

Characterizing Correlated Latency Anomalies in Broadband Access Networks

Swati Roy and Nick Feamster
School of Computer Science, Georgia Tech
Atlanta, GA, USA
swati.gt@gatech.edu, feamster@cc.gatech.edu

ABSTRACT

The growing prevalence of broadband Internet access around the world has made understanding the performance and reliability of broadband access networks extremely important. To better understand the performance anomalies that arise in broadband access networks, we have deployed hundreds of routers in home broadband access networks around the world and are studying the performance of these networks. One of the performance pathologies that we have observed is correlated, sudden latency increases simultaneously and to multiple destinations. In this work, we provide an preliminary glimpse into these sudden latency increases and attempt to understand their causes. Although we do not isolate root cause in this study, observing the sets of destinations that experience correlated latency increases can provide important clues as to the locations in the network that may be inducing these pathologies. We present an algorithm to better identify the network locations that are likely responsible for these pathologies. We then analyze latency data from one month across our home router deployment to determine where in the network latency issues are arising, and how those pathologies differ across regions, ISPs, and countries. Our preliminary analysis suggests that most latency pathologies are to a single destination and a relatively small percentage of these pathologies are likely in the last mile, suggesting that peering within the network may be a more likely culprit for these pathologies than access link problems.

Categories and Subject Descriptors

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

General Terms

Measurement, Performance

Keywords

Measurement; Active probing; Performance

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGCOMM'13, August 12–16, 2013, Hong Kong, China.
ACM 978-1-4503-2056-6/13/08.

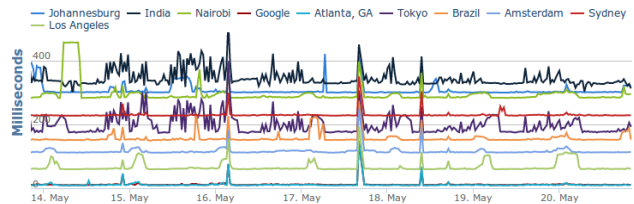


Figure 1: An example of a home broadband access network experiencing correlated latency increases to multiple destinations simultaneously. We are developing location algorithms to determine rough network locations for where these pathologies are occurring.

1. MOTIVATION

Despite the growing prevalence and penetration of broadband access networks around the world, little is known about the performance pathologies that these networks experience over time. To better understand longitudinal performance of broadband access networks, we have deployed hundreds of customized home routers in broadband access networks around the world, as part of a testbed that we call BISmark (Broadband Internet Service Benchmark) [2]. We have been collecting performance data from this deployment for more than two years; this deployment thus allows us to answer questions about performance problems that arise over time, as well as the potential sources of these pathologies.

One of the more compelling persistent performance pathologies that we have observed is that of correlated latency increases from a broadband access network to multiple destinations. Figure 1 shows an example of several such events, as observed from a single broadband access network. Although it is tempting to conclude that these pathologies are most likely caused by problems close to the access router, even a casual observer will note that many of the latency increases depicted in Figure 1 are observed to only some fraction of destinations, not all—indicating that the problem is perhaps not, in fact, caused by the access link, but rather by problems somewhere along the end-to-end path.

Our goal in this study is to provide a first look into these correlated latency increases. Based on whether a latency increase is experienced across one or more routers and is observed to one or more servers, we classify each observed latency increase into one of the following three categories: (1) last-mile access pathology; (2) last-mile server pathology; (3) middle-of-path pathology. Our preliminary analysis suggests that many of the latency problems that we observe are in fact only observed to a single destination, suggesting that broadband access link latency bottlenecks may not be as common as other types of latency bottlenecks.

Algorithm 1 Correlation-based diagnosis.

Input: End-to-end and last-mile latency measurements from home routers.

Output: Detect spikes and classify them into three categories: a) spikes in last-mile latency b) spikes in RTT to a single server c) spikes in RTT to multiple servers.

Ensure: Latency value crossing a pre-determined threshold is classified as latency spike.

```
while Not all routers are covered do
  Read latency spike
  if RTT spike to all the servers simultaneously then
    if Spike in the last-mile latency value then
      Classify as last-mile anomaly
    else
      Classify as middle-mile anomaly
    end if
  else
    Classify as server-side anomaly
  end if
end while
```

2. APPROACH

Our approach for analyzing correlated latency increases has two steps: (1) anomaly detection; (2) diagnosis, based on correlation. The anomaly detection step simply determines whether a latency increase observed by the router to a particular destination constitutes an anomaly, based on whether the increase is uncharacteristically large. The diagnosis step attempts to determine whether the latency anomaly occurred in the last mile or elsewhere in the network, based on whether the latency increase was observed to all destination servers or only a subset of these servers. We briefly describe each of our approaches below.

Anomaly detection. To detect latency anomalies, we perform round-trip latency measurements from each of our deployed BISMARK routers to globally distributed Measurement Lab servers, as well as to Google’s anycast DNS service. We also measure the *last-mile latency* from each of these routers, which is the round-trip latency to the first IP hop along the path, as determined using a traceroute measurement.

We measure the average latency for each day and declare a latency anomaly to have occurred if the router observes a latency increase that is more than one standard deviation greater than the mean latency for that day. After declaring an anomaly to have occurred for a particular access network and server, we then proceed to determine whether that anomaly correlates with latency anomalies observed on other paths using the diagnosis algorithm described below.

Correlation-based diagnosis. To narrow the likely location that is inducing a sudden increase in latency, we apply Algorithm 1, which infers a rough location responsible for an anomaly based on which vantage points observed the anomaly. If a single BISMARK router observes a latency anomaly to all destination servers that also correlates with a spike in last-mile latency, we declare the pathology to be a *last-mile* problem; if the router observes a latency anomaly to multiple (or all) destinations that does not correlate with an increase in last-mile latency, we classify that problem as a *middle-mile* latency problem—perhaps caused by congestion in an ISP along the path or at a peering point. If the latency anomaly does not occur to multiple servers simultaneously, we classify the anomaly as a *server-side* latency anomaly.

Latency data is collected from numerous BISMARK routers distributed around the world and connected to various ISPs. Hence any single method of selecting the pre-determined threshold to classify a latency value as a spike might not work for all the routers.

	Middle-Mile	Last-Mile	Server-Side
Atlanta	19.53±12.65	14.42±10.07	66.05±15.56
Comcast	18.64±7.43	16.20±9.40	65.63±13.22
Developed	27.70±8.41	15.73±7.15	56.57±10.40
Developing	35.92±20.85	13.03±12.15	51.05±22.14

Table 1: Categories of latency anomalies for different classes of access networks.

One simple approach is to choose the threshold for a particular router-server pair as the average latency plus the standard deviation of RTT observed to that server over the entire time period of the data being analyzed. This threshold approach is naive and has a limitation in that a few outliers can lead to a high threshold value causing latency spikes of lower magnitude to go undetected by the algorithm. We plan to experiment with multiple approaches to select the threshold value. We evaluate the efficacy of each threshold selection method against a manually obtained spike count and choose the method which performs optimally.

3. PRELIMINARY RESULTS

In our preliminary analysis, we applied the detection algorithm to latency data from September 2012, using 24 home routers in the Atlanta area and 27 home routers in Comcast across the United States. We also performed our analysis for collections of routers deployed in developing and developed countries to determine whether developing countries experience more problems in different portions of the network. Table 1 shows the average percentage of latency anomalies for each time, and their corresponding 95% confidence intervals. We observe that far fewer latency anomalies than we expected appear to be in the last mile. Interestingly, developing countries experienced relatively more middle-mile anomalies suggesting perhaps that peering along the paths from developing countries to various destinations may have significant room for improvement.

In our ongoing work, we are using Paris traceroutes [1] from the home router to the servers to determine whether correlated latency anomalies also have common elements along an end-to-end path. For example, we expect middle-mile anomalies to be reflected in paths to multiple servers that may share one or more traceroute hops in common—thus pointing to the likely cause or location of the latency anomaly. Another area for investigation is the relatively large fraction of anomalies that appear to be server-side (*i.e.*, only corresponding to a single destination server); given that the Measurement Lab servers are supposedly on well-provisioned networks, this finding is somewhat surprising, and we are examining this result in more detail in our ongoing work.

Acknowledgments

This work was funded by NSF awards CNS-1059350 and CNS-1261357 and a Google Focused Research Award.

REFERENCES

- [1] Paris-Traceroute.
<http://www.paris-traceroute.net/>.
- [2] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescape. Broadband Internet Performance: A View From the Gateway. *SIGCOMM*, Aug. 2011. <http://conferences.sigcomm.org/sigcomm/2011/papers/sigcomm/p134.pdf>.