Outages from natural disasters, political events, software or hardware issues, and human error [2] place a huge cost on e-commerce ($66k/minute at Amazon [1]).

While several existing systems detect Internet outages [4–8], these systems often too inflexible, fixed parameters across the whole internet with CUSUM-like change detection. We instead propose a system using passive data, to cover both IPv4 and IPv6, customizing parameters for each block to optimize the performance of our Bayesian inference model.

Our poster describes our three contributions: First, we show how customizing parameters allows us often to detect outages that are at both fine timescales (5 minutes) and fine spatial resolutions (/24 IPv4 and /48 IPv6 blocks). Our second contribution is to show that, by tuning parameters different for different blocks, we can scale back temporal precision to cover more challenging blocks. Finally, we show our approach extends to IPv6 and provide the first reports of IPv6 outages.

Approach summary: Our approach uses passive traffic observations from network-wide services. In our evaluation we consider traffic from B-root’s DNS service, but in principal we could use data from a large website (like Wikipedia, Google, or Amazon) or other infrastructure (like NTP). We build a model of historical traffic from each source to the service, then detect interruptions that violate model history, using with Bayesian inference to detect outages.

Some strong sources directly support detection, but when possible, we correlate multiple signals from the same region to corroborate our observations. From network-wide services. In our evaluation we consider B-root observations from B-root and Trinocular’s observations.

Table 1: Confusion matrix for long-duration outages (in seconds)

<table>
<thead>
<tr>
<th>Observation (B-root)</th>
<th>Ground truth (Trinocular)</th>
</tr>
</thead>
<tbody>
<tr>
<td>availability outage</td>
<td>availability outage</td>
</tr>
<tr>
<td>TP = ta = 52525765695</td>
<td>TP = ta = 2471178</td>
</tr>
<tr>
<td>FN = fo = 78163261</td>
<td>FN = fo = 13147965</td>
</tr>
<tr>
<td>Recall 0.9985</td>
<td>TNR 0.84178</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix for long-duration outages on dense blocks (in seconds)

<table>
<thead>
<tr>
<th>Observation (B-root)</th>
<th>Ground truth (Trinocular)</th>
</tr>
</thead>
<tbody>
<tr>
<td>availability outage</td>
<td>availability outage</td>
</tr>
<tr>
<td>TP = ta = 7644527262</td>
<td>TP = fa = 77152</td>
</tr>
<tr>
<td>FN = fo = 387011</td>
<td>FN = to = 2233042</td>
</tr>
<tr>
<td>Recall 0.99</td>
<td>TNR 0.96</td>
</tr>
</tbody>
</table>

Table 3: Confusion matrix for short-duration outages on dense blocks (in seconds)

<table>
<thead>
<tr>
<th>Observation (B-root)</th>
<th>Ground truth (Trinocular)</th>
</tr>
</thead>
<tbody>
<tr>
<td>availability outage</td>
<td>availability outage</td>
</tr>
<tr>
<td>TP = ta = 154956784</td>
<td>TP = fa = 154956784</td>
</tr>
<tr>
<td>FN = fo = 387011</td>
<td>FN = to = 2233042</td>
</tr>
<tr>
<td>Recall 0.99</td>
<td>TNR 0.96</td>
</tr>
</tbody>
</table>

Detecting short outages: Prior work either detects 11-minute outages at fine spatial scales (typically /24 blocks, [6, 7]) or 5-minute outages, but for coarse spatial scales (entire ASes, [4, 8]), or very fast reaction but with a very large amount of input data (seconds, but requiring all TCP flows [5]). In each case, prior systems only improve temporal resolution by increasing active traffic, passive spatial scale or input data. Our new approach interprets passive data and can employ exact timestamps of observed data, allowing both fine spatial and temporal precision in many cases.

Although our approach can apply to many traffic sources, we quantify its effectiveness for both long-duration and short-duration outages, by evaluate B-root [9] as a passive data source. We test against Trinocular active outage detection [6] and observations from RIPE Atlas (inspired by Chocolatine [4]) as ground truth. Since B-root coverage is limited, we compare only /24 IPv4 blocks that overlap between our observations from B-root and Trinocular’s observations.

We evaluate our accuracy in the confusion matrix in Table 1. We define a false outage (fo) as a prediction of down when it’s really up in Trinocular, with analogous definitions of false availability (fa), true availability (ta), and true outages (to). High precision means the outages that we report are true, and strong recall means we correct estimate duration. TNR suggests that we often find shorter outages than Trinocular. This difference may be due to Trinocular’s precision (±330 s), while using exact timestamps of data allows us to be more precise and often shorter. To avoid uncertainty and show our model works for finding the maximum number of outages we also test on only dense data having high frequency of traffics in Table 2. This shows that we have very good precision and recall for very dense blocks and TNR shows that we can detect 96% of the outages. Evaluation of short outages is challenging: precision of ±180 s hides differences in uncertainty for short outages (300 s or less). The poster will compare short outages by events (not time) to factor out imprecision in timing.

In the Table 3 we report outage events that are 5 minutes in length for both B-root and RIPE data. Our results show that we have great precision (0.9979) and recall (0.9479) indicating our model has good accuracy in long-duration outages (11 minutes or more). Similarly, we have great precision (0.9769) and recall (0.9453) for short-duration outages (5 minutes or more). Our measurements show that on 2019-01-10, around 5% of total blocks that have 5 minute outages that were not seen in prior work. These short outages add up—when we add the outages from 5 to 11 minutes that were previously omitted to observations, we see that total outage duration increases by 20%.

Optimizing across a diverse Internet: Because the Internet is so diverse, outage detection systems need to be tuned to operate differently for differently-behaving regions. We describe the first passive system that optimizes parameters to each block to provide fine spatial and temporal precision when possible, but falling back on coarser temporal precision when necessary. By contrast, prior
Table 3: Confusion matrix for short-duration outages (events)

<table>
<thead>
<tr>
<th>Coverage (B-root)</th>
<th>Ground truth (RIPE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>availability (events)</td>
<td>outage (events)</td>
</tr>
<tr>
<td>4445</td>
<td>105</td>
</tr>
<tr>
<td>257</td>
<td>290</td>
</tr>
<tr>
<td>0.97692</td>
<td></td>
</tr>
<tr>
<td>0.9453</td>
<td>0.7341</td>
</tr>
</tbody>
</table>

Figure 1: Trading temporal and spatial precision.

IPv6: IPv6 is a growing part of the Internet today, and of course it has outages, but prior outage-detection systems have not extended to IPv6. Prior active monitoring systems cannot possibly probe all unicast IPv6 address, since 2^128 addresses requires centuries to scan, and privacy-preserving addressing makes most client addresses ephemeral. Our new approach extends coverage to IPv6, by analyzing passive data, allowing the active addresses to come to us.

We evaluate our IPv6 coverage based on one representative day of passive data from B-root, comparing results in IPv4 and IPv6. In Figure 2a we see 11,918 /48 IPv6 blocks that are measurable (they have enough data to provide a reliable outage signal), and we see at least one 10 minute outages in 1338 (12% of measurable blocks). By comparison, the same system sees 167,851 /24 IPv4 blocks that are measurable, and 8689 /24 IPv4 blocks have one 10 minute outage (5.5% of measurable blocks). The absolute number of IPv4 outages is larger than IPv6 because there are far more measurable IPv4 blocks. However, the outage rate (the percentage measurable blocks with outages) for IPv6 seems somewhat greater than for IPv4, suggesting IPv6 reliability can improve.

Our coverage in IPv6 is surprisingly large: Figure 2b compares coverage relative to best prior system. For IPv6, our approach with B-root sees about 12,765 IPv6 /48 blocks, 17% of the 74,373 /48 blocks in the Gasser IPv6 hitlist [3]. This coverage is similar to what we see in IPv4, where the 1M /24s blocks our system with B-root is about 20% of the 5.1M in Trinocular. In both cases, B-root coverage is limited (it sees only recursive resolvers), but it seems about the same fraction of IPv6 as IPv4. We expect to add additional passive sources to increase IPv6 coverage.

Although our work is still in progress, our early results suggest a significant advance in the ability to observe both long and short outages of the network edge for both IPv4 and IPv6. We show that users can tradeoff between spatial and temporal precision depending on the type of data. We show our IPv6 provides coverage similar to our IPv4 coverage, and provide the first report IPv6 outages.

This work was supported by NSF grant CNS-2007106 (EIEIO).

REFERENCES


Internet Outage Detection Using Passive Analysis

Asma Enayet
enayet@usc.edu

John Heidemann
johnh@isi.edu

Introduction

Our goal is to detect both short and long network outages in IPv4 and IPv6. Current outage detection systems are limited to longer outages (or large networks) and IPv4. Detecting all outages in all networks is important to understand outages caused by natural disasters, political events, software and hardware issues, and human error, and their cost on today’s e-commerce (an outage costs Amazon $66k/minute).

We propose a new, principled approach to outage detection using passive data, to cover both IPv4 and IPv6, customizing parameters for each block to optimize the performance of our Bayesian inference model.

Contributions:

• We can detect short-duration outages as we control time precision
• We are the first to report outages in IPv6 address space because we see data from clients that exists.
• We are the first to exploit a trade-off between spatial and temporal precision

Problem Statement

We know short-duration outage occur, but current systems do not measure them:
• Prior active detection (like Trinocular and Thunderping) cannot increase temporal precision without becoming overly intrusive.
• Prior passive detection systems (like Chocolatine) can detect 5-minute (short) outages, but only for coarse (AS-level) spatial precision.
• Our approach allow fine spatial and temporal precision (24 and 5 minutes) where possible through adaptive estimation of exact timing of passive data.

Because the Internet is so diverse, outage detection systems need to be tuned to operate differently for differently-behaving regions.
• Prior passive systems are homogeneous, with the same parameters across all blocks, yielding only coarse spatial coverage or limited coverage.
• We exploit the ability to trade-off between spatial and temporal precision.
• Each block uses different, custom parameters.
• Different regions vary temporal and spatial precision to retain coverage.
• We provide broad coverage while providing fine precision when possible by lowering temporal precision when required and adapting parameters per block.

Of course there are outages in IPv6, but prior outage-detection systems have not reported on IPv6, a growing part of the Internet today:
• Prior active monitoring systems cannot cover all unicast IPv6 (it’s too big!).
• Prior passive systems were not parameterized for IPv6.
• We provide the first results for IPv6, using our parameterized passive system.

Detection Algorithm

• Goal: detect short and long outages for the overall space
• Input: address blocks with second-level timing for traffic arrival from an Internet wide service (we use B-Root DNS, but big websites would work).
• Output: availability and outages, with start time and duration.
• Procedure:
  • P(t) the rate at which traffic appears
  • Two extremes of P(t) is dense and sparse blocks
  • Bayesian inference to calculate the belief B(a) of the next time bin.
  • Finally, judge blocks as either down or up
  • Belief, B(a) ranges from 0(DOWN) to 1(UP)

Example below show two extremes: a block with dense traffic (top) and sparse traffic (bottom) where belief often varies.

Examples below show two extremes: a block with dense traffic (top) and sparse traffic (bottom) where belief often varies.

Introduction

Our goal is to detect both short and long network outages in IPv4 and IPv6. Current outage detection systems are limited to longer outages (or large networks) and IPv4. Detecting all outages in all networks is important to understand outages caused by natural disasters, political events, software and hardware issues, and human error, and their cost on today’s e-commerce (an outage costs Amazon $66k/minute).

We propose a new, principled approach to outage detection using passive data, to cover both IPv4 and IPv6, customizing parameters for each block to optimize the performance of our Bayesian inference model.

Contributions:

• We can detect short-duration outages as we control time precision
• We are the first to report outages in IPv6 address space because we see data from clients that exists.
• We are the first to exploit a trade-off between spatial and temporal precision

Problem Statement

We know short-duration outage occur, but current systems do not measure them:
• Prior active detection (like Trinocular and Thunderping) cannot increase temporal precision without becoming overly intrusive.
• Prior passive detection systems (like Chocolatine) can detect 5-minute (short) outages, but only for coarse (AS-level) spatial precision.
• Our approach allow fine spatial and temporal precision (24 and 5 minutes) where possible through adaptive estimation of exact timing of passive data.

Because the Internet is so diverse, outage detection systems need to be tuned to operate differently for differently-behaving regions.
• Prior passive systems are homogeneous, with the same parameters across all blocks, yielding only coarse spatial coverage or limited coverage.
• We exploit the ability to trade-off between spatial and temporal precision.
• Each block uses different, custom parameters.
• Different regions vary temporal and spatial precision to retain coverage.
• We provide broad coverage while providing fine precision when possible by lowering temporal precision when required and adapting parameters per block.

Of course there are outages in IPv6, but prior outage-detection systems have not reported on IPv6, a growing part of the Internet today:
• Prior active monitoring systems cannot cover all unicast IPv6 (it’s too big!).
• Prior passive systems were not parameterized for IPv6.
• We provide the first results for IPv6, using our parameterized passive system.

Detection Algorithm

• Goal: detect short and long outages for the overall space
• Input: address blocks with second-level timing for traffic arrival from an Internet wide service (we use B-Root DNS, but big websites would work).
• Output: availability and outages, with start time and duration.
• Procedure:
  • P(t) the rate at which traffic appears
  • Two extremes of P(t) is dense and sparse blocks
  • Bayesian inference to calculate the belief B(a) of the next time bin.
  • Finally, judge blocks as either down or up
  • Belief, B(a) ranges from 0(DOWN) to 1(UP)

Example below show two extremes: a block with dense traffic (top) and sparse traffic (bottom) where belief often varies.

Example below show two extremes: a block with dense traffic (top) and sparse traffic (bottom) where belief often varies.

Validation

We compare seven days (2019-01-09 to 2019-01-15) covering around 900,000 blocks, evaluating outage duration.

Long outages: Comparing against active probing
• We compare our system to Trinocular data for one day (2019-01-10) to validate.
• We maximize the duration of true outages detection.

We have great precision and recall (>0.99) : our model has good accuracy.

The True Negative Rate (0.84) means that we can detect 84% of the outages.

Trading between temporal and spatial precision

• We have good precision for dense blocks, but reduce precision for sparse blocks

⇒ Researchers can trade spatial or temporal precision for coverage as desired

Results

Can we detect short-duration outages?
• Detects short outages (5 min.) for fine blocks (24s), unlike prior work.
• Here we compare our 5-minute outages against RIPE Atlas to validate our method.
• We compare 24 hours on 2019-01-10, for 600 blocks having traffic from both B-root and RIPE.
• Here we compare outage events, since timing precision (≤180) hides differences in duration

We have great precision (0.971) and recall (0.949) for short outages.

The good True Negative rate (0.734) indicates that we can detect 73% of the outages...

Extending to IPv6

Coverage report- IPv4 vs IPv6:
• The fractions of best prior works’ coverage are almost similar for both IPv4 (19.6% of Trinocular’s) and IPv6 (17% of Gasser’s)

We provide the first published reports on IPv6 reliability.

We find that IPv6 is not as reliable as IPv4.

Plans and Conclusions

• We plan to extend this work to other passive data sources (such as darknets) to increase coverage.

Conclusions:
• We are able to detect short outages (5 min.) with fine spatial precision (~24s), and can relax precision to increase coverage.
• We provide the first published results of IPv6 network reliability.
• The key to our approach is adapting parameters to match each block.

For data and more information, see https://ant.isi.edu/datasets/outage_passive/