Open access • Proceedings Article • DOI:10.1145/3098822.3098858

# The Impact of Router Outages on the AS-level Internet — Source link

Matthew Luckie, Robert Beverly

**Institutions:** University of Waikato, Naval Postgraduate School

**Published on:** 07 Aug 2017 - ACM Special Interest Group on Data Communication

**Topics:** Core router, One-armed router, Router and Default-free zone

# The Impact of Router Outages on the AS-level Internet

Matthew Luckie
University of Waikato
mjl@wand.net.nz

Robert Beverly
Naval Postgraduate School
rbeverly@nps.edu

## ABSTRACT

We propose and evaluate a new metric for understanding the dependence of the Internet on *individual* routers. Whereas prior work uses large volumes of reachability probes to infer outages, we design an efficient active probing technique that directly and unambiguously reveals router restarts. We use our technique to survey 149,560 routers across the Internet for 2.5 years. 59,175 of the surveyed routers (40%) experience at least one reboot, and we quantify the resulting impact of each router outage on global IPv4 and IPv6 BGP reachability.

Our technique complements existing data and control plane outage analysis methods by providing a causal link from BGP reachability failures to the responsible router(s) and multi-homing configurations. While we found the Internet core to be largely robust, we identified specific routers that were *single points of failure* for the prefixes they advertised. In total, 2,385 routers – 4.0% of the routers that restarted over the course of 2.5 years of probing – were single points of failure for 3,396 IPv6 prefixes announced by 1,708 ASes. We inferred 59% of these routers were the customer-edge border router. 2,374 (70%) of the withdrawn prefixes were not covered by a less specific prefix, so 1,726 routers (2.9%) of those that restarted were single points of failure for at least one network. However, a covering route did not imply reachability during a router outage, as no previously-responsive address in a withdrawn more specific prefix responded during a one-week sample. We validate our reboot and single point of failure inference techniques with four networks, finding no false positive or false negative reboots, but find some false negatives in our single point of failure inferences.

## CCS CONCEPTS

• **Networks** → **Network measurement**; **Network reliability**;

## KEYWORDS

Internet reliability, single points of failure, BGP, routing

**ACM Reference format:**
Matthew Luckie and Robert Beverly. 2017. The Impact of Router Outages on the AS-level Internet. In *Proceedings of SIGCOMM '17, Los Angeles, CA, USA, August 21-25, 2017,* 14 pages.
https://doi.org/10.1145/3098822.3098858


*SIGCOMM '17, August 21-25, 2017, Los Angeles, CA, USA*

ACM ISBN 978-1-4503-4653-5/17/08…$15.00
https://doi.org/10.1145/3098822.3098858

## 1 INTRODUCTION

While the Internet is well-established as critical infrastructure, its complex ecosystem of systems, users, applications, and service providers make an accurate assessment of resilience difficult. At the infrastructure substrate, networks may utilize routing protocols and multiple routers, links, and providers to take advantage of available diversity, and to gain redundancy and resilience to failure. In this paper, we take a new step towards understanding Internet reliability by examining *individual router* reboots and measuring the broader global impact of these outages.

Graph-theoretic approaches to understanding and characterizing resilience [1, 12] are notoriously fragile [31, 49] due to the underlying difficulty in obtaining accurate Internet topologies [11, 23, 44]. While a router may appear central to a network's connectivity, significant redundancy may be invisible to measurement. Despite a pressing need for complete and sound data on Internet infrastructure resilience, the inherently global, unregulated, complex, and decentralized nature of the Internet makes obtaining such data challenging. Attempts to quantify the network's resilience have largely come from researchers attempting to infer outages, attacks, and their causes from noisy and incomplete data. For instance, prior work has shown that continual, high-volume probing of edge devices can expose more reachability failures than control-plane analysis alone [39, 43], while [17] uses control plane analysis to identify failures and tomography to isolate their causes.

In contrast, we seek to empirically quantify the dependence of interdomain routing on individual routers. We employ an efficient active probing technique that elicits IPv6 fragment identifiers from routers. These 32-bit fragment identifiers are commonly monotonic, *unambiguously and directly* exposing router reboots when there is a discontinuity in the sampled sequence [7]. While our restart inference technique is only effective for routers with an IPv6 control plane, we show in §5 that the restart events we find also frequently manifest as IPv4 BGP churn and withdrawals due to the preponderance of IPv4/IPv6 infrastructure sharing [16].

We survey 149,560 distinct routers over 2.5 years. We associate each router with a set of destinations (IPv4 and IPv6 network prefixes) reachable via that router, as indicated by analysis of CAIDA's traceroute topology data [10]. We opportunistically correlate each observed router outage with BGP control plane activity – the sequence of globally visible BGP update and withdrawal messages for the associated prefixes observed at the Routeviews looking glass – to shed light on the impact of more than 2.5 years of such events on global IPv4 and IPv6 BGP reachability.

In addition to localizing BGP reachability failures to responsible routers and configurations, we identify specific routers that are *single points of failure* for their networks. Surprisingly, we find that 4.0% of the surveyed routers were relied on by their AS to announce a prefix in BGP. In summary, our contributions include:

(1) A new perspective on characterizing the importance of individual routers on Internet resilience.
(2) An Internet-wide large-scale measurement campaign to identify and characterize router restarts over 2.5 years.
(3) Correlation of restarts against ≈12M BGP update messages to quantify their global IPv4 and IPv6 reachability impact.
(4) Identification of 2,385 individual routers in the global topology that represented single-points of failure for 1,708 ASes.
(5) Week-long high-frequency active probing of networks behind rebooting routers to assess the impact of the reboot on the data-plane.
(6) Public release of the source code for our adaptive probing engine as part of scamper [32].

The remainder of this work is organized as follows. Section 2 provides background and context, while §3 gives an overview of the extensive prior work in measuring and mitigating network outages. Section 4 details our methodology and data, while §5 provides the results of our extended, Internet-wide measurement study. We conclude in §6 with a discussion on the broader implications of our findings and suggestions for future research.

## 2 BACKGROUND

We start with background on our technique for inferring router restarts, a discussion of our use of BGP in understanding the effect of router restarts, as well as a summary of the myriad ways in which networks interconnect, deploy redundancy, and attempt to provide resilience. Based on the different multi-homing configurations, we discuss the expected effects of a router restart to provide context for our results in §5.

### 2.1 Router Restarts

In contrast to prior efforts that indirectly localize outages, blackholes, and reachability faults (e.g., via traceroutes, topology atlases, tomography, etc), we directly interrogate the control-plane IPv6 stack on routers in order to unambiguously determine whether the router has restarted. At a high-level, we construct a per-router time series of IP Identifier (IPID) values from fragmented IPv6 response packets, and use discontinuities in the series to infer that the router restarted some time during the polling interval.

We focus on IPv6 routers for two reasons. First, as we show in §4.1.1, probing routers affords significant efficiencies as compared to probing edge systems, where even finding a stable set of edge IPv6 addresses presents a practical difficulty. Second, probing IPv6 routers allows us to unambiguously determine that a router rebooted, rather than relying on fragile tomography inferences. Only IPv6 routers provide a clean signal that allows us to reliably infer that a restart occurred. While the IPv4 stack on routers also has an IPID field, it is both small (16-bits) and has natural velocity since every control packet the router originates must have an IPID [5] in case the packet is fragmented in the network. In contrast, the IPv6 IPID field is 32-bits in size with no natural velocity because routers rarely send fragmented packets, and only the sender can fragment a packet. The small size of the IPv4 IPID field, combined with the fact that a router may send packets in bursts due to operational and control traffic, makes it infeasible to distinguish an IPv4 router restart from a counter wrap.

The IPID field is an optional IPv6 extension header. To obtain fragmented responses, we use our method from [7] and send 1300-byte ICMP6 echo request packets to router interfaces; upon receipt of an ICMP6 echo reply larger than 1280 bytes, we respond with an ICMP6 packet too big message asking the receiver to reduce its packet size to 1280 bytes. This process induces the router to fragment subsequent ICMP6 echo replies, exposing the router's IPv6 IPID counter values. In our data, 27.1% of router interfaces that responded for at least 14 days during our 2.5 years of probing assigned IPID values from a counter; others assigned values from a pseudo random number generator, did not send fragmented responses to our probes, had fragments blocked along the reverse path, or ICMP packet too big messages were blocked along the forward path. A responsive router with a monotonic sequence that resets therefore represents a restart. Section 4.1 provides our full algorithm for inferring a router outage.

### 2.2 BGP

The Border Gateway Protocol version 4 (BGP) is the protocol used by ASes to organize interdomain routing [40, 41]. To quantify the impact of an inferred router outage, we examine the sequence of BGP messages received at the Routeviews looking glass. In this section, we discuss subtleties of BGP that manifest in the BGP update and withdrawal messages we can observe, and how they can impact our ability to correlate BGP activity with router outages.

An AS exchanges network-layer reachability information with neighboring ASes according to the configuration applied by the router's operator. An AS establishes a session with a neighbor using TCP, and each peer keeps state regarding which prefixes are reachable; a neighbor may announce new routes, withdraw existing routes, or update routes.

**Ending a BGP session:** A router may end a BGP session gracefully by notifying the peer that it is closing the session, or by closing the TCP connection. A graceful shutdown occurs when an operator halts, reboots, or changes the configuration of a router to remove the peer. A router will not shutdown a BGP session gracefully if the neighbor router crashes or fails, or if a physical link between the peers is broken and the router does not detect the link failure. A router infers that a peer has failed if it does not receive a BGP keep-alive message from the peer within the *hold time* period agreed by the peers using the BGP protocol [41]. The BGP specification suggests that routers send keep-alive messages at 30 second intervals, and wait 90 seconds after receiving a keep-alive message before deciding the BGP session has failed. However, the Cisco default is to send keep-alive messages at 60 second intervals, and wait 180 seconds before deciding the BGP session has failed. A router may infer a BGP session with a peer has failed more quickly if the operator configures a shorter hold time for the peer, or if the operator configures Bidirectional Forwarding Detection [25], which allows the router to verify the peer is responsive at the network layer.

**Path hunting:** When a router's BGP session with a peer fails or is gracefully shutdown, the router removes the routes it received from that peer from its Routing Information Base (RIB). The router then runs a decision process to select alternative routes for the withdrawn routes that were previously the best available for the prefixes. If the router has no other route available for a prefix, it

sends a withdrawal notice to its peers, causing them to remove the route from their RIBs and to run the BGP route decision process. Depending on the availability of alternative routes, routers may begin path hunting or path exploration, where a router replaces the withdrawn route with what it believes to be a valid route with a different path. However, if the replacement route is via a path affected by the same failure, the replacement route may also be subsequently withdrawn or updated as routing information converges. This activity typically induces a burst of withdrawal and update messages to propagate through the Internet, so-called "churn."

**Route flap damping:** Excessive BGP control traffic and churn is a source of instability in the network if BGP routers are unable to process BGP control traffic at the same rate it arrives. The BGP specifications [40, 41] recommend that BGP routers use a Minimum Route Advertisement Interval (MRAI) of 30 seconds for routes exchanged between ASes. The initial BGP specification recommends that the MRAI not apply to withdrawn routes [40], however an update to the specification includes withdrawn routes in the MRAI [41]. In practice, the MRAI is applied per-neighbor, rather than per prefix. To further reduce the impact of unstable routes sustained over long time-scales, the IETF standardized a protocol for route flap damping [47]. A router that uses route flap damping (RFD) scores each prefix according to recent activity, penalizing prefixes that contribute significant BGP load by suppressing updated routes for the prefix at increasingly long durations, up to one hour. Individual network operators make their own decisions about the configuration and deployment of MRAI and RFD.

### 2.3 Network Redundancy

Network redundancy is implemented in a wide variety of ways, and the resulting control-plane artifacts and failure-modes can be both subtle and complex. A customer, which we informally define as an entity at the network edge that pays for access from a service provider, can have redundant i) routers; ii) links; and iii) providers. This redundancy may be utilized via load-balancing, only during failures, or deployed in more complex configurations (e.g., via selective advertisement of more specific prefix announcements) [3]. In this work, we focus on the role routers play in providing network resilience, and ignore failures of links connecting routers (e.g., a severed long-haul optical link).

The types of customer connectivity range from the cross-product of single versus multiple customer edge routers, single versus multiple provider edge routers, and the use of a single or multiple providers. In the most basic case, no redundancy exists: a single customer edge router connects to a single provider edge router, and requires neither BGP nor provider-independent address space. We expect outages of these customer edge routers to have no impact or discernible effect on the global BGP control-plane.

The more interesting case, and the focus of this paper, involves a customer with their own AS and address space. Regional Internet Registries (RIRs) require a multi-homing plan for an organization to obtain an AS number. Perhaps surprisingly, we uncover a non-trivial number of singly-homed customers in this class. An outage of such a customer or their provider's router will manifest as a series of withdrawal and update messages in the global BGP system during path exploration. Eventually, we observe all Routeviews peers withdrawing the customer's prefix. A significant component of our work is determining, for a given customer's network prefix, the provider and customer edge routers (see §4.2).

More complex scenarios invoke more subtle behaviors. While we do not know the cause of a router reboot, a single router outage should not cause a complete prefix withdrawal if the network is multi-homed with at least two routers. For instance, we observe significant BGP churn for customers with multiple routers connected to either a single or multiple provider. The presence of an available alternate path is implied by BGP activity at a remote looking glass stabilizing after the outage as BGP converges to an alternate path. We therefore classify the resultant per-prefix global BGP activity due to a router outage as either: i) none; ii) churn; iii) partial withdrawal, where a fraction of the peers withdraw; or iv) complete withdrawal, where all peers withdraw such that the looking glass has no available path (see §4.3).

## 3 RELATED WORK

Significant prior work examines network resilience in general, and Internet outages in particular. Our research extends this body of prior work in several crucial ways, by: i) providing wide breadth and scope, including 2.5 years worth of Internet-wide measurements; ii) taking a first-step toward understanding the impact of router outages; and iii) empirically identifying routers that represent single points of failure for their connected customers. In this section, we provide an abbreviated taxonomy of prior work and its relation to our research.

Empirical work on outages, reachability, and failure includes passive and active measurement at both the control plane and data plane. These techniques have been applied within a single AS, across vantage points, and across the entire Internet.

For instance, passive analysis, clustering, and temporal grouping of BGP routing messages can reveal insights into events of interest [50] and their origins [20, 45]. As purely passive techniques are most effective with a complete view of all routing activity within a single AS [22], localizing and inferring causes of outages outside of a single domain is challenging [9]. Dainotti *et al.* [15] employ a novel technique that synthesizes two sources of passive data: network telescopes and control-plane activity. Their work shows that backscatter observed at their network telescope can correlate with BGP activity and known macro-level outages [6].

Both Trinocular [39] and Pingin' in the Rain [43] send continual data-plane probes to the network edge from dedicated measurement nodes to discover reachability issues, while Choffnes *et al.* enlist BitTorrent nodes to crowdsource measurements[13]. These systems reveal insights into the real-world availability experienced by end nodes, and demonstrate that reachability correlates with natural disasters, severe weather, and known political events. However, these methods require significant volumes of probing traffic and cannot directly implicate a particular router as the root cause.

Prior work showed promise toward fusing data and control plane analysis. For instance, Feamster *et al.* [19] and Wang *et al.* [48] utilized continual active measurements between a mesh of dedicated measurement nodes, and correlated BGP activity observed with reachability problems. Our work takes an Internet-wide view of outages, without requiring a full-mesh of measurement nodes.

A large body of research investigated, localized, and proposed methods to mitigate network reachability blackholes – instances of end-to-end path failures that are avoidable, in that a policy-compliant route exists but is not used by the forwarding path [14, 17, 26, 27, 29]. Many of these works used tomography algorithms to locate faults. Rather than examining silent reachability faults, our work centers on router outages. While Iannaccone *et al.* [24] provided a detailed analysis of router failures within the Sprint backbone, and [46] examined availability in a regional network via logfile analysis, our focus is on Internet-wide identification of router outages and their impact.

In this work we extend our method in [7] to perform direct and unambiguous measurement of IPv6 router restarts, rather than relying on error-prone inferences. Whereas [7] simply characterized the prevalence of observed router reboots via coarse-grained probing every six hours, we implement an adaptive probing algorithm based on each router's behavior to enable efficient fine-grained sampling. This study is significantly broader in scope and duration. Further, we close the causal loop by tying observed reboots with the resulting impact (or lack of impact) on BGP – thereby providing a new way to concretely differentiate critical (i.e. single point of failure) routers from more resilient configurations.

## 4 METHODOLOGY

Our methodology involves addressing three distinct aspects of router outage characterization:

(1) **Identifying router restarts (§4.1):** We utilize macroscopic traceroute data to identify IPv6 router interfaces, a novel adaptive-rate active probing technique that identifies router restarts, and IPv6 alias resolution techniques to reduce interfaces to routers.
(2) **Associating networks with routers (§4.2):** We again use traceroute data to find interfaces along the forward path to each network prefix. We then label routers with their relative distance from the customer edge border router. We use a similar procedure to map routers to the IPv4 prefixes for which they were responsible.
(3) **Associating router restarts with BGP activity (§4.3):** Finally, we correlate global BGP data, both RIBs and BGP messages, from the Routeviews looking glass as part of a large-scale data-fusion effort. We juxtapose these three sources of data to correlate inferred outages with their impact on the global routing system.

Figure 1 depicts the components of our methodology for fusing large uptime IPID time series with Routeviews BGP updates and traceroute data in order to identify single points of failure.

### 4.1 Inferring Router Outages

Our router outage inference consists of two components: i) active router probing; and ii) identifying reboots from the responses.

*4.1.1 Probing.* Beverly *et al.* performed a five month uptime study of 66,471 IPv6 interfaces [7], and sent six probes every six hours regardless of the nature of previous responses. In this work, we use macroscopic CAIDA IPv6 traceroute topology data [10] to identify a much larger set of IPv6 router interfaces as probe targets, which mandates a more adaptive probing approach. Bandwidth spent probing interfaces that produce no or random responses is effectively wasted. However, over the 2.5 year duration of our

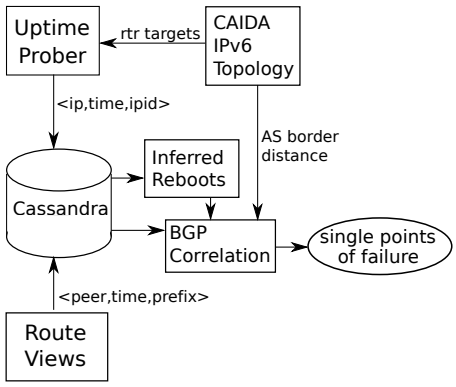



**Figure 1: Method overview: correlation between inferred reboots of AS border routers and BGP updates identify routers that are single points of failure for prefixes they advertise.**

study, we found the way interfaces respond can change, e.g., from predictable to random or vice-versa; while we wish to minimize probing of interfaces that are not amenable to probing, we must periodically check them.

Therefore, we implemented an optimized prober that regularly samples interfaces that respond with monotonically increasing IPID sequences, and periodically samples interfaces that do not. Our prober operates in rounds, and solicits a single fragmented response from each interface that responds with a monotonic sequence. If the value in the IPID field is less than the previous value our prober received from the same interface, our prober sends six further probes to determine if the router is still assigning IPID values from a counter. If an interface that was sending a monotonic sequence goes silent, our prober continues to solicit a single fragmented response for up to two hours, before scheduling the interface for probing less frequently.

The measurement parameters we used during this study evolved over the 2.5 year period. To start with, we used a static list of 83,393 interface addresses that appeared in CAIDA IPv6 traceroutes for December 2014, and probed these at 100pps from 18 January 2015 until 18 October 2016. Systems that did not respond with an incrementing IPID value for more than two hours were subsequently probed every 12-24 hours. From 18 October 2016 until 24 February 2017, we used a static list of 1,086,047 addresses observed in CAIDA's June to October 2016 traceroute data. To maintain a 15-minute round, we increased our probing rate to 225 pps. Finally, on 24 February 2017, we transitioned to a more dynamic probing algorithm, where we added addresses newly observed in traceroute to the list every two hours, and to further reduce our probing burden, we probed addresses that did not send incrementing IPID values every 7-14 days. By 30 May 2017, the list had grown to ≈2.4M addresses.

The second and third sets are much larger than the first because they mostly include interfaces in two IPv6 /32 prefixes operated by the same access network; this portion of the router IP address list is almost entirely (99%) unresponsive to our probes, and are likely home gateways rather than an ISP's BGP-speaking routers. Table 1 summarizes our measurement parameters, and figure 2 shows the size of our probe list and the number of interfaces that sent incrementing IPID values over time.

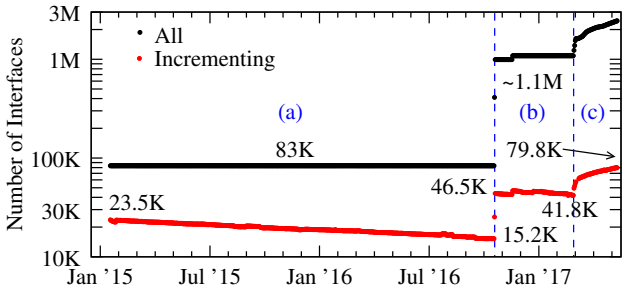



**Figure 2: Number of interfaces in probing set, and number of interfaces that respond with an incrementing IPID value, over time. We label the three different measurement parameter periods defined in table 1.**

| | |
|---|---|
| (a) | **18 Jan 2015 – 18 Oct 2016**<br>Static list: 83,393 interfaces seen in CAIDA traceroutes for Dec 2014.<br>Probing rate: 100 pps.<br>Probing of not applicable addresses: every 12-24 hours. |
| (b) | **18 Oct 2016 – 24 Feb 2017**<br>Static list: 1,086,047 interfaces seen in CAIDA traceroutes for June to October 2016, and addresses in (a).<br>Probing rate: 225 pps.<br>Probing of not applicable addresses: every 12-24 hours. |
| (c) | **24 Feb 2017 – 30 May 2017**<br>Dynamic list: obtained from CAIDA traceroutes every two hours, and addresses in (b).<br>Probing rate: 200 pps.<br>Probing of not applicable addresses: every hour for 24 hours, then every 7-14 days. |

**Table 1: Measurement parameters for acquiring interface addresses, and how the prober handles interfaces that do not send incrementing IPID values (i.e. interfaces where our method is not applicable).**

The prober runs continuously in a tight loop such that it begins a new round of probing as soon as the prior round finishes. The prober shuffles the set of addresses to probe in each round, and maintains inter-round state in an SQLite database. The prober sends packets asynchronously and maintains a constant packet rate. We conducted all probing in this work from a single host on a well-provisioned, native IPv6 academic network in the United States.

By directly probing routers, rather than edge systems or networks, our adaptive prober achieves significant efficiency. For example, on December 30th 2016, we sent approximately 17.4M packets toward 1,086,055 unique router interfaces. Given our 1300 byte probe packet size, this equates to ~ 200pps (~ 2Mbps). By comparison, existing approaches to outage inference, e.g., [39], require two orders of magnitude higher packet per second probing rate.

Further, methods that directly measure the edge require a relatively stable set of known responsive nodes. Whereas prior work has successfully built such "hitlists" for IPv4 [39], the exponentially larger IPv6 address space, combined with the popular use of private, temporal IPv6 addresses [36] presents practical difficulties in simply obtaining a set of edge IPv6 hosts to probe [38]. By directly probing IPv6 routers (which can much more easily be discovered via traceroute), we sidestep the stable IPv6 hitlist problem entirely.

**Algorithm 1** findOutage($id[], tx[]$)

```
v ← 0                                                    ▷ Velocity
min_id ← 0
reboots ← {}
for i ∈ |id|:
    v' = (id[i] − id[i − 1])/(tx[i] − tx[i − 1])
    v = 0.8v + 0.2v'
    E[spin] = (tx[i] − tx[i − 1]) ∗ v
    if id[i] < id[i − 1]
        if (|id[i] − min_id| < 2^16) ∧ (id[i] > 2^16)
            Cyclic
        else if rand_seq(id[i, i + 10])
            Random
        else
            reboots ← reboots ∪ tx
    else if id[i] > id[i − 1] + E[spin] ∗ 10
        if rand_seq(id[i, i + 10])
            Random
        else
            min_id ← id[i]                               ▷ Cyclic reboot
            reboots ← reboots ∪ tx
    if id[i] < min_id
        min_id ← id[i]
return reboots
```

*4.1.2 Reboot Inference.* Our probing (§4.1.1) produces approximately 10M IPID samples per day. To facilitate efficient range-based queries, we place all time series data, including IPIDs and BGP updates, in an Apache Cassandra NoSQL database [30], with the target and timestamp as primary keys.

Algorithm 1 describes how we infer a router reboot from an IPID time series. If the router appears to assign IPID values from a counter, then a discontinuity in the time series implies the counter reset. Recall that the IPv6 IPID is a 32-bit value, and the counter only increments for fragments, so overflow is rare. An IPID less that the previous IPID from the interface indicates a potential reboot, however this non-monotonicity can also be due to either random or cyclic counters.

Some routers based on Linux produce time series that are cyclic; these routers maintain a counter state per host to which they send fragmented packets. The counter state is initialized to value $X$ derived from a secret cryptographically generated at system boot, and the destination host's IP address. When the router removes the per-host state and then re-creates the state when we probe it next, the router re-initializes the IPID state to $X$ – we term these "cyclic" interfaces. We infer a cyclic reboot when the router chooses a different value $X'$ to initialize the per-destination state, a consequence of the router's secret changing when it rebooted.

A cyclic reboot may result in an IPID initialized to an $X$ which is larger than the previous IPID sent by the router. To avoid a false positive due to interfaces with a large natural IPID velocity, Algorithm 1 computes $E[spin]$, the expected amount of IPID change given a weighted moving average of the historical IPID velocity. If

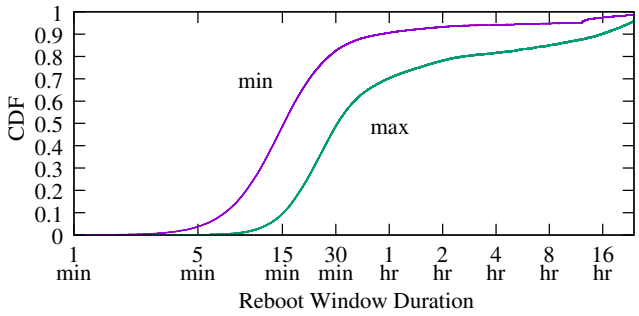



**Figure 3: CDF of minimum and maximum outage window lengths measured. Half of the maximum outage windows were shorter than 31 minutes.**

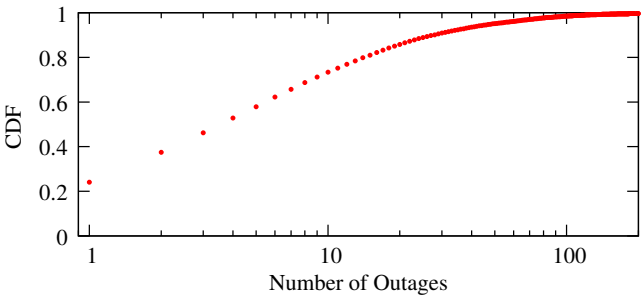



**Figure 4: CDF outages per router for the 59,175 routers that experienced an outage. 71% had less than ten outages during our probing.**

the discontinuity is larger than an order of magnitude more than the expected change, and the sequence is non-random, we infer a cyclic reboot.

Our method allows us to infer an outage window – a period of time in which the router restarted. Figure 3 shows the minimum and maximum outage window lengths for each address that was monotonic or cyclic during our probing. At least half of the maximum outage windows were shorter than 31 minutes, though another 22% of the windows were at least 2 hours, and 4% of the maximum outage windows in our dataset were more than 24 hours. Figure 4 shows the number of router outages inferred per router, for those routers that had at least one inferred outage. We inferred 24% of these routers had one outage during our measurements, 53% had four outages, and 71% had fewer than ten outages during our probing. These short and infrequent outages is consistent with these devices operating as critical infrastructure.

Figure 5 shows the daily number of rebooting routers. There is some decay in the number of reboots we inferred per day between January 2015 and February 2017, as our router sample was mostly static over this time period, and some routers became unresponsive. Because routers rarely restart owing to their nature as critical infrastructure, a low-rate probing balances our ability to capture outages from the probing cost in doing so.

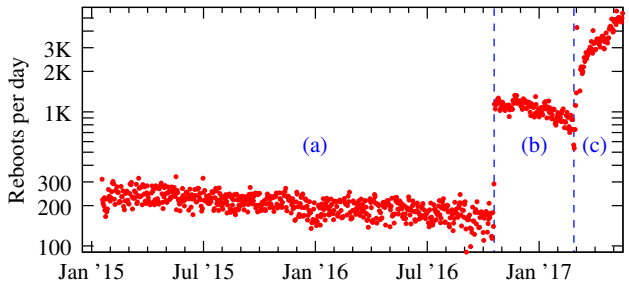



**Figure 5: Number of inferred outages, per day. We inferred ≈200 per day between January 2015 and October 2016, and ≈1K per day after we updated the probe list in October 2016, and up to ≈5K per day after we began dynamically updating the probe list in February 2017. We label the three different measurement parameter periods defined in table 1.**

## 4.2 Associating networks with routers

The second fundamental problem is determining which networks are routed by which routers. We again use CAIDA traceroutes to build a router graph that associates each interface in the graph with destination prefixes the interface is in the path towards. For instance, assume interface $I$ is observed in the traceroute to destination $D$. We find the longest matching prefix $P$ to which $D$ belongs, and note that $I$ is involved in providing reachability to $P$.

In December 2016, the Ark project used a set of 55 IPv6 vantage points distributed around the world, each of which probed two addresses in each routed prefix, the first (::1), as well as a randomly generated address. Naturally, the closer $I$ is to $P$, the more $P$ depends on $I$ (and, conversely, interfaces close to our vantage points are weakly associated with a large number of prefixes). Therefore, we also labeled each interface with the distance, in IP hops, the interface was from the AS originating the prefix.

Figure 6 illustrates our distance computation. We define distance as the number of IP hops a router interface found by traceroute is from the customer edge border router for the destination probed by traceroute. The first router belonging to the destination AS has a distance of zero, routers with negative distance are within the destination AS; routers with positive distance are before the destination AS. The first router in the destination AS will often appear in traceroute using an address assigned by their provider for the interconnecting link. Therefore, we infer the destination AS border router in traceroute is the hop prior to the first hop that used an address BGP-originated by the destination AS.

## 4.3 Correlating with BGP Control Plane

We used the University of Oregon's Routeviews archived routing table snapshots and update messages to provide a BGP-level view of routing. We ingested into a Cassandra time series database RIB snapshots at midnight UTC for each day from 18 January 2015 to 30 May 2017, as well as all update messages between each RIB snapshot. We used all peers providing data to the "routeviews6" collector, where multiple Tier-1 IPv6 ISPs provide a global view of routing from their perspectives, including AT&T, Verizon, Hurricane Electric, CenturyLink, GTT Communications, and NTT.

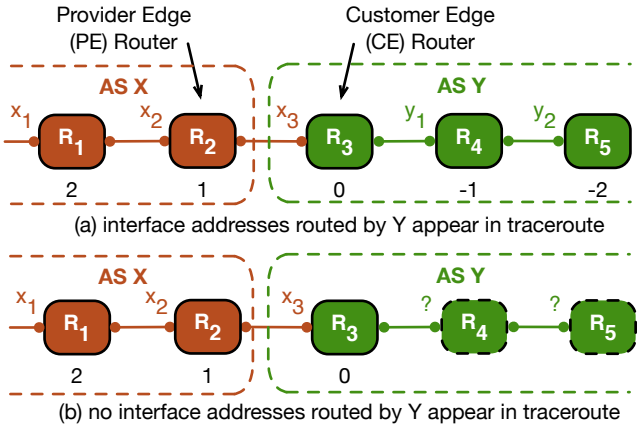



**Figure 6: Computing distance of router reboot from AS announcing route in BGP. A router with distance zero is likely to be the AS's border router, a router with negative distance is within the AS, and a router with positive distance is outside the AS.**

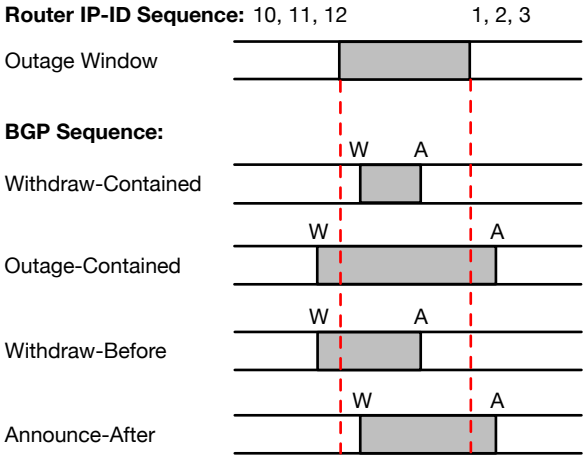



**Figure 7: Correlating router outages with BGP control plane (W)ithdrawal and (A)nnouncements. More router outage windows in our data contain withdrawals, rather than withdrawals before or announcements after outage windows.**

A router outage can manifest itself in the BGP control plane in four different ways, depending on the number of routes advertised by that router to neighboring BGP routers that were selected as the best path. Because of the information-hiding properties of BGP, where only the best route is advertised to external peers, a router outage topologically close to the advertised prefix is more likely to result in control plane activity because it is more likely to be in the best path for neighboring ASes. In this work, we consider four event types, and we empirically correlate them in time and distance from the AS announcing the route.

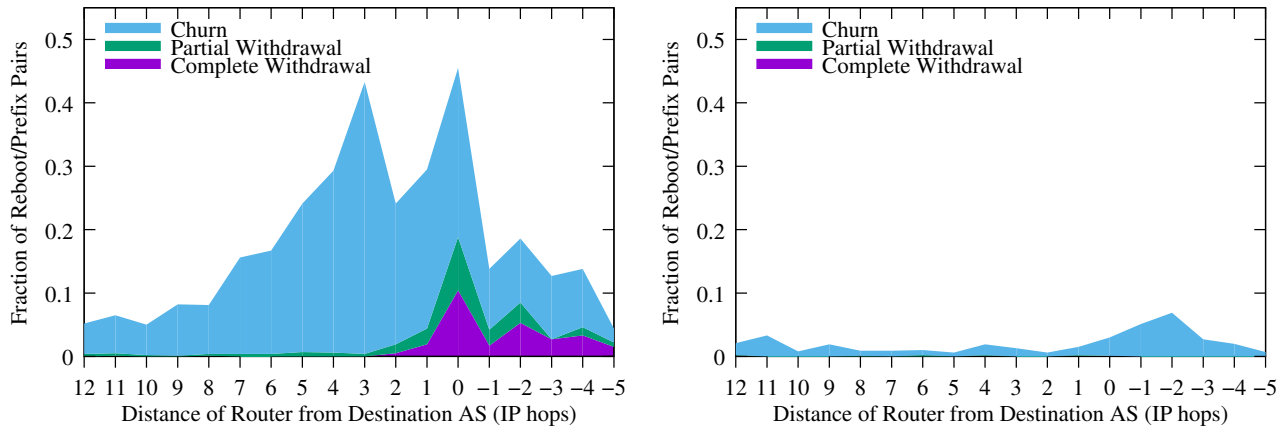
**Event Types:** First, a *complete withdrawal* occurs when a router that has an outage is present in all external BGP paths towards the prefix. In this work, we classify a prefix as being completely withdrawn if no vantage point has a route for 70 consecutive seconds. A complete withdrawal occurs when the router is a single point of failure for that prefix. Second, a *partial withdrawal* occurs when a router that has an outage does not cause all VPs to withdraw the route, but at least some do withdraw, each for 70 consecutive seconds. A partial withdrawal can occur when some VPs carry the route, but other VPs do not converge on the same path because they did not receive it through the export policy of other ASes. In this work, we also observed individual Routeviews looking glass peers carrying routes while all others withdrew the prefix for multiple hours, masking a complete withdrawal. Third, *churn* occurs when a router that has an outage is present in the path for some peers, but those peers were able to converge on an alternative route during the outage. Finally, *no BGP activity* can occur during a router outage when other equal cost paths are available with the same AS path, so no BGP control message is required.

**Time Effects:** Figure 7 shows the four ways that a router outage can correlate in time with a BGP withdrawal. First, and most common in our data, the withdrawal event is contained within the router outage window. Specifically, the time that the last peer withdraws is after the left edge of the outage window and the time that the first peer announces is before the right edge of the outage window. This is the most common effect in our data, as most withdrawals more than 70 seconds in length were still less than the approximate 15-minute probing intervals of our prober. We believe these short withdrawals are mostly caused by operators upgrading the software and rebooting the router.

Second, *outage-contained* occurs when the operator shuts down all BGP sessions with the router's peers before restarting the router, and the sessions come back up after we detected the router outage ended. Because of our 15-minute granularity of probing, we likely missed some outage-contained correlations. Third, *withdraw-before* occurs when the operator shuts down all BGP sessions with the router's peers before restarting the router, we probe the router before it restarts, and the router restarts BGP sessions with neighbors before we probe the router again. Finally, *announce-after* occurs when the first announcement after the outage occurs after we detected the router had restarted.

**Distance Effects:** Figure 8 shows the effect of an inferred router outage on BGP control plane activity for outages occurring during February 2015. We counted each reboot/prefix pair once, to reduce the effect that a few routers with frequent outages could have. The closer the router is to the destination AS, the more likely the outage will result in BGP churn and prefix withdrawal. In particular, 10% of border routers operated by a destination AS resulted in a complete withdrawal. We investigate the nature of these outages thoroughly in §5; we include figure 8 here to provide evidence that our technique yields reasonable inferences.

Figure 8 also includes withdrawal and churn measurements for the six-hour period *before* the router outage for the same outage window length, as a control. In the control figure, there is substantially less BGP activity for the affected prefixes, and there were no withdrawals of the prefixes lasting at least 70 seconds. We therefore believe the elevated level of churn and withdrawals were caused by the inferred router outages.




**(a) During Router Outage** **(b) Six Hours Before Router Outage**

**Figure 8: Impact of router outages on BGP routing during an inferred router outage as a function of distance the router is from the border router of the destination AS, and a control of six hours before the outage for the same duration, during February 2015. A router outage was more likely to correlate with churn and withdrawal the closer the router is to the border of the AS.**

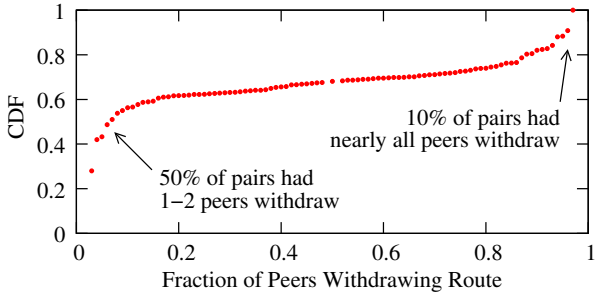



**Figure 9: The fraction of Routeviews peers that withdrew a prefix when not all peers withdrew. 10% of router-prefix pairs had nearly all peers withdraw the prefix.**

Figure 9 shows the fraction of peers that withdrew a prefix when not all peers withdrew a prefix. As with figure 8, we only plot each router/prefix pair once. For half of the pairs, only 1-2 peers withdrew a prefix; however, 10% of pairs had nearly all peers withdraw the prefix, but did not meet the criteria for a full withdrawal.

## 4.4 Limitations

Measurement of network resilience at the router-level is fraught with challenges because the available data is imperfect. In this section, we discuss the most important challenges and limitations.

**Limited to IPv6.** Our IPID reboot inference is limited to probing routers with an IPv6 interface. While IPv4 packets also contain an IPID field, and router implementations frequently reset the IPID counter to zero and populate the IPID field monotonically, there are two fundamental problems with using the IPID for IPv4 reboot inference. First, whereas IPID is an optional extension header in IPv6, IPID is a required IPv4 header value, so every IPv4 control-plane packet requires a unique value. Many routers therefore have high IPID velocity due to their participation in routing protocols and network management. Second, whereas the IPID is a 32-bit counter in IPv6, it is only 16-bits in IPv4. High-velocities combined with small counters imply that we would have to frequently sample the IPv4 counter to avoid misinterpreting a counter-wrap as a reboot. While the IPID velocity for a particular router may typically be low, any probing system must handle large bursts of control traffic, for instance when the router is exchanging routing information, or when the router is polled by a management system. Conservatively assuming a maximum control-plane transmission rate of 100Mbps and 1500 byte packets, the Nyquist rate dictates that each router must be probed approximately once every four seconds. While prior work has shown the feasibility of high-rate IPID probing [8], the scale required for Internet-wide uptime monitoring precludes IPv4 probing. Further, such a probing rate is effectively equivalent to high-rate reachability probing, and yet it would still be impossible to differentiate any period of unreachability from an actual reboot.

**Requires incrementing IPID.** Even when restricting our technique to routers with IPv6 interfaces, our method only works for those routers that respond and send fragments with an IPID value assigned from a counter. Therefore, our method is only able to detect a fraction of all router outages in the Internet. In prior work, we found that routers manufactured by Cisco, Huawei, Vyatta, HP, and Mikrotik returned sequential fragment identifiers [33]. However, routers manufactured by Juniper return random identifiers.

**Complex events could be masked.** Our method might incorrectly associate a complex event, such as a power outage causing multiple simultaneous router outages, with a single router. While we can detect multiple overlapping router outages when the routers involved all assign IPID values from a counter, we were unable to detect this complex event if only one router involved assigns IPID values from a counter. Further, we were only able to probe routers observed in traceroute; other routers only on a backup path that simultaneously failed were therefore not probed.

**BGP dynamics can mask single points of failure.** Correctly inferring a router is a single point of failure is challenging due to the complexities in topology, routing, and configuration. As discussed in §2, BGP routers can use a MRAI timer to reduce churn during path exploration by not announcing a route for a prefix more regularly than the timer, and operators may deploy route flap damping to further suppress updates. Routers may use a MRAI timer of 30 seconds, though large transit networks often reduce the

| Network | Correct | | Incorrect | | Not Validated | |
|---|---|---|---|---|---|---|
| | R | S | R | S | R | S |
| US University | 7 | 7 | 0 | 0 | 8 | 8 |
| US R&E backbone #1 | 2 | 3 | 0 | 2 | 3 | 0 |
| US R&E backbone #2 | 3 | 0 | 0 | 0 | 1 | 4 |
| NZ R&E backbone | 11 | 4 | 0 | 2 | 22 | 27 |
| Total: | 23 | 14 | 0 | 4 | 34 | 39 |

**Table 2: Summary of validated reboot windows (R) and single point of failures (S). Most reboot events were not validated due to the difficulty of confirming events prior to the most recent.**

timer to allow faster convergence. In this work, we inferred that a Routeviews peer had no route if the peer did not report a route for the prefix for at least 70 seconds, which allows for two default MRAI timer values, as well as a 10-second fudge. Router outages that completed more quickly than 70 seconds would be classified as churn, understating the importance of the router to that prefix.

**Infrequent update of destination prefixes.** Prior to February 2017, the set of destination prefixes used to seed CAIDA's IPv6 traceroute data collection was infrequently updated – every few months, rather than continuously as new IPv6 prefixes were announced. Our process to associate network prefixes to routers (§4.2) requires traceroutes towards an address in the prefix in order to establish a dependence between a router that appears in a traceroute and a corresponding prefix.

## 4.5 Validation

We contacted operators at six different networks for validation of our inferences. We focused on operators of research and education (R&E) networks, as we believed they would be willing to provide feedback on our work. Further, we believed they would be willing to refer us to their customer network operators, in the cases where only the customer could supply ground truth on the root cause of the outage. Five operators responded, and four provided feedback; the fifth declined to provide feedback without permission of the members involved, and the sixth did not respond to our request.

We asked the operators to validate that our reboot inferences were correct (true positives), and that we did not miss any reboots (false negatives). We also asked the operators to validate our single point of failure inferences; for routers where we did not detect any full withdrawals, we asked them to confirm the routers were not a single point of failure. Table 2 summarizes the validation outcome. The validation shows the technique is able to correctly infer router reboot windows when the router assigns IPID values from a counter. Where the operators were able to, they confirmed we did not miss any outages; however, the operators had difficulty confirming outages prior to the last outage. Our technique is also usually able to correctly infer dependence of a prefix on a router.

**US University:** We contacted the network operator of a large campus network for feedback on six different routers that had rebooted. In total, we detected 15 reboots associated with the routers, and no prefix withdrawals correlated with any of the reboots. They used SNMP queries for the sysUpTime counter to validate four of the six outage windows, and logged into two routers where the counter was unable to validate because the counter had rolled over after 497 days of uptime [35].

They had no reasonable way to confirm inferences prior to the last reboot of the six routers we identified, with the exception of one core router, where they validated two reboot events. The routers were all internal to their network, and provided connectivity to small subnets of the larger IPv6 address space. The operator validated our inference that the routers were not a single point of failure for the larger prefixes advertised in BGP. All of the previous six reboot inferences were the last time the routers in question had rebooted; the seventh reboot inference was validated through personal knowledge of the operator.

**US R&E backbone #1:** We discussed five router reboot events involving four routers with a US state-level R&E network backbone. They explained that larger universities have two routers and redundant connections to the R&E backbone, but the smaller members have a single connection to a single router in the R&E backbone. They used their trouble ticket system to provide validation and context on the reboots we inferred.

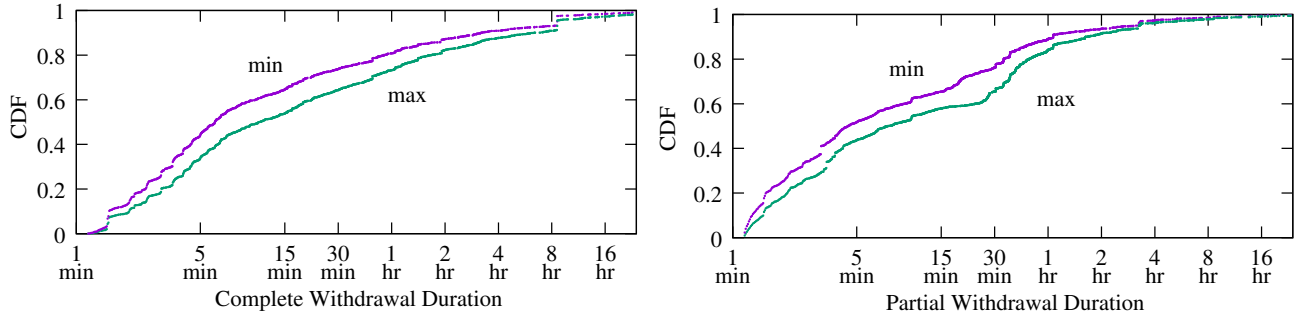
For the first router, they reported their provider-edge router within a University data center was down for more than two hours as part of a data center-wide power outage. They confirmed the reboot, and that the router was a single point of failure for a prefix that their customer announced in BGP. However, because all but one Routeviews peer withdrew the prefix, our system did not classify the router as a single point of failure. During the outage window, the single Routeviews peer periodically reported updates for the route, though these updates could not have come from the provider-edge router, as the router was powered off at the time.

For the second router, they confirmed a customer router that was a single point of failure for one prefix likely rebooted within the inferred outage window, as their BGP session with the router restarted within the window. They reported the BGP session was quickly re-established; while all 20 Routeviews peers withdrew the prefix for at least some period of time, only three were without a route for more than 30 seconds, and our system did not classify the router as a single point of failure.

For the third router, they reported the router was within their customer's network, and they believed it was not a single point of failure for any prefixes, which is congruent with the absence of BGP activity during this time. However, they did not have information that would allow them to validate the two reboots involving the device. They were unable to provide any information about a fourth router that rebooted more than two years prior, but that we were correct in not inferring a single point of failure for any prefix.

**US R&E backbone #2:** We sent the operator of a second US R&E backbone a list of 11 routers with interfaces using addresses from within their BGP-announced prefixes. Due to the labor involved in searching their systems for archived SNMP data, they only provided feedback for four of the routers, three of which we inferred to be a single point of failure in BGP for at least one prefix. While they did not provide any comment on the inferred single points of failure, they did provide feedback on router reboot events.

For the first router, they had no record of a reboot occurring during the inferred window. They had a record for a crash four days earlier, however we had not been probing the router during that time. Nevertheless, the reboot was correlated with the complete withdrawal of a prefix.




**(a) Complete Withdrawals** **(b) Partial Withdrawals**

**Figure 10: CDF of minimum and maximum withdrawal durations, per router/prefix pair, for complete and partial durations.**

For the second router, they reported that it was one of their routers, and confirmed that they had an alert for the router during the window, and that it had probably rebooted. They provided three further events for the router; however, the router had stopped responding to our probes before the subsequent events, so we were not able to infer any behavior for the router.

For the third router, they reported the interface was on their customer's router, and that the link went down to the router during the outage window. This event was correlated with the detected complete withdrawal of a customer prefix.

For the fourth router, they confirmed the reboot we reported was a known reboot event. They provided another eight other events that might also have been reboot events, however the IPID sequence from the router suggests they were not reboot events.

**NZ R&E backbone:** We discussed 32 reboot events across 8 routers with the operator of a New Zealand R&E backbone, and approached two of their members for further elaboration. The operator confirmed they had BGP outages correlated with reboot events archived in their monitoring infrastructure for four routers we inferred as having rebooted once each. They were unable to help with five events for two other routers, as the outages were in January and April 2015 when they had a different network architecture.

We approached the operators of two other routers individually. The seventh router had five reboots inferred; the operator had configured external data-plane monitoring at five-minute intervals, and the reboots correlated with four data-plane outages. The operator was unable to confirm the fifth outage because the monitoring granularity was too wide. The operator confirmed the router was a single point of failure for a more specific prefix they announced in BGP using the backbone's ASN. Even though there was a covering prefix in BGP announced by the backbone, there was no reachability to the addresses covered by the more specific.

The eighth router had 18 reboot events inferred across the 2.5 years of probing. However, the operator could only confirm the final two reboots in 2017, as they had no data allowing them to validate prior events. The router was a single point of failure in BGP for a single prefix; however, we had not observed the router in traceroute paths in 2017, so while we correctly inferred the reboots, we had not inferred the router was a single point of failure.

**Summary:** In total (table 2), we validated 23 reboot events with no false positives or false negatives, and 14 single point of failures with four false negatives, owing to limitations discussed in §4.4.

# 5 RESULTS

In total, we inferred 749,451 reboots involving 59,175 (40%) of 149,560 responsive routers between January 18th 2015 and May 30th 2017. As discussed in §4, most of the detected router outages were shorter than 2 hours (figure 3) and most routers that restarted did so fewer than two times (figure 4). Importantly, we were only able to correlate 2,385 routers – 4.0% of the routers that had outages in our dataset – with a prefix being completely withdrawn, after excluding more complex events involving multiple routers overlapping in time (§5.5). Figure 8 shows that, where necessary, BGP was able to converge on alternative paths for the majority of router outages, and outages that caused a complete withdrawal were confined to the edge of the network – either topologically adjacent to the affected network, or within the affected network. In our data, we inferred that 59% of all single points of failure were the customer-edge border router, 8% were the provider-edge border router, and 29% were within the destination AS itself. These properties are consistent with previously published work describing the likely behavior of networks based on a first-principles approach that included router-level topologies from two academic networks [31]; our work is the first to measure the actual impact of router outages on the Internet's BGP routing system at scale.

Among the 2,385 single point of failure routers, we used Maxmind [34] to geolocate the ::1 address of the IPv6 prefixes involved in these failures to 90 different countries. 25% of the prefixes geolocate to the United States, 13% to Brazil, 8% to Great Britain, 6% to India, and 2-3% each to Russia, Australia, Indonesia, Germany, and Ukraine. This geodistribution largely mirrored the overall distribution of countries in our monitored dataset, implying that outages were no more prevalent in any given country than another.

## 5.1 Prefix Withdrawal Duration

Figure 10a shows the distribution of complete prefix withdrawal durations for each router/prefix pair. For each prefix that was withdrawn, we computed the minimum and maximum duration the prefix was withdrawn in BGP. 50% of the minimum durations were less than 5 minutes in length, suggesting most router outages were short lived, and likely caused by a router restart due to maintenance. However, the tail is long, as 30% of prefixes were withdrawn for at least 30 minutes.

Similarly, figure 10b shows the distribution of partial prefix withdrawal durations for each router/prefix pair. As with the duration of

the full withdrawals, the partial withdrawals were short: 50% were less than five minutes. As with the complete withdrawals, the tail is long, and manual investigation of the longer withdrawal durations, particularly when nearly all peers withdrew the prefix (figure 9) revealed apparent bugs in BGP implementations. For example, we confirmed with one R&E network operator that an inferred multi-hour router outage was a single point of failure for their customer, despite the fact that one Routeviews peer retained a route to the prefix, and periodically reported updates for the route. However, these updates could not have come from the provider-edge router, as the router was powered off for data center maintenance.

## 5.2 Presence of Covering Prefixes

The 2,385 routers that we were able to correlate with a prefix withdrawal were in paths toward 3,396 prefixes. Of these 3,396 withdrawn prefixes, 1,022 (30%) were covered by a less specific prefix that was not simultaneously withdrawn, suggesting that the more specific prefix that was withdrawn could be present in the routing table for traffic engineering purposes. Indeed, 699 (68%) of the more specific prefixes had a covering prefix announced by the same AS. The 2,374 prefixes that were not covered by a less specific prefix when they were withdrawn were correlated with 1,726 routers that had an outage – 2.9% of the routers that had outages in our dataset.

## 5.3 Impact on AS-level Reachability

The list of routers we probed was updated at the end of October 2016, coinciding with CAIDA's probe list being updated, so for the following analysis we used router outages inferred during November and December 2016, as we have traceroute paths for all IPv6 routed prefixes in October 2016. This allows us to correlate all announced prefixes with routers, rather than the subset probed before October 2016. In total, we were able to infer prefix withdrawals correlated with router outages for 149 ASes during these two months; 82 of these ASes (55%) were completely unrouted in BGP during these router outages. The routers were single points of failure for these 82 ASes.

## 5.4 Impact of AS-level Multi-homing

To better understand where single points of failure reside, we used a January 16, 2017 Routeviews RIB to construct the IPv6 AS-level graph. We then found the degree of the ASes corresponding to the IPv6 prefixes involved in the identified single points of failure. Figure 11 shows the cumulative fraction of ASes originating prefixes that represented single points of failure as a function of AS degree. For context, we also plot the AS degree distribution for the monitored population, which includes IPv6 router interfaces that responded with fragment identifiers in the final 24 hours of our dataset. Here, we use longest prefix matching against the same Routeviews RIB to map each router interface to the AS originating the corresponding IPv6 prefix.

As seen in Figure 11, 21% of the routers in the monitored population belong to degree-one stub ASes, while 42% of the single points of failure are in degree-one stub ASes. Although we intuitively expect stub ASes to be less resilient to single router failures, it is interesting to note that we observe single points of failure in non-stub ASes. For instance, we found that 16% of the ASes that contained

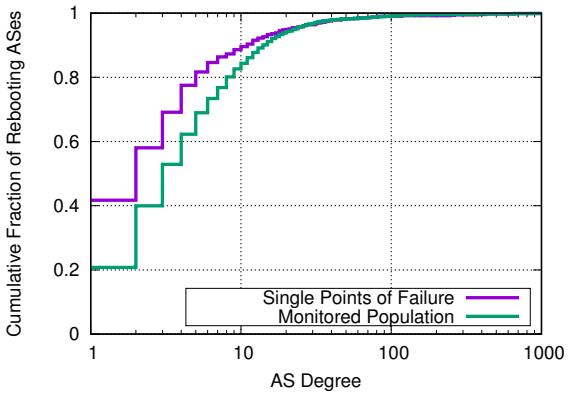



**Figure 11: Cumulative distribution of ASes as function degree. While only 21% of the routers in the monitored population belong to degree one ASes, 42% of the single points of failure are in degree one ASes.**

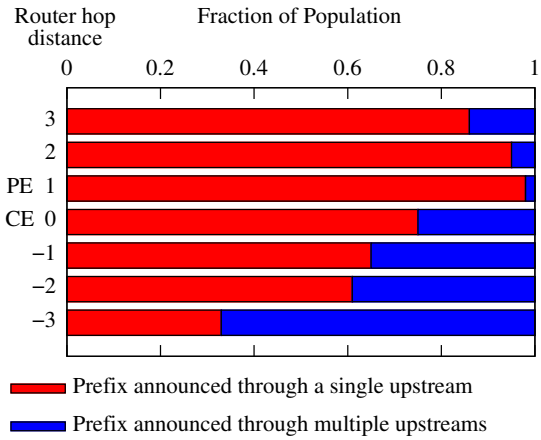
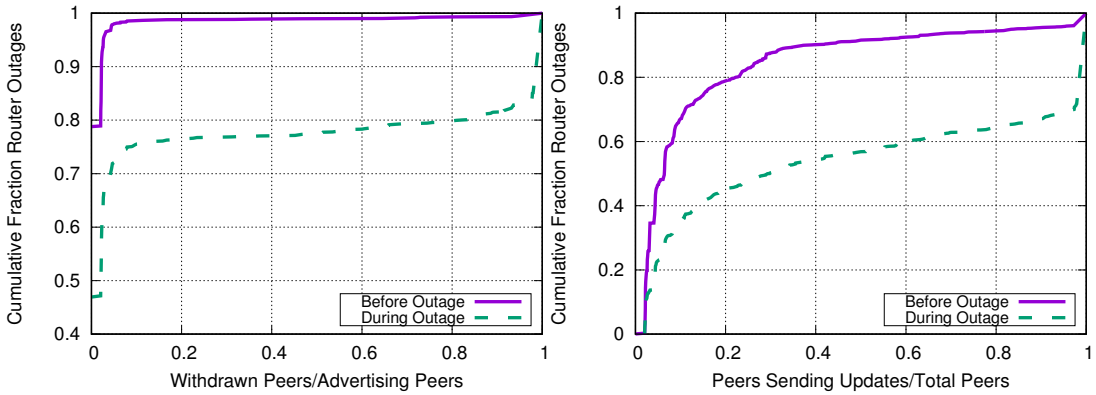



**Figure 12: The effect of AS-level multi-homing on prefix/router pairs. In our data, prefixes that were completely withdrawn as a result of a router outage outside the network (distances 1-3) were nearly exclusively propagated through a single upstream AS.**

routers representing single points of failure had degree-two – suggesting that a single router peered with two different providers. Further, while the fraction of high-degree ASes was small, the instances of single points of failure we found in these ASes is likely attributable to exchange points and other highly interconnected infrastructure. These findings underscore the fact that analyzing the AS-level topology without the router-level interconnection context is insufficient to show a network is resilient to failure.

Figure 12 shows the effect of AS-level multi-homing on prefix/router pairs for our 2.5 years of router outage data. For each router outage where the router was within three IP hops of the AS announcing the prefix, we classified the BGP event type according to whether the prefix was announced via a single upstream or via multiple upstreams. In our data, prefixes that were completely withdrawn as a result of a router outage outside the network (distances 1-3) were nearly exclusively propagated through a single upstream




**(a) Max number of peers concurrently withdrawing** **(b) Number of peers sending an update**

**Figure 13: Correlating router outages with IPv4 prefix withdrawals and churn: distribution of the fraction of Routeviews peers for the associated IPv4 prefix during versus outside the inferred outage window.**

AS. A router outage at distance 1, the provider-edge (PE) router, implied the connecting AS was single-homed to a single PE router.

Router outages at distance 0, the customer-edge (CE) router, usually (63.5% of outages) implied the prefix was announced through a single upstream AS. However, the remaining 36.5% of outages where a CE router experienced an outage were BGP-announced through multiple upstream ASes, implying the same CE router was used to connect through multiple upstreams. Even though the affected ASes had provider diversity at the AS-level, our router-level measurements still inferred their CE router to be a single point of failure, further showing the benefit of synthesizing BGP and router-level data when analyzing the resilience of ASes in BGP.

## 5.5 Overlapping Router Outages

In our data, some routers that were in the path towards the same prefix had outages that overlapped in time, and we filtered these events from our analysis of single points of failure. There are multiple possible explanations for these events, including power outages that impact multiple neighbors, and operator router maintenance activities that overlap. Further, it is possible the two routers had non-overlapping outages, but our probing granularity was not fine-grained enough to capture this effect. Because router outages were rarely correlated with a prefix withdrawal if they were more than three hops from the AS announcing the prefix, we focused on overlapping router outages no further than three hops from the destination AS, as well as router outages within the destination network. In total, we inferred 865 overlapping router outages that we correlated with 619 prefixes.

We examined the adjacency of these routers using the distance metric we described in §4.2. Only 15% of these events involved routers the same distance from the edge of the destination AS: 30% and 27% of impacted routers were separated by one and two hops, respectively, implying a localized outage. We detected two routers with overlapping outage windows in 64% of the events; 33% had three separate routers involved. We emphasize the fraction of complete withdrawal events containing overlapping outage windows is a small fraction of the overall set, consistent with the relatively small number of short events per day (figures 3 and 5).

## 5.6 Correlation with IPv4 Outages

As the IPv6 Internet matures, its topology grows increasingly congruent with the IPv4 Internet [16]. To understand the extent to which IPv6 router restarts impacted IPv4, we examined 28637 reboots experienced by 2665 IPv6 interfaces – a ~ 5% sample of all interfaces where we inferred at least one reboot. We associated IPv6 router interfaces with a set of IPv4 prefixes as follows. For each IPv6 interface, we determined the origin AS of the IPv6 prefixes associated with that interface in §4.2. Using daily RIB snapshots from Routeviews, we found all IPv4 prefixes announced by that origin AS on that day. Using the method described in §4.3, we examined the set of BGP update messages received at the Routeviews looking glass to determine the activity for these IPv4 prefixes during the time period of the IPv6 router restart. Our algorithm links IPv4 prefixes to the router most likely to originate the prefix. Thus, if the router suffers an outage and represents an IPv4 single point of failure, we expect to lose reachability to the prefix.

We examined two distinct time periods for each interface in our analysis: one within our inferred outage window, and a second period, of the same duration, two days prior to the outage. Using the daily RIB IPv4 snapshots, we determined the number of Routeviews peers advertising the prefix, and then replayed the BGP update messages to maintain state over the number of peers concurrently advertising or not advertising the IPv4 prefix. To determine the extent of the outage, we used the ratio of the maximum number of peers that withdrew the route to the number of peers that had been advertising the route.

Figure 13a shows the distribution of router outages as a function of ratio of affected IPv4 Routeviews peers, before and during the inferred outage. 47% of the within-window outages resulted in no peers withdrawing the IPv4 route, suggesting that these prefixes were robust to the failure of single routers. However, we observed a distinct correlation between withdrawals during the outage as compared to outside the outage. Similarly, Figure 13b shows distinctly more update churn for the IPv4 prefixes during the outage window. While none of the IPv4 prefixes we examined were fully withdrawn outside the outage, approximately 10% of the outages resulted in more than 90% of the peers withdrawing the prefix.

### 5.7 Impact of Outages on Application Services

A consequence of an outage is that both the customers of the affected provider, as well as the application-layer services they provide to the Internet, become unreachable. While there do not exist good IPv6 address census data due to the sheer size of the address space, we make use of the daily IPv4 censys.io scanning data [18] as an indicator of the outage impact for a single point of failure.

In particular, we examined those IPv4 prefixes tied to an inferred IPv6 router reboot (§5.6) where at least 90% of the Routeviews peers withdrew that prefix in the inferred outage time window. For each IPv4 address in an April 2017 censys.io daily scan, we used a Routeviews BGP table to find the longest matching prefix and determine whether the host belonged to a prefix that experienced an outage. For the IPv4 prefixes that experienced an outage in our sample, we found 39,107 active hosts in the censys data, 25,592 hosts listening on port 80 (HTTP), 16,321 hosts listening on port 443 (HTTPS), 7383 SMTP listeners, 5127 IMAP servers, 11,277 SSH servers, and 7922 responsive DNS servers. While this methodology yields only a rough approximation of an outage impact, it shows that non-trivial numbers of applications were affected by the single points of failure we discovered.

### 5.8 Impact of Outages on the Data Plane

To better understand how our work complements end-host data-plane probing approaches, we sought to quantify the relationship between outages, data-plane reachability, and BGP activity. We identified responsive IPv6 targets within each of the 41K globally routed prefixes by sending ICMP6 echo requests to the first 16 addresses within each prefix. From this simple survey, we discovered at least one active target within 11,003 prefixes (~27% of the 41K total prefixes in May 2017). We then probed a single address in each of the 11,003 prefixes at 30-second intervals between May 23-30 2017. To counter the effect of single probe loss, we sent a second probe if there was no response to the first probe.

Then, for each router outage inferred during the week of probing, we examined the impact, if any, on data-plane responsiveness. Our binary metric of correlation between end-host data-plane probing and inferred router outages determined if: (1) the loss of active probes overlapped with the inferred router outage window; (2) the router was responsive before and after; and (3) the responsive address was not an alias of the router that rebooted, so that we were probing an end-host reachable via the router, rather than another interface on the router itself.

We first examined the set of router outages over the May 23-30 2017 window inferred via IPID sequence discontinuities that identified a single point of failure, i.e. those outages that resulted in a BGP withdrawal. While we might expect a loss of data-plane reachability due to the BGP withdrawal, the withdrawn prefix may be a more specific announcement that belonged to a larger aggregate (a "covering prefix") that was not withdrawn.

In total, there were 36 distinct router outage / complete BGP withdrawal pairs that we correlated with loss of active probes; 13 of these had a covering prefix that was not simultaneously withdrawn. While our sample size is small, we found no evidence that a covering IPv6 prefix provided data-plane reachability when a more specific prefix was withdrawn due to a router outage.

Second, we examined inferred router outages during our data-plane probing window that were not clear single points of failure; that is, the outage did not result in a complete BGP withdrawal for the associated prefixes. We found 138 router outages that correlated with a loss of active probes to unique prefixes, despite the prefix remaining in the global BGP table. Thus, while our technique did not deem these routers to be BGP-level single points of failure, our probing identified them as data-plane single points of failure.

Our result that nearly four times more router outages were correlated with active probe loss than with complete BGP withdrawal is congruent with prior work. Using Hubble [26], Katz-Bassett *et al.* reported that the majority of reachability problems they found were uncovered with active probes, with reachability problems often not correlated with BGP activity for the prefix. Similarly, Trinocular [39] finds that control-plane measurements underestimate outages.

However, by tying network outages to individual router outages, our technique complements existing control-plane and data-plane methods – whereas data-plane probing can identify reachability issues, it cannot accurately identify the root-cause of the outage. In future work, we plan to more closely couple data-plane probing with our router outage detection framework.

## 6 CONCLUSION

The resilience of the Internet to individual points of failure has been argued in the literature for nearly two decades, and there has been significant work to undermine the theory that the Internet has high-degree hubs crucial to overall network connectedness [1, 12, 31, 49]. However, there has been a dearth of empirical data allowing researchers and policymakers to understand AS-level reliability cognizant of the underlying router-level interconnection. In this paper, we take a first empirical step towards understanding the resilience of the Internet to individual router outages through the opportunistic correlation of router outages with BGP routing information. In our data, collected from a survey of 149,560 IPv6 routers responsive to our method, we inferred that 59,175 (40%) had an outage during our study. Only 2,385 routers (4.0%) were correlated with complete withdrawals involving 3,396 prefixes where the routers appeared in traceroute paths towards the prefixes. Further, 2,374 (70%) of the withdrawn prefixes were not covered by a less specific prefix, so only 1,726 routers (2.9%) of those that restarted were BGP-level single points of failure for at least one network, and routers that were single points of failures were over-represented in stub ASes. We were also able to correlate IPv6 router outages with IPv4 control plane instability, reinforcing that while our study was applied to IPv6 routing out of methodological need, failures were present in IPv4 as well.

Our method and data also has a network security impact, and therefore also ethical considerations [37]. During the course of our Internet-wide measurement campaign over 2.5 years, which we conducted at a low rate, we received a single query from a network operator about our measurements, suggesting that the method could also be used by an attacker in a stealthy manner. The security implications of predictable fragment identification values are well known [21], as they provide, for example, a side-channel that allows an adversary to infer the number of systems behind a middle-box [4], and the packet sending rate and open ports of

an end-host [2]. To improve privacy, some Linux kernels use a hash-based algorithm to select an initial identifier value, and then assign identifier values sequentially from the initial value. However, Knockel *et al.* were able to leverage an artifact of the implementation to infer if two third-party computers were communicating [28]. In this work, we also show the fragment identifier side channel can be used to infer the impact that the failure of a single router has on the Internet's routing system.

Our method and data has utility beyond our initial study of network resilience. Our method and data could also be used, for example, to study the deployment of BGP configurations that limit routing convergence, data plane performance during routing convergence, study the evolution of network resilience, develop actionable information for national CERT bodies to assist network operators to deploy more resilient routing, and to cross-validate other outage detection methods. In order to balance the research utility of our dataset with the possibility it can be used to do harm until operators have deployed routers without the side channel, we will release our raw data to researchers via a suitable legal framework [42].

## ACKNOWLEDGMENTS

We thank kc claffy, Young Hyun, Daniel Andersen, the network operators who helped us with validation, our shepherd Ratul Mahajan, and the anonymous reviewers for their feedback. Views and conclusions are those of the authors and should not be interpreted as representing the official policies or position of the U.S. government. The CAIDA Ark infrastructure is supported by NSF CNS-1513283 and DHS S&T/CSD HHSP233201600010C.

## REFERENCES

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. 2000. Error and attack tolerance of complex networks. *Nature* 406 (June 2000).
[2] antirez. 1998. new tcp scan method. (1998). http://seclists.org/bugtraq/1998/Dec/79.
[3] T. Bates and Y. Rekhter. 1998. *Scalable Support for Multi-homed Multi-provider Connectivity.* RFC 2260.
[4] Steven M. Bellovin. 2002. A Technique for Counting NATted Hosts. In *IMW.*
[5] Adam Bender, Rob Sherwood, and Neil Spring. 2008. Fixing Ally's Growing Pains with Velocity Modeling. In *IMC.*
[6] Karyn Benson, Alberto Dainotti, kc claffy, and Emile Aben. 2013. Gaining insight into AS-level outages through analysis of Internet Background Radiation. In *INFOCOM Workshops.*
[7] Robert Beverly, Matthew Luckie, Lorenza Mosley, and kc claffy. 2015. Measuring and Characterizing IPv6 Router Availability. In *PAM.*
[8] Timm Böttger, Félix Cuadrado, Gareth Tyson, Ignacio Castro, and Steve Uhlig. 2016. *Open Connect Everywhere: A Glimpse at the Internet Ecosystem through the Lens of the Netflix CDN.* Technical Report.
[9] Randy Bush, Olaf Maennel, Matthew Roughan, and Steve Uhlig. 2009. Internet Optometry: Assessing the Broken Glasses in Internet Reachability. In *IMC.*
[10] CAIDA. 2016. The CAIDA UCSD IPv6 Topology Dataset. (2016). http://www.caida.org/data/active/ipv6_allpref_topology_dataset.xml.
[11] Nikolaos Chatzis, Georgios Smaragdakis, Jan Böttger, Thomas Krenc, and Anja Feldmann. 2013. On the benefits of using a large IXP as an Internet vantage point. In *IMC.*
[12] Qian Chen, Hyunseok Chang, Ramesh Govindan, and Sugih Jamin. 2002. The origin of power laws in Internet topologies revisited. In *INFOCOM.*
[13] David R. Choffnes, Fabián E. Bustamante, and Zihui Ge. 2010. Crowdsourcing Service-level Network Event Monitoring. In *SIGCOMM.*
[14] Ítalo Cunha, Renata Teixeira, Nick Feamster, and Christophe Diot. 2009. Measurement Methods for Fast and Accurate Blackhole Identification with Binary Tomography. In *IMC.*
[15] Alberto Dainotti, Claudio Squarcella, Emile Aben, kc claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. 2011. Analysis of country-wide Internet outages caused by censorship. In *IMC.*
[16] Amogh Dhamdhere, Matthew Luckie, Bradley Huffaker, kc claffy, Ahmed Elmokashfi, and Emile Aben. 2012. Measuring the Deployment of IPv6: Topology, Routing and Performance. In *IMC.*
[17] Amogh Dhamdhere, Renata Teixeira, Constantine Dovrolis, and Christophe Diot. 2007. NetDiagnoser: Troubleshooting Network Unreachabilities Using End-to-end Probes and Routing Data. In *CoNEXT.*
[18] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *CCS.*
[19] Nick Feamster, David G. Andersen, Hari Balakrishnan, and M. Frans Kaashoek. 2003. Measuring the Effects of Internet Path Faults on Reactive Routing. In *SIGMETRICS.*
[20] Anja Feldmann, Olaf Maennel, Z. Morley Mao, Arthur Berger, and Bruce Maggs. 2004. Locating Internet Routing Instabilities. In *SIGCOMM.*
[21] F. Gont. 2016. *Security Implications of Predictable Fragment Identification Values.* RFC 7739.
[22] Yiyi Huang, Nick Feamster, Anukool Lakhina, and Jim (Jun) Xu. 2007. Diagnosing Network Disruptions with Network-wide Analysis. In *SIGMETRICS.*
[23] B. Huffaker, M. Fomenkov, and k. claffy. 2012. *Internet Topology Data Comparison.* Technical Report. Cooperative Association for Internet Data Analysis (CAIDA).
[24] Gianluca Iannaccone, Chen-nee Chuah, Richard Mortier, Supratik Bhattacharyya, and Christophe Diot. 2002. Analysis of Link Failures in an IP Backbone. In *IMW.*
[25] D. Katz and D. Ward. 2010. *Bidirectional Forwarding Detection (BFD).* RFC 5880.
[26] Ethan Katz-Bassett, Harsha V Madhyastha, John P John, Arvind Krishnamurthy, David Wetherall, and Thomas E Anderson. 2008. Studying Black Holes in the Internet with Hubble. In *NSDI.*
[27] Ethan Katz-Bassett, Colin Scott, David R. Choffnes, Ítalo Cunha, Vytautas Valancius, Nick Feamster, Harsha V. Madhyastha, Thomas Anderson, and Arvind Krishnamurthy. 2012. LIFEGUARD: Practical Repair of Persistent Route Failures. In *SIGCOMM.*
[28] Jeffrey Knockel and Jedidiah R. Crandall. 2014. Counting Packets Sent Between Arbitrary Internet Hosts. In *USENIX FOCI.*
[29] R. R. Kompella, J. Yates, A. Greenberg, and A. C. Snoeren. 2007. Detection and Localization of Network Black Holes. In *INFOCOM.*
[30] Avinash Lakshman. 2008. Cassandra. (2008). http://cassandra.apache.org/.
[31] Lun Li, David Alderson, Walter Willinger, and John Doyle. 2004. A First-Principles Approach to Understanding the Internet's Router-level Topology. In *SIGCOMM.*
[32] Matthew Luckie. 2010. Scamper: a Scalable and Extensible Packet Prober for Active Measurement of the Internet. In *IMC.*
[33] Matthew Luckie, Robert Beverly, William Brinkmeyer, and k claffy. 2013. Speedtrap: Internet-scale IPv6 Alias Resolution. In *IMC.*
[34] MaxMind. 2017. GeoIP2City. (2017). http://www.maxmind.com/.
[35] K. McCloghrie and M. Rose. 1991. *Management Information Base for Network Management of TCP/IP-based internets: MIB-II.* RFC 1213.
[36] T. Narten, R. Draves, and S. Krishnan. 2007. *Privacy Extensions for Stateless Address Autoconfiguration in IPv6.* RFC 4941.
[37] Craig Partridge and Mark Allman. 2016. Ethical Considerations in Network Measurement Papers. *CACM* 59, 10 (Oct. 2016).
[38] David Plonka and Arthur Berger. 2015. Temporal and Spatial Classification of Active IPv6 Addresses. In *IMC.*
[39] Lin Quan, John Heidemann, and Yuri Pradkin. 2013. Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *SIGCOMM.*
[40] Y. Rekhter and T. Li. 1995. *A Border Gateway Protocol 4 (BGP-4).* RFC 1771.
[41] Y. Rekhter, T. Li, and S. Hares. 2006. *A Border Gateway Protocol 4 (BGP-4).* RFC 4271.
[42] Charlotte Scheper and Susanna Cantor. 2015. PREDICT: An Important Resource for the Science of Security. In *HotSoS.*
[43] Aaron Schulman and Neil Spring. 2011. Pingin' in the rain. In *IMC.*
[44] Neil Spring, Ratul Mahajan, and David Wetherall. 2003. Measuring ISP topologies with Rocketfuel. In *SIGCOMM.*
[45] Renata Teixeira and Jennifer Rexford. 2004. A Measurement Framework for Pin-pointing Routing Changes. In *SIGCOMM NetTs Workshop.*
[46] Daniel Turner, Kirill Levchenko, Alex C. Snoeren, and Stefan Savage. 2010. California fault lines: understanding the causes and impact of network failures. In *SIGCOMM.*
[47] C. Villamizar, R. Chandra, and R. Govindan. 1998. *BGP Route Flap Damping.* RFC 2439.
[48] Feng Wang, Zhuoqing Morley Mao, Jia Wang, Lixin Gao, and Randy Bush. 2006. A Measurement Study on the Impact of Routing Events on End-to-End Internet Path Performance. In *SIGCOMM.*
[49] Walter Willinger, David Alderson, and John C Doyle. 2009. Mathematics and the Internet: A source of enormous confusion and great potential. *Notices of the AMS* 56, 5 (May 2009).
[50] Jian Wu, Zhuoqing Morley Mao, Jennifer Rexford, and Jia Wang. 2005. Finding a Needle in a Haystack: Pinpointing Significant BGP Routing Changes in an IP Network. In *NSDI.*