ORION: Observatory for Cyber-Risk Insights and Outages of Networks

Alexander Hsia and Jianbin Zhang (University of Michigan--Ann Arbor CSE), Michael G. Kallitsis (University of Michigan--Ann Arbor, Merit)

merit

LSA MICHIGAN RESEARCH AND DISCOVERY SCHOLARS UNIVERSITY OF MICHIGAN

UROP UNIVERSITY OF MICHIGAN

Abstract

Every hour, network telescope systems record gigabytes of internet traffic directed to routed but unused internet address space. They provide insight on global Internet behavior and are one of the main sources of data for network and security analysts to better understand malware propagation, Distributed Denial of Service (DDoS) attacks, network scans, internet outages, and other abnormal behavior. Merit Network, Inc., a research-and-education academic institution serving the state of Michigan, has operated one of the largest research network telescopes for just over the past decade. Merit's empirical data has enabled researchers to uncover the Mirai botnet, responsible for launching one of the largest DDoS attack in public record, and to understand its operation and device composition.

The main goal of Project ORION (Observatory for Cyber-Risk Insights and Outages of Networks) is to implement a real-time database from the information collected from Merit Network's Telescope that allows researchers to observe and analyze meaningful shifts in the internet's behavior. The immediate goal of our work is to identify and visually present verifiable high-impact events such as major power outages, malware infections and outbreaks, etc., via Merit Network's darknet telescope data.

Introduction: Merit's Network Telescope

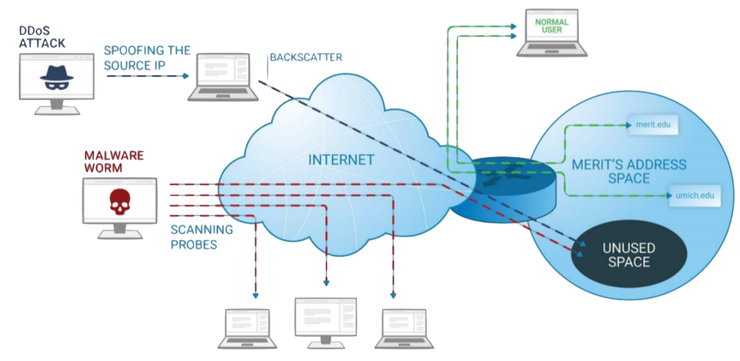

Figure 1. Diagram of how malicious traffic enters Merit's Network Telescope [1] as well as the main types of traffic that flows into the darknet.

Internet Service Providers (ISP's) such as Merit Network own and operate large chunks of IP address space. IP (internet protocol) addresses are unique sequences of integers identifying internet devices and applications. Only a fraction of Merit's available IP addresses are used. Merit runs a network telescope in the unused address space, capturing unsolicited packets. Altogether, the network telescope collects ~50 gigabytes of network traffic [1]. The majority of which comes from port scans and backscatters.

- Port scans identify open or unprotected ports on network devices. A port on a computer is an access point for the device to send and receive information. Recent malware, such as botnets capable of launching DDoS attacks, rely on port scans to find new potential hosts. DDoS attacks are when large volumes of traffic from many sources are sent to a single destination, preventing users from accessing the application and creating a "denial of service" [2]. A notable example of malware used to launch DDoS attacks is the Mirai virus, which used its botnet of IoT (Internet-of-Things) devices to attack targets.
- Backscatters are also collected in the darknet. When threat actors launch DDoS attacks, they often spoof their IP address or alter it to mask their identity. When victims receive the spoofed packets, they respond to the packet's altered IP address. Whenever the attacker spoofs their IP address to one in Merit's darknet, the victim's response reaches the network telescope.

Network telescopes are also useful for analyzing catastrophic damage to internet infrastructure. Network scanning is so ubiquitous that any power outage causes a significant drop in network traffic [3]. This is because power outages take down regional internet service providers, preventing people in the affected area from accessing the internet.

Methodology: Analysis of large-scale data

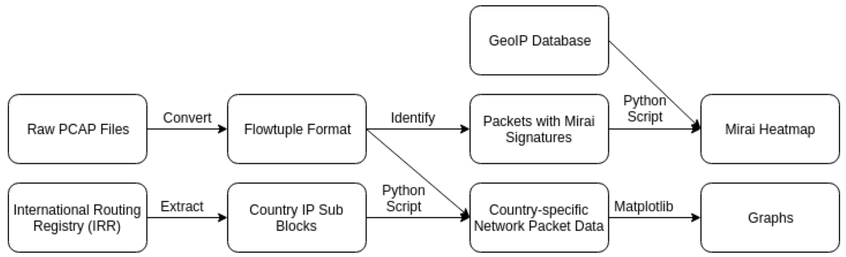

Figure 2. Darknet data visualization methodology. This flowchart shows how we created the graphs and the Mirai heatmap.

1. Captured network packets were recorded as raw PCAP (packet capture) files.
2. In order to efficiently store and process this information, the data was then converted into CAIDA's flowtuple-formatted files. The latter disregards the packets' payload, or the message the packet carried, and only includes the network packet's header, including the origin and volume of traffic, timestamp, ports number, etc [4].
3. We then parsed through the chronologically-organized data with Python scripts to look for packets from areas of interest. Each country has allocated IP sub blocks, made public by the Internet Routing Registry. To find packets from a specific country, our program isolated packets with source IP prefixes in that area's sub block.
4. We then used Matplotlib, a python package, to plot the data over the interval.
5. To create the Mirai heatmap in Figure 4, we used GeoIP, a geolocation lookup table, to generate a heatmap of the Mirai data. Once the coordinates were produced, we plotted them using Folium: a python data mapping and visualization package.

Findings: Network Outages and the Mirai Outbreak

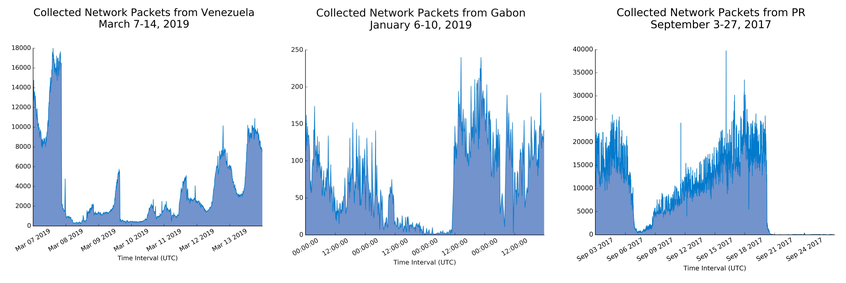

Figure 3. [L] Collected network packet count in Merit's darknet from Venezuela during March 7-13, 2019. [M] Collected network packet count in Merit's darknet from Gabon during January 6-7, 2019. [R] Collected network packet count from Puerto Rico during September 7-28, 2017. Each point is an aggregate packet count with time interval of 10 minutes.

1. The first case study we conducted aimed at Venezuela's recent power outage during March 7-14, 2019. Despite some fluctuations due to the diurnal pattern of internet activity, there is a significant drop in the packet count on the evening of March 7: the beginning of the blackout. We can also observe a gradual increase in the packet count, which is due to Venezuela's repairs on their power grid.
2. Another reason for internet blackouts is government censorship. A recent example of this is the 2019 Gabonese coup d'état. On January 7, 2019, Gabon had an attempted coup d'etat, during which rebels shut down local telecom companies. The graph shows the internet blackout, with a near-zero packet count for parts of January 7th and January 8th.
3. The last case study was aimed at Puerto Rico's power grid failure. Hurricane Irma made landfall in Puerto Rico around September 6, 2017, knocking its grid offline. Two weeks, Hurricane Maria soon followed on September 20th, 2019, taking out power for the whole territory. The outage can be seen, with two distinct intervals with near-zero packet counts.
4. We also analyzed Mirai botnet activity, responsible for the infamous Dyn DDoS Attack on October 12th, 2016. The massive DDoS attack left much of the U.S.' east coast without internet. Merit was one of the leading observers of the attack and thus has the historical data of the attack. We processed this data into a heatmap, "uncovering" the hidden Mirai botnet population and displaying its concentrations.

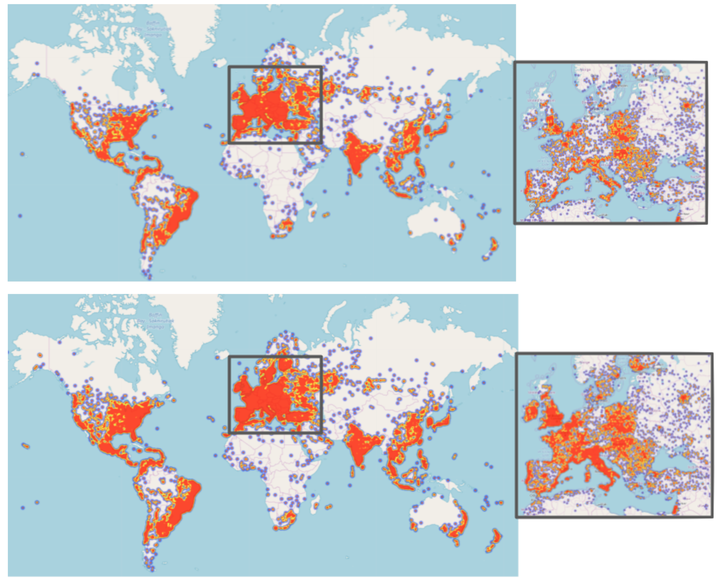

Figure 4. [T] Heatmap of active, scanning infected Mirai devices in the first week of September 2016 (9/1/2016-9/7/2016), the beginning of the Mirai infections. [B] Heatmap of infected Mirai devices in the last week of November: the time in which the number of port scans detected in Merit's network telescope peaked (see Figure 5). Europe is enlarged to display the large increase in infected devices.

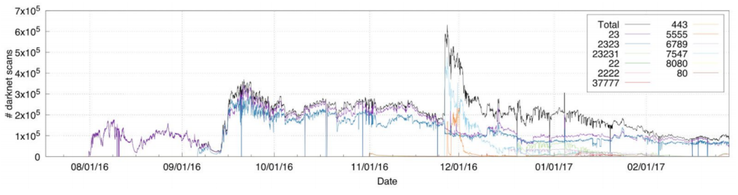

Figure 5. Collected network packets in the darknet with the Mirai signature. The sharp increase in port scans is due to the Mirai worm [2].

From these cases, we can conclude that our program can accurately log and process internet traffic data to reflect major events. Our findings suggest Merit's darknet provides a valuable window into the internet's behavior.

Future Work

Since this is the first phase of Project ORION, the main focus of our work is to test and verify that it is possible to process the internet traffic data and yield verifiable results. Our work confirmed that network telescope data provides accurate information of large-scale power outages and of DDoS Attacks. This conclusion allows us to move further with Project ORION. The ultimate goals of which are twofold:

1. To produce a data pipeline that can predict threats to the internet, such as DDoS attacks and power outages.
2. To create a real-time database for researchers to use.

References

[1] Michael Kallitsis, Mark Weiman, *Sharing Network Telescope Data for Education*, 2018.
[2] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou, *Understanding the mirai botnet*, 26th USENIX Security Symposium (USENIX Security 17) (Vancouver, BC), USENIX Association, 2017, pp. 1093–1110.
[4] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. Internet background radiation revisited. In Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, IMC '10, pages 62–74, New York, NY, USA, 2010. ACM.
[4] CAIDA Corsaro Flowtuple Documentation: http://www.caida.org/tools/measurement/corsaro/docs/formats.html#formats_conventions