A Scalable Platform for Enabling the Forensic Investigation of Exploited IoT Devices and Their Generated Unsolicited Activities

Sadegh Torabi[1], Elias Bou-Harb[2], Chadi Assi[1], and Mourad Debbabi[1]

1. *Security Research Centre, Concordia University, Montreal, Canada*
2. *The Cyber Center for Security and Analytics, University of Texas at San Antonio, San Antonio, U.S.*

Presented by:

SADEGH TORABI

UNIVERSITÉ Concordia UNIVERSITY

GINA CODY SCHOOL OF ENGINEERING AND COMPUTER SCIENCE

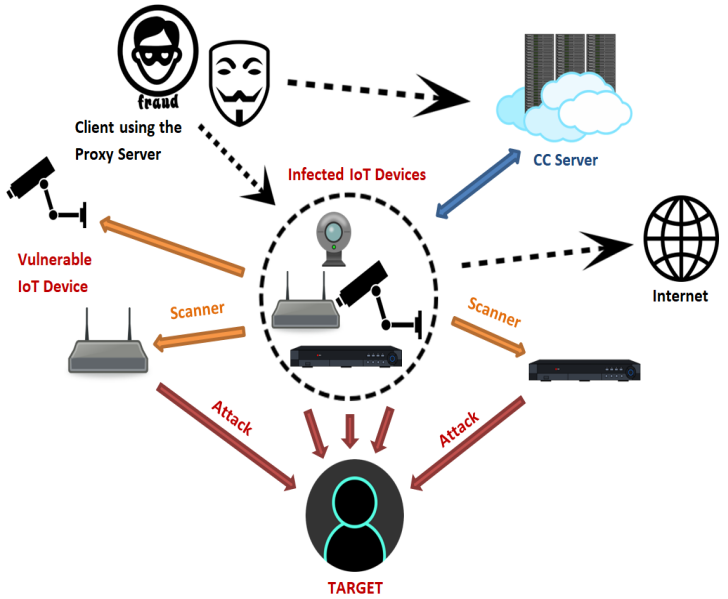
Security Research Centre

UTSA

The Cyber Center for Security & Analytics

Introduction

- Internet of Things (IoT) device are widely used in our daily activities
  - Facilitate data collection, monitoring, and information sharing
- Despite their benefits, IoT devices are used as effective attack enablers
- The rise of IoT-driven cyber attacks was marked by the Mirai botnet [1-2]
  - Propagates by exploiting vulnerable IoT devices (e.g., weak/default credentials)
  - Utilizes compromised IoT devices to perform Internet-scale attacks (e.g., DDoS)
- To mitigate such attacks, we need to possess an Internet-scale perspectives of compromised IoT devices and their activities (Challenging)
  - Lack of empirical data on deployed IoT devices
  - Lack of knowledge about their unsolicited behaviors
- Leverage passive network measurements as an alternative approach for inferring and characterizing IoT threats

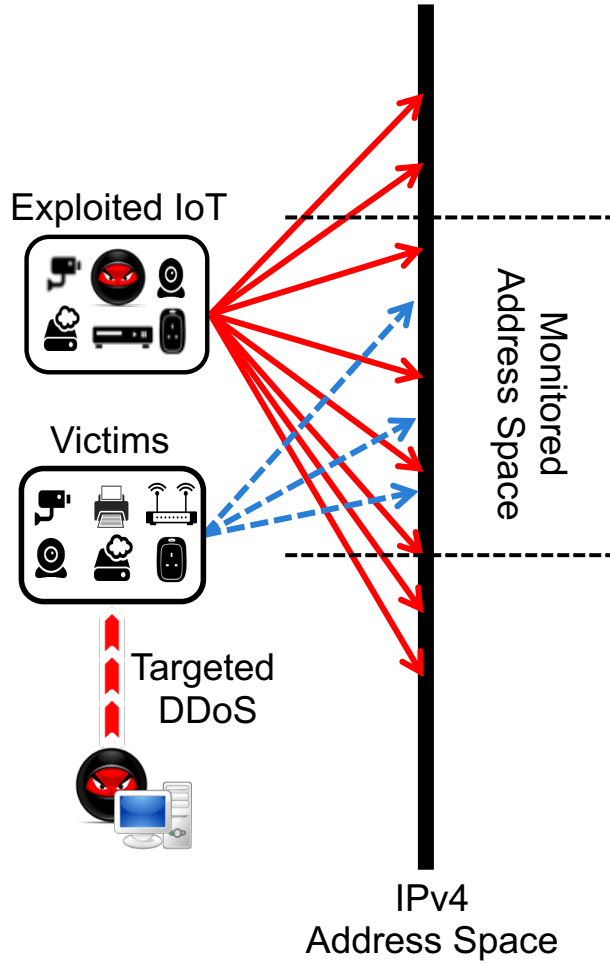


The architecture of the Mirai botnet [2]

[1] Antonakakis, M., et al., 2017. Understanding the Mirai botnet. In: 26th USENIX Security Symp. Vancouver, BC, pp. 1093--1110.
[2] https://www.fortinet.com/blog/threat-research/omg--mirai-based-bot-turns-iot-devices-into-proxy-servers.html

Background

- Data-driven methodologies for detecting compromised IoT devices [1]
  - Correlating IoT device information and passive network measurements
- IoT device information through active scanning and banner analysis (e.g., Shodan [2])
- Passive network measurements (network telescope or darknet):
  - Traffic captured at unused, yet routable IP addresses
  - Mainly Internet scanning and backscatter traffic (a byproduct of targeted DDoS attacks with spoofed IP addresses)
  - E.g., CAIDA's darknet (one of the largest existing resources with 16.7M IPs) [3]



[1] Torabi, S., Bou-Harb, E., Assi, C., Galluscio, M., Boukhtouta, A., Debbabi, M., June 2018. Inferring, characterizing, and investigating internet-scale malicious IoT device activities: a network telescope perspective. In: Proc. Of the 48th Annual IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN), pp. 562--573.
[2] https://www.shodan.io/
[3] The CAIDA UCSD Real-Time Network Telescope Data. UCSD - Center for Applied Internet Data Analysis. http://www.caida.org/data/passive/telescope-near-real-time_dataset.xml.

IoT (In)Security

Motivated by:

- Insecurity of IoT devices at scale [1]
- Rising number of IoT-tailored malware as a major threat [2-3]

Problem:

- Address the lack of scalable cyber-threat intelligence reporting and analysis capabilities that can trigger informed decisions for in-depth forensic investigations

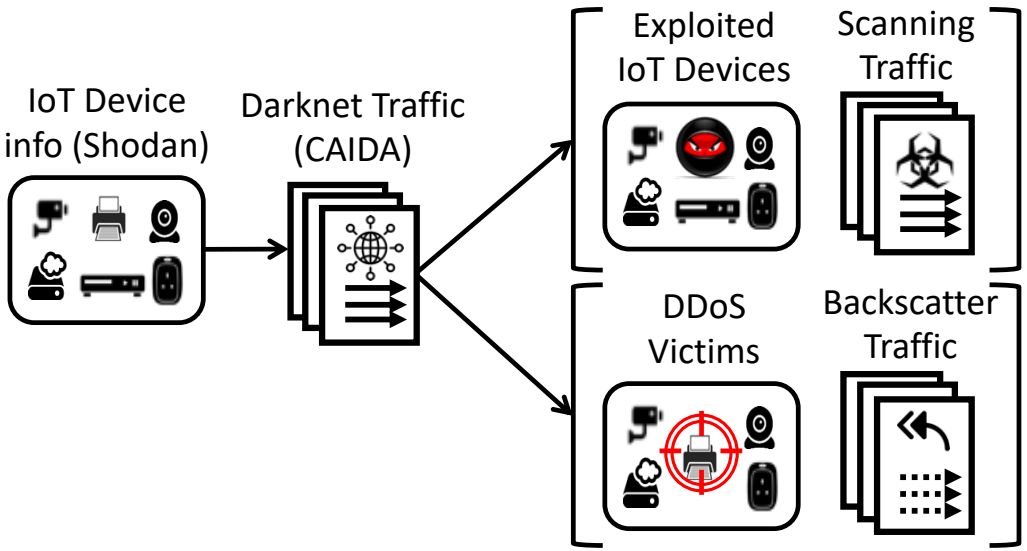
Approach:

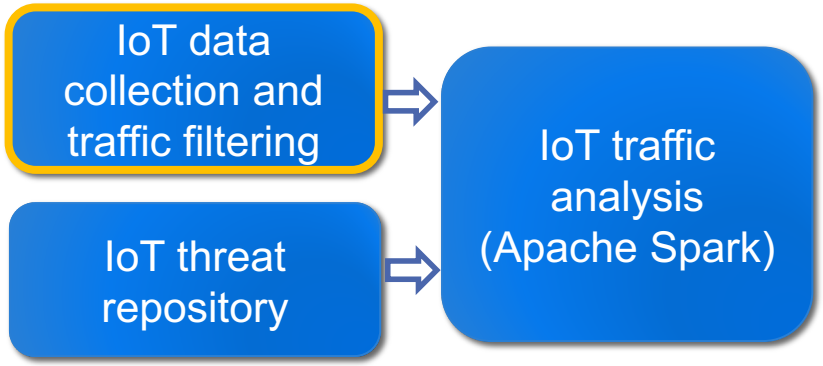
- Leverage data-driven methodologies, passive network measurements, and IoT device information
- Develop a system prototype using a big data analytics framework (Apache Spark [4]) to enable scalable and timely IoT threat detection and analysis

[1] https://www.helpnetsecurity.com/2020/02/26/shadow-iot-enterprise/
[2] Reports by Checkpoint Security Inc. (March, 2020) https://www.globenewswire.com/news-release/2020/03/11/1998560/0/en/February-2020-s-Most-Wanted-Malware-Increase-in-Exploits-Spreading-the-Mirai-Botnet-to-IoT-Devices.html
[3] https://www.comparitech.com/antivirus/malware-statistics-facts/
[4] https://spark.apache.org/

System Architecture and Components

IoT device information collection and traffic filtering
- Collect IoT device information from Shodan [1]
- Filter IoT-generated traffic on the darknet [2]
- IoT-generated traffic is processed as flowtuples

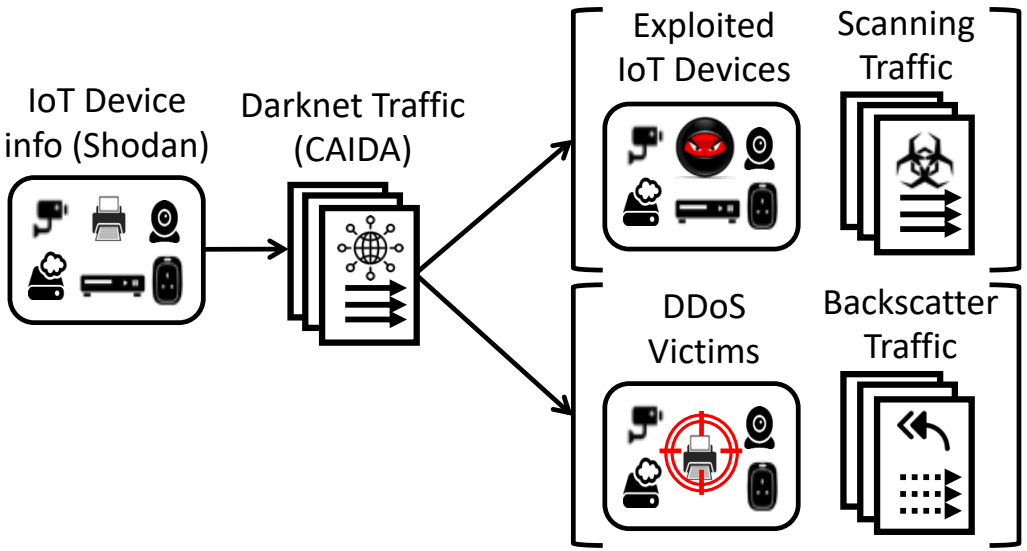


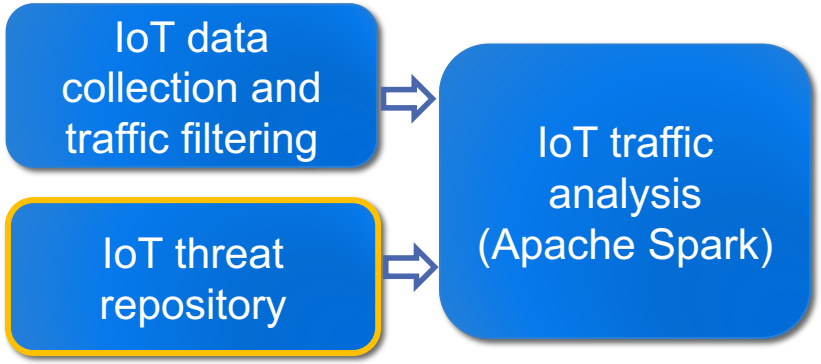


[1] https://www.shodan.io/
[2] The CAIDA UCSD Real-Time Network Telescope Data. UCSD - Center for Applied Internet Data Analysis. http://www.caida.org/data/ passive/telescope-near-real-time_dataset.xml.

System Architecture and Components

IoT device information collection and traffic filtering
- Collect IoT device information from Shodan [1]
- Filter IoT-generated traffic on the darknet [2]
- IoT-generated traffic is processed as flowtuples





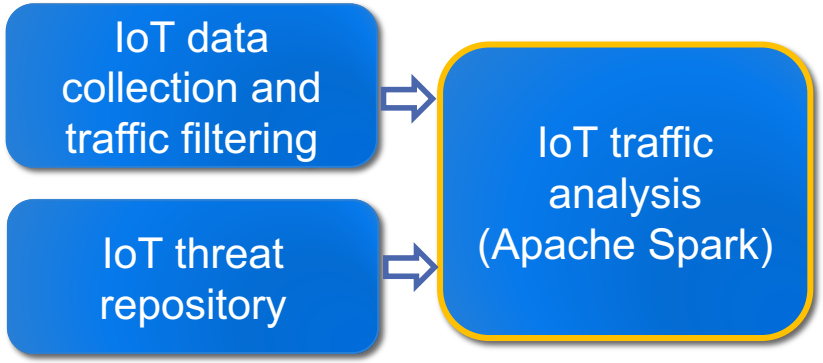
IoT threat repository (ongoing work)
- Collect IoT malware binaries/executables
- Dynamic malware analysis and attribution

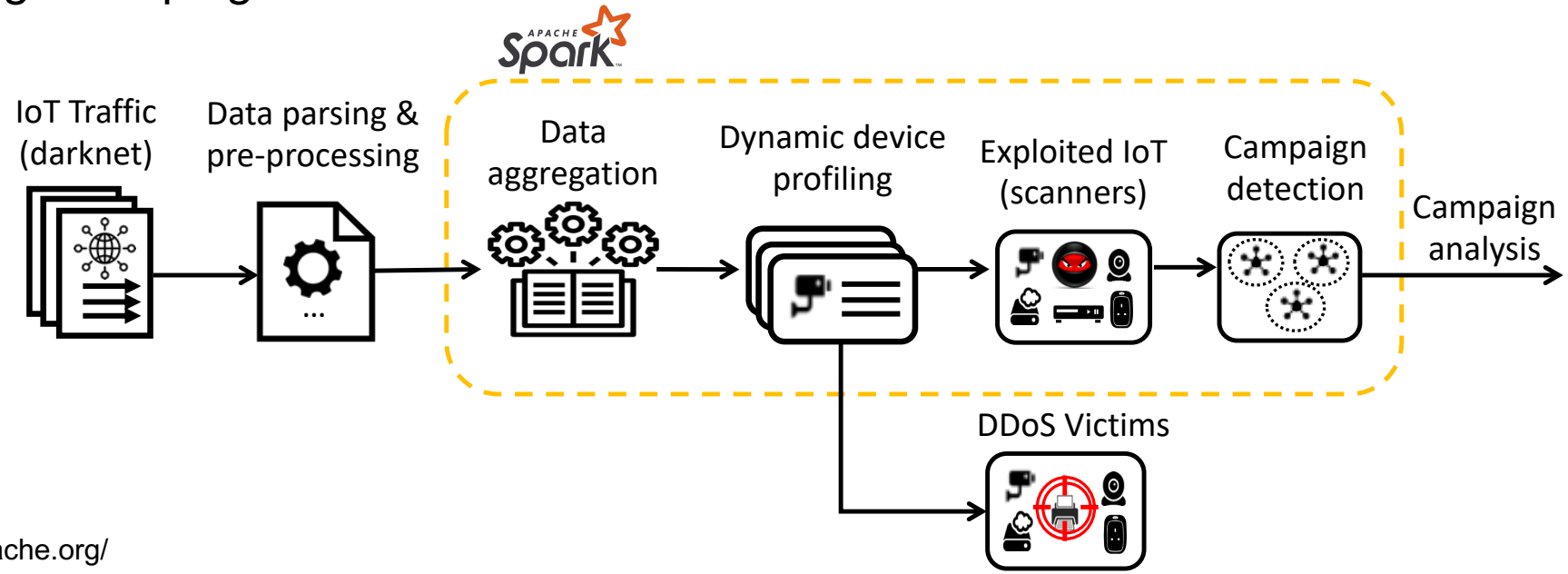
[1] https://www.shodan.io/
[2] The CAIDA UCSD Real-Time Network Telescope Data. UCSD - Center for Applied Internet Data Analysis. http://www.caida.org/data/ passive/telescope-near-real-time_dataset.xml.

System Architecture and Components

IoT traffic analysis (main component)

- Deployed in Apache Spark [1] to support fast and scalable operations
- Data parsing and pre-processing
- Data aggregation (over different time intervals)
- Dynamic device profiling with aggregate flow features
- Multi-stage campaign detection and attribution





[1] https://spark.apache.org/

Experimental Results

- Collected/Processed data

IoT device info	~400K devices (Shodan)	Consumer IoT devices (routers, IP cameras, WAP, etc.)
IoT traffic	4TB of darknet data over 5 days	308M packets (flowtuples), mainly TCPSYN (87%)
Compromised IoT	27,849 devices	~300M scanning packets (97% of all traffic)

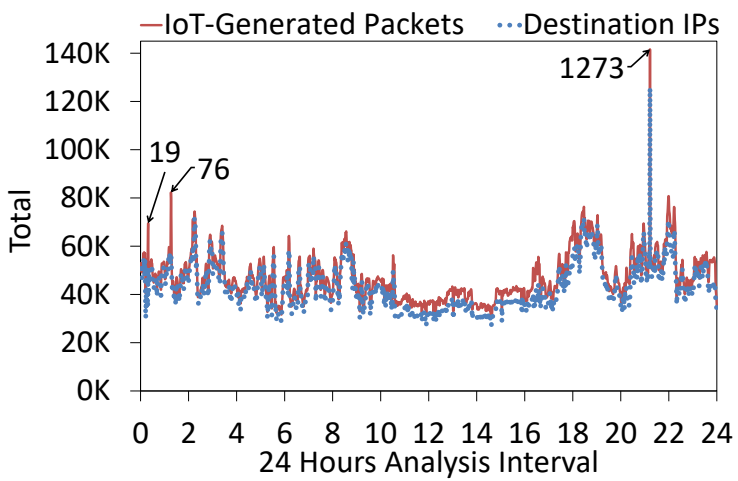
- Experimental setup
  - Deployed Apache Spark using PySpark in a standalone mode on a single node
  - Debian Operation System (Ubuntu 18.04 version), 8 CPU cores (Intel® Xeon(R) CPU E3-1240 v5 @ 3.50GHz), 64GB memory

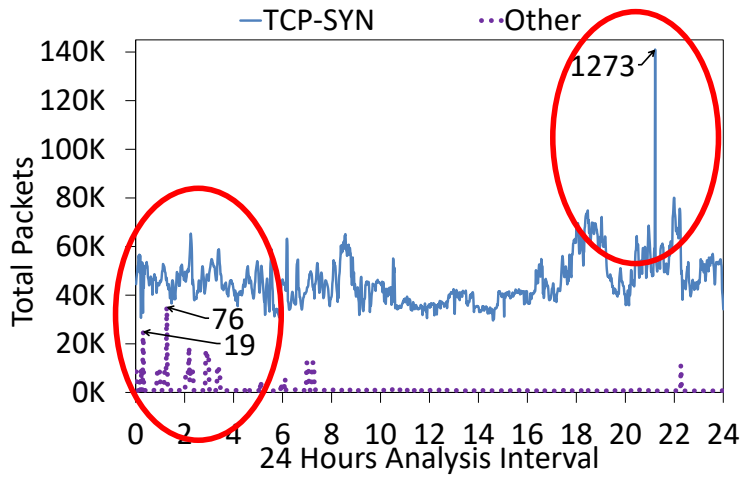
- Present examples of the network forensic capabilities and applications

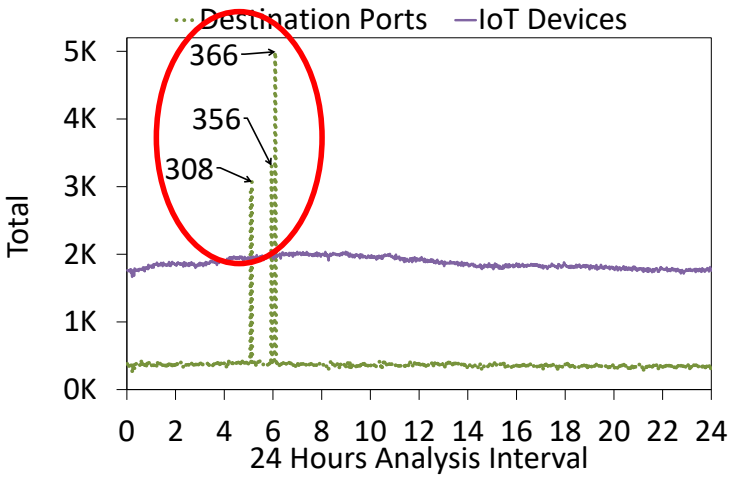
Monitoring Unsolicited Activities

High level macroscopic views in terms of IoT-generated flows, targeted destination IP addresses, distribution of the packets, targeted destination ports, and total IoT devices

- Overall trends and correlation between the number of generated packets and the targeted IP addresses (reflect Internet scanning activities)
- Highlight increased activities in certain periods (intense scanning campaigns and/or DDoS activities)
- Detecting port scanning activities (e.g., minutes 308, 356, and 366)







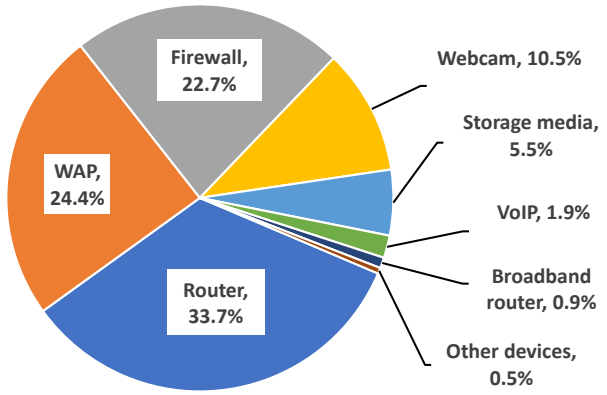
Detecting Compromised IoT Devices

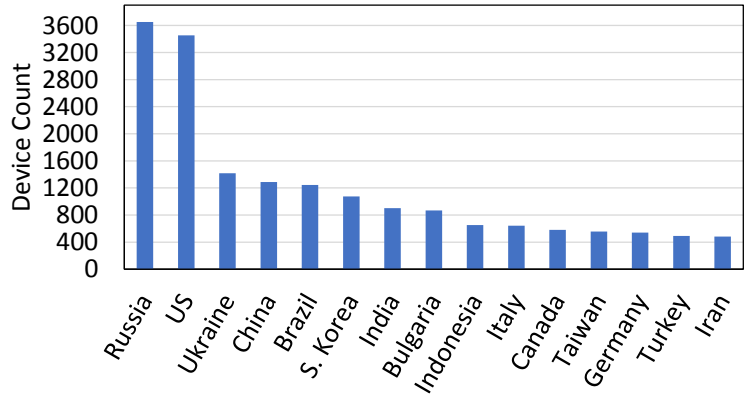
Detected about 27K compromised IoT devices that were sending scanning packets (TCP-SYN, UDP, and ICMP-REQ)

- In-depth analysis of the involved IoT devices
- Distribution of scanning packets and compromised devices per protocol
- Intensity of TCP-SYN scans (fewer devices producing significantly larger traffic)
- Distribution of compromised devices per type and hosting countries (may indicate malware outbreak)

Compromised IoT devices and their generated scanning traffic type(s).

Scanning Traffic	Devices		Packets	
	Count	(%)	Count (M)	(%)
UDP	**14,314**	**51.40**	33.21	10.32
TCP-SYN	3,770	13.54	**167.88**	**52.19**
ICMP-REQ	23	0.08	0.71	0.22
TCP-SYN/UDP	9,728	34.93	118.38	36.80
UDP/ICMP-REQ	40	0.14	1.83	0.57
TCP-SYN/ICMP-REQ	36	0.13	0.97	0.30
All types	31	0.11	1.05	0.32





Inferring and Monitoring Scanning Campaigns

Identify scanning campaigns by analyzing common scanning objectives (targeted ports)

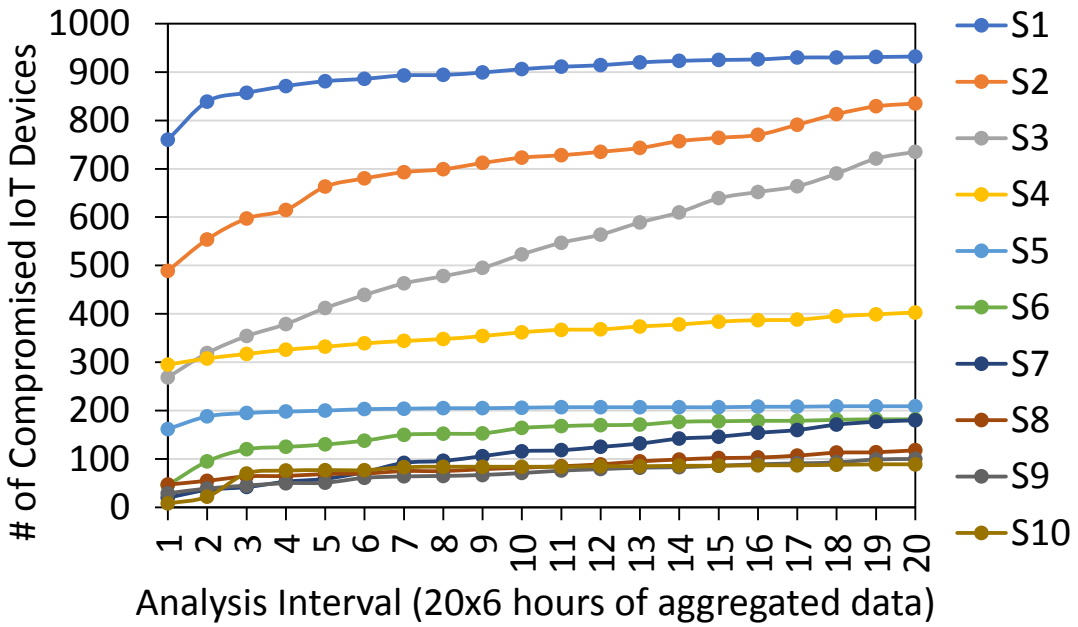
- The majority of IoT devices scanned a very small list of known ports (e.g., Telnet and HTTP)
- These port sets are associated with known IoT malware (e.g., Mirai)
- UDP/TCP ports comparison in terms of involved IoT devices and the generated scanning traffic
- Presence of targeted ports associated with emerging IoT malware (e.g., port 5555/ADB.Miner)

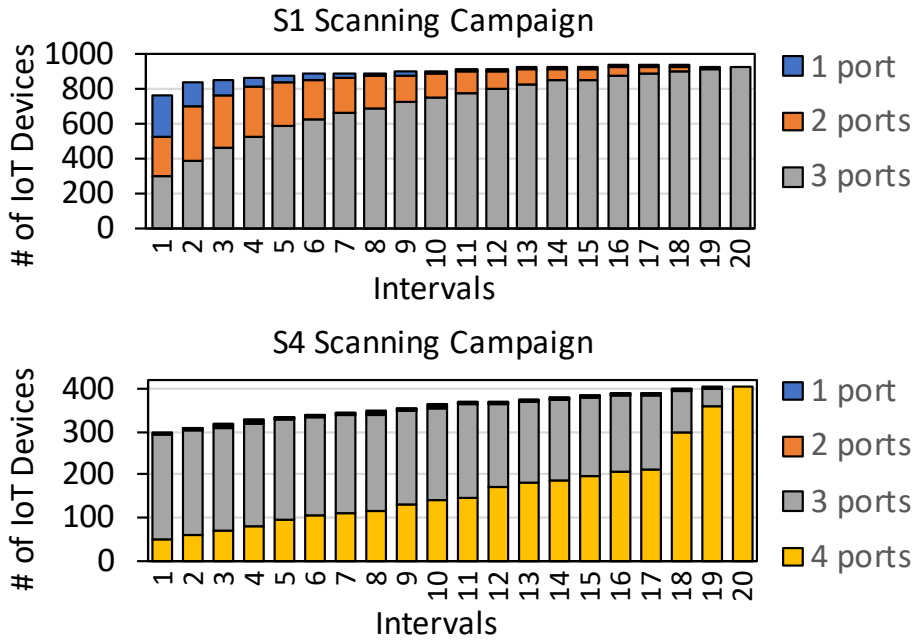
Top 10 identified scanning objectives (S').

	S_i	TCP/UDP Ports	Devices (%)	Packets (M)
UDP ports	1	28183, 32124, 37547	**932 (6.33)**	0.300
	2	445	835 (5.67)	7.687
	3	23, 80, 8080	735 (4.99)	11.200
	4	23, 80, 8080, 37547	403 (2.74)	15.809
	5	28183, 32124	209 (1.42)	0.007
	6	37547	182 (1.24)	0.015
TCP ports	7	23, 2323	180 (1.22)	**16.849**
	8	80, 8080	118 (0.80)	1.122
	9	80	100 (0.68)	1.607
	10	80, 443, 8080	89 (0.60)	0.019

Temporal Analysis and Campaign Evolution

- Granular overview of campaign evolution in terms of involved IoT devices and targeted ports
- Campaign dynamics (scanning rate, saturation, involved device types, etc.)
- Infer intensive malware propagation campaigns (e.g., S3 ports 23/80/8080)



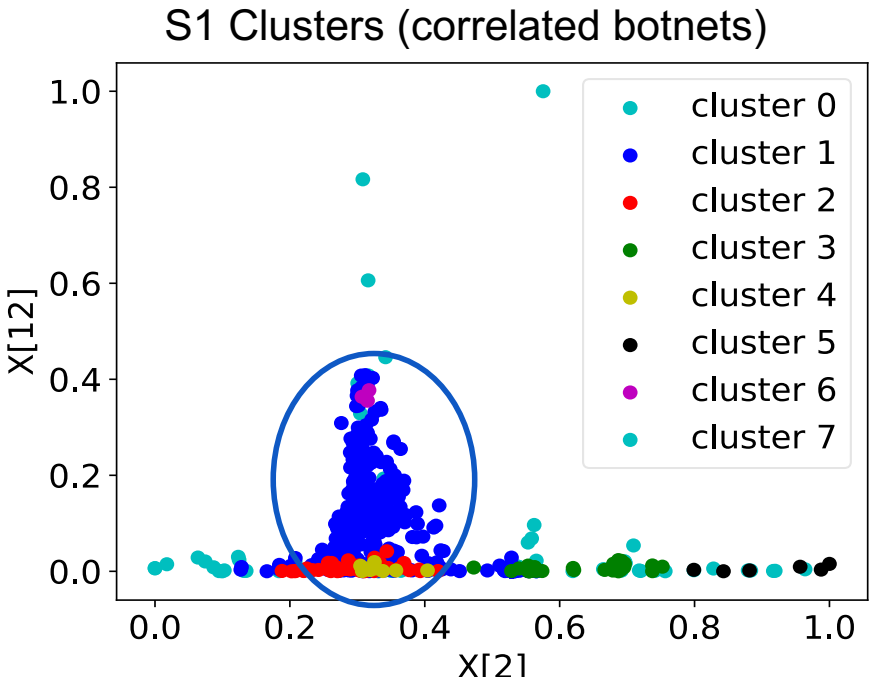


Inferring IoT Botnets

Identify correlated devices (possible botnets) within scanning campaigns

- Clustering analysis (DBSCAN) using 16 raw/aggregate flow features
- Detecting botnets of correlated devices with similar behavioral characteristics/features (e.g., 7 clusters within S1 scanning campaign)
- Analysis of devices within botnets may indicate targeted or vulnerable device types/models)

f_i	**Selected Features**
1–3	$U_{i,m}$: number of scanning packets from each type (m)
4	$S_P = \sum_m U_{i,m}$: combined scanning packets
5–7	$\alpha_{i,m}$: discrete prob. dist. representing the fraction of each scanning packet to scans
8	N': number of active intervals (minutes)
9	$A_R = \frac{b_i - a_i}{N'_i}$: activity rate
10	$S_R = \frac{S_P}{N'_i}$: scan rate
11	$\overline{TTL}$: average TTL value
12	$\overline{P}_{size}$: average packet size
13	Src_{Ports}: number of source ports
14	Dst_{IPs}: number of destination IP addresses
15	$Dst_R = \frac{S_P}{Dst_{IPs}}$: per destination packet rate
16	Dst_{Ports}: number of scanned destination ports

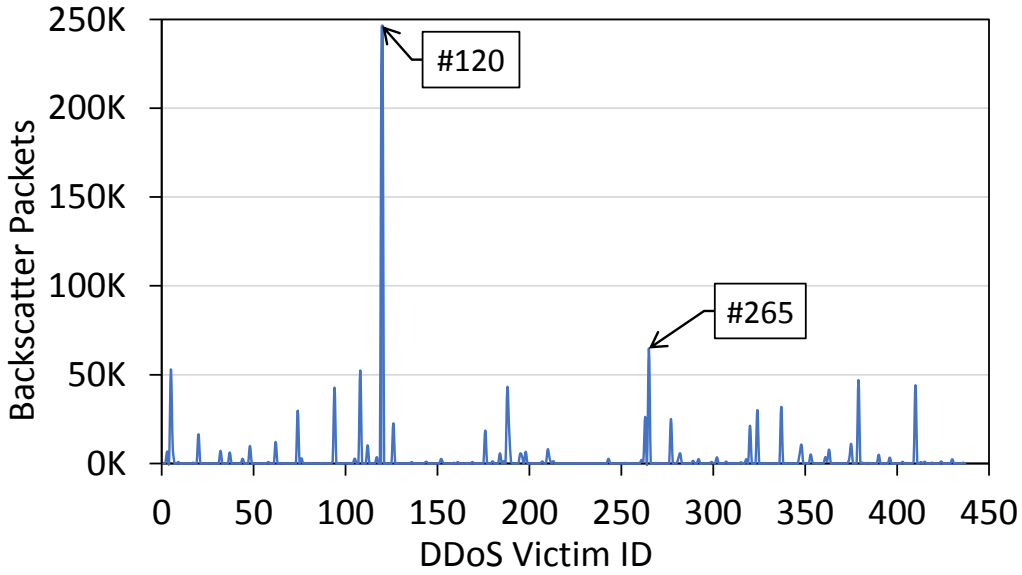


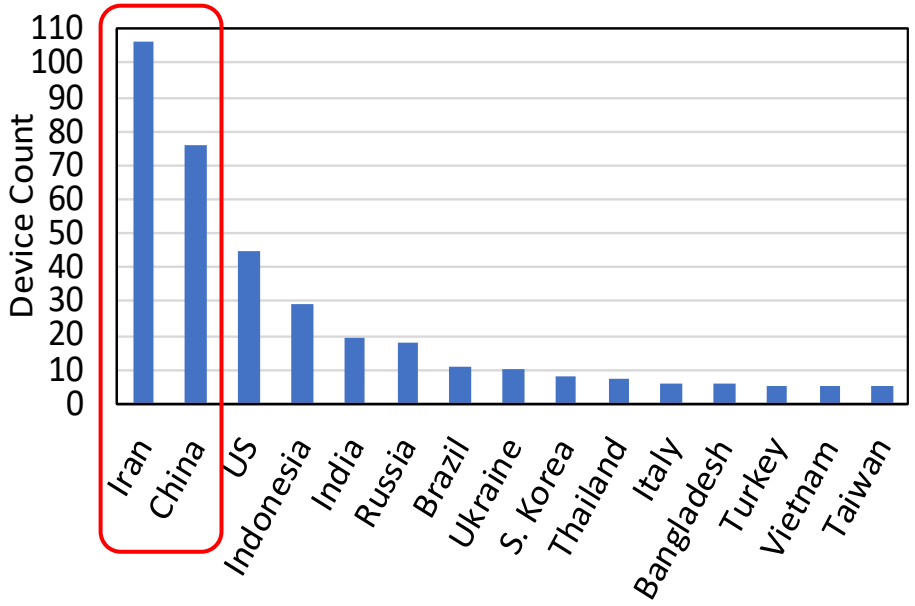
Clusters	Members
0 (outliers)	60
1	753
2	45
3	53
4	9
5	6
6	3
7	3

Identifying DDoS Victims

IoT devices that are targeted by DDoS attacks using spoofed IP addresses, which happen to be within the darknet, generate backscatter replies towards the darknet

- Targeted DDoS attacks (e.g., device #120/Radware firewall located in China and #265/MikroTik router from Iran)
- Hosting countries with the most targeted DDoS victims
- Indication of targeted attacks towards certain device models and/or countries



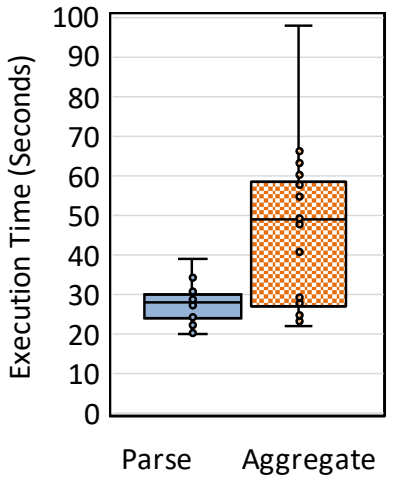


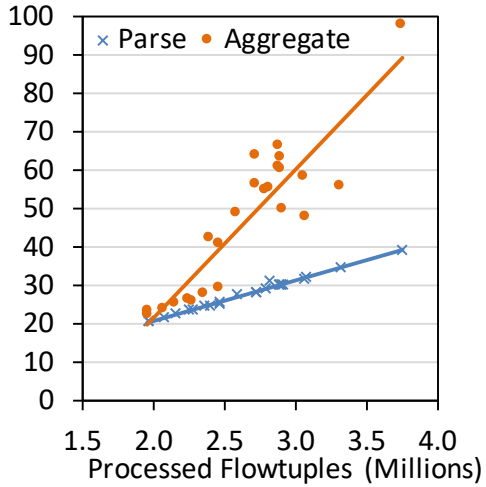
Performance Evaluation: Execution Times

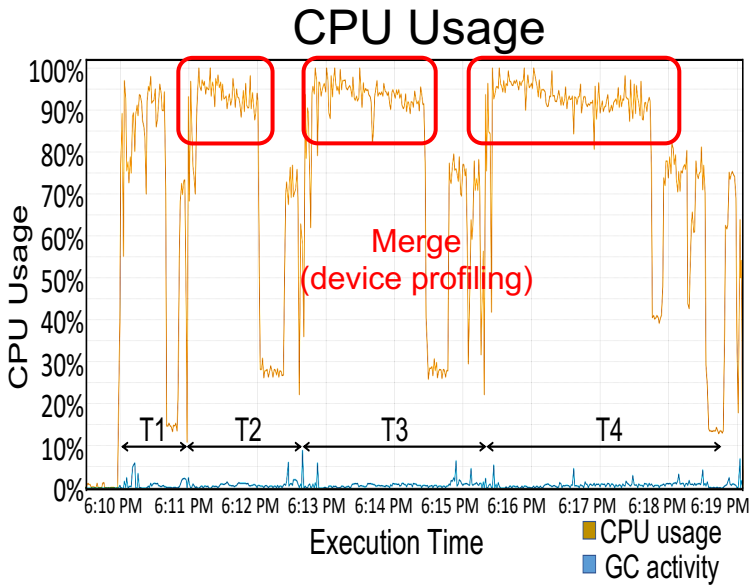
Evaluation:
- 24 hours data sample (~64M packets)
- Hourly data aggregation/merging

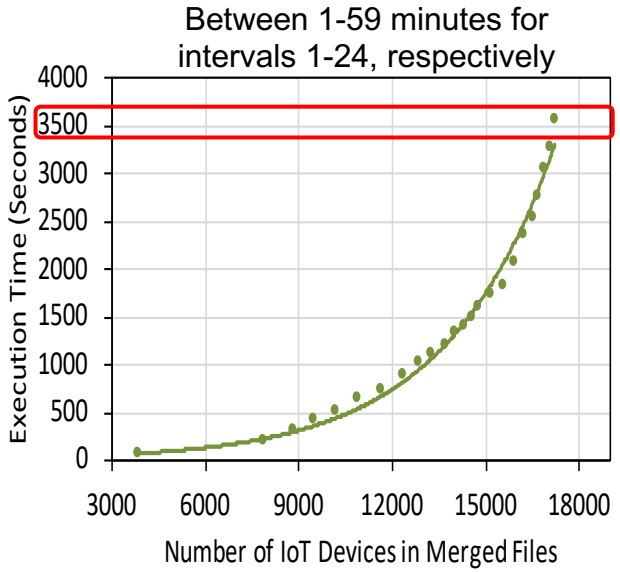
Parse/Aggregate:
- Relatively short time (mean<50s)
- Linear correlation between execution times and the processed flows (<2 minutes for processing 3.8M flowtuples)

Device profiling (merge):
- Requires the longest time (exponential increase with cumulative number of devices)
- Less than 59 minutes to perform aggregation and device profiling for a full day (~17K Devices)
- Can be reduced with a multi-cluster implementation



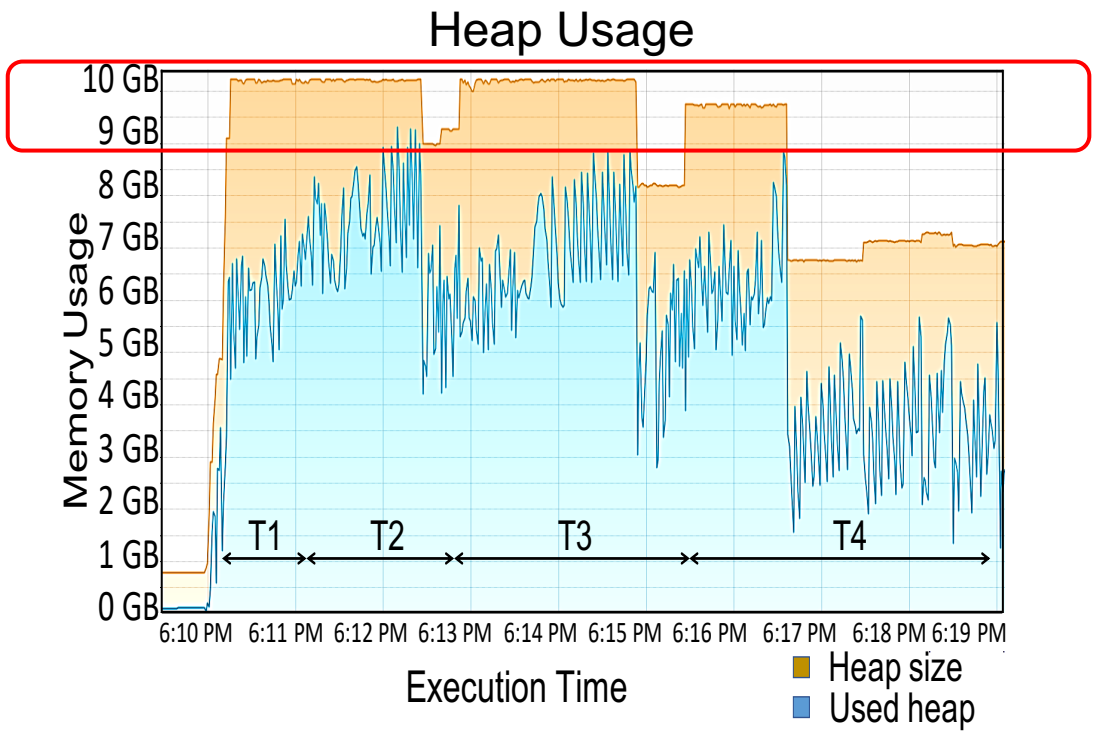


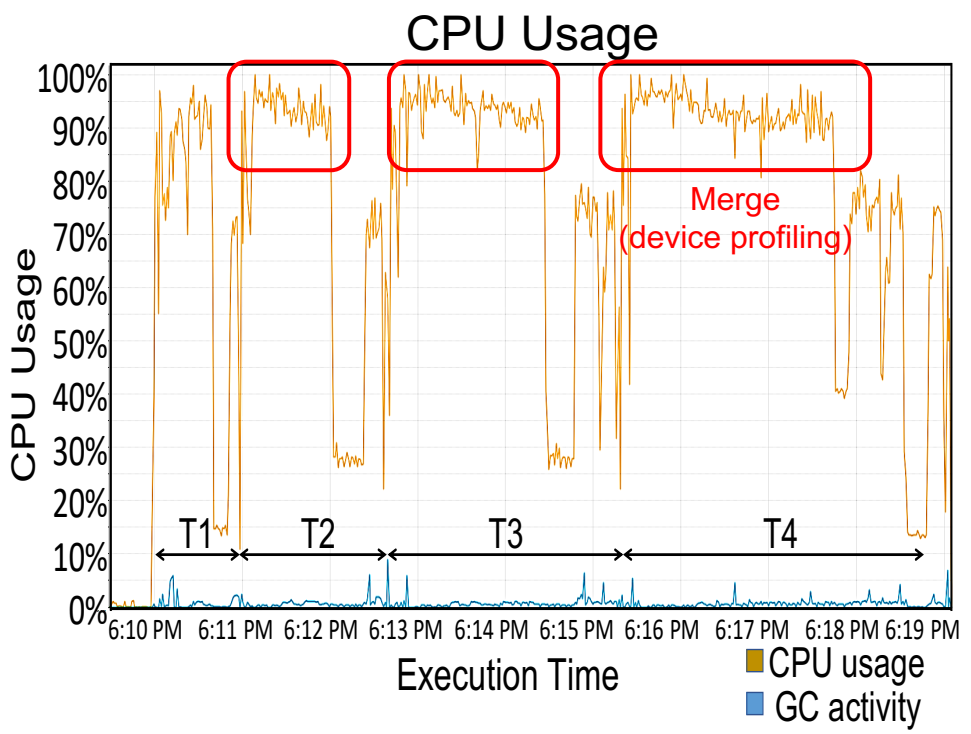




Memory/CPU Usage

Reasonable Memory/CPU usage
- Scalable operations with less than 10 GB of required memory
- Experience extended periods of CPU intensive operations with cumulative IoT devices/traffic, which can be reduced through a multi-cluster implementation





Memory/CPU usage during the first four intervals T1-T4 (hours)

Main Takeaways

- Proposed and evaluated an effective and scalable system prototype for IoT-centric cyber forensic investigations by leveraging
  - Big data analytics frameworks such as Apache Spark
  - Data-driven methodologies using passive network traffic and IoT device information
- Addressed main operational challenges such as process automation, scalability, and fast operations
- Demonstrated the capabilities of the system as an infrastructure for enabling cyber-forensic investigations
- Leveraged empirical data to examine the effectiveness of the system and evaluate its performance with traffic generated by compromised IoT devices in the wild

Thank you

For further information, contact the corresponding author (Sadegh Torabi) at:

sa_tora@encs.concordia.ca