Visualizing Sparse Internet Events: Network Outages and Route Changes

Lin Quan · John Heidemann · Yuri Pradkin

Received: 30 August 2012 / Accepted: 8 November 2012

Abstract To understand network behavior, researchers and enterprise network operators must interpret large amounts of network data. To understand and manage network events such as outages, route instability, and spam campaigns, they must interpret data that covers a range of networks and evolves over time. We propose a simple clustering algorithm that helps identify spatial clusters of network events based on correlations in event timing, producing 2-D visualizations. We show that these visualizations where they reveal the extent, timing, and dynamics of network outages such as January 2011 Egyptian change of government, and the March 2011 Japanese earthquake. We also show they reveal correlations in routing changes that are hidden from AS-path analysis.

1 Introduction

Researchers and network operators must interpret large amounts of network data each day to understand and manage their networks. Many different types of events happen in the Internet: networks become accessible and inaccessible [6, 11, 17, 19, 23, 24], congestion occurs and subsides, routes change [12, 13], spam campaigns and denial-of-service attacks wax and wane [7, 16]. In many cases, these events have common root causes: a shared router that fails [25], or a common change to software or a configuration [18], or distributed botnet with common external control.

One can infer common root causes of network events by studying their occurrence over space and time. By *space*, we mean the Internet topology or address space, since root causes are often associated with particular routers, Autonomous Systems (ASes), or address blocks. By *time*, we mean events can be reduced to discrete times or ranges, like network outages or bursts of spam messages. Events that are continuous (such as degree of congestion) are not our primary focus, but could be studied by looking for threshold changes.

USC/Information Sciences Institute
4676 Admiralty Way, Suite 1001, Marina del Rey, CA
E-mail: {linquan, johnh, yuri}@isi.edu

Our goal is to detect repeated correlations in time and space. We assume that events that are repeated in time suggest a common root cause. We wish to know how widely correlated they are in space to understand the potential extent of that cause. While correlation does not guarantee a cause, and the step from correlation to causation is necessarily specific to each event, repeated correlation can help network operators narrow the search space of problems. We intentionally make few assumptions about network topology, considering each observation independently and not considering topological hints offered by AS paths or address structure. We choose this "hands-off" approach since it can reveal *hidden* correlations—cases where networks with different addresses and AS paths share potentially related failure causes.

We propose a clustering algorithm that groups events that happen at similar *times* in network address *space*, supporting a two-dimensional visualization that reveals patterns (Section 2). The contribution of this paper is to show that simple clustering is helpful at determining correlated network events. We support this claim with examples of such events for network outages (Section 3) and for routing changes (Section 4). We find that clustering reveals the size and dynamics of network outages, and our approach to clustering routing changes provides evidence of correlations suggesting possible shared root causes that are not visible in AS-path-data alone.

Data from this paper is available at no cost from the authors, as described on-line [22].

2 Visualizing Correlated Events

We begin by describing how we identify clusters in a timeseries $\Omega_b(i)$ for an array of blocks b and discrete times (or *rounds*) i. Each element of $\Omega_b(i)$ takes a binary value indicating the presence or absence of an event. We later apply this method to two datasets: network outages (Section 3), where blocks are /24 address prefixes with 11-minute rounds; and network route changes (Section 4), where blocks are variable-size address prefixes from BGP with 2-hour rounds.

2.1 Clustering Visualization of Network Data

Our simple clustering algorithm groups the network event timeseries ($\Omega(\cdot)$) in two dimensions: time and space (Algorithm 1). We order blocks based on Hamming distance. For blocks m and n, with binary-valued timeseries $\Omega_m(i)$ and $\Omega_n(i)$, we define distance:

$$d_h(m, n) = \sum \Omega_m(i) \oplus \Omega_n(i).$$

Perfect temporal correlation occurs if $d_h(m, n) = 0$.

Since network events often require some time to propagate [12], we consider blocks with small variance of event times (less than a parameter θ) to be *close*. This approach may fail if there are two unrelated events with similar timing, but we believe that timing alone is often sufficient to correlate to larger events in today's Internet, provided we use a conservative θ.

Table 1 lists the survey datasets we use in this paper. Figure 1 shows the result of visualization clustering for network outages in Survey S_{38c}. This dataset uses a

Algorithm 1 Clustering of blocks for visualization

Input: A: the set of blocks, with $\Omega(\cdot)$ event timeseries
Output: B: reordered list of blocks, by distance

start with block $m \in A$ with smallest $\sum \Omega_m(i)$ (number of events)
$A = A \setminus \{m\}$
B.append(m)
while $A \neq \varnothing$ **do**
  for all n, s.t. $d_h(m,n) \leq \theta$ **do**
    $A = A \setminus \{n\}$
    B.append(n)
  end for
  // pick the next most similar block:
  find m' s.t. $d_h(m,m') \leq d_h(m,n) \forall n \in A$
  $A = A \setminus \{m'\}$
  B.append(m')
  $m = m'$
end while
return B

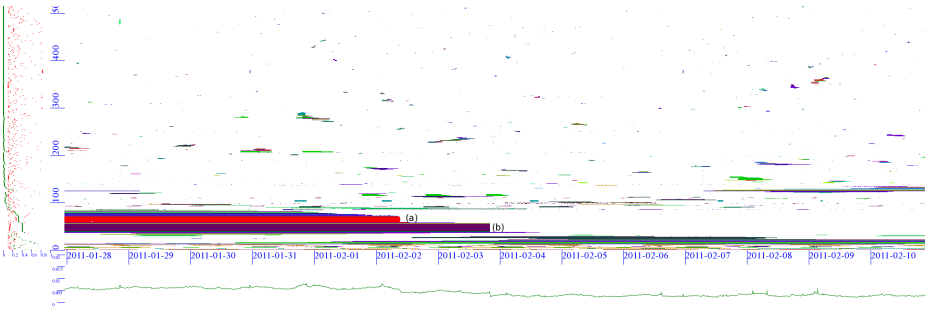

Fig. 1: The 500 largest outages of Survey S_{38c}, x axis: time, y axis: address space (blocks). Colors represent countries. Subgraphs on X and Y axis show marginal distributions (green line) and overall block responsiveness (red dots). Full interactive graph available as Supplement 1 in [22].

Survey	Start Date	Duration (days)	Blocks	Analyzable
S_{30w}	2009-12-23	14	22381	10629
S_{38c}	2011-01-27	14	22373	10553
S_{39w}	2011-02-20	16	22375	11585
S_{39c}	2011-03-08	14	22375	10955
whole Internet study	2011-09-28	1	—	2.5M

Table 1: Internet surveys used in this paper, with dates and durations. Survey numbers are sequential with a letter indicating collection location (w: ISI-west in Marina del Rey, CA; c: Colorado State U. in Ft. Collins, CO; j: Keio University, Fujisawa, Japan).

study of routing outages in about 10,600 blocks as described later in Section 3.1. The x-axis is time, each row shows the Ω_b downtime for a different /24 block b. Due to space constraints, we plot only the 500 blocks with most outages; we provide the full graph as Supplement 1 [22]. Color is keyed to the country to whom each block is allocated.

As an example of what clustering shows, we see that clusters of blocks that have near-identical outage end times. The cluster labeled (a) covers 19 /24s that are down for the first third of the survey; it corresponds to the Feb. 2011 Egyptian Internet shutdown. The cluster labeled (b) covers 21 /24 blocks for a slightly longer duration; it is an outage in Australia concurrent with flooding in the eastern coast. Since all large countries have disjoint address space, our clustering algorithm is essential to identify these large events; they are dispersed and therefore invisible if one views the data sorted by address or similar factors. Beyond this one example, Section 3.2 and Section 4.2 provides a more detailed discussion of insights from visualization.

Performance and Alternatives: Our clustering algorithm is a simple greedy algorithm; its performance is $O(b^2)$ for b blocks. Although we would prefer a faster algorithm, the largest possible b for IPv4 is 2^{24}; a value within the reach of current computers. We have done clustering for 2.5M blocks, processing time is about 35 minutes on a single core of a 2.2GHz Intel Xeon CPU.

We considered other standard clustering algorithms, including k-means and hierarchical agglomerative clustering, but found neither suitable for our problem. The k-means algorithm cannot be used because we do not know how many k clusters exist beforehand. Hierarchical agglomerative clustering has runtime $O(b^3)$, pushing its performance out of reach for our larger datasets.

2.2 Choice of the Closeness Threshold

Our clustering algorithm depends on θ, the threshold determining when blocks are close enough to cluster. We want to use a conservative θ, to be safe when declaring two blocks are close in temporal behavior. We use $\theta = 2$ rounds (22 minutes for outages, 4 hours for route changes), in clustering events. We have also studied much larger $\theta = 10$ (110 minutes for outages), finding the main difference is it tends to group more single-round events together [21] (omitted due to space).

2.3 Marginal Distributions

In addition to basic clustering, we find the *marginal distributions* are important to characterize the size of an event relative to the whole network. To evaluate events over the Internet as a whole, we next define statistical measures, how many and for how long blocks are experiencing events.

Figure 2, with data from S_{30w}, shows an example of marginal distributions (full figure as Supplement 5 in [22]). We see outages that affect many blocks for a short period (event (c) here, about 20 minutes), while others like (d) and (e) affect fewer blocks but for longer periods of time (here 2 to 3 hours).

Given N_b blocks and N_i times in an observation, the marginal distributions are the time- and space-specific sums:

$$\bar{\Omega}_I(i) = \sum_{b=1}^{N_b} \Omega_b(i) \qquad \bar{\Omega}_B(b) = \sum_{i=1}^{N_i} \Omega_b(i)$$

We normalize $\bar{\Omega}_I(i)$ by N_b and $\bar{\Omega}_B(b)$ by N_i in the subgraphs of some plots (such as Figure 1).

We show $\bar{\Omega}_I(i)$ along the x-axis. In our S_{30w} (Supplement 5 [22]), we see bumps in $\bar{\Omega}_I(i)$ late in the day on 2009-12-29, and midday on 2010-01-05. We show $\bar{\Omega}_B(b)$

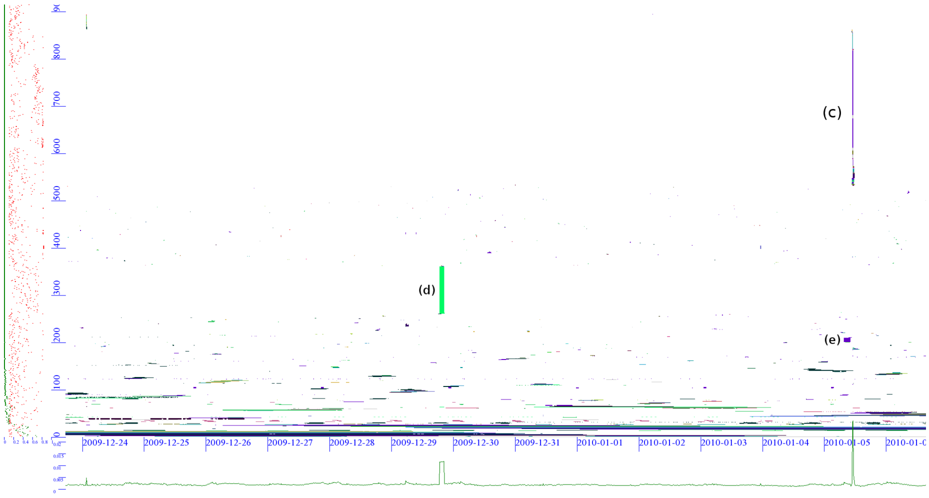

Fig. 2: The 900 largest outages of Survey S_{30w}, x axis: time, y axis: address space (blocks). Colors represent countries. Subgraphs on X and Y axis show marginal distributions (green line) and overall block responsiveness (red dots). Full interactive graph available as Supplement 5 in [22].

as the solid green line to the left of the y-axis. Because we sort by degree of outages, the largest $\bar{\Omega}_B(b)$ appear at the bottom of the graph. (For network outages, we also show the responsiveness of the target block in the margin, since sparse blocks can give misleading outage values. They appear as a speckle of red dots.)

When blocks are different sizes, as with route changes, we can compute the marginal distribution $\bar{\Omega}_I(i)$ in terms of numbers of prefixes and weighted by numbers of addresses. (Currently we show unweighted values.)

The above metrics are useful to characterize the degree of network events in today's Internet. We consider long-term trends and validation of these metrics as part of outage computation [21].

2.4 Handling Large Images

The IPv4 address space covers thousands of blocks, and events can span large parts of that space. Even at one pixel per block and round, visualizations are large: of 20k×1k for 2-week outage survey; 2.5M×130 for a 24-hour, whole Internet outage study (Supplement 6 in [22]); and 130k×360 for a 1-month routing change study. These large images quickly become difficult to view in traditional tools, where viewers typically assume the image fits in memory, and more robust tools (like Photoshop) are encumbered with editing. We have therefore implemented a custom, Google-maps-style web-based browser to make them more accessible. We use OpenLayers and the "slippy map" from Open Street Map [14]. Interactive examples of our visualizations are on the web in our supplemental data [22].

3 Visualizing Network Outages

Our first application is to visualize network outages. We use existing outage information derived from active probing [21]; we describe this source briefly below. Visualization is useful to characterize both large and small outages, and evaluate outage onset and recovery.

3.1 Data Sources: Detecting Outages

We visualize outage data covering from 10k to 2.5M /24 blocks from two data sources.

Outage Detection Through Probing: We draw on probing data from two sources. First, we consider Internet address surveys: active probing of all addresses in a sample of about 22k /24 address blocks in the IPv4 address space for two weeks [9]. Second, we probe 20 addresses in each /24 block for the entire measurable IPv4 address space (blocks with addresses that will respond), 2.5M blocks, for 24 hours [21].

In both cases, addresses are probed in a pseudo-random order, spreading probes out over an 11 minute round. We are careful to consider clock drift when mapping probes to rounds. We observe similar results from data taken from locations in California, Colorado, and Japan.

Detecting block-level outages: The overall responsiveness of block b at round i is defined as the fraction of responding addresses. We watch the change of the overall responsiveness at round i against recent behavior, and conclude an outage starts if we see a threshold-determined, dramatic drop in responsiveness (likewise for outage ends, with a dramatic increase of responsiveness). The output of this step is the binary timeseries $\Omega_b(i)$.

Details of data collection and validation of the method are outside the scope of this paper on visualization, but are available elsewhere [21]. In principle, our visualization methods can be applied to other sources of outage input, such as that from iPlane [17], Hubble [11], or localized methods [23]; such integration is future work.

3.2 Learning from Outage Visualization

Visualization has been important in developing our outage detection methods. We use visualization to quickly identify both large and typical events and to gain intuition about the dynamics of outages. Visualization is also important to diagnose problems in data collection and analysis. We describe each of these cases below.

Identifying Large Events From Background: The most important benefit of visualization is to provide an intuitive view of a large amount of data: quickly scanning a two-week survey allows one to "eye-ball" about 20M data points, yet large events and trends jump out from noisy background data.

As one example, the region marked (a) in Figure 1 corresponds to the Egyptian Internet outage in response to protests that resulted in the resignation of the Mubarak government by 2011-02-11 [24]. This survey began 2011-01-27 T23:07 +0000, just missing the beginning of the Egyptian network shutdown, and observed

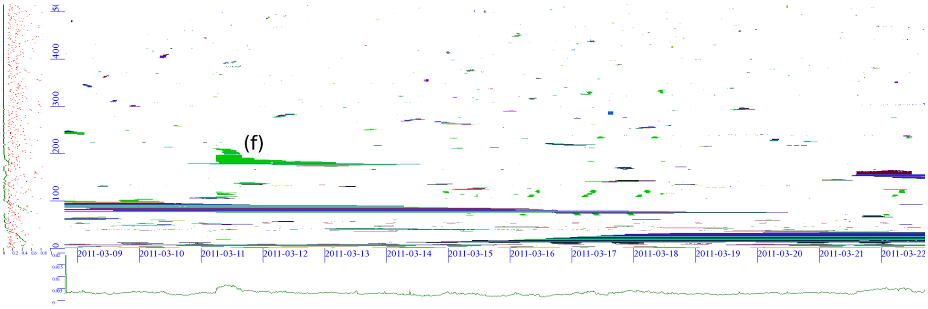

Fig. 3: The 500 largest outages in Survey S_{39c}, x axis: time, y axis: address space (blocks). Colors represent countries. Subgraphs on X and Y axis show marginal distributions (green line) and overall block responsiveness (red dots). Full interactive graph available as Supplement 2 in [22].

the restoration of network service around 2011-02-02 T09:28 +0000. Our survey covered 19 responsive /24 blocks in the Egyptian Internet (about 1% of responsive networks in Egypt). The use of country-specific coloring helps this outage stand out in our visualization.

A second interesting example is a significant Australian outage in the same dataset, marked (b) in Figure 1. We are able to locate these blocks in the east coast of Australia, including Sydney and Brisbane. Private communications [1] and the AusNOG mailing list [2] suggest this outage may be related to mid-January flooding in eastern Australia. The recovery of the network seems consistent with news reports about telecommunications repairs [10]. Visualization is important in this case to demonstrate that the outage was about *as large and long-lasting* as the Egyptian outage, yet the Egyptian Internet outage made global news while the Australian outage got little discussion. The Egyptian outage was more newsworthy both because of the political significance, and because it represented nearly all Egyptian traffic. Australia, by comparison, has eight times more allocated IPv4 addresses than Egypt, so though the Australian outage may be as large as the Egyptian one, it does not have the same country-wide impact. Visualization was the primary reason this outage came to our attention, as there was minimal news coverage, and nearly none that emphasized network problems.

Although our visualization methods help large events to stand out, and can provide an overview of small events, ultimately automated tools are necessary to consistently find and characterize events. We believe the large events described here show the usefulness of our visualization, but we refer interested readers to our technical report for discussion of tools to quantify large and small outages, and for analysis of a random sample of events [21].

Understanding Dynamics of Outages: In addition to presence of outages, visualization is ideal to reveal the *dynamics of complicated outages*. For example, in Survey S_{39c}, we see Japanese network outages related to the Tōhoku Japanese earthquake of 2011-03-11 [19], marked (f) in Figure 3. Unlike most other outages we observe, both the start and recovery from this outage vary in time. For many blocks, the outage begins at the exact time of the earthquake, but for some it occurs two hours later. Recovery for most blocks occurs within ten hours, but a few remain down for several days. Analysis by IIJ provides details about Japanese network outages and recoveries from the perspective of a network operator [5], confirming our observations. The contribution of our visualization method is to provide some information about the impact of this natural disaster on the network, but observable externally by non-experts on Japanese network topology.

Quantifying Typical Outages: In addition to major events, visualization also helps characterize typical network behavior. We can see this partly in the scattering of outages in Figure 1. We can also see that outages come in different "shapes", with many small and short, some long duration, like (a) and (b) in this figure, and others are "taller" (affecting more networks), like (g) in Figure 4.

We have found ground truth for several regional outages of different shapes. Survey S_{30w}, omitted here for space but available as Supplement 5 in [22] and as Figure 3 in [21], shows three regional network outages: (c) a Verizon outage affecting many (331) /24 blocks for a short time (22 minutes); (d) an AT&T/Comcast outage affecting fewer (12) blocks for longer time (165 minutes); and (e), a large Mexico outage affecting 105 blocks. These three events show how visualization quickly characterizes the different shapes of typical events, from tall events like (c) that affect many networks and multiple countries, to fatter, longer outages like (d) and (e). It also shows the value of marginal distributions; both of these significant regional events show up as bumps in $\bar{\Omega}_I$ running along the x-axis.

Detecting Problems in Local Networks and Data Collection: Finally, we have found visualization vital to gaining confidence in our measurement approach by revealing problems in our network and software. Figure 4 (from Survey S_{39w}) shows examples of three problems we detected with visualization and then corrected.

First, event (g) shows a long-term outage of over 200 networks, all at USC. In diagnosing this problem, we discovered that USC's network operators chose to block our probes, in spite of pre-authorization. Visualization helped us detect and repair this misunderstanding.

Events (h) and (i) both show outages across all monitored blocks (although this visualization shows only 500 of them), corresponding to planned maintenance activities. Event (h) was planned maintenance in our server room; the blue color indicates absence of data because we temporarily suspended data collection. Event (i) was a second planned power outage that took down a router near our survey machines, although probes continued running. More subtly, the smattering of white "up" blocks in event (i) shows a data collection problem. The combination of actively probing 20k blocks, *and* all probes timing out due to a powered-off second-hop router, results in the overflow of connection tracking (iptables) at our probing machine. Such overflow produces random ICMP network unreachable error replies at the probing host, which are then interpreted incorrectly as a response from the remote network. We have since corrected this problem by disabling ICMP connection tracking, but it was detected only because of differences in visualization.

4 Visualizing Routing Changes

We have also used our clustering algorithm to visualize BGP path changes. We find that visualization is helpful in finding correlated BGP path changes, sometimes showing common failure sources that are not obvious from simple inspection of AS

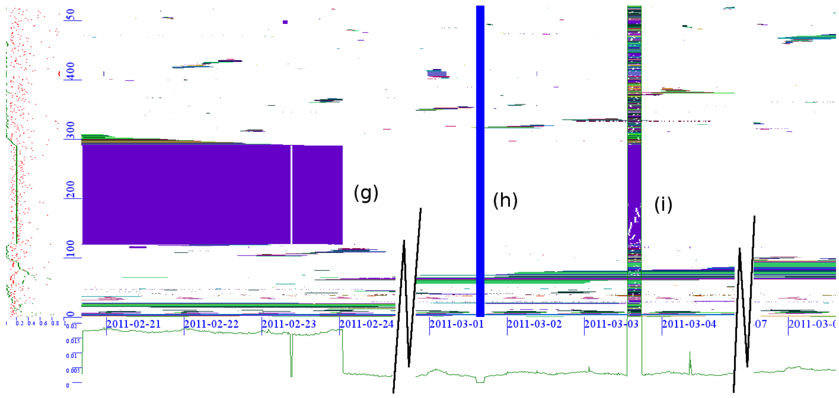

Fig. 4: The 500 largest outages in S_{39w}, x axis: time, y axis: address space (blocks). Colors represent countries. Subgraphs on X and Y axis show marginal distributions (green line) and overall block responsiveness (red dots). Full interactive graph available as Supplement 3 in [22].

paths. Visualization of path changes also shows that our methodology generalizes to other kinds of network timeseries.

4.1 Data Sources: Detecting BGP Path Changes

We collect BGP snapshots with BGPmon [27], at three sites in California, Colorado, and Japan, taking snapshots every two hours. To visualize this data, we consider each announced, routable IPv4 prefix as a *block*, and the two-hour collection interval as *round* duration. (Finer time resolution is future work.)

To generate our $\Omega_b(i)$ timeseries, we compare the AS path at round i with round $i-1$ for each prefix (block) b, setting $\Omega_b(i) = 1$ if the paths differ. We apply our clustering visualization algorithm to $\Omega_b(i)$, to find groups of BGP prefixes that show similar timing across repeated route changes.

4.2 Learning from Visualizing Route Changes

Just as with network outages (Section 3.2), we find that visualization of route changes can be used to quickly identify large events, to understand dynamics, and to quantify these changes with marginal distributions.

Comparing Route Changes to Outages: Comparing route changes to outages, visualization makes it clear that route changes affect many more prefixes and more frequent but shorter than network outages. For example, route changes in Figure 5 affect hundreds of blocks, each of which is often much larger than a /24, while outages (such as Figure 3) typically affect only a few /24s, but often for much longer. Such comparisons must be made carefully, though, since the time and spatial scales of our visualizations are different. In both cases, though, repeated temporal correlations are useful to suggest blocks that behave similarly.

Finding Hidden Prefix Correlations: Prior approaches have studied AS paths to detect common points of failure [4]. Our approach, however, uses only

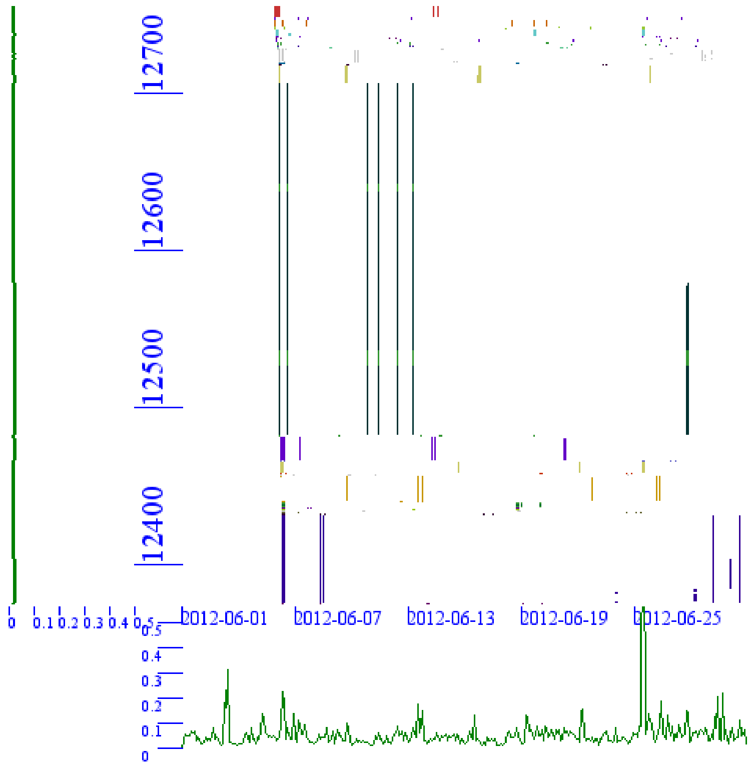

Fig. 5: Sample cluster: correlated BGP changes for China prefixes, in June 2012. X-axis: time at 2 hour rounds. Y-axis: prefixes. Full interactive graph available as Supplement 4 in [22].

timing to cluster blocks into events, thus it can reveal correlations that are not apparent from AS paths alone. We describe several examples below using data from our Japanese routing vantage point taken in June 2012.

An Example of Correlated Routing Problems: To illustrate a correlation seen in our data, we observe several correlated route changes affecting 242 routing prefixes on 3 days, as shown in Figure 5.

While correlated routing changes are expected, but it is surprising when correlated appear to be caused at different points in the AS path. Of the cluster, 101 prefixes switch between AS paths *2500 2914 4837 17816* and *2500 2914 3356 4837 17816*, suggesting that AS2914 (NTT) is selecting between AS4837 (China Unicom) and AS3356 (Level3), while 131 prefixes switch between *2500 2914 4837 17816* and *2500 2914 4837 4837 17816*, apparently due to a problem in AS4837. These differences would not be seen from AS path analysis only, since the AS paths of the first block suggest the problem is between AS2914 and AS4837, while in the second path it appears to be between AS4837 and AS17816.

A similar problem affects 116 prefixes in Australia, where many 115 prefixes oscillate between paths *2500 2914 4713 2516 4637 1221 38285 10113* and *2500 2914 4713 2516 4637 1221 38285 38285 38285 10113*, suggesting changes in AS38285;

while another oscillates between two very different paths: *2500 2914 4713 2516 4637 1221 38285* and *2500 2914 3257 7473 7474 38285*.

Frequency of Hidden Correlations: Hidden correlations are not uncommon. We examined all 8716 clusters in our June 2012 data and found that about 11% showed such hidden prefix correlations. Analyzing the root causes of such correlations is future work, outside the scope of this visualization paper, but we believe it demonstrates that visualization and our simple clustering algorithm can reveal hidden correlations.

5 Related Work

Although there is significant prior work on routing outages and route changes, there has been much less work *visualizing* network phenomena.

Several prior papers have visualized numbers of outages by time, including Markopolou et al. [20], Turner et al. [25], and Gill et al. [8]. While showing time-series, none of these attempt to cluster the blocks to bring out correlations. Our clustering would make the correlations they discuss more obvious.

Data clustering is a well established field and there are many generic clustering algorithms (refer to [15, 26] for detailed surveys), we employ a simple greedy algorithm most suitable for our problem (Section 2.1).

Somewhat related to our visualization work is CAIDA's AS dispersion graphs, which cluster and visualize based on the common AS paths from one point to many destinations [3]. This intuitive visualization shows the AS level connectivity seen at a vantage point. We complement their work by clustering the changes of AS paths, instead of static analysis.

For a review of related work beyond visualization, we refer readers to our evaluation of outage detection [21].

6 Conclusions

We show that clustering and visualization provides a useful way to analyze and characterize network timeseries data, such as block-level outages found by active probing, and BGP path changes. Visualization and simple clustering provides new insights for both domains.

Acknowledgments

We thank Jim Koda (ISI), Brian Yamaguchi (USC), and CSU network operations for providing BGP feeds to assist our evaluation, and Dan Massey, Christos Papadopoulos, Mikhail Strizhov for assisting with BGPmon and at CSU. We also thank Katsuhiro Horiba (WIDE) for providing probing infrastructure and BGP feeds. This work was reviewed by USC's IRB (IIR00000975) and identified as non-human subjects research.

References

1. Grenville Armitage. Private communications, Jul. 2011.

2. AusNOG. Discussions about Australia flooding, Jan. 2011. http://lists.ausnog.net/pipermail/ausnog/2011-January.
3. CAIDA. As dispersion graphs, 2012. http://www.caida.org/projects/ark/statistics/san-us/as_dispersion_by_as.html.
4. Di-Fa Chang, Ramesh Govindan, and John Heidemann. The Temporal and Topological Characteristics of BGP Path Changes. In *Proc. of ICNP*, November 2003.
5. Kenjiro Cho, Cristel Pelsser, Randy Bush, and Youngjoon Won. The Japan earthquake: the impact on traffic and routing observed by a local ISP. In *Proc. of ACM Special Workshop on Internet and Disasters*, pages 2:1–2:8, Tokyo, Japan, December 2011. ACM.
6. Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide Internet outages caused by censorship. In *ACM IMC*, 2011.
7. Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y. Zhao. Detecting and characterizing social spam campaigns. In *ACM IMC*, pages 35–47, 2010.
8. Phillipa Gill, Navendu Jain, and Nachiappan Nagappan. Understanding network failures in data centers: measurement, analysis, and implications. In *Proceedings of the ACM SIGCOMM 2011 conference*, SIGCOMM, pages 350–361, 2011.
9. John Heidemann, Yuri Pradkin, Ramesh Govindan, Christos Papadopoulos, Genevieve Bartlett, and Joseph Bannister. Census and Survey of the Visible Internet. In *Proc. of ACM IMC*, Oct. 2008.
10. International Business Times. Optus, Telstra see service outages after Cyclone Yasi, 2011. http://hken.ibtimes.com/articles/108249/20110203/optus-telstra-see-service-outages-after-cyclone-yasi.htm.
11. Ethan Katz-Bassett, Harsha V. Madhyastha, John P. John, Arvind Krishnamurthy, David Wetherall, and Thomas Anderson. Studying black holes in the Internet with Hubble. In *NSDI*, 2008.
12. Craig Labovitz, Abha Ahuja, Abhijit Bose, and Farnam Jahanian. Delayed Internet routing convergence. In *Proc. of SIGCOMM*, 2000.
13. Craig Labovitz, G. Robert Malan, and Farnam Jahanian. Internet routing instability. In *Proc. of SIGCOMM*, 1997.
14. Open Layers. Openlayers: Free maps for the web. web site http://openlayers.org, 2012.
15. T. Warren Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857 – 1874, 2005.
16. Xin Liu, Xiaowei Yang, and Yanbin Lu. To filter or to authorize: network-layer DoS defense against multimillion-node botnets. In *SIGCOMM*, pages 195–206, 2008.
17. Harsha V. Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani. iPlane: an information plane for distributed services. In *OSDI*, 2006.
18. Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP misconfiguration. In *Proc. of SIGCOMM*, 2002.
19. Om Malik. In Japan, many undersea cables are damaged. GigaOM blog, http://gigaom.com/broadband/in-japan-many-under-sea-cables-are-damaged/, Mar. 14 2011.
20. Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen nee Chuah, and Christophe Diot. Characterization of Failures in an IP Backbone. In *Proc. of IEEE Infocom*, 2004.
21. Lin Quan, John Heidemann, and Yuri Pradkin. Detecting Internet outages with precise active probing (extended). Technical Report ISI-TR-2012-678, February 2012.
22. Lin Quan, John Heidemann, and Yuri Pradkin. Supplemental data and visualizations, September 2012. http://www.isi.edu/ant/pubs/network-vis/.
23. Aaron Schulman and Neil Spring. Pingin' in the rain. In *Proc. of ACM IMC*, pages 19–28, 2011.
24. New York Times. Egypt cuts off most Internet and cell service. http://www.nytimes.com/2011/01/29/technology/internet/29cutoff.html.
25. Daniel Turner, Kirill Levchenko, Alex C. Snoeren, and Stefan Savage. California fault lines: understanding the causes and impact of network failures. In *Proc. of SIGCOMM*, 2010.
26. Rui Xu and II Wunsch, D. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645 –678, May 2005.
27. He Yan, Ricardo Oliveira, Kevin Burnett, Dave Matthews, Lixia Zhang, and Dan Massey. BGPmon: A real-time, scalable, extensible monitoring system. In *Proc. of IEEE CATCH*, March 2009.