California Fault Lines: Understanding the Causes and Impact of Network Failures

Daniel Turner, Kirill Levchenko, Alex C. Snoeren, and Stefan Savage
{djturner, klevchen, snoeren, savage}@cs.ucsd.edu

Department of Computer Science and Engineering
University of California, San Diego

ABSTRACT

Of the major factors affecting end-to-end service availability, network component failure is perhaps the least well understood. How often do failures occur, how long do they last, what are their causes, and how do they impact customers? Traditionally, answering questions such as these has required dedicated (and often expensive) instrumentation broadly deployed across a network.

We propose an alternative approach: opportunistically mining "low-quality" data sources that are already available in modern network environments. We describe a methodology for recreating a succinct history of failure events in an IP network using a combination of structured data (router configurations and syslogs) and semi-structured data (email logs). Using this technique we analyze over five years of failure events in a large regional network consisting of over 200 routers; to our knowledge, this is the largest study of its kind.

Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations

General Terms

Measurement, Reliability

1. INTRODUCTION

Today's network-centric enterprises are built on the promise of uninterrupted service availability. However, delivering on this promise is a challenging task because availability is not an intrinsic design property of a system; instead, a system must accommodate the underlying failure properties of its components. Thus, providing availability first requires understanding failure: how long are failures, what causes them, and how well are they masked? This is particularly true for networks, which have been increasingly identified as the leading cause of end-to-end service disruption [2, 9, 15, 24, 30], as they exhibit complex failure modes.

Unfortunately, such analysis is rarely performed in practice as common means of measuring network failures at fine grain presuppose measurement mechanisms (e.g., IGP logging [23], pervasive high-frequency SNMP polling, passive link monitoring [8], and pair-wise active probing [26]) that are not universally available outside focused research-motivated efforts and which can incur significant capital and operational expense. Indeed, even in the research community, it is common to use arbitrary synthetic failure models due to the dearth of available empirical data [3, 20, 25, 28].

As a step toward addressing this issue, we describe a "cheap and dirty" approach to extracting the requisite measurement data from "lowest common denominator" data sources commonly found in production networks today. In particular, we demonstrate a methodology for reconstructing historical network failure events inside of an autonomous system using three near-universal, yet under-appreciated, data sources: router configuration files, syslog archives, and operational mailing list announcements.

Router configuration files (e.g., as used to configure Cisco IOS and JunOS routers) describe the *static* topology of a network at a point in time and are commonly logged in networks of significant size to support configuration management. Each configuration file describes the set of interfaces enabled on a router and typically enough information to infer its connectivity (e.g., via the short IP network prefixes commonly assigned to point-to-point links). It is by no means a perfect data source; it may omit topology out of its purview (e.g., transparent optical cross-connects) and may include topology that is illusory (e.g., entries can persist in a config file long after a link has been decommissioned). However, in aggregate and when combined with additional data it provides broad topological coverage.

However, the long-term topology of a network by itself tells us little about its failures. Here we turn to syslog which, as typically implemented on modern routers, logs a plethora of *events* including link status changes to a remote server. Thus, it complements the router configuration data by describing the *dynamic* state of the network—the status of all active links at every point in time. However, reconstructing this state can be painful: First, the unstructured quality of syslog messages requires parsing a diverse assortment of message formats and correlating these events with interface configuration records to obtain a complete description of an event. Second, because of the "best effort" in-band nature of syslog, some messages are necessarily lost (in particular, when a link on the shortest path to the syslog server has failed). Yet, in our experience, by exploiting natural reporting redundancy (i.e., a link failure is usually reported by both endpoints), we can recover instantaneous link status almost 90% of the time.

Finally, it is nearly universal practice for network operations staff to maintain mailing lists and trouble ticket systems to share changes in operational state (e.g., to document the response to a failure, advertise a planned maintenance activity, and so on). Such free-form

SIGCOMM'10, August 30–September 3, 2010, New Delhi, India.

natural language data is both rich and incomplete: on the one hand, it provides information not available from syslog, such as the *cause* of a failure, but at the same time it is generated by non-deterministic social processes and, thus, only reflects failures that are of a subjectively sufficient magnitude and duration to warrant a broader notice. However, this duality allows such announcement data to serve two distinct purposes: as a classifier for failure causes and as an independent source of "ground truth" that can be used to validate our analyses. Following the methodology of Feamster and Balakrishnan [12], we use a combination of keyword searches, regular expressions and manual inspection to analyze announcements.

Taking these sources together we have analyzed five years of archival data from the CENIC network—a production IP network consisting of over two hundred routers serving most of the public education and research institutions in California. Using syslog and router configuration data, we extract failure events over this period, infer causes from administrator email logs, and check our results for consistency against three independent sources of network failure data: active probes of our network from the CAIDA Skitter/Ark effort, BGP logs collected by the Route Views Project, and the administrative announcements from the CENIC operators. Finally, we use our reconstructed failure log to present concrete analyses of failure duration, cause, and impact, validating a number of widely held beliefs about network failure (e.g., the dominance of link "flapping") as well as describing new findings for our dataset (e.g., the relative importance of planned maintenance vs. unplanned failures and the role of third-party telco providers in flapping episodes).

In summary, we believe our main contributions are:

- A methodology for combining router configuration files, syslog messages and human-generated network operations logs to derive a topological and dynamic failure history of a network.
- A detailed analysis of over five years of such data for a large-scale network.

The rest of the paper is organized as follows. We begin in Section 2 by discussing related work. Section 3 introduces the CENIC network and the particular datasets we use in our study. Section 4 describes our methodology followed by a discussion of validation methods in Section 5. Section 6 presents our analysis before we conclude in Section 7 with a summary of our contributions.

2. RELATED WORK

The designers of computer networks have had to contend with frequent failures—link failures, router failures, interface failures and so on—since the first networks were built [4]. However, for practical reasons, most *measurements* of failure have taken place from the edge of the network [6, 10, 11, 16, 18, 22, 32, 35]. Unfortunately such *tomographic* techniques do not provide a complete picture of the network; "a gap remains between research in network tomography and practical systems for scalable network monitoring," according to Huang, Feamster and Teixeira [16]. Direct measurement remains the gold standard for network failure data.

We are aware of three direct measurement studies in the last decade. Shaikh *et al.* [27] studied OSPF behavior in a large enterprise network. Their dataset tracks 205 routers over a month in 2002. Although the aim of the study was OSPF behavior itself, it also provided a valuable insight into the underlying component failure characteristics. In particular, the authors observe that the majority of apparent link failures in their network were caused by a single misconfigured router. A similar study monitoring OSPF behavior in a regional service provider was conducted by Watson, Jahanian and Labovitz [33]. Their study tracked a network of fifty routers, including internal and customer links, over the course of one year. They observe that a small number of routers contribute disproportionately to network instability, and that flapping links are the predominant source of instability.

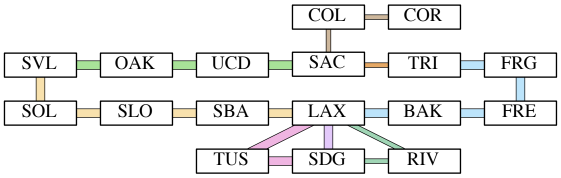


Figure 1: CENIC hub sites are connected by an optical backbone. Long-haul fiber routes SVL to SAC, SVL to LAX, TRI to LAX, LAX–SDG, and LAX–TUS–SDG use the Cisco 15808 DWDM & 15454 platforms. Metro-area networks use a combination of CWDM and Cisco 15530 & 15540 DWDM equipment.

Markopoulou *et al.* [23] studied failure in the Sprint backbone. Using passive optical taps and high-speed packet capture hardware, they collected six months of IS-IS routing protocol messages and were able to monitor "hundreds" of nodes interconnected by a DWDM (dense wave-division multiplexing) optical backbone—a design shared with our network. In addition to providing a characterization of time-to-failure and time-to-repair distributions for the Sprint backbone, they also observe that 2.5% of all links are responsible for more than half of the individual failures. Furthermore, a number of these links exhibit flapping behavior.

However, each of these previous studies have required extensive—and often expensive—instrumentation not commonly present in today's production networks. In contrast, we focus on using frequently available sources of implicit data, such as router configurations, email archives and syslog records. While others have also exploited these data sources (for example, Feamster and Balakrishnan parse router configs to find BGP errors [12], Labovitz *et al.* combine SNMP queries with operational logs to analyze failures in the backbone of a regional service provider [19], and Xu *et al.* parse syslog records to identify anomalies in datacenter operations [34]), we believe ours is the first effort that uses this information to systematically identify and characterize network failures.

3. DATA SOURCES

While we intend for our methodology to be generally applicable, our current study focuses on one particular network, where we have been able to obtain several years worth of configuration and log information. In order to set the context for our analysis, we begin by describing the network itself, and then detail the particular data sources available.

3.1 The CENIC network

CENIC, the Corporation for Education Network Initiatives in California, operates a common state-wide network providing Internet access to California public education and research institutions. Its members, with a combined enrollment of over six million, include the University of California system, the California State University system, community colleges, and K-12 school districts. Physically, the CENIC network is an optical backbone with over 2,700 miles of fiber, connecting hub sites in major cities, as shown in Figure 1. In addition, CENIC also manages equipment located outside the hub sites for some of its smaller members.

```
interface GigabitEthernet0/0/0.23
 description lax-sw-1 3/2 lax-isp ge-0/2/0.23
 ip address 137.164.22.8 255.255.255.254
 ip router isis
```

Figure 2: A Cisco 12410 configuration file entry describing a Gigabit Ethernet interface. The `description` line is free-form text; in the CENIC network, it used to record the endpoints of the connection.

Administratively, the CENIC network can be divided into three major components: the Digital California (DC) network, the High-Performance Research (HPR) network, and customer-premises equipment (CPE), each described below.

- **DC network.** The Digital California (DC) network (AS 2152) is CENIC's production network, providing Internet connectivity to University of California schools, California State Universities, California community colleges, a number of private universities, and primary schools via their respective County Offices of Education. At the end of our measurement period (December 2009) the core network consisted of 53 routers (mostly Cisco 12000 series) connected by 178 links. We refer to these links as DC (core) links. The DC network uses the IS-IS routing protocol for intra-domain routing.
- **HPR network.** In addition to the production network, CENIC also operates a High Performance Research (HPR) network (AS 2153), which interconnects major California research institutions at 10 Gb/s. It offers "leading-edge services for large application users" [7]. At the end of 2009, it consisted of six Cisco 12000 routers at the SAC, OAK, SVL, SLO, LAX, and RIV hub sites connected by seven logical links over the optical backbone. The HPR network runs its own instance of the IS-IS routing protocol.
- **CPE network.** CENIC also manages customer-premises equipment (CPE) for some of its smaller customers. A number of CPE routers (mainly those with redundant connectivity) run IS-IS on links to DC routers and other CPE routers. There were 102 such routers and 223 links at the end of 2009. We refer to these customer access links as CPE links.

There are also several statically configured access links in the CENIC network. For these links, only events from the physical layer and data link layer are recorded in syslog, as they are not monitored by the routing protocol. In the absence of a network-layer connectivity test provided by IS-IS, link semantics are unclear: interfaces plugged into a switch or DWDM device may appear "up" without the other endpoint being reachable. Given this fundamental ambiguity, we do not include static links in our analysis.

3.2 Historical data

Our study uses three distinct forms of log information from the CENIC network extending from late 2004 to the end of 2009.

- **Equipment configuration files.** CENIC uses RANCID [29], a popular open-source system that automatically tracks changes to router configurations. All changes are committed to a revision control system, making it possible to recover the configuration history of any router in the network. We were granted access to this repository, consisting of 41,867 configuration file revisions between June 2004 and December 2009. Figure 2 shows an example of an interface description for a Cisco 12410-series router.

	Network		
Parameter	HPR[1]	DC	CPE
Routers	7	84	128
IS-IS links	14	300	228
Avg. config changes per router	255	178	54
Avg. syslog entries per link	748	187	595
Avg. BGP announcements per prefix	N/A	5504	4202

[1] Excludes RIV–SAC link (see Section 6.1).

Table 1: Summary of the CENIC network dataset.

```
Mar  6 15:55:46 lax-core1.cenic.net 767:          ▷
 RP/0/RP1/CPU0: Mar 6 16:56:08.660: IS-IS[237]:  ▷
 ROUTING-ISIS-4-ADJCHANGE: Adjacency to          ▷
  lax-core2 (TenGigE0/2/0/7) (L2) Up, Restarted
```

Figure 3: A syslog message generated by a Cisco CRS-8/S router (split into multiple lines to fit). The message indicates that IS-IS routing protocol has transitioned the `lax-core1` link to the `lax-core2` router on interface `TenGigE0/2/0/7` to the up state.

```
This message is to alert you that the CENIC      ▷
network engineering team has scheduled           ▷
PLANNED MAINTENANCE:

START 0001 PDT, FRI 8/17/07
END 0200 PDT, FRI 8/17/07
SCOPE:  Power breaker upgrade
IMPACT: Loss of power redundancy at Level 3/     ▷
 Triangle Court, Sacramento

COMMENTS
CENIC Engineering team has scheduled             ▷
remote-hands at Level 3/ Triangle Court,         ▷
Sacramento to swap out breakers.
```

Figure 4: An operational announcement. Announcements are a combination of fixed-format elements (`START` and `END`) and free-form text.

- **Syslog messages.** All CENIC network routers are configured to send syslog [21] messages over the network itself to a central server located at the Tustin (TUS) hub site. The messages announce link failures at the physical link layer, link protocol layer, and network layer (IS-IS), covering the first three layers of the network protocol hierarchy. Unlike many local logs, messages in the centralized syslog are timestamped to only whole-second granularity. We obtained an archive of these messages from November 2004 to December 2009, of which 217,498 pertained to the networks in this study. Unfortunately 176 days of syslog data (9/23/2007 to 3/17/2008) are absent from the archive. Figure 3 shows a syslog message generated by IS-IS.
- **Administrator notices.** We also obtained archives of two mailing lists used to disseminate announcements about the network. Together the mailing lists contained 7465 announcements covering 3505 distinct events from November 2004 to December 2009. Figure 4 shows a typical administrator notice.

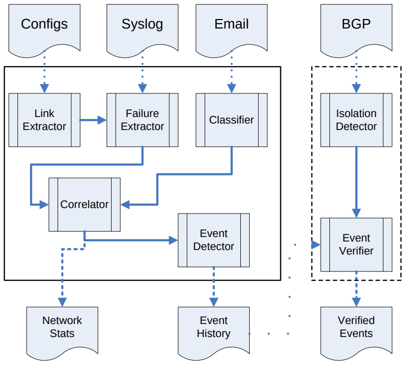


Figure 5: Our failure event reconstruction work flow. The BGP validation process is described in Section 5.

Finally, in order to help validate our conclusions about when failures occur, we extract BGP announcements relating to the CENIC networks from the Route Views Project [31]. In particular, we collect all BGP messages received by the Peering and Internet Exchange (PAIX) listener in Palo Alto, California that pertain to address blocks belonging to or serviced by the CENIC network. Note that our analyses do not depend on the BGP data—we instead use it as ground-truth regarding the small subset of externally visible failures. Table 1 provides a brief overview of the cumulative dataset that we consider for the remainder of the paper.

4. METHODOLOGY

The main goal of our work is to develop a general procedure to mine the three "low quality" sources of historical data to construct a crisp timeline of failures, where each failure event is annotated with a start and end time, set of involved links, and, if possible, a potential cause. Moreover, where appropriate, we seek to aggregate multiple simultaneous link failures into larger events such as router and point-of-presence (PoP) failures, and coalesce frequent back-to-back link failures into enclosing flapping episodes. In addition to this annotated failure timeline, we also produce statistical information about link lifetimes that serves as input to our analysis (Section 6). Figure 5 depicts the extraction process, as well as our validation experiments discussed in the next section.

4.1 Recovering the topology

Before beginning to catalog failures, we must first build a topological model of the network under study. While it is possible that a map may be readily available (indeed, the current CENIC topology is available on the Web[1]), we instead choose to infer the topology from the historical data. Our reasons are two-fold: First, previous work has shown that topological databases rapidly become out of date as operators change the physical topology to increase capacity or in response to failures [17]. Second, we need to cross-correlate syslog information with physical network entities; extracting the actual identifiers used in the configuration files significantly simplifies this task.

We begin by attacking the latter issue. In particular, we map entities described in the syslog files to individual routers and links in the network topology. This process is not entirely straightforward however, as layer-3 syslog messages identify both endpoint *routers* of a link, but only the particular *interface* for the router that generates the syslog message. In several cases this is insufficient to fully describe a link, for example when two routers have multiple links between them. To accurately identify the interfaces at both ends of a link, we consult the router configurations. Each configuration file describes the kinds of interfaces present on the router and how they are configured; Figure 2 shows an example interface description.

Our collection of router configuration files is not just a single snapshot in time, but rather a series of configuration files for each router, where each file version is annotated with its update time. Thus, router configurations give us a meaningful way to define link "lifetime" as the period of time between its first mention in a configuration file and its last.

We identify the ports associated with each link using a straightforward iterative process similar to previous work on on extracting global network state from configuration files [13, 14]. For each active interface running IS-IS, we determine the set of IP addresses on the same subnet. The overwhelmingly common case is that an interface's subnet is 255.255.255.254 (i.e., a point-to-point link) making it obvious which interfaces are communicating with each other. An important caveat is that IP addresses are often changed and re-used on different routers, so it is critical to allow interfaces to be part of multiple different links throughout the analysis.

4.2 Identifying failures

Armed with the set of links in the network, we process the failure history of the network in several steps. We begin with the syslog archive under the assumption that it contains an accurate—if incomplete—enumeration of link failures.

4.2.1 *Defining failure*

For our purposes, a *failure* is any event that causes a routing-state change (layer-3) syslog message to be recorded. As a result, our reconstructed event history reflects the routing state of the network, i.e., a link is considered to have failed whenever a router refuses to send traffic over it. As such, our event history may not accurately capture the physical state of the network components. For example, a router may refuse to route traffic over a link because a hold-down timer has yet to expire rather than because of an actual disconnection. We define the duration of a failure event to extend from the first layer-3 "down" message in syslog (we may receive messages from routers at both ends of the link) to the first "up" referring to the same link.

Recall that syslog also contains failure messages generated at the physical link layer and at the link protocol layer. We choose to focus on the network layer, as opposed to the link layer, because it more faithfully captures the state we are interested in, namely whether the link can be used to carry traffic. The bias is, of course, one-sided: if the physical layer reports the link is "down," then it is necessarily also "down" at the network layer; on the other hand, a link may be "up" at the physical layer, but not at the network layer (e.g., an Ethernet link plugged into a switch with incorrectly configured VLANs).

4.2.2 *Grouping*

Once individual failure events have been identified, we further consider whether failure events overlap. We define simultaneous failures to be two or more failures on distinct links occurring or healing within 15 seconds of each other. In keeping with the literature [23], we identify three types of simultaneous failures: router-related, PoP-related, and other. A simultaneous failure is deemed router-related if all involved links share a common router, PoP-related if all links share a common PoP but not a common router, and "other" if the failures has no common PoPs.

In addition to grouping simultaneous failures across multiple links, we also aggregate back-to-back failure events on a single

[1] http://noc.cenic.org/maps/

Classification	Example causes or explanations
Power	"City-wide power failure", "UPS failure"
Hardware	"Replacing line card" , "Replacing optical amplifier"
External	Failure of non-CENIC equipment (e.g., leased fiber)
Software	"Upgrading IOS"
Configuration	"Modifying IGP metrics" , "adding IPv6 capability"
Other	"DoS attack" , "flooded machine room"
Unknown	Failures with unknown or unreported causes

Table 2: Operational announcement failure classification.

link into an enclosing *flapping* event. Link flapping has long been understood as a challenge for routing protocols [33]. Based on our experience with the CENIC dataset in this study, we liberally define flapping as two or more up/down state changes where the down-to-up periods last no longer than 10 minutes. (We justify our particular parameter choice in Section 6.2.4.)

4.2.3 Dealing with loss

In contrast to Markopoulou *et al.* [23], who used a specialized IS-IS listener, we glean routing state information entirely from syslog messages generated by the routers themselves. Unfortunately, because of the unreliable nature syslog's UDP-based transport, not all router log messages make it to the syslog server. As a result, it is common to hear about a failure from one side of a link but not the other. For this reason, we consider a link down if at least one endpoint reports it being down, and up if at least one endpoint reports it coming up.

It is also common to see two "up" messages with no intervening "down" message, and vice versa. In one instance, for example, the syslog shows a link between the LAX and RIV routers in the HPR network fail (reported "down" by RIV but not LAX), and then, 36 days later, the same link is reported down by LAX, with no intervening messages about the link. We discard such anomalous periods—between consecutive "up-up" or "down-down" messages where it was impossible to infer when a link changed state—from the dataset. We choose this conservative approach in order to favor correctness over completeness. In the case of our dataset, the excluded time periods account for 12.6% of the link-hours on HPR links, 9.5% of the link-hours on DC links, and 16% of the link-hours on CPE links.

4.3 Categorizing failures

So far, we have determined when failures occur and how long they last, but nothing more. Inferring the probable *cause* of these failures requires additional inference and additional data.

Over and above the syslog entries, operational announcement archives contain a wealth of information that, when available, can turn simple failures into well described events (for example, see Figure 4). After manually reviewing a number of announcements, we observed that most events can be categorized into a small number of classes.

We classify the administrator notices into seven categories, listed in Table 2. We manually labeled each announcement based on matching keywords, phrases, and regular expressions. In some instances, there may be multiple announcements pertaining to the same failure event. Grouping these multiple announcements into a single event requires some piece of information to be repeated in each announcement. Luckily, the first announcement about an event contains the start time of the event in an easy to identify and parse format. From there, each additional announcement either contains the original announcement or restates the start time of the event. Also the final announcement contains the time that the event officially ended.

Armed with failure start and end times from syslog as well as failure causes tagged with start time and end time from the operational announcements archives, we use temporal correlation to match failures (computed by processing the syslog) with potential causes (based on administrator notices). To find matches, we widen the start and end times from the operational announcements by fifteen minutes to compensate for potential issues with clock synchronization, conservative time windows, and delayed reporting. Unfortunately, blindly assigning causes to syslog failures that fall within an announcement's window leads to a large number of false positives. To minimize such errors we extract router names or abbreviations from announcement messages and ensure that at least one router in a link was mentioned in the message before matching it to a corresponding syslog-based inference. For our CENIC dataset, we discard 1,335 of the 2,511 messages (53%) that, while contemporaneous with a failure event in syslog, do not explicitly mention a router involved in the failure. It is likely that manual inspection could salvage a significant percentage of these.

5. VALIDATION

Unfortunately, there is no free lunch. Syslog data was never intended to be used for our purposes; consequently certain omissions and ambiguities are inevitable. Validating network failure data in general is challenging, and especially so when dealing with events five years in the past. Thus, we take an opportunistic approach, checking for consistency against data we *do* have with an understanding that there will be noise and errors that reflect the different vantage points between these diverse data sources. In particular, our approach has two major shortcomings: it is neither complete nor 100% accurate: there are likely to be failures that our log does not include, and it may be the case that failures we do include are spurious, misclassified, or improperly timestamped. We discuss the potential biases that result from our choice of data source, as well what we did to validate our results and help quantify our errors.

5.1 Measurement bias

As discussed earlier, some link status change messages may be missing from the syslog due to its unreliable nature. Thus, a "down" link state transition may not have a preceding "up" or vice versa. In our use to date we have found that such gaps are relatively minor (accounting for less than 13% of link time) but this could also be an artifact of our particular network and hardware.

Additionally, our definition of link failure is based on adjacency status reported by the underlying routing protocol. For example, to ensure connectivity, the IS-IS protocol requires routers to send and receive *hello* messages. By default, a router sends a *hello* message once every ten seconds and declares a link disconnected if no *hello* message is received for thirty seconds. Hence, we may under-count failure duration by up to thirty seconds per failure. Conversely, IS-IS considers a repaired link down until a configurable hold-down timer expires (this process is dynamic, but should create biases of similar magnitude).

Another ambiguity arises in pinpointing the "age" of each link to allow annualized statistics to be calculated. The natural definition of age is simply the amount of time between when a link was added to the network and when it was removed. One minor issue with this definition is that some links are added to the network before our syslog data begins (left censored), are not removed until after our syslog data runs out, and/or continue to operate during the months where syslog data was lost (right censored). To combat these issues we do not allow any links to be added to the network before the syslog data starts, remove all links from the network after the syslog data ends, and ignore any operational time for the period of time missing in our syslog data. A second version of this problem that cannot be overcome directly is the granularity with which router configuration files are maintained. Since interfaces are not tagged with creation or removal times, we rely on the first and last configuration file that contains a valid interface description for these times. Unfortunately, configuration updates are logged periodically—rather than instantaneously—thus, we are prone to add links to our network later than they have actually been added and remove them after they have likely been removed.

5.2 Internal consistency

Because our data is historical, and the CENIC network operators did not collect or maintain any additional logs that we can use as ground truth regarding the timing or causes of failure, we are forced to search for alternate means of validation. We use two qualitatively different approaches. The first is to cross-validate the records we do have; any inconsistencies or disagreement between syslog and the operational email announcements increases the likelihood of error. While we cannot say for certain that the lack of inconsistency implies correctness, we can quantify the degree of inconsistency to provide an approximate upper bound on the accuracy of our approach. Second, certain failures may be externally visible, in which case we can leverage logs collected by third parties.

Focusing first on internal consistency, we use the administrator notices (Section 4.2.3) to validate the event history reconstructed from the syslog archive. In reconstructing this history, we used the administrator notices to label failures with causes when available—in particular, if there is an announcement that pertains to the particular failure. Understandably, only a small subset of the link failures are discussed by the operators on the email list. Here, we attempt the opposite mapping. Specifically, we check whether the reconstructed event history also records the corresponding event.

Ideally, we would confirm that each of the 3,505 distinct events mentioned in an administrative announcement appears in the log. Due to the difficulties in extracting precise details from free-form email messages, the matching must be done manually. Hence, we verify a random subset of the events. Of the 35 (roughly 1%) events we inspected, only one could not be matched to a corresponding (set of) failure(s) in the event history (i.e., 97% accuracy).

5.3 Externally visible events

In a well-designed network, most failures are masked by redundant links and protocols. Hence, while network operators are clearly interested in knowing about failures so they can address the fault and restore proper operation, users of the network may not even notice when failures occur. A certain class of catastrophic failures, however, cannot be hidden: those that result in network partitions. The CENIC networks are connected to the larger Internet and, hence, any network partitions in those networks would be observable from the commercial Internet.

We are aware of two publicly available datasets concerning reachability that go back far enough in the past to validate our failure logs: the CAIDA Skitter/Ark active traceroute measurements, and the University of Oregon's Route Views BGP logs. Here, we develop a methodology to validate our failure log—at least in the limited case of failures that result in a network partition—by checking against publicly available traceroute and BGP records.

5.3.1 CAIDA Ark/Skitter traceroute

One direct method of ascertaining whether a link is down or not is to attempt to use it. Most commercial network operators conduct periodic active end-to-end probes [26] to do just that for their own networks. CAIDA's Ark (né Skitter) project conducts sporadic traceroutes to numerous destinations throughout the Internet from various traceroute servers [5]. Occasionally, Skitter probes destinations within the CENIC network. While the actual route itself is of little interest to us, the reachability of the end point is. In particular, for all Skitter probes to a destination within CENIC, we can validate our failure log by comparing the success or failure of the Skitter probe to our event records: for all successful Skitter probes, we verify that all of the links traversed (which are conveniently enumerated by the Skitter record) were "up" at the time of the probe according to our failure log. Conversely, should a Skitter probe fail, we verify that either 1) the probe failed before reaching or after passing through the CENIC network, or 2) the link leaving the last successful hop was "down" at the time of the probe according to our log.

CAIDA provided us with the Skitter traceroute data covering six months (January–June 2007) of our study—already over four gigabytes of compressed data. From the data, we extracted 75,493,637 probes directed at 301 distinct destinations within the CENIC network from 17 different traceroute servers, covering 131 links and 584 distinct paths through the CENIC network. The outcome of each of these 75 million Skitter probes was consistent with the link states reflected in our event history. Unfortunately, none of the Skitter probes failed within the CENIC network itself—in other words, while the log is completely consistent with the Skitter data, Skitter does not positively confirm any failure events in the log.

5.3.2 Route Views BGP archive

Unlike traceroute, which requires active probing to detect failures, passive BGP listeners are asynchronously informed of reachability information. Hence, to the extent a link's connectivity is monitored by BGP, its failure history is likely to be far more complete. The University of Oregon's Route Views project has deployed ten BGP listeners throughout the world to collect BGP updates, and makes their logs publicly available. The main challenge with BGP data, however, is its coarse granularity. BGP speaks in terms of networks or IP prefixes as opposed to individual layer-3 links like traceroute. Hence, a BGP listener will only detect when an entire network becomes unreachable.

When considering the particular case of the CENIC network, we must bear in mind that multiple core routers in multiple cities would have to fail simultaneously to partition the core of the network. It is not surprising, then, that we do not observe a partition in the CENIC core network during the course of our study. However, most customer sites—places with CPE routers—have only one router and only one or two links to the CENIC core. Therefore, if all of the links between CENIC and a CPE router fail, the site becomes partitioned from the network. Such events are infrequent, but do occasionally occur.

We identified the IP prefixes for 60 distinct networks (i.e., customer sites) served by CENIC. Unfortunately, we can only use BGP to validate a subset of these sites because CENIC does not withdraw prefixes of customers residing in CENIC address space (these are typically small customers like K-12 school districts). We identified 19 customer sites in the CENIC failure logs that have their own autonomous system (AS) and for which CENIC generates BGP withdraw messages. We identify network partitions for these sites in our reconstructed event history by searching for multi-link failure events that involve all of a CPE router's links to CENIC. We declare such customer sites to be *isolated* for the duration of the failure. One issue with this approach is that some customers may be multi-homed—in other words, have access links to networks other than CENIC. In such an instance, we would assert that a site is isolated when in fact it is only suffering degraded service. We have

	Sites	Events	Pw	Hw	Sw	N/A
Isolation	14	51	2	1	13	36
Path change	19	105	4	2	24	73

Table 3: Summary of the CENIC network partitions validated against the Route Views BGP data.

not, however, uncovered any evidence of such sites in our logs or interactions with CENIC operators.

The geographically closest Route Views BGP listener to the CENIC network is housed at the Peering and Internet Exchange (PAIX) in Palo Alto, California. Unfortunately, the BGP listener's network (AS6447) does not directly peer with the CENIC network (AS2152), but it does peer with several ASes that directly peer with CENIC. To accommodate the idiosyncrasies of BGP convergence between all of the peers of the Route Views listener[2], we declare a CENIC site isolated according to BGP if at least four peer ASes withdraw all of the site's prefixes. In addition to these isolation events, we also observe instances where two or three ASes withdraw all of a site's prefixes but several other ASes will advertise multiple paths of monotonically increasing length to a site's prefixes. We refer to this second type of event as a *BGP path change*. While isolation is a strong proof of network partition, BGP path change events are also likely due to externally visible failures within the CENIC network and are therefore also useful for validating our error log.

Of the 147 isolating events in our event history that should be visible in BGP (see Table 7 for a breakdown), we were able to match 51 to complete BGP isolations—i.e., fully converged route withdrawals (Table 3). If we more conservatively consider BGP path changes, however, we are able to confirm 105 of the 147 events. Notably, of the remaining 42 events, 23 of them pertain to a single link. It is possible that this link is backed up by a statically configured link that is not reflected in our IS-IS dataset.

6. ANALYSIS

By applying the methodology described in the previous sections to the CENIC syslogs and operational announcement email logs, we obtain over half-a-decade worth of failure data for a moderately sized production network. Hence, we are in the position to ask several fundamental questions regarding the operation of a real network. In particular, we consider:

- How often do failures occur? How long do they last?
- What are the *causes* of failures? Are some types of links more failure-prone than others?
- What is the *impact* of link failures? Does the network adapt without significant service disruption?

While we make no claims regarding the generality of our results, we believe that a study of this scope and duration is unprecedented in the literature, and that our findings are likely to be representative of a larger class of educational and research networks.

6.1 Event history at a glance

Figure 6 shows the reconstructed event history at a glance. Links are ordered lexicographically along the y axis. Each failure is represented by a single point on the plot, located according to the start of the event. Two aspects bear note:

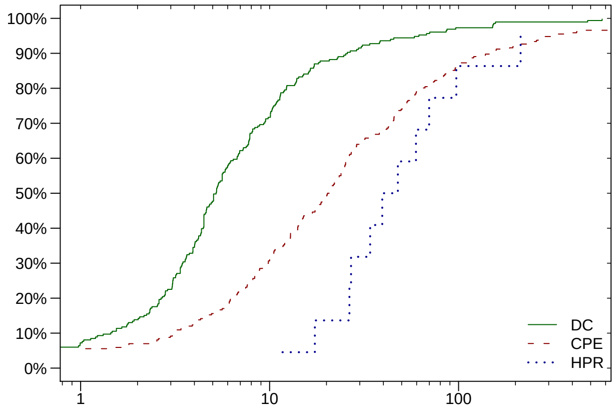


Figure 7: Failures per link per year, excluding links up for less than 30 days.

Vertical banding. Several vertical bands are apparent in the figure, which correspond to system-wide events. For example, the band in September 2005 labeled V_1 in the figure is a network-wide IS-IS configuration change requiring a router restart. (The scale of the figure makes the link failures appear simultaneous; the band actually spans about a week.) Another band in March 2007 (labeled V_2) is the result of a network-wide software upgrade. The third band, V_3, occurs in February 2009 as a network-wide configuration change in preparation for IPv6.

Horizontal banding. Figure 6 also contains several horizontal segments. The nearly solid segment labeled H_1 in the middle of the figure corresponds to a series of failures on a link between a core router and a County of Education office. The segment is made up of many short failures happening only a few times a day. After at least one unsuccessful attempt to diagnose the problem, the cause was ultimately found to be faulty hardware.

The horizontal segment labeled H_2 in the figure corresponds to a RIV-SAC link between two HPR routers. Between July 2006 and January 2007 this link experienced over 33,000 short-duration failures. While the initial cause was a fiber cut, the repair process damaged an optical device leading to instability that was difficult to diagnose. Because this single flapping event accounts for 93% of all link failures in the HPR network, we remove it from the data set to avoid skewing further analyses.

6.2 Aggregate statistics

We begin our analysis by computing aggregate statistics about the frequency and duration of failures on a per-link basis, both in terms of individual failure events and cumulative link downtime. Table 4 shows the average, median, and 95th percentile of each distribution. For all annualized statistics, we excluded links in operation fewer than 30 days because of their inflated variance.

6.2.1 Failure rate

Perhaps the most natural first question we might ask is, "How many failures are there?" Figure 7 shows the cumulative distribution function (CDF) of the number failures per link per year. We compute the number of failures per year for each link by dividing the number of failures by the lifetime of the link (excluding links in operation for less than 30 days).

In the DC network, most links experience few failures, as one might expect of a production network. The CPE network, consisting of access links and routers on customer premises, is somewhat less reliable, with a median annual failure rate of 20.5 failures per link. The HPR network experienced considerably more failures.

[2] BGP announcements are modulated by local router policy, which may differ between peers.

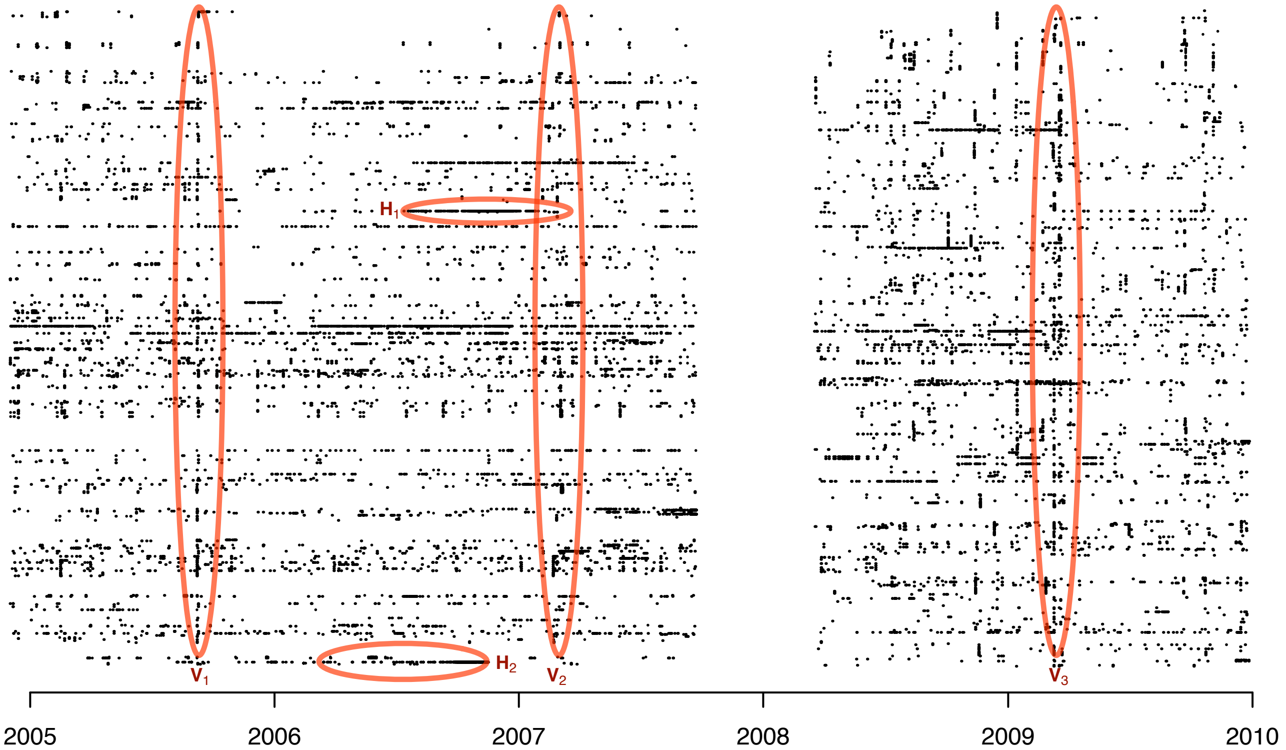


Figure 6: Overview of reconstructed link failure dataset. Links are enumerated along the *y* axis. A mark indicates the beginning of a link failure event. No syslog data is available for the seven-month period from late 2007 to early 2008.

	Annual failures			Annual downtime			Time to repair		
	Avg	Med	95%	Avg	Med	95%	Avg	Med	95%
DC	16.2	5.1	57.7	2.7 d	24 m	34 h	17.4 m	13.0 s	15.2 m
CPE	302.0	20.2	276.7	1.6 d	72 m	163 h	3.6 m	3.0 s	2.6 m
HPR	58.5	39.5	155.1	1.2 d	497 m	125 h	16.1 m	4.0 s	23.7 m
Internet2	12.9	14.3	29.1	0.2 d	112 m	19 h	20.7 m	54.0 s	75.5 m

Table 4: Annual number of failures per link and annual link downtime for links in operation 30 days or more. As noted in Section 6.1, the HPR network statistics exclude the RIV-SAC. The Internet2 Network is discussed in the Appendix.

In all three networks, however, the distributions are heavy-tailed: the 95th percentile statistics (Table 4) reveal that the most failure-prone links are roughly an order of magnitude less reliable than the majority of links.

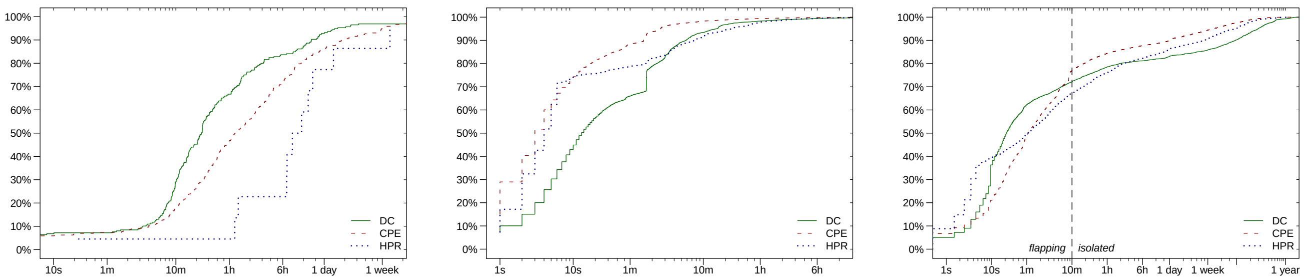
Overall, we found that five links are responsible for more than half of the link failures observed. This result is consistent with past studies documenting the same phenomenon. Watson *et al.* [33] observe that a few routers are responsible for most of the updates in a regional ISP, a finding also documented by Markopoulou *et al.* [23] in the Sprint backbone, and Shaikh *et al.* [27] in a large enterprise network (there, a single router was found to be responsible for most of the apparent failures in the entire network). In some sense, this is good news because it suggests that there are a small number of "hot spots" in the network that require attention.

6.2.2 Downtime

We can also quantify a link's reliability in terms of its total *downtime* (i.e., the sum of repair times). Figure 8(a) shows the cumulative distribution of annual link downtime by network, again excluding short-lived links. The median annual downtimes for the DC and CPE network are 24 and 72 minutes, respectively, corresponding to four nines of reliability. This difference confirms an intuitive belief we held, namely that backbone links would be more reliable than access links. The most natural explanation for this is that backbone links affect more customers and are thus better maintained and more closely monitored than access links. The median link in the HPR network, on the other hand, is closer to three nines of reliability—most likely reflecting its experimental status and more frequent upgrades. Once again, the 95th percentile shows all distributions have a long tail (note the tails are in terms of hours, not minutes).

6.2.3 Time to repair

The annual downtime and failure rate statistics suggest that the "outlier" links in the long tail have much shorter failures than "normal" links (represented by the median). Figure 8(b) shows the cu-



(a) Annualized link downtime. (b) Time to repair. (c) Time between failures.

Figure 8: CDFs of individual failure events, by network, for links in operation 30 days or more. (Log-scale *x*-axes.)

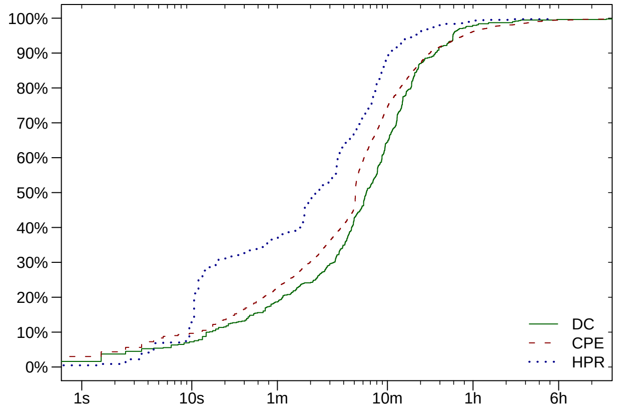


Figure 9: Total downtime within a flapping episode.

mulative distribution of individual repair times. The sharp spike in the DC network is due to a single highly periodic flapping link. Other than this anomaly, the most striking feature of the graph is the low failure durations. Remarkably, over 70% of the failures in the CPE and HPR networks last less than 10 seconds. In the DC network, 70% of failures last less than 100 seconds, with a median duration of 13 seconds.

6.2.4 *Grouping*

So far, we have considered each link failure independently. As discussed in Section 4.2.2, however, we also group link failures into larger events based upon temporal correlation. In particular, we aggregate simultaneous failures, when appropriate, into PoP or router failures, and combine back-to-back failures into flapping episodes. In the case of the CENIC network, however, the former are relatively infrequent, so we focus exclusively on the latter.

Figure 8(c) plots the CDF of time between failure events on a single link. We draw a vertical line at 10 minutes, which serves as our definition of "flapping:" two or more consecutive failure events on the same link separated by less than 10 minutes are grouped together into a larger flapping episode. 10 minutes is just past the knee of the curve for each network—the distributions appear memoryless for longer intervals. More than 50% of all flapping episodes constructed in this manner consist of only two failures, but 5% of the episodes contain more than 19 individual failures (not shown). Figure 9 shows the amount of downtime within flapping episodes—note that this is *not* the duration of the episode, only the periods within the episode when the link was actually down. Comparing to Figure 8(b), we see that flapping episodes, on the whole, are more disruptive than typical failure events.

Again, our findings reinforce those of prior studies. Both Watson *et al.* [33] and Markopoulou *et al.* [23] also find that link flapping is a predominant source of instability. It is unlikely that all three studies reflect anomalous networks and instead we suggest that short time scale and oscillatory behavior may simply be "normal" in large networks. Thus, network protocols and routing algorithms should be prepared to handling flapping as a common case.

6.3 Causes of failure

Now that we have quantified how often failure occurs, we turn our attention to its causes. We consider whether particular types of links are more likely to fail, and then examine the instances where operators explicitly place blame.

6.3.1 *Link type*

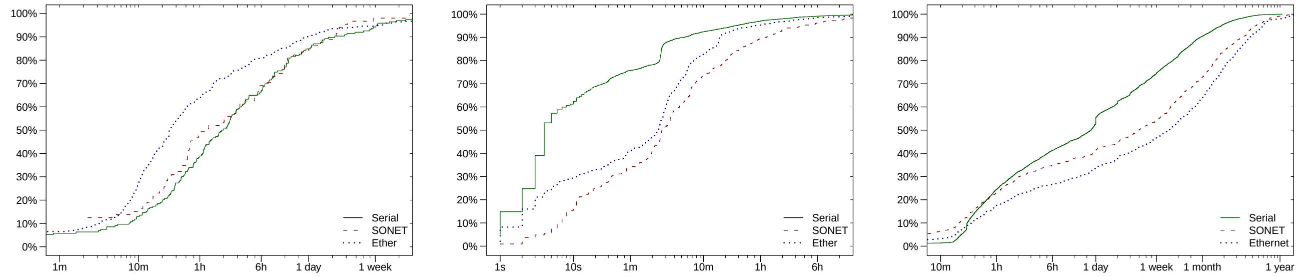
Each constituent CENIC network is composed of a number of different link technologies, including Ethernet, SONET, and serial lines. Figure 10 breaks down the individual failure events not by network (c.f. Figure 8), but instead by the type of hardware involved. Figure 10(a) suggests that Ethernet links are more reliable than other technologies. Figure 10(b) shows that while Ethernet failures are not as quick to repair as serial lines, they are far less frequent (Figure 10(c)). This is undoubtedly in part due to Ethernet's predominance for short-haul links, which are less exposed to external failure processes.

Figure 11 presents a similar breakdown, separating links into intra-PoP and long haul. Perhaps not surprisingly, Figure 11(a) shows a clear separation in reliability, with intra-PoP links being markedly more available than long-haul links. This may be due to the fact that many intra-PoP links are carried over Ethernet; indeed, comparing Figures 11(b) and 11(c) to Figures 10(b) and 10(c) suggests that long-haul failures are dominated by serial links.

6.3.2 *Labeled causes*

For a subset of link failures, we are able to annotate them with information regarding their causes by matching them to administrator notices. We were able to match 5,237 (out of 19,046) events to such a notice, accounting for 37.5% of the total downtime. Figure 12 shows the breakdown of these events according to the stated cause. The plurality of failure events are due to software upgrades, with hardware upgrades the next most frequent cause. Figure 13, however, shows that while hardware-related events account for the lion's share of the downtime, software upgrades are responsible for much less of the total downtime; indeed, external factors including power disruptions have a more significant footprint. The data is also summarized in Table 6.

Table 5 provides some basic statistics regarding the duration of individual failure events for each category. Most events are

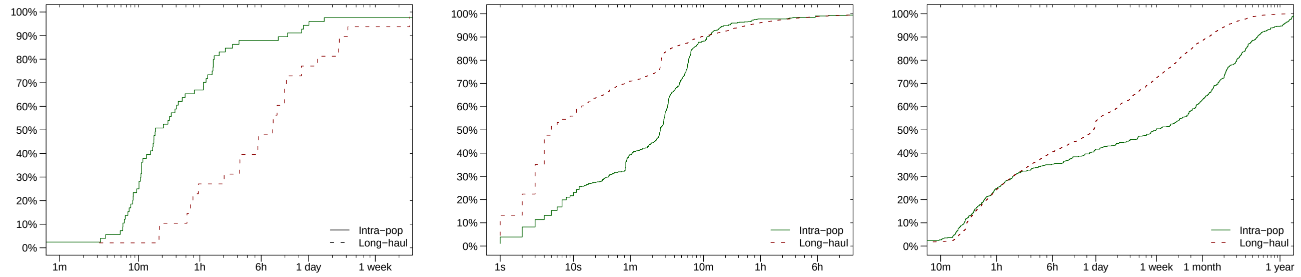


(a) Annualized link downtime.

(b) Time to repair.

(c) Time between failures.

Figure 10: CDFs of individual failure events, by link hardware type, for links in operation 30 days or more. (Log-scale *x*-axes.)



(a) Annualized link downtime.

(b) Time to repair.

(c) Time between failures.

Figure 11: CDFs of individual failure events, by link class, for links in operation 30 days or more. (Log-scale *x*-axes.)

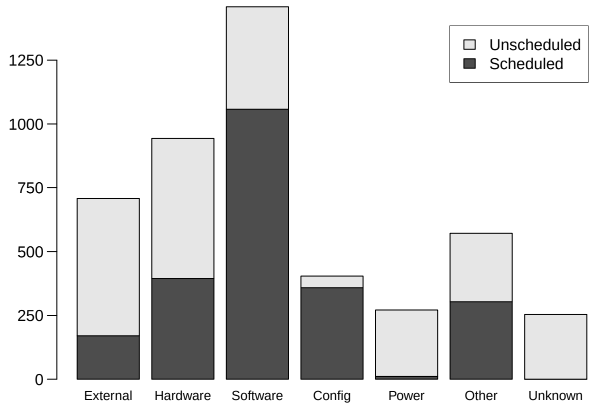


Figure 12: Failure events that matched administrator notices during the entire measurement period, broken down by cause.

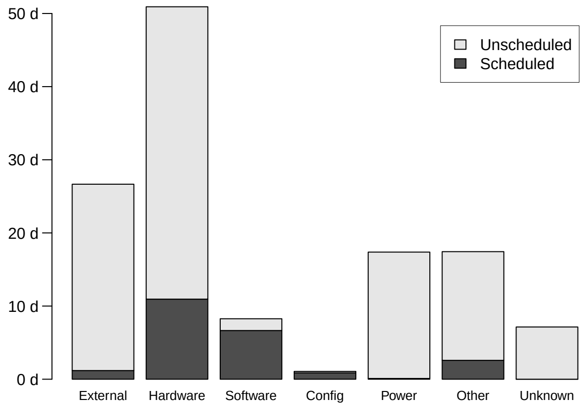


Figure 13: Cumulative downtime of the failures that matched administrator notices over the entire measurement period, categorized by failure.

short, but the median hardware and power outages are substantially longer—over twenty minutes. Almost all categories have heavy tails, however, which cause the average failure duration to be an order of magnitude longer than the median.

In addition to identifying the cause of the failure, administrator notices also indicate whether or not the failure is anticipated or "scheduled". While most of the failure *events* found in the administrator announcements are scheduled, most of the actual downtime can be attributed to unexpected failures—likely because the operators take care to make sure planned downtime is as limited as possible. Indeed, the median planned outage lasts less than 5 minutes (not shown). Interestingly, it appears network operators frequently are not notified by external entities ahead of incidents that impact the network's operation.

6.4 Failure impact

In general, it is extremely difficult for us to tell from the failure log what—if any—impact a failure had on users of the network. For the set of events that are annotated with administrator notices, however, we can report if the notice explicitly stated whether or not the event was supposed to have an impact on the network. The third column of Table 6 indicates what fraction of the events were supposed to have some impact—however brief—on the network. In almost all cases, the operators indicate some link downtime may result. This phenomenon is perhaps due to self selection on the part of the operators, however. Non-impacting failure events—especially unscheduled ones—seem far less likely to motivate an operator to send an announcement.

Cause	Events	Time to repair Avg	 Med
Hardware	20%	95 m	5 m
Power	6%	93 m	18 m
External	15%	61 m	4.6 m
Software	32%	10 m	4 m
Configuration	9%	5 m	1 m
Other	12%	46 m	6 m
Unknown	5%	52 m	6 m

Table 5: Major causes of failure according to administrator announcements, ordered by median time to repair.

Cause	Notices	Scheduled	Impacting
Hardware	25%	65%	71%
Power	20%	4%	99%
External	15%	29%	95%
Software	12%	84%	99%
Other	12%	69%	82%
Configuration	8%	91%	45%
Unknown	7%	0%	99%

Table 6: Breakdown of administrator notices by failure cause.

	Sites	Events	Pw	Hw	Sw	N/A
AS2152	41	361	28	12	39	287
Other AS	19	147	6	5	26	105

Table 7: Summary of the CENIC network partitions.

Cause	Avg	Med	95%
Power	5 h	20.6 m	33 h
Hardware	8.2 h	32 m	3.7 d
Software	6 m	2.7 m	13.9 m
N/A	8 h	32 m	3.7 d

Table 8: Duration of network partition for all isolating events.

As discussed in Section 5, the only type of impact we can infer are isolating network partitions. Table 7 presents the 508 isolating failures we identify in the failure log, separates them into networks with and without their own AS, and provides the cause annotation if available. Interestingly, the breakdown of failure causes for partition events is somewhat different than for all events—here, software failures dominate. Table 8 summarizes the distribution of times to repair for different causes of isolating failures, independent of the AS involved. As with non-isolating events, power and hardware events have significantly longer durations than those caused by software failures.

6.5 Time dynamics

Like most complex systems, the CENIC network is continually evolving. The most significant change, starting in 2008, was to designate some DC routers as "core" routers and the rest as "access" routers. This resulted in the decommissioning of 235 links and the introduction of 432 new links. A natural question, then, is whether the network qualities examined earlier have changed as well. Figure 14 shows the annualized link downtime in the DC network for each year in the measurement period. We expected to find a trend in the basic statistics presented in Section 6.2. In fact, we found that these performance indicators varied from year to year with no discernible trend. The year 2006, and to a lesser extent 2008, stands out for having lower link downtimes than the preceding and following years. The annualized number of failures per link varied accordingly, with the lowest median of 0.0 in 2006 and the highest median of 6.0 in 2005.

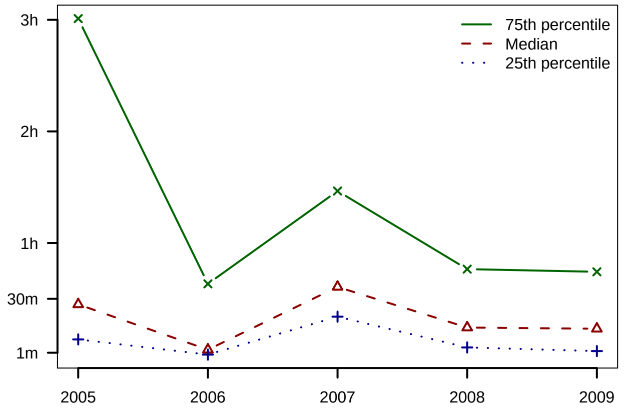


Figure 14: Annualized link downtime, in seconds, in the DC network, by year. From top to bottom, the lines show the 75th percentile, median, and 25th percentile.

Investigating further, we find that the distribution of causes studied in Section 6.3.2 varies as well. Several network-wide events are responsible for a significant variation in the number of link failures. Most notably, software-related link failures and configuration changes were a significant source of link failures in some years and not others. The three vertical bands in Figure 6 due to network-wide upgrades and configuration changes (see Section 6.1) had a significant impact on the median number of failures and median link downtime in 2005, 2007, and 2009. Longitudinal trends, if present, are thus dwarfed by major but infrequent events.

7. CONCLUSION

In this paper we present a methodology for inferring and analyzing the link failure history of a network absent dedicated monitoring infrastructure. In particular, we show that existing "low quality" data sources already widely gathered in production networks—syslog, router configs and operational mailing lists—can be opportunistically combined to reconstruct topology, dynamic state and failure causes. Using this approach we have analyzed five years of link failure history from the CENIC network, a large California Internet service provider, and both validated existing understandings about failure (e.g., the prevalence of link flapping) and documented less appreciated issues (e.g., the large amounts of downtime attributable to 3rd-party leased line problems). We believe our overall approach is fairly general and should be straightforward to adapt to a wide variety of IP networks.

Acknowledgments

The authors would like to thank Brain Court, Darrell Newcomb, Jim Madden, and our shepherd, Aman Shaikh, for their advice and suggestions. This work is supported in part by a grant from the UCSD Center for Networked Systems.

8. REFERENCES

[1] Internet2. http://www.internet2.edu.

[2] A. Akella, J. Pang, B. Maggs, S. Seshan, and A. Shaikh. A comparison of overlay routing and multihoming route control. In *Proceedings of SIGCOMM*, pages 93–106, 2004.

[3] M. Balakrishnan and A. Reibman. Characterizing a lumping heuristic for a Markov network reliability model. In *Proceedings of FTCS*, pages 56–65, 1993.

[4] P. Baran. On distributed communications networks. *IEEE Transactions on Communications Systems*, 12(1):1–9, March 1964.

[5] K. Claffy, T. Monk, and D. McRobb. Internet tomography. *Nature, Web Matters*, January 1999.

[6] M. Coates, R. Castro, and R. Nowak. Maximum likelihood network topology identification from edge-based unicast measurements. In *Proceedings of SIGMETRICS*, pages 11–20, 2002.

[7] Corporation for Education Network Initiatives in California. The CalREN network. http://www.cenic.org/calren/index.html.

[8] C. Cranor, T. Johnson, O. Spataschek, and V. Shkapenyuk. Gigascope: a stream database for network applications. In *Proceedings of SIGMOD*, pages 647–651, 2003.

[9] M. Dahlin, B. B. V. Chandra, L. Gao, and A. Nayate. End-to-end WAN service availability. *IEEE/ACM Transactions on Networking*, 11(2):300–313, April 2003.

[10] A. Dhamdhere, R. Teixeira, C. Dovrolis, and C. Diot. NetDiagnoser: Troubleshooting network unreachabilities using end-to-end probes and routing data. In *Proceedings of CoNEXT*, 2007.

[11] N. Duffield. Network tomography of binary network performance characteristics. *IEEE Transactions on Information Theory*, 52(12):5373–5388, 2006.

[12] N. Feamster and H. Balakrishnan. Detecting BGP configuration faults with static analysis. In *Proceedings of NSDI*, pages 43–56, 2005.

[13] A. Feldmann. Netdb: IP network configuration debugger/database. Technical report, AT&T, 1999.

[14] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford. Netscope: Traffic engineering for IP networks. *IEEE Network*, 14(2):11–19, 2000.

[15] K. P. Gummadi, H. V. Madhyastha, S. D. Gribble, H. M. Levy, and D. Wetherall. Improving the reliability of Internet paths with one-hop source routing. In *Proceedings of OSDI*, pages 13–13, 2004.

[16] Y. Huang, N. Feamster, and R. Teixeira. Practical issues with using network tomography for fault diagnosis. *Computer Communication Review*, 38(5):53–57, October 2008.

[17] R. R. Kompella, J. Yates, A. Greenberg, and A. C. Snoeren. IP fault localization via risk modeling. In *Proceedings of NSDI*, pages 57–70, 2005.

[18] R. R. Kompella, J. Yates, A. Greenberg, and A. C. Snoeren. Detection and localization of network black holes. In *Proceedings of INFOCOM*, pages 2180–2188, 2007.

[19] C. Labovitz, A. Ahuja, and F. Jahanian. Experimental study of Internet stability and backbone failures. In *Proceedings of FTCS*, pages 278–285, 1999.

[20] K. Levchenko, G. M. Voelker, R. Paturi, and S. Savage. XL: An efficient network routing algorithm. In *Proceedings of SIGCOMM*, pages 15–26, 2008.

[21] C. Lonvick. RFC 3164: The BSD syslog protocol, August 2001.

[22] Y. Mao, H. Jamjoom, S. Tao, and J. M. Smith. NetworkMD: Topology inference and failure diagnosis in the last mile. In *Proceedings of IMC*, pages 189–202, 2007.

[23] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, Y. Ganjali, and C. Diot. Characterization of failures in an operational IP backbone network. *Transactions on Networking*, 16(4), 2008.

[24] V. N. Padmanabhan, S. Ramabhadran, S. Agarwal, and J. Padhye. A study of end-to-end web access failures. In *Proceedings of CoNEXT*, pages 1–13, 2006.

[25] V. D. Park and M. S. Corson. A performance comparison of the temporally-ordered routing algorithm and ideal link-state routing. In *Proceedings of ISCC*, pages 592–598, 1998.

[26] V. Paxson. End-to-end routing behavior in the Internet. In *Proceedings of SIGCOMM*, pages 25–38, 1996.

[27] A. Shaikh, C. Isett, A. Greenberg, M. Roughan, and J. Gottlieb. A case study of OSPF behavior in a large enterprise network. In *Proceedings of IMC*, pages 217–230, 2002.

[28] A. U. Shankar, C. Alaettinoğlu, I. Matta, and K. Dussa-Zieger. Performance comparison of routing protocols using MaRS: Distance vector versus link-state. *ACM SIGMETRICS Performance Evaluation Review*, 20(1):181–192, June 1992.

[29] Shrubbery Networks, Inc. RANCID. http://www.shrubbery.net/rancid/.

[30] L. Tang, J. Li, Y. Li, and S. Shenker. An investigation of the Internet's IP-layer connectivity. *Computer Communications*, 32(5):913–926, 2009.

[31] University of Oregon. University of Oregon Route Views project. http://www.routeviews.org.

[32] F. Wang, Z. M. Mao, J. Wang, L. Gao, and R. Bush. A measurement study on the impact of routing events on end-to-end Internet path performance. In *Proceedings of SIGCOMM*, pages 375–386, 2006.

[33] D. Watson, F. Jahanian, and C. Labovitz. Experiences with monitoring OSPF on a regional service provider network. In *Proceedings ICDCS*, pages 204–212, 2003.

[34] W. Xu, L. Huang, A. Fox, D. Paterson, and M. Jordan. Detecting large-scale system problems by mining console logs. In *Proceedings of SOSP*, 2009.

[35] M. Zhang, C. Zhang, V. Pai, L. Peterson, and R. Wang. PlanetSeer: Internet path failure monitoring and characterization in wide-area services. In *Proceedings of OSDI*, pages 167–182, 2004.

APPENDIX

While the primary focus of this paper is the CENIC network, we also applied our methodology to the Internet2 network to establish a point of comparison [1]. As an added benefit, this allowed us to assess the effort required to apply our methodology to a different network. Here we briefly describe the Internet2 network and its measurement data, how we adapt the methodology for this dataset, and highlights of the results.

The Internet2 network. The Internet2 network (AS 11537) consists of nine Juniper T640 routers located in nine major US cities. These routers are interconnected with either 1-Gb/s or 10-Gb/s Ethernet links. The Internet2 network most closely resembles the CENIC HPR network, however HPR is more experimental in nature, while the Internet2 network is a production network. Like the CENIC network, the Internet2 network uses the IS-IS protocol. The network has been in operation since 1996 and serves 330 member institutions. We obtained router configuration snapshots and syslog message logs for the period 01/01/2009 to 12/31/2009. Internet2 operational announcements were also available, but because these required additional manual labor and were not essential to the analysis, we did not use them. We also did not use available IS-IS LSA logs because our methodology does not use this data source (it was unavailable for the CENIC network).

Adapting the methodology. The main challenge to processing Internet2 data is dealing with a different data format: the Internet2 network uses Juniper routers while the CENIC network uses Cisco routers. This required writing about 300 lines of new parsing code.

Brief summary. Table 4 in Section 6.2 shows the number of failures, downtime, and link time to repair in the Internet2 network. Its performance is somewhere between the DC network and HPR network with respect to number of failures and annual link downtime, with a longer time to repair. The Internet2 network also differs from the CENIC networks in the shape of the distribution of the time between failures (not shown): the CENIC networks have many short failures—indicative of flapping—while the Internet2 network has no such bias (the 25th percentile time between failures in the latter is over a minute and about 10 seconds in the DC network).