Arcus Internet Disruption Dashboard and Report

Violet Lingenfelter [*]
Northeastern University

Robert Gove, Chae A. Clark, Anthony Wong [†]
Two Six Labs, LLC

Abstract

Internet disruptions inconvenience everyone, and being able to quickly visualize geographic and temporal disruption patterns allows stakeholders to make effective responses to disruption events, like those caused by natural disasters and cyber attacks. We present a set of visualizations for communicating internet disruption data: a dashboard for internal system developers and a report for a more general audience. The dashboard visualizes disruptions temporally and geographically at the country, city, and organizational level. We also cover an example analysis of a disruption event.

1 Introduction

Timely, detailed information on ongoing and current internet disruptions—ranging from full network outages to internet provider slowdowns—enables quick and effective response by interested parties, such as network administrators and disaster relief organizers. In this work, we visualize the output of Arcus, a network anomaly detection system, to identify and communicate internet disruption events to both technical and non-technical audiences.

Arcus aggregates raw Border Gateway Protocol (BGP) data [3, 4] to the Classless Inter-Domain Routing (CIDR) block level, and applies an isolation forest to mark anomalous CIDR blocks. CIDR blocks are geo-located using Maxmind Geolite2 database [2]. Every 10 minutes approximately 1 million unique records are collected, so aggregation and visualization is required to make this data interpretable.

In this poster we discuss users who would consume the Arcus output, the user requirements to analyze the output, and a set of work-in-progress visualizations derived from the user requirements that allow both technical and non-technical audiences to interpret the output. We created a highly interactive dashboard to assist in network disruption detection research, and a dynamically generated report to communicate identified disruption events. This draws on some prior work such as CAIDA's IODA Outage Severity Overview[1] and Down Detector[2], but was designed specifically for the data generated by Arcus.

2 User groups and requirements

There are two distinct user groups for visualizations of Arcus output data: (1) data scientists and system developers seeking to refine the algorithm, identify disruption events, and validate output data; and (2) a more broad non-technical audience, such as journalists, who want curated information on past and current disruption events. Neither audience can interpret the raw output of Arcus, which is at the CIDR level. Both audiences need to see data aggregated to either the organizational level or to some geographical level, such as country, region, or city.

Arcus data scientists and system developers have several requirements: be able to dig into the data to identify and confirm disruption events in order to verify system output and identify necessary algorithm design changes to assist in network disruption detection research; be able to identify high level disruptions from the algorithm's output, as well as the entities—such as organizations, cities, and countries—composing or affected by the higher level disruption; and be able to quickly understand trends for entities over time, drill down to smaller time slices, and search for specific entities.

The broader non-technical audience require a more curated view of the data showing only what entities are disrupted, with as much noise eliminated from the data and interface as possible.

3 User interfaces

We have distinct user groups, so we split the tool into two interfaces: an internal dashboard and a public facing report. Each interface is composed of three major components: a set of time series visualizations showing disruption over time, a map showing geographic disruption, and summary tables for all entities (Figure 1A). The two interfaces have varying levels of interactivity, and the automatically generated reports contain an additional section summarizing the disruption event.

3.1 Dashboard

Disruptions Over Time The uppermost time series chart shows a count of affected Internet Protocol (IP) addresses over a given time range. This chart has an interactive brush that users can use to drill down to smaller time slices, updating the other dashboard components to display data from the selected sub time range. Beneath the affected IP time series chart are two more time series charts: the top chart shows a count of affected cities over time, and the one below shows a count of affected organizations over time. These two charts have highlighted regions that indicate the selected sub time range.

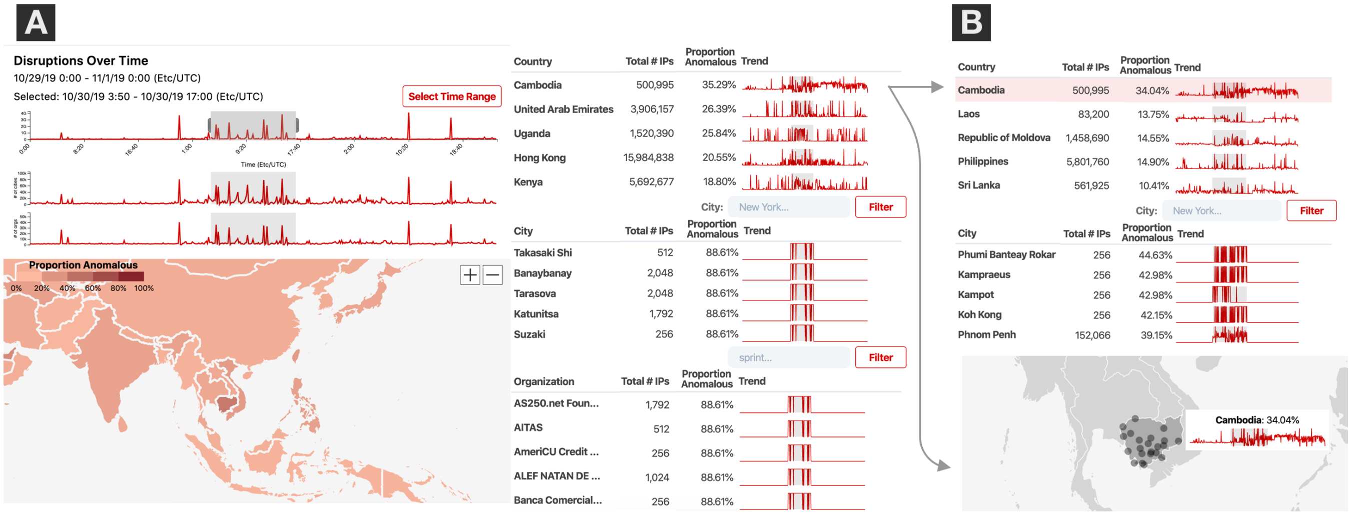
Geographic Disruptions Beneath the time series charts is a choropleth map showing the proportion of anomalous IP addresses to total number of IP addresses for countries ("proportion anomalous") for the given sub time range. The user can select a country and the top 1000 affected cities within that country are added to the map. When the country is deselected those cities are removed from the map. When a user hovers over a country polygon or a city circle, they can see its name and its average proportion anomalous over the sub time range in a tool tip. When hovering over a country, the tool-tip also includes a small trend line showing the change in the country's proportion anomalous over the overall time range (Figure 1B). This allows developers to quickly compare disruption patterns between neighboring countries to identify geographic disruption trends.

Entity Tables There are three entity tables: one for countries, cities, and organizations of adjustable length. For each row, there is the entity name, total number of IP addresses affected, proportion anomalous, and a trend line showing change in proportion anomalous for the overall time range. The table rows are ordered by proportion anomalous. The top cities and top organizations tables have search bars for filtering on entity name, allowing developers to find a specific entity that might not appear in the table.

When a user clicks on a row in the top countries table, the map zooms to that country, and the top 1000 affected cities in that country are added to the map and to the top cities table. The country table is also reordered by similarity to the selected country's trend

[*] e-mail: lingenfelter.v@northeastern.edu
[†] e-mail: {robert.gove, chae.clark, anthony.wong}@twosixlabs.com

[1] https://ioda.caida.org/ioda/dashboard
[2] https://downdetector.com/



Country	Total # IPs	Proportion Anomalous	Trend
Cambodia	500,995	35.29%	
United Arab Emirates	3,906,157	26.39%	
Uganda	1,520,390	25.84%	
Hong Kong	15,984,838	20.55%	
Kenya	5,692,677	18.80%	

City: New York... Filter

City	Total # IPs	Proportion Anomalous	Trend
Takasaki Shi	512	88.61%	
Banaybanay	2,048	88.61%	
Tarasova	2,048	88.61%	
Katunitsa	1,792	88.61%	
Suzaki	256	88.61%	

sprint... Filter

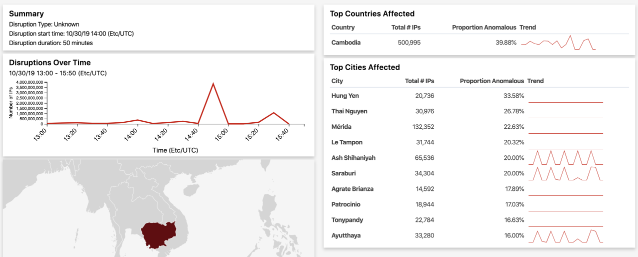
Organization	Total # IPs	Proportion Anomalous	Trend
AS250.net Foun...	1,792	88.61%	
AITAS	512	88.61%	
AmeriCU Credit ...	256	88.61%	
ALEF NATAN DE ...	1,024	88.61%	
Banca Comercial...	256	88.61%	

Country	Total # IPs	Proportion Anomalous	Trend
Cambodia	500,995	34.04%	
Laos	83,200	13.75%	
Republic of Moldova	1,458,690	14.55%	
Philippines	5,801,760	14.90%	
Sri Lanka	561,925	10.41%	

City: New York... Filter

City	Total # IPs	Proportion Anomalous	Trend
Phumi Banteay Rokar	256	44.63%	
Kampraeus	256	42.98%	
Kampot	256	42.98%	
Koh Kong	256	42.15%	
Phnom Penh	152,066	39.15%	

Figure 1: A) The Arcus dashboard showing data from October 29, 2019 to November 1, 2019. Top left: time series charts showing change in count of disrupted IP addresses, cities, and organizations. Bottom left: choropleth map showing proportion of anomalous IP addresses for countries. Right: tables showing top five affected countries, cities, and organizations. B) Top: Dashboard tables (top countries and top cities) and map after Cambodia has been selected. Bottom: Dashboard map showing affected cities in Cambodia.



Summary
Disruption Type: Unknown
Disruption start time: 10/30/19 14:00 (Etc/UTC)
Disruption duration: 50 minutes

Disruptions Over Time
10/30/19 13:00 - 15:50 (Etc/UTC)

Top Countries Affected

Country	Total # IPs	Proportion Anomalous	Trend
Cambodia	500,995	39.88%	

Top Cities Affected

City	Total # IPs	Proportion Anomalous	Trend
Hung Yen	20,736	33.58%	
Thai Nguyen	30,976	26.78%	
Mérida	132,352	22.63%	
Le Tampon	31,744	20.32%	
Ash Shihaniyah	65,536	20.00%	
Saraburi	34,304	20.00%	
Agrate Brianza	14,592	17.89%	
Patrocinio	18,944	17.03%	
Tonypandy	22,784	16.63%	
Ayutthaya	33,280	16.00%	

Figure 2: An automatically generated report a disruption event that occurred on October 30, 2019 at 2:00 p.m.

line (Figure 1B). This is useful for identifying countries that have experienced similar disruption patterns.

3.2 Automatically Generated Dynamic Reports

The report interface is a simplified version of the dashboard that allows users to quickly understand what occurred during a disruption event by cutting out unnecessary information. A disruption event report can be generated manually by developers or automatically by the system, via the dynamic template. By default each report contains a time series chart displaying the number of anomalous IP addresses and a disruption summary text. The report template dynamically adds relevant components, so reports may contain additional time series charts, tables, and/or a map depending on the disruption.

4 Example Analysis

Figure 2 shows an automatically generated report of a disruption event in Cambodia that occurred on October 30, 2019 at 2:00 p.m. (GMT). We investigated further by looking at the surrounding days with the dashboard (Figure 1A). After selecting Cambodia, we could see that other countries in South East Asia, Laos and the Philippines, had similar trend lines, indicating a potential geographic disruption (Figure 1B). We later learned that Tropical Storm Matmo, which made landfall in Cambodia on October 30, 2019 [1], coincides with this automatically detected disruption event.

5 Future Work

This tool shows promise for analyzing disruptions at the CIDR level, but it is a work in progress. Future work would address several known shortcomings. Sorting the tables by proportion anomalous means that smaller entities, which often experience higher proportional disruption when affected, are often at the top of the table. However, larger entities, such as large cities and internet service providers, could be more useful for identifying meaningful disruption patterns than small entities that may be unknown to users. Developing better sorting criteria may eliminate this problem.

Additionally, with the current UI, users cannot see geographic disruption at an intermediate level between city and country. Subregions, such as states and territories, may not exist in small countries, and data that maps IP ranges to sub-regions is difficult to work with for some countries. Future work will explore automatically clustering cities into cohesive regions that could prove useful for detecting sub-country disruptions and regional disruptions that cross country borders.

Acknowledgments

Thanks to Matthew DeLetto for system development support. This work was supported by Department of Homeland Security contract 70RSAT18CB0000018.

References

[1] Global Disaster Alert and Coordination System. *Overall Red Tropical Cyclone alert for MATMO-19/BULBUL-19 in Bangladesh, India, Myanmar, Thailand, Cambodia, Laos, Viet Nam from 29 Oct 2019 18:00 UTC to 10 Nov 2019 12:00 UTC*, 2019 (accessed Dec 6, 2019). https://www.gdacs.org/report.aspx?eventtype=TC&eventid=1000624.

[2] Maxmind, INC. *GeoLite2 Free Downloadable Databases*, 2019 (accessed Sept 4, 2019). https://dev.maxmind.com/geoip/geoip2/geolite2/.

[3] RipeNCC. *Routing Information Service (RIS)*, 2019 (accessed Dec 6, 2019). https://www.ripe.net/analyse/internet-measurements.

[4] RouteViews. *Route Views Project*, 2019 (accessed Dec 6, 2019). http://www.routeviews.org/.