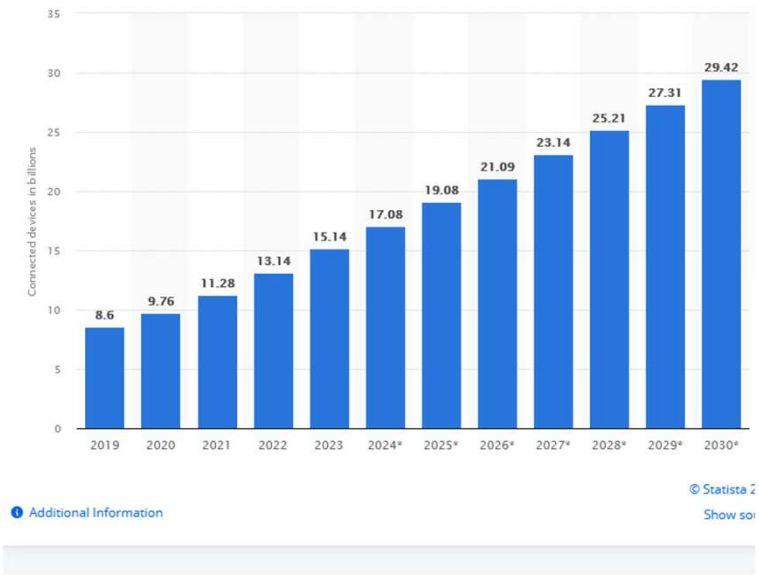
ProvIoT: Detecting Stealthy Attacks in IoT through Federated Edge-Cloud Security

Kunal Mukherjee, Joshua Wiedemeier, Qi Wang, Junpei Kamimura, John Junghwan Rhee, James Wei, Zhichun Li, Xiao Yu, Lu-An Tang, Jiaping Gui, and Kangkook Jee

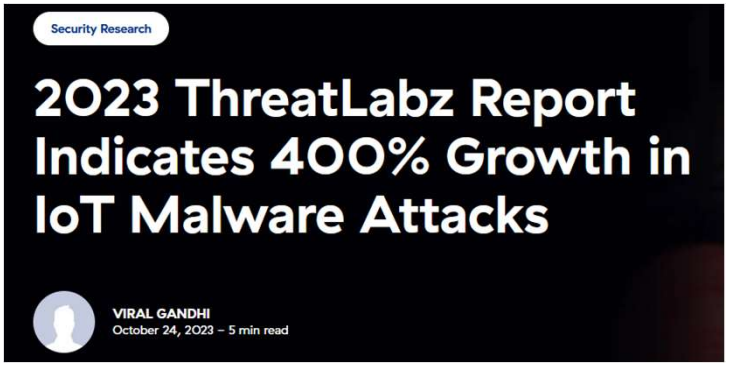
22nd International Conference on Applied Cryptography and Network Security 2024

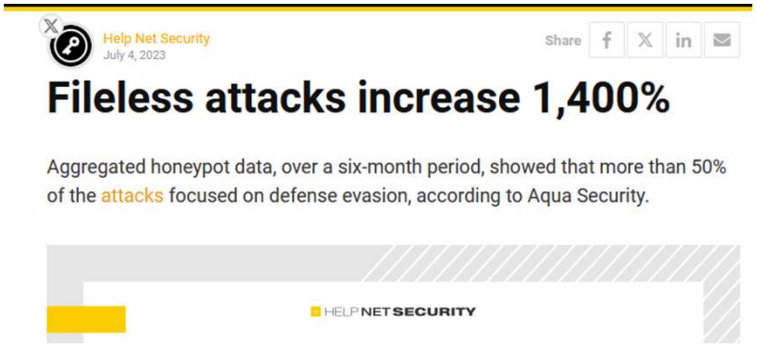
Motivation



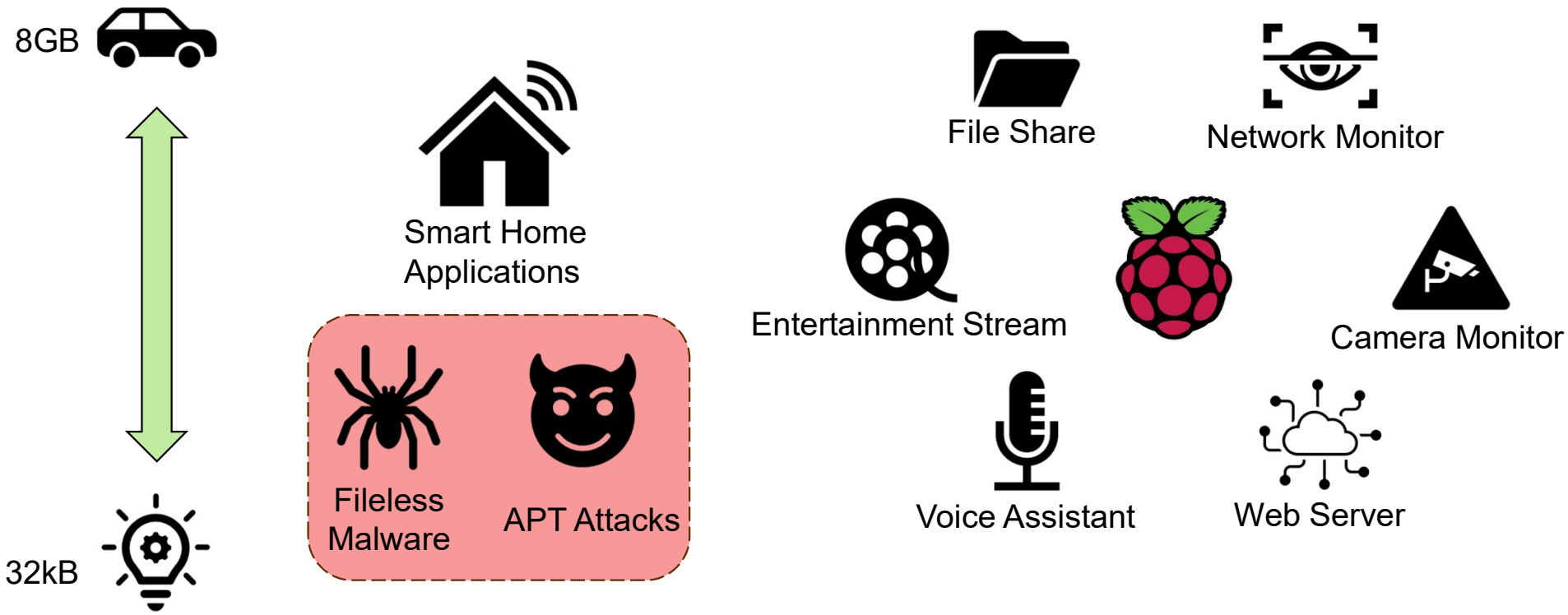
Number of IoT connected devices worldwide 2019-2023, with forecasts to 2030

Published by Lionel Sujay Vailshery, Jul 27, 2023





Scope


8GB
32kB
Smart Home Applications
Fileless Malware
APT Attacks
File Share
Network Monitor
Entertainment Stream
Camera Monitor
Voice Assistant
Web Server

Fileless Malware

Memory of Running Process 

 Monitored File System

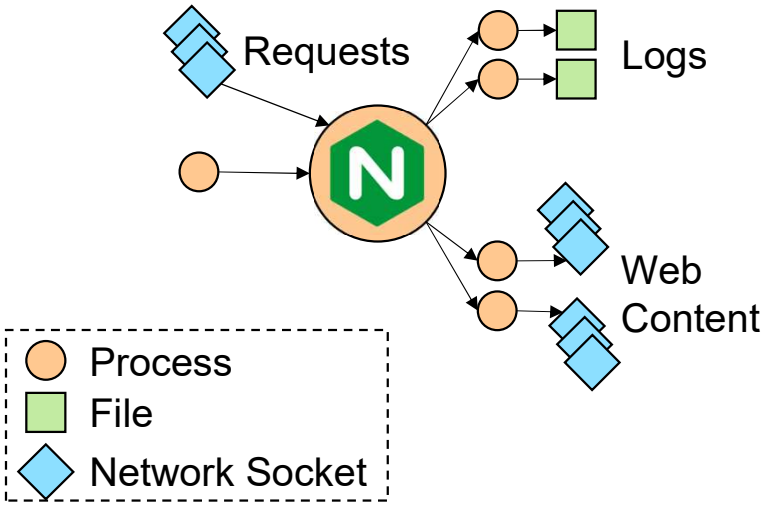
Fileless attacks are **invisible** to static IDS [Dang '19]

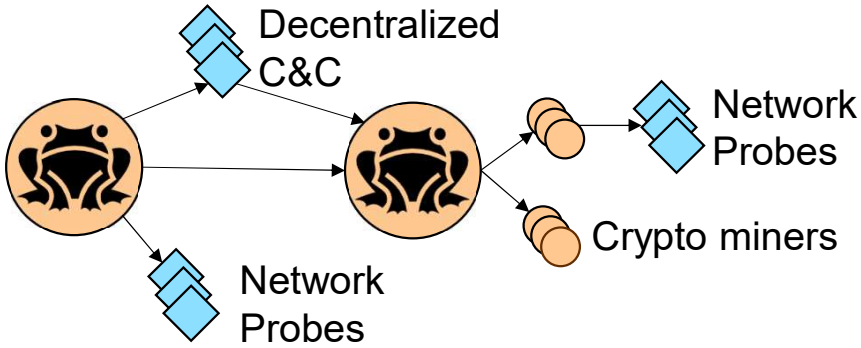
The FritzFrog botnet has claimed **over 20,000 victims** [FritzFrog]

Exploits Log4j and weak SSH passwords to impersonate vulnerable webservers

Provenance Graphs

- Use dynamic **runtime behavior** to detect fileless malware [Hassan '19]
- System provenance causally connects system resources

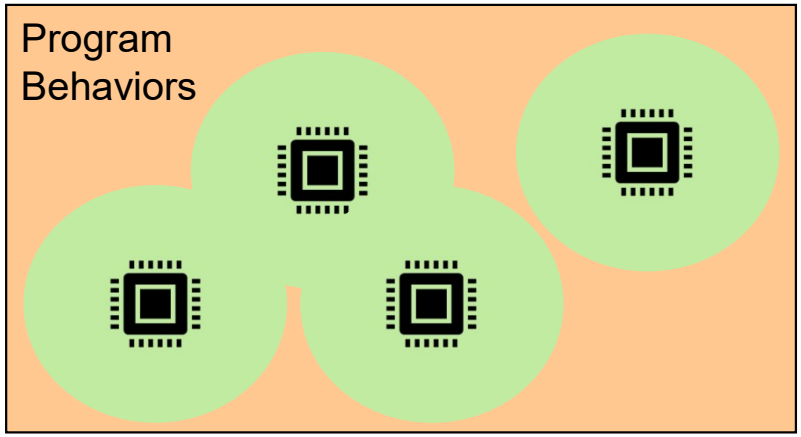




Why hasn't system provenance been adopted in IoT?

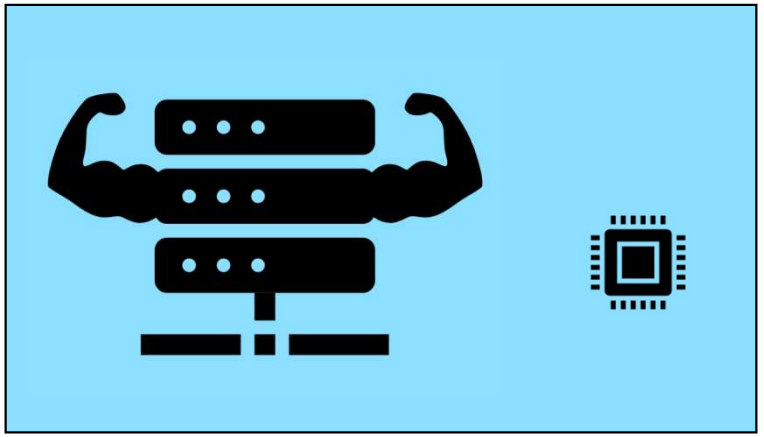
Challenges to system provenance in IoT

Exposure Limitation



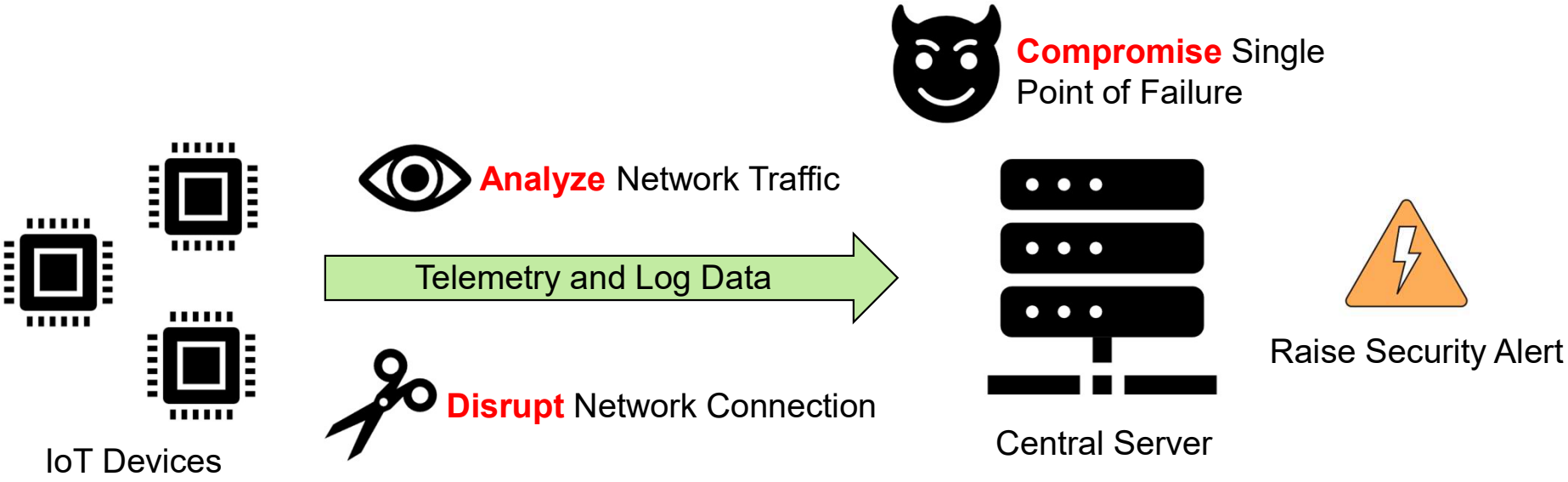
Each device has **limited exposure** to program behaviors

Resource Limitation



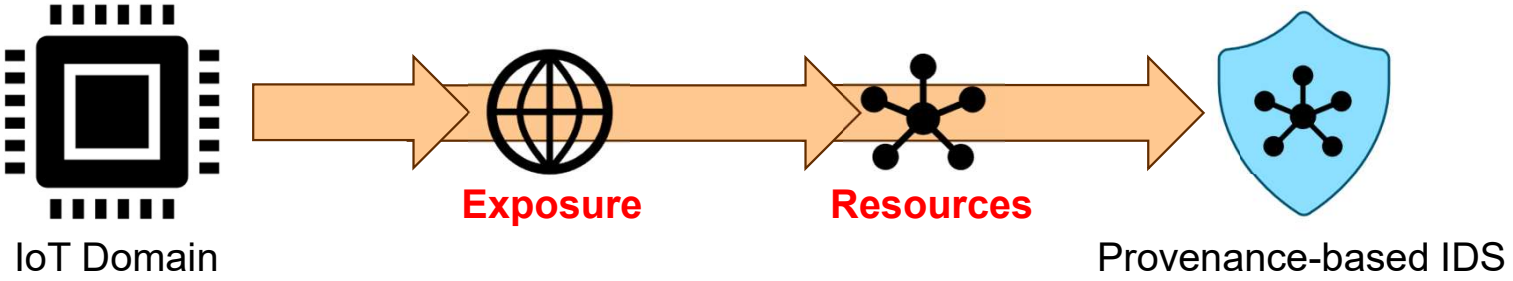
GNN-based graph analysis is **resource-intensive**

Traditional Method of Addressing IoT Limitations

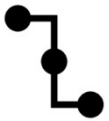


[Nguyen ’19, Cossan ‘21, Reiger ‘23]

ProvIoT - Design

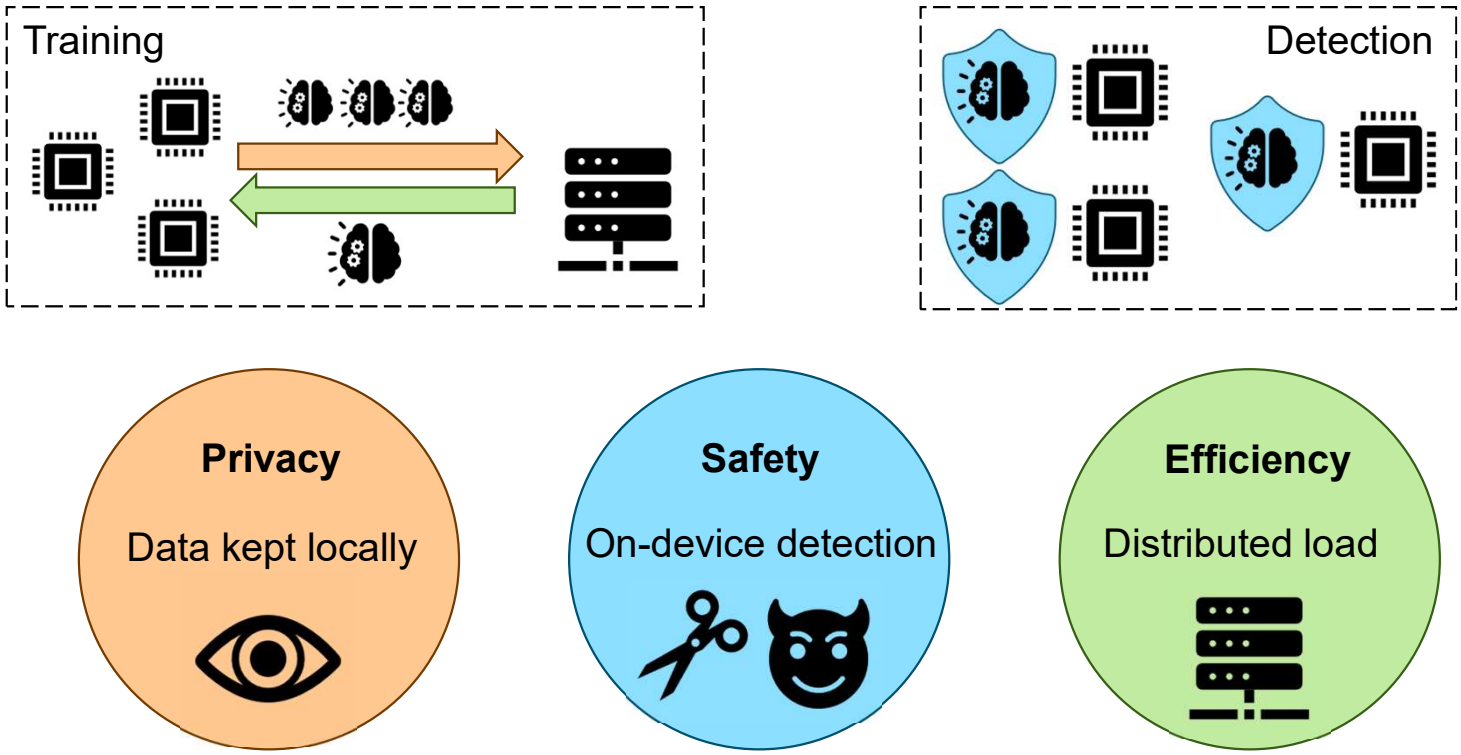


Federated Learning

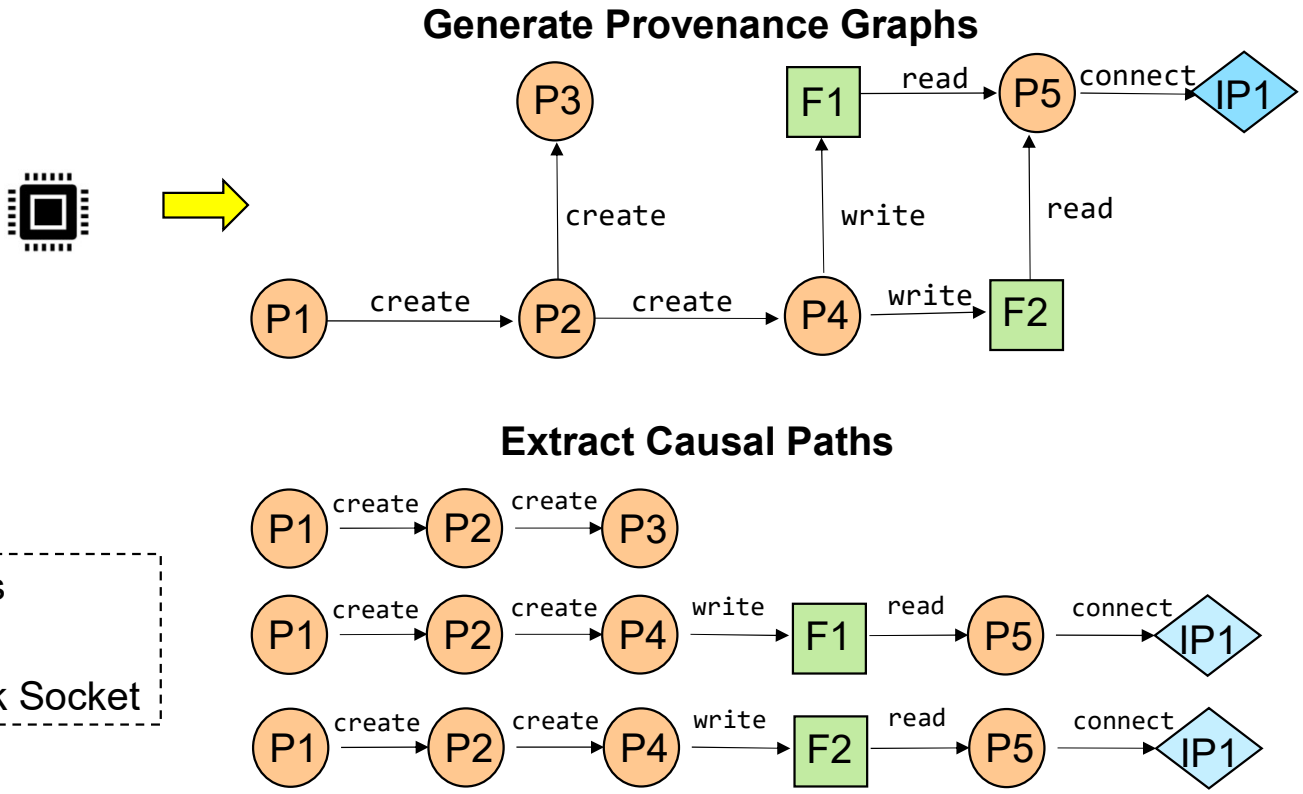


Lightweight Provenance Analysis

ProvIoT – Federated Learning



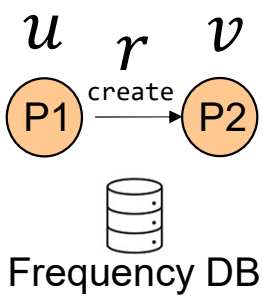
ProvIoT – Data Preparation (Path Extraction)



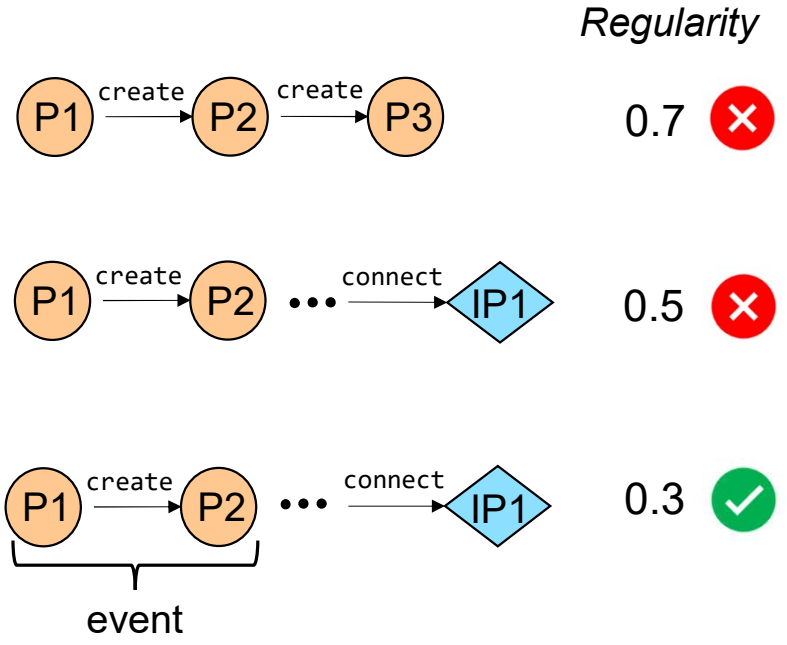
ProvIoT – Data Preparation (Regularity Filtering)

Calculate Regularity Score
[Hassan '19, Wang '20]

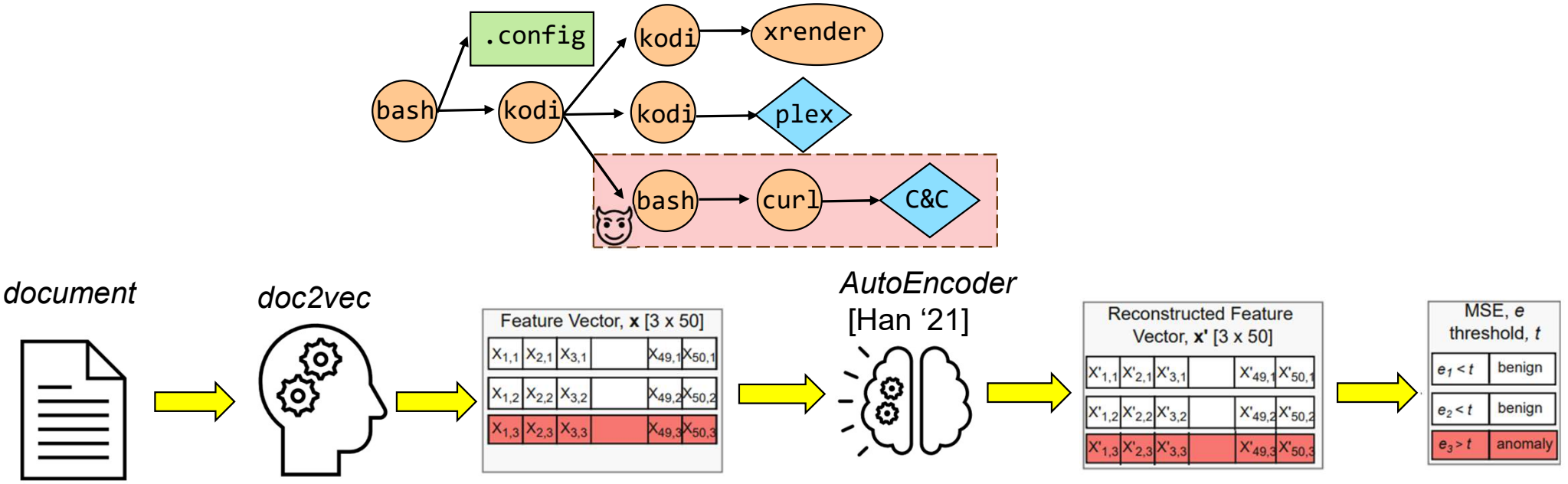
$$R\big(e=(u,v,r)\big)=\frac{|Freq(u,v,r)|}{|Freq(u,*,r)|}$$



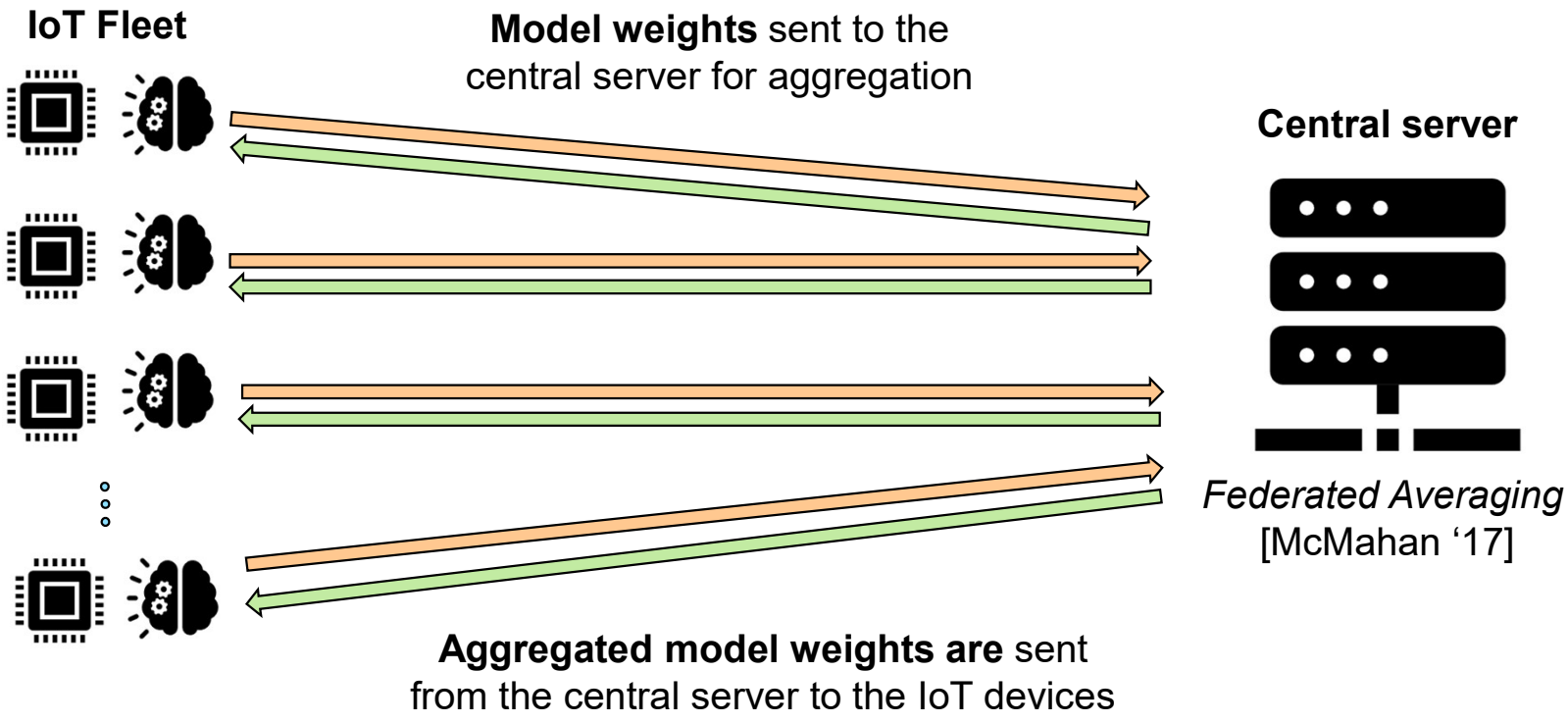
Select Low Regularity Paths



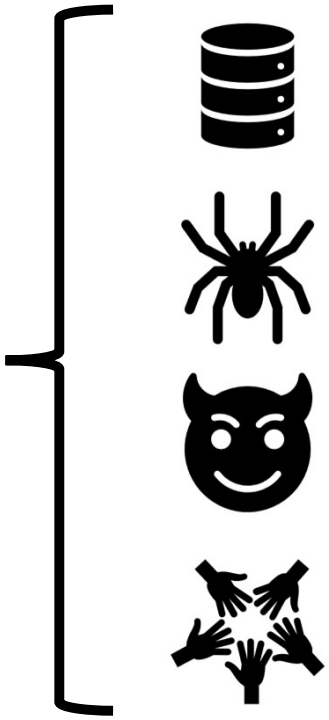
ProvIoT – Anomaly Detection Pipeline



ProvIoT – Model Synchronization



Evaluation



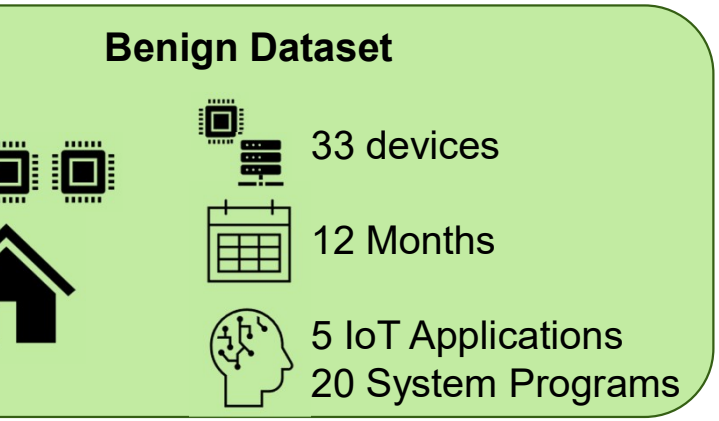
Training Dataset

Fileless Malware Detection

APT Detection

Federated Benefit and Overhead

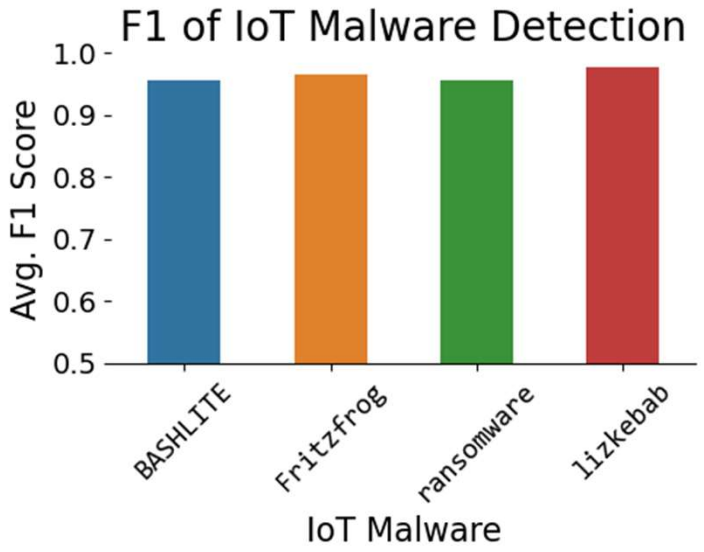
Training Data: `syscall` traces



IoT Applications: `google, kodi, motion, samba, zeek`

System Programs: `bash, cat, cp, cron, dash, dbus-daemon, dd, firefox, grep, java, ls, nginx, perl, ps, python, rm, service, sh, smbd, sshd`

IoT Malware Detection Result



Accurate

Detection rates >95%

Dynamic

Detects fileless malware behavior

General

Evaluated across 20 system programs

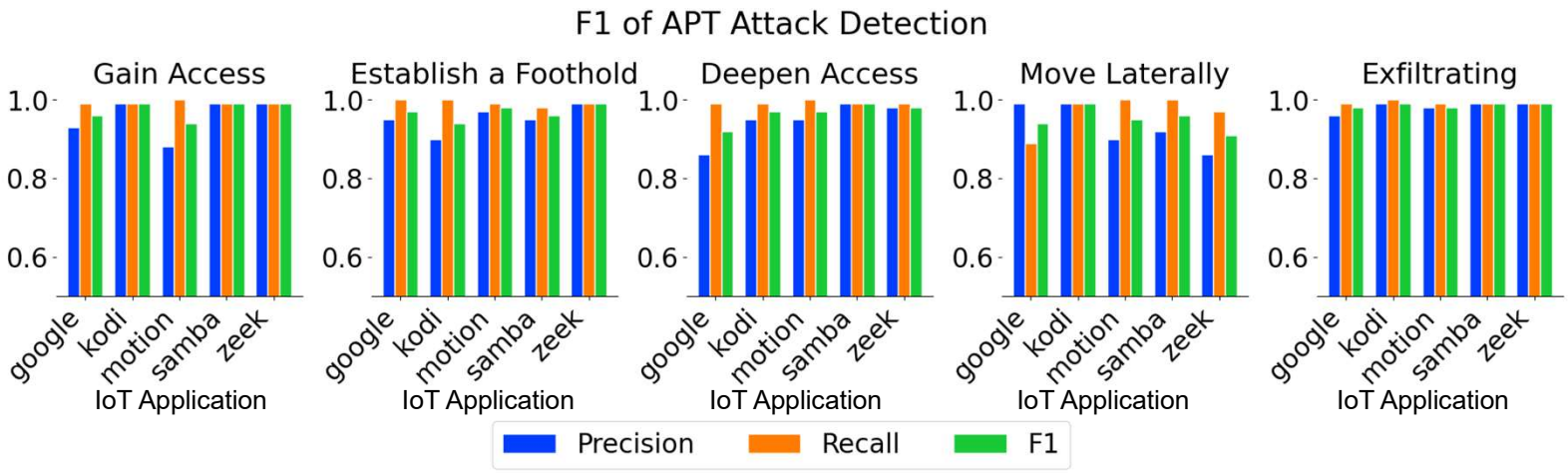
APT Attack Detection Result

APT Attacks

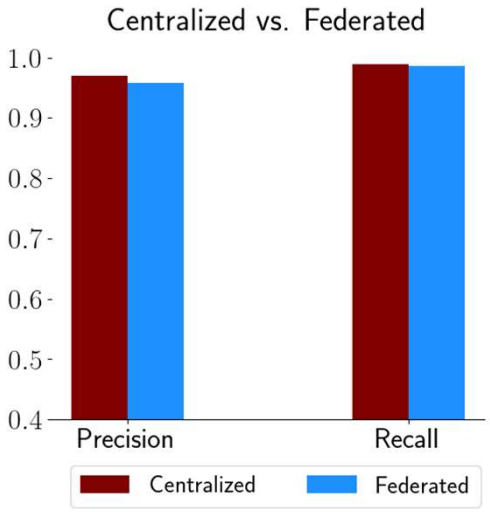
Attack compliant with the MITRE ATT&CK framework

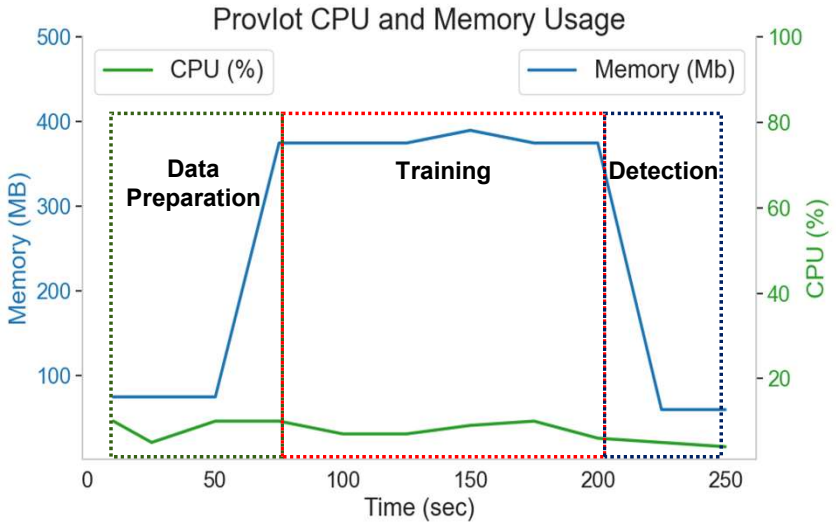
MITRE ATT&CK™

[Mukherjee '23]



Federated Benefit and Overhead





Performance
Competitive federated performance

Cost
Low data preparation and detection overhead

Flexibility
Adaptive training schedule

Conclusion

ProvIoT detects fileless malware in IoT devices **without relying on a central detection server**

97%
average fileless malware detection rate

99%
average APT detection rate

Edge-cloud collaborative security framework

Provenance analysis in IoT devices

THANK YOU

Please forward any questions, comments and future collaboration opportunities to kunal.mukherjee@utdallas.edu

Scan the QR code to access the paper*

References

McMahan ‘17 - McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. Proceedings of Machine Learning Research, 2017.

Dang ’19 - Dang, Fan, et al. "Understanding fileless attacks on linux-based iot devices with honeycloud." Proceedings of Annual International Conference on Mobile Systems, Applications, and Services. 2019.

Hassan ’19 - Hassan, Wajih Ul, et al. "Nodoze: Combatting threat alert fatigue with automated provenance triage." Network and Distributed Systems Security Symposium. 2019.

Nguyen ‘19 – Nguyen, Thien Duc, et al. “DÏoT: A Federated Self-learning Anomaly Detection System for IoT.” IEEE 39th International Conference on Distributed Computing Systems. 2019.

Wang ’20 - Wang, Qi, et al. "You Are What You Do: Hunting Stealthy Malware via Data Provenance Analysis." Network and Distributed Systems Security Symposium. 2020.

Cosson ‘21 – Cossan, Adrien, et al. “Sentinel: A Robust Intrusion Detection System for IoT Networks Using Kernel-Level System Information.” Proceedings of the International Conference on Internet-of-Things Design and Implementation. 2021.

Han ‘21 - Han, Xueyuan, et al. "SIGL: Securing Software Installations Through Deep Graph Learning." USENIX Security Symposium. 2021.

Mukherjee ‘23 – Mukherjee, Kunal, et al. “Evading Provenance-Based ML Detectors with Adversarial System Actions." USENIX Security Symposium. 2023.

Rieger ‘23 – Rieger, Phillip, et al. “ARGUS: Context-Based Detection of Stealthy IoT Infiltration Attacks.” USENIX Security Symposium. 2023.

FritzFrog – David, Ori. “Frog4Shell — FritzFrog Botnet Adds One-Days to Its Arsenal“. Akamai. 2024. https://www.akamai.com/blog/security-research/fritzfrog-botnet-new-capabilities-log4shell.

Backup – Dataset

	Avg. # of causal paths	Avg. # of total vertices / edges	Avg. # of forward vertices / edges	Avg. # of backward vertices / edges
IoT Application				
google	571,052.33	159.0 / 314.0	95.67 / 216.0	63.33 / 98.0
kodi	29,946.89	210.33 / 273.78	149.33 / 176.89	61.0 / 96.89
motion	9,113.0	179.0 / 504.0	5.0 / 4.0	174.0 / 500.0
samba	85,347.0	2,537.0 / 2,857.0	76.4 / 120.8	2,460.6 / 736.2
zeek	494,160.0	2,149.5 / 3,724.5	1,032.5 / 1,124.5	1,117.0 / 2,600.0
average	**237,923.84**	**1,046.97 / 1,534.66**	**271.78 / 328.44**	**775.19 / 1,206.22**
System Program				
bash	166,355.43	454.25 / 510.76	10.57 / 9.31	443.68 / 501.45
cat	184,346.43	310.51 / 210.9	9.0 / 6.99	301.51 / 203.91
cp	175,636.86	193.42 / 212.7	179.09 / 184.69	14.33 / 28.01
cron	214,827.71	327.16 / 241.85	10.27 / 9.96	316.89 / 231.89
dash	153,808.57	371.87 / 381.97	211.61 / 206.44	160.26 / 175.53
dbus-daemon	156,713.0	20.16 / 20.04	9.02 / 6.42	11.14 / 13.62
dd	213,601.29	995.5 / 1,003.6	551.68 / 501.81	443.82 / 501.79
firefox	176,843.86	194.22 / 504.56	15.84 / 18.78	178.38 / 485.78
grep	212,413.86	191.51 / 502.32	13.51 / 16.43	178.0 / 485.89
java	169,180.71	133.94 / 222.4	17.44 / 19.63	116.5 / 202.77
ls	179,185.86	213.62 / 356.47	10.25 / 9.3	203.37 / 347.17
nginx	258,367.17	514.27 / 514.13	500.76 / 501.26	13.51 / 12.87
perl	809.0	25.01 / 23.22	11.95 / 12.05	13.06 / 11.17
ps	181,846.43	834.01 / 998.14	369.21 / 501.77	464.8 / 496.37
python	161,755.57	365.71 / 348.31	11.51 / 8.14	354.2 / 340.17
rm	174,590.43	452.89 / 440.38	15.06 / 18.5	437.83 / 421.88
service	231.43	18.32 / 21.24	15.32 / 18.55	3.0 / 2.69
sh	208,367.43	445.01 / 851.27	4.16 / 357.78	440.85 / 493.49
smbd	201,559.57	355.37 / 371.15	9.69 / 3.39	345.68 / 367.76
sshd	182,601.57	233.04 / 234.15	9.35 / 6.6	223.69 / 227.55
average	**168,652.11**	**332.49 / 398.48**	**99.26 / 120.89**	**233.23 / 277.59**

	Avg. # of causal paths	Avg. # of total vertices / edges	Avg. # of forward vertices / edges	Avg. # of backward vertices / edges
IoT Malware				
BASHLITE	110.5	21.0 / 21.0	4.0 / 3.0	17.0 / 18.0
FritzFrog	46,253.8	751.0 / 747.4	248.6 / 246.8	502.4 / 500.6
lizkebab	293.2	29.0 / 33.0	6.0 / 4.0	23.0 / 29.0
randomware	250.4	28.0 / 44.0	8.0 / 12.0	20.0 / 32.0
average	**11,726.98**	**207.25 / 211.35**	**66.65 / 66.45**	**140.6 / 144.9**
APT Kill Chain Scenario				
Gain Access (S1)	2,789.6	510.6 / 554.8	495.2 / 537.6	15.4 / 17.2
Establish a Foothold (S2)	46,763.75	470.25 / 550.12	398.38 / 429.5	71.88 / 120.62
Deepen Access (S3)	1,192.4	171.0 / 202.6	164.0 / 195.0	7.0 / 7.6
Move Laterally (S4)	27,314.33	97.5 / 116.0	70.17 / 84.83	27.33 / 31.17
Look, Learn and Remain (S5)	20,521.75	928.12 / 983.5	897.38 / 929.62	30.75 / 53.88
average	**19,716.37**	**435.49 / 481.40**	**405.03 / 435.31**	**30.47 / 46.09**

IoT Malware Detection Result