Carnegie Mellon University

Mapping a Flash Crowd Incident:

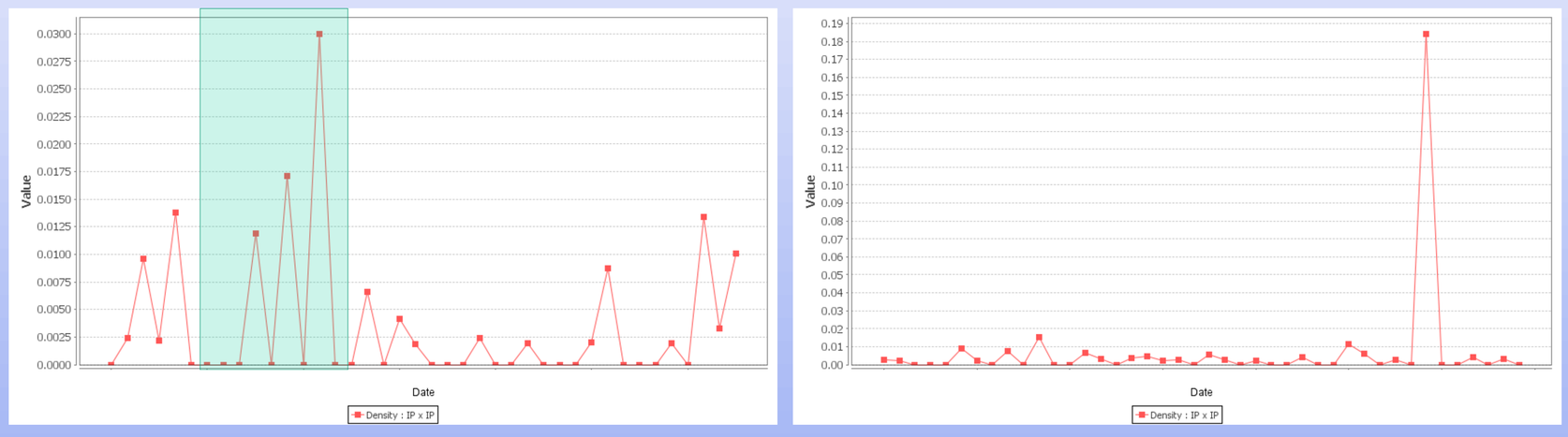
Filtering and characterizing the behavior of an enterprise network

Adam Tse
atse1@andrew.cmu.edu

Tim J. Shimeall
tjs@cert.org

Prof. Kathleen M. Carley
kathleen.carley@cs.cmu.edu

Introduction: Current methods of analysis on enterprise networks relies on either signature detection or anomaly detection. Signature detection is unscalable with zero-day threats and anomaly detection's feature set is limited mostly focusing on port usage and bandwidth. This work attempts to answer the question if behavioral features such as network science measurements can be used as features detecting anomalies. For this, a case study was conducted of a flash crowd incident that resulted in a denial of service in a few corporate machines. A flash crowd is a surge in traffic upon a machine from legitimate users that results in dramatic performance reduction or even crashing of the machine. Using a real event that is of interest to network administrators would show the value of behavioral measurements in analyzing enterprise IT networks. By understanding the behavior of networks over similar incidents, modelling techniques can be applied creating real-time systems at detecting and classifying network events.



Dataset: Data taken from large corporate network. In total, 3 weeks of data were taken; the incident week and 2 randomly picked normal weeks. The data was binned into 5 network protocols, TCP, UDP, ICMP, ESP, and GRE, because each protocol had its own set of uses that exhibit different behavior.

Figure 1: Example figure of density measurements over the incident and a normal week for TCP. Density increased significantly during the event indicating more connected communities or collaboration between machines during the event. Possible work day sprint to try to fix the problem.

In addition to the comparative study of network science measurements over events, a second focus was on Internet Background Radiation, useless traffic that can both be malicious or benign that takes up the majority of communication over many IT networks. Analysis was done comparing the behavior of key IPs exhibiting the behavior with the network as a whole. This is a precursor to creating ways of classifying IPs by their behavior so automatic blacklisting and whitelisting methods can be created.

Figure 2: Reporting on AbuseIPDB and the attributes recorded

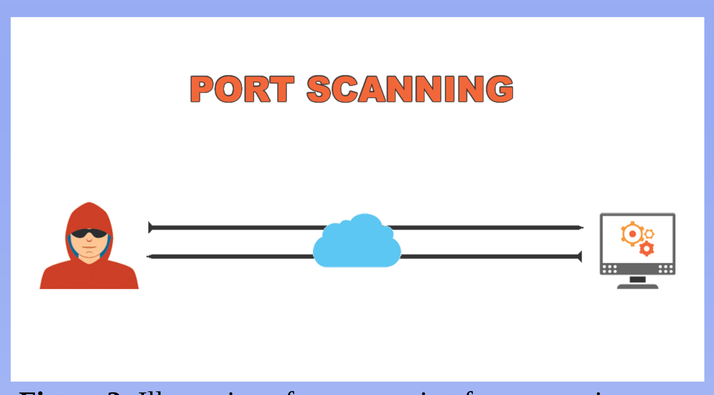


Figure 3: Illustration of port scanning for reconnaissance and vulnerability scanning

Gathering Attribute Data: Additional study was done to find out what kind of IPs were extracted from the two pruning strategies. AbuseIPDB, a database where network administrators can report IPs for suspicious behavior within their networks, was used to gather attribute information on the IPs interacting within the company's network. These attributes include ISP, organization, hostname, country, city, and number of reports. Additional study was done by creating ego-networks for the most reported IPs within the company's network.

Pruning the Dataset: Methods were created to reduce the size of the data and highlight effects that were hypothesized based off of the nature of flash crowd events. The following metrics were used to extract the data and the top 20% of each were used to create new reduced sized networks:

- Difference in total-degree centrality
- Increased Sending rate and decreased receiving rate

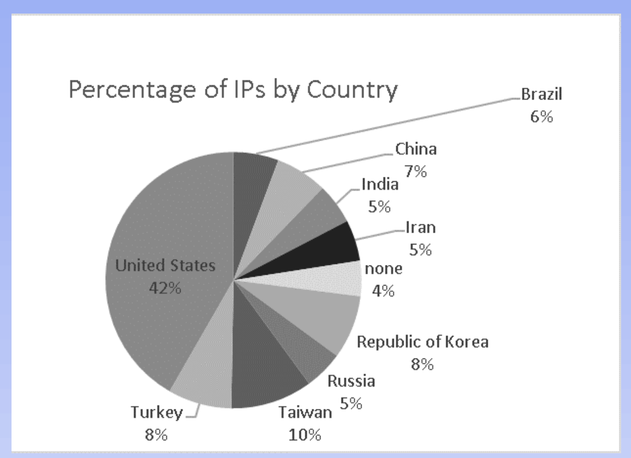


Figure 4: Distribution of TCP IPs by country

Key Findings:

- Network science measurements had less dramatic effect than originally predicted
- Difference in total-degree pruning revealed more cyclic normal behaviors
- Sending rate and decreased receiving rate revealed more background radiation IPs
- Overall best measurements that showed different behavior were fragmentation/clustering measurements
- Ego networks exhibited very erratic behavior during the event

Future Work:

- Analyze ego-networks for corporate machines, private machines, and legitimate suspicious machines (Amazon, Google)
- See how clusters change over time of the incident
- Create method for classifying machines by internet background radiation behavior
- Conduct similar analysis with ICMP

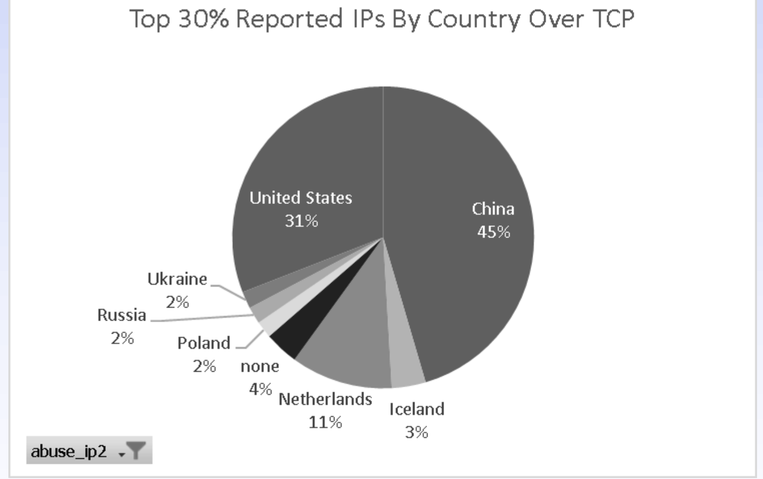


Figure 5: Distribution of Top ReportedTCP IPs by country

Sponsor and grant information

This project was funded in part by CASOS, the center for Computational Analysis of Social and Organizational Systems and in part by the Office of Naval Research - MURI N00014081186 . The views and proposal contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the U.S. government.