD journey. First and foremost, this journey was made possible through the continuous support of my supervisor, Johanna Ullrich. I first started working with her during my bachelor's studies, where I rewrote ZMap to conduct IPv6 measurements. This collaboration led to conducting IPv6-wide scans during my master's studies and ultimately to pursuing a PhD in the field of Internet measurements. I am deeply grateful for her guidance, trust, and support throughout these years. Since the beginning of our research group, two colleagues have consistently been by my side. Researching, challenging new ideas together, and running measurement infrastructure would not have been the same without you, Gabriel Gegenhuber and Markus Maier. I also thank Edgar Weippl for supporting me on this path and for creating an environment in which this research could thrive, and Barbara Limbeck-Lilienau for sharing her knowledge and her help with administrative tasks.

I gratefully acknowledge netidee for supporting this thesis and my research on Internet measurement through a stipend.

I would like to thank Georgios Smaragdakis and Oliver Hohlfeld for serving on my committee and for reviewing this thesis.

I am especially grateful to my partner, Viola Garstenauer, for her endless enthusiasm in listening to my talks and reading early drafts of my papers. I feel truly fortunate to have you in my life. Finally, I am thankful for the support of my family during times of hardship: my sister Johanna, my father Josef, and my mother Michaela, whose memory remains with us.

List of Publications

This cumulative dissertation is based on three peer-reviewed papers published in conference proceedings of the *USENIX Annual Technical Conference* and the *ACM Internet Measurement Conference*.

[HULF22] Holzbauer, F., Ullrich, J., Lindorfer, M., & Fiebig, T. (2022). *Not that Simple: Email Delivery in the 21st Century.* In 2022 USENIX Annual Technical Conference (USENIX ATC 22) (pp. 295-308).

[HMU24] Holzbauer, F., Maier, M., & Ullrich, J. (2024, November). *Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources.* In Proceedings of the 2024 ACM on Internet Measurement Conference (IMC 2024) (pp. 280-294).

[HSU25] Holzbauer, F., Strobl, S., & Ullrich, J. (2025, October). *Tracking Internet Disruptions in Ukraine: Insights from Three Years of Active Full Block Scans.* In Proceedings of the 2025 ACM on Internet Measurement Conference (IMC 2025) (pp. 474 - 492).

In addition to the results presented within this dissertation, several contributions on related research topics in the field of Internet measurements have been made, published at top-tier security conferences and journals:

[GMH+23] Gegenhuber, G. K., Maier, M., Holzbauer, F., Mayer, W., Merzdovnik, G., Weippl, E., & Ullrich, J. (2023). *An Extended View on Measuring Tor AS-level Adversaries.* Computers & Security, 103302. **(Not part of the thesis)**

[GHF+24] Gegenhuber, G. K., Holzbauer, F., Frenzel, P. É., Weippl, E., & Dabrowski, A. (2024). *Diffie-Hellman Picture Show: Key Exchange Stories from Commercial VoWiFi Deployments.* In 33rd USENIX Security Symposium (USENIX Security 24) (pp. 451-468) . **(Not part of the thesis)**

[GGM+25] Gegenhuber, G. K., Günther, M., Maier, M., Judmayer, A., Holzbauer, F., Frenzel, P., & Ullrich, J. (2025, October). *Careless Whisper: Exploiting Silent Delivery Receipts to Monitor Users on Mobile Instant Messengers.* In Proceedings of the 28th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2025). **(Not part of the thesis)**

Abstract

Delivery failures occur both in the postal system and on the Internet when packets cannot be forwarded to their destination. Active Internet measurements probe Internet-connected systems by sending packets and observing their responses. In practice, not all probes reach their target. When delivery fails, an error message may be returned indicating the cause of the failure, such as a nonexistent address or an unreachable system. Despite their diagnostic value, Internet measurements traditionally focus on successful delivery. As the Internet continues to evolve, including the transition to IPv6 and the deployment of new protocols, new measurement methods are required to observe previously unmeasured parts of the Internet. This thesis shows how delivery failures can be systematically leveraged to expand the observable Internet.

The thesis introduces three measurement approaches, email-security-scans.org, BValue Steps, and CountryMonitor, which use delivery failures to observe infrastructure and security properties that go beyond the capabilities of current state-of-the-art measurement techniques. These approaches extend measurement coverage in three domains: email delivery, IPv6 deployment, and Internet outage detection.

The thesis distinguishes between intentional and unintentional delivery failures. Intentional delivery failures are deliberately triggered by the measurement methodology to elicit protocol-defined responses, whereas unintentional delivery failures arise from real-world conditions such as host unavailability or network outages and are observed through missing responses. Email-security-scans.org and BValue Steps rely on intentional delivery failures, while CountryMonitor is based on unintentional delivery failures.

In the domain of email measurements, delivery failures are used to extend observability beyond inbound email infrastructure. By intentionally triggering validation and rejection mechanisms, email-security-scans.org measures the configuration and adoption of security-relevant standards of outbound email servers at scale, including authentication and validation mechanisms that are otherwise difficult to observe. This enables a sender-side perspective on global email security practices.

For IPv6 measurements, the large address space makes random probing of active hosts infeasible. The thesis shows that delivery failures in the form of ICMPv6 error messages can be used to infer the presence of active networks and network boundaries without requiring successful packet delivery. By analyzing error message types and timing behavior, this approach extends the enumeration of active networks beyond previously observable scopes.

Finally, delivery failures that manifest as missing responses are leveraged to monitor Internet disruptions. CountryMonitor uses loss of responsiveness to detect and track Internet outages over time in Ukraine, providing a longitudinal view of connectivity disruptions affecting numerous smaller Internet operators during an ongoing large-scale geopolitical conflict.

Kurzfassung

Aktive Internetmessungen erfassen Informationen über mit dem Internet verbundene Geräte, indem gezielt Pakete an diese gesendet werden. Dabei erreichen nicht alle Pakete ihr vorgesehenes Ziel und der Sender wird im Idealfall über entsprechende Zustellfehler benachrichtigt. In der Praxis konzentrieren sich viele Internetmessungen jedoch auf erfolgreich zugestellte Pakete, obwohl diese Benachrichtigungen wertvolle Hinweise auf die Ursache von Fehlern liefern, etwa nicht existierende Adressen oder nicht erreichbare Empfänger. Angesichts des fortlaufenden Wachstums des Internets, insbesondere des Übergangs zu IPv6 und der zunehmenden Einführung neuer Protokolle, ist eine kontinuierliche Weiterentwicklung bestehender Messmethoden erforderlich. Diese Dissertation zeigt, wie Zustellfehler gezielt für Internetmessungen genutzt werden können, um deren Reichweite und Aussagekraft zu erweitern.

Die Dissertation stellt drei neuartige Messmethoden vor, email-security-scans.org, BValue Steps und CountryMonitor,die Zustellfehler systematisch auswerten, um etablierte Ansätze in den Bereichen E-Mail-Zustellung, IPv6-Adaption und Erkennung von Internetausfällen zu ergänzen. Der Fokus liegt auf der quantitativen Erfassung der Verbreitung sowie der Sicherheitsaspekte von Internetstandards. Es wird dabei zwischen beabsichtigten und unbeabsichtigten Zustellfehlern unterschieden. Beabsichtigte Fehler werden gezielt durch die Messmethodik erzeugt, etwa durch das Senden von Paketen an nicht existierende Empfänger. Unbeabsichtigte Zustellfehler entstehen hingegen durch Netzwerkausfälle oder Nichterreichbarkeit von Systemen. Während email-security-scans.org und BValue Steps auf beabsichtigten Zustellfehlern basieren, misst CountryMonitor unbeabsichtigte Zustellfehler.

Email-security-scans.org nutzt die gezielte Fehlkonfiguration von Sicherheitsprotokollen und eingeschränkte Erreichbarkeit auf der Empfängerseite, um zu analysieren, ob E-Mails zugestellt werden oder ob die Fehler korrekt erkannt werden und die Zustellung unterbunden wird. Dadurch wird eine senderseitige Analyse globaler E-Mail-Sicherheitspraktiken ermöglicht, die auch Mechanismen sichtbar macht, die mit konventionellen Messansätzen nur eingeschränkt beobachtbar sind. BValue Steps klassifiziert den IPv6-Adressraum für aktive Messungen als aktiv und inaktiv. Dabei wird von einer bekannten IPv6-Adresse die Größe des aktiven Netzwerkes abgeleitet und erkannt, ob inaktive Bereiche in dem Netzwerk vorhanden sind. Die Klassifizierung erfolgt anhand der ICMPv6-Fehlernachrichten, welche absichtlich durch das Senden von Paketen an nicht existierende IPv6-Adressen in dem Netzwerk ausgelöst werden. CountryMonitor schickt in regelmäßigen Intervallen ICMP Echo Request-Pakete an alle IPv4-Adressen des ukrainischen Adressraums. Der Vergleich der Anzahl an unbeabsichtigten Zustellfehlern durch das Ausbleiben von Antworten mit dem Durchschnitt der Vorwoche gibt Rückschlüsse auf Internetausfälle in der Ukraine. Die Methode ermöglicht Einblicke in zahlreiche kleinere Internetanbieter in der Ukraine, die von bisherigen Analysemethoden nicht erfasst wurden.

Contents

1 Introduction

1.1 Motivation

> *Although the Internet is now a planet-wide communication medium, we have remarkably little quantitative understanding of it.* - Mark Crovella, 2006, [CK06]

Internet measurements provide quantitative insight into the state of the global network. However, no single tool can measure the entire Internet, so each measurement technique captures only a specific viewpoint. Since the first interconnection of a few university networks in the United States in 1969, the Internet has grown into one of the most complex systems ever built, extending its physical reach into space. Mark Crovella published one of the early foundational books on Internet measurements in 2006, establishing a basis for a field that has since evolved rapidly [CK06]. Although the quote on the lack of quantitative understanding dates from 2006, it remains valid due to the continued growth and increasing complexity of the Internet. The ongoing deployment of new protocols further requires constant updates to measurement methods.

Measuring the Internet requires actively sending or passively collecting network packets and traffic. Therefore, Internet measurements are commonly divided into active and passive approaches. Active measurements collect information by sending packets to Internet-connected devices and evaluating their responses. In 2013 ZMap was developed as the first high-speed asynchronous scanner that allows sending packets to the entire public IPv4 address space in under 45 minutes [DWH13]. Active measurements are widely used for vulnerability scanning [CER, RB19] and for monitoring protocol deployment at scale. For example, connecting to email servers reveals whether they support encrypted transport during message transmission [DAM+15a]. However, active probes observe only what responds; devices behind firewalls or not reachable from the public Internet remain invisible.

Passive measurements observe existing traffic or control-plane data at privileged vantage points. For example, Domain Name System (DNS) resolvers can reveal historical name-to-IP mappings and resolution patterns, while Border Gateway Protocol (BGP) collectors provide visibility into routing dynamics [RIP, Net25]. Although they provide valuable visibility into network behavior, they miss silent devices that do not actively send traffic and rely on infrastructure that is often inaccessible. This limits reproducibility and restricts who can perform such measurements.

Together, the limitations of both approaches highlight the need to select measurement techniques carefully and to understand what each approach can and cannot

cover. They also motivate the search for additional signals that extend the reach of measurements. Delivery failures often result in error messages that are returned when packets cannot be delivered to its original destination. Examining these failures provides insights even when a target does not respond directly, thereby expanding the observable Internet.

Ubiquitous networking has also introduced new challenges. First, expectations for service availability have increased. Prior work has documented the continuing shift toward cloud hosting and the decline of on-premise infrastructure [FGG+21]. This consolidation places the responsibility for security and uptime on a small number of operators. While protocol designers increasingly incorporate security into new standards, operators sometimes adopt these features slowly, particularly when complexity increases [MYM].

> *"The DNS Camel, or, how many features can we add to this protocol before it breaks."* — Bert Hubert, 2018 [Hub18]

This quotation illustrates a broader concern about the pace of DNS feature development raised at the Internet Engineering Task Force (IETF) meeting 101. Hubert observes that many DNS implementations lack support for the many features standardized in recent years [Hub18]. This highlights a persistent gap between the design of Internet standards and their real-world deployment. Second, users increasingly assume that data is properly secured, whether in transit or at rest. Yet many widely used standards originated decades ago and were later extended with additional mechanisms to meet modern requirements. This underscores the need for measurements that assess whether operators deploy these extensions effectively and in compliance with standards.

The underlying network layer has also changed. The transition from 32-bit IPv4 addresses to 128-bit IPv6 addresses expanded the address space from 4.3 billion to an astronomically larger number (2^{128}). As a consequence, traditional active IPv6 scanning was long considered infeasible. While passive measurements remain unaffected, new research is required to develop practical active measurement techniques for IPv6 networks.

Finally, as society becomes more interconnected, safeguarding the digital environment has become crucial. The ongoing war in Ukraine demonstrates that national defense must also account for disruptions in a country's digital landscape. Measurements can help track and characterize these disruptions. By assigning IP addresses to their physical area of operation, the behavior of IP addresses can be linked to events in the physical world, and vice versa. With a sufficient number of IP-to-location observations over time, changes in the status of Internet-connected devices can provide valuable information about real-world events. Early work demonstrated how Internet measurements could capture the impact of Hurricane Sandy [DWH13], but accuracy and event detection capabilities remain limited. Work on outage detection aggregates IPs to /24 blocks and samples these blocks in higher intervals to increase outage granularity [QHP13]. Further research is needed to improve the precision of such measurements and to identify which device and network properties best reflect external events such as wars or natural disasters.

1.2 Research Goal & Questions

Traditional Internet scanning typically focuses only on successful responses from targets: a probe is sent, and any reply from the destination is collected and analyzed. However, in practice, communication can fail at multiple points, both on the network layer and the application layer. On the network layer, packet delivery might fail on the path to the destination due to missing routes, misconfigurations on routers, or network failures along the path. On the application layer, connections can be dropped because of issues such as failed Transport Layer Security (TLS) negotiation or unsuccessful authentication of the communicating parties. These observations raise an important question: what can be learned from such delivery failures about Internet protocols and reachability? The thesis explores this often-ignored dimension in Internet measurements. The thesis designs and evaluates measurement approaches that treat non-delivery, both of packets and of application-layer messages, as a valuable signal to create new and extend existing measurement methodologies.

To measure Internet reachability, first, a definition of reachability in this context is needed. In graph theory, reachability describes whether a path exists between two nodes; however, the term is less commonly used to describe Internet connectivity. In this thesis, Internet reachability is defined as the ability of host A to reach host B using a specific protocol and configuration. The focus is solely on whether a connection can be established. No assumptions are made about the Quality of Experience (QoE) or performance of the connection. Consequently, reachability is treated as a binary property, represented as either reachable (true) or not reachable (false).

The expected results of the thesis are: (1) new and adapted concepts for Internet measurement design, specifically developed to integrate delivery failures into measurement methodologies; (2) the experimental implementation and execution of these Internet measurement designs in real-world measurement campaigns; (3) the collection of raw data sets resulting from these experimental campaigns; and (4) analysis results obtained by linking the collected measurement data with external data sets.

The thesis distinguishes between intentional and unintentional delivery failures. Figure 1.1 visualizes the types of delivery failures that are leveraged by each of the publications. Intentionally triggered delivery failures are deliberately induced by the measurement methodology to elicit protocol-defined responses. In contrast, unintentional delivery failures arise from real-world conditions, such as host unavailability or network outages, and are observed through the absence of responses. The thesis led to the discovery of new active and passive measurement methods for both of the delivery failure types. Both types of failures provide complementary insights into Internet reachability and protocol behavior. While both capture reachability, intentional delivery failures are primarily used to track protocol adoption, whereas unintentional delivery failures reveal Internet disruptions.

Intentional Delivery Failures. Delivery failures can occur at many layers of the protocol stack. To keep the scope focused, this thesis focuses on the use of delivery

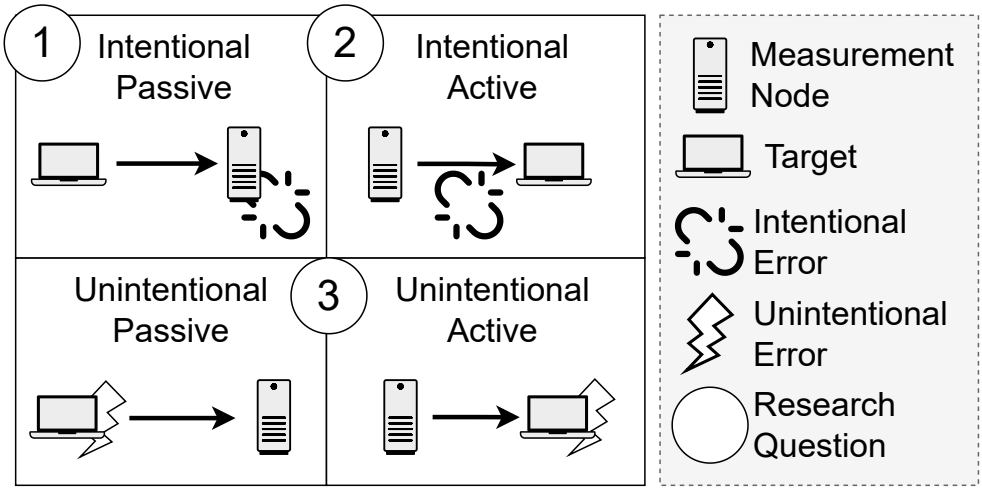


Figure 1.1: Categorization of delivery failures. Intentional and unintentional delivery failures captured through active and passive measurements are examined across the publications included in this thesis. Research questions 1-3 correspond to the ones mentioned in this chapter.

failures for two fundamental protocols in the Internet originally designed by Jonathan P. Postel: the Simple Mail Transfer Protocol (SMTP) first standardized in RFC821 in 1982 [Pos82], which is essential for email delivery, and the Internet Control Message Protocol (ICMP) standardized by RFC792 [Pos81] in 1981 and its successor ICMPv6 first defined by RFC1885 [CD95] in 1995, which is widely used to assess network connectivity between Internet-connected devices.

SMTP is now embedded in a set of delivery-relevant standards that have evolved over time. As the number of Internet-connected devices increased, the requirements for SMTP changed accordingly. Early email servers were commonly deployed without authentication, allowing open relaying. As this behavior was increasingly abused for sending unsolicited emails, additional mechanisms and policies were introduced. Because replacing a well-established protocol such as SMTP is impractical within the Internet ecosystem, these changes were implemented as incremental extensions. This thesis investigates how operators adopt these protocol additions and evaluates if email servers are operated in compliance with current standards. The thesis tackles the following research question.

RQ_1 How can failures in email delivery be leveraged to measure the adoption of email security and transport protocols across email providers? - To answer this question, in chapter 2 we introduce an email delivery testbed to study sender-side capabilities by configuring a set of target addresses with intentional configuration errors and network layer reachability restrictions to verify the connectivity and configuration of outbound email servers around the world. Delivery failures are intentionally introduced on the receiver side to measure the sender's configuration.

ICMP is best known for its use by the ping utility. In addition to echo requests and replies, ICMP defines a set of error messages that allow routers to notify the sender when packet delivery fails along the path to the destination. Such failures can be intentionally triggered, for example, by sending packets to destination addresses that do not exist. This approach is particularly useful in IPv6, where the host address space is often sparsely populated while the size of a single /64 end user allocation encompasses 2^{32} times the total IPv4 address space. These characteristics enable systematic probing to elicit ICMPv6 error messages. This leads to the following research question for active IPv6 measurements:

RQ_2 What do ICMPv6 error messages reveal about their sources? - To answer this research question, in chapter 3 we show that by default router vendors configure the router operating system to return ICMPv6 error messages in case of delivery errors. We perform an active measurement study in the IPv6 Internet, showing that apart from one third of networks that do not return any response, IPv6 networks in the wild also return error messages. We show that for networks with a known IPv6 address, border value steps can be used to infer the prefix size of the active network. Based on the error message types, we classify networks as active (including host addresses, valuable for further measurements), inactive (missing routes, null routes and routing loops; not valuable), or ambiguous. In addition to target networks, we study the rate at which error messages are returned by different router vendors. The routers under test deploy source-based rate limits, which can be depleted without affecting the error messages of other users. We find rate limits of core routers to often be vendor-specific, while edge routers show Linux-kernel defaults, allowing to fingerprint router vendors in the Internet core and kernel versions on the edge.

Unintentional Delivery Failures. In contrast to intentionally triggered failures, the thesis also investigates delivery failures that arise without configuring target systems or exploiting specific protocol properties. Instead, these failures originate from the target itself or from disruptions along the delivery path. In such cases, packet delivery may fail silently, without eliciting any explicit error message. Even when a target address exists and routers maintain valid routes to the destination, delivery can still fail if the target host is inactive, unreachable, or otherwise unable to respond, resulting in the absence of a reply. The thesis studies these unintentional delivery failures by analyzing missing responses to active probes. In particular, it examines how the absence of ICMP Echo Replies can serve as an indicator of large-scale connectivity disruptions and Internet outages.

RQ_3: How can the absence of ICMP Echo Replies be leveraged to detect and comprehensively track Internet outages in a conflict-affected country? - To address this research question, chapter 4 presents a longitudinal full-block measurement of the Ukrainian IPv4 address space. We demonstrate that the absence of responses can be leveraged to identify Internet disruptions at both regional and provider levels. By employing full-block scans rather than sampling-based approaches, outage coverage is extended to the long tail of smaller providers operating within Ukraine.

1.3 Research Methodology

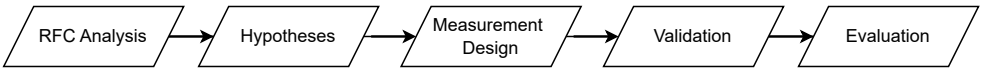


Figure 1.2: Five-step approach to measuring protocol adoption. Hypotheses are derived from RFC documents and validated through quantitative measurement experiments.

The dissertation adopts a structured research methodology to measure protocol adoption. As illustrated in Figure 1.2, the methodology consists of five phases that are repeated for each protocol.

RFC Analysis. Request For Comments (RFC)s, managed by the IETF, describe the protocols used on the Internet and are published as textual documents. The earliest RFCs date back to the beginnings of the Internet and have been used to formally specify Internet protocols ever since [Cro69]. These documents serve as blueprints for implementing software that adheres to the rules defined by the respective standards. Depending on their maturity, RFCs fall into different categories, ranging from Internet-Drafts to Internet Standards. Not every document advances to an Internet Standard, as this progression requires widespread adoption and operational experience. Through the analysis of RFCs, we derive how standards are specified to operate in principle. Modern RFCs typically include a dedicated security considerations section, which evaluates how attackers could interfere with the respective protocols. RFCs can also update existing standards, for example, by introducing additional features such as authentication or encryption to protocols that were originally designed without them [KFR+95]. The analysis focuses on generating testable conditions whether standards and additions are implemented as supposed in real-world implementations or not.

1. Identify relevant RFCs or standards.
2. Extract statements (*MAY*, *SHOULD* and *MUST*) that define expected protocol behavior [Bra97].
3. Translate expected protocol behavior into testable conditions.

Hypotheses. One or more hypotheses are formulated around delivery failures that follow the protocol workflow as specified by the Internet standard.

Measurement Design. This step includes the design of active and passive measurements to collect data on the hypothesis. The proper measurement methodology in terms of setup and scope is evaluated.

Validation. The validation step compares collected measurement data against a ground truth dataset. Different nodes might be online over different scan periods. Active and passive measurements may miss certain data, for example, systems that are not listed in the DNS or IP addresses that are unreachable from outside the network. By comparing our measurement results with ground truth network topologies, we can assess the coverage and completeness of our findings. The completeness is limited to the representativeness of these ground truth networks.

Evaluation. In this step, measurement data is further attributed and statistically evaluated. IP attribution allows assigning IPs to the Autonomous System (AS) that is announcing the respective IP range to the global routing tables, or by geolocation to assign an IP to a physical region.

Several datasets are used to attribute our IP-level data to entities; the most important ones are:

1. **Organizations through AS numbers**. Every IP address on the Internet belongs to an AS, which is a collection of IP address ranges (prefixes) managed by a network operator such as an Internet Service Provider (ISP), cloud provider, or large enterprise. By mapping IPs to their AS numbers, we can identify which organization is responsible for announcing and operating those IP addresses. For example, IPs in AS15895 belong to Kyivstar, the largest telecom operator in Ukraine.

2. **Physical location through IP geolocation** Geolocation databases map IP addresses to estimated physical locations (country, region, city, and estimated latitude/longitude). This provides information about where an IP address is operated, where data is likely stored or where users are accessing services from.

Ethics. This dissertation encompasses active measurements conducted with good Internet citizenship in mind [DWH13]. We limit traffic rates and bandwidth to avoid overloading target networks and distribute probes evenly across targets. All scans are performed from a dedicated vantage point with a reverse DNS pointer to a website explaining the measurements and providing an opt-out mechanism.

The initial phase of learning how to conduct Internet measurements yielded important insights, including the need to carefully select vantage points. Since then, all scans have been operated exclusively from dedicated infrastructure and are instrumented by a central blocklist that prevents packets from being sent to networks that have opted out.

1.4 Research Contributions

The initial phase of this dissertation contributed to the availability of Internet measurement infrastructure in Austria. Together with my supervisor, we established a dedicated entry point for active Internet measurements by deploying a specialized scanning infrastructure in collaboration with Nextlayer. This infrastructure is now publicly available and accessible to other researchers and students via ◎AIM - aim.sba-research.org. The infrastructure has since been used in follow-up work, including a study that improved the understanding of routing loops in the Internet [MU23]. Building on this foundation, the dissertation led to the development of novel measurement approaches across five thematic areas represented at the Internet Measurement Conference 2025 [ACM25]. These topics span a broad range of domains, including Privacy, Cloud Computing, Passive Traffic Analysis, **Mapping Resources and Infrastructure [HSU25]**, AI, Blockchain Security, 5G and Optical Networks, **IPv6 [HMU24]**, **Protocols and Compliance** (**All**), BGP and Routing Security,

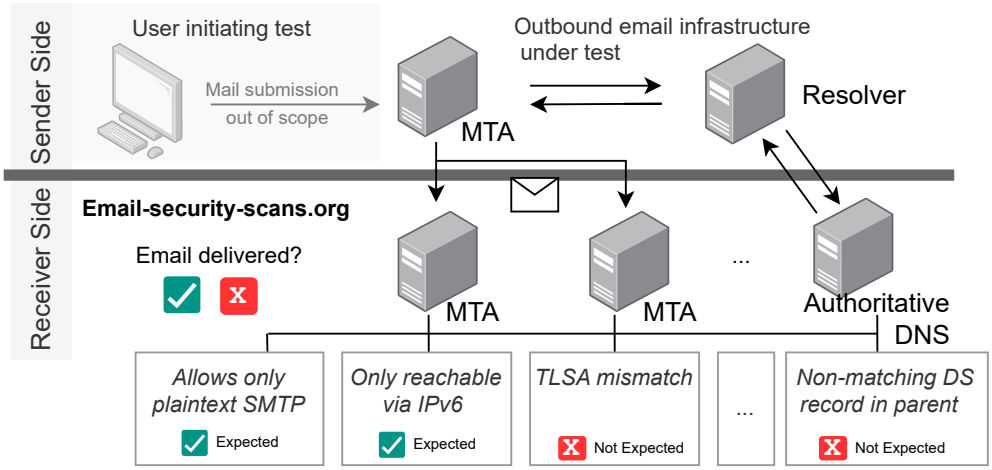


Figure 1.3: Email delivery testbed works by operating receiver-side email servers and authoritative name servers with intentional restrictions and misconfigurations to measure sender-side email delivery capabilities.

Security, Web and Mobile Systems, Routing and Tunneling, Satellite Networks, and **Email and DNS Security [HULF22]**.

Overall, the measurement studies presented in this thesis provide quantitative insights into protocol deployment, operational practices, and security properties of today's Internet.

Email Delivery Testbed

Chapter 2 covers the design and deployment of the publicly accessible email delivery testbed `email-security-scans.org`. The testbed actively engages users by asking them to send emails to a predefined set of measurement addresses. Assuming successful message submission, the focus lies on evaluating sender-side email delivery behavior and the adoption of security standards during mail transfer.

In its initial deployment, the testbed consisted of 11 target email addresses hosted across three authoritative DNS servers and four mail servers. At the time of writing, the platform has been extended to 30 target addresses to support additional measurements, including Mail Transfer Agent-Strict Transport Security (MTA-STS) enforcement, resolver identification, extended TLS cipher suite testing, and delimiter handling.

To address *RQ1*, the measurement targets use delivery failures by intentionally including misconfigurations for which delivery is expected to fail. For example, the address *measurement@v4-mail.dnssec-broken.example.com* is hosted in a subdomain with deliberately broken Domain Name System Security Extensions (DNSSEC), created by publishing a non-matching DS record in the parent zone. A DNSSEC-validating resolver should therefore return a `SERVFAIL`. Based on this, we can use delivery failures in email to verify if the sending email server relies on a DNSSEC-validating resolver or not. The same principle can and is applied to measure other protocols and email security extensions. Apart from this, the original 11 measurement targets allow measuring the IPv6 readiness of email, including DNS resolution, the

use of opportunistic and strict TLS configuration and validation of certificates and the effectiveness of the anti-spam measure greylisting.

Figure 1.3 presents a subset of the measurement addresses and indicates whether successful delivery is expected. Plaintext-only delivery is treated as a special case, while email servers shall fall back on plaintext communication if no other option exists, we find operators that enforce TLS actively avoiding sending emails in plaintext.

The test procedure requires that measurement addresses are explicitly communicated to participating users. The initial study listed target addresses on a website and promoted this study worldwide via addressing several mailing lists and social media channels. In the current version the process has been simplified for users; by entering their email they are sent an instruction email and the test is initiated by reply-all to this email. Successful delivery to at least one address is required to enumerate the remaining targets. If a target address is missing in the *To* header from the received email, it is impossible to distinguish whether the user intentionally removed the address or whether delivery failed.

The email delivery testbed enables controlled measurements of sender-side email configurations by introducing intentional delivery failures. Evaluating delivery to 11 target addresses, the following key findings emerge:

- **Email lags behind DNS in IPv6 adoption:** While 65 % of email providers can resolve hosts over IPv6, only 44 % are able to deliver emails via IPv6.
- **Limited adoption of security features:** Only 57.35 % of providers rely on DNSSEC-validating resolvers and encryption remains largely opportunistic. Enforcing TLS in 2020/21 lead to approximately 10% of regular providers being unable to deliver emails.
- **Large providers prioritize reachability:** While more than half of regular providers rely on DNSSEC-validating resolvers, only four out of the 13 tested large providers did so. Notably, the most prominent providers, Google and Microsoft, switched to DNSSEC-validating resolvers after the initial study in 2020/21.
- **Low validation rates for strict TLS through DNS-based Authentication of Named Entities (DANE):** Despite enabling non-opportunistic encryption for email delivery, DANE sees limited enforcement. When the Transport Layer Security Authentication (TLSA) record did not match the presented certificate, the mismatch was only validated by about one fifth of providers, 21.55 % of regular (TLS-supporting) providers, and 23.1 % of large providers.
- **Spammers favor the path of least resistance:** No spam emails were observed over our IPv6-only email servers. Enforcing TLS reduced spam volume to approximately one third, whereas the traditional anti-spam mechanism of greylisting reduced spam to 63.1 % of the original volume.

Longitudinal Measurement Platform Operating the email-security-scans.org measurement infrastructure beyond the original study period enables additional insights into longitudinal trends in email security adoption. Therefore, we analyze data collected by our platform between 2023 and 2025. For each request domain

and year, we consider the first and last conducted test in order to capture both improvements and adoption dynamics over time. This results in a total of 3 610 first-last test pairs, of which 2 781 domains conducted a single test per year and 829 conducted multiple tests per year. For domains with multiple tests, we observe an improvement in the last test compared to the first test in 81% of cases on average.

Analyzing which protocols were most frequently improved, we find DKIM, DANE, IPv6 support (in 2023 and 2024), SPF validity, DMARC reporting, and MTA-STS (in 2025) among the top five across years. DKIM ranks highest in both 2023 and 2024 and peaks in 2024. This coincides with Gmail sender guidelines that, since February 2024, require sending MTAs to sign outgoing email using DomainKeys Identified Mail (DKIM) [Goo26]. The same guidelines mandate the use of TLS for email transport to Google since December 2023. In comparison to our initial study, in 2023 only 2.9% of tests fail to encrypt email in transit, decreasing further to 0.5% in 2025.

The impact of selecting either the first or the last test of a request domain on measured adoption rates ranges from 0.08 percentage points for IPv4 delivery in 2023 to 3.33 percentage points for DKIM support in 2024. In the following, we therefore report adoption rates based on the last test for the 23% of users who conducted multiple tests per year. Using this approach, we identify the following positive trends in the adoption of email security standards:

1. **First restrictions on cleartext transmission.** In 2023, 3% of providers already refrain from transmitting email in plaintext. This share increases to 9.5% in 2025, indicating that nearly one tenth of providers no longer fall back to cleartext delivery.

2. **Increase in IPv6 adoption.** The share of IPv6-capable mail servers, previously measured at 44% in 2020/21, rises to nearly 60% in 2025.

3. **Growing adoption of strict TLS mechanisms.** DANE support increases to 38.43%. In addition, we extend our testbed to measure MTA-STS deployment, which grows from 13.9% in 2023 to 24.4% in 2025.

4. **Widespread adoption of DKIM and DMARC.** Sender guidelines have significantly accelerated the deployment of email authentication standards. DMARC reaches 90% adoption in 2025, while DKIM reaches 74.4%, representing the largest increase since 2023 with a gain of 16.5 percentage points.

At the same time, we observe unintended side effects of standard adoption. Over the three-year period, the share of authenticated senders decreases, as the prevalence of valid SPF configurations and correct Forward Confirmed Reverse DNS (FCRDNS) declines. In particular, we identify cases where IPv6 mail transport is enabled without corresponding SPF or PTR record configuration, leaving newly introduced IPv6 sending addresses unauthenticated. Improvements observed between first and last tests partially mitigate this effect, with SPF validity improvements peaking in 2025.
Artifacts. The source code of the measurement infrastructure and configuration of email servers and authoritative DNS servers is available as open-source software at `email-measurement-toolchain`.

Scenario	Act. Netw.	Inact. Netw.	Act. Netw. ACL	Inact. Netw. ACL	Null Route	Routing Loop
NR (S2)	○ 0	● 14	● 1	● 2	● 2	○ 0
AP (S3,S4)	○ 0	○ 0	● 4	● 5	● 3	○ 0
AU (S1)	● 14	○ 0	○ 0	○ 0	● 1	○ 0
PU ()	○ 0	○ 0	● 3	● 2	○ 0	○ 0
FP (S3,S4)	○ 0	● 1	● 1	● 2	○ 0	○ 0
RR (S5)	○ 0	○ 0	○ 0	○ 0	● 2	○ 0
TX (S6)	○ 0	○ 0	○ 0	○ 0	○ 0	● 15
∅	● 1	○ 0	● 4	● 3	● 9	○ 0

Table 1.1: ICMPv6 error message defaults of router images under test in six different routing scenarios. Two-letter abbreviations are used for Destination Unreachable codes and *TX* referring to *Time Exceeded*.

IPv6 Reconnaissance through ICMPv6 Error Messages

Chapter 3 investigates how ICMPv6 error messages can be leveraged to extend the observable IPv6 Internet. We follow a three-step methodology addressing *RQ2*. First, we study protocol defaults in a controlled environment with full control over router configurations. Second, we conduct active measurements in the IPv6 Internet to validate whether real-world networks exhibit the same behavior. Finally, we perform large-scale measurements to quantify the prevalence and impact of our findings.

To address *RQ2*, which asks what ICMPv6 error messages reveal about their sources, we construct a virtual router laboratory, as the primary source of ICMPv6 error messages are routers. In this controlled environment, we systematically enumerate router defaults by deploying multiple router operating systems within the network simulator GNS3 (gns3.com). Our setup includes images from core router vendors (Cisco, Juniper, Huawei) as well as devices commonly used at the network edge (OpenWRT, Mikrotik). For each router, we evaluate compliance with the ICMPv6 error message specifications defined in RFC 4443 [CDG06] across six routing scenarios.

Table 1.1 summarizes the results. By default, nearly all tested routers return ICMPv6 error messages, with only one out of fifteen vendors configured to suppress such responses. This observation confirms assumptions derived from preliminary IPv6 Internet test scans. Prior work on detecting IPv6 periphery routers collects source addresses of destination unreachable messages without differentiating error types [LLZ+21]. Our work goes a step further by analyzing what the specific error types reveals about the target network.

We observe that routers behave similarly in scenarios involving active and inactive networks as well as routing loops, but exhibit more diverse behavior in the presence of Access Control Lists (ACLs). The goal of this analysis is to determine whether ICMPv6 error message types can be used to distinguish active from inactive IPv6 networks, as only active networks contain hosts that are relevant for subsequent measurements. We classify an error message as *ambiguous* if it is returned for both active and inactive networks and therefore cannot be used to reliably distinguish between the two.

Under this definition, no single ICMPv6 error message type uniquely identifies active networks. For example, the error type *AU – Address Unreachable* is returned for active networks in scenario S1 by 14 out of 15 routers, but is also returned for inactive networks with null routes in scenario S5 (Juniper). However, we find that analyzing the timing of error messages enables a reliable distinction between these cases. In active networks, *AU* responses are returned only after a delay of two to three seconds, corresponding to the time required for neighbor discovery attempts. In contrast, *AU* responses in scenario S5 are returned immediately.

Having established these scenarios in a controlled environment, we next examine whether they are observable in the IPv6 Internet. This requires a ground truth distinguishing active from inactive IPv6 networks; however, no such dataset exists. The most comprehensive data source available for IPv6 measurements are hitlists, which contain collections of IPv6 addresses known to be active. We therefore develop an address-seeded approach, referred to as *BValue Steps*, which derives active and inactive networks from hitlist addresses. We apply this approach to a set of 47.9 K IPv6 addresses from the IPv6 hitlist service [GSF+18, ZSS+22, SKZ+23], finding that in 95% of cases, IPv6 networks labeled as active behave according to the findings in our controlled environment.

Finally, we conduct two large-scale measurements to quantify the overall extent of these observations in the routed /48 and /64 address space finding atleast one active /64 in 61% of responsive networks.

We further investigate what ICMPv6 error messages reveal about the router that returned the error message. During experiments in the virtual router lab, we observe that ICMPv6 configurations are often vendor-dependent. While error message types alone can already provide hints about the router vendor, real-world deployments typically expose only a subset of routing scenarios. We therefore identify the rate at which error messages are returned as a more robust fingerprinting signal. Vendor-specific rate limits allow us to infer router operating systems in the Internet core, as well as Linux kernel versions for routers deployed at the network edge.

In summary, the thesis shows that ICMPv6 error messages reveal:

1. **IPv6 networks are likely to return ICMPv6 error messages.** Our measurements show that 61% of IPv6 prefixes return error messages and 39% remain silent with ICMP probes. This further increases by 3 percentage points for UDP and is lowest for TCP (52.9%).

2. **BValue Steps classifies prefixes and prefix sizes from a known IPv6 address.** We detect suballocation borders in 72% (ICMP) of responsive prefixes while for the remaining 28 % the active network border equals the routed network border.

3. **Address Unreachable with a delay of more than a second indicates active networks**. We observe that *Address Unreachable* messages returned with a delay of more than one second are indicative of active networks, as they are triggered by the Neighbor Discovery process. In contrast, *Address Unreachable* messages returned without a noticeable delay indicate inactive networks.

4. **Active IPv6 networks are scarce.** We comprehensively probe the routed IPv6 Internet by probing 5Bn /48 prefixes and 6Bn /64 prefixes. At /48 granularity,

only 1.7% of IPv6 networks are identified as active. At /64 granularity, this share increases to 12%. Moreover, at least one active /64 network is observed in 61% of responsive /48 prefixes.

5. **Rate limiting of error messages reveals router vendors or kernel versions**. We measure ICMPv6 error message rate limits for 1.4 million routers in the IPv6 Internet. Using the number of paths on which a router resides, we compute a centrality score. For routers with a centrality greater than one, vendor-specific rate limits can be identified, whereas for routers with a centrality of one, rate-limiting behavior reveals differences between Linux kernel versions.

Artifacts. The source code for the measurements, the measurement data and Jupyter notebooks to replicate the results are publicly available under:
`icmpv6-destination-reachable`.

CountryMonitor

Chapter 4 introduces CountryMonitor, which leverages the absence of responses to detect regional- and provider-level Internet disruptions. This methodology does not require measurement infrastructure within the target country, assuming that IP-to-location data is available to assign disruptions to specific regions.

The detection of Internet disruptions has evolved from IP-based methods and high-speed probing [DWH13] to block-based outage detection algorithms such as Trinocular [QHP13]. To improve detection accuracy in sparse blocks, defined as blocks with a low number of responsive IP addresses, full block scans were developed to reconstruct block state from multiple Trinocular rounds [BH20].

This thesis deploys full block scans not derived from Trinocular data but conducted as active measurements. Specifically, the full state of Ukrainian /24 networks has been surveyed every two hours since the seventh day of the invasion of Ukraine. This methodology addresses *RQ3* by capturing unintentional delivery failures caused by damage to network infrastructure or power loss, thereby enabling the quantification of Internet disruptions in Ukraine.

Figure 1.4 visualizes the different outage signals that are collected to capture Internet disruptions. Unintentional delivery failures are captured by active ICMP probing while passively by monitoring prefix visibility in RouteViews BGP looking glasses. This allows CountryMonitor to track different type of outages. For example, when the border router fails and BGP prefixes become invisible indicates more severe outages, whereas power losses in Ukraine lead to partial outages affecting the loss of individual IPs while the number of /24 blocks remains constant.

The full block scans of the Ukrainian address space show:

1. **Visible IP churn:** IP addresses leave frontline oblasts at a faster pace than non-frontline regions. In Kherson, the number of geolocated IP addresses decreases by 62%. Of the 141K IPs observed in 2022, only 26% remain within the oblast.

2. **Increased provider coverage:** Measuring every IP address within a block, rather than relying on sampling, yields a more granular outage signal. This approach detects disruptions affecting many smaller operators, covering 1 674 out of 1 743 providers, whereas IODA covers only 19 % (333 providers).

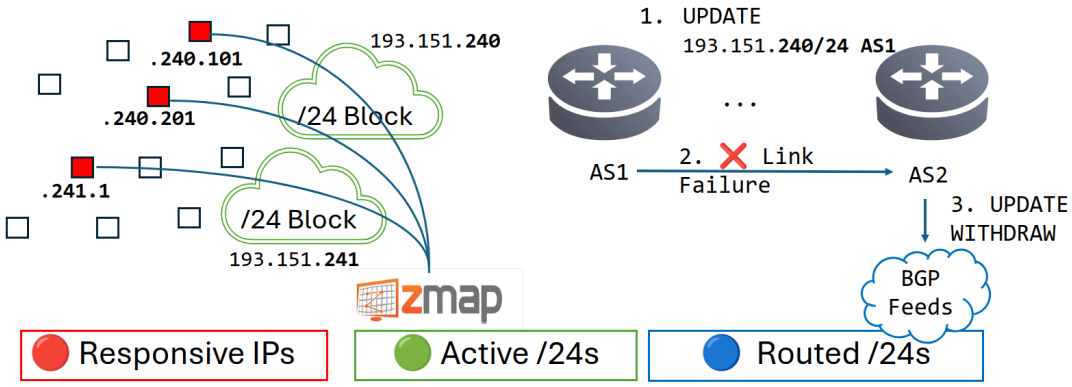


Figure 1.4: CountryMonitor captures two active based on active probing and one passive outage signal based on prefix visibility in routing tables.

3. **Periods of significant strain:** We identify periods during which the Ukrainian Internet experienced substantial strain. Internet disruptions peak in the winter of 2022/23 and throughout 2024, correlating with an increasing number of attacks on Ukraine's power infrastructure.

4. **Outages in non-frontline regions correlate with power outages:** We observe a substantial overlap between days with Internet disruptions (maximum 2 822 hours, average 686 hours) and reported power outages (1 951 hours) affecting more than 50 % of oblasts, as reported by Ukraine's transmission system operator Ukrenergo. This overlap yields a Pearson correlation coefficient of 0.725.

5. **Insights into providers in Kherson:** Regional classification detects long-term trends in IP geolocation and enables the assignment of IP blocks and ASes to regions within Ukraine. During a telecommunications search in May 2022 that affected one of Status ISP's offices, we observe a corresponding outage in our dataset, corroborating video footage recorded by the provider. Following the recapture of Kherson city, the three /24 blocks in Kherson experience a ten-day outage, followed by a gradual reconnection exhibiting diurnal power cycles between 08:00 and 18:00.

Artifacts. Due to the sensitive nature of IP-level data, the access to measurement data is restricted for research purposes only and includes either block level availability data or anonymized IP level responsiveness if required:
`countrymonitor.github.io`.

1.5 Structure of this Work

Chapter 1 introduces the problem of delivery failures, motivates the research, formulates the research questions, describes the methodology, and summarizes the main contributions of the thesis.

The introduction is followed by the individual publications. Each publication is preceded by a summary page. Chapters 2 and 3 focus on intentional delivery failures.

Chapter 2 presents the email delivery testbed and a measurement study demonstrating how delivery failures can be leveraged to assess security and protocol adoption in email delivery. Chapter 3 examines ICMPv6 error messages beyond their source address and shows how active measurements can intentionally trigger delivery failures to classify IPv6 networks and routers.

Chapter 4 addresses unintentional delivery failures using a combination of active and passive signals. By applying full block scans to the Ukrainian address space over a three-year period, this chapter demonstrates how such measurements can be used to quantify Internet disruptions.

Finally, Chapter 5 summarizes the impact of delivery failures on extending the observable Internet. It combines insights from all three publications on the Internet, discusses limitations of the presented methodology, as well as directions for future work.

2 Measuring Email Delivery

This chapter is based on the first publication, which focuses on measuring email delivery from the sender's perspective.

Title	Not that Simple: Email Delivery in the 21st Century
Authors	Florian Holzbauer, Johanna Ullrich, Martina Lindorfer and Tobias Fiebig
Publication Status	This paper is included in the Proceedings of the 2022 USENIX Annual Technical Conference. Pages 295-308. (ISBN: 978-1-939133-29-8). CORE2023-Ranking: A. Acceptance Rate (Long Papers): 16.71%
Pointer	`https://usenix.org/system/files/atc22-holzbauer.pdf`
Author Contributions	Florian Holzbauer: Main author, data evaluation (statistics, visualizations), large provider and spam measurement setup and execution, paper writing. Johanna Ullrich: Supervision, paper writing and participation in weekly discussions. Martina Lindorfer: Supervision and participation in weekly discussions. Tobias Fiebig: Supervision and technical implementation of email servers and measurement setup, promotion of measurement campaign, participation in weekly discussions and paper writing. Technical advice on email standards and centralization.
Artifacts	Email Testbed: `https://email-security-scans.org` Code: `https://github.com/ichdasich/email-measurement-toolchain`
Reference	[HULF22]

Not that Simple: Email Delivery in the 21st Century

Florian Holzbauer
SBA Research

Johanna Ullrich
University of Vienna[*]

Martina Lindorfer
TU Wien

Tobias Fiebig
Max-Planck-Institut für Informatik

Abstract

Over the past two decades, the number of RFCs related to email and its security has exploded from below 100 to nearly 500. This embedded the Simple Mail Transfer Protocol (SMPT) into a tree of interdependent and delivery-relevant standards. In this paper, we investigate how far real-world deployments keep up with this increasing complexity of delivery- and security options. To gain an in-depth picture of email delivery apart from the giants in the ecosystem (Gmail, Outlook, etc.), we engage people to send emails to eleven differently configured target domains. Our measurements allow us to evaluate core aspects of email delivery, including security features, DNS configuration, and IP version support on the sending side across different types of providers.

We find that novel technologies are often insufficiently supported, even by large providers. For example, while 65.4% of email providers can resolve hosts via IPv6, only 44.3% can also deliver emails via IPv6. Concerning security features, we observe that less than half (41.5%) of all providers rely on DNSSEC validating resolvers, and encryption is mostly opportunistic, with 89.7% of providers accepting invalid certificates. TLSA, as a DNS-based certificate verification method, is only used by 31.7% of the providers in our study. Finally, we turned our eye to the impact modern standards have on unsolicited bulk email (SPAM). We found that greylisting is effective, reducing the SPAM volume by roughly half while not impacting regular delivery. However, and interestingly, SPAM delivery currently seems to focus on plaintext IPv4 connections, making IPv6-only, TLS-enforcing inbound email servers a more effective anti-SPAM measure—even though it also means rejecting a major portion of legitimate emails.

1 Introduction

Electronic mail (email) relies on the Simple Mail Transfer Protocol (SMTP) for delivery. This protocol was first specified in 1982 in RFC 821 and is now close to celebrating its 40th birthday [39]. SMTP had two design goals, namely to allow *reliable* and *efficient* delivery of emails. As with many protocols of the time, security and authenticity were not priorities [16]. In fact, anyone could relay emails through an SMTP server, which was the default configuration for many email servers – like Sendmail – until the late 1990s [3].

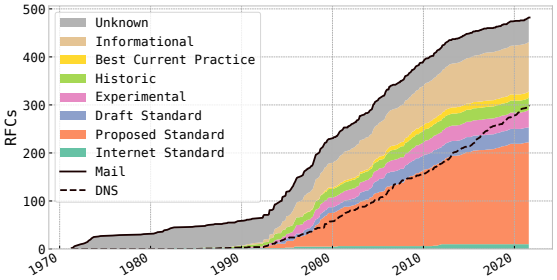


Figure 1: Overview of the explosion of email-related standards ("SMTP Camel"), compared to DNS-related standards.

However, the practical reality of the Internet led to increased security and authenticity requirements [16]. Since the mid-1990s, hundreds of protocols and extensions have been introduced to cover these gaps, as illustrated in Figure 1. In order to authenticate email, attempts mostly rely on the *Domain Name System* (DNS), which, in turn, suffers from authenticity issues. To address those issues, the *DNS Security Extensions* (DNSSEC) were introduced in 1999, which enabled signing DNS entries [1]. Besides authenticity, the original email protocol faced other security-related challenges, most notably confidentiality, as emails were exchanged in plaintext. In addition to end-to-end encryption approaches like Pretty Good Privacy (PGP) [7], this led to an extension of SMTP for Transport Layer Security (TLS) [18]. Finally, like all protocols on the Internet, SMTP was also affected by the introduction of IPv6.

All these factors have turned the *simple* from SMTP to *complex*. To outline this increase in complexity, we created the *SMTP Camel* in Figure 1 (after the famous DNS Camel

[*] Christian Doppler Laboratory for Security and Quality Improvement in the Production System Lifecycle, Security & Privacy Group, Faculty of Computer Science

of Bert Huber, who illustrated the complexity of DNS with *"How many features can we add to this protocol before it breaks?"* [23]). Figure 1 visualizes RFCs related to email – and, for reference, DNS. We compiled this list by performing a title/keyword search on all RFCs on September 28, 2021.[1] In total, we found 481 email-related RFCs compared to 298 DNS-related ones. Among these, more than half of the RFCs belong to the standards track, representing mature standards. We see no development in draft standards as they were declared as deprecated in 2010 [21]. In June 2021, we reached a total of 225 proposed standards. Proposed standards only advance to Internet standards once they have *"widespread deployment of multiple implementations from different code bases"* [21]. Currently, only eleven email-related RFCs have met this requirement, and also the handling of this guideline by the Internet Engineering Task Force (IETF) varies. This indicates that the development of new standards has outpaced their implementation. Furthermore, since the latest email measurement study in 2020 [31], seven new email-related RFCs have been published.

In this paper, we investigate how the increasing number of additional standards has influenced email delivery in the wider ecosystem. Related work already demonstrated that adoption rates of email-related standards are low and implementations often rely on insecure defaults [8, 14, 17, 22, 26, 31, 37, 45]. However, previous work predominantly focused on large operators, such as Google (Gmail) or Microsoft (Outlook), and did not investigate fundamental aspects of email standards, like supported IP versions and the DNS infrastructure of sending systems. We take a step back and investigate the most fundamental aspects of email in transit across a wide sample going beyond major email providers.

To accomplish this, we introduced eleven target address configurations to verify how email providers implement email-related standards and protocols, i.e., we set up systems that – depending on the remote server's configuration and implementation – either do or do not receive measurement emails. Our measurement technique allows us to measure IP support, STARTTLS configuration, DNSSEC validation, and how different SMTP applications react to greylisting, an anti-SPAM technique by which incoming emails are initially rejected. Our focus is on protocols that influence email delivery once an email has been submitted. To increase the providers' coverage, we crowdsourced the sending of emails to participants recruited through mailing lists and social media.

As a result, we collect emails from three different sources, spanning (1) small participants in the email ecosystem, (2) large providers, and (3) unsolicited bulk email, aka SPAM. We are the first to discuss the impact of new and established standards on email delivery, as – in contrast to most related measurements – we rely on actively collecting emails, allowing us a more in-depth view of email server configurations.

[1] https://www.rfc-editor.org/search/rfc_search_detail.php

In summary, we make the following contributions:

- We introduce a new ranking method using passive data to find the top 15 email providers. Our results highly overlap with Liu et al. [32], while causing significantly less measurement overhead (see Section 3).
- We illustrate challenges in the interoperability between large centralized operators and smaller operators, including how the ability to deliver emails as the main objective limits the adoption of new network and security protocols. We describe how our datasets cover different actors in the email ecosystem in Section 4.
- We are the first to measure and connect the impact of protocol extensions in protocols email relies on – DNS(SEC) and IPv6 – to email delivery and the contrast between smaller and larger providers (see Section 5).
- We illustrate protocol support and compliance in the heavy-tail of the email ecosystem, i.e., in a large set of smaller email operators, and contrast this to earlier work and patterns found in large providers (see Section 6).
- Based on our results, we derive recommendations for email system operators on how they can utilize modern protocol compliance to – currently – reduce SPAM delivery (see Section 7).

Artifacts: Our measurement can be executed using any valid domain and a set of machines connected to the Internet. Along with our paper, we publish a setup-documentation and the scripts we used to receive and analyze emails sent to our systems at https://github.com/ichdasich/email-measurement-toolchain. For privacy reasons, we cannot publish our email dataset. This also applies to the SPAM dataset, as even SPAM may contain PII, for example in the recipient addresses.

2 Background: Protocols and Standards

In this paper, we focus on standards influencing email delivery between email servers, i.e., the Mail Transfer Agent (MTA). Email submission, e.g., the communication between Mail User Agent (MUA) and Mail Submission Agent (MSA), is not part of our study. We focus on IP- and DNS-related mechanisms that impact delivery. Interpretations of higher-level delivery security features, like the Sender Policy Framework (SPF) [27], DomainKeys Identified Mail (DKIM) [9], Authenticated Received Chain (ARC) [4], and Domain-based Message Authentication, Reporting, and Conformance (DMARC) [30] are out of scope for our study, as they only influence the receiver's decision on whether to accept incoming emails or not. We also did not include MTA Strict Transport Security (MTA-STS) in our study as this RFC was too recent

when we set up our infrastructure [33], but Section 3 describes how our work can be extended to include it in the future.

IPv4 [38] and IPv6 [10]. Since addresses in the 2^{32} bit address space of the Internet Protocol Version 4 (IPv4) are running out [41], Internet Protocol Version 6 (IPv6) with a 2^{128} bit address space was introduced in the late 1990s. Two concurrent IP versions introduce a great challenge in terms of interoperability on the network layer, especially as the adoption of IPv6 is still slow [25]. IP version support impacts email delivery *indirectly* via DNS support, i.e., the authoritative and recursive servers support the same IP version, and *directly*, i.e., in terms of whether the involved email servers both support the same IP version. Servers can support IPv4, IPv6, or both—also referred to as "dual-stack."

DNSSEC [5]. The DNS-Security Extensions (DNSSEC) provide authenticity to DNS responses by signing DNS entries via a keychain along the path of the DNS tree. A DNSSEC validating recursor responds with `SERVFAIL` in case of a validation error. As a consequence, the target domain cannot be resolved, and email delivery fails. Hence, in case of misconfigurations – common in system operations [11] – or attacks, the DNSSEC validation behavior of DNS resolvers at email-sending servers becomes important for email delivery. Similarly, DNSSEC is a prerequisite for DANE (see below).

STARTTLS [19]. The SMTP Service Extension for Secure SMTP over TLS (STARTTLS) enables TLS for email delivery. The connection is established on the same port as SMTP. The original SMTP handshake remains in cleartext. Sending- and receiving servers can (1) not support TLS, (2) support TLS and cleartext, (3) enforce TLS. TLS can be configured either in an (a) opportunistic or (b) strict manner. While opportunistic TLS configurations allow for encrypted connections not validating the remote certificate, strict configurations cause email delivery to fail in case of (1) invalid certificates, (2) not supporting mandatory ciphers, or (3) a connection to a non-TLS-supporting server. In turn, this can then impact email delivery, depending on whether a connection can be established or not.

DANE [20]. The DNS-Based Authentication of Named Entities (DANE) prevents MTA-to-MTA transport encryption from downgrade attacks, even in the absence of certificates signed by a certificate authority (CA); this is done through recording valid CA or end-entity certificates for a domain name via the TLSA DNS record. Trusting/guaranteeing the authenticity of TLSA records (i.e., preventing MITM and DNS cache poisoning scenarios) requires the use of DNSSEC, as described above. Several email server implementations, including Sendmail and Microsoft Exchange, do not yet support requesting TLSA records, in contrast to for example, Postfix and Exim [31].[2] DANE can be implemented similar to TLS in an opportunistic or mandatory manner. Email delivery fails for both opportunistic and mandatory configurations if a signed TLSA record is available but certificate validation fails or for mandatory configurations if no TLSA record can be found.

[2]Microsoft announced support after our measurement period in Feb, 2022 (see https://techcommunity.microsoft.com/t5/exchange-team-blog/releasing-outbound-smtp-dane-with-dnssec/ba-p/3100920)

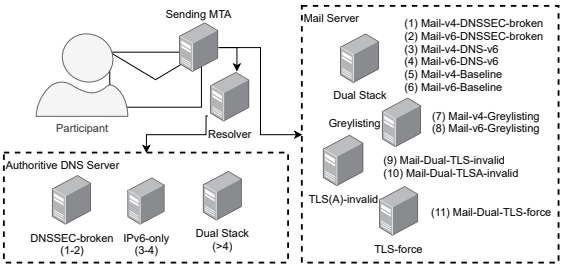


Figure 2: Overview of our measurement setup: 3 DNS servers serve 4 email servers with 11 differently configured target addresses.

Anti-SPAM (Greylisting [29]). Greylisting is one of the most simplistic approaches to reduce SPAM emails. It works by initially responding with SMTP code `4xx temporary failure`. While reputable servers usually re-attempt email delivery after several minutes, many SPAM senders do not keep enough state for this. For email delivery, greylisting introduces delays, and email delivery fails if an implementation does not attempt redelivery.

3 Methodology

Measurement Platform. Our measurement setup consists of four email servers running Postfix 3.6 [40] on OpenBSD 6.7 [36] in OpenBSD virtual machines (VMM). As we conduct non-performance bound network measurements, the exact type and model of the used hardware are not relevant to our measurement platform. Furthermore, we rely on three PowerDNS authoritative nameservers in version 4.3.1 to measure the impact of different DNS server setups. We configured a non-default TTL of 300 seconds for all entries in our DNS zones to minimize the impact of caching, i.e., a DNS resolver used by multiple study participants. This also affects our weekly spam domain rotations, pointing them at different measurement target addresses. However, we consider a maximum overlap of five minutes in comparison to a one-week measurement period negligible. IPv6 connectivity to our systems was provided via a Hurrican Electric IPv6 tunnel, while IPv4 connectivity was provided via dedicated IP space from the RIPE region. On these systems, we set up eleven email addresses, as shown in Figure 2. For each of these addresses, we applied different configuration states, which either enable or prevent remote servers from sending emails to them, depending on their own configuration state. This allows us to measure the remote servers' email delivery capabilities and protocol use by measuring whether they are able to deliver

```
measurement@v4-mail.example.com
measurement@v6-mail.example.com
measurement@v4-mail.v6only.example.com
measurement@v6-mail.v6only.example.com
measurement@v4-mail.dnssec-broken.example.com
measurement@v6-mail.dnssec-broken.example.com
measurement@v4-mail-greylisting.example.com
measurement@v6-mail-greylisting.example.com
measurement@mail-tls-force.example.com
measurement@mail-tls-invalid.example.com
measurement@mail-tlsa-invalid.example.com
```

Figure 3: List of email addresses for the 11 target configurations.

emails to these email addresses. We then asked participants to send *one* email with all measurement addresses in the `To:` field. If we do not receive a message at a specific target address but see in our baseline that the target is included in the `To:` header, we know that the respective feature is not supported. The target addresses can be easily extended to cover new protocols, e.g., MTA-STS [33] was introduced as a barrier against downgrade or interception attacks for domains that are unable to deploy DNSSEC. MTA-STS can be measured by adding two new target addresses in the future. One could implement the TLS-RPT standard to measure TLS reporting frequency, and the other could measure if providers still deliver emails in case of an enforced MTA-STS policy with non-matching MX records.

3.1 Target Address Configurations

We configured the following eleven different email addresses at the unique destination domains listed in Figure 3. Below, we describe the purpose of each of these addresses, i.e., which configuration parameters we tested with them:

IP Support. In order to test basic delivery behavior, we created for both IPv4 (`measurement@v4-mail.`, *Mail-v4-Baseline*) and IPv6 (`measurement@v6-mail.`, *Mail-v6-Baseline*) one address which is configured with no restrictions on delivery. Similarly, we created distinct IPv4- and IPv6 addresses for the DNS and greylisting measurements described below. Note that during our study, we noticed that our choice not to support STARTTLS on this system did indeed introduce an unexpected parameter in the case of senders that enforce STARTTLS use. In turn, this allowed us to detect six providers that enforce STARTTLS for outgoing emails.

DNS Recursion IPv6 Support. To test whether the recursive resolvers of an email sending host support IPv6, we created a subdomain that can only be resolved via IPv6, i.e., the zone had only `AAAA` glue records, and the hosts in the zone's `NS` records also only have `AAAA` records. Under that domain, we then again created two addresses for IPv4 and IPv6 delivery (`measurement@v4-mail.v6only.`, *Mail-v4-DNS-v6* and `measurement@v6-mail.v6only.`, *Mail-v6-DNS-v6*).

DNSSEC Validation. To test if the remote site validates DNSSEC, we set up a subdomain with a non-matching DS RRset in the parent, i.e., we provide a public key in the parent zone that does not match the key with which records are signed in our zone. Hence, a DNS recursive resolve validating DNSSEC is unable to validate DNSSEC for our domain and should therefore refuse to resolve it. Thus, an email server using a validating resolver cannot deliver emails to that domain. Under that domain, we again created two addresses for IPv4- and IPv6 delivery (`measurement@v4-mail.dnssec-broken.`, *Mail-v4-DNSSEC-broken* and `measurement@v6-mail.dnssec-broken.`, *Mail-v6-DNSSEC-broken*).

TLS Configuration. In order to test the TLS and TLSA behavior of sending hosts, we configured three email addresses that required the use of TLS to deliver emails:

- `measurement@mail-tls-force.` *Mail-Dual-TLS-force* on a correctly configured TLS enabled server.
- `measurement@mail-tls-invalid.` *Mail-Dual-TLS-invalid* on a server that provides a certificate with a non-matching CN/DNS0 entry.
- `measurement@mail-tlsa-invalid.` *Mail-Dual-TLSA-invalid* on a server that has a TLSA record configured, which does not match the supplied certificate.

This setup allows us to verify if systems (1) support STARTTLS, (2) perform opportunistic encryption, and (3) verify TLSA records. Due to a misconfiguration, these systems initially did not support TLS1.3. Hence, remote systems that only support TLS1.3 would be unable to deliver their emails. We were able to isolate the affected cases (76 emails from 29 providers) and reconstructed the actual state from the stored SMTP sessions, as the abort conditions differ between 'not supporting TLS,' 'rejecting the certificate/TLSA record,' and 'not having a matching cipher.'

Anti-SPAM (Greylisting). To identify RFC-compliant SMTP implementations, and as an additional control, we set up Postgrey that performs greylisting as an anti-SPAM measure (`measurement@v4-mail-greylisting.`, *Mail-v4-Greylisting* and `measurement@v6-mail-greylisting.`, *Mail-v6-Greylisting*). By configuring these addresses, we can test the impact of greylisting on average SPAM received and check whether legitimate email servers support multiple delivery attempts.

3.2 Email Collection and Recruitment

In order to provide different views on email delivery, we target three types of actors in the email ecosystem: (1) Regular providers by actively engaging users to send emails to our measurement system. (2) A set of top-ranked email providers

Table 1: Recruitment channels for study participants.

Type	Name	Description
Blogs	RIPE Labs	Article in RIPE's Research Blog/Newsfeed
	APNIC	Article in APNIC's Blog/Newsfeed
Social Media	Twitter	Tweets by researchers involved in the project
	LinkedIn	Posts by researchers involved in the project
	Reddit	Reddit post to /selfhosted
Mailing Lists	NANOG	North American Network Operator List
	INNOG	Indian Network Operator List
	AFNOG	African Network Operator List
	SAFNOG	South African Network Operator List
	DENOG	German Network Operator List
	NLNOG	Dutch Network Operator List
	IRTF-MAPRG	Network Research Interest Group at IETF/IRTF
	MAIL-OPS	Global Mail Operator List
Presentations	Internet.nl	Presentation at an organization promoting the adoption of security standards
Personal	-	Colleagues and personal networks, especially in the APNIC and LACNIC regions

by registering user accounts and sending emails. (3) Spammers by registering expired domains and collecting unsolicited emails targeting these domains.

Regular Providers. To collect emails, we actively engaged Internet users to participate in our study. We recruited participants via a social media campaign on Twitter, LinkedIn, and Reddit, via mailing lists focusing on email and network operators, blog articles promoted by Internet governance bodies, and our personal networks (see Table 1). Our recruitment message asked users to visit our website, which provided instructions on how the reader can participate in our study, what the purpose of our study is, and what data access and deletion rights they have. One critical aspect was to ensure that we would be able to distinguish whether an email to one of our measurement hosts was sent and not delivered or not sent at all. Thus, we instructed participants to add all measurement addresses to the `To:` field of a single email. In case a participant's provider performed pre-filtering, e.g., did not accept delivery to domains they cannot resolve, we removed affected emails from the dataset.

Large Providers. In order to rank email providers, we rely on the passively collected Farsight SIE DNS dataset [43]. This enables us to count email servers to which a lot of domains point their `MX` records, i.e., email servers used for a lot of domains. We assume that the number of domains using a provider's email servers correlates to the provider's size. For our ranking, we use DNSDB MX data extracted for November 2020, which includes data of 73,705,268 different `MX` lookups. We do not rank providers based on the amount of `MX` lookups, as low TTLs or different DNS resolver setups might bias the number of lookups. For each `MX`, we extract the public suffix, i.e., 'example.com' for 'mail.example.com' and 'example.co.uk' for 'mail.example.co.uk' using the Public Suffix List [35]. This results in 23,378,583 different public suffixes. We rank public suffixes of `MX` records by counting the number of different domains pointing their `MX` records towards them. We then register accounts at the top 15 providers according to this ranking to send emails to our target domains, as done in prior work [17, 22, 31, 32, 45]. This enables us to compare email delivery from regular providers with an exclusive set of large providers, but also to compare the results of our measurement pipeline to the results of prior work.

Table 2: Categories of domains from ExpiredDomains.

Category	Description
1990s	Domains with the first screenshot available on Archive.org between 1990 and 2000 (= "birth year")
alexa	Domains selected based on Alexa traffic rank
backlinks	Domains based on number of Majestic external backlinks
dmoz	Domains found in the latest snapshot of dmoz.org (~2017)
majestic	Domains with low Majestic million global rank
wiki	Domains with high numbers of Wikipedia links

Spammers. To collect SPAM emails, we registered expired domains that are still likely to receive SPAM. To do so, we relied on `expireddomains.net` for a list of domains [42]. To increase the likeliness that respective domains still receive SPAM, we chose them from different categories, based on their age ("birth year," i.e., the first entry in `Archive.org`), their popularity according to rankings from Alexa and Majestic, and the number of links from Wikipedia and the (now defunct) DMOZ content directory. Table 2 lists these categories; Table 3 lists the domains in each category, as well as the volume of SPAM we received during our measurements.

Once registered, we pointed `MX` records of respective domains at our target domains. To identify if domains still receive SPAM, we executed a three-week baseline measurement. During this period, all 50 re-registered domains pointed their `MX` records to the `MX` of *Mail-v4-Baseline*, i.e., our most basic configuration. We classified the domains' value for our measurement based on the amount of SPAM received as *high* (multiple times a week), *low* (once a week), and *none* (none received). To verify that received messages are SPAM, we consulted four active DNS blocklists: `bl.spamcop.net`, `ips.backscatterer.org`, `pbl.spamhaus.org` and `sbl.spamhaus.org`. We continuously verified the liveness of these blocklists by requesting IP 127.0.0.2 as a test record.

In total, we found 26% of domains receive SPAM on a regular basis, thus falling into category *high*. In the next step, we pointed high-value SPAM domains towards a set of our target addresses in a weekly rotation until each domain had been pointed at each target at least once. This allowed us to monitor the change in SPAM volume based on the corresponding test conditions. For these measurements, we relied on a reduced set of target addresses. As we only received individual emails and did not simultaneously measure all conditions for each sender, we did not differentiate IPv6 behavior for different target addresses. We only verified general IPv6 support (*Mail-v6-Baseline*), IPv4 sending for IPv6 only DNS (*Mail-*

Table 3: Re-registered domains for SPAM collection and the amount of SPAM emails we received for each of them.

	Category	Domain	Spam Frequency
1	1990s	anx-chicago-rawhide.com	low
2	1990s	intecconstruction.com	high
3	1990s	michael-rauch.com	-
4	1990s	mmf-maintenance.com	high
5	1990s	sapphire-controls.co.uk	high
6	1990s	stratos-bde.com	low
7	alexa	inkpreneur.com	-
8	alexa	jsmmf.org	-
9	alexa	kenyamalikmotors.com	-
10	alexa	lafdo.com	high
11	alexa	nepaltravelcentre.com	high
12	alexa	olakassen.com	-
13	alexa	onmylevelchey.com	-
14	backlinks	18Chaa.com	low
15	backlinks	521qiangwweisizu.com	-
16	backlinks	cretms.com	low
17	backlinks	fotiis.com	-
18	backlinks	g6china.com	-
19	backlinks	io365f.com	-
20	backlinks	io365i.com	-
21	backlinks	theproxylist.co.uk	-
22	backlinks	tuncayparlak.com	low
23	backlinks	vous-y-etes.com	-
24	dmoz	beechamsdrivingschool.co.uk	high
25	dmoz	bilder-touren-allgaeu.de	-
26	dmoz	costatehogrally.com	low
27	dmoz	djk-handball-coesfeld.de	-
28	dmoz	leben-ohne-alkohol.eu	low
29	dmoz	navesprefabricadassprint.com	high
30	dmoz	parissi.eu	-
31	dmoz	pringfieldfarms.co.uk	-
32	dmoz	printshopleeds.co.uk	low
33	dmoz	smugglegame.com	high
34	dmoz	sotralentz.es	high
35	dmoz	survivalschool.ch	high
36	dmoz	thermoboss.net	low
37	majestic	djmzengaman.com	-
38	majestic	eiecan.eu	-
39	majestic	hkmxdna.com	-
40	majestic	keerthiwrites.com	-
41	majestic	kientrucnghethuatduongdai.com	-
42	majestic	printspixelz.com	-
43	majestic	studiopaeez.com	low
44	majestic	thi-marprojects.be	high
45	wiki	catholic-church-corfu.org	low
46	wiki	grandeguerrafvg.org	-
47	wiki	iranairlinenews.com	-
48	wiki	mosul-network.org	-
49	wiki	unaf-foot.com	-
50	wiki	worldipcomgroup.com	low

v4-DNS-v6), DNSSEC behavior (*Mail-v4-DNSSEC-broken*), as well as our three TLS configurations.

3.3 Ethical Considerations

As our measurements focus on the technical aspects of the involved email setups, this study was not within the scope of our local human subject research ethics council. Nevertheless, we informed participants about the purpose of our data collection, which information we collected, and that they could withdraw from the study at any time. We received one request to be removed from the dataset and complied with this request immediately. In addition, we followed network measurement best practices as outlined in the Menlo report [6, 12].

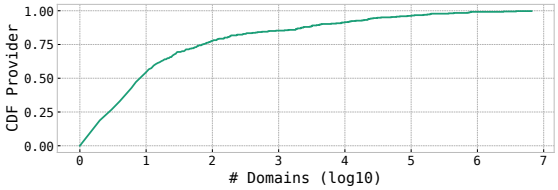


Figure 4: Validation of regular study participants tend to be/use small email providers. We match regular providers to the passive DNS ranking.

This means that we took the necessary technical precautions to protect the only Personally Identifiable Information (PII) we collect, i.e., the sending email addresses. We removed these addresses from our dataset as soon as possible before we started the aggregation of our collected data. Also, since the provider name might reveal PII, we do not publish or share provider names of smaller providers. For our measurements of large providers, we registered accounts ourselves and published their names for better comparison to related work, in accordance with common practice for email-related measurements [17, 22, 31, 32, 45].

4 Datasets

By following our approach, we collected three datasets covering (a) regular providers by volunteers sending emails to our measurement infrastructure, (b) large providers by registering accounts and sending emails ourselves, and (c) spammers by collecting unsolicited emails sent to re-registered domains.

(a) Regular Providers. Between July 4, 2020 and October 29, 2021 we received a total of 5,847 emails. After filtering emails that do not cover all eleven target addresses in the `To:` field, a total of 4,660 emails sent by 622 study participants remained for further analysis. There is a clear dominance of emails from European countries, see Table 5, a consequence of recruiting via our personal channels (e.g., on Twitter).

Multiple participants used the same infrastructure to send emails; beyond, emails of the same user might be sent by multiple servers in the same domain (e.g. `server1.domain.any` and `server2.domain.any`). Thus, we grouped the data set using the email servers' first-level domain (EHLO name) at the granularity of providers. This yields a total of 436 providers.

(b) Large Providers. Analysis of the Farsight SIE DNS dataset revealed the top 15 providers as presented in Table 4. We noticed a large gap in served domains even within the top 15 providers, ranging from 14.1% (Google) to 0.68% (1&1) of first-level domains (FLDs) in our passive DNS dataset. The top 15 providers jointly serve 33.8% of all FLDs with `MX` hosts. To gain an overview of provider sizes in our regular dataset, we matched regular providers with domains in the

Table 4: Top 15 providers based on passive DNS data. Providers greyed out have no online email service, e.g., *Above.com* is a domain broker.

										IP support				DNSSEC		Spam		TLS		
NR	Provider	2015 Durumeric [14]	2015 Foster [17]	2018 Hu [22]	2020 Lee [31]	2021 Tatang [45]	2021 Liu [32]	# Dom.	% Dom.	Mail-v4-Baseline	Mail-v6-Baseline	Mail-v4-DNS-v6	Mail-v6-DNS-v6	Mail-v4-DNSSEC-broken	Mail-v6-DNSSEC-broken	Mail-v4-Greylisting	Mail-v6-Greylisting	Mail-Dual-TLS-force	Mail-Dual-TLS-invalid	Mail-Dual-TLSA-invalid
1	Google	⋆	△	•	□	◇	⊙	9,148,093	14.08	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	Microsoft	⋆			□	◇	⊙	3,869,507	5.95	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	GoDaddy	⋆					⊙	2,453,911	3.78	✓				✓		✓		✓	✓	✓
4	OVHCloud	⋆					⊙	1,292,615	1.99	✓						✓		✓	✓	✓
5	Enom						⊙	871,527	1.34	✓				✓		✓		✓	✓	✓
6	One.com							797,194	1.23	✓						✓		✓	✓	
7	Namecheap						⊙	784,486	1.21	✓		✓		✓		✓		✓	✓	✓
8	Strato						⊙	762,923	1.17	✓	✓	✓	✓			✓	✓	✓	✓	
9	Yandex	⋆	△				⊙	759,482	1.17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	SiteGround						⊙	712,418	1.10	✓		✓		✓		✓		✓	✓	
11	H-email.net							575,451	0.89											
12	Above.com							469,500	0.72											
13	Beget						⊙	447,284	0.69	✓				✓		✓		✓	✓	✓
14	Tencent	⋆	△				⊙	442,064	0.68	✓		✓		✓		✓		✓	✓	✓
15	1&1							440,558	0.68	✓	✓	✓	✓			✓	✓	✓	✓	✓
Optimal Configuration										✓	✓	✓	✓			✓	✓	✓	✓	

Table 5: Number of countries/emails/AS per region. Our social media promotion led to an increased number of emails from European countries. We skipped large providers as geographical data has no impact on our provider ranking.

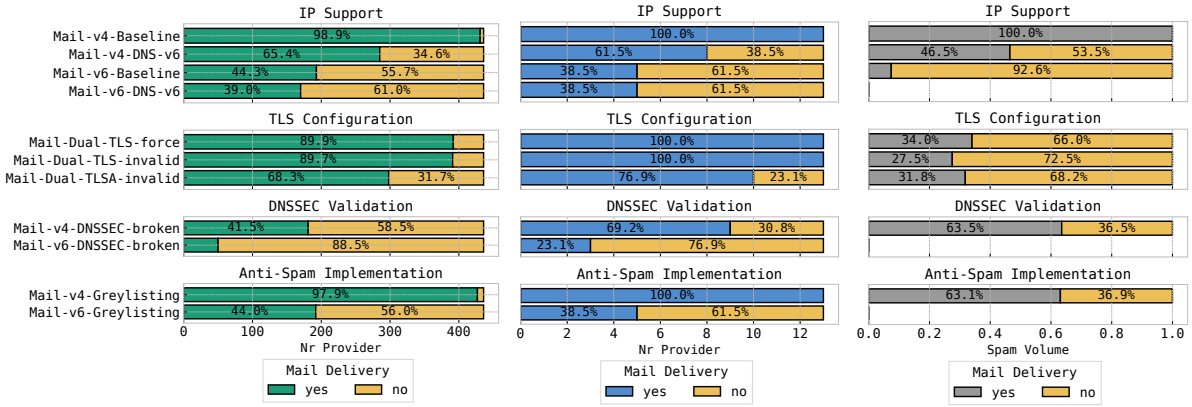
	Region	Africa	Asia	Europe	N. America	Oceania	S. America
Regular	Countries	5	12	30	2	1	3
	Emails	48	168	3,368	1,045	1	30
	ASes	5	19	202	60	1	3
SPAM	Countries	22	32	36	15	2	11
	Emails	95	2,056	1,963	2,437	17	204
	ASes	50	254	234	170	9	119

passive DNS dataset. Figure 4 shows the amount of FLDs pointing at each of the study participants' domains for email. 80% of regular providers have less than 150 domains relying on them for email service. Comparing our top 15 providers with previous work, we find the largest overlap, namely eleven providers, with Liu et al. [32], who used a five-step approach including MX records, Banner/EHLO messages, and TLS certificates to detect large email providers. Previous work relying on manual ranking results in less overlaps, namely six [14], three [17], two [31, 45], and one [22] (see Table 4), and suggests that human perception of providers is different from their actual dominance in the email ecosystem.

(c) Spammers. We executed SPAM measurements in three phases. First, we conducted a baseline measurement from March 30, 2021 to April 6, 2021. Next, we pointed SPAM domains to our other target addresses in a weekly rotation. Finally, we did another baseline measurement to ensure that the baselines remained stable over our observation time. We received a total of 6,772 unsolicited emails. Thereof, 4,442 (65.7%) were classified as SPAM by one of our four DNS blocklists, suggesting that emails towards the re-registered domains are indeed SPAM. We included all received emails in our further analysis. In comparison to our regular provider dataset, SPAM emails are not dominated by a single region (see Table 5). In comparison to regular and large providers, we can only measure the SPAM volume and its reduction in dependence of the different configurations.

5 Results

For each of the three datasets, namely (a) *regular providers*, (b) *large providers*, and (c) *spammers*, Figure 5 shows the ratio of delivered to undelivered emails per target address. We provide the individual results for the top 15 providers, including a line indicating the optimal configuration, in Table 4. The optimal configuration includes IPv4- and IPv6 support for both email servers and DNS resolvers. Regarding TLS, providers should implement opportunistic STARTTLS, i.e., still use transport encryption when facing self-signed or expired certificates.



(a) *Regular Providers* (b) *Large Providers* (c) *Spammers*

Figure 5: Impact of different target address configurations on email delivery. For our investigation of spammers we skipped the IPv6 target addresses other than the baseline (this affects greylisting, DNSv6, DNSSEC).

However, they should validate TLSA records and reject email delivery in case of an invalid record. As a foundation for DANE and other DNS-based security standards, a provider should rely on a DNSSEC supporting and -validating resolver. Looking at the top 15 providers, we find major discrepancies for even the largest providers. We discuss our measurement results on IP support, TLS configuration, DNSSEC validation, and anti-SPAM implementation in the following sections.

5.1 IP Support

Email Servers. The *Mail-v4-Baseline* is configured without any restrictions on email delivery. For regular providers, however, this baseline is reduced by 5/436 (1.2%) as five providers enforced TLS causing undeliverability (see also Section 3.1). For large providers, the baseline is met by all providers. For spammers, the baseline is necessary to estimate the number of SPAM emails that are typically sent to the investigated domains. For all three populations, the delivery to *Mail-v6-Baseline* is reduced compared to the IPv4 baseline, implying limited deployment of IPv6 at email servers. Differences among regular and large providers remain small – the first received IPv6-only mails in 193/436 (44.3%) of the cases, the latter in 5/13 (38.5%) –, however, SPAM towards the IPv6 target is drastically reduced and accounts for 7.4% of the IPv4 baseline.

DNS Resolvers. Both targets, *Mail-v4-DNS-v6* and *Mail-v6-DNS-v6*, rely on an IPv6-only authoritative nameserver and allow to infer whether resolvers are capable of IPv6. The number of successfully delivered emails to *Mail-v4-DNS-v6* is consistently higher than for IPv6-only email servers (*Mail-v6-Baseline*) – 285/436 (65.4%) vs. 193/436 (44.3%) (regular providers), 8/13 (61.5%) vs. 5/13 (38.5%) (large providers), and 46.5% vs. 7.4% (spammers) – and lead to the conclusion that IPv6 support is more prevalent among DNS resolvers than among email servers. The difference is particularly remarkable for SPAM, and suggests that spammers rely on external DNS resolvers. In comparison to *Mail-v6-Baseline*, delivery towards *Mail-v6-DNS-v6* is, if at all, only slightly reduced – 193/436 (44.3%) vs. 170/436 (39%) (regular providers), and 5/13 (38.5%) vs. 5/13 (38.5%) (large providers) –, i.e., IPv6 support at the email server typically implies IPv6 support at the respective DNS resolver. For the regular providers, Table 6 shows interdependencies concerning IP support: Most dominant are dual stack implementations 171/436 (39.2%) resp. IPv4-only configurations for email and DNS 125/436 (28.7%), as well as IPv4-only email servers with dual stack DNS resolvers 116/436 (26.6%).

Table 6: DNS and email server IP support levels (IPv4 only, IPv6 only or dual stack) of regular providers; reads f.e. 22 (5.0%) have dual stack email servers, but IPv4-only DNS resolver.

		Email					
		IPv4		**IPv6**		**Dual**	
DNS	**IPv4**	125	28.7%	1	0.2%	22	5.0%
	IPv6	0	0.0%	0	0.0%	1	0.2%
	Dual	116	26.6%	0	0.0%	171	39.2%

Key Findings. In summary, we find that less than half of all regular email providers support IPv6 for their email deployments. Interestingly, IPv6 support for DNS is more frequent, even for providers that do not support IPv6 for their email servers. We conjecture that this is connected to – especially in smaller setups – using public resolvers like the commonly known Cloudflare (1.1.1.1) or Google (8.8.8.8) instances. Interestingly, we also find that 23/436 (5.3%) of the observed

providers *do* use IPv6 for their email setup while *not* using it for their DNS resolvers. Even though finding this case is not unsurprising – PowerDNS, for example, does not perform IPv6 resolution by default—it still means that these operators are not able to deliver emails to IPv6-only zones, even though their email servers support IPv6.

5.2 TLS Configuration

TLS Enforcement. If our target *Mail-Dual-TLS-force* enforces the use of TLS, 392/436 (89.9%) of the regular and all large providers behave accordingly. These numbers indicate a high prevalence of TLS capability among email servers. Concerning SPAM, TLS enforcement has a considerable effect and reduces the number of emails to 34.0%.

TLS Validation. In the presence of invalid certificates, as provided by *Mail-Dual-TLS-invalid*, a similar picture emerges for regular and large providers. As common practice suggests [13] providers regularly fall back on opportunistic STARTTLS. Just one of the regular providers is more strictly configured and rejects email delivery in the case of a certificate with a non-matching CD/DNS0 entry. TLSA mismatch as caused by *Mail-Dual-TLSA-invalid* should technically prevent opportunistic encryption from being used. However, we find that only 138/436 (31.7%) of regular providers and 3/13 (23.1%) of large providers honor the TLSA record and refuse delivery. When we turn our eye to SPAM delivery, we find that enforcing TLS has a significant impact on the number of received emails. On our two TLS-enforcing targets, only 27.5% (*Mail-Dual-TLS-Force*) and 31.8% (*Mail-Dual-TLSA-Invalid*) of the baseline values of emails are received.

Key Findings. The broad majority of providers support TLS. However, emails from 10.1% of regular providers in our dataset would be lost in case of enforcing it. Providers fulfilling TLS enforcement typically also fall back on opportunistic encryption in case of invalid certificates. TLSA – a method to move beyond opportunistic encryption, even in the absence of CA-signed certificates – is sadly ignored by the majority of providers. At the same time, TLS enforcement does not only increase security, but it also reduces SPAM by more than 65%. While spammers could implement TLS quickly, it still would force them to adopt more costly TLS handshakes.

5.3 DNSSEC Validation

Targets *Mail-v4-DNSSEC-broken* and *Mail-v6-DNSSEC-broken* allow to infer the prevalence of resolvers validating DNS records. For regular providers, 181/436 (41.5%) delivered emails to our first target. The remaining 255/436 (58.5%) of all providers conducted a thorough validation for DNSSEC. Among the large providers, DNSSEC validation appears less prevalent: Only 4/13 (30.8%) (IPv4) and 2/5 (40.0%) (IPv6) of providers validate DNSSEC. We suspect that operators refrain from deploying DNSSEC to avoid customers missing emails or being unable to send emails due to misconfigurations. Furthermore, we observed a significant SPAM reduction for domains with broken DNSSEC. We conjecture that this is due to common open resolvers that validate DNSSEC being regularly used by spammers. This suspicion was confirmed when we revisited our DNS servers' logs to identify the most commonly used DNS resolvers. Query logs are, however, not fully available as log rotations removed some logs due to high response numbers. Still this enabled us to identify the most commonly used DNS resolvers. We were able to match resolvers for 2839/4660 (61%) regular emails and for 3399/6772 (50.2%) of emails sent by spammers. We found 1,443 unique resolver IPs for regular providers and 1,774 for spammers. Relying on MaxMind's public GeoLite AS database, we lookuped AS information for each IP. This resulted in 259 unique ASes used for DNS resolution for regular providers and 269 for spammers. Comparing the DNS servers used by regular and large providers with those used by spammers revealed an overlap of 138 IPs and 62 ASes.

Key Findings. DNSSEC validation is performed in 255/436 (58.5%) (IPv4) and 143/193 (74.0%) (IPv6) and regular providers. The numbers for large providers are lower, i.e., 4/13 (30.8%) (IPv4) and 2/5 (40.0%) (IPv6). In comparison, previous work [8] found DNSSEC to be less common; however, those measurements focused on zones using DNSSEC. The numbers for DNSSEC validation among spammers are – surprisingly – comparable to those of large providers. However, this connects to spammers regularly using public resolvers that already validate DNSSEC.

5.4 Anti-SPAM (Greylisting)

The greylisting targets *Mail-v4-Greylisting* and *Mail-v6-Greylisting* provoked an error in delivery the first time and accepted the email in a second – delayed – attempt. Legitimate providers reattempt to deliver emails in case of a failure, and our measurements indeed show that this is the case. Only 4/436 (0.9%) (IPv4) and 1/193 (0.5%) (IPv6) of the regular providers refrain from retransmission, and no large provider does so. However, greylisting reduces the number of received SPAM emails by 36.9%. Interestingly, this makes greylisting a less effective anti-SPAM measure than enforcing TLS.

Key Findings. Greylisting reduces the SPAM volume by 36.9% and does not introduce delivery problems for legitimate email. However, greylisting has less impact than TLS enforcement, which reduces SPAM by over 65%.

6 Related Work

In the past years, email has been receiving significant attention from the research community. In this section, we systematize eleven email-related measurement studies from 2014 onward.

Table 7: Measured adoption rates by related work. Percentages are collected for domains with MX records. SPF, DKIM and DMARC are included for comparison only as they merely influence the receiver's decision to accept incoming emails.

Citation	Year	Active Meas.	Domains	Sample Size	SPF	DKIM	DMARC	DNSSEC	DANE	TLS (inc.)
Adkins et al. [2]	2014		Facebook	/	-	-	-	-	-	76%
Foster et al. [17]	2015		Alexa	1M	42.3%	-	1%	3.4%	-	-
Foster et al. [17]	2015		Adobe	1M	43.6%	-	0.9%	2.8%	-	54%
Durumeric et al. [14]	2015	•	Gmail	/		-	-	-	-	80%
Durumeric et al. [14]	2015		Alexa	1M	47%	-	1.1%	-	-	81.8%
Hu et al. [22]	2018		Alexa	1M	44.9%	-	5.1%	-	-	-
SIDN [44]	2019		.nl	5.9M	44.2%	18.6%	8%	53%	-	62%
Kambourakis et al. [26]	2019/20	•	Custom	3236	80.7%	59.4%	51.3%	23.2%	17.6%	97.6%
Lee et al. [31]	2020		Alexa	100K	-	-	-	-	0.5%	-
Tatang et al. [45]	2021		*x*	2.04M	50%	13%	11%	-	-	-
Yajima et al. [34]	2021		Tranco	10K	88.7%	-	54.1%	7.7%	0.8%	-
Our work	2020/21	•	Custom	417	91.3%	63%	53.5%	57.4%∗	21.6%∗	89.9%

∗: We can only verify the percentage of DNSSEC resolvers and TLSA validating email servers.
•: Studies with active measurements
x: Mix of Alexa top 1M, Tranco, Majestics

We find that these studies use different sample sets and measurement methodologies. Sample sets range from top 1M domain lists to email collections with sample sizes from a million domains to a few thousand. However, using different methodologies, they all ultimately report comparable adoption rates of security-related email protocols, including SPF, DKIM, and DMARC. Hence, we compare their adoption rates and findings to our results in Table 7 to validate our methodology and provide a comprehensive picture of current providers' email delivery capabilities. Related work on email delivery so far primarily focused on large providers and did not consider the transport perspective – especially IPv6 and DNS – highlighting the gap our work fills.

Adoption Rates. Looking at the reported adoption rates from related work, we do find an upward trend in adoption, especially for security-related standards. We can also observe the difference in adoption rates per region. For example, `.nl` sees a 53% adoption rate of DNSSEC, which is significantly higher than the, e.g., 7.67% adoption rate for DNSSEC for Tranco Top 10K domains reported by Yajima et al. [34]. We attribute this high adoption rate to the Registrar Scorecard, a campaign incentivizing the deployment of standards by the Dutch domain name registrar SIDN, responsible for the .nl top-level domain [44]. In contrast to the number of DNSSEC-enabled zones, we find the number of validating resolvers to be considerably higher. We find a 57.35% of participants in our study rely on DNSSEC-validating resolvers, mostly due to common public resolvers, for example, the popular 8.8.8.8 resolver offered by Google.

Large Providers. Related work uses several methods for identifying and ranking large email providers (see Table 8): Durumeric et al. [14], Hu et al. [22], and Tatang et al. [45] used manual rankings by relying on their own expertise. However, this might induce bias towards the researcher's experience and location. Foster et al. [17], and Lee et al. [31] relied on email address domains from the leak of Adobe user records in 2013 [28] to rank email providers. However, this approach is limited to a one-time data dump and in completeness as it cannot detect different domains pointing their `MX` records at the same provider. Liu et al. [32] proposed a more comprehensive approach to detect and rank email providers in 2021. One of their major components is certificate information gathered through Internet-wide SMTP handshakes. In contrast, we introduce a new ranking method based on already existing passive DNS data from DNSDB (see Section 3). Based on this ranking we list the top 15 providers in Table 4. Our method thereby overlaps highly with the results of Liu et al., while introducing significantly less measurement overhead and revealing additional providers.

Table 8: Overview of large provider sets used in related work.

Year	Rel. W.	Overlap	Size	Method
2015	Durumeric et al. [14]	6	19	Manually
2015	Foster et al. [17]	3	22	Adobe leak
2018	Hu et al. [22]	1	35	Manually
2020	Lee et al. [31]	2	29	Adobe leak
2021	Tatang et al. [45]	2	25	Manually
2021	Liu et al. [32]	11	15	Custom
2021	Our work		15	passive DNS

Sender-side Evaluation. We only found two related measurement studies relevant to the sender-side aspects of email delivery [8, 31]. Chung et al. [8] performed a study focusing on DNSSEC adoption independent of email delivery setups in 2017. They set up ten differently misconfigured target domains (missing, incorrect, expired RRSIGS; missing DNSKEYs; incorrect DS; etc.), collecting data from 4,427 DNSSEC capable resolvers (DO bit set) from the Luminati proxy service. They found that 3,635 (81.1%) failed to validate DNSSEC responses. Only 543 (12.2%) resolvers did handle all ten different scenarios correctly. As we did not focus on DNSSEC validation specifically, but only wanted to test if validation was attempted, we relied on a single DNSSEC

setup for our measurement. Similar to us, Lee et al. [31] used 14 target domains to measure DNSSEC, STARTTLS, and DANE validation in 2020. However, they only measured the top 29 providers ranked by email addresses in the Adobe leak. The measurement setup is similar to ours, but contrary to Lee et al., we actively engaged participants to send emails to our target domains. Hence, we were able to cover a wider range of providers. Our set of large providers also differs from Lee et. al. as we used a more comprehensive ranking method, similar to that of Liu et al. [32]. Other studies evaluate email-related protocols from the receiver's perspective [2,14,17,26], i.e., evaluating emails once they are successfully delivered. For example, studying DNS TXT records between 2015 and 2018, van der Toorn et al. [46] observed a rise in the adoption of email security standards, such as SPF and DKIM, and attributed this to stricter policies from large email providers. However, this line of work generally finds similar problems on the receiver side as we observed on the sender side, e.g., the high complexity of standards, generally low adoption, and therefore, low validation rates. Durumeric [14] found that SPF network ranges are usually configured overly broad, e.g., nearly 30% of domains allow IPv4 address ranges of more than a /16 to originate emails. Furthermore, SPF inclusions are not used carefully, and a multitude of domains trust the same handful of cloud providers. Hu et al. [22] found that 34 of 35 (97%) of popular email providers deliver forged emails to inboxes even if validation of either one or multiples of SPF/DKIM/DMARC failed. Tatang et al. [45] compiled a list of DKIM selectors and found that domains do not only commonly share the same selector, but also the same key.

Standard Complexity. In 2021, Yajima et al. [34] first discussed how standards' complexity influences their adoption rate. They measured DNS-based security mechanisms and found that setup difficulty influences the adoption rate. Their rating of setup difficulty awards points for the following configuration aspects: DNS record (1pt); DNS server configuration (2pt); email server configuration (2pt); web server configuration (2pt); required third party (3pt). DNSSEC and DANE score the highest with 6 points. While DANE is a relatively new standard introduced in 2012, DNSSEC was introduced in 1999 and still faces a relatively low adoption and validation rate. Potential causes include a (perceived) high risk of service disruptions due to misconfigurations – even in 2021, we still regularly see outages of top-level domains due to misconfigured DNSSEC [24] – and complexity in maintaining DNSSEC. Further investigating the complexity of DNSSEC key material handling, Chung et al. [8] found that a majority of domains roll keys too infrequently, use weak keys, or do not perform rollovers correctly.

7 Discussion

Successful system operation includes design, implementation, and maintenance. In a world of ubiquitous networking, systems like the email ecosystem cannot be redesigned from scratch, but have to be carefully adapted. This means that successful further development has to consider the impact of improvements on the existing ecosystem. Hence, our measurement provides a perspective on the current state of email.

Our measurements pinpoint an apparent gap between the email ecosystem as standardized by the IETF and its actual deployment. Recently introduced standards such as TLSA (validation) have not made it into practice. Thus, our results suggest that the development of new email standards has to be accompanied by strategies fostering their actual deployment.

7.1 Heavy-tail Email

A pattern that emerges in our measurements as well as in the work of, e.g., Liu et al. [32] is the heavy-tail nature of email: As Table 4 shows, a small portion of operators provide email services to the majority of users and domains on the Internet. Our investigation of related work also shows that studies often focus only on this top part of email providers. However, when we want to understand the email ecosystem, the major challenge is identifying and measuring the diverse tail of email providers and small self-hosted email instances. This becomes particularly challenging if – like in our measurements – user participation is necessary, and might lead to a situation where smaller providers are less investigated with potentially negative impact on their security, resilience, etc.

In a more techno-philosophical dimension, this development also raises concerns in the context of centralization. For example, in 2021 Fiebig et al. measured the migration of universities to large cloud providers, including their email infrastructures [15]. Centralization might accelerate the adoption of standards (e.g., if the relevant players are directly involved in standardization), but this can also potentially enforce the deployment of burdensome standards by small operators, effectively creating a walled garden. Beyond, failure of a single large provider, either due to an accidental error or a deliberate attack, affects a large share of users/domains, emphasizing the importance of decentralization and diversity for the resilience of the overall email ecosystem.

What we certainly highlight – if we want to keep a distributed Internet – is that future development efforts should not only focus on improving standards themselves, but also make it easier to follow these standards and enable operators to run their email infrastructure in full standard compliance. We encourage RFCs drafted by the IETF to be accompanied by *technical and organizational measures facilitating implementation*, reducing the gap between standardization and deployment.

7.2 Delivery vs. Adoption

Looking at the large provider dataset in our study, we find that currently especially large providers prioritize email delivery over security, e.g., DNSSEC validation is enabled for

Google's public DNS service, but not for the resolvers Gmail relies on. This is understandable from an operational standpoint but suggests that security is still considered subordinate to functional goals. We conjecture that Google prioritizes the deliverability of emails over strict enforcement of DNSSEC. The status-quo appears to represent an upside-down world: Precisely for large providers, the deployment of a new security feature appears manageable; yet, they refrain from doing so in a strict manner. At the same time, small operators implement the respective features at a disproportionate operational overhead.

This divergence of the email ecosystem ultimately creates challenges, as new security features often do address actual problems. Hence, the operations community must discuss how this divide can be addressed in the future. The Registrar Scorecard has already proven that financial incentives are successful [44]. Thus, we suggest including the design of such systems already during standardization. The Internet Governance Forum also recommends financial incentives by translation of standards into business cases [47]. However, this poses various challenges, among others the collaboration of multiple stakeholders, funding, and the operation of respective evaluation systems, which have to be solved by future work.

7.3 Standard Deployment and SPAM

In our study, we find that TLS enforcement and IPv6-only delivery have a significant impact on the amount of SPAM systems receive. While IPv6-only delivery naturally has a significant negative impact on legitimate emails being delivered, this impact is smaller when enforcing TLS. According to our measurements, emails from about 10% of regular providers would be affected. However, it is hard to determine an adoption threshold for which enforcement of standards is justified. On the one hand, TLS is an old and well-understood standard, fully supported by large providers which represent the driving force in standard deployment; also the implementation effort is low compared to other standards like DNSSEC or DANE. On the other hand, it is unclear why 10.1% of these providers have not implemented (START)TLS. If this is the case because delivery is still possible without, enforcement of TLS should take place; if the reasons are rooted in structural aspects (e.g., lacking support for certain types of systems or adequately educated staff), we suggest to target these root causes first, again requiring additional technical and organizational measures accompanying RFCs.

8 Conclusion

We investigated email delivery, especially in terms of protocol use (IPv4 vs. IPv6, recursive DNS servers' configuration, TLS sending support) and thereby complement existing related work, which mostly investigated the receiving side of the email ecosystem. Together with a review of related work on email delivery, this allows us to paint a comprehensive picture of the complexity of email delivery in 2021.

We find that 'new' protocols and extensions relevant to email delivery, like IPv6 and DNSSEC, lack adoption. The overall ecosystem is slow in this regard, especially since large email providers prioritize email delivery and – while trying to offer as many options as possible to receive emails – take a conservative stance when trying to deliver emails to others. This highlights the importance of including the heavy-tail of smaller providers in email-related measurements. Our results show that standard deployment is lower than it could be. At the same time, we know that financial incentives work well to increase deployment rates. Hence, we suggest that such incentive systems should accompany Internet standards. However, continuous funding appears to be difficult; thus, future work should also address the impact of non-financial incentives.

Acknowledgements

This material is based upon work partially supported by (1) the Christian-Doppler-Laboratory for Security and Quality Improvement in the Production System Lifecycle; the financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged; (2) SBA Research (SBA-K1), a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the province of Vienna. The COMET Programme is managed by FFG; (3) Project 877110 2big2fail funded by the Program "BRIDGE 1" (FFG); (4) Project FO999887504 DynAISEC funded by the Program "ICT of the Future"—an initiative of the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology; (5) the European Commission through the H2020 project CyberSecurity4Europe (Grant No. #830929); and (6) by the Vienna Science and Technology Fund (WWTF) through project ICT19-056.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of their host institutions or those of the European Commission.

References

[1] Donald E. Eastlake 3rd. Domain Name System Security Extensions. RFC 2535, RFC Editor, March 1999. http://www.rfc-editor.org/rfc/rfc2535.txt.

[2] M. Adkins. The Current State of SMTP STARTTLS Deployment, 2014. Retrieved Sept. 16, 2021 from https://www.facebook.com/notes/1453015901605223.

[3] Eric Allman. sendmail 8.9.0 released. Retrieved Sept. 20, 2021 from https://www.sendmail.org/~ca/email/releases/sm890announce.html.

[4] Kurt Andersen, Brandon Long, Seth Blank, and Murray Kucherawy. The Authenticated Received Chain (ARC) Protocol. RFC 8617, July 2019. https://www.rfc-editor.org/info/rfc8617.

[5] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. DNS Security Introduction and Requirements. RFC 4033, RFC Editor, March 2005. http://www.rfc-editor.org/rfc/rfc4033.txt.

[6] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The Menlo Report. *IEEE Security & Privacy*, 10(2):71–75, 2012.

[7] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer. OpenPGP Message Format. RFC 4880, RFC Editor, November 2007. http://www.rfc-editor.org/rfc/rfc4880.txt.

[8] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, and Christo Wilson. A longitudinal, end-to-end view of the DNSSEC ecosystem. In *Proceedings of the USENIX Security Symposium (USENIX Security 17)*, 2017.

[9] D. Crocker, T. Hansen, and M. Kucherawy. DomainKeys Identified Mail (DKIM) Signatures. STD 76, RFC Editor, September 2011. http://www.rfc-editor.org/rfc/rfc6376.txt.

[10] Dr. Steve E. Deering and Bob Hinden. Internet Protocol, Version 6 (IPv6) Specification. RFC 8200, July 2017. https://rfc-editor.org/rfc/rfc8200.txt.

[11] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. Investigating system operators' perspective on security misconfigurations. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2018.

[12] David Dittrich and Erin Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, 2012. https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.

[13] Viktor Dukhovni. Opportunistic Security: Some Protection Most of the Time. RFC 7435, December 2014. https://rfc-editor.org/rfc/rfc7435.txt.

[14] Zakir Durumeric, David Adrian, Ariana Mirian, James Kasten, Elie Bursztein, Nicolas Lidzborski, Kurt Thomas, Vijay Eranti, Michael Bailey, and J Alex Halderman. Neither snow nor rain nor MITM... an empirical analysis of email delivery security. In *Proceedings of the Internet Measurement Conference (IMC)*, 2015.

[15] Tobias Fiebig, Seda Gürses, Carlos H Gañán, Erna Kotkamp, Fernando Kuipers, Martina Lindorfer, Menghua Prisse, and Taritha Sari. Heads in the clouds: Measuring the implications of universities migrating to public clouds. *arXiv preprint arXiv:2104.09462*, 2021.

[16] Tobias Fiebig, Franziska Lichtblau, Florian Streibelt, Thorben Krüger, Pieter Lexis, Randy Bush, and Anja Feldmann. Learning from the past: designing secure network protocols. In *Cybersecurity Best Practices*. Springer, 2018.

[17] Ian D Foster, Jon Larson, Max Masich, Alex C Snoeren, Stefan Savage, and Kirill Levchenko. Security by any other name: On the effectiveness of provider based email security. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2015.

[18] P. Hoffman. SMTP Service Extension for Secure SMTP over TLS. RFC 2487, RFC Editor, January 1999. http://www.rfc-editor.org/rfc/rfc2487.txt.

[19] P. Hoffman. SMTP Service Extension for Secure SMTP over Transport Layer Security. RFC 3207, RFC Editor, February 2002. http://www.rfc-editor.org/rfc/rfc3207.txt.

[20] P. Hoffman and J. Schlyter. The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. RFC 6698, RFC Editor, August 2012. http://www.rfc-editor.org/rfc/rfc6698.txt.

[21] R. Housley, D. Crocker, and E. Burger. Reducing the Standards Track to Two Maturity Levels. BCP 9, RFC Editor, October 2011. http://www.rfc-editor.org/rfc/rfc6410.txt.

[22] Hang Hu and Gang Wang. End-to-end measurements of email spoofing attacks. In *Proceedings of the USENIX Security Symposium (USENIX Security 18)*, 2018.

[23] Bert Hubert. DNS-Camel, 2018. Retrieved Jan. 13, 2022 from https://blog.apnic.net/2018/03/29/the-dns-camel/.

[24] IANIX. Major DNSSEC Outages and Validation Failures, November 2021. Retrieved Nov. 16, 2021 from https://ianix.com/pub/dnssec-outages.html.

[25] Siyuan Jia, Matthew Luckie, Bradley Huffaker, Ahmed Elmokashfi, Emile Aben, Kimberly Claffy, and Amogh Dhamdhere. Tracking the deployment of IPv6: Topology, routing and performance. *Computer Networks*, 165:106947, 2019.

[26] G. Kambourakis, G. Draper, and I. Sanchez. What Email Servers Can Tell to Johnny: An Empirical Study of Provider-to-Provider Email Security. *IEEE Access*, 8:130066–130081, 2020.

[27] S. Kitterman. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208, RFC Editor, April 2014. http://www.rfc-editor.org/rfc/rfc7208.txt.

[28] Brian Krebs. Adobe To Announce Source Code, Customer Data Breach, October 2013. Retrieved Jun. 6, 2022 from https://krebsonsecurity.com/2013/10/adobe-to-announce-source-code-customer-data-breach/.

[29] M. Kucherawy and D. Crocker. Email Greylisting: An Applicability Statement for SMTP. RFC 6647, RFC Editor, June 2012. http://www.rfc-editor.org/rfc/rfc6647.txt.

[30] M. Kucherawy and E. Zwicky. Domain-based Message Authentication, Reporting, and Conformance (DMARC). RFC 7489, RFC Editor, March 2015. http://www.rfc-editor.org/rfc/rfc7489.txt.

[31] Hyeonmin Lee, Aniketh Girish, Roland van Rijswijk-Deij, Taekyoung Kwon, and Taejoong Chung. A Longitudinal and Comprehensive Study of the DANE Ecosystem in Email. In *Proceedings of the USENIX Security Symposium (USENIX Security 20)*, 2020.

[32] Enze Liu, Gautam Akiwate, Mattijs Jonker, Ariana Mirian, Stefan Savage, and Geoffrey M Voelker. Who's Got Your Mail? Characterizing Mail Service Provider Usage. In *Proceedings of the ACM Internet Measurement Conference*, 2021.

[33] D. Margolis, M. Risher, B. Ramakrishnan, A. Brotman, and J. Jones. SMTP MTA Strict Transport Security (MTA-STS). RFC 8461, RFC Editor, September 2018. http://www.rfc-editor.org/rfc/rfc8461.txt.

[34] Yoshiro Yoneya Masanori Yajima, Daiki Chiba and Tatsuya Mori. How prevalent is the operation of DNS security mechanisms? Retrieved Sept. 15, 2021 from https://indico.dns-oarc.net/event/39/contributions/867/.

[35] Mozilla. Public Suffix List, 2021. Retrieved Nov. 24, 2021 from https://publicsuffix.org/list/public_suffix_list.dat.

[36] OpenBSD. OpenBSD 6.7. Retrieved Oct.12, 2021 from https://www.openbsd.org/67.html.

[37] Damian Poddebniak, Fabian Ising, Hanno Böck, and Sebastian Schinzel. Why TLS is better without STARTTLS: A Security Analysis of STARTTLS in the Email Context. In *Proceedings of the USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2021.

[38] Jonathan B. Postel. Internet Protocol. RFC 791, September 1981. https://www.rfc-editor.org/info/rfc791.

[39] Jonathan B. Postel. Simple Mail Transfer Protocol. STD 10, RFC Editor, August 1982. http://www.rfc-editor.org/rfc/rfc821.txt.

[40] Postfix. Postfix stable release 3.6.0. Retrieved Oct. 12, 2021 from http://www.postfix.org/announcements/postfix-3.6.0.html.

[41] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. A primer on IPv4 scarcity. *ACM SIGCOMM Computer Communication Review*, 45(2):21–31, 2015.

[42] Marco Schmidt. Expired Domains, 2021. Retrieved March 15. 2021 from https://www.expireddomains.net/.

[43] Farsight Security. Passive DNS historical internet database: Farsight DNSDB, 2021. Retrieved Nov. 24, 2021 from https://www.farsightsecurity.com/solutions/dnsdb/.

[44] SIDN. Registrar Scorecard yields great results. Retrieved Sept. 16, 2021 from https://www.sidn.nl/en/news-and-blogs/registrar-scorecard-yields-great-results.

[45] Dennis Tatang, Florian Zettl, and Thorsten Holz. The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws. In *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses*, 2021.

[46] Olivier van der Toorn, Roland van Rijswijk-Deij, Tobias Fiebig, Martina Lindorfer, and Anna Sperotto. TXTing 101: Finding Security Issues in the Long Tail of DNS TXT Records. In *Proceedings of the International Workshop on Traffic Measurements for Cybersecurity (WTMC)*, 2020.

[47] De Natris Consult Wout de Natris. Setting the Standard for a more Secure and Trustworthy Internet, 2020. Retrieved from https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/9615/2023.

3 Measuring Active IPv6 Networks

This chapter is covered by the second publication, which examines implementations of ICMPv6 error messages.

Title	Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources
Authors	Florian Holzbauer, Markus Maier and Johanna Ullrich
Publication Status	This paper is included in the Proceedings of the 2024 ACM on Internet Measurement Conference, Pages 280–294. CORE2023-Ranking: A Acceptance Rate (Long Paper): 18.99%
DOI	`https://doi.org/10.1145/3646547.3688420`
Author Contributions	Florian Holzbauer: Is the first and main author of this publication. He is responsible for the measurements design, execution, evaluation, and paper writing. Markus Maier: Design of the router lab and contribution of data to chapter 4.1 by implementing the GNS3 router lab and conducting experiments in the router lab to measure ICMP error message defaults. Providing measurement data for Chapter 4.3 from an IPv6-wide tracerouting campaign. Johanna Ullrich: Supervision and paper writing.
Artifacts	Jupyter & Code: `https://github.com/sbaresearch/icmpv6-destination-reachable`
Reference	[HMU24]

Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources

Florian Holzbauer
Faculty of Computer Science
Doctoral School Computer Science
University of Vienna
Vienna, Austria
florian.holzbauer@univie.ac.at

Markus Maier
SBA Research
Vienna, Austria
mmaier@sba-research.org

Johanna Ullrich
University of Vienna
Vienna, Austria
johanna.ullrich@univie.ac.at

Abstract

The probability of hitting an active IPv6 address by chance is virtually zero; instead, it appears more promising to analyze ICMPv6 error messages that are returned in case of an undeliverable packet. In this paper, we investigate the implementation of ICMPv6 error messages by different router vendors, whether a remote network's deployment status might be inferred from them, and analyze ICMPv6 error messaging behavior of routers in the IPv6 Internet. We find that Address Unreachable with a delay of more than a second indicates active networks, whereas Time Exceeded, Reject Route and Address Unreachable with short delays pinpoint inactive networks. Furthermore, we found that ICMPv6 rate-limiting implementations, used to protect routers, allow the fingerprinting of vendors and OS-versions. This enabled us to detect more than a million periphery routers relying on Linux kernels from 2018 (or before); these kernels have reached end of life (EOL) and no longer receive security updates.

CCS Concepts

• **Networks** → **Network protocols**; **Routers**; **Network measurement**; **Public Internet**.

Keywords

IPv6, ICMPv6 Error Messages, Network Activity Classification, Router Classification, Rate Limiting

ACM Reference Format:
Florian Holzbauer, Markus Maier, and Johanna Ullrich. 2024. Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3646547.3688420

1 Introduction

An exhaustive scan of the IPv4 Internet takes less than an hour [1, 11], which remains infeasible with the successor protocol IPv6 due to the sheer size of the address space. Alternative approaches are needed. In fact, the probability of hitting an alive IPv6 host by chance is virtually zero [12, 18, 32] and it appears more promising to analyze the numerous ICMPv6 error messages that are returned in case of undeliverability as they provide insight into remote networks.

ICMP error messages have been collected before, both for IPv4 and IPv6. Best known are topology discovery [6, 9, 16], also known as tracerouting, and routing loop detection [22, 23]. Also, ICMPv6 messages have been intentionally triggered to extract source addresses, thus collecting millions of IPv6 addresses of periphery devices [22, 33].

We take a different stance and analyze ICMPv6 error messages beyond their source addresses. The main goal of this paper is to (I) examine the different error message types that are returned by active and inactive networks on the Internet and (II) classify routers within these networks. Based on the type, code, and timing of an error message, we infer routing scenarios other than routing loops such as active networks, ACL filtering, or null routes. Our analysis considers aspects such as the responsiveness of routers, given that most ICMPv6 error message types are sent voluntarily, the compliance of routers with the ICMPv6 specification [10], and whether variances in type usage and rate limiting implementations allow to classify router and OS versions.

We follow a threefold methodology for both aspects, namely (I) *network activity classification* and (II) *router classification*. First, we observe ICMPv6 error messaging behavior of eleven router vendors in the network simulator GNS3, providing full control over the setup. Second, we use labeled datasets [2, 14] to verify whether the behavior observed in the virtual setup is congruent with that of actual routers in the IPv6 Internet. Third, we conduct Internet measurements to gain insight into the current state of ICMPv6 error message implementations across routed networks and to enumerate ICMPv6 error message handling within a more diverse set of networks. Our research makes several key contributions, detailed in the following sections. The code for our measurements is publicly available [1].

Network Activity Classification (§4.1, §4.2) We associate ICMPv6 error message types with the activity status of a remote network. The receipt of message type *Address Unreachable* with long delays is found to indicate active networks with a probability of 95.1%, while *Time Exceeded*, *Reject Route* and *Address Unreachable* with short delays indicate inactive networks with a probability of 79.5%.

BValue Steps (§4.2) To validate our network activity detection, we develop the BValue step method to create datasets of addresses in

[1] https://github.com/sbaresearch/icmpv6-destination-reachable

IMC '24, November 4–6, 2024, Madrid, Spain

ACM ISBN 979-8-4007-0592-2/24/11
https://doi.org/10.1145/3646547.3688420

active and inactive networks from the IPv6 Hitlist Service [15]. This method is also useful to investigate the error message responding behavior of individual networks in case an active address is known.

Network Activity Scans(§4.3) We collect and classify ICMPv6 error messages across a wide portion of routed IPv6 Internet. In two measurements, we discovered 83M (of a total of 5Bn) /48s and 356M (of a total of 6Bn) /64s to be active. Our methodology is useful in order to guide host discovery toward more promising parts of the Internet.

Router Classification (§5.1, §5.2) In our GNS3 environment, we identify different ICMPv6 rate limiting behavior of routers and exploit it to remotely classify them (vendor/operating system). Relying on SNMPv3-vendor labels for ground truth [2], we were able to verify our approach on the Internet and to extend it with additional fingerprints. Our method fills a gap as previous work based on varying iTTL values [3, 36] is not applicable since the Hop Limit in IPv6 has been increasingly harmonized [8].

Linux-Based Routers at the End of Life (§5.3) In a large-scale measurement study, we classified 1.4M routers applying our method on the Internet and found 1M periphery routers relying on Linux kernels from 2018 or before. These kernel versions have reached end of life at latest by January 2023 and pose a potential security risk.

2 Terminology

Routed, Active, and Inactive Networks If a network prefix is available in routing tables, packets towards its addresses can be forwarded. We consider these addresses to be *routed*. The mere routability is not sufficient for successful delivery. On the receiver's side, it also requires a last-hop router attached to the local network that forwards the packets to their final destinations. Therefore, the router conducts Neighbor Discovery [25] to resolve IP addresses into link-layer addresses. In this work, we denote such networks as *active* networks. If a last hop router is not prevalent or if it is discarding traffic towards the destination network, we refer to these networks as *inactive*. The distinction between *active* and *inactive* prefixes facilitates reconnaissance since responsive IPs can only exist in active networks.

Assigned, Unassigned and Responsive Addresses In an active network, only a fraction of its addresses are assigned to individual hosts and might be used to communicate with others. We refer to these addresses as *assigned* addresses. If such an address is returning packets (e.g., *ICMPv6 Echo Replies*) upon request (e.g., *ICMPv6 Echo Requests*), it is considered to be a *responsive* address. An address that is not assigned to any host – and thus cannot be used in communication – is referred to as an *unassigned* address.

ICMPv6 Error Messages RFC4443 [10] defines two informational message types – (*Echo Request* and *Echo Reply*), used for diagnosis (ping) – and four error message types. The message types and their subcodes are listed in Table 1. For readability, we use two-letter abbreviations instead of the messages' full name. If no response is received, we use the symbol ∅. The RFC defines processing of ICMPv6 messages as follows: (1) Only *TB* and *TX* messages are mandatory; the others are optional. (2) ICMPv6 error messages include the packet triggering the error as a payload. This allows the extraction of the initial request's destination. (3) Neighbor Discovery [25] uses the ICMPv6 message format. A router sends a *Neighbor Solicitation* to resolve an IPv6 into a link-layer address. Per address to be resolved, the sending of only one such message per second is allowed. If unresolved after three attempts, the router should return *AU*. (4) Rate limiting of ICMPv6 messages is mandatory, and a token bucket algorithm is proposed. For each message sent, a token is removed. If the bucket is empty, messages are discarded until a refill.

ICMPv6 Types and Codes	Abbr.
Destination Unreachable	
No route to destination	*NR*
Admin. prohibited	*AP*
Beyond scope of source address	*BS*
Address unreachable	*AU*
Port unreachable	*PU*
Ingress/egress policy	*FP*
Reject route to destination	*RR*
Time Exceeded	*TX*
Packet Too Big	*TB*
Parameter Problem	*PP*
Echo Request	*EQ*
Echo Reply	*ER*
Unresponsive	∅

Table 1: ICMPv6 error message types from RFC4443 and abbreviations used in the paper.

3 Methods Overview

In this paper, we rely on a three-step methodology for both of our goals, (I) the classification of a remote network's activity status based on the received ICMPv6 error message types in Section 4, and (II) the router classification based on ICMPv6 rate limiting behavior in Section 5. In particular, we

- (M1) investigate ICMPv6 error messaging behavior of eleven router vendors in a virtual GNS3 setup, facilitating full control of the router configurations[2].
- (M2) validate if routers on the IPv6 Internet behave in the same way as observed in our fully controlled laboratory environment. Our validations build upon labeled datasets [2, 14].
- (M3) perform measurements on the IPv6 Internet to show the extent of our findings and provide insights into the current state of the Internet's deployment.

Network Activity Classification We analyze ICMPv6 error message types and classify them to draw conclusions about a remote network's activity status. (M1) (§4.1) In our virtual GNS3 setup, we tested 15 routers and firewalls from 11 vendors in six different routing scenarios – such as forwarding packets to unassigned IP addresses, lacking routing table entries, or null routes – to see whether they show coherent behavior among each other as well as with regard to RFC4443 [10]. Based on the results, we associate ICMPv6 error message types with the activity status (active, inactive, ambiguous) of the remote network that has returned this message. (M2) (§4.2) The virtual setup is by definition limited in variety. Consequently, we performed a measurement to verify whether our observations are congruent with the diverse routers on the Internet. Applying our BValue steps method, we therefore inferred

[2] https://github.com/sbaresearch/router-lab

data sets of (unassigned) addresses in active and inactive networks from the IPv6 Hitlist Service [14, 34, 38]; then, we compared them against our assumptions on the correlation between ICMPv6 error message type and network activity status. (M3) (§4.3) Finally, we conducted two measurements. Relying on yarrp [6], the first probes all BGP-announced prefixes at the granularity of /48, resulting in 5 Billion traces. In this measurement, shorter prefixes (e.g. a /32) are resolved into multiple /48 prefixes. Prefixes less specific than /24 are prescanned for promising /24s by scanning 2 targets per included /32 and take those for which we receive a response. The second measurement used ZMap [11] to exhaustively probe the 92,856 /48 prefixes that are announced in BGP (as of November 2023) at /64 granularity. In the latter measurement, less-specific prefixes announced in BGP were ignored.

Router Classification We exploit ICMPv6 rate limiting behavior to identify a remote router's vendor and/or OS version. (M1) (§5.1) In our virtual setup, we measured the vendors' default settings for ICMPv6 rate limiting as a baseline for comparison with real-world behavior on the Internet. Many routers are based on Linux or BSD; thus, we additionally investigated different kernel versions to understand their default behavior. (M2) (§5.2) Previous work found that certain routers unintentionally reveal vendor and other information by responding to unsolicited and unauthenticated SNMPv3 requests [2]. We were able to measure 50K routers with SNMPv3 vendor labels available, allowing us to verify our results against ground truth and to additionally extend our database of fingerprints. (M3) (§5.3) After confirming the congruence of our router classification against ground truth, we measured and classified a total of 1.4M routers with regard to their vendors and/or operating systems. In addition, we were able to group them into Internet core- and periphery devices – depending on the number of paths they appeared in the previous measurement with yarrp – thus revealing different router populations.

Limitations *(I) Router Coverage.* Our testbed is, by definition, limited to vendors and versions available in GNS3. This also includes configuration options as some were restricted, e.g., ACLs for two RUTs or Null Routes for another ones. We marked these scenarios with (-) in Table 9 in Appendix B. We countered this limitation by extending our validation to routers on the IPv6 Internet. *(II) Validation in the Internet.* We did not have access to ground truth other than virtual appliances in our laboratory setup and labeled datasets from related work [2, 14]. We used these data sets to validate if our findings are representative by comparing if routers on the Internet behave similarly to those in the laboratory setup. *(III) Network Coverage.* We could not fully cover the routed IPv6 address space in our prefix-seeded measurements. For routed /48 networks, we can cover the subnet space up to /64. For networks larger than /48, we sampled their subnet space. In the second measurement, we prioritized coverage of many networks instead of going in-depth into single networks.

4 Network Activity Classification

Our measurements in the virtual laboratory (§4.1) reveal that the router implementations of ICMPv6 error messages deviate from the specifications outlined in RFC4443 [10]. These discrepancies led to a reassessment of our classification of error messages, but also allow to fingerprint a router's vendor. In our validation (§4.2), we found the same distinctive behavior for routers in the IPv6 Internet. We performed two measurements to find active networks on today's Internet (§4.3), reducing the search space for host discovery to 1.7% and 12%, respectively.

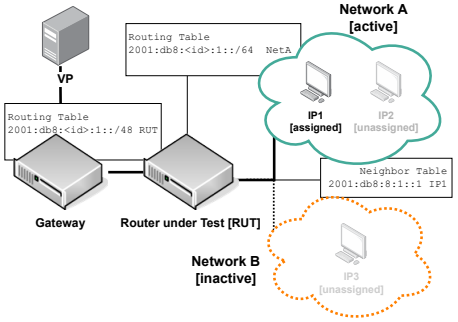


Figure 1: GNS3 laboratory setup. As common for IPv6 [22, 28], the RUT routes traffic to the active /64 network A, but not to inactive network B.

4.1 ICMPv6 Error Message Defaults

In a virtual setup, we analyzed the default ICMPv6 responding behavior of 15 routers and firewalls in six routing scenarios. Implementations show variance and deviations from RFC 4443 [10].

Router Laboratory In the network emulator GNS3, we set up a test network, see Figure 1. The gateway forwards traffic towards a /48 prefix to the router-under-test (RUT), but the RUT is only configured as a last-hop router for a /64 subnetwork (network *A*). According to our terminology, the /48 prefix is routed, but only network *A* is active. In network *A*, address *IP1* is assigned to an alive host and responsive, while *IP2*, belonging to the same network range, remains unassigned. In contrast, network *B* is inactive due to the RUT not being configured to handle traffic for network *B* and, thus, lacking a last-hop router conducting Neighbor Discovery. *IP3* represents an address within the inactive network's range *B*.

Routing Scenarios We configured six different routing scenarios, (S1) to (S6), which trigger ICMPv6 error messages based on the specification in RFC4443. We use 15 virtual images for the RUT to reveal the different routers' default ICMPv6 messaging behavior. If ICMPv6 error messages are not sent by default, we enable them for our experiments. We probe IP addresses using ICMPv6 Echo Requests, TCP SYNs, and UDP requests to verify protocol-specific response behavior. We list the expected response types based on the specification of RFC4443 for each scenario next to the scenario's name.

- (S1) **Active Network** - *AU*. Network *A* is directly configured on one of the RUT's interfaces. Requests towards IP2 reveal the ICMPv6 error message in case of an unassigned address in an active network.

	S1	Active Network	S2	Inactive Network	S3	Active Netw. ACL	S4	Inactive Netw. ACL	S5	Null Route	S6	Routing Loop
NR	○	0	●	14	●	1	●	2	●	2	○	0
AP	○	0	○	0	●	4	●	5	●	3	○	0
AU	●	14	○	0	○	0	○	0	●	1	○	0
PU	○	0	○	0	●	3	●	2	○	0	○	0
FP	○	0	●	1	●	1	●	2	○	0	○	0
RR	○	0	○	0	○	0	○	0	●	2	○	0
TX	○	0	○	0	○	0	○	0	○	0	●	15
∅	●	1	○	0	●	4	●	3	●	9	○	0

NOTE: Number = # of routers that return the error message type in a scenario; a single RUT can return multiple error message types if more than one configuration option is available.

Table 2: ICMPv6 error messages from 15 RUTs in 6 routing scenarios. The expected error message is indicated in gray. We list individual RUTs in Table 9 in Appendix B.

- (S2) **Inactive Network** - *NR*. The RUT receives a packet for which it has no entry in its routing table. Requests towards IP3 reveal the ICMPv6 error message in case of an inactive network.
- (S3) **Active Network with ACL** - *AP*,*FP*. We configure ACLs that either filter packets (I) towards network A (destination-based filtering) or (II) from our vantage point (source-based filtering). We probe IPs in network *A* to reveal the ICMPv6 error message for an active network with ACL.
- (S4) **Inactive Network with ACL** - *AP*,*FP*. An ACL for network *B* is configured to verify whether differences among active and inactive networks with ACLs are observable. Requests towards IP3 reveal the error message in case of an inactive network with ACL.
- (S5) **Null Routes** - *RR*. A null route is configured, discarding/rejecting all packets towards network B, and address IP3 is probed.
- (S6) **Routing Loops** - *TX*. The RUT maintains a default route towards the gateway. As network B is not routed, requests towards IP3 will be routed back via the incoming interface, forming a routing loop.

Results Table 2 provides an overview on the received ICMPv6 error messages per routing scenario. Focusing on *implementation coherency*, we found congruent behavior among the different routers for (S1), (S2), and (S6) – with single exceptions for (S1) and (S2). For the remaining scenarios, we see five different message types each due to vendor-specific filtering implementations. We also found *differences between scenarios* (S3) and (S4). For routers that rely on forward chain filters, the routing decision is made before the filter is applied. This results in three RUTs that are more likely to be used in the Internet edge, returning the same error message type as in (S2). Regarding *response timings*, there is a peculiarity for *AU*. For (S1), we notice delays of 2, 3, and 18 seconds as the messages are only returned after the Neighbor Discovery's timeout. We discovered that a delay of 2s is unique to Juniper, while 18s to Cisco XRv, allowing to fingerprint these vendors based on the delays. The other routers show delays of 3s as proposed in the RFC. For (S5), *AU* is returned immediately. Also, all other message types are returned immediately. Comparing the *request protocols*, we only find differences in the presence of ACLs. Two RUTs try to mimic protocol-specific responses from the target host for TCP and UDP.

Status	*NR*	*AP*	$AU_{>1s}$	$AU_{<1s}$	*PU*	*FP*	*RR*	*TX*
active	○	○	●	○	○	○	○	○
inactive	○	○	○	●	○	○	●	●
ambig.	●	●	○	○	●	●	○	○

Table 3: Classification of ICMPv6 error message types indicating activity/inactivity of a remote network.

Network Activity Classification Our goal is to differentiate ICMPv6 messages indicating active remote networks from those indicating an inactive ones. Based on the results in Table 2, we classify messages that have only been returned for active networks ((S1), (S3)) as *active*, those that have only been returned for inactive networks ((S2), (S4), (S5), (S6)) as *inactive*, and those appearing in both cases as *ambiguous*.

AU is consistently received for active networks ((S1)) but has also been returned by Juniper routers for (S5). Thus, we would have to consider *AU* to be ambiguous. Yet, the difference in timing – (S1) always causes a delay of multiple seconds that is longer than typical RTTs on the Internet – allows differentiation. For the remainder of the paper, we differentiate between $AU_{RTT>1s}$ indicating an active network, and $AU_{RTT<1s}$ indicating an inactive network. Table 3 summarizes our classification of ICMPv6 error message types.

Compliance with the RFC Testing router vendor implementations, we found behavior that deviates from the specification in RFC4443 [10]. This has an impact on the diagnostic value of the error message types, as one cannot simply rely on the RFC, but needs to know the vendor-specific behavior. This leads to a different network activity classification than if we had based our classification solely on the RFC.

The affected types and scenarios are *FP* for inactive networks ((S2)), *NR* and *PU* for filtering ((S3) and (S4)) and *NR*, *AP* and *AU* for null routing ((S5)). Based on the RFC, *PU* and *AU* should be returned for active networks. In addition, *PU* should only be returned by destination nodes only, i.e., assigned IPs. In our measurement, however, we found one of the firewalls using *PU* to mimic responses from the target host. Next to reporting a failure in Neighbor Discovery for unassigned addresses, we found one RUT implementing *AU* instead of *RR* for null routing. Following the specification, *NR* should be used for inactive networks due to a lacking entry in the routing table. However, we also found one RUT to return *NR* for active networks with ACL in (S3) and null routes in (S5).

4.2 ICMPv6 Error Messages in the Internet

The virtual setup, as used in the previous section, is limited to the availability of router images and does not fully reflect the variety of routers on the Internet. For validation of our network activity classification in Table 3, we need unassigned addresses from networks on the Internet, categorized as active and inactive. Probing these addresses helps us collect error messages specific to each category. Since no datasets of addresses exist for these categories, we developed a method called BValue Steps to separate addresses in active and inactive networks.

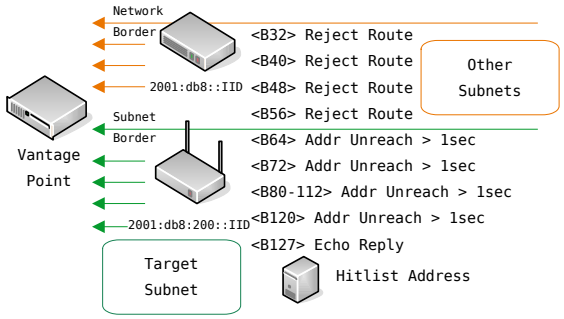


Figure 2: BValue Steps aim for a change in ICMPv6 error messages. Message types before the change represent active, those after the change inactive networks.

Data Set Generation Based on our terminology, a responsive IP address, as those present in hitlists, resides in an active network. To collect ICMPv6 error messages for unassigned addresses in the same active network, we derived addresses from the responsive address by randomizing their lower bits. With more and more randomized bits, we eventually reached the network border and probed addresses outside the active network. This way, we can collect ICMPv6 error message for other, likely inactive parts of the BGP-announced prefix. We measured addresses from hitlists this way, and included those with a change in received ICMPv6 error message types into our analysis. For these networks, we label the error message type *before* the change to represent addresses in active networks, and the one *after* the change to represent addresses in inactive networks, see Figure 2.

BValue Steps Assuming knowledge of an assigned IPv6 address and its respective routed network prefix length, we take the address and replace its lower bits – in multiples of eight bits – with random values. Figure 3 shows our approach with an example address. These addresses are referred to as BValue (Border value) addresses for bit 120, 112, 104, etc. (short B120, B112, B104, etc.). The number indicates the highest randomized bit. If the network border – in our example bit 32 – is reached, the process stops. In total, five addresses are generated for each BValue step. This allows to compensate the loss of individual responses or rare positive replies from hitting an assigned address/active network by chance. For higher BValue steps, there is a higher chance of targeting other assigned addresses close to the hitlist address. Therefore, for each step, a majority vote decides on the error message type, ignoring protocol-specific positive responses such as *ER*, *TCPACK*, *RST*. Additionally, we measure *B*127, an address congruent with the seed address, flipping only the last bit.

Measurement Setup We apply our method of BValue Steps to the IPv6 Hitlist Service [14] which provides responsive addresses. For the network borders, we use RIPE RIS BGP looking glass [26]. Preventing bias from networks with many addresses in the hitlist, we only take a single address per BGP-announced prefix. In total, we probed 47,923 addresses with three protocols (ICMPv6, TCP - Port 443, and UDP - Port 53) from two vantage points on five successive days in March 2023 (2023-03-14 to 2023-03-18).

```
Original hitlist address:
          2001:db8:1234:abcd:1234:abcd:1234:0101
Generated addresses:
          <original bits>            <random bits>
B127      2001:db8:1234:abcd:1234:abcd:1234:0100
B120      2001:db8:1234:abcd:1234:abcd:1234:01e8
B112      2001:db8:1234:abcd:1234:abcd:1234:6aa1
B104      2001:db8:1234:abcd:1234:abcd:1221:f38d
...
B48       2001:db8:abcd:5276:d080:ccd6:7fc3:311c
B40       2001:db8:ab3e:3eb7:4c66:7f16:ade5:2b3d
B32       2001:db8:7438:221f:b244:476c:66bb:8da5
```

Figure 3: BValue Steps address generation. From an active address, more and more bits are randomized in steps of 8.

		Vantage 1 (σ)			Vantage 2 (σ)		
W. Ch.	ICMPv6	21,070	(79)	44.2%	20,847	(30)	44.1%
	TCP	18,393	(57)	38.6%	18,142	(25)	38.3%
	UDP	24,620	(108)	51.7%	24,287	(33)	51.3%
W/o. Ch.	ICMPv6	8,165	(41)	17.1%	8,014	(24)	16.9%
	TCP	6,808	(28)	14.3%	6,727	(26)	14.2%
	UDP	6,005	(25)	12.6%	5,879	(29)	12.4%
∅	ICMPv6	18,407	(62)	38.6%	18,461	(28)	39.0%
	TCP	22,441	(89)	47.1%	22,452	(53)	47.4%
	UDP	17,017	(59)	35.7%	17,156	(24)	36.3%

NOTE: # of Networks = mean and σ = standard deviation of five days.

Table 4: As a basis for validation, BValue differentiates networks (i) with a change in ICMPv6 error message, (ii) without such a change, and (iii) unresponsive networks.

Data Set For 44% (ICMP), 38% (TCP), and 52% (UDP) of the seed addresses, we were able to differentiate active from inactive networks by observing at least one change in ICMPv6 error messages, see Table 4. Depending on the protocol, around 12% to 17% of prefixes show no change in error message types and 36% to 47% do not return error messages at all. The results are consistent across both vantage points.

Comparing source addresses at ICMPv6 error message type changes, in 86% of the cases, the change in response type aligns with a change in the source address (and consequently the responding router), further supporting our assumption on network borders. For the other cases a single router serves the target network and other network ranges.

Figure 4 shows the BValue after which the message type changes have been observed. 71.6% are found at B64+, reflecting well-known IPv6 address assignment strategies [30] and supporting our hypotheses. While the overall percentage is low, we also detect networks with multiple network borders. This does not directly impact our labeling, but verifies common network borders used in IPv6. 5% of the networks with a first change also show a second change at the /56 or /48 border. In addition, 0.1% show a third change at the /48 or /40 border. For the remainder of the analysis, we label the message types received for the higher BValues (from B127, i.e., before the first change) to represent active networks, and lower BValues (up to Bxxx, i.e., after the first change) to represent inactive networks.

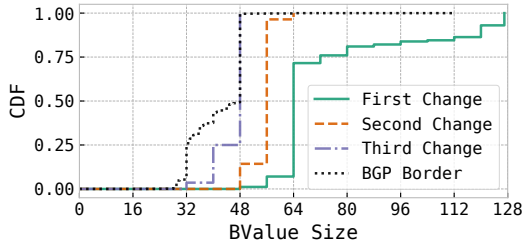
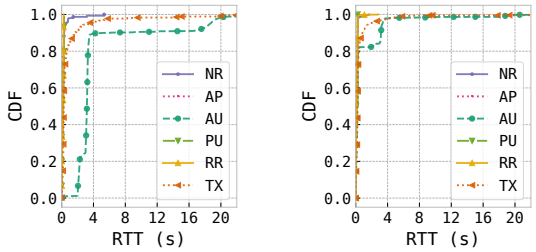


Figure 4: Inferred distribution of IPv6 suballocation sizes for 21,184 (44.2% of measured) IPv6 networks. Results for ICMPv6 on 2023-03-14.



(a) Active Networks **(b) Inactive Networks**

Figure 5: Also on the Internet, *AU* is delayed by multiple seconds for active networks and returned immediately for inactive ones.

Validation - Error Message Response Timings In a first step, we analyze whether the delays of multiple seconds for *AU* – allowing to differentiate $AU_{RTT<1s}$ from $AU_{RTT>1s}$ – are also observed on the Internet. Figure 5 shows RTTs for the ICMPv6 error message types, separated for active and inactive networks. We see sharp increases at 2, 3, and 18 seconds for active networks (22.25% 2s, 68.5% 3s, 9.25% 18s), reflecting the same delays due to Neighbor Discovery timeout that were also observed for router appliances in the laboratory setup. This implies that also on the Internet it is feasible to distinguish *AU* for active networks from those for inactive networks.

Validation - Error Messages for Active Networks In the left column of Table 5, the error message types for probing addresses in active networks are shown. With a probability of 95%, classification is successful as we received message types associated with active networks for ICMPv6. In only 2% of the cases, these networks were classified as ambiguous, i.e., no decision can be made. In 3% of the cases, the networks are incorrectly classified as inactive. We reach comparable classification rates for TCP. UDP performs worse with only 56% of networks labeled as active also classified as active. The reason is as follows: For UDP, we cannot verify if *PU* error messages came from the target itself or were caused by a filter. Thus, we categorize the remote network ambiguous instead of active. For many networks we target assigned IPs close to the hitlist address, resulting in *PU* being returned by assigned IPs close to the hitlist address. A difference of nearly 40 percentage points indicates a large share of networks is affected by this. While this negatively impacts our labeling for UDP, it supports our claim that host discovery in these networks is feasible. Still, we cannot classify *PU* as active due to its usage for firewalling. This renders ICMP the preferred protocol for the task of network activity classification.

Validation - Error Messages for Inactive Networks In the right column of Table 5, the error message types for probing addresses in inactive networks are shown. For ICMPv6, the networks are correctly classified as inactive in 80% of the cases. No classification is feasible in 16% of the cases, and in 5% of the cases they are incorrect. Again TCP shows similar and UDP worse performance.

		labeled active			labeled inactive		
		Netw.	σ	%	Netw.	σ	%
active	ICMPv6	17,361	109	95.1%	471	11	4.6%
	TCP	14,522	112	93.7%	620	12	7.4%
	UDP	12,490	82	56.2%	3,687	35	32.0%
ambig.	ICMPv6	352	10	1.9%	1,645	12	15.9%
	TCP	566	10	3.7%	1,552	14	18.6%
	UDP	9,377	91	42.2%	1,455	7	12.6%
inactive	ICMPv6	537	13	2.9%	8,230	34	79.5%
	TCP	405	8	2.6%	6,191	26	74.0%
	UDP	337	12	1.5%	6,396	49	55.4%

NOTE: σ Standard deviation over five days.

Table 5: Network activity classification (active, ambiguous and inactive) for networks (active, inactive) labeled by BValue Steps.

4.3 Network Activity Scans

In a final step, we performed two prefix-seeded measurements, not including any ground truth, to collect and classify ICMPv6 error messages across a wide portion of the routed IPv6 Internet. In measurement M1, we probed a random address in each routed /48 prefix. Thereby, prefixes of shorter length (e.g., /32) are split in multiple /48 prefixes. In measurement M2, we took only those prefixes that are announced as a /48 in BGP, and exhaustively probe them at the granularity of /64. The first measurement prioritizes breadth over depth and targets more towards the Internet's core, the second measurement focuses on depth instead of breadth and the Internet's periphery.

M1 - Sampling the Internet at /48 Granularity We take a total of 45,434 prefixes with a prefix length of /48 or shorter. We traceroute a random address within each /48 prefix using yarrp [6, 7], resulting in a total of 5Bn destinations measured from our vantage point 1 between 2023-03-16 and 2023-04-05. Figure 6 visualizes the distribution of active, inactive, ambiguous and unresponsive /48 prefixes. We received 616M responses, representing 12% of all destinations. Classifying the received error messages, we find 83M active and 341M inactive /48s. 192M remain ambiguous. For the detailed share of responses, we refer to Table 6. In comparison to previous measurements, the share of unresponsive destinations appears to be high; however, aggregating them to BGP prefixes, only 39% (17,580) of them do not respond at all. This number is comparable to the previous experiment.

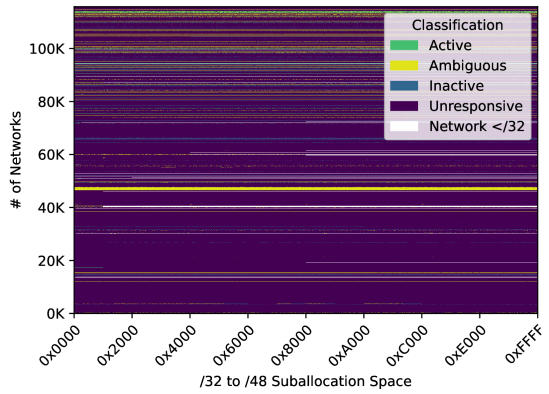


Figure 6: Sampling the Internet at /48 granularity. Each row represents a /32 network and each column one /48 network inside the /32.

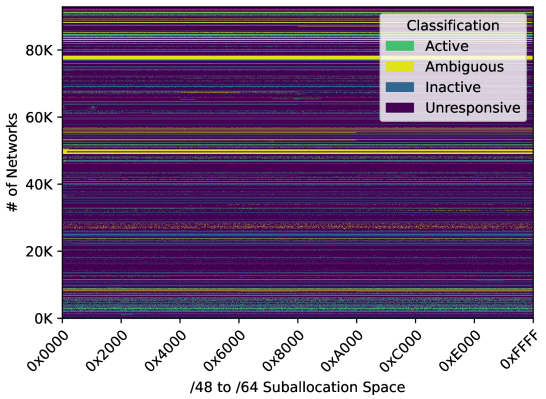


Figure 7: Exhaustive probing of BGP-announced /48 prefixes. Each row represents a /48 prefix and each column a /64 inside the /48.

M2 - Exhaustive Probing of /48s We focus on the 92,856 networks that were announced as a /48 prefix in BGP (2023-11-01) as we are able to exhaustively probe them at the granularity of /64 and investigate behavior of the Internet periphery. Using ZMap, we probe a random address in each encompassed /64 prefix, resulting in a total of 6Bn destinations. Figure 7 visualizes the distribution of message types that are classified as active, inactive, ambiguous and unresponsive. We received 1.4Bn responses, representing 23% of all destinations. We classified 356M as active, 802M as inactive and 210M remain ambiguous. In comparison to M1, we received a higher share of responses that are classified as active. We discovered 45.3M unique sources of error messages, with 14M periphery routers that perform Neighbor Discovery. Of those 4M rely on EUI-64 addresses, with the most represented vendors (>10K routers) being Huawei, ZTE, T3, Dasan, DZS, PPC Broadband, Taicang, Nokia and Netlink. Assigning the responses to the individual BGP-announced prefixes, we see again that 39% of the BGP prefixes do not respond at all, a number similar to the one in M1. It also shows that similar to results from related work, inactive address space is often not routed correctly, leading to routing loops in over 62.9% of prefixes that return error messages [22, 23].

Message Types Table 6 outlines the contribution of the individual error message types to the classification of our network activity scans. In M1 - core we see a higher share of null routing through *RR* (33.3%) and $AU_{RTT<1s}$ (13.1%) while for M2 -periphery we see a higher share (32.8%) of routing loops (*TX*) and active networks indicated by $AU_{RTT>1s}$ (26%). For target networks classified as ambiguous that could both be active or inactive *NR* contributes the most with 20.3% in M1 and 13.6% in M2.

Network Activity We found active networks to account for 1.7% of the IPv6 Internet at /48 granularity. We find a higher share (12%) of active networks for /64 periphery networks. This 12% of active networks are divided across 34,924 – equal to 61% – of responsive /48 prefixes. The respective error messages indicate that the request was forwarded and triggered Neighbor Discovery. This makes them a priority target for further reconnaissance efforts. Narrowing down the search space to these networks is however

Type	M1 - Core	M2 - Periphery
$AU_{RTT>1s}$	13.5%	26.0%
NR	20.3%	13.6%
AP	4.3%	1.6%
FP	0.0%	0.0%
PU	6.5%	0.0%
$AU_{RTT<1s}$	13.1%	16.7%
RR	33.3%	9.1%
TX	8.9%	32.8%
Total	616M	1368M

active, ambiguous and inactive networks.

Table 6: Share of ICMPv6 error message types received in measurements M1 and M2.

tricky as we cannot guarantee that these networks are the only active networks inside the target network range. Active networks with filters might discard our requests and remain silent, i.e., our results have to be considered to represent a lower bound for the number of active networks. We also find 22% of prefixes return error messages for inactive networks only. While these networks can be excluded for host discovery, periphery and subnet discovery does not depend on the returned response type [7]. However, we find that periphery discovery based on error messages is not a solution for every network [22]. For around 38% to 39% of prefixes in the IPv6 Hitlist Service, M1 and M2 do not return error messages. For the remainder of the paper, we focus on networks that return error messages. As response behavior varies between different network equipment vendors, this necessitates the identification of the router type used within a network.

5 Router Classification

Measuring ICMPv6 rate limiting behavior, we found varying behavior among vendors in our lab (§) that could be validated against SNMPv3 labels from routers in the IPv6 Internet (§). Our approach extends router classification to IPv6 routers that are not SNMPv3 responsive. In a final measurement (§), we fingerprinted routers on the Internet. For periphery routers, fingerprinting vendors is

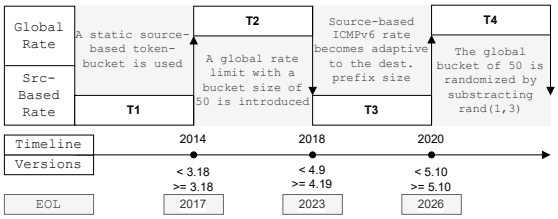


Figure 8: ICMPv6 rate-limiting behavior for different Linuxs kernel versions.

– in comparison to core routers – limited as they mainly show Linux default behavior. Yet, we are able to estimate Kernel versions, detecting many routers that have reached end of life.

5.1 Router Defaults

We rely again on our GNS3 setup, but instead of a single request we send ICMPv6 Echo requests at 200 pps for a time period of ten seconds to (I) unassigned addresses (see IP2 in Figure 1) in active network A triggering $AU_{RTT>1s}$ at the RUT, (II) addresses in inactive network B (IP 3 in Figure 1) triggering *NR*, or (III) with Hop Limits triggering *TX* as a response (as in Scenario (S6) in Section 4.1). The requests contain ascending sequence numbers as a payload and allow to check which requests remain unanswered. The responses are shaped by ICMPv6 rate-limiting behavior, typically implemented with a token bucket algorithm [10]; we infer its parameters as follows:

- *Bucket size:* Missing packets pinpoint the bucket's depletion. We determine the first missing response; its sequence number is equivalent to the bucket size.
- *Refill size:* The refill size is equivalent to the number of replies between two successive depletions. We count this number for all successive depletions and take the median of the collected values as refill size.
- *Refill interval:* The refill interval is equivalent to the time between two refills. Therefore, we infer the time spans between successive responses, remove the ones reflecting our measurement rate (5ms), and take the median. This value represents the pause between two bursts. Combining it with the duration of the previous burst, we infer the refill interval.
- *Number of error messages*: As a simplistic indicator, including the other parameters of a router's rate-limiting behavior, we count the total number of error messages received in a time span of ten seconds.

We conduct this measurement from a single source, and repeat it with two source addresses to see whether rate limits are configured globally or per source address.

Vendor Defaults Table 8 shows our results in detail. Seven routers apply rate limiting per source address, another six only apply a global limit, and two do not limit ICMPv6 error messages at all. We observe differences among vendors – though not among all (e.g., the Linux-based Mikrotik, OpenWRT, VyOS, and Aruba) – but also between routers/versions of the same vendors (e.g., Cisco XRV9000 and Cisco IOS 15.9), and in some cases, even between the different error message types from the same routers (e.g., Juniper and Huawei). While Linux-based routers show a token-bucket rate limit algorithm, FreeBSD ones show generic rate limits, where the refill size equals the bucket size. Another peculiarity has been observed for Huawei; the bucket size is randomly chosen between 100 and 200. This appears to be a countermeasure against idle scanning or exploiting routers as remote vantage points for scanning [4, 28].

Prefix Size	Refill Interval (ms)			# Error Messages
Kernel HZ ->	100	250	1000	
0	60	60	62	165-167
1-32	120	124	125	85-86
33-64	248	248	250	45-46
65-96	500	500	500	25-26
97-128	1,000	1,000	1,000	15-16

Table 7: Since kernel 4.19, the refill interval depends on the IPv6 prefix length and the kernel tick rate.

Linux Kernel Defaults For Mikrotik routers relying on the Linux kernel rate limiting, we observed a difference in behavior between version 6.48 and 7.7, and consequently investigated the limiting behavior of Linux kernels using Debian live images in more detail. This led to detecting a change in the peer-based rate limiting behavior (for completeness we list the values for all tested kernel versions in Table 12 in the Appendix), which is congruent with the change between Mikrotik version 6 and 7.

Linux changed its peer-based rate-limiting behavior between kernel 4.9 and 4.19 (between 2016 and 2018). Before the change, the rate limits behaved static; now, it is dependent on the router's assigned prefix size, see Table 7. Figure 8 shows the evolution of ICMPv6 rate limiting in the Linux kernel over time. The code for the prefix-based rate limit exists since kernel 2.1.111, but was not effective until 4.9/19. In this paper we focus on measuring the peer-based rate limit. Measuring the global rate limits is more invasive as it requires to bypass the peer-based rate limit by measuring with multiple source addresses in parallel. Pan et al. already showed that the global rate limit can be measured this way [28]. Also, hosts with global rate limits were exploited as remote vantage points for scanning [4, 28]. This led to a new behavior of the Linux kernel, similar to Huawei routers, by subtracting a random integer of up to 3 from the default bucket size of 50. Both the introduction and the randomization of the global rate limit provide additional steps to fingerprint the Linux kernel versions. In this paper we aim to separate routers, relying on the Linux kernel rate limiting, based on the peer-based rate limit into T2 or before and T3 and after.

5.2 Rate Limits in the Internet

As a next step, we aim to verify if we see the same fingerprints on the Internet. To do this, we use a dataset with responses to unauthenticated, unsolicited SNMPv3 requests revealing vendor information for 476K IPv6 addresses [2]. We chose to elicit *TX* messages at routers, as these are mandatory according to RFC4433. However, we need a suitable combination of destination address and Hop Limit to trigger *TX* at a certain router. We checked whether the SNMPv3 dataset addresses used as ground truth were also present in our M1 dataset collected by yarrp (see Section 4.3), enabling us to set the destination address and Hop Limit in our requests

	Router OS	iTTL	Delay	Bucket Size			Refill Interval (σ)			Refill Size			# Error Messages			Per
		All	AU	TX	NR	AU	TX	NR	AU	TX	NR	AU	TX	NR	AU	Src
Diff AU/NR/TX	CiscoXRV9000	64	18	10	10	10	1,000	1,000	1,000	1	1	1	19	19	0★	
	CiscoIOS 15.9	64	3	10	10	10	~100	~100	3,800★	1	1	10	~105	~105	22★	
	CiscoCSR1000 17.03	64	3	10	10	10	~100	~100	3,000★	1	1	10	~105	~105	22★	
	Juniper 17.1	64	2	52	12	12	~1,000	10,000	10,000	52	12	12	~520◇	12	12	
	HPE VSR1000	64	3	∞	∞	★	∞	∞	★	∞	∞	★	∞	∞	★	
	Huawei NE40	64	3	100-200	8	/	1,000	1,000	/	100	8	/	1,000-1,100	88	/	
No diff for AU/NR/TX	Arista 4.28	64	3		∞			∞			∞			∞		
	VyOS 1.3	64	3		6			250*			1			45*		✓
	Mikrotik 6.48	64,255	3		6			1000			1			15		✓
	Mikrotik 7.7	64	3		6			250*			1			45*		✓
	OpenWRT 19.07	64	3		6			250*			1			45*		✓
	OpenWRT 21.02	64	3		6			250*			1			45*		✓
	ArubaOS 10.09	64	3		6			250*			1			45*		✓
	Fortigate 7.2.0	255	3		6			10			1			1000		✓
	PfSense 2.6.0	64	3		100			1000			100			1000		

~ ... Refill interval is less stable / ... The response type is not returned by the RUT. ★ ...Affected by the Neighbor Discovery Process. * ... /48 destination prefix; for other prefix sizes see Table 7 ∞ ... RUT is either not rate-limited or > $scanrate$ (tested up to 10K pps). ◇ ... Juniper's Neighbor Discovery for hop limit 0 packets causes a 2-second delay also for TX.

Kernels: Linux, Wind River Linux and FreeBSD.

Table 8: ICMPv6 rate limiting behavior of routers observed in GNS3 laboratory setup. Parameters vary among vendors, versions, and sometimes even for message types.

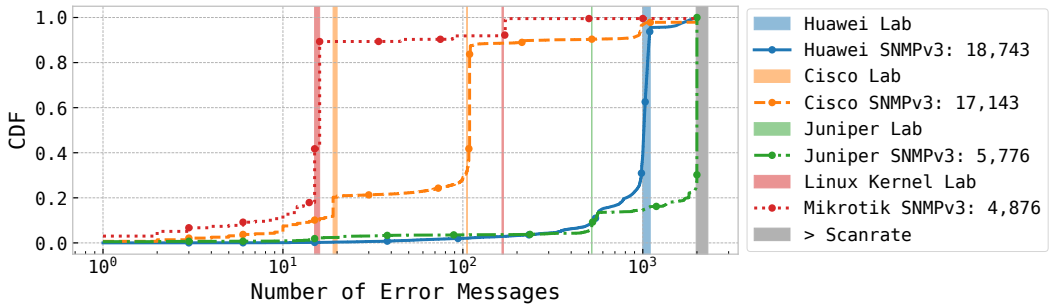


Figure 9: No. of error messages in 10 s for SNMPv3 routers matches the laboratory results (marked vertically).

accordingly. As before, we sent requests at a rate of 200 pps over a time period of 10 s. This way, we could validate the behavior of 50,952 IPv6 addresses against SNMPv3 labels.

Classification To match router rate limits to recorded vendor fingerprints, we rely on a more elaborate approach than comparing the number of received error messages. In the first step, classification is based on one-dimensional vectors, each element describing the number of received ICMPv6 error messages per second. If the distance between a router's behavior and the collected labels lies within a predefined threshold, we assign the respective label. The threshold is adaptive based on the total number of error messages received ranging from 10 (<100 error messages) to 100 (<2,000 error messages). Only if labels from different routers overlap, we compare the token bucket algorithm's parameters of refill interval and refill size in a second step. From the fingerprints that match all these, the vendor fingerprint with the lowest distance from the one-dimensional vector is selected. If this is not the case, we classify it as *New Pattern*. Some routers in the Internet measurements appear to apply a dual double token bucket algorithm for rate limiting, including two refill intervals and sizes. With the refill interval being the median of refill intervals, we rely on the skewness measure $abs(1 - mean/median) > 0.5$ to check for a second refill interval and label these routers accordingly.

Comparison with Virtual Laboratory Figure 9 compares the total number of ICMPv6 error messages returned by routers recorded in the lab compared to the number of error messages collected for SNMPv3-labeled routers in the IPv6 Internet. Each vertical line represents the number of error messages we saw for the specific vendor in our laboratory. We see overlapping behavior with our results from the virtual laboratory setup: Applying our classification, the rate-limit patterns observed in our laboratory account for 70% of the Cisco, 51% of the Huawei, and 91% of the Mikrotik routers in our Internet measurement. In contrast, the Juniper router (Junos 17.1) from the laboratory only accounts for 5% of the Juniper-labeled routers, and HPE label for 7%. This is, however, not surprising as our laboratory setup contains only a limited set of routers. Juniper routers' rate-limiting implementations seem to vary more across different versions than for other vendors [21], and we found that 82% of Juniper-labeled routers are rate-limited above our scanrate of 200 pps. However, we do not conduct scans with higher pps due to ethical considerations.

Additional Fingerprints SNMPv3 labels allow to extend our fingerprints from the laboratory setup. We relied on clustering with varying k-values from 2 to 10 [19] on the one-dimensional vector to detect rate-limiting patterns for each vendor. We used the elbow method to detect the number of different error message

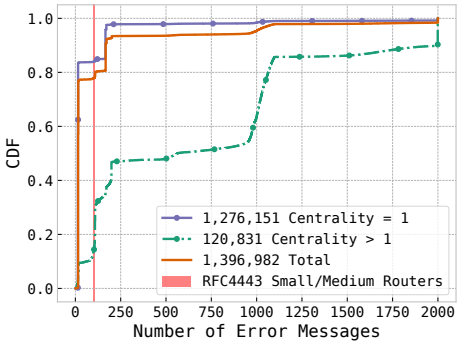


Figure 10: Routers on multiple paths (centrality>1) have higher rate limits than ones on one path (centrality=1).

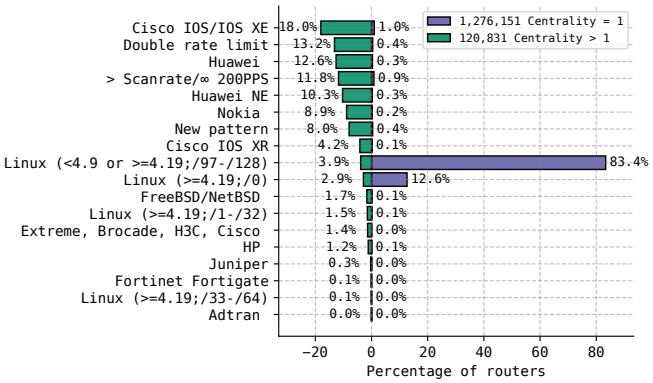


Figure 11: Router classification. Core routers (centrality > 1) are diverse, periphery routers (centrality=1) mostly Linux-based.

rates for each vendor. We found that vendors show a maximum of four different rate-limiting patterns. Based on the patterns, we manually inferred additional fingerprints for Nokia (Number of error messages over 10 seconds=100-200), HP (NR10=5), Adtran (NR10=42), and Huawei (NR10=1000-1100,550) routers. In addition to NR10 we also extract the bucket size, refill intervals and refill sizes.

Multi Vendor Fingerprints We also detected overlapping behavior. The SNMPv3-label H3C reveals the same fingerprint as for Cisco IOS and IOS XE, but there remains a subtle difference – H3C is more likely to show 11 initial responses – facilitating their separation with improved classification. For sure, we will never be able to differentiate vendors if all the rate-limiting parameters are identical. This is the case for Extreme, Brocade, H3C, and Cisco sharing a common fingerprint of a random bucket size between 10 and 20, a refill interval of 100ms and a refill size of 10.

5.3 Router Classification on the Internet

Finally, we conducted a large-scale study of router types on the Internet. From the M1 dataset (see Section 4.3), we extracted all addresses responding with a *TX* as we consider them to be router addresses. We sent requests at a constant rate of 200 pps to the destination addresses with the respective router en route and set the Hop Limit accordingly; both were also inferred from the tracerouting data set. This way, we could infer a rate limit for 1,396,982 IP addresses.

Results Figure 10 shows the total number of returned *TX* messages over a period of 10 seconds for all these routers and dominant behavior at 15 packets. We separate routers into two groups: those with $centrality = 1$, assumed to be on the Internet's periphery (appearing on a single path), and those with $centrality > 1$, located closer to the Internet's core (appearing on multiple paths). Figure 10 shows distinct results for these two groups and suggests that they consist of different router types.

Figure 11 shows our results: Core routers come from different vendors, including Cisco (multiple fingerprints combined: 22.2%), Huawei (multiple fingerprints combined: 22.9%), and Nokia (8.9%). Meanwhile, 83.4% of the periphery routers either rely on Linux kernel version 4.9 (or even older) or a current kernel version with an assigned prefix length of /97-/128. However, as such long prefix lengths are not very common on the Internet [27], this implies that up to 1,066,856 routers in our measurement reached the end of life in January 2023. Only 12.6% of the periphery routers run newer kernel versions.

Comparison with SNMPv3 and LFP Our approach of ICMPv6 error message based router classification extends SNMPv3 router labels similarly to LFP for IPv4 [3] especially since only 476,000 IPv6 routers were found to be SNMPv3 responsive [2]. Unlike unsolicited SNMPv3 responses, rate limiting is unlikely to be disabled in the future, and routers must return *TX* error messages per RFC4443. However, we also faced drawbacks similar to those of LFP. Our classification does not fully cover every router vendor. Juniper is underrepresented in our dataset, as most Juniper routers on the Internet are rate-limited above our scan rate. In contrast to SNMPv3 engineIDs that uniquely identify router vendors, our classification also includes multi-vendor labels, and we cannot distinguish vendors that rely on the Linux kernel default. Notably, Linux kernel version fingerprinting is a new application not previously attempted.

6 Related Work

ICMP Error Messages For both protocols, IPv4 and IPv6, the most prominent use of ICMP error messages has been topology discovery (tracerouting) [6, 9, 16] and analysis of error message types mainly focused on routing loop detection [23, 31]. For the IPv4 Internet, Bano et al. [5] and Rüth et al. [31] investigated ICMPv4 error messages that had been received as a byproduct of ZMap-based measurements. Depending on the protocol for probing, 0.7% (ICMPv4), 9.0% (TCP), resp. 81.3% (UDP) of all responses were ICMPv4 error messages [5]. In IPv6, the share of error messages tends to be higher: For addresses close to responsive addresses, ICMPv6 error messages have a share of 60% and this number rapidly increases to 99% for addresses with more random bits, see Table 10 in the Appendix.

IPv6 Reconnaissance Initial approaches of IPv6 reconnaissance relied on the collection of addresses from public sources [12, 15] or algorithms generating target addresses based on a data set of known addresses [7, 13, 24, 35]. The most successful passive approach by Rye et al. relied on NTP servers to collect active addresses from 7.2M /48s [32]. Rye and Beverly performed active measurements by intentionally triggering error messages to collect 64 million IPv6 addresses at the Internet periphery [33], and Li et al. [22] scanned 15 IPv6 ISP network ranges to trigger *Destination Unreachable* messages at periphery routers collecting 52 million IPv6 addresses. Both approaches triggered ICMP error messages and extracted source addresses, but did not further classify the periphery as we did by interpreting distinct error types and timings. Our measurements also point out a limitation. Across all of our measurements, we find approx. 38% of IPv6 prefixes that do not return ICMPv6 error messages, rendering all above-mentioned approaches useless for these networks.

Router Classification Vanauble et al. [36] derived router signatures from the initial *TTL* (iTTL) values of *Time Exceeded* resp. *Echo Reply* messages; this way, the authors could infer a router's vendor through remote measurements. Holland et al. [20] combined banner grabbing (SSH, SNMP, telnet) and active probing of routers to train an automated classifier. While the classifier was trained on diverse features, the iTTL remained the most distinctive feature to distinguish vendors. Meanwhile, iTTLs are harmonized among the majority of vendors (see Table 8 and [8]). Consequently, a new methodology is necessary to distinguish vendors. The most recent approach for router classification found 476,000 routers that reply to unauthenticated SNMPv3 requests with replies including vendor-specific engineIDs [2]. Albakour et al. extended their approach for non-SNMPv3 responsive IPv4 routers by including protocol header specific information such as the IPID [3]. However, this methodology cannot be applied for IPv6 due to the required fields missing in IPv6 and harmonized iTTL values. Our methodology complements [2] for non-SNMPv3 responsive IPv6 routers by developing a classification method based on ICMPv6 error message rate limiting behavior of routers.

ICMPv6 Rate Limits Ravaioli et al. classified ICMP rate limiting in different categories such as on-off behavior or generically rate-limited routers. They also explored the effects of increasing probing rates and found that higher probing rates lead to more irregular on-off behavior of routers [29]. While we keep the probing rates low, we also noticed irregular behavior, but assume it is triggered by other entities impacting the routers global rate limit. Previous work exploited routers' rate-limiting behavior for purposes other than router vendor classification. Vermeulen et al. [37] conducted alias resolution, i.e., identifying IP addresses belonging to the same router. As aliased addresses are subject to the same rate limit, probing them simultaneously triggers rate limiting and distinct loss patterns. Security-wise, Pan et al. [28] showed how to exploit remote routers' global error message rate limits to use them as vantage points for network scans. The same concept is used by Albrecht et al. [4] to perform UDP idle scans through remote routers. To the best of our knowledge, we are first to explore ICMPv6 error message rate limiting in such detail, as well as its exploitability for router classification.

7 Discussion

Error Message Classification Our classification of error message types does not provide a guarantee that every router behaves accordingly. While we used BValue steps to quantify the response behavior of active and inactive IPv6 networks, the classification could be impacted by either the correctness of inferred active and inactive networks or routers that misuse the respective error message type. However, we cannot distinguish between those two cases. Most of the classified IPv6 networks behave accordingly and for 95% of networks identified as active we receive $AU_{RTT>1s}$.

Prefix Boundary Precision We used BValue steps to separate active from inactive networks, but the precision of network border detection could be further improved. First, the randomization of the address bits might results in an overlap with the original hitlist address. The probability for the first bit to overlap is 50%, the probability for the eight bits to overlap 0.4%. A pseudo-random address flipping the first bit of the BValue would increase precision as the generated address would not overlap. However, we generate five addresses per BValue step indicating that on average 2.5 addresses deviate already in the first bit. Second, the step width impacts prefix boundary precision. We opted for eight bits as a trade-off to cover major prefix boundaries. A change at a non-eight-bit prefix boundary, e.g., /60, is misclassified as a /56. However, the occurrences appear to be limited in practice, Table 11 in Appendix C shows that we received only one response type in 97% of BValues, suggesting that changes within a BValue step are a minor phenomena.

Unique Vendor Identification Our approach allows us to determine a router's vendor and/or operating systems based on the ICMPv6 rate limiting behavior, turning a protection mechanism into a privacy leak. There is room for improvement: First, we were limited to the fingerprints collected in our laboratory setup and the SNMPv3 data set. These fingerprints do not cover the whole Internet population as our Internet measurement revealed unknown patterns. Second, certain fingerprints overlap. While this does not allow unambiguous classification, it still allows to narrow the field down to a few vendors. This information could then be combined with other approaches. For example, one could investigate the rate limiting behavior of different error message types and compare with Table 9.

Old Kernels used in Periphery Routers In our measurements, we identified 1.2M routers operating a Linux version of 2018 or older that have already reached end of life. This does not mean that they are exploitable, but in case of a vulnerability no updates will be made available for this significant share of periphery routers.

Countermeasures Strict adherence to RFC4443 [10] only facilitates network activity classification by making router behavior more consistent. For router classification, the consequences are the opposite. More congruent ICMPv6 rate limiting (e.g. more specific values could be proposed by the RFC) would hinder classification of vendor and operating system. Disabling ICMPv6 error messaging mitigates both, Network Activity Classification and Router Classification, and is also compliant with the specification. In fact, this is the case for 38% of the investigated networks. For these networks,

our approaches were not successful, but also previous work collecting addresses from ICMPv6 error messages [22] would fail for such networks. In addition, disabling hinders network diagnosis by administrators. Removal of the rate limits would, according to the specification, put the routers at risk of denial-of-service attacks; still, some of investigated routers appear to operate properly without such limits.

8 Conclusion

In this paper, we developed two new measurement methods, exploiting ICMPv6 error messages beyond the mere extraction of source addresses, to to gain more insight into remote networks. For each method, we first established our hypothesis in a virtual laboratory setup, then validated the results via measurements involving ground truth, and, finally, conducted exemplary Internet measurements. Our work is summarized as follows: (I) Routers return ICMPv6 messages in non-specified ways with RFC4443 which negatively impacts the messages' diagnostic value. Nevertheless, we were able to classify them regarding the activity status of the remote network by combining ICMPv6 message type, subcode, and timing behavior. Our method is able to guide scanning efforts towards active networks where responsive IPv6 addresses reside. (II) ICMPv6 rate limits protect routers against denial-of-service, but the implementations vary significantly among different router vendors and operating systems. We use them for router classification and, by measuring 1.4 million routers on the Internet, we discovered different populations for core and periphery routers; the latter are primarily Linux-based (96.0%). Most prevalent (83.3%) is Linux kernel version 4.9 and older which have reached end of life in January 2023.

Acknowledgements

This material is based upon work partially supported by (1) the Christian-Doppler-Laboratory for Security and Quality Improvement in the Production System Lifecycle; the financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged; (2) SBA Research (SBA-K1), a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the province of Vienna.(3) Project DynAISEC FO999887504 funded by the Program "ICT of the Future" – an initiative of the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology, (4) the Austrian Science Fund (FWF) (SFB SPyCoDe F85).

References

[1] David Adrian, Zakir Durumeric, Gulshan Singh, and J Alex Halderman. 2014. Zippier zmap: internet-wide scanning at 10 gbps. In *8th USENIX Workshop on Offensive Technologies (WOOT 14)*.

[2] Taha Albakour, Oliver Gasser, Robert Beverly, and Georgios Smaragdakis. 2021. Third time's not a charm: Exploiting SNMPv3 for router fingerprinting. In *Proceedings of the 21st ACM Internet Measurement Conference*. 150–164.

[3] Taha Albakour, Oliver Gasser, Robert Beverly, and Georgios Smaragdakis. 2023. Illuminating Router Vendor Diversity Within Providers and Along Network Paths. In *Proceedings of the 2023 ACM on Internet Measurement Conference*. 89–103.

[4] Martin Albrecht. 2019. UDP idle scanning. Retrieved Nov 20, 2023 from https://martinralbrecht.wordpress.com/2019/10/25/udp-idle-scanning.

[5] Shehar Bano, Philipp Richter, Mobin Javed, Srikanth Sundaresan, Zakir Durumeric, Steven J Murdoch, Richard Mortier, and Vern Paxson. 2018. Scanning the internet for liveness. *ACM SIGCOMM Computer Communication Review* 48, 2, 2–9.

[6] Robert Beverly. 2016. Yarrp'ing the Internet: Randomized High-Speed Active Topology Discovery. In *Proceedings of the 2016 Internet Measurement Conference*. ACM, 413–420.

[7] Robert Beverly, Ramakrishnan Durairajan, David Plonka, and Justin P Rohrer. 2018. In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In *Proceedings of the Internet Measurement Conference 2018*. ACM, 308–321.

[8] Wesley G Bofman and Fernando Maniego. 2019. *Fingerprinting IPv4 and IPv6 routers using ICMP*. Ph. D. Dissertation. Monterey, CA; Naval Postgraduate School.

[9] Caida. 2007. Archipelago (Ark) Measurement Infrastructure. http://www.caida.org/projects/ark/

[10] A. Conta, S. Deering, and M. Gupta (Ed.). 2006. Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification. RFC 4443 (Internet Standard). https://doi.org/10.17487/RFC4443 Updated by RFC 4884.

[11] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide scanning and its security applications. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. 605–620.

[12] Tobias Fiebig, Kevin Borgolte, Shuang Hao, Christopher Kruegel, and Giovanni Vigna. 2017. Something from Nothing (There): Collecting Global IPv6 Datasets from DNS. In *Passive and Active Measurement*, Mohamed Ali Kaafar, Steve Uhlig, and Johanna Amann (Eds.). Springer International Publishing, Cham, 30–43.

[13] Pawel Foremski, David Plonka, and Arthur Berger. 2016. Entropy/IP: Uncovering Structure in IPv6 Addresses. In *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC '16)*. ACM, New York, NY, USA, 167–181. https://doi.org/10.1145/2987443.2987445

[14] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczynski, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proceedings of the 2018 Internet Measurement Conference* (Boston, MA, USA). ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3278532.3278564

[15] Oliver Gasser, Quirin Scheitle, Sebastian Gebhard, and Georg Carle. 2016. Scanning the IPv6 Internet: Towards a Comprehensive Hitlist. In *Proc. of 8th Int. Workshop on Traffic Monitoring and Analysis*. Louvain-la-Neuve, Belgium.

[16] Eric W Gaston. 2017. *High-frequency mapping of the IPv6 Internet using Yarrp*. Technical Report. Naval Postgraduate School Monterey United States.

[17] Fernando Gont. 2013. scan6 - An IPv6 host scanner. http://manpages.ubuntu.com/manpages/cosmic/man1/scan6.1.html

[18] F. Gont and T. Chown. 2016. Network Reconnaissance in IPv6 Networks. RFC 7707 (Informational). https://doi.org/10.17487/RFC7707

[19] Allan Grønlund, Kasper Green Larsen, Alexander Mathiasen, Jesper Sindahl Nielsen, Stefan Schneider, and Mingzhou Song. 2017. Fast exact k-means, k-medians and Bregman divergence clustering in 1D. *arXiv preprint arXiv:1701.07204* (2017).

[20] Jordan Holland, Ross Teixeria, Paul Schmitt, Kevin Borgolte, Jennifer Rexford, Nick Feamster, and Jonathan Mayer. 2020. Classifying Network Vendors at Internet scale. *arXiv preprint arXiv:2006.13086* (2020).

[21] Juniper. 2022. ICMP Features. Retrieved Mai 25, 2023 from https://www.juniper.net/documentation/us/en/software/junos/transport-ip/topics/topic-map/icmp.html.

[22] Xiang Li, Baojun Liu, Xiaofeng Zheng, Haixin Duan, Qi Li, and Youjun Huang. 2021. Fast IPv6 Network Periphery Discovery and Security Implications. In *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 88–100.

[23] Markus Maier and Johanna Ullrich. 2023. In the loop: A measurement study of persistent routing loops on the IPv4/IPv6 Internet. *Computer Networks* 221 (2023), 109500.

[24] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. 2017. Target generation for Internet-wide IPv6 scanning. In *Proceedings of the 2017 Internet Measurement Conference*. ACM, 242–253.

[25] T. Narten, E. Nordmark, W. Simpson, and H. Soliman. 2007. Neighbor Discovery for IP version 6 (IPv6). RFC 4861 (Draft Standard). https://doi.org/10.17487/RFC4861 Updated by RFCs 5942, 6980, 7048, 7527, 7559, 8028, 8319, 8425.

[26] "RIPE NCC". 2018. Routing Information Service (RIS). https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris

[27] Ramakrishna Padmanabhan, John P Rula, Philipp Richter, Stephen D Strowes, and Alberto Dainotti. 2020. DynamIPs: Analyzing address assignment practices in IPv4 and IPv6. In *Proceedings of the 16th international conference on emerging networking experiments and technologies*. 55–70.

[28] Long Pan, Jiahai Yang, Lin He, Zhiliang Wang, Leyao Nie, Guanglei Song, and Yaozhong Liu. 2023. Your Router is My Prober: Measuring IPv6 Networks via ICMP Rate Limiting Side Channels. In *30th Annual Network and Distributed System Security Symposium, NDSS 2023, San Diego, California, USA, February 27 - March 3, 2023*. The Internet Society. https://www.ndss-symposium.org/ndss-paper/your-router-is-my-prober-measuring-ipv6-networks-via-icmp-rate-limiting-side-channels/

[29] R. Ravaioli, G. Urvoy-Keller, and C. Barakat. 2015. Characterizing ICMP rate limitation on routers. In *2015 IEEE International Conference on Communications (ICC)*. 6043–6049. https://doi.org/10.1109/ICC.2015.7249285
[30] RIPE. 2020. *IPv6 Address Allocation and Assignment Policy*. Technical Report. https://www.ripe.net/publications/docs/ripe-738/
[31] Jan Rüth, Torsten Zimmermann, and Oliver Hohlfeld. 2019. Hidden Treasures – Recycling Large-Scale Internet Measurements to Study the Internet's Control Plane. In *Passive and Active Measurement*, David Choffnes and Marinho Barcellos (Eds.). Springer International Publishing, Cham, 51–67.
[32] Erik Rye and Dave Levin. 2023. Ipv6 hitlists at scale: Be careful what you wish for. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 904–916.
[33] Erik C. Rye and Robert Beverly. 2020. Discovering the IPv6 Network Periphery. In *Passive and Active Measurement*, Anna Sperotto, Alberto Dainotti, and Burkhard Stiller (Eds.). Springer International Publishing, Cham, 3–18.
[34] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. 2023. Target Acquired? Evaluating Target Generation Algorithms for IPv6. In *Proceedings of the Network Traffic Measurement and Analysis Conference (TMA)* (Naples, Italy).
[35] J. Ullrich, P. Kieseberg, K. Krombholz, and E. Weippl. 2015. On Reconnaissance with IPv6: A Pattern-Based Scanning Approach. In *2015 10th International Conference on Availability, Reliability and Security*. 186–192. https://doi.org/10.1109/ARES.2015.48
[36] Yves Vanaubel, Jean-Jacques Pansiot, Pascal Mérindol, and Benoit Donnet. 2013. Network fingerprinting: TTL-based router signatures. In *Proceedings of the 2013 conference on Internet measurement conference*. 369–376.
[37] Kevin Vermeulen, Burim Ljuma, Vamsi Addanki, Matthieu Gouel, Olivier Fourmaux, Timur Friedman, and Reza Rejaie. 2020. Alias resolution based on ICMP rate limiting. In *Passive and Active Measurement: 21st International Conference, PAM 2020, Eugene, Oregon, USA, March 30–31, 2020, Proceedings 21*. Springer, 231–248.
[38] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. 2022. Rusty clusters? dusting an ipv6 research foundation. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 395–409.

A Ethics

In our measurements, we followed the rules of good Internet citizenship [11]. We received a single request for opt-out and acted accordingly. For both measurements, we only sent requests which are typical for Internet traffic. Overall, for network activity classification, the number of requests has been moderate. For validation with BValue Steps, we send 62 requests to a /32 prefix; in our Internet measurements, we send a single request per /48 and /64 prefix respectively. The targets were randomized to prevent the overloading of individual routers. For router classification, we intentionally triggered ICMPv6 rate limiting behavior of routers. Measurements exploiting ICMP rate limiting have been conducted before and related work found no significant performance degrades on routers [28, 37]. Our measurements do not impact the forwarding abilities of routers, it could only deny the origination of ICMPv6 error messages for routers with global rate limits. For example, a potential consequence might be that the affected routers would not reply to tracerouting efforts of other Internet hosts during our measurement period. We found such global rate limited routers to be present in the Internet-core. Periphery routers apply peer-based rate limits and only refrain from returning further ICMPv6 error messages to our measurement host. To minimize our impact on the routers, we limited our measurements to 200 pps and a maximum measurement period of 10 s, resulting in a total of 2,000 requests. Thereby, we did our best to minimize the impact on the router: Packet sizes were kept small to minimize bandwidth consumption; as payload, they only included a request ID and the sent timestamp. Beyond that, we decided to elicit *TX* like in tracerouting campaigns (instead of *AU*) to prevent stateful and, thus, more resource-intensive address resolution. It has to be highlighted that IPv6 rate limits are a protection measure. They prevent routers from sending too many ICMPv6 error messages in order to maintain their main functionality of packet forwarding.

B Vendor Coverage

We provide details on vendor coverage for tested routers and state how the router images relate to each vendor. The individual vendors are listed in Table 9. A lot of router operating systems nowadays rely on the Linux kernel. However, operating systems are still customized based on the vendor's needs. We will take a closer look at each RUT. We test three different images for ***Cisco***. Cisco IOS (Internet Operating System) version 15.9 (2019) is the original monolithic operating system developed by Cisco. The Cisco CSR1000v represents a virtual router series that has been designed for cloud services. It runs a subset of Cisco IOS XE, which supports the same commands as IOS as it runs IOS as a separate process, but is built on Linux. In contrast, IOS XR has a completely different codebase. The XRv 9000 Version 7.2.1 is a virtualized router implementing the feature set of IOS XR. IOS XR is originally based on a microkernel provided by QNX, but has changed to Wind River Linux since version 6. With these images, our lab setup covers Cisco's main networking software for routers. The remaining Cisco NX-OS is the operating system for a series of Cisco switches. We found different ICMPv6 error message implementations for all three router OSes with IOS and IOS-XE being more similar. IOS-XR shows a more diverse behavior with unique Neighbor Discovery timings and response type usage in filtering scenarios. ***Juniper*** runs Junos as a single operating system across its router and switches. The image in our lab is Junos VMx Version 17.1. In contrast to the Cisco operating systems it is based on FreeBSD. It is the second appliance

	Protocols	(S1) Active Network	(S2) Inactive Network	(S3) Active Network with ACL	(S4) Inactive Network with ACL	(S5) Null Route	(S6) Routing Loop
Cisco IOS XR (XRv 9000 7.2.1)	All	AU [18s]	NR	∅	AP	∅	TX
Cisco IOS (15.9 M3)	All	AU [3s]	NR	AP/FP•	AP/FP•	RR	TX
Cisco IOS-XE (CSR1000v17)	All	AU [3s]	NR	AP	AP	RR	TX
Juniper Junos (VMx 17.1)	All	AU [2s]	NR	AP	AP	AU/∅•	TX
HPE (VSR1000)*	All	AU [3s]	NR	AP	AP	∅	TX
Huawei (NE40)	All	∅	NR	-	-	∅	TX
Arista (vEOS 4.28)	All	AU [3s]	NR	-	-	∅	TX
VyOs (1.3)	All	AU [3s]	NR	PU	NR★	∅	TX
Mikrotik (6.48/7.7)	All	AU [3s]	NR	NR	NR★	NR/AP/	TX
OpenWRT (19.07/21.02)	ICMP/U	AU [3s]	FP	PU	FP★	NR/AP/	TX
	TCP	AU [3s]	FP	RST	FP★	NR/AP/	TX
ArubaOS (OS-CX)	All	AU [3s]	NR	∅	∅	AP	TX
Fortigate (7.2.0)*	All	AU [3s]	NR	∅	∅	∅	TX
PfSense (2.6.0)*	ICMP	AU [3s]	NR	∅	∅	-	TX
	TCP	AU [3s]	NR	∅/RST•	∅/RST•	-	TX
	UDP	AU [3s]	NR	∅/PU•	∅/PU•	-	TX

• Multiple ACL/route options. [] Minimum delay. - Not supported. ★ ACL on forward chain. Error messages indicating active, ambiguous and inactive networks.

Table 9: ICMPv6 error message behavior of routers as observed in GNS3 laboratory setup. For router configurations see https://github.com/sbaresearch/router-lab.

BValues	active	ambiguous				inactive			ER	Responsive	Targets
	$AU_{RTT>1s}$	NR	AP	FP	PU	$AU_{RTT<1s}$	RR	TX			
B127	49.3%	3.1%	1.4%	0.1%	0.0%	2.2%	0.9%	2.8%	40.2%	20,319	47,922
B120	71.3%	3.7%	1.1%	0.1%	0.0%	3.9%	1.8%	7.4%	11.1%	28,693	47,922
B112	78.3%	4.2%	1.3%	0.1%	0.0%	4.5%	2.2%	8.7%	0.7%	25,447	47,922
B64	77.4%	4.4%	1.3%	0.1%	0.0%	4.7%	2.4%	9.6%	0.1%	24,981	47,879
B56	29.2%	11.8%	1.7%	0.1%	0.0%	16.1%	9.1%	31.8%	0.1%	19,102	46,847
B48	24.5%	12.8%	1.6%	0.1%	0.0%	17.6%	9.9%	33.4%	0.2%	18,176	46,743
B40	13.3%	16.3%	1.7%	0.0%	0.0%	14.8%	15.6%	38.1%	0.1%	6,246	19,585
B32	12.2%	15.9%	0.9%	0.1%	0.0%	16.4%	18.5%	35.9%	0.1%	3,670	10,669

Table 10: Selected BValue steps for 47,922 IPv6 prefixes showing the transition from active to inactive error message types. The two rightmost columns show the number of responsive targets vs. the total ones.

No. of message types	Protocol	No. of responses 1	2	3	4	5
1	ICMPV6	4.0%	3.0%	4.0%	6.0%	80.0%
	TCP	4.0%	4.0%	5.0%	7.0%	79.0%
	UDP	4.0%	4.0%	4.0%	6.0%	79.0%
2	ICMPV6	1.0%	0.0%	0.0%	1.0%	0.0%
	TCP	1.0%	0.0%	0.0%	0.0%	0.0%
	UDP	1.0%	0.0%	0.0%	1.0%	0.0%
3	ICMPV6	0.0%	0.0%	0.0%	0.0%	0.0%
	TCP	0.0%	0.0%	0.0%	0.0%	0.0%
	UDP	0.0%	0.0%	0.0%	0.0%	0.0%

Table 11: Mean number responses in relation to number of message types for BValue Steps.

that shows unique response timings by returning *AU* after a timeout of 2 seconds. It is also the only appliance that returns *AU* in the presence of null routes.

The ***HP*** image in our lab belongs to the virtual router series (VSR1000). The virtual router runs the same Comware version 7 operating system as HPE routers and switches. Since version 7 the operating system has switched to the Linux kernel. HP was the only vendor where ICMPv6 error messages were deactivated by default. After enabling them it shows simlar behavior to previous vendors. ***Huawei*** runs its own operating system Versatile Routing Platform (VRP) on its routers. The Huawei image is a virtualized version of the NetEngine 40E series. However, the image has limited capabilities, i.e. configuring ACLs is not possible. It is also the only image that does not return *AU* for unassigned IP addresses. Networks with huawei routers behaving in a similar way could therefore be missing in our list of active networks. Howerver, an analysis of EUI-64 addresses for M2 in §5.3 yielded Huawei routers to be the most prominent for active periphery. Thus other versions of VRP might behave differently in regards of active networks. The same limiation for ACL configuration accounts for ***Arista*** vEOS 4.28. Arista's Extensible Operating System (EOS) which is again a Linux-based network operating system.

VyOS is an open virtual Linux-based network operating system. It arose as a community fork from Vyatta which was based on the Debian Linux-distribution. ***Mikrotik*** products are more targeted towards small office, home office (SOHO) customers. Mikrotik features its own router operating system RouterOS. We include both version 6.48 and 7.7.1 in our lab, as they run different versions of the Linux kernel. While we see no difference in error message type, we found the rate limiting behavior to change between these versions. With Mikrotik we also begin to see a clear change in error message behavior for scenarios including ACLs and null routes. ***OpenWrt*** is a vendor-independent network operating system based on Linux for embedded devices. We include both version 19.07 which is based on kernel version 4.14 and 21.02 which is based on kernel version 5.4. OpenWRT is the only appliance to return *FP* in (S2). ***Aruba*** ArubaOS-CX is a virtual switch simulator implementing the features of the Linux-based AruboaOS-CX operating system. However, its layer 3 functionality allows it to be tested in the lab setup. To show that also firewall appliances return ICMPv6 error messages we included ***PfSense*** and ***Fortigate***. While by default any inbound traffic is rejected, we configured rules to explicitly forward traffic to the target network.

C BValue Steps Responsiveness & Borders

We provide details about the share of different message types for our BValue Steps approach. Table 10 gives an overview of the received ICMPv6 responses for the different generated addresses, sorted by message types associated with active respective inactive networks and ambiguous ones. For B127 we find that from the original 47,922 prefixes only 20,319 are responsive. This highlights that traffic in these networks is only forwarded to the hitlist address. For the 42% of prefixes we receive a response in 40% of cases we target an assigned IP, while in the other 60% we target an unassigned IPs. For B120 we also see a higher share of responses from assigned IPs. This shows the presence of other assigned IPs close to the hitlist address. Tools like scan6 [17] can exploit these address patterns to perform host discovery in these networks. We also notice a clear shift from networks returning $AU_{RTT>1s}$ from B127 to B64 to networks returning *NR*, $AU_{RTT<1s}$, *RR* and *TX* for B56 to B32. *BValue Step Width* In the beginning we experimented with step widths of 4, 8 and 16 bits. We decided for a step width of 8-bit as a trade-off in number of probes and covering the major prefix boundaries. A change at a prefix boundary such as /60 currently results in a change in B56 in Figure 4. However, the overall occurrence of such non 8-bit boundaries appear to be limited as highlighted in Table 11. In 97% of the BValue steps, we receive a single error message type suggesting that these cases are . Future work could repeat the measurements with lower step sizes for more fine-grained prefix boundary detection. For most BValue Steps (80%) one message type and five responses are received.

D Linux & BSD Kernel Default Behavior

Table 12 shows the change in response behavior between Linux kernel versions 4.9 and 4.19. We automated the network configuration of Debian-live CDs (https://cdimage.debian.org/mirror/cdimage/archive/) in qemu by redirecting the serial console. This way, we can trigger error messages and measure the behavior of the underlying Linux kernels beginning in 2014. With each major Debian version,

	Kernel Version	Release	IPv4	IPv6
Linux	2.6.26-1-2	2008	15	15
	3.16.0-4-6	2014	15	15
	4.9.0-3-13	2016	15	15
	4.19.0-5-21	2018	15	45
	5.10.0-8-22	2020	15	45
	6.1.0-9	2022	15	45
Freebsd	11.0	2016	2000	1000
Netbsd	8.2	2020	1000	1000

Table 12: Error messages (NR(10)) for *TX* for IPv4 and IPv6 of different Linux kernels yielding a change between version 4.9 and 4.19.

the underlying Linux kernel version also changed. We tested Debian version 5 with Linux kernel 3.16.0 up to 12 with Linux kernel version 6.1.9. We failed to automate different versions for FreeBSD and manually verified the rate-limiting behavior of version 11. The error rate matches that of PfSense, which we tested in our GNS3 lab. Similarly, we also checked the rate limit of NetBSD version 8.2. For NetBSD we find overlapping behavior with FreeBSD. We classify this fingerprint similar to a multi vendor fingerprint as *FreeBSD/NetBSD*.

4 Measuring Internet Outages

This chapter is covered by the third publication, which presents longitudinal full-block scans to detect Internet outages in Ukraine.

Title	Tracking Internet Disruptions in Ukraine: Three Years of Active Full Block Scans.
Authors	Florian Holzbauer, Sebastian Strobl and Johanna Ullrich
Publication Status	This paper is included in the Proceedings of the 2025 ACM on Internet Measurement Conference, Pages 474-492. CORE2023-Ranking: A Acceptance Rate (Long Paper): 17.44%
DOI	`https://doi.org/10.1145/3730567.3764449`
Author Contributions	Florian Holzbauer: Is the first and main author of this publication. He is responsible for the measurement design, execution, evaluation and paper writing. Sebastian Strobl: Experiments on storage solutions and database schemas to store responsiveness data. Johanna Ullrich: Supervision and paper writing.
Artifacts	Data: `https://countrymonitor.github.io`
Reference	[HSU25]

Tracking Internet Disruptions in Ukraine: Insights from Three Years of Active Full Block Scans

Florian Holzbauer
Faculty of Computer Science
Doctoral School Computer Science
University of Vienna
Vienna, Austria
florian.holzbauer@univie.ac.at

Sebastian Strobl
SBA Research
Vienna, Austria
sstrobl@sba-research.org

Johanna Ullrich
University of Vienna
Vienna, Austria
johanna.ullrich@univie.ac.at

Abstract

Numerous disruptions to Internet access have been reported during the war in Ukraine, including large-scale outages, damage to network infrastructure, surveillance, and censorship measures. However, most observations rely on local reports or monitoring systems within Ukraine. In this paper, we investigate whether the conflict's impact on Internet connectivity can be observed externally, from a vantage point outside Ukraine. Focusing on the Kherson region, which has remained on the frontline for over three years, we conduct an active measurement campaign probing the Ukrainian address space at two-hour intervals since March 2, 2022, the 7th day of the invasion, resulting in a country-wide dataset that spans the full duration of the conflict. Extending existing outage detection approaches, we infer three signals to detect Internet disruptions and refine the mapping of ASes and address blocks to specific regions. This allows us to assign disruptions to oblasts with greater confidence. Our results demonstrate that Internet disruptions caused by the war can be measured remotely by any host connected to the Internet. Our analysis provides new insights into the resilience of small regional providers and identifies periods when Ukraine's Internet infrastructure was under significant strain.

CCS Concepts

• **Networks** → **Network measurement**; **Public Internet**.

Keywords

Outage Detection; ICMP; Full Block Scans; Ukraine; Kherson

ACM Reference Format:
Florian Holzbauer, Sebastian Strobl, and Johanna Ullrich. 2025. Tracking Internet Disruptions in Ukraine: Insights from Three Years of Active Full Block Scans. In *Proceedings of the 2025 ACM Internet Measurement Conference (IMC '25), October 28–31, 2025, Madison, WI, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3730567.3764449

1 Introduction

The Internet is a critical infrastructure for communication, both in everyday life and during crises. It enables people to stay in touch with family, friends, and colleagues while also serving as a primary medium for accessing information, such as news and governmental advisories. Since the war in 2022, Internet connectivity in Ukraine has been repeatedly disrupted and has undergone continuous efforts of restoration.

Several analyses have examined the impact of war on Internet connectivity in Ukraine. Many of these studies [19, 26, 30, 31] require connection initiation from within Ukraine. In contrast, active measurement campaigns that send probes to IP addresses enable data collection from outside the country. However, existing outage detection platforms, such as IODA [17], rely on Trinocular [36], which uses a limited set of representative IP addresses per /24 block.

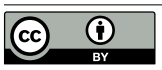
In this work, we extend existing research on full-block [3, 4] scans by conducting our own active measurements of the Ukrainian address space. Thereby, we capture the full block state directly by probing the entire address space using ICMP, rather than inferring it from the sampled Trinocular data. This approach has several advantages: we collect the full block state every two hours, and probing every IP allows us to introduce an additional outage signal based on responsive IPs to also detect partial outages.

A central challenge, also encountered in related work, is the attribution of outages to specific regions. In §4, we address this by leveraging long-term trends in IP geolocation to improve confidence in block-level location assignments. This allows us to better enumerate Internet disruptions at the regional level, allowing us to distinguish between outages in frontline and non-frontline areas. Together, our methodology and dataset offer improved insights into Internet disruptions, particularly in countries with high address churn, such as Ukraine. In summary, we make the following contributions:

Unique Full Block Dataset (§2,§3) We collected responsiveness and round-trip-time data for the entire Ukrainian address space. While the advantage of probing all addresses over sampling for Internet outage detection has been demonstrated in small-scale case studies before, it is the first time that it is applied to a country at war over a period of three years, showcasing its relevance. Access to the dataset can be requested at `https://countrymonitor.github.io`; it is provided for research purposes only.

IP Address Churn (§4) We discover churn of IP addresses across the different regions in Ukraine. Particularly, IP addresses leave frontline regions at a faster pace than other regions. This motivates the improvement of selection strategies to identify IP addresses that are representative of individual ASes or regions.

Regional Evaluation (§4) We refine the assignment of ASes and address blocks from the national to the regional level, enabling



IMC '25, Madison, WI, USA

ACM ISBN 979-8-4007-1860-1/2025/10
https://doi.org/10.1145/3730567.3764449

more localized analysis of measurement data, as demonstrated for Kherson Oblast. Among 118 ASes with IPs in the region, 34 operate regional blocks, making their outages more representative. We validate this classification against IPInfo's geolocation confidence metric, finding that regional blocks generally exhibit higher geolocation precision.

Internet Disruptions (§5) Based on three outage signals inferred from our collected data, we derive periods where Internet access was disrupted in Ukraine and specifically Kherson Oblast. Our approach is able to detect Internet outages across a larger set of providers than previous work, particularly small providers. We find that during Winter 2022/23 and 2024/25, Internet disruptions were widespread across Ukraine. During the remaining time periods, outages are specific to frontline regions.

Verification of Results (§5) We verified our results on Internet outages with regard to multiple aspects. For large ASes, our results correlate with those of Trinocular. Beyond that, we were able to verify our list of regional ASes in Kherson Oblast with a regional administrator, and we were also able to relate their outages to reported events such as cable cuts, the destruction of a dam or the seizure of infrastructure. Finally, we show that outages in non-frontline regions strongly correlate with power outages.

2 Background and Related Work

In this section, we provide background on Ukraine and reported Internet disruptions caused by the full-scale invasion. Then, we compare previous Internet measurement studies on Ukraine during wartime.

2.1 Ukraine and Verified Internet Outages

Ukraine is a country in Eastern Europe and is divided into 24 oblasts, two cities with special status, and an autonomous region. We refer to all these entities as regions or oblasts, irrespective of the detailed administrative differences, and use them interchangeably. The country's capital, Kyiv, is a city with special status and surrounded by an oblast of the same name, which we consider to be a single region for the purpose of our work, resulting in a total of 26 regions in our analysis. On February 24th, 2022, neighboring Russia started a full-scale invasion of the country. The initial military advance on the capital failed in April 2022, and Kherson Oblast was partly liberated again in November 2022. Since then, the frontline is practically stable.

Frontline and Non-Frontline Regions We consequently differentiate between frontline regions, i.e., oblasts marking the border between Ukrainian and Russian troops and experiencing continuous war actions since 2022, and non-frontline regions. Frontline regions are the oblasts of Chernihiv, Donetsk, Kharkiv, Kherson, Luhansk, Sumy, and Zaporizhzhia. All other oblasts are considered to be non-frontline regions. The latter also includes Kyiv and Mykolaiv, which experienced active combat only during the initial advance at the full-scale invasion's beginning.

Kherson Oblast as an Example Region We investigate Internet outages at the regional level, and Kherson Oblast, connecting Crimea with the Ukrainian mainland, serves as the example region for our work. We chose it as it was fully occupied by Russia in 2022 and partially liberated again in the same year. Since then, the Dnipro River marks the frontline – the right bank is controlled by Ukraine, the left bank by Russia. Its administrative center, Kherson city – a port city with 280,000 inhabitants before the war – resides on the liberated right bank. As of early 2025, only 70,000 are estimated to live in the city [16, 44].

Internet Disruptions affecting Kherson Oblast According to reports, the region's Internet connection was disrupted and reconstructed multiple times as a consequence of kinetic warfare. The following timeline provides an overview of the events. In this work, we are able to verify these reported events based on our detected outages (indicated with ☑), see Section 5 for our detailed results. For some of the events, we also provide additional insights beyond the reports (indicated with ⚙).

- ⚙ **March–May 2022**: Russian troops searched ISP offices in Kherson and seized infrastructure [8]. In collaboration with a local ISP (Status), we verified an Internet outage caused by the seizure.
- ☑ **April 30, 2022**: An oblast-wide outage occurred due to damage to the last functioning backbone cable [20, 33]. Our dataset allowed us to pinpoint 24 active ASes that were affected by this incident.
- ☑ **May–November 2022**: Internet from Kherson was routed over Russian upstream, leading to higher RTTs [27, 41]. We confirm these RTT increases for regional ASes, and additionally observe several disconnections of non-regional ASes for addresses regional to Kherson Oblast.
- ☑ **November 11, 2022**: Ukrainian forces liberated Kherson city and its surroundings [23]. Based on our contact with the regional ISP Status, we track this event at the granularity of address blocks, revealing a ten-day outage followed by gradual service restoration for their address blocks in Kherson.
- ⚙ **June 6, 2023**: The Kakhovka dam was destroyed [24]. While only the outage of a single AS is documented [35], we show that the resulting flooding had a broader impact, with a timely disruption visible for Viner Telecom, Digicom, and TLC-K.
- ⚙ **June/July 2024 and Winter 2024/25**: Airstrikes on energy infrastructure caused electricity outages [11]. We find a strong correlation between Internet outages and power outages in non-frontline regions, suggesting that they primarily arise from a lack of electricity in these regions.

2.2 Existing Measurements Ukraine

Several Internet measurement studies attempted to gain insight into Ukraine and the consequences of warfare on the country.

Passive Measurements Cloudflare monitored HTTP request volumes in its data centers and detected rolling power outages after targeted strikes against Ukrainian power infrastructure [2, 6]. Other studies analyzed (a) BGP data from route collectors to identify periods of unreachability of Ukrainian networks [26] as well as to quantify Crimea's dependency on certain, predominantly state-operated Russian ASes [14], (b) results from Measurement Lab's Network Diagnostic Tool (NDT), initiated by Ukrainian users, to detect the war's impact on Internet performance [19], and (c) web

Dataset	Singla et al. [42]	Klick et al. [22]	IODA/Trinocular [17]	This Work	Cloudflare [2]
Measurement Type	active	active	active	active	passive
IP/Block-based	IP	IP	/24	/24	IP
Protocols	DNP3, Modbus	60+	ICMP	ICMP	HTTP, DNS
Vantage Points	1	>1	approx. 20	1	330 cities [5]
Measurement Interval	24 hours	4 hours	10 min	2 hours	<1 min
Probes per /24 Block	256	up to 256	up to 15	256	-
Block Eligibility	-	-	$E(b) \geq 15$ & $A \geq 0.1$	$E(b) \geq 3$	-
Geolocation Confidence	Low	High	Low	High	Moderate
Target Set	UA delegated	400K static IPs	IPv4-wide	UA delegated	UA clients
Avg. Responsive IPs	435K	-	-	1.5M	-
War Period Coverage	6 Months in 2022	Until March 2023	Since 2022	Since 2022	Since 2022

$E(b)$ refers to the number of ever-active addresses in a /24 address block b, $A(E(b))$ to the long-term probability that the ever-active addresses reply.

Table 1: Comparison of methods used for Internet outage detection, with a focus on Ukraine. We compare four active measurement approaches, including this work, and one passive method (Cloudflare).

analytics data such as those provided by Google and Cloudflare, again to assess Internet performance as well as to trace the flow of refugees to Ukraine's neighboring countries [30, 31]. Passive measurements analyze already available Internet traffic instead of generating additional data transmission, but require a privileged position, e.g., in a content delivery network or at a BGP collector, to observe sufficient amounts of traffic in the first place.

Active Measurements As an alternative, active measurements might be run from vantage points in the affected regions. Ukrainian nodes that are part of the RIPE Atlas platform are an option, e.g., to assess round-trip times between Ukraine and Russia [26]. These measurements, however, require the setup of vantage points in the region of interest and notably limit scalability. RIPE Atlas operates about 200 nodes in Ukraine [31], covering a limited number of geographic locations. In contrast, it is more flexible and scalable to send requests from a vantage point, whether inside or outside of Ukraine, to the Ukrainian address space. IODA [7, 17, 36], Singla et al. [42], and our work follow this principle, see Table 1 for a comparison of active approaches with Cloudfare's passive one.

Detecting Internet Disruptions in Ukraine Table 1 compares existing approaches for measuring Internet disruptions in Ukraine. Singla et al. [42] probed the entire Ukrainian address space using industrial protocols (Modbus, DNP3), but only once every 24 hours and for six months in 2022. Klick et al. [22] focused on a set of 400k static IPs probed at an interval of four hours. By targeting static IP addresses, this approach reduced noise from dynamic address changes and increased confidence in assigning outages to regions. Similarly, we improve geolocation precision. Instead of probing only static, we evaluate long-term geolocation trends to assign blocks to regions with high confidence (§4).

The IODA platform applies the Trinocular method, probing up to 15 IPs per /24 block. By trading comprehensive probing for fewer addresses, it achieves the highest probing frequency among active approaches, with measurements every ten minutes [17, 36]. While block-based probing increases outage confidence over single-IP approaches [40], its reliance on few IPs can yield unstable results. Full-block scanning (FBS) addresses this by aggregating responses across rounds, reducing the eligibility threshold to three ever-active IPs ($E(b) \geq 3$), though it has only been evaluated in case studies and is not actively used by IODA [3, 4].

Cloudflare instead monitors traffic volumes passively [2], benefiting from wide vantage coverage and high temporal resolution, but relying on proprietary data. Our dataset complements these methods by actively probing 10.5M Ukrainian IPs, with about 1.5M responding per round, every two hours (§3). This is the first long-term study to apply FBS in practice, extending coverage and stability beyond other active approaches. Unlike passive measurements, this data can be collected by any host connected to the Internet.

3 Methodology

Only seven days after the beginning of the full-scale invasion of Ukraine, we began probing the entire Ukrainian IPv4 address space. We tried to minimize the impact on Ukrainian networks and state ethical considerations in Appendix A. By combining this active measurement data with external sources, we generate three distinct outage signals. We first describe our measurement setup and signal generation (§ 3.1). Since IP responsiveness alone is insufficient to reliably infer outages and assign them to regions, we then discuss the external datasets integrated into our analysis (§ 3.2).

3.1 Setup and Signals

Data Collection We probe all Ukrainian IPv4 addresses at a two-hour interval using ZMap [13] from our vantage point located in a European data center, approximately 1000km from Ukraine's capital Kyiv. The ICMP-based measurements started on March 2nd 2022, at 10 p.m., i.e., the 7th day of the full-scale invasion, and have been running since. For the work at hand, we analyze data from the beginning of our measurements to February 24th, 2025, the invasion's third anniversary.

Internet Availability Signals We combine our collected data with external datasets to generate three Internet availability signals, aggregated at the AS or regional level. The first two signals align with those of IODA [17], though the second signal *FBS* is generated from our results by comprehensively measuring the address space

Level	BGP ★	FBS ■	IPS ▲
AS	< 95%	< 80% (if IPS < 95%)	< 80%
Regional	< 95%	< 95% (if IPS < 95%)	< 90%

Table 2: Static Internet disruption detection thresholds relative to a seven-day moving average.

instead of sampling. We extend these signals by a third signal that is only feasible due to comprehensively probing the Ukrainian IP address space.

(1) *BGP* ★ provides the number of routed /24 address blocks per AS. As we also develop a method to assign ASes to a region, we are also able to generate this signal per region.
(2) *FBS* ■ provides the number of active /24 address blocks, again either grouped per AS or region, and is equivalent to the number of address blocks meeting the eligibility criteria of at least three ever-active addresses per month.
(3) *IPS* ▲ provides the number of responsive IP addresses per AS or region and enables us to also capture partial outages, i.e., decreases in IP activity while block reachability remains stable. We limit this signal to months where the average number of responsive IP addresses exceeds 10.

Signal Properties *FBS* ■ and *IPS* ▲ are based on the regional share of IPs in blocks classified as regional. To detect outages, we compare current values with the moving average of the previous week. Based on the level of aggregation, i.e., AS or region, we defined different thresholds for outages, see Table 2. The rationale is that more granular aggregations (e.g., ASes in comparison to regions) involve fewer entities (IPs, blocks). Consequently, they are assigned more relaxed thresholds to avoid false positives. Due to the sliding window, the moving average adapts to the new baseline after an outage, causing the outage criteria to no longer be met. To still capture such long-lasting outages, we add a flag to the *BGP* ★ signal – if no routed /24 is visible for the ASN or region, the outage period is considered ongoing, even after the moving average stabilizes. Finally, we also employ ISP availability sensing as proposed by Baltra et al. [3] to avoid false positives in the *FBS* ■ signal. This way, we filter out false positives caused by dynamic IP reallocations, rather than mistaking them for outages.

Limitation - Single Vantage Point. All measurements were conducted from a single vantage point located outside Ukraine. This design reflects both the urgency of setting up the monitoring infrastructure during the early stages of the war and the storage constraints of processing a more comprehensive *FBS* ■ signal than previous Trinocular signals. Previous campaigns did not reveal systematic limitations associated with this vantage point. However, as with any single-source measurement system, data is unavailable when the vantage point is offline. These outages occurred on the following dates and are indicated in the figures (March 6th-7th, 2022, March 14th-28th, 2022, October 12th-19th, 2022, March 5th - April 2nd, 2024, July 13th, 2024, August 7th-19th, 2024, and September 16th, 2024).

Limitation - Bi-Hourly Probing Interval To minimize measurement load on the Ukrainian Internet, we adopted a probing interval of two hours. This choice reduces potential network strain but limits our temporal resolution. Specifically, outages that begin and resolve between two consecutive probing windows may go undetected. Each probing session spans approximately 20 minutes, so the maximum undetected outage duration is bounded by the remaining 100 minutes between sessions. Other approaches, like Trinocular, measure at shorter intervals, e.g., 10 minutes, promising higher resolution. We quantify the limitation in Section 5.4 by enumerating outages between our probing intervals. However, they only probe a fraction of the addresses in each round and do not report outages for small regional providers that our exhaustive measurements are able to detect.

3.2 External Datasets

For our measurement study, we extend our probing data with external datasets. As an input for our ZMAP measurements, we relied on **RIPE Delegations** [38], which contain allocated and assigned IP address ranges, among others, for European countries. In the delegations files from December 14th, 2021 the most recent at the start of the invasion, we found 3,085 address ranges with a total of 10,508,960 addresses with Ukraine's country code UA. We also rely on **RouteViews BGP dumps** [45], which, like our measurement data, are available in two-hour intervals. These are used to create the Internet availability signal *BGP*★ based on the number of routed /24 blocks. Additionally, this data enables the aggregation of IP addresses per AS. For geolocation, we use the commercial service **IPInfo** [18] to assign IP addresses to geographic locations both on the regional and country level. We obtained access to the full database on the first day of each month. While an IP might be located in different geolocations over a given month, we focus on long-term trends by detecting providers that remain in the same region across multiple months. This approach reduces geolocation noise in the detection of outages.

To validate our findings, we use **IODA** [17], which detects and reports Internet outages in the IPv4 Internet. Unlike other platforms such as Netblocks [34] or Kentik [21], IODA provides access to raw signal and outage data [25]. We rely on their API to compare and validate outages detected for Ukraine. Finally, we compare outages with **Energy Map** data provided by the national power company Ukrenergo [32], which includes information on power grid outages affecting more than 50% of Ukraine's Oblasts between January 1st, 2023 and January 20th, 2025.

Limitation - Leased Prefixes and Churn Our measurements relied on a static snapshot of RIPE delegations from December 14th, 2021, which we used unaltered throughout the three-year period. However, RIPE notes that delegation data may not always reflect actual usage, particularly due to leased prefixes—i.e., address blocks registered in one country but used in another. At the start of our measurements, 350K addresses (3.3%) delegated to Ukraine were geolocated outside the country. To assess the reverse case, we relied on IP geolocation to reveal 773K addresses (7.4%) used in Ukraine but delegated to other countries. Assessing churn in delegated prefixes, 348 Ukrainian prefixes (12%) changed their registered country code, with one-third reallocated to Russia and the rest to various

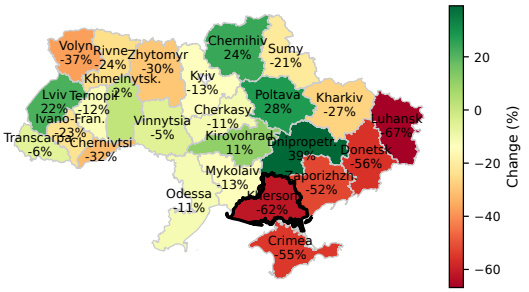


Figure 1: Relative change in IPv4 address counts per oblast (February 1, 2022 – February 1, 2025). IP addresses shift away from frontline regions to other parts of Ukraine or to foreign countries.

European countries and the U.S. We also observed a net decline in prefix allocations to Ukraine: only 198 new prefixes (7%) were assigned over three years (see Figure 18 in Appendix B). Based on these dynamics, we estimate RIPE delegations to consistently cover at least 93% of Ukraine's active address space during our measurements, making them a reliable baseline for IP-level monitoring.

4 Regional Classification

We refine the assignment of ASes and /24 address blocks from the national to the regional level using Ukraine's administrative divisions (oblasts). However, direct application of geolocation services reveals significant address churn, with IPs shifting between regions or countries (see §4.1). To avoid misclassifying regional outages, we introduce a stricter definition of regionality based on sustained address presence over time (§4.2). We then verify this classification (§4.3) and assess its impact on our dataset by analyzing address responsiveness (§4.4).

4.1 Regional Address Churn

We geolocated all probed IP addresses using IPinfo and compared their regional assignment from before the war (February 1, 2022) to three years later (February 1, 2025). Our results reveal substantial address churn, with many IPs moving between regions or leaving Ukraine entirely. This motivates a stricter definition of regionality for both ASes and address blocks.

Churn across Ukraine Figure 1 shows the relative change in IPv4 address counts per oblast. Nineteen of 26 regions saw declines, with the sharpest losses in frontline areas: Luhansk (-67%), Kherson (-62%), Donetsk (-56%), Zaporizhzhia (-52%), Kharkiv (-27%), and Sumy (-21%). Only Chernihiv recorded a net increase (+24%). Churn also occurred in non-frontline oblasts such as Volyn (-37%), Zhytomyr (-30%), and Rivne (-24%). Appendix C, Figure 20 shows increasing IPv6 adoption, which may help extend outage detection in sparse regions once suitable methods exist. Geolocation confidence, measured by IPInfo's radius metric (5 to 5,000 km, with increasing step widths), also declined—its median for Ukrainian IPs rose from 100 km in 2022 to 500 km thereafter. Of 3.73M IPs that changed location, 2.24M moved between Ukrainian regions, primarily driven by national ISPs like Ukrtelecom (697K), Kyivstar (341K), Vodafone (243K), and Vega (67K), reflecting dynamic address assignment. Another 1.5M addresses were geolocated abroad, mostly to the US (926K), Russia (110K), and Germany (60K). Notably, AS16509 (Amazon) now announces 519K of these, about one-third of the externally reassigned IPs.

Churn in the Kherson Region We also found this trend in Kherson. Of the 141K IP addresses initially geolocated to Kherson, only 36K (26%) remained there after three years. 63K (45%) moved to another Ukrainian oblast, and 41K (29%) were geolocated abroad. This includes 33K IPs previously held by Volia (AS25229), now announced by Amazon.

Key Takeaways

(1) Between 2022 and 2025, 3.7M IP addresses changed location, indicating substantial churn.
(2) Of these, 2.2M moved within Ukraine (mainly due to national ISPs), while 1.5M were reassigned abroad, primarily to Amazon, the US, or Russia.
(3) Frontline regions lost more IPs than non-frontline regions; in Kherson, only 26% of IPs remained.

4.2 Definition of Regionality

Address churn motivates a stricter definition of regionality in favor of reducing the distortion of our results. Therefore, we aggregate IP addresses to AS or /24 address blocks, respectively, and decide on their regionality depending on their share of addresses in a region over time. Based on our definition, we define them as regional, i.e., primarily operating in a single region, or non-regional, i.e., operating in multiple regions.

Formal Definition. Let E_{total} denote the set of entities, either ASes or /24 address blocks, with at least one geolocated IP address in the investigated region over the observed period T. For each entity $e \in E_{\text{total}}$ at time t, we define the share $s_t(e) = \frac{n_t(e)}{N(e)}$, where $n_t(e)$ is the number of geolocated IPs in the region and $N(e)$ is its maximum possible number of addresses (for ASes their addresses in Ukraine; for /24 blocks $N(e) = 256$). We classify entity e as regional for a region if its share meets the threshold M in at least T_{perc} of routed months T_{routed}. As an example, Figure 2 shows a /24 block classified as regional.

$$E_{\text{reg}} = \left\{ e \in E_{\text{total}} \;\middle|\; \sum_{t=1}^{T_{\text{routed}}} \mathbf{1}\big(s_t(e) \geq M\big) \;\geq\; \left\lfloor T_{\text{perc}} \cdot T_{\text{routed}} \right\rceil \right\}$$

To assess the sensitivity of our classification for different parameter choices, we vary M and T_{perc} from 0.1 to 1 in steps of 0.1, see Figures 22 and 23 in Appendix D on the resulting numbers of regional ASes and blocks. For the remainder of this work, we selected thresholds of $M = 0.7$ and $T_{\text{perc}} = 0.7$.

We illustrate the impact of our parameter choices by two ISPs in Kherson. With a strict threshold of $M = 0.9$ and $T_{\text{perc}} = 0.9$, *ISP Status* would be classified as non-regional as one of its four /24 subblocks is located in Kyiv. By contrast, relaxed thresholds

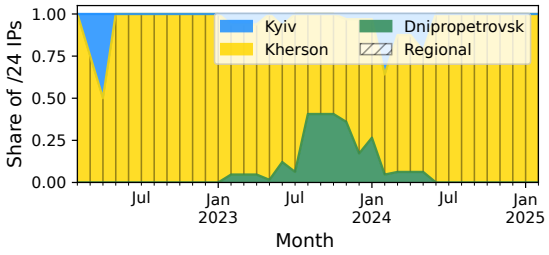


Figure 2: The exemplary /24 block 176.8.28 belonging to Kyivstar meets the regional threshold of M=0.7 in more than 70% of routed months in Kherson.

	Ukraine			Kherson		
Category	**ASes**	**$\overline{IPS}$**	**$\overline{/24s}$**	**ASes**	**$\overline{IPS}$**	**$\overline{/24s}$**
Total	2024	8.99M	35.2K	118	73.8K	512
Reg.	1428	4.17M	16.7K	13	8.3K	33
Non-Reg.	484	4.80M	19.4K	40	64.9K	465
Temporal	112	17.9K	313	65	596	15
Target Set	1773	7.15M	28.5K	34	41.7K	168

Table 3: Classification of regional, non-regional, and temporal ASes, including average monthly counts of ASes, IP addresses, and /24 blocks in Ukraine and Kherson (2022–2025). The final row highlights the target set: Regional and non-regional ASes with regional /24 blocks that are suitable for outage detection.

of 0.5 would classify providers operating across multiple oblasts, such as *Digicom*, as regional. Our chosen tradeoff of 0.7 balances these extremes: it permits a share of IPs to be operated outside the main oblast, while still capturing the regional nature of the provider. The same rationale applies at the block level. However, in contrast to ASes, /24s pointing to multiple locations and oblasts are less prominent. From the total of 35.2K /24s we find a mean of 78% pointing to a single location during a given time. This number increasees to 86% on the oblast level. Figure 21 in Appendix D shows that for multi-local /24s there is usually a dominant share that points to one region. We aim to detect temporal assignments as an additional filter on non-regional ASes, as temporal geolocation assignments (a few IPs, only one month) are likely caused by noise in geolocation and are not valuable measurement targets. We define non-regional ASes as temporal if they fail to reach a certain number of IPs in the target region (< 256, equaling one /24) or the regional share exceeds 10% for at least one month.

Separate Classification of ASes and Blocks ASes, both regional and non-regional, might encompass regional and non-regional address blocks. The blocks' classification has two advantages. First, regional ASes might predominantly serve one region, but often also other (neighboring) regions to a certain extent. Exclusion of the non-regional blocks improves the result's significance for the region

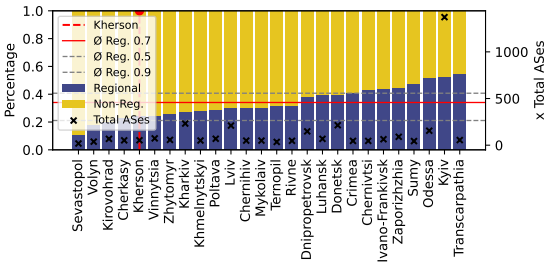


Figure 3: Regional classification helps to identify important ASes per region. Regional ASes account on average for 34% of the ASes with at least one address geolocated to a region. From 118 ASes in Kherson, we identified 40 non-regional and 13 regional ones, while 65 only show temporary presence in the region.

of interest. For example, AS25482 (Status) is regional in Kherson. Three of its four/24 blocks meet the regionality criteria; the fourth is regional in Kyiv and would distort our results for Kherson. Second, non-regional ASes like national ISPs might contain regional blocks, providing insight into a region. From AS15895 (Kyivstar) 299 /24s located once in the Kherson region, of which only 52 are regional.

Mitigated Noise to Outage Attribution There is inherent noise in geolocation due to the mobility of IP addresses and the dynamic usage of address blocks. Introducing regional classification to outage attribution, we mitigate these common scenarios. While this approach still relies on IP geolocation, it does so only on the regional level and not on the more granular city level. By identifying and excluding dynamic blocks, regional dependency is limited to long-term stable blocks. Summarizing, we mitigate the following scenarios: *(II) IP drift:* If one or more IP addresses of a /24 temporarily geolocate to a different region, our outage detection approach only includes the IPs belonging to the regional part of the block. *(II) Block drift:* If a /24 block geolocates to another subdivision for a limited period, we do not attribute outages to this block. *(III) Regional churn:* If multiple IP addresses and blocks leave a region and are somewhere else, the block is entirely classified as non-regional and excluded from outage detection. If it still passes the threshold, it is evaluated only during the months in which it is considered to be regional. Finally, our approach cannot prevent *systematic misattribution*, i.e., IPInfo assigning the block to an incorrect region for ≥ 70% of months across the 3-year window.

Key Takeaways

(1) We classify ASes and /24 blocks as *regional* or *non-regional* depending on long-term trends in geolocation.
(2) Regional classification mitigates noise from IP and block shifts, reducing misattribution in outage detection.
(3) Separating AS and block classification improves precision: regional ASes may still contain some non-regional blocks, and non-regional ASes may include regional blocks.

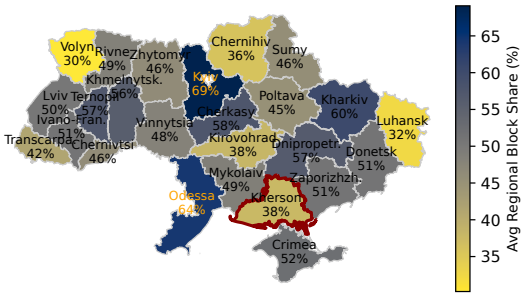


Figure 4: Share of regional /24 address blocks. On average, 50% of the blocks with at least one address geolocated to the region are classified as regional.

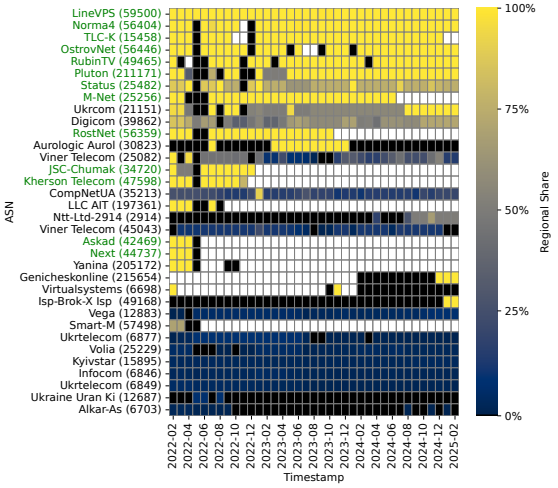


Figure 5: ASes with regional /24 address blocks in Kherson. Periods without BGP visibility are indicated in white.

4.3 Evaluation

According to our definition, we classified ASes and /24 address blocks into regional and non-regional.

Regional Classification Ukraine Table 3 shows that 1428 ASes (serving 16.7K /24 blocks) are regional for at least one of the 26 oblasts, while 484 are non-regional and 112 are temporal. Comparing IPInfos' confidence metric reveals a notable geolocation gap: IPs from non-regional /24s show a stable median radius of 500km across years, whereas regional /24s show higher precision, 50km in 2022, increasing to 200km by 2025. This reflects more accurate geolocation of regional networks, often tied to fixed sites like data centers or government offices, unlike mobile or carrier IPs [1]. Prior work supports the high prevalence of regional networks, noting Ukraine's unusually fragmented Internet infrastructure [12]. Figure 3 shows the number of regional ASes per oblast: 46% of ASes with at least one geolocated address are eventually classified as regional. However, some ASes are regional in one oblast and non-regional in others, leading to lower regional shares than the totals in Table 3. Figure 4 displays the share of regional blocks. This share ranges from 69% in Kyiv to 30% in Volyn. In total, we classify 28.5K /24 blocks, covering 7.15M IPs and 1773 ASes, as regional (Table 3), making them valuable for detecting regional outages. Temporal ASes account for 112 (5.5%) nationally, but this rises sharply in Kherson, where 118 ASes have at least one IP, and 65 (55%) appear only temporarily.

Regional Classification: Kherson We identified 34 ASes with regional blocks in Kherson, as summarized in Table 3. Figure 5 presents these ASes ordered by their regional share of IP addresses, with higher shares indicating stronger association with Kherson. The figure shows a clear visual distinction between regional and non-regional ASes: regional providers appear at the top, while non-regional ones are concentrated at the bottom. It also highlights ASes that were active during only part of the measurement period. These are accurately captured by our method and appear with white gaps, indicating that the AS was no longer BGP-routed at those times. In total, seven ASes show discontinued service: `15458`, `25256`, `56359`, `34720`, `47598`, `42469`, and `44737`.

Verification in Kherson We contacted a local ISP and one of their administrators named the active regional providers in Kherson city and its surroundings, i.e., the liberated area on the right bank. This way, we verified that the ASes classified by our approach are indeed regional for Kherson oblast. We missed two providers in Kherson city, namely `AS42782` (Stream Kherson), and `AS39667` (Online Net). Their addresses are leased from AlfaTelecom, thus attributed to the Czech Republic in the RIPE delegations files, and eventually not considered by our input data set, see limitations.

Key Takeaways

(1) In Ukraine, we identify 1,428 regional ASes and 28.5K regional /24 blocks (7.15M IPs).
(2) Regional blocks show higher geolocation precision (median radius 50 km in 2022, 200 km in 2025) than non-regional ones (stable at 500 km).
(3) In Kherson, classification distinguishes 13 regional and 40 non-regional ASes, validated with a local ISP.

4.4 Regional Responsiveness

Outage detection on the regional level requires limitation to regional IP addresses as defined in the previous subsections, but also reduces the number of responses from our dataset, collected in the Internet measurements, for analysis. Consequently, we examine the responsiveness of regional blocks in Ukraine.

Responsiveness of Regional IPs Probing 10.5M Ukrainian IP addresses, we received on average 1.47M replies in 2022. By 2025, this number dropped to 1.21 million replies, a reduction by −18%. Regional IPs accounted for 1.31M in 2022 (89% of all responses) and 1M in 2025 (83% of all responses), i.e., regional addresses return more responses than non-regional ones. Figure 6 provides absolute and relative numbers of responsive IPs within regional blocks per region.

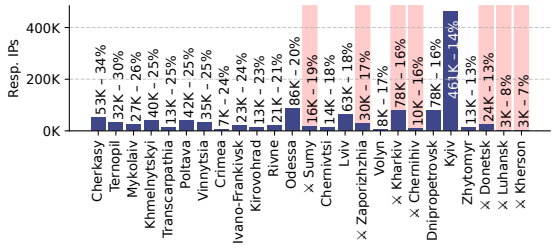


Figure 6: Share of responsive IP addresses per oblast. Labels indicate the average number and percentage of responsive IPs among all regional IPs in each oblast. Frontline regions are marked in red.

Frontline oblasts show lower responsiveness with the lowest share in Kherson oblast – from 41.7K IP addresses in regional blocks, 4.5K (10.7%) were responsive in 2022 and 1.4K (3.4%) in 2025, impacting the eligibility of blocks for outage detection.

Filtering Measurable Blocks We build on the established method of full block scans [4] for outage analysis that requires at least three ever-active IPs per block and month ($E(b) \geq 3$) as eligibility criteria. From the 21.4K responsive /24 address blocks in our measurement data set, 20.4K, on average, meet this criterion. Figure 7 shows the distribution of measurable blocks across the different regions and highlights changes observed between 2022 and 2025. In frontline regions, we observe a strong correlation with recorded IP churn. Although the overall number of responsive IPs has decreased and most blocks are now concentrated in the capital, Kyiv, measurable blocks remain present in every oblast as of 2025. In Kherson, 1,400 responsive IPs are observed across 89 regional /24 blocks. These blocks form the basis for our outage detection in the region.

Comparing Block Eligibility to Trinocular. Table 4 compares the number of eligible blocks for full block scans, as done in our work, with Trinocular [36] for all regions of Ukraine. Trinocular poses stricter block eligibility criteria ($E(b) \geq 15 \wedge A \geq 0.1$). Yet, with 18.1K eligible blocks, the number of eligible blocks remains comparable. However, this number needs contextualization, considering Trinocular's known limitations: 4K blocks exhibit indeterminate belief ($A < 0.3$) [4], i.e., are more likely not to lead to a definitive belief whether the block is up or down. Additionally, Richter et al. [37] decided to exclude sparse blocks with five or more outages in three months further as they have shown fluctuating results. Consequently, many blocks might have to be filtered despite their initial eligibility (see §5.4 for a comparison).

Key Takeaways

(1) Despite an overall decline in replies (–18% from 2022 to 2025), regional IPs consistently respond more often than non-regional ones; frontline oblasts show the lowest responsiveness, with Kherson at the bottom.
(2) In 2025, all oblasts still show responsive eligible blocks.
(3) Compared to Trinocular, FBS preserves a higher number of eligible blocks and avoids indeterminate belief.

Category	Regional		Non-Regional	
All Regional Blocks	28,458	100.0%	10,650	100.0%
Responsive	21,542	76.0%	5,993	56.0%
-> Full Block Scans [4]	20,603	96.0%	5,628	94.0%
-> Trinocular [36]	18,138	84.0%	4,314	72.0%
Thereof Indetermined [36]	4,376	24.0%	2,549	59.0%

Table 4: Eligible blocks comparing regional to the filtered non-regional for outage detection.

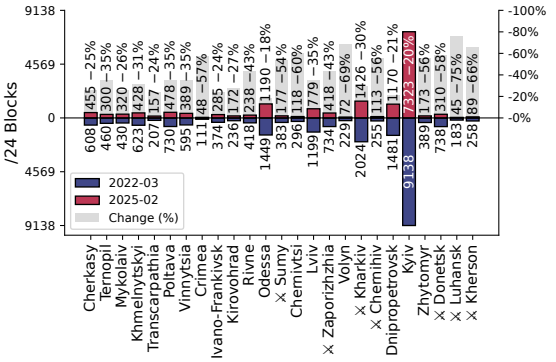


Figure 7: Comparison of responsive /24 address blocks (2022-03 vs. 2025-02).

5 Internet Disruptions

For a set of 1,773 ASes with regional address blocks (see Table 3) we analyze Internet outages across Ukraine, gradually increasing the granularity of our analysis. We start with a national view, finding that outages in non-frontline areas are largely driven by electricity outages (§ 5.1). We then focus on three key events in Kherson–the Mykolaiv cable damage, traffic rerouting during Russian occupation, and the Kakhovka dam destruction–highlighting their distinct impacts on distinct ASes in the region (§ 5.2). Through direct communication with Status AS, a local ISP, we confirmed visibility in provider-level outages (§ 5.3). Finally, we compare our findings with IODA data, showing broader outage coverage, especially for smaller ASes (§ 5.4).

5.1 Disruptions in Ukraine

Figure 8 presents the Internet outages per Ukrainian oblasts separated by the three signals – BGP reachability (*BGP* ★), active /24 blocks (*FBS* ■), and responsive IPs (*IPS* ▲) – each detecting different types of outages. For two periods, we observe a decline in responsive IPs affecting practically the entire Ukraine, while the number of active /24 blocks remains stable. This pattern persists even when applying thresholds as high as 99% for block activity, suggesting that blocks remained active despite lower numbers of responsive IPs. The figure further shows that most outages are not detectable through BGP alone. Instead, the majority are revealed by the FBS and IP responsiveness signals. In contrast, the IODA-based Figure 25 (replicated in Appendix G) portrays a different picture,

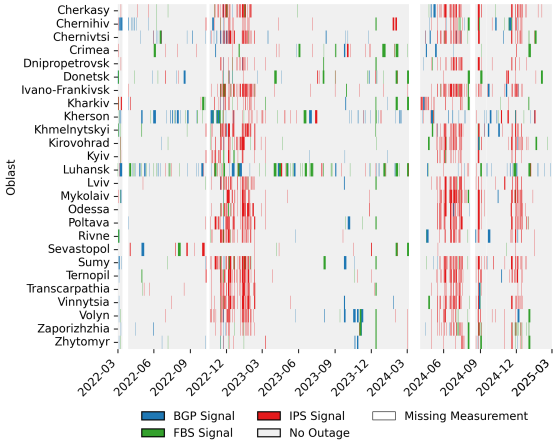


Figure 8: Internet disruptions detected by region. The three outage signals on routed /24 address blocks *BGP* ★, active blocks *FBS* ■, and responsive IPs *IPS* ▲ detect different outages.

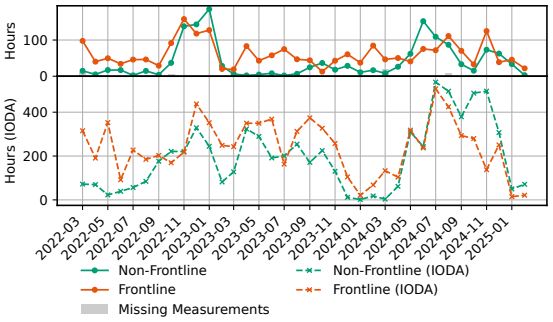


Figure 9: Monthly aggregated hours affected by Internet outages, comparing our measurements (top) with estimates based on Trinocular data from IODA (bottom).

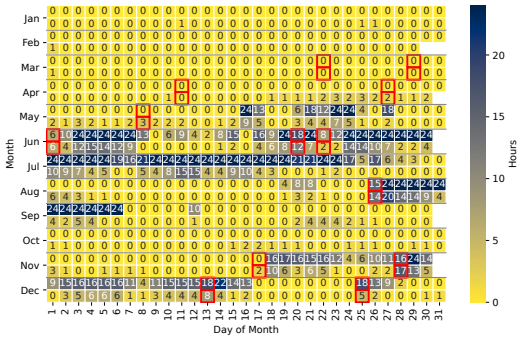


Figure 10: 1 Average hours of power and Internet outages per day for non-frontline regions (2024). Power outages as reported by Ukrainian power grid operator Ukrenergo [32] (top row) and Internet disruptions detected in this work (bottom row) correspond with each other. Days marked in red correspond to attacks on the power grid documented by [11].

dominated by long-term losses in BGP visibility across oblasts. We attribute this difference to the absence of regional classification in IODA's data model. Since IODA maps both regional and non-regional ASes to oblasts, BGP outages affecting large, non-regional providers can manifest as simultaneous outages across multiple regions.

Frontline vs. Non-Frontline Regions In Figure 8, oblasts at the frontline, e.g., Kherson or Luhansk, experience recurring outages throughout the entire three-year measurement period. Non-frontline regions are primarily affected by outages during the winter months of 2022/23, and again in 2024/25. Figure 9 presents the average number of Internet outages per month separately for frontline and non-frontline regions. The figure also compares our results with those from IODA. For IODA, the non-frontline regions resemble the pattern from the frontline regions, suggesting that their ability to attribute outages to individual regions might be affected by IP churn. Beyond, IODA reports more hours of downtime. In some months, they account for 450 hours that would be equivalent to 63% downtime. IODA's higher outage hours appear to stem from long-term BGP visibility losses, which inflate the total hours of reported downtime.

Disruptions in Non-Frontline Regions The outages detected in the winter months of 2022/23 and again in 2024/25 appear to affect all oblasts, see again Figure 8. A closer look, however, reveals that Crimea and Sevastopol, which are also on the Crimean peninsula, did not experience these outages. Both are occupied by Russia since 2014, and unlike the other Ukrainian regions connected to the Russian power grid [43].

1 ***June, July, and Winter 2024*** In Figure 10, we consequently compare the Internet disruptions, as detected by our data, with the periods of electricity outages reported by Ukrenergo, the Ukrainian power grid operator [32]. This data is only available from 2023 onward and shows a clear increase in 2024 with 1951 hours without electricity. In the figure, we also indicate 13 dates of confirmed large-scale attacks on Ukraine's energy infrastructure as reported by Dixigroup [11]. Comparing the number of Internet and power outages, we detected on average 686 hours with Internet disruptions; the latter were typically shorter, possibly due to backup power solutions in place and power disruptions not affecting all oblasts equally. When looking at the worst case scenario, the maximum outage hours per day in which any of the oblast is affected, we detect up to 2,822 hours, including days where no power outages were reported by Ukrenergo. For instance, Kyivstar can maintain mobile services for up to four hours without electricity, and its core infrastructure even longer [10].

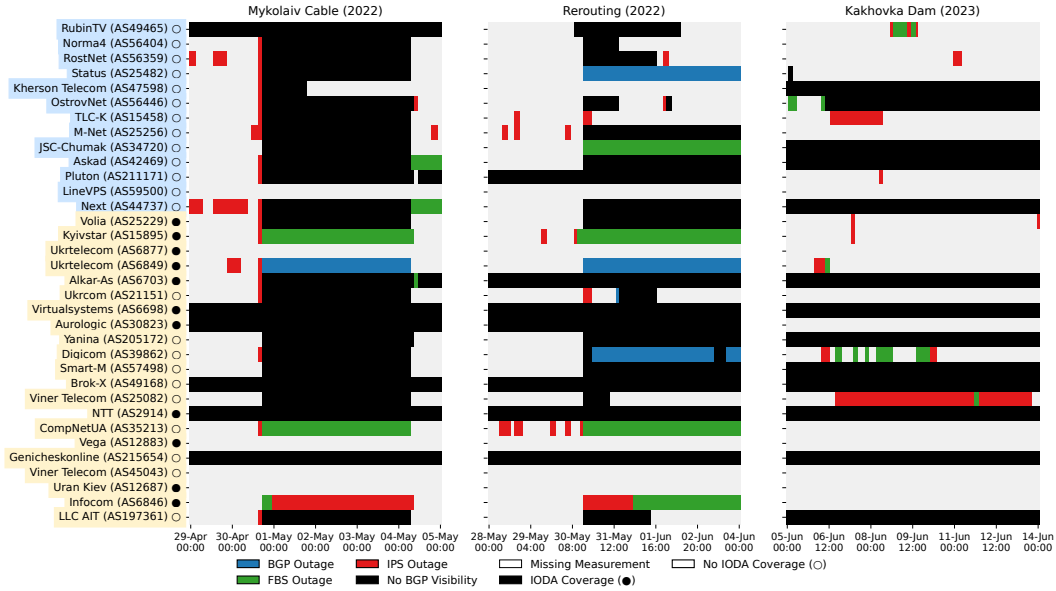


Figure 11: Internet disruptions recorded by any of the three signals for ASes in Kherson for three events to validate regional outages.

With a Pearson coefficient of $r = 0.725$, there is indeed a strong positive correlation between the hours of Internet and power outages in non-frontline regions. In comparison, frontline oblasts yield lower correlations of $r = 0.298$. This indicates that in frontline regions, Internet outages are less directly tied to power outages and are rather caused by other factors such as damage to the network infrastructure. Replicating this analysis with IODA data, see Figure 26 (Appendix G), shows weak correlation, both for non-frontline ($r = 0.328$) and frontline ($r = 0.394$) regions. Moreover, the similar correlation values observed in the IODA dataset for both frontline and non-frontline regions might again be a consequence of IODA being unable to precisely distinguish between the two classes.

5.2 Disruptions in Kherson

In this subsection, we now shift focus to the AS level, analyzing ASes in the region of Kherson; one of the seven frontline regions. We show that our data covers three major Internet disruptions, namely the damage to the Mykolaiv cable, traffic rerouting in Russian-occupied regions, and the destruction of the Kakhovka dam. AS-level analysis introduces additional challenges. Specifically, the *IPS* ▲ signal might be unreliable for ASes with only a few responsive IPs. Consequently, we only consider this signal for months in which the average of responsive IPs exceeds 10. This excluded Digicom (2 months), Infocom (2), Ukrtelecom (2) and Genicheskonline (1).

Timeline of Observed Disruptions Figure 11 shows how our data covers the three events, distinguishing regional ASes in blue from non-regional ones in yellow. Six (Mykolaiv cable), seven (traffic rerouting), and twelve (Kakhovka dam) ASes were already invisible before the events – indicated by black bars in the figure – we only attribute a disruption if BGP visibility was lost after the event. A complete version of the figure, covering the entire three years of our measurements, as well as a table with information on the number of analyzed regional /24 address blocks and headquarter per AS is provided in Appendix D (see Figure 28, Table 5). Valid outage signals were recorded for 30 out of 34 ASes, indicating high responsiveness, even among ASes with only a single /24 block. Notably, IODA only reports outages for non-regional ASes, emphasizing its limited coverage in the Kherson region.

2 ***April 30, 2022*** Kherson experienced a three-day oblast-wide Internet outage due to damage to the last backbone cable connecting the city [20, 33]. Our outage signal aligns clearly with the timing of the incident. Initially, it caused a drop in responsive IPs and eventually resulted in a loss of BGP visibility for 24 ASes. Most ASes recovered after three days, with the exception of Pluton and Alkar remaining offline afterwards.

3 ***May – November, 2022*** As a consequence of Russian occupation, Internet traffic was rerouted through Russian upstream providers and *mir-telekom* was introduced as a mobile network operator [9, 28]. Cloudflare reported rerouting for 15 ASes by identifying Russian ASes on the BGP paths of networks in Kherson. Kentik [27] confirmed this by increased round-trip times (RTTs) for affected ASes. Based on our measurement data, figure 11 shows that 21 ASes experienced outages during this period. While regional ASes

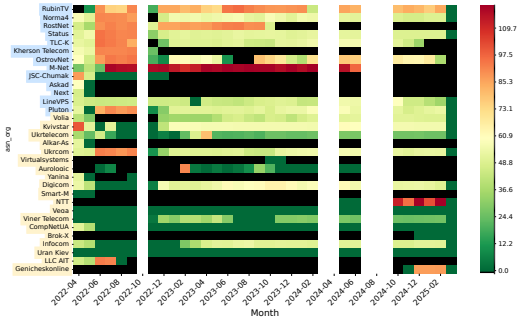


Figure 12: 3 Average monthly Round-Trip-Time (RTT) of ASes in Kherson.

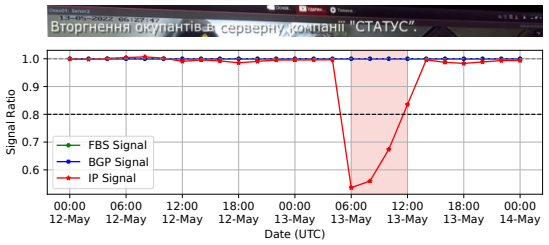


Figure 13: 5 Outage signals for Status (AS25482) on May 12th to May 14th with an incident recorded on May 13th, 06:28. Our dataset verifies that the action in the footage led to a disruption for the service provider.

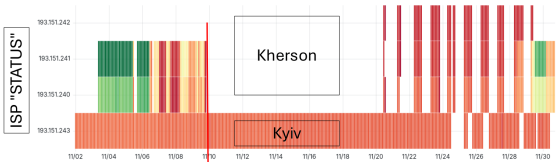


Figure 14: 6 *IPS* ▲ signal for each of the four /24 blocks of Status. The ISP went offline during the Russian retreat from Kherson city and came back online 10 days later using an emergency power supply.

were only temporarily affected and regained BGP visibility later, that was less prevalent for non-regional ASes, see the figure's complete version in the Appendix. Results from our RTT measurements, see Figure 12, confirm increased delays for the regional providers RubinTV, Norma4, Rostnet, Status, TLC-K, Kherson Telecom, OstrovNet, M-Net. For three ASes, namely RubinTV, RostNet, and M-NET, these elevated RTTs persisted even after the Ukrainian liberation of Kherson city end of 2022. The reason might be that their headquarters are, according to their websites, in Kakhovka, Oleshky, and Henichesk, respectively. All cities are in the still Russian-occupied part of Kherson Oblast on Dnipro's left bank. Beyond regional address blocks, several non-regional ASes were also disconnected, see Figure 28. This includes Askad, Next, Volia, Yanina, and Smart-M. Our dataset further confirms increased RTTs for Ukrcom, and LLC AIT.

4 ***June 6, 2023*** The destruction of the Kakhovka Dam and successive flooding led to significant regional disruption [24]. While Netblocks reported only a single flood-related outage affecting Volia on June 14 [35], we detected additional outages. OstrovNet, headquartered in Kherson city's port district on Korabel Island, appeared to be severely affected by flooding. According to our data, it took three months to restore connectivity, with services resuming in September 2023. Interestingly, the *IPS* ▲ signal remained largely unaffected for providers that retained BGP connectivity. This suggests that most responsive IPs were located outside the flooded areas. We observed disruptions in the *FBS* ■ or *IPS* ▲ signal for Viner Telecom, TLC-K, and Digicom not only operating in Kherson city.

5.3 Disruptions at the Status ISP

Finally, we focus on AS25482 Status, a local ISP in Kherson, mainly operating in Kherson city. We were in contact with one of the operators and were able to verify provider-level events in our measurement data.

5 ***March - May 2022*** In the first months of the occupation, Russian troops seized local ISPs. For Status, there is video footage of soldiers entering the provider's server rooms. Figure 13 displays the video's timestamp alongside our outage signals. At that time of this event, the *IPS* ▲ signal decreased, while *BGP* ★ and *FBS* ■ remained stable. This demonstrates the capability of our measurement approach to detect localized, provider-specific outages and emphasizes the sensitivity of the *IPS* ▲ signal to such events. This highlights another important use of the dataset: in a conflict where misinformation is widespread, the data can help verify the authenticity of reported incidents and related footage.

6 ***November 11, 2022*** Ukrainian forces recaptured Kherson city, eventually restoring control over local infrastructure [23]. Before, Russian troops destroyed infrastructure to disguise their retreat. We confirm outages, including Status ISP. Figure 14 shows the operator's four blocks, three of which are regional to Kherson and another one to Kyiv. According to our results, two blocks in Kherson stopped responding on November 11, while the block in Kyiv remained responsive. Ten days later, the Kherson blocks became responsive again, but only with clear diurnal cycles, potentially reflecting only limited availability of electricity during daylight hours.

In conclusion, our evaluation confirms the validity of our outage signals at multiple levels, from regional and AS-level incidents to outages impacting individual providers and subblocks.

Key Insights from Our Dataset

(1) Longitudinal view. Ukraine's Internet disruptions peaked in winter 2022/23 and throughout 2024 (*IPS* ▲ outages).
(2) In 2024, a total of 1,951 hours of power outages were reported. Our dataset shows non-frontline regions experienced on average 686 hours of Internet disruptions, and a worst-case maximum of 2,822 hours in which at least one oblast was affected. Outage days show a strong correlation with power cuts ($r = 0.725$).
(3) AS-level insights in Kherson: (I) Apr 2022 Mykolaiv cable cut (24 ASes offline), (II) May–Nov 2022 Russian enforcement (21 ASes, RTT spikes for 8 ISPs), (III) Jun 2023 Kakhovka flooding (OstrovNet offline 3 months).
(4) Insights into small regional ISPs. Our dataset verifies video footage from Status ISP recorded on May 13, 2022. The office search caused interference resulting in a visible *IPS* ▲ outage.

5.4 IODA AS-level Signal Comparison

As with the regional analysis, we compare our detected outages for 1,773 ASes with regional /24 blocks in Ukraine against those reported by IODA, using data retrieved via the IODA API [25]. We include all /24 blocks per AS, as our focus is on AS-wide outages rather than region-specific ones, making them more comparable to IODA. Otherwise, IODA can report outages for non-regional blocks that we do not measure.

Extended AS Coverage We first evaluate for how many ASes in Ukraine, outages are reported. For comparability, we exclude outages from the *IPS* ▲ signal (absent in IODA) as well as IODA's Merit network telescope signal, which accounts for only 2% of its outages and is not present in our dataset. Figure 15 shows a CDF of the reported outages ordered by increasing AS size (in /24 blocks). Our approach reports 77.6K outages in 1,674 ASes. IODA reports 31.9K outages for 333 ASes and none for the other 1,440. Feedback from IODA confirms that it only reports outages for ASes with 20 or more /24s, affecting many smaller regional ASes in Ukraine.

Probing Interval We evaluate the effect of the bi-hourly probing interval on our results by quantifying IODA outages (any signal type) that occur during the 100 minutes between our measurements. Out of a total of 31.9K outages, on average 70.5% fall within one of our probing intervals. Examining the signals separately, 23.7% of *BGP* ★ outages and 29.5% of *TRIN* ■ outages occur in the 100-minute gap between measurement cycles. This limitation could be mitigated by reducing the probing interval in future *FBS* ■ measurements. For instance, hourly scans would miss only 9.5% of outages, though at the cost of doubling storage requirements and necessitating real-time BGP table tracking, as historic RouteViews data is available only in bi-hourly intervals. A 30-minute probing interval with a 10-minute gap would further reduce missed outages to 0.1%. Alternatively, *FBS* ■ scans could explore lowering the scan rate, thereby distributing probes across a longer period.

Comparison of Common Outages To align our comparison with IODA, we narrow the set to 182 ASes that reach high coverage in our measurements (target share > 0.9).

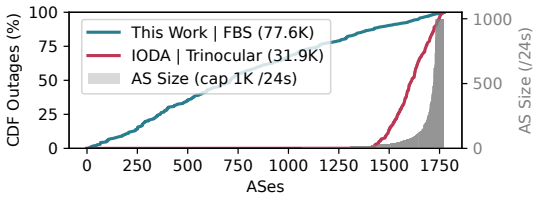


Figure 15: Comparison of AS outage coverage with IODA. CDF of outages with ASes ranked by their their size.

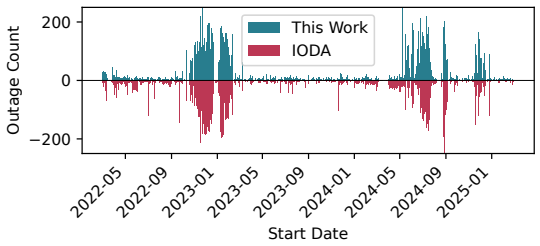


Figure 16: Number of outages starting per day reported by this work and IODA for a set of 182 ASes covered by both datasets.

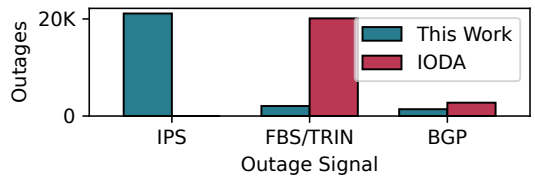


Figure 17: Signals and their share on total outages for the set of common ASes.

For comparability, we subtract our missing measurement periods from IODA outage periods and vice versa. We also incorporate the third signal of responsive IP addresses (*IPS* ▲, see Section 3.1).

Figure 16 shows strong agreement with Trinocular-based IODA data ($r = 0.85$), while Figure 17 details the contributions of individual signals. IODA primarily detects outages via active /24 blocks (*TRIN* ■), whereas our full-block scans rely on responsive IPs (*IPS* ▲). This suggests that many *TRIN* ■ events correspond to partial, not full, outages. In our data, *FBS* ■ contributes only 2,063 outages compared to 20,113 outages from *TRIN* ■ in IODA, since we require full block unresponsiveness, while Trinocular flags a block if only a few probed IPs fail. Our additional signal *IPS* ▲ detects 21,120 outages, capturing sudden loss of responsiveness among previously reachable IPs and indicating that IODA often classifies partial failures as block-wide outages.

Undetected Outages We now examine cases where the set of ever-active IPs per month changes, and the *IPS* ▲ signal captures an outage while *TRIN* ■ does not. Therefore, we compare the outages reported by both signals and quantify the number of days that an outage was detected in one dataset but not the other. In favor of IODA, we observe 6,943 cases in which *TRIN* ■ reported an outage and *IPS* ▲ did not, primarily due to short-lived outages lasting less than two hours. In contrast, we identified 12,088 instances in which *IPS* ▲ reported an outage and IODA did not detect a corresponding event. In summary, our approach sacrifices some temporal granularity and therefore provides richer detail and broader coverage of Ukrainian ASes.

Key Takeaways

(1) *FBS* ■ provides a stable signal that is particularly useful in countries with many small regional providers and helps increase AS coverage (1,674 vs. 333, IODA).
(2) While we probe more frequently than other IP-based approaches [22, 42], the bi-hourly schedule misses ~30% of short-lived outages; future *FBS* ■ scans with 30–60 min intervals could further reduce this gap.
(3) Results correlate strongly with IODA ($r = 0.85$), but signals differ: *TRIN* ■ often shows partial outages, while *FBS* ■ capture more significant and *IPS* ▲ still allows to detect partial ones.

6 Discussion

Internet Disruption Characteristics With an address churn of up to 67%, detecting Internet outages on the regional level for Ukrainian oblasts is challenging. Consequently, we develop a novel method that assigns IP blocks only to regions that remain continuously associated with them. By the example of Kherson oblast, regional blocks account for only 38% of all blocks, and only 7% of their IP addresses respond to our measurements, the lowest among all regions. Despite this low responsiveness, we observed outage signals, even for providers with a handful of address blocks, and were able to verify them against reported events.

Advantages Through Regional Classification A central challenge, also encountered in related work, is the attribution of outages to specific regions. In §4, we address this by leveraging long-term trends in IP geolocation to improve confidence in block-level location assignments. We see a clear advantage in §5.1, where we find a strong correlation (Pearson $r = 0.725$) between Internet outages and recorded power outages in non-frontline regions, significantly higher than the correlation observed in IODA data ($r = 0.328$), indicating that our outage data more accurately represents these regions. Together, our methodology and dataset offer improved insights into Internet disruptions, particularly in countries with high address churn, such as Ukraine.

Insights from Kherson Internet outages in non-frontline regions appear to be predominantly caused by power outages, whereas the frontline oblast of Kherson is additionally affected by damage to communication lines and network infrastructure, including cable cuts, equipment seizure, and traffic rerouting. In view of the circumstances, the Internet in Kherson was surprisingly resilient. Our personal exchange with a local operator revealed three key aspects, namely (I) the local Kherson Internet Exchange (KS-IX) to share (while originally not intended to) upstream connectivity among operators, (II) the deployment of redundant servers, links, and emergency power systems, and (III) the use of passive optical networks (PON) reducing dependency on electricity.

Advances in Outage Detection We extended outage detection by a novel signal on responsive IP addresses (*IPS* ▲), which is only feasible when comprehensively probing the address space, to detect partial outages. This way, we are able to extend coverage from 330, as covered by IODA, to 1,674 ASes in Ukraine. This particularly includes outages at smaller ASes, such as our example Status ISP in Kherson, that would remain undetected otherwise, and is particularly relevant for countries as Ukraine with a highly fragmented Internet provider structure. Only a few of the investigated ASes exhibit clear day-night cycles, suggesting that outages of end users might be underrepresented. A promising direction for future work is the integration of IPv6-based signals, especially as we identified growth in its deployment in Ukraine, see Figure 20. Identifying home routers by NTP [39] or ICMPv6 error messages [15] would offer improved visibility on residential networks as they are not hidden behind NAT. This study relies on fixed probing intervals or rates for FBS, future work could further explore the impact of different intervals or explore dynamic thresholds for outage detection.

7 Conclusion

In this work, we demonstrated that Internet outages are reliably detected by active measurements from a single vantage point. Sending ICMP requests to all Ukrainian IP addresses at a two-hour interval, we gained detailed insight into Internet disruptions in the presence of kinetic warfare, and only a single opt-out request was received while measuring a country at war. IP churn motivated a more sophisticated approach to assigning probed addresses to regions in Ukraine. Focusing on regional ASes, we derive three distinct outage signals: BGP visibility, the number of responsive full blocks, and responsive IP counts. This multi-signal approach enabled us to validate disruptions against known events, uncover previously undocumented outages, and correlate connectivity loss with infrastructure damage and power failures. By combining active measurements with regional attribution, we provide a practical dataset that reveals Internet disruptions in Ukraine, which we found not to be fully captured by existing methods.

Acknowledgements

We thank Alexander Kovalenko from ISP Status for sharing valuable insights into ISPs in Kherson throughout the three-year war period. We are also grateful to Zachary Bischof and the IODA team for providing the platform and sharing insights on outage detection. Finally, we thank the anonymous reviewers for their constructive feedback on this paper. This research is supported in part by the Austrian Science Fund (FWF SFB project SPyCoDe 10.55776/F85), the Austrian Research Promotion Agency (FFG TestCat 911248) and SBA-K1 NGC, a COMET Center within the COMET – Competence Centers for Excellent Technologies Programme, funded by BMIMI, BMWET, and the federal state of Vienna.

References

[1] Abdullah. 2023. The radius field in the IP to Geolocation extended database explained. https://community.ipinfo.io/t/the-radius-field-in-the-ip-to-geolocation-extended-database-explained/694 Accessed: 2025-05-16.

[2] Carlos Azevedo. 2023. Gone offline: how Cloudflare Radar detects Internet outages. https://blog.cloudflare.com/detecting-internet-outages/ Accessed: 2025-04-09.

[3] Guillermo Baltra and John Heidemann. 2019. *Improving the optics of active outage detection (extended).* Technical Report. Technical Report ISI-TR-733. USC/Information Sciences Institute.

[4] Guillermo Baltra and John Heidemann. 2020. Improving Coverage of Internet Outage Detection in Sparse Blocks. In *Proceedings of the Passive and Active Measurement Workshop*. Springer, Eugene, Oregon, USA.

[5] David Belson. 2025. A diversity of downtime: the Q4 2024 Internet disruption summary. https://blog.cloudflare.com/q4-2024-internet-disruption-summary/ Accessed: 2025-04-09.

[6] David Belson. 2025. *New year, no shutdowns: the Q1 2025 Internet disruption summary.* Cloudflare. https://blog.cloudflare.com/q1-2025-internet-disruption-summary/ Accessed: 2025-04-22.

[7] Zachary S Bischof, Kennedy Pitcher, Esteban Carisimo, Amanda Meng, Rafael Bezerra Nunes, Ramakrishna Padmanabhan, Margaret E Roberts, Alex C Snoeren, and Alberto Dainotti. 2023. Destination unreachable: Characterizing internet outages and shutdowns. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 608–621.

[8] Center for Countering Disinformation. 2023. *Occupiers use blackmail and threats to force Ukrainian providers to connect to Russian networks.* https://cip.gov.ua/ua/news/okupanti-shantazhem-i-pogrozami-zmushuyut-ukrayinskikh-provaideriv-pidklyuchatisya-do-rosiiskikh-merezh Accessed: 2025-04-24.

[9] Inc. Cloudflare. 2023. *One Year of War in Ukraine.* Cloudflare Blog. https://blog.cloudflare.com/one-year-of-war-in-ukraine/ Accessed: 2025-05-05.

[10] Developing Telecoms. 2024. *Kyivstar starts next phase of network backup-power upgrades.* https://developingtelecoms.com/telecom-technology/energy-sustainability/17551-kyivstar-starts-next-phase-of-network-backup-power-upgrades.html Accessed: 2025-05-12.

[11] DiXi Group. 2025. *Electricity outages lasted almost 2 thousand hours in 2024.* https://dixigroup.org/en/electricity-outages-lasted-2-thousand-hours-for-ukrainian-households-in-2024/ Accessed: 2025-04-22.

[12] Frédérick Douzet, Louis Pétiniaud, Loqman Salamatian, Kevin Limonier, Kavé Salamatian, and Thibaut Alchus. 2020. Measuring the fragmentation of the Internet: the case of the Border Gateway Protocol (BGP) during the Ukrainian crisis. In *2020 12th international conference on cyber conflict (CyCon)*, Vol. 1300. IEEE, 157–182.

[13] 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.

[14] Romain Fontugne, Ksenia Ermoshina, and Emile Aben. 2020. The Internet in Crimea: a Case Study on Routing Interregnum. In *2020 IFIP Networking Conference*. Paris, France. https://hal.archives-ouvertes.fr/hal-03100247

[15] Florian Holzbauer, Markus Maier, and Johanna Ullrich. 2024. Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources. In *Proceedings of the 2024 ACM on Internet Measurement Conference*. 280–294.

[16] Intent Press. 2025. *Russians Shell 13 Settlements in Kherson Region, Killing 3 Civilians.* https://intent.press/en/news/war/2025/russians-shell-13-settlements-in-kherson-region-killing-3-civilians/ Accessed: 2025-04-22.

[17] IODA. 2023. Internet Outage Detection and Analysis (IODA). Retrieved Dec 12 2023 from https://www.caida.org/projects/ioda.

[18] IPinfo, Inc. 2025. IPinfo IP Geolocation API. https://ipinfo.io/products/ip-geolocation-api Accessed: 2025-03-19.

[19] Akshath Jain, Deepayan Patra, Peijing Xu, Justine Sherry, and Phillipa Gill. 2022. The ukrainian internet under attack: an NDT perspective. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 166–178.

[20] João Tomé and David Belson. 2022. Tracking Shifts in Internet Connectivity in Kherson, Ukraine. Cloudflare Blog. Retrieved October 14, 2024, from https://blog.cloudflare.com/tracking-shifts-in-internet-connectivity-in-kherson-ukraine/.

[21] Kentik. 2015. Network Traffic Intelligence at Tomorrow's Scale. https://assets.ctfassets.net/6yom6slo28h2/6ByuKfku9UGGuwoKisU4go/b7299a843ccb30872d6b893286297b2b/Kentik-Overview-whitepaper-Jul2015.pdf

[22] Johannes Klick. 2023. How to use Internet scans and passive measurements to analyze Russian attacks and their impact in Ukraine. In *Chaos Communication Camp 2023* (Milliways, Chaos Communication Camp). Slides available.

[23] Mick Krever, Anna Chernova, Teele Rebane, Gianluca Mezzofiore, Tim Lister, and Sophie Tanno. 2022. *Ukrainian troops sweep into key city of Kherson after Russian forces retreat, dealing blow to Putin.* CNN. https://edition.cnn.com/2022/11/12/europe/kherson-city-ukraine-russia-intl/index.html Accessed: 2025-04-22.

[24] Kseniia Kunakh. 2024. *Kakhovka Dam Flooding Detection In Ukraine For PEJ.* EOS Data Analytics. https://eos.com/blog/kakhovka-dam-flooding-detection-in-ukraine-for-pej/ Accessed: 2025-04-22.

[25] Internet Intelligence Research Lab. 2025. Internet Outage Detection and Analysis (IODA) API v2. https://api.ioda.inetintel.cc.gatech.edu/v2/. Accessed: 2025-05-04.

[26] Valerio Luconi and Alessio Vecchio. 2023. Impact of the first months of war on routing and latency in Ukraine. *Computer Networks* 224 (2023), 109596.

[27] Doug Madory. 2022. Rerouting of Kherson follows familiar gameplan. Retrieved Jan 16 2024 from https://www.kentik.com/blog/rerouting-of-kherson-follows-familiar-gameplan/.

[28] Doug Madory. 2023. *Ukraine's Wartime Internet from the Inside.* https://www.kentik.com/blog/ukraines-wartime-internet-from-the-inside/ Accessed: 2025-05-15.

[29] Amanda Meng and Tara Kelly. 2025. *How Russia's Recent Attacks on Ukraine's Energy Grid Impacted its Internet Connectivity.* https://ioda.inetintel.cc.gatech.edu/reports/how-russias-recent-attacks-on-ukraines-energy-grid-impacted-its-internet-connectivity-2/ Accessed: 2025-05-16.

[30] Tal Mizrahi and Jose Yallouz. 2022. Internet Performance in the 2022 Conflict in Ukraine: An Asymmetric Analysis. *arXiv preprint arXiv:2205.08912* (2022).

[31] Tal Mizrahi and Jose Yallouz. 2022. Using Internet Measurements to Map the 2022 Ukrainian Refugee Crisis. *arXiv preprint arXiv:2205.08903* (2022).

[32] National Power Company Ukrenergo. 2025. Information on electricity consumption limitation measures. https://map.ua-energy.org/en/resources/0f8f9882-1fb2-47c6-81dc-31fbad914f16/. https://map.ua-energy.org/en/resources/0f8f9882-1fb2-47c6-81dc-31fbad914f16/ Dataset covering planned stabilization electricity outages for households across Ukraine. Data spans from January 1, 2023, to January 20, 2025. Last updated on January 21, 2025..

[33] Netblocks. 2022. Internet disruptions registered as Russia moves in on Ukraine. Retrieved Jan 16 2024 from https://netblocks.org/reports/internet-disruptions-registered-as-russia-moves-in-on-ukraine-W80p4k8K.

[34] Netblocks. 2022. Mobile internet disrupted in Luhansk, Ukraine amid heightened tensions with Russia. Retrieved Dec 12 2023 from https://netblocks.org/reports/mobile-internet-disrupted-in-luhansk-ukraine-amid-heightened-tensions-with-russia-l8Wx7LAO.

[35] NetBlocks. 2023. *Confirmed: Metrics indicate internet provider Volia in Kherson is experiencing a major outage.* https://x.com/netblocks/status/1668910973011279872 Accessed: 2025-04-24.

[36] 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.

[37] Philipp Richter, Ramakrishna Padmanabhan, Neil Spring, Arthur Berger, and David Clark. 2018. Advancing the art of internet edge outage detection. In *Proceedings of the Internet Measurement Conference 2018*. 350–363.

[38] RIPE NCC. 2024. RIPE NCC Statistics. Retrieved from https://ftp.ripe.net/pub/stats/ripencc/. Accessed: 2022-12-14.

[39] Erik Rye and Dave Levin. 2023. Ipv6 hitlists at scale: Be careful what you wish for. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 904–916.

[40] Aaron Schulman and Neil Spring. 2011. Pingin'in the rain. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 19–28.

[41] James Shires and Isabella Wilkinson. 2024. *Internet resilience in Ukraine.* Technical Report. Chatham House. doi:10.55317/9781784136123 Accessed: 2025-04-22.

[42] Rishabh Singla, Shreyas Srinivasa, Narasimha Reddy, Jens Myrup Pedersen, Emmanouil Vasilomanolakis, and Riccardo Bettati. 2023. An analysis of war impact on Ukrainian critical infrastructure through network measurements. In *7th Network Traffic Measurement and Analysis Conference 2023*. IFIP.

[43] TASS. 2022. Integration of power grids of Crimea, rest of Russia completed — Deputy Energy Minister. *TASS* (29 December 2022). https://tass.com/politics/1557425 Accessed: 2025-05-10.

[44] Ukrinform. 2025. Regional chief reveals current population of Kherson region. https://www.ukrinform.net/rubric-society/3953685-regional-chief-reveals-current-population-of-kherson-region.html. Accessed: 23 April 2025.

[45] University of Oregon. 2024. University of Oregon Route Views Project. http://www.routeviews.org/routeviews/. Accessed: 2024-10-14.

A Ethics

We planned our measurements so as not to put additional load on the Internet of a country already at war. This includes only a single probe per two hours, randomized targets, a low probing rate of 8,000 packets per second, i.e., around 500KB/s from a single vantage point spread across all Ukrainian IP ranges and only minimal resources of these systems were used (e.g., ICMP instead of the stateful TCP), while platforms such as Censys are known to

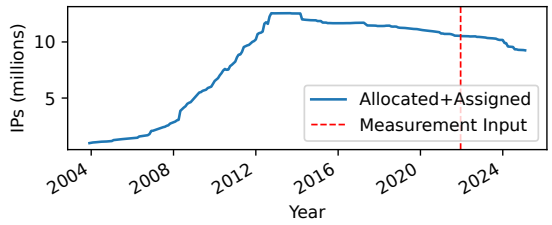


Figure 18: IPv4 address ranges with status allocated or assigned to UA from RIPE NCC over time.

probe multiple ports multiple times a day. Furthermore, we offered additional information, contact details, and the option to opt out via a web server. We received only one opt-out request, not due to resource strain, but because the requester preferred not to be included in our data.

Our results offer insight into the resilience and state of the Ukrainian Internet during wartime, but they do not reveal information that could be directly exploited to further endanger the country. Access to the underlying data is therefore carefully controlled. While Internet outage measurements can be independently collected by others, unrestricted release could enable adversaries to assess the effectiveness of attacks on power supply or network infrastructure (e.g., in frontline regions) without relying on other sources such as satellite imagery or reconnaissance.

At the same time, the data is of clear scientific and societal value: it allows researchers to quantify and verify the impact of concrete events on Internet availability, such as the seizure of infrastructure, as was evaluated in this paper. To balance these interests, we provide block-level availability data to the research community and if requested anonymized IP-level responsiveness, which avoids privacy risks while enabling meaningful analysis. Any further release of more detailed data is coordinated in consultation with the national CERT of Ukraine.

B Country-specific IP ranges

Using a one-time snapshot of the delegated files as input lets us track prefixes over time. While BGP-announced prefixes are subject to frequent changes, allocated and assigned prefixes tend to be more stable.

For Ukraine, we observed the following trends: out of the initial 3,085 allocated ranges, 3,026 (98%) still existed as of January 2025, with 2,678 (87%) remaining allocated to Ukraine. This means that 348 prefixes (12%) have changed country codes. Among these, 31% are now assigned to Russia (*RU*), 13.5% to the United States (*US*), 11% to Poland (*PL*), 9% to Latvia (*LV*), and the remaining third to other, mostly European, countries. In the second snapshot, we identified a total of 2,876 IP ranges still belonging to Ukraine.

Two key observations emerge from this analysis. First, the total number of IP allocations in Ukraine has decreased by 7%. We find the impact of new allocations to be minor since our snapshot. Figure 18 shows the development since 2004. Second, 12% of previously Ukrainian prefixes have now been reassigned to different country codes, but might still be valuable measurement targets.

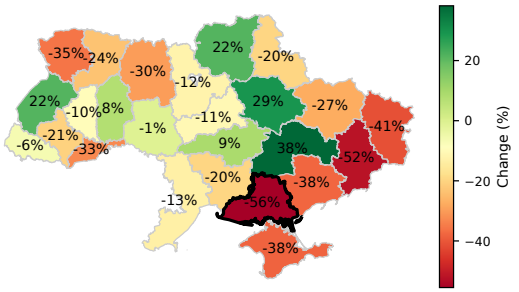


Figure 19: Churn of all IPv4 addresses locating to oblasts in Ukraine, comparing 2022-02-01 to 2025-02-01.

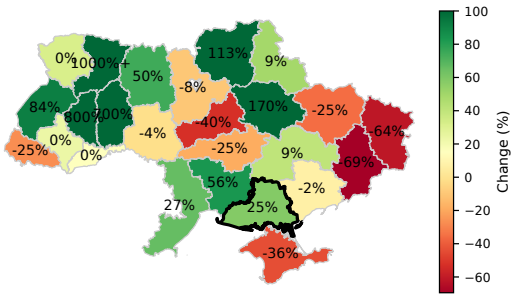


Figure 20: Churn of all IPv6 addresses locating to oblasts in Ukraine, comparing 2022-02-01 to 2025-02-01.

C Extended IP Churn

To compare IP churn of measurement targets to all IPv4 addresses, we include Figure 19, which is the same as Figure 1, without the restriction to our measurement targets in the Appendix. While most oblasts show similar churn, Luhansk differs by 26 percentage points, Crimea by 17, Zaporizhia by 14, Mykolaiv by 7, Khmelnytskyi and Kherson by six, and all others by below five.

IPv6 Churn . While there is a noticeable decrease in IPv4 addresses across Ukraine, a different picture emerges for IPv6. We replicate Figure 19 in Figure 20 for IPv6. We find a noticeable increase in IPv6 adoption across Ukraine. Regions with low or no IPv6 adoption show a high increase in percentage points (Rivne, Ternopil, and Khmelnytskyi). It could be especially interesting for regions with noticeable decreases in IPv4 addresses, such as Kherson, Mykolaiv, and Sumy, to include IPv6 measurements in the future.

D Regional ASes and Blocks

Mapping blocks to regions in Ukraine shows around 14% of blocks that point to multiple regions. If the block meets the regional criteria in one region, we will only consider the part of the block that is

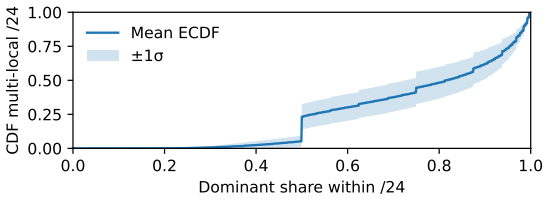


Figure 21: CDF of blocks highlighting the share of IPs pointing to the dominant location for multi-local /24 blocks.

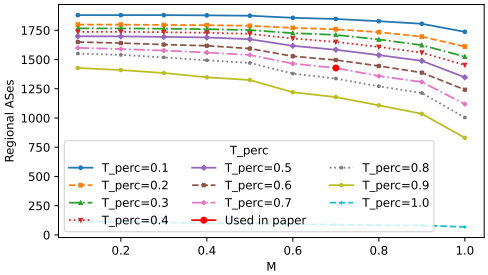


Figure 22: Choice of parameters M and T_perc and their impact on the number of regional ASes.

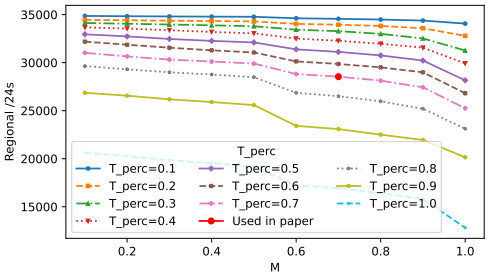


Figure 23: Choice of parameters M and T_perc and their impact on the number of regional /24s

locating to the target region. However, as Figure 21 shows there is usually a majority of IPs in the block that geolocate to the dominant region. To better limit the impact of noise in geolocation data, we separate regional from non-regional blocks and ASes in Ukraine. Figure 23 illustrates the sensitivity on the block level while Figure 22 on the AS level to varying values of the geolocation threshold (M) and the required percentage of routed months (T_{perc}). We test both parameters, ranging from 0 to 1 in steps of 0.1.

Using the strictest setting ($M = 0.9$, $T_{perc} = 0.9$), we classify 1036 out of 2024 ASes (51.19%) as regional, resulting in 21,952 regional subblocks. A majority setting ($M = 0.5$,$T_{perc} = 0.5$) yields 1674 regional ASes (82.71%) and 32,107 regional subblocks. We

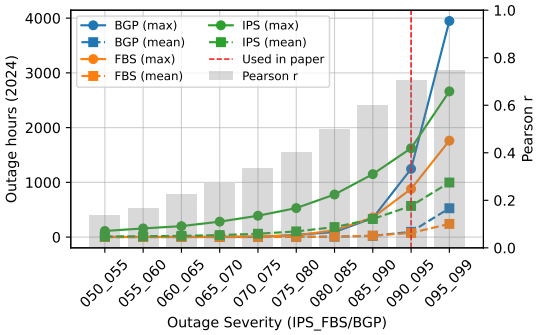


Figure 24: Different regional outage severity thresholds and the respective outage hours in 2024 in non-frontline regions and the correlation coefficient with power outages.

adopt $M = 0.7$, $T_{perc} = 0.7$ as a balanced configuration, identifying 1428 regional ASes (70.55%) and 28,541 regional subblocks. This compromise avoids over-classification due to noisy geolocation while still capturing consistent regional behavior.

E Outage Severity and Threshold Sensitivity

We evaluate how different outage thresholds affect the number of reported outage hours. Our analysis focuses on Internet outages detected throughout 2024 in non-frontline regions, where we can correlate Internet disruptions with reported power outages from Ukrenergo (see Section 5.1). Outage severity is measured as the deviation from the moving average over the previous week. Figure 24 shows results for thresholds ranging from 50% to 99%. The *IPS* ▲ signal applies a threshold that is five percentage points stricter than the other outage signals, since it is more volatile and IPs typically fail before entire blocks do. We observe that only a small fraction of outages affect 50% or more of IPs or blocks in a region. At the other end of the spectrum, the most sensitive threshold, which is triggered when just 5% of IPs or 1% of blocks go offline, likely overestimates outage hours. We reach a similar correlation with reported power outages already at lower thresholds, namely 10% IP loss or 5% block loss (through unresponsiveness or loss of BGP visibility). Using these thresholds reduces the number of reported outage hours while capturing more relevant events.

F AS-level Disruptions Kherson

Target ASes Table 5 lists ASes with regional /24s in the Kherson oblast. ASes are split into regional and non-regional and ranked by their number of regional /24s inside the category. For all 34 ASes, we manually investigated the location of their headquarters. Most regional ASes are headquartered in the Kherson oblast, with only one based in Kyiv. In contrast, the majority of non-regional ASes are headquartered in Kyiv. Note that we did not restrict non-regional ASes to those registered in Ukraine; as a result, two foreign ASes appear in the non-regional group. We observe six regional ASes in Kherson that each have only a single regional /24. Similarly,

ASN	/24s	Reg.	Org.	HQ	IODA [17]	RU [20]	∅ [45]
49465	16	16	RubinTV	N. Kakhov	○	●	○
56404	8	8	Norma4	Kherson	○	●	○
56359	5	5	RostNet	Oleshky	○	●	●
25482	4	3	Status	Kherson	○	●	○
15458	2	2	TLC-K	Kherson	○	●	●
47598	3	2	Kherson Telecom	Kherson	○	●	●
56446	2	2	OstrovNet	Kherson	○	●	○
25256	1	1	M-Net	Henichesk	○	○	●
34720	1	1	JSC-Chumak	Kyiv	○	○	●
42469	1	1	Askad	Skadovsk	○	○	●
44737	1	1	Next	Kherson	○	○	●
59500	1	1	LineVPS	Kherson	○	○	○
211171	1	1	Pluton	Kherson	○	●	○
25229	190	160	Volia	Kyiv	●	○	○
15895	299	52	Kyivstar	Kyiv	●	○	○
6877	239	49	Ukrtelecom	Kyiv	●	○	○
6849	682	31	Ukrtelecom	Kyiv	●	○	○
6703	29	12	Vega	Kyiv	●	○	○
21151	18	10	Ukrcom	Kherson	○	●	○
6698	16	9	Virtualsystems	Kyiv	●	○	○
30823	6	6	Aurologic	Langen(DE)	●	○	○
205172	6	6	Yanina	Kherson	○	○	●
39862	7	4	Digicom	Kherson	○	○	○
57498	4	3	Smart-M	Kherson	○	○	●
2914	2	2	NTT	Redmond(US)	●	○	○
12883	8	2	Vega	Kyiv	●	○	○
25082	12	2	Viner Telecom	Kherson	○	●	○
35213	12	2	CompNetUA	Kherson	○	○	○
49168	2	2	Brok-X	Kherson	○	●	○
6846	7	1	Infocom	Kyiv	●	○	○
12687	1	1	Uran Kiev	Kyiv	●	○	○
45043	4	1	Viner Telecom	Kherson	○	○	○
197361	1	1	LLC AIT	Kherson	○	●	●
215654	1	1	Genicheskonline	Henichesk	○	○	○

/24s in Ukraine, Regional, Non-Regional, ∅ No BGP prefixes in 2025

Table 5: Regional and non-regional ASes in Kherson, showing number of regional /24s, headquarters, IODA coverage, reports on rerouting, and whether they announced any prefixes in 2025.

five non-regional ASes, mostly larger ISPs, also show a single regional /24 in the oblast. Despite the limited number of responsive IPs (1,400 in 2025), we find that most ASes in Kherson remained responsive if they maintained their service in the region. We also examine third-party reporting. IODA [25] has reported outages for a subset of larger, non-regional ASes in Kherson (see § 5.4). Cloudflare [9] identified 15 ASes in 2022 as rerouting traffic via Russian upstreams; 12 of them are included in Table 5. By 2025, RouteViews BGP tables [45] reveal that seven of the 13 regional ASes had ceased announcing any prefixes, suggesting that many local operators were either decommissioned or had permanently shut down operations in the region.

Disruption Timeline Figure 28 extends Figure 11 to cover the entire measurement period. Despite only 7% responsiveness for regional IPs in Kherson, we are still able to observe outage signals for most ASes in the region, including those with only a single regional /24. Over the three-year period, a significant portion of regional /24s belonging to non-regional ASes are not visible in BGP. During the Russian occupation, many of these address ranges were disconnected—particularly for ASes such as Vega (Alkar), Smart-M, and Yanina—which remained offline for extended periods. Volia was also disconnected but reappeared after the liberation of the right bank. While for non-regional we see that blocks initially not

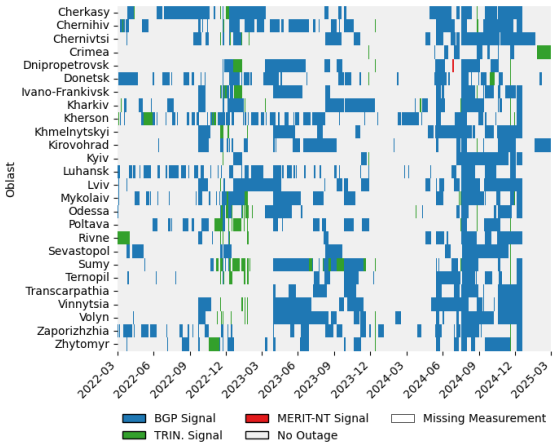


Figure 25: IODA outages reported for regions in Ukraine.

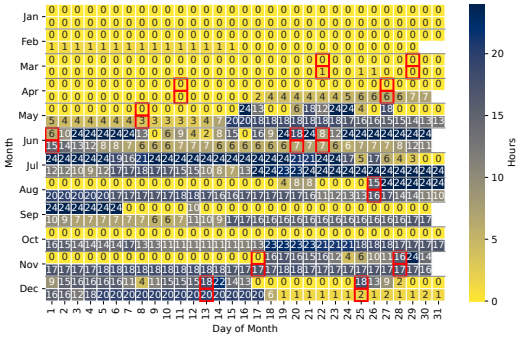


Figure 26: Comparison of daily average hours of power outages reported by Ukrenergo [32] (top row) and Internet disruptions detected by IODA (bottom row) for non-frontline regions in 2024. Days marked in red correspond to reported attacks as documented by [11].

visible in BGP were announced during the measurement period, as visible for Brok-X, Genicheskonline, NTT. Regional providers show the opposite of being active in Kherson first and then discontinuing their service, probably due to falling subscriber bases, as was reported by the Status ISP. We observe that connectivity in Kherson experienced repeated cycles of disruption and restoration. Major outage periods coincide with known events such as the severing of the Mykolaiv cable and the destruction of infrastructure during the Russian retreat. Additionally, a clearly visible multi-AS disruption occurred on November 28th, as was captured by the IODA regional signal [29] for Kherson.

G IODA Signal

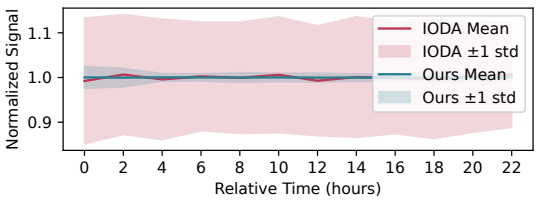


Figure 27: Signal deviation from the mean of one day (March 2, 2023) for the IODA (Trinocular) signal and our signal across 1,073 ASes without signal loss (# of active /24s = 0).

Signal Stability We further evaluate possible reasons for the differences in AS coverage. First, we look at the number of ASes that IODA includes active /24s (*TRIN*■) data. For this, we request raw data for one day in each year from 2022 to 2025 from [25] for each of the ASes (YYYY-03-02). We find that IODA includes data, in at least one year, for 90% of ASes (1,597 out of 1,773). Block eligibility does not seem to be the issue here. As we found that IODA does not report outages for smaller providers, i.e. with less than 20 /24s, we will look at signal stability in the next step. Figure 27 visualizes the IODA and our signal recorded over one day for ASes that include values for each bi-hourly interval. We find the signal spread to be much more prominent for Trinocular (avg Signal to Noise Ratio=7.6) than for our signal (avg. SNR=99.7; higher=clearer signal). This can be a problem when detecting disruptions for smaller ASes, as unresponsive blocks in ASes with few /24 blocks are more likely to trigger thresholds (80% warning, 50% critical). This likely caused the filter by Richter et al. [37] to exclude blocks with many down events.

IODA Outage Events To compare outage events with IODA on a regional level, we replicated Figure 8 with IODA outage events reported for the different oblasts in Ukraine in Figure 25, showing the number of recorded outages over the three years for each of the outage signals. Compared to Figure 8 showing detected outages by our dataset, which only shows shorter outage periods on the oblast level, IODA shows long-lasting BGP signal-driven outages on the oblast level. We assume the reason to be the following: through assigning both regional and non-regional ASes to oblasts in Ukraine, BPG outages for non-regional ASes have a stronger effect on IODA data. However, this means that the BGP outage of a single non-regional provider can affect outage data in multiple regions. Apart from this, the active probing signal from Trinocular is visible in green, reporting some additional outages not visible in BGP. The merit signal is based on traffic originating from regions in Ukraine to a so-called Darknet, that is routed address space monitored for incoming traffic. The contribution of the two other signals compared to BGP on the regional level is, however, marginal.

Non-Frontline Disruptions We further investigate reported outages for non-frontline regions, replicating Figure 9 with IODA data in Figure 26. We aggregate hours affected by Internet outages in non-frontline regions and plot them on a daily basis in 2024. The figure visualizes the lower Pearson coefficient for IODA data in relation to reported power outages in non-frontline regions, which is highly likely caused by IODA's long-lasting BGP outages.

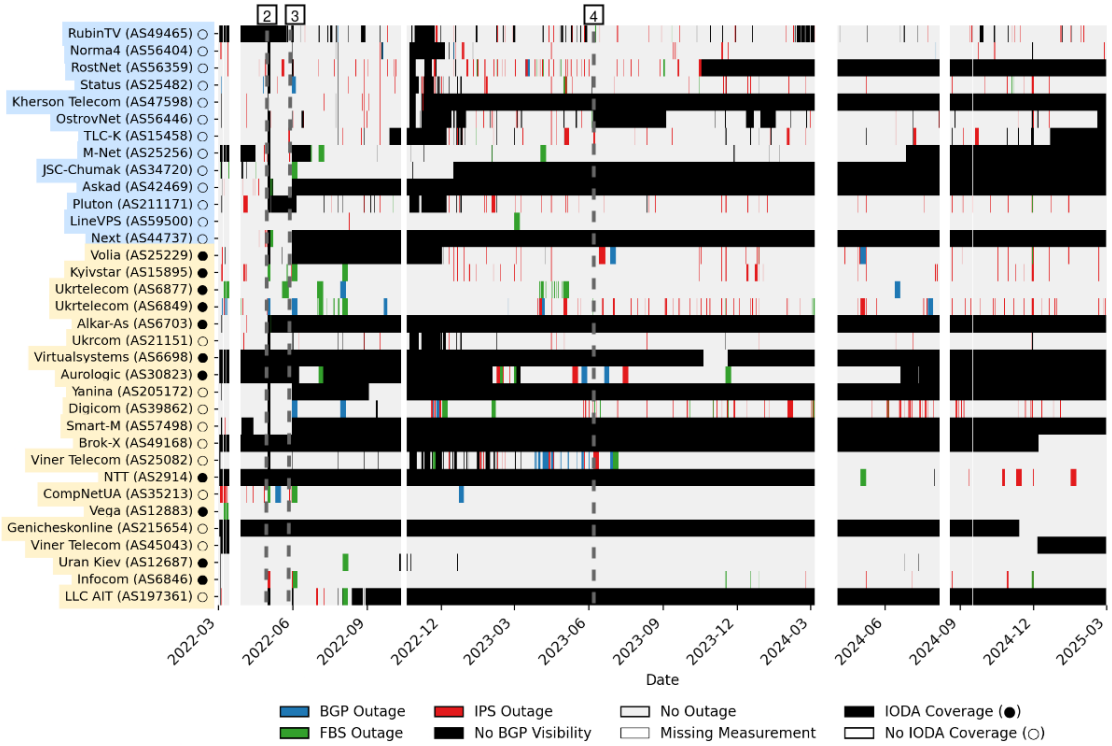


Figure 28: Internet disruptions recorded by any of the three signals for ASes in Kherson from 2022 to 2025. Most ASes show long-lasting BGP visibility loss in the region.

5 Conclusion

The thesis presents three novel Internet measurement approaches that are explicitly designed around delivery failures. The resulting methods span three domains: email delivery, IPv6 reconnaissance, and Internet outage detection.

The first publication shows how intentional delivery failures can be leveraged to measure the adoption of Internet standards on the sender side of email delivery by systematically misconfiguring or restricting receiving email and DNS infrastructure. This approach enables large-scale measurement of properties such as DNSSEC validation, opportunistic versus enforced transport encryption, and IPv6 deployment. The second publication explores ICMPv6 error messages as a source of information for IPv6 reconnaissance. We demonstrate that *Destination Unreachable* and *Time Exceeded* messages can be interpreted beyond their source address and used to distinguish between active and inactive IPv6 networks. In particular, the timing characteristics of *Address Unreachable* messages allow the identification of active networks by triggering the Neighbor Discovery process. The third publication focuses on unintentional delivery failures for Internet outage detection. By conducting full block scans of the Ukrainian address space over three years, starting on the seventh day of the war in Ukraine, we show that existing outage detection platforms, such as IODA, fail to capture disruptions affecting smaller operators.

A key contribution of the thesis is the longitudinal and continuous operation of the measurement platforms. The email delivery measurements were conducted in 2020 and 2021 and have remained publicly available since then. The ICMPv6 measurements were carried out over multiple months throughout 2023, using a methodology that is easily repeatable and transferable to other vantage points. The Ukrainian outage measurements have been running continuously since March 2, 2022, providing long-term insights into Internet disruptions during an ongoing conflict.

Taken together, these measurement platforms demonstrate that delivery failures are not merely side effects of misconfiguration or outages, but can serve as a valuable measurement signal. They enable both users and network operators to verify whether systems behave as expected and whether deployments adhere to protocol standards and best practices.

5.1 Insights on the Internet

This section summarizes and combines insights from the presented measurements. While each publication includes its own discussion, the following paragraphs highlight combined observations.

Challenges in email delivery. Email-security-scans.org provides insights into persistent challenges in email delivery. We briefly discuss several of them in more detail,

including the adoption of new standards, the trade-off between security and reachability, the role of DNS, transport security, and unsolicited bulk email, commonly referred to as spam.

The first challenge concerns the **adoption of new standards**. The standards discussed are not recent, having been published more than a decade ago. To authenticate whether a sender is authorized to send email for a domain, three sender-side authentication mechanisms were introduced: Sender Policy Framework (SPF) in 2006, DKIM in 2007, and Domain-based Message Authentication, Reporting, and Conformance (DMARC) in 2015. Our results show that SPF adoption increased from a mean of 44% in 2015 to 91% in 2020/21. Similarly, DMARC adoption rose from 1% to 53%. While no historical reference values are available for DKIM, our dataset shows that 63% of email tests sign emails using DKIM.

Successful adoption of these standards, particularly on the sender side, positively influences spam detection systems such as SpamAssassin. As a result, senders benefit directly from deploying SPF, DKIM, and DMARC, which have seen increasing support over time. In contrast, standards that primarily improve security rather than reachability tend to see lower adoption, especially as configuration complexity increases. Examples include IPv6, DNSSEC, and DANE. Related work reports adoption rates of approximately one-fifth of providers for DNSSEC (23%) and DANE (18%) in 2020 [KDS20]. Our measurements confirm this trend on the validation side. Only 57% of senders validated our DNSSEC misconfiguration, and just 22% detected the TLSA mismatch.

This leads to the challenge of **enabling security while preserving reachability**. Many providers avoid enforcing strict security policies because doing so would prevent delivery to misconfigured domains. A prominent example is Google and Microsoft, which, until after our initial study in 2020/21, did not validate DNSSEC errors on the resolvers used for their email infrastructure.

DNS has become a central component of email delivery. Sender validation relies on DNS queries for SPF, DKIM, and DMARC records, while correct MX resolution is essential to ensure delivery to the intended destination. By operating authoritative DNS servers, we were able to identify the resolvers used by senders. We found that even large providers often rely on closed resolvers for email delivery, which exhibit different behavior than public resolvers. Until 2020/21, Google performed DNSSEC validation only on its public resolvers, not on those used by Gmail. Related work found that around 65% of resolvers are closed resolvers, and that some of these are still vulnerable to the Kaminsky attack [Lon22b]. This vulnerability allows DNS responses to be spoofed due to the use of non-randomized source ports. Raising awareness for security issues from closed resolvers remains an ongoing challenge. Providers and users should therefore verify the security properties of the resolvers on which their email infrastructure depends. Email-security-scans.org currently identifies resolver IPs and checks if DNSSEC validation is performed by those resolvers. To test for non-random source ports used by resolvers, artifacts by related work exist [Lon22a].

In contrast to instant messaging and other end-to-end encrypted communication platforms, **email transport encryption remains largely opportunistic**. Communication starts in plaintext and is upgraded to encryption only if supported. This opens up opportunity for misuse. In the most severe case, STARTTLS stripping has been observed for 96% of emails sent from Tunisia to Gmail [DAM+15b]. Enforcing

TLS on the receiving side could mitigate this issue, but providers hesitate to do so. Our measurements show that even in 2020/21, approximately 10% of regular providers delivered email exclusively over plaintext. While DANE can be used to enforce TLS via DNS, it depends on DNSSEC deployment. Its alternative for unsigned zones, MTA-STS, also shows low adoption due to increased complexity, as found in survey results [AFC25].

These conditions leave **spammers** with a minimal path to implementing only what is necessary. By re-registering 50 expired domains and comparing spam volume against plaintext IPv4 delivery as a baseline, we measured standard adoption among spam-sending entities. Greylisting reduced spam volume by 36.9%, while enforcing TLS reduced spam by over 65%. In 2020/21, we observed zero spam volume for an IPv6-only reachable email server. Overall, the findings confirm that spam delivery infrastructures lag behind best practices in email transport security. Even modest measures, such as greylisting or enforced TLS, exclude a substantial fraction of spam traffic, while IPv6-only reachability effectively eliminated spam during the observation period. These results suggest that stricter transport-layer policies can simultaneously promote protocol adoption and significantly reduce unsolicited bulk email.

IPv6 Adoption. IPv6 adoption is examined across publications, including email delivery readiness, IPv6 reconnaissance, and IP churn in Ukraine. The measurements lead to new insights into IPv6 adoption.

Domain-specific adoption rates. IPv6 adoption progresses at different rates across domains. Email-security-scans.org shows that while 60% of DNS resolvers were IPv6-capable in 2020, only 40% of email servers supported IPv6 delivery. A common argument against IPv6 adoption in email is the challenge of adapting IP-based reputation systems to deal with a vastly larger address space. For example, this complicates blocklist management, as the blockage of individual IPs can be bypassed more easily than in IPv4. This motivates the usage of domain-based reputation systems in IPv6. Comparing email to the web, Google reported that IPv6 accounted for approximately 25% of web traffic in 2020 and over 50% by 2025, indicating substantial growth. Increased IPv6 connectivity among end users incentivizes IPv6 adoption in many other domains.

Steps towards active IPv6 measurements. As IPv6 adoption grows, new measurement techniques are required. Hitlist compilation remains one of the most effective approaches, mostly through passive approaches such as collecting DNS data [GSF+18] or NTP server deployments [RL23]. Our work introduces the use of ICMPv6 error messages, enabling the identification of at least one active network in 61% of responsive networks without prior knowledge of any address within the network. Future work will explore applying this approach to Ukraine to compensate for regions with sparse IPv4 addressing.

Vendor-specific implementations. RFC 4443 allows multiple implementation choices for ICMPv6 error message handling. Especially in routing scenarios around firewalling, we find considerable diversity across vendors. Our results show that access-control list (ACL) based routing scenarios affect the diagnostic value of Destination Unreachable subcodes. While this limits interpretability, it also enables router vendor fingerprinting. In combination with vendor-specific rate limiting of error messages, this

complements existing fingerprinting techniques based on banner grabbing or SNMPv3 vendor labels [AGBS21] and does not depend on explicit management-plane exposure, as ICMPv6 error messages, particularly *Time Exceeded*, are returned by most routers.

Surprising IPv6 side effects. Our measurements reveal the effects of IPv6 adoption on spam, censorship, and routing loops. As observed in email-security-scans.org, spam volume dropped from the IPv4 plaintext baseline to zero for IPv6-only reachable mail servers. While this situation is likely to change in the foreseeable future, it suggests that, during the observation period, spam delivery infrastructures had not yet adapted their operations to IPv6 email transport.

The IP churn measurements in Ukraine further show that while many regions experience IPv4 churn, IPv6 usage is increasing in many of these regions. IPv4 address leasing and resale provide financial incentives, whereas IPv6 address space remains inexpensive. This motivates extending outage detection approaches to IPv6. A side-effect of the IPv6 adoption was observed in one of the frontline regions. Collaboration with a local ISP in Kherson through CountryMonitor revealed that early Russian censorship measures failed to block IPv6 traffic.

Additionally, ICMPv6 measurements show a large fraction of *Time Exceeded* messages, indicating routing loops. Routing loops are investigated in more detail by related work [MU23, LLZ+21]. However, a common cause of routing loops in IPv6 is the assignment of a larger prefix to a customer, while only a more specific subprefix is actually routed or in use at the customer edge. This issue does not occur in IPv4 due to address scarcity and CGNAT, where multiple customers are routed over a single public IPv4 address. In IPv6, customers are often assigned at least a /64 prefix, or larger [LLZ+21]. In a loop scenario, a customer may be assigned a /56 prefix while only using a single /64. Packets destined for the remaining $2^8 - 1$ subprefixes are forwarded between the provider and the customer router until the hop limit is exhausted. A simple way to avoid routing loops is to deploy a null route for the unused prefixes. Still, routing loops are very prevalent in IPv6. In a periphery scan of 6 billion /64 prefixes, we observed 1.4 billion ICMPv6 error messages, of which nearly one third are *Time Exceeded* responses.

The Ukrainian Internet under attack. Monitoring Internet disruptions in Ukraine provided insights into the structure and operation of networks that have been repeatedly disrupted and rebuilt throughout the war. Understanding the factors that contribute to resilience is essential for adapting successful strategies.

Partial outages during power cuts. Using three independent signals, we observed that increased attacks on power infrastructure during the winter of 2022/23 and throughout 2024 manifest as partial outages in responsive IP measurements. Partial outages indicate that while IPs become unreachable, at least one within the same /24 prefix remains active. We find one reason to be the availability of backup power supply at different parts of the network. For example, Kyivstar reported in 2024 that its core infrastructure can operate without power for multiple days, while edge infrastructure is limited to several hours [LNM25].

Long-lasting outages and traffic rerouting. Long-lasting outages can arise from multiple causes. We observe operators shutting down services due to declining subscriber bases, while in occupied areas, providers are taken offline or their traffic

being rerouted through Russian upstream networks. These effects are also visible in the measurement data, where they manifest as increases in round-trip times and persistent path changes.

Kherson, a frontline region. CountryMonitor led to collaboration with the local ISP *STATUS*, operating in Kherson city and surrounding areas. This collaboration provides insights into a region that has remained on the frontline for more than three years. *STATUS* reports that during the initial phase of the war and the capture of Kherson city, providers have shared a final uplink via a local Internet exchange after their upstreams failed. In the measurement data, we observe the longest Internet outage occurring in late April 2022, resulting in a complete loss of Internet access for multiple consecutive days. During a search of telecommunication operators, *STATUS* recorded video footage of Russian soldiers entering one of their offices on May 13th, 2022, after which infrastructure was taken offline. This event coincided with a measurable drop in active measurement signals recorded by CountryMonitor. One of the most significant outages occurred after the Russian retreat in November 2022, when critical infrastructure was destroyed and underground tunnels were mined to hinder rapid repairs, leaving Kherson without power and Internet access for several days [Reu22]. After ten days, we again observed active signals from the partner ISP. Infrastructure then operated on generator-supplied electricity, which was reflected in the measurement signal being active primarily between 8 a.m. and 6 p.m. On the hardware side, *STATUS* reports several measures that have helped maintain connectivity throughout the war, including redundancy in network nodes and the availability of emergency power supply on multiple layers of their network (core and edge), the use of transport channels from different suppliers, and the absence of vendor lock-in in their passive optical network. Together with continuous repairs carried out under life-threatening conditions, these measures enabled Kherson to remain connected throughout the war.

5.2 Limitations

The results presented in this thesis are limited to the protocols investigated and to the vantage points from which the measurements were conducted. As active measurements inherently depend on the vantage point's location, routing, and connectivity, the observed behavior may differ slightly when measured from other vantage points or when applied to additional protocols. First, the scope of delivery failures explored in this thesis is based on to those covered in the publications. While we focused on SMTP and ICMP(v6) as fundamental protocols in the Internet, other forms of delivery failures exist that can be explored in the future. These remain unexplored in this work. Second, the number of vantage points is limited. The email measurement platform relies on voluntary participation by users worldwide. After the initial promotion phase, a large fraction of observed traffic originated from major providers such as Google and Microsoft, which operate a significant share of global email infrastructure, effectively reducing the sample size of regular providers after the initial study. The active IPv6 measurement campaign is conducted from two vantage points, which yield consistent results but still limit geographic diversity. The Ukrainian outage measurements are conducted from a single vantage point located in Vienna and connected via Nextlayer

using a dedicated IP range without traffic filtering. While this vantage point has shown no measurable bias in prior studies and benefits from reduced geoblocking compared to non-European locations, it nevertheless represents a single external perspective.

5.3 Future Work

Delivery failures could be further integrated into Internet measurements. First, delivery failure measurements can be extended to additional protocols beyond SMTP and ICMP(v6). This thesis establishes a baseline for applying the concept to fundamental Internet protocols, but similar approaches may be applicable to other protocols. Second, intentional and unintentional delivery failures can be measured in combination. One promising direction is to intentionally trigger ICMPv6 error messages in order to discover active residential IPv6 prefixes that can then be monitored over time for unintentional delivery failures. This is particularly relevant in regions with high IPv4 churn, such as was measured in frontline regions in Ukraine. Third, outage detection based on full block scans can be expanded beyond Ukraine. While Ukraine provides a unique and well-documented case study, applying the same methodology to other countries would allow a more global view on Internet disruptions.

5.4 Final Remarks

This thesis demonstrates three complementary ways in which delivery failures can be used as a signal in Internet measurements. First, intentional misconfiguration of email receivers enables the measurement of protocol and security adoption on the sender side. Second, ICMPv6 error messages allow IPv6 network and router classification when interpreted in terms of type and timing. Third, loss of responsiveness across the Ukrainian address space enables the detection and quantification of Internet disruptions that remain invisible to existing platforms.

Overall, delivery failures offer a new approach to studying Internet deployment, security, and resilience, and they open a promising field in future Internet measurement research.

Bibliography

[ACM25] ACM (Association for Computing Machinery). *IMC '25: Proceedings of the 2025 ACM Internet Measurement Conference*, New York, NY, USA, 2025.

[AFC25] Md Ishtiaq Ashiq, Tobias Fiebig, and Taejoong Chung. Unraveling the complexities of mta-sts deployment and management in securing email. In *Proceedings of the 2025 ACM Internet Measurement Conference*, pages 1–16, 2025.

[AGBS21] Taha Albakour, Oliver Gasser, Robert Beverly, and Georgios Smaragdakis. Third time's not a charm: exploiting snmpv3 for router fingerprinting. In *Proceedings of the 21st ACM internet measurement conference*, pages 150–164, 2021.

[BH20] Guillermo Baltra and John Heidemann. Improving coverage of internet outage detection in sparse blocks. In *International Conference on Passive and Active Network Measurement*, pages 19–36. Springer, 2020.

[Bra97] S. Bradner. Key words for use in RFCs to Indicate Requirement Levels. RFC 2119 (Best Current Practice), March 1997. Updated by RFC 8174.

[CD95] A. Conta and S. Deering. Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6). RFC 1885 (Proposed Standard), December 1995. Obsoleted by RFC 2463.

[CDG06] A. Conta, S. Deering, and M. Gupta (Ed.). Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification. RFC 4443 (Internet Standard), March 2006. Updated by RFC 4884.

[CER] CERT.at GmbH. Computer emergency response team austria. Accessed: 2025-12-04 from `https://www.cert.at/de`.

[CK06] Mark Crovella and Balachander Krishnamurthy. *Internet Measurement: Infrastructure, Traffic and Applications*. John Wiley & Sons, Inc., USA, 2006.

[Cro69] S. Crocker. Host Software. RFC 1, April 1969.

[DAM+15a] Zakir Durumeric, David Adrian, Ariana Mirian, James Kasten, Elie Bursztein, Nicolas Lidzborski, Kurt Thomas, Vijay Eranti, Michael Bailey, and J Alex Halderman. Neither snow nor rain nor mitm... an empirical analysis of email delivery security. In *Proceedings of the 2015 Internet Measurement Conference*, pages 27–39, 2015.

[DAM+15b] Zakir Durumeric, David Adrian, Ariana Mirian, James Kasten, Elie Bursztein, Nicolas Lidzborski, Kurt Thomas, Vijay Eranti, Michael Bailey, and J Alex Halderman. Neither snow nor rain nor MITM... an empirical analysis of email delivery security. In *Proceedings of the Internet Measurement Conference (IMC)*, 2015.

[DWH13] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. Zmap: Fast internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 605–620, 2013.

[FGG+21] Tobias Fiebig, Seda Gürses, Carlos H Gañán, Erna Kotkamp, Fernando Kuipers, Martina Lindorfer, Menghua Prisse, and Taritha Sari. Heads in the clouds: measuring the implications of universities migrating to public clouds. *arXiv preprint arXiv:2104.09462*, 2021.

[GGM+25] Gabriel Karl Gegenhuber, Maximilian Günther, Markus Maier, Aljosha Judmayer, Florian Holzbauer, Philipp Frenzel, and Johanna Ullrich. Careless whisper: Exploiting silent delivery receipts to monitor users on mobile instant messengers. In *28th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2025)*, October 2025.

[GHF+24] Gabriel K Gegenhuber, Florian Holzbauer, Philipp É Frenzel, Edgar Weippl, and Adrian Dabrowski. Diffie-hellman picture show: Key exchange stories from commercial vowifi deployments. In *33rd USENIX Security Symposium (USENIX Security 24)*, 2024.

[GMH+23] Gabriel K Gegenhuber, Markus Maier, Florian Holzbauer, Wilfried Mayer, Georg Merzdovnik, Edgar Weippl, and Johanna Ullrich. An extended view on measuring tor as-level adversaries. *Computers & Security*, 132:103302, 2023.

[Goo26] Google Workspace Admin Help. Email sender guidelines. `https://support.google.com/a/answer/81126?hl=en`, 2026. Accessed: 2026-01-19.

[GSF+18] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczynski, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proceedings of the 2018 Internet Measurement Conference*, New York, NY, USA, 2018. ACM. event-place: Boston, MA, USA.

[HMU24] Florian Holzbauer, Markus Maier, and Johanna Ullrich. Destination reachable: What icmpv6 error messages reveal about their sources. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24)*, pages 1–15. ACM, 2024.

[HSU25] Florian Holzbauer, Sebastian Strobl, and Johanna Ullrich. Tracking internet disruptions in ukraine: Insights from three years of active full block scans. In *Proceedings of the 2025 ACM Internet Measurement Conference*, IMC '25, page 474–492, New York, NY, USA, 2025. Association for Computing Machinery.

[Hub18] Bert Hubert. Herding the dns camel, November 2018. Accessed: 2025-12-04 from `https://www.ietf.org/blog/herding-dns-camel`.

[HULF22] Florian Holzbauer, Johanna Ullrich, Martina Lindorfer, and Tobias Fiebig. Not that simple: Email delivery in the 21st century. In *2022 USENIX Annual Technical Conference (USENIX ATC 22)*, pages 295–308, 2022.

[KDS20] G. Kambourakis, G. Draper, and I. Sanchez. What Email Servers Can Tell to Johnny: An Empirical Study of Provider-to-Provider Email Security. *IEEE Access*, 8:130066–130081, 2020.

[KFR+95] J. Klensin, N. Freed, M. Rose, E. Stefferud, and D. Crocker. SMTP Service Extensions. RFC 1869 (Internet Standard), November 1995. Obsoleted by RFC 2821.

[LLZ+21] Xiang Li, Baojun Liu, Xiaofeng Zheng, Haixin Duan, Qi Li, and Youjun Huang. Fast ipv6 network periphery discovery and security implications. In *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 88–100. IEEE, 2021.

[LNM25] Gianluca Lo Nostro and Leo Marchandon. Ukraine's kyivstar adds backup power as russian strikes strain power grid, ceo says. *Reuters*, November 2025. Accessed: 2026-01-13.

[Lon22a] Timo Longin. Dns-analysis-server: Tools to assess dns security, 2022. Accessed: 2026-01-11 from `https://github.com/The-Login/DNS-Analysis-Server`.

[Lon22b] Timo Longin. Melting the dns iceberg: Taking over your infrastructure kaminsky style, October 2022. Accessed: 2026-01-05 from `https://sec-consult.com/blog/detail/melting-the-dns-iceberg-taking-over-your-infrastructure-kaminsky-style`.

[MU23] Markus Maier and Johanna Ullrich. In the loop: A measurement study of persistent routing loops on the ipv4/ipv6 internet. *Computer Networks*, 221:109500, 2023.

[MYM] Yoshiro Yoneya Masanori Yajima, Daiki Chiba and Tatsuya Mori. How prevalent is the operation of DNS security mechanisms? Accessed: 2021-09-15 from `https://indico.dns-oarc.net/event/39/contributions/867`.

[Net25] Network Startup Resource Center. Routeviews, 2025. `https://doi.org/10.7264/1Y7V-2D90`.

[Pos81] J. Postel. Internet Control Message Protocol. RFC 792 (Internet Standard), September 1981. Updated by RFCs 950, 4884, 6633, 6918.

[Pos82] J. Postel. Simple Mail Transfer Protocol. RFC 821 (Internet Standard), August 1982. Obsoleted by RFC 2821.

[QHP13] Lin Quan, John Heidemann, and Yuri Pradkin. Trinocular: Understanding internet reliability through adaptive probing. *ACM SIGCOMM Computer Communication Review*, 43(4):255–266, 2013.

[RB19] Philipp Richter and Arthur Berger. Scanning the scanners: Sensing the internet from a massively distributed network telescope. In *Proceedings of the Internet Measurement Conference*, pages 144–157, 2019.

[Reu22] Reuters. Russian troops damage kherson tv centre, infrastructure - reports. *Reuters*, November 2022. Accessed: 2026-01-13.

[RIP] RIPE NCC. Routing information service (ris). Accessed: 2025-12-04 from `https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris`.

[RL23] Erik Rye and Dave Levin. Ipv6 hitlists at scale: Be careful what you wish for. In *Proceedings of the ACM SIGCOMM 2023 Conference*, pages 904–916, 2023.

[SKZ+23] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. Target acquired? evaluating target generation algorithms for ipv6. In *Proceedings of the Network Traffic Measurement and Analysis Conference (TMA)*, June 2023.

[ZSS+22] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. Rusty clusters? dusting an ipv6 research foundation. In *Proceedings of the 2022 Internet Measurement Conference*, New York, NY, USA, 2022. ACM.