# 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,

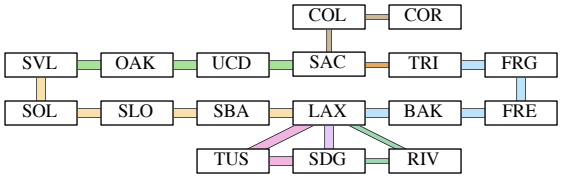


**Figure 1: CENIC hub sites are connected by an optical backbone. Long-haul fiber routes SVL to SAC, SVL to LAX, TRI to LAX, LAX–SDG, and LAX–TUS–SDG use the Cisco 15808 DWDM & 15454 platforms. Metro-area networks use a combination of CWDM and Cisco 15530 & 15540 DWDM equipment.**

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.

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.