THE UNIVERSITY OF TEXAS AT DALLAS

# IoT Integration, Adversarial Attacks, and Threat Explanations in Provenance-Based Intrusion Detection Systems


**Kunal Mukherjee**

PhD Student

Computer Science


Advisor: Dr. Kangkook Jee

Co-Advisor: Dr. Murat Kantarcioğlu

Committee Member: Dr. Bhavani Thuraisingham

Committee Member: Dr. Feng Chen

# Agenda

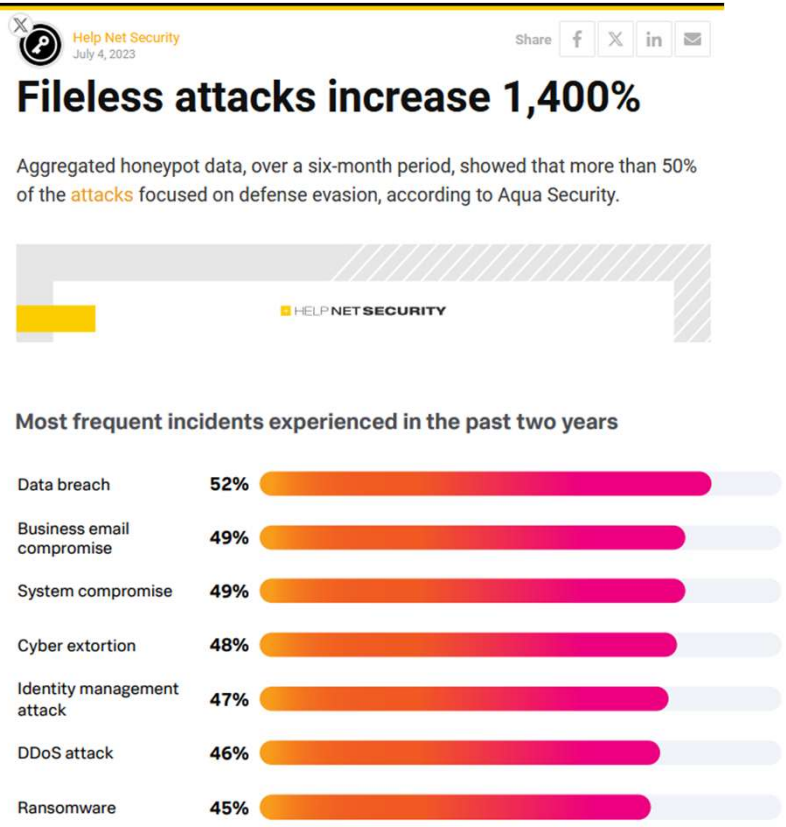
1. Background
2. Motivation
3. **Scope**: ProvIoT
4. **Robustness**: ProvNinja
5. **Explainability:** ProvExplainer
6. Research Contribution
7. Future Work
8. Conclusion

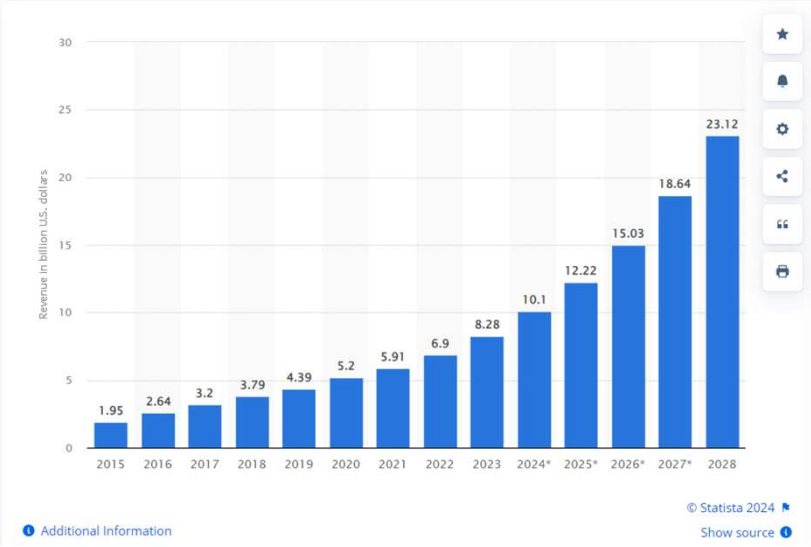
# Agenda

1. **Background**
2. Motivation
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. Research Contribution
7. Future Work
8. Conclusion

Background:

# Stealthy Attacks



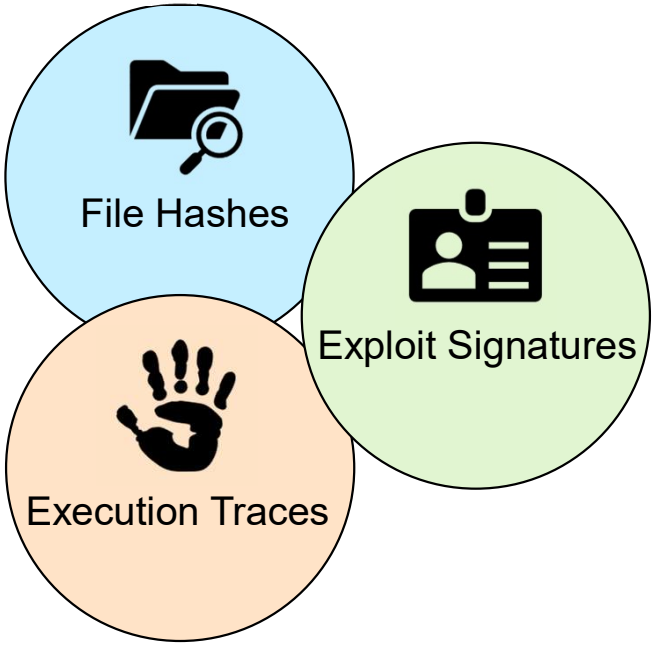




Revenue from advanced persistent threat (APT) from 2015 to 2027

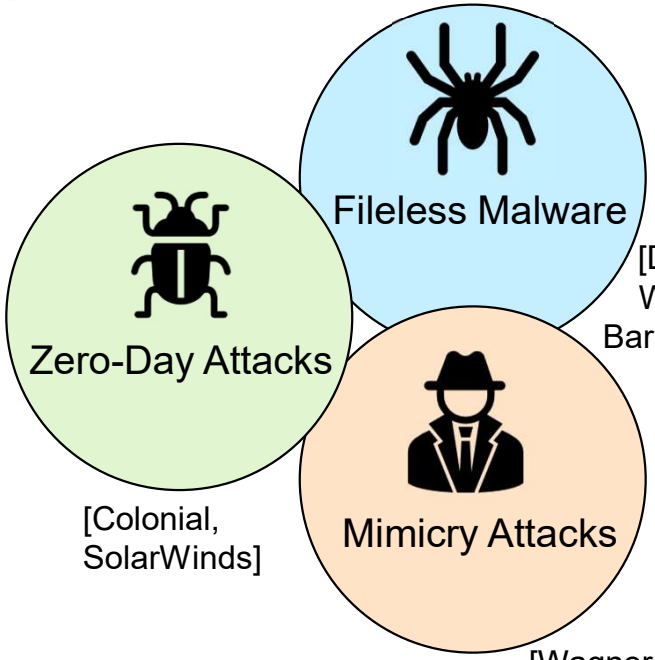
Background:

# Static Host Defenses

Traditional Host *Intrusion Detection System (IDS)* detects **static artifacts**




Adversaries evade detection with **stealthy techniques**




[Dang '19, Wang '20, Barr-Smith '21]

[Colonial, SolarWinds]

[Wagner & Soto '02, Tan & Maxion '03]

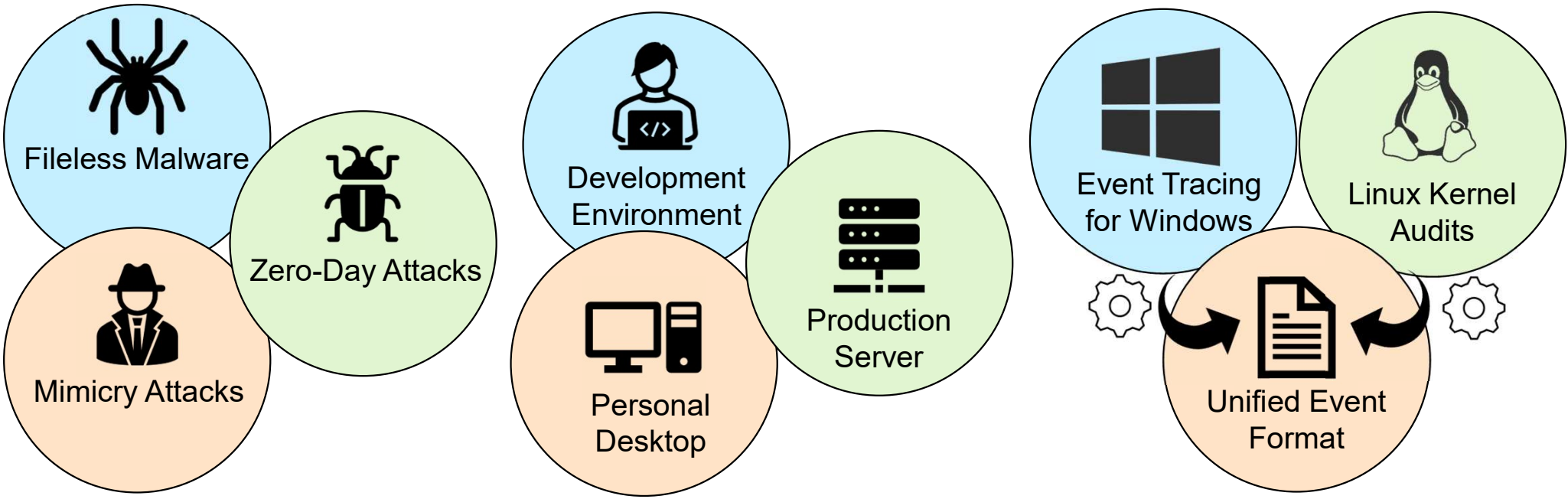
Background:

# Dynamic Host Defenses

Capture dynamic **runtime behaviors**

**Fine-tune** to different environments

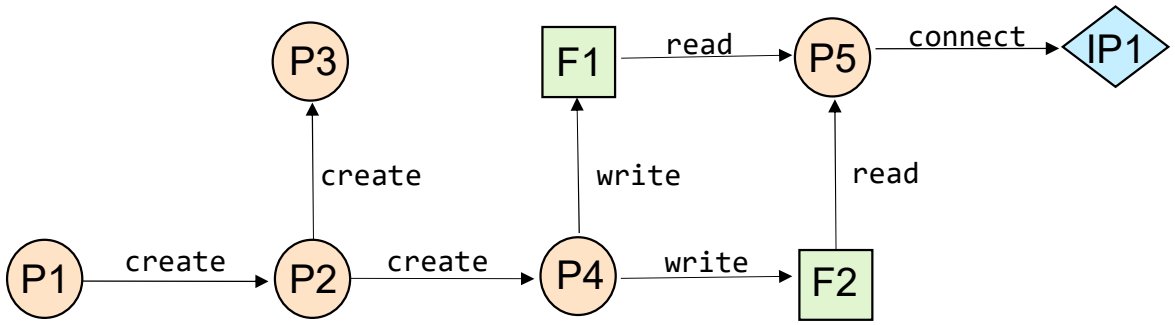
**Platform independent** through generic event collection frameworks



Background:

# System Provenance

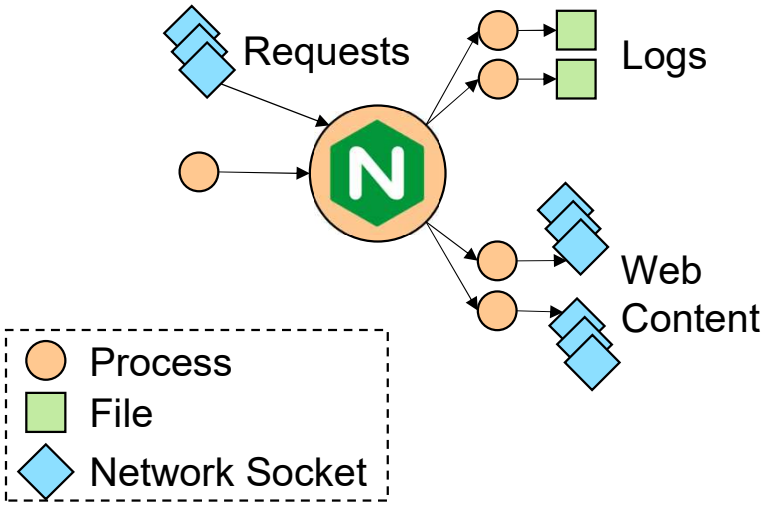
- **System Provenance** championed as a *host-based* dynamic defense
  - Influential works [Hassan '19, Wang '20, Han ‘21, Rehman ’24, Goyal ‘24, ]
- System Provenance *causally* connects system resources
  - Captures *dynamic* control and data dependencies




**How can System Provenance help detect stealthy attacks?**

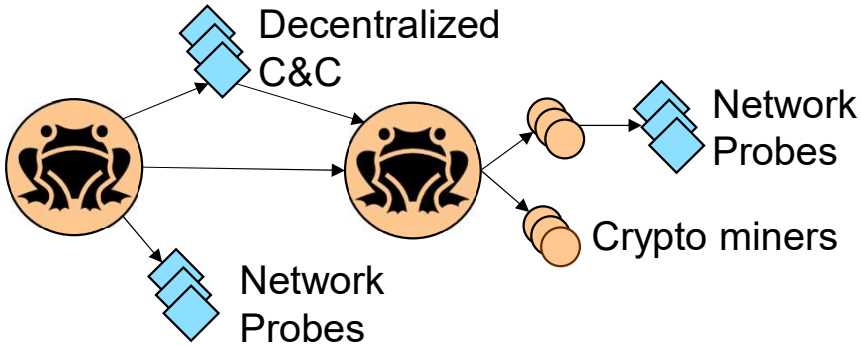
Background:

# Fileless Malware Example




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

Exploits Log4j and weak SSH passwords to impersonate vulnerable Nginx webservers



# Agenda

1. Background
2. **Motivation**
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. Research Contribution
7. Future Work
8. Conclusion

Motivation

# ***Scope*** of Provenance-Based IDS

Provenance-based IDS

Prediction Granularity

| | |
|---|---|
| Velickovic '17 (GNN) | Graph-level detection |
| Wang '20 (LOF) | Path-level detection |
| Han '21 (AutoEncoder) | Path-level detection |
| Zeng ‘22, Cheng ‘24, Rehman ’24 (GNN) | Interaction-level detection |

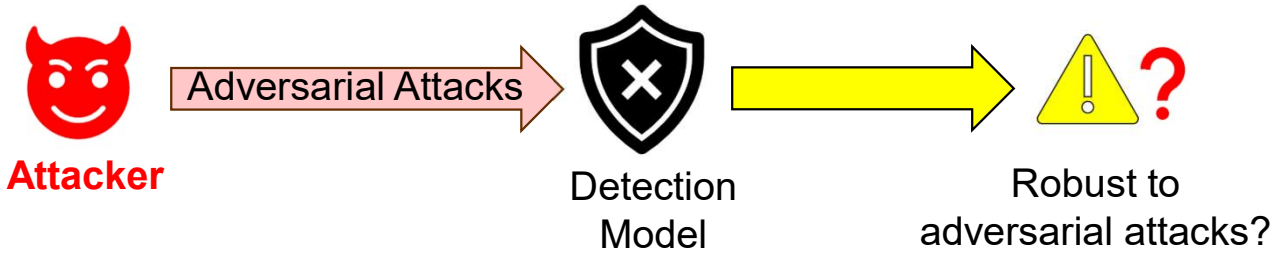
- Fine-Grained View
- Dynamic Behavior
- Long-Range Dependencies

Provenance-based IDS are **resource-intensive**

**Resource-constrained** devices still need protection from stealthy attackers

Motivation

# ***Robustness*** of Provenance-Based IDS

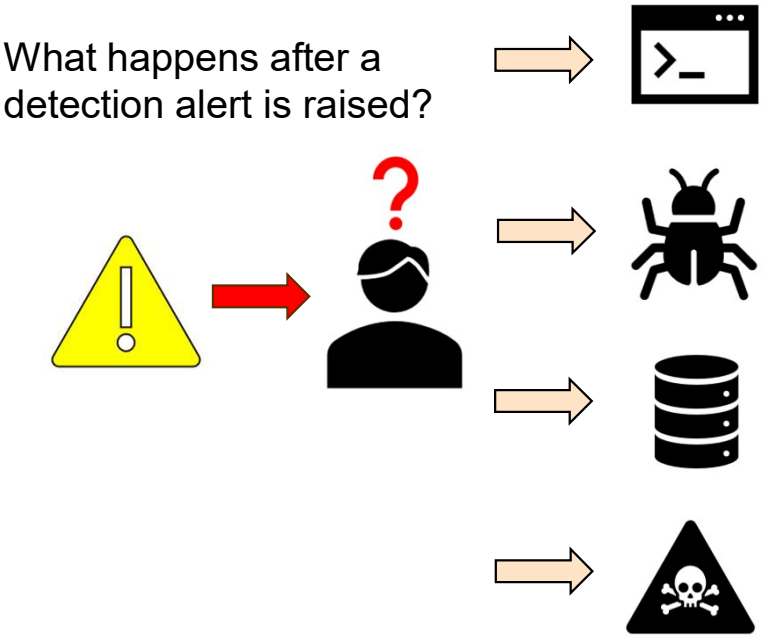



**Robustness** against dedicated adversaries has **not been verified**

Machine learning models need to accurately **detect adversarial attacks**

Motivation

# ***Explainability*** of Provenance-Based IDS




**Trust** in Provenance-based IDS has **not been established**

Human operators need **justification** and **guidance** to triage and respond to alerts

# Agenda

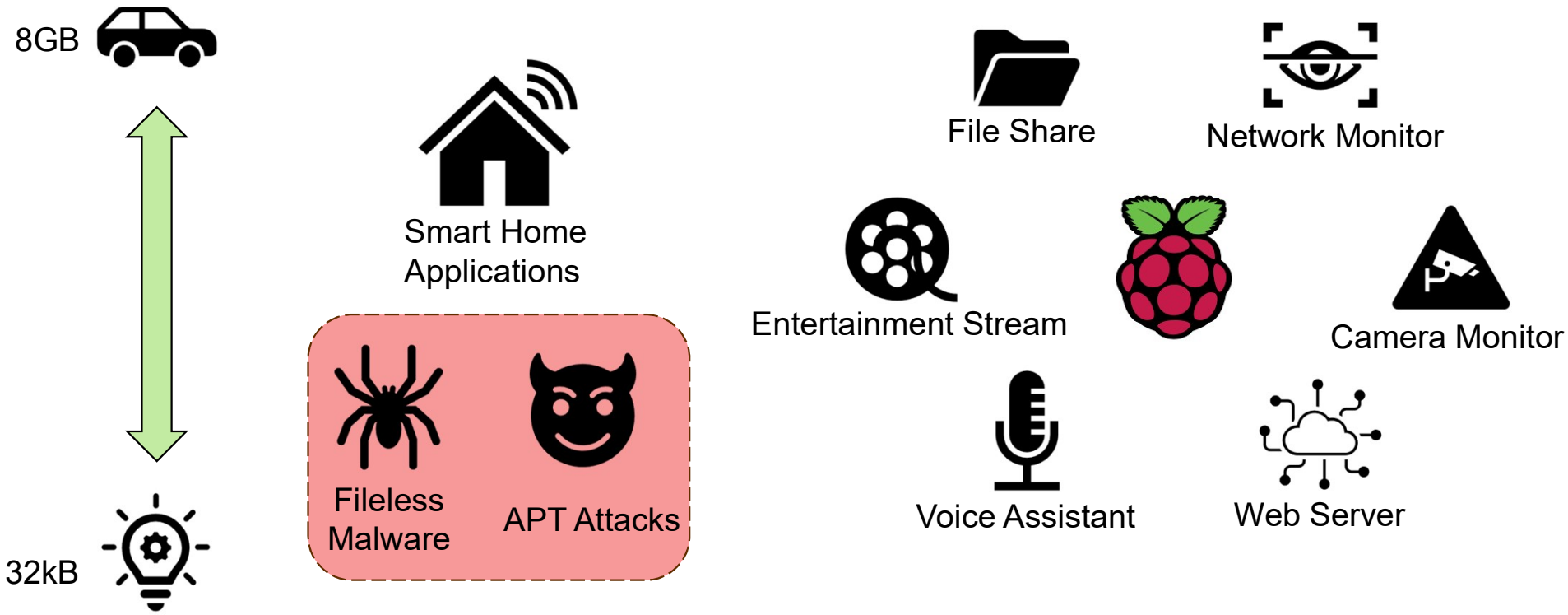
1. Background
2. Motivation
3. **Scope: *ProvIoT***
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. Research Contribution
7. Future Work
8. Conclusion

**Kunal Mukherjee**, Kangkook Jee, et.al. *"ProvIoT:Detecting Stealthy Attacks in IoT through Federated Edge-Cloud Security,"* in ACNS '24

**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. "ProvIoT:Detecting Stealthy Attacks in IoT through Federated Edge-Cloud Security," in *International Conference on Applied Cryptography and Network Security*, Springer, 2024.

***Scope***: ProvIoT

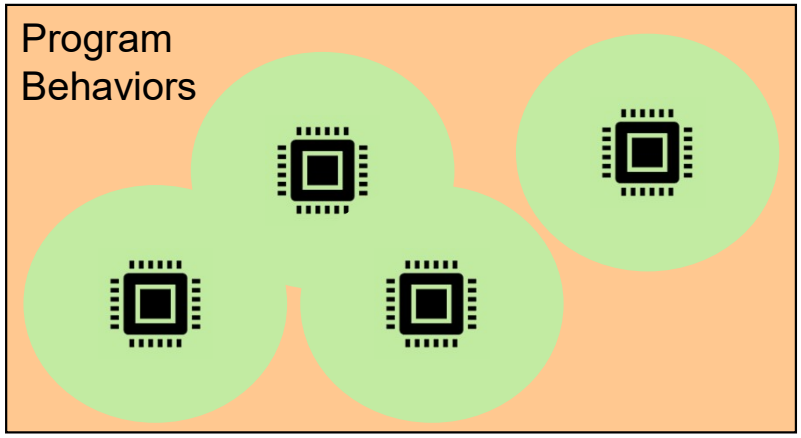
# Motivation



***Scope***: ProvIoT

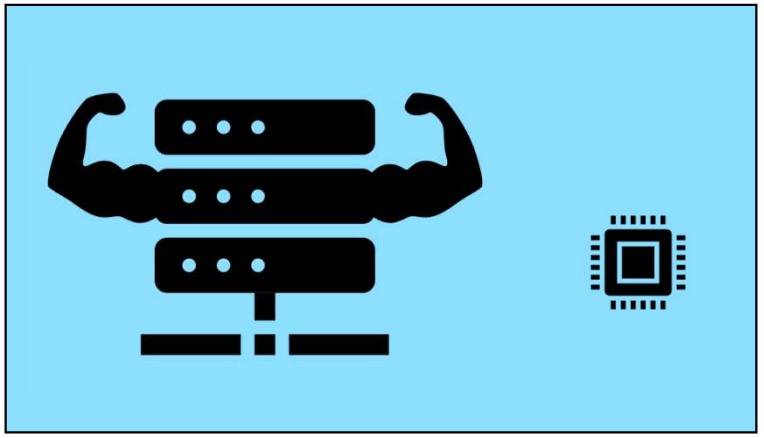
# Motivation : *Challenges*

Exposure Limitation




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

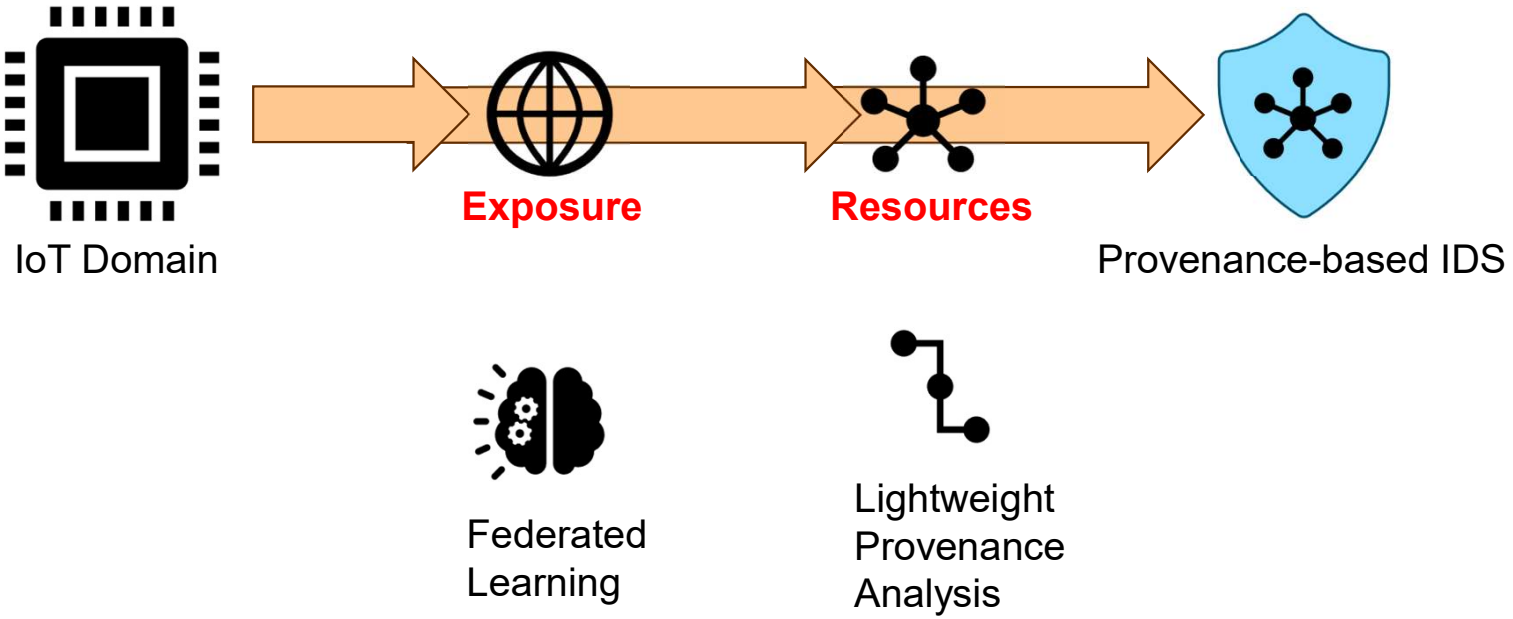
Resource Limitation



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

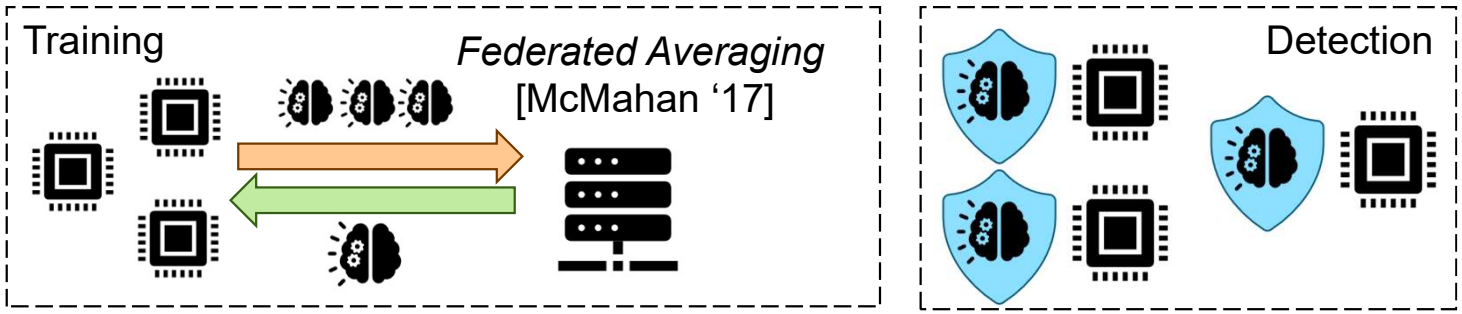
***Scope***: ProvIoT

# Design



***Scope***: ProvIoT

# Design : *Federated Learning*




**Privacy**

Data kept locally

**Safety**

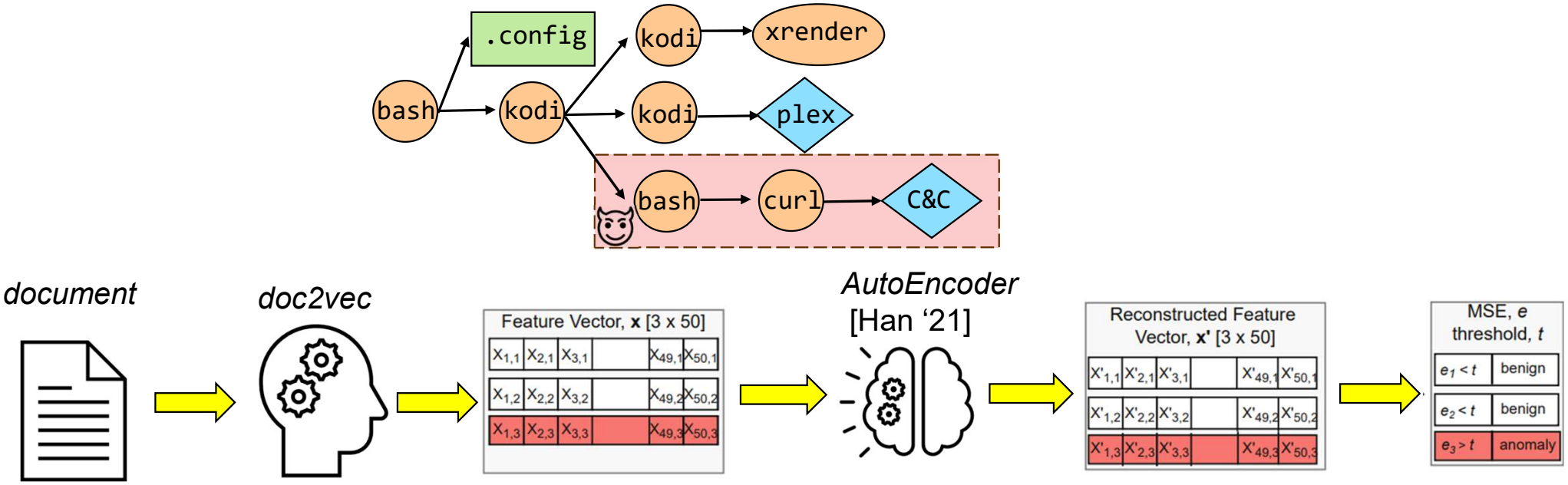
On-device detection

**Efficiency**

Distributed load

***Scope***: ProvIoT

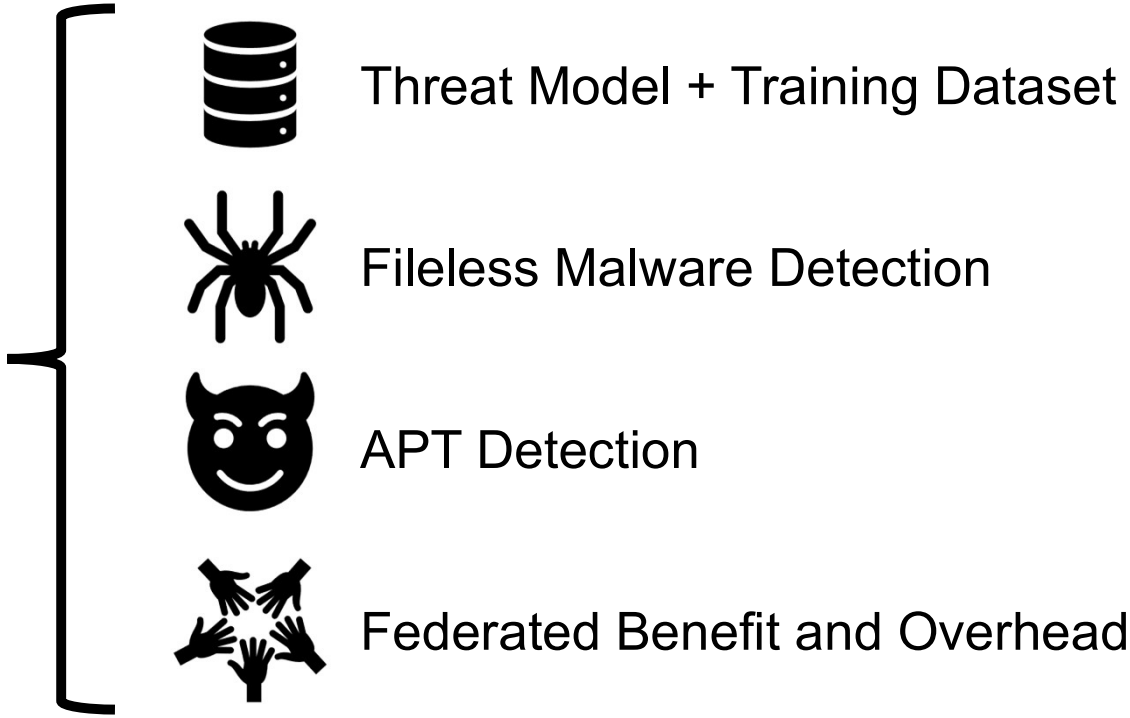
# Design : *Anomaly Detection*



***Scope***: ProvIoT

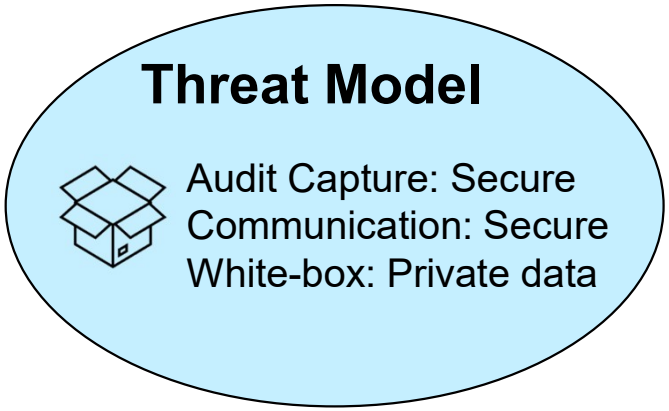
# Evaluation

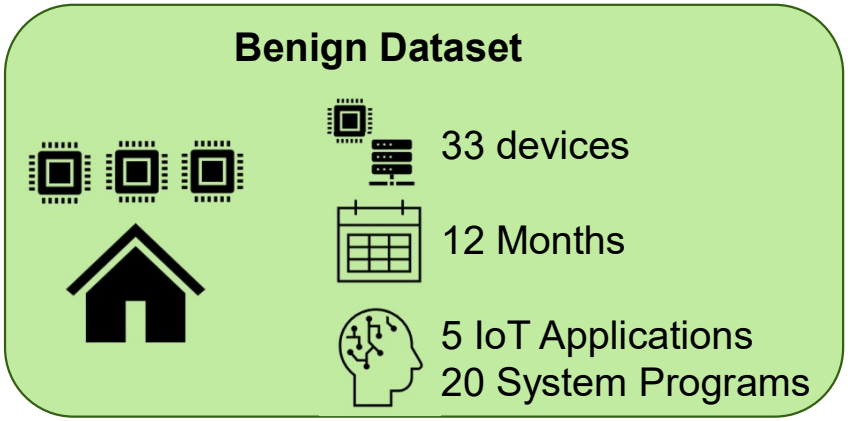
Evaluation



***Scope***: ProvIoT

# Evaluation :*Threat Model* and *Training Data*



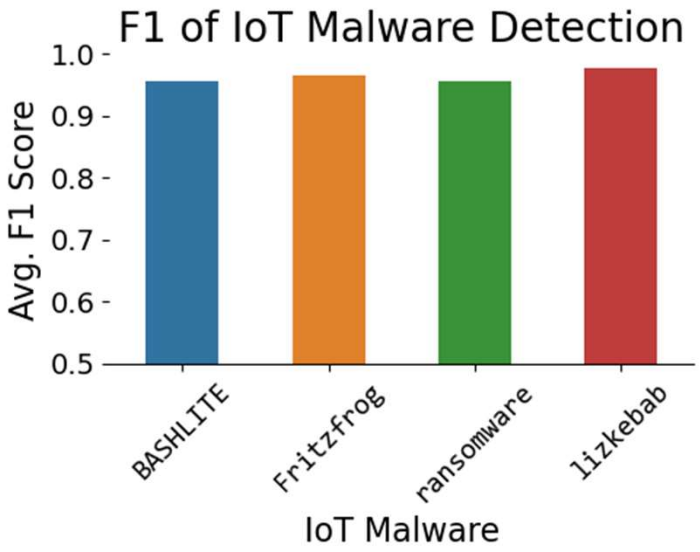




**IoT Applications**: `google, kodi, motion, samba, zeek`

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

***Scope***: ProvIoT

# Evaluation : *Malware Detection*




**Accurate**

Detection rates >95%

**Dynamic**

Detects fileless malware behavior

**General**

Evaluated across 20 system programs

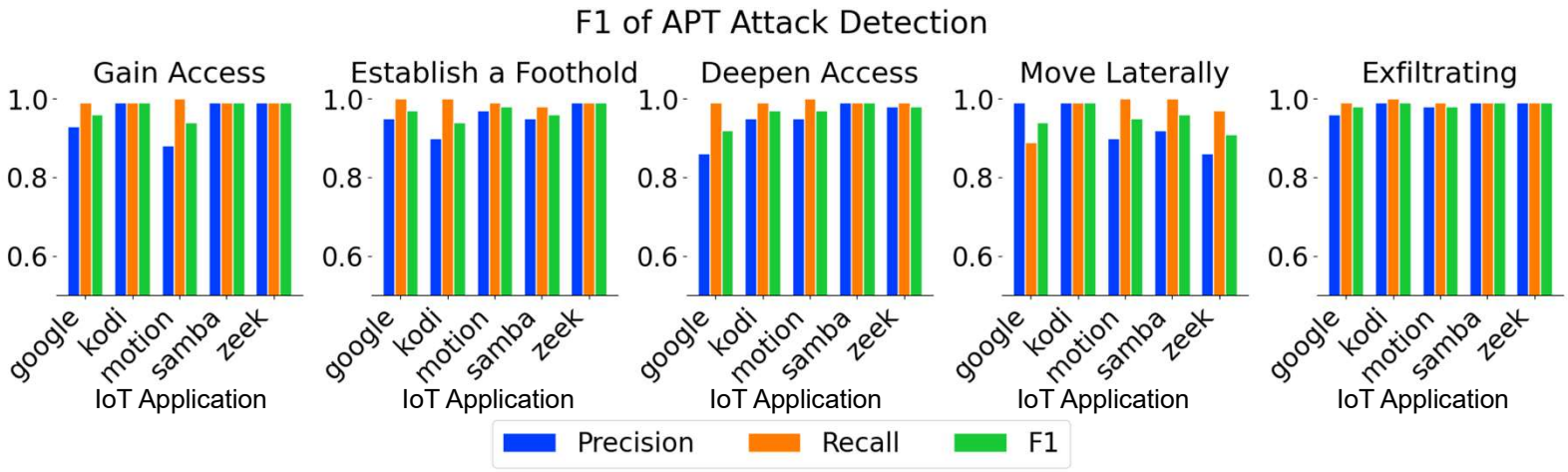
***Scope***: ProvIoT

# Evaluation : *APT Detection*

**APT Attack Stages**

Compliant with the MITRE ATT&CK framework

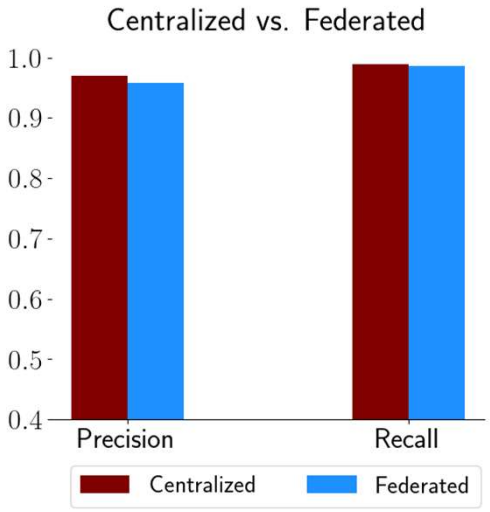
MITRE ATT&CK™

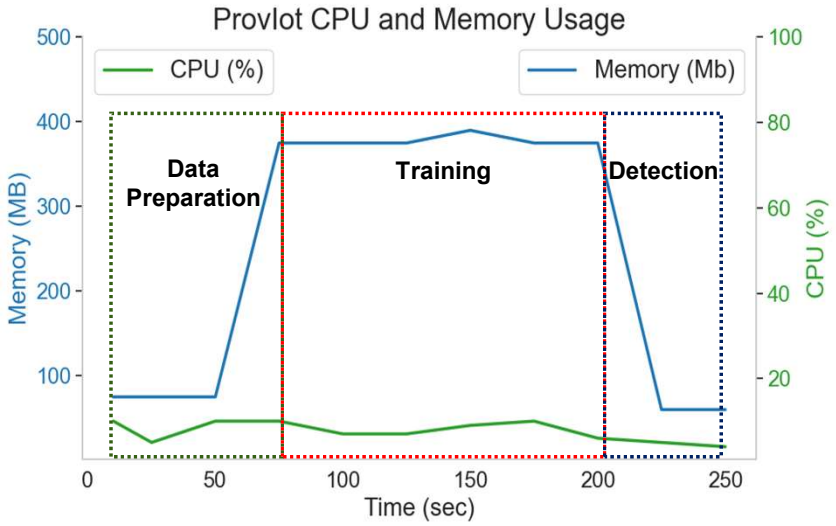



ProvIoT achieves a **high detection rate** across various APT stages and diverse IoT applications.

***Scope***: ProvIoT

# Evaluation : *Federated Benefit and Overhead*




**Performance**
Competitive federated performance

**Cost**
Low data preparation and detection overhead

**Flexibility**
Adaptive training schedule

***Scope***: ProvIoT

# Project Summary

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

# Agenda

1. Background
2. Motivation
3. Scope: ProvIoT
4. **Robustness: *ProvNinja***
5. Explainability: ProvExplainer
6. Research Contribution
7. Future Work
8. Conclusion

**Kunal Mukherjee**, Feng Chen, Murat Kantarcioglu, Kangkook Jee, and et.al. "Evading Provenance-Based ML Detectors with Adversarial System Actions," *USENIX Security 23*

**Kunal Mukherjee**, Joshua Wiedemeier, Tianhao Wang, James Wei, Feng Chen, Muhyun Kim, Murat Kantarcioglu, and Kangkook Jee. "Evading Provenance-Based ML Detectors withAdversarial System Actions," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023.

***Robustness***: ProvNinja

# Motivation

**Trust** in Provenance-based IDS has **not been established**

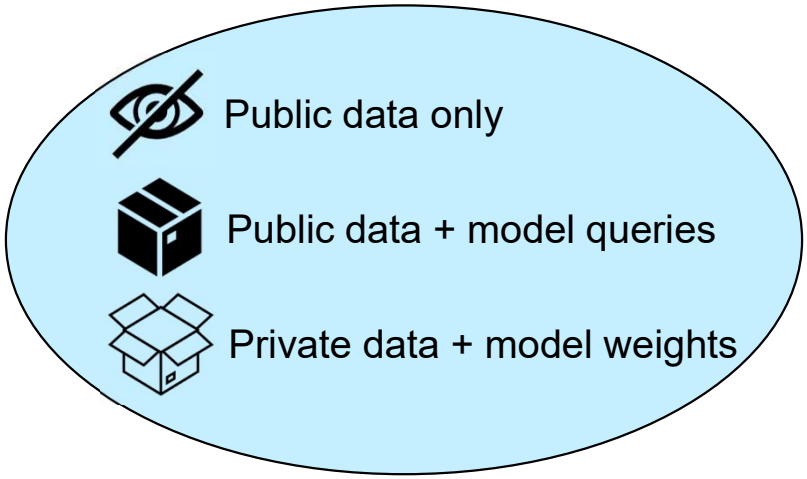
**Robustness** against dedicated adversaries has **not been verified**

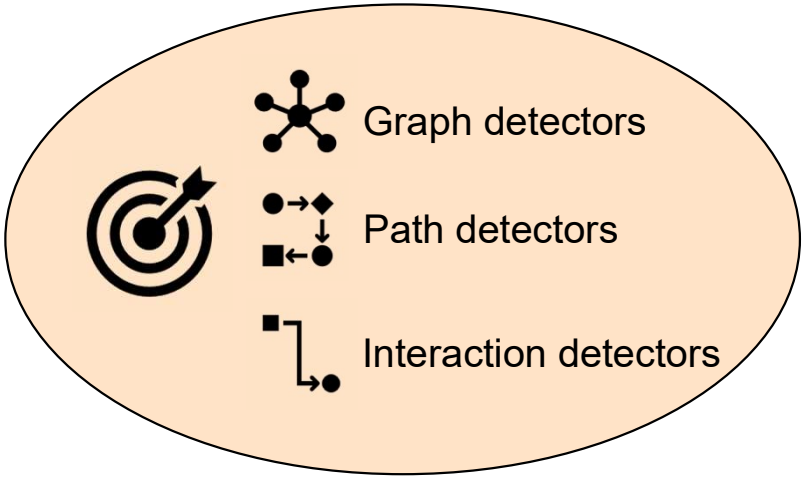
**Adversarial validation** is an established way to **prove robustness**

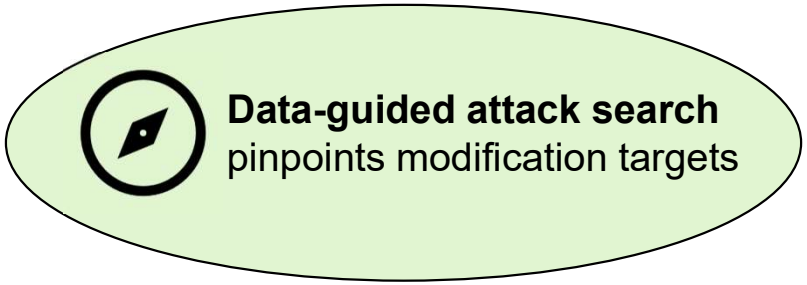
***Robustness***: ProvNinja

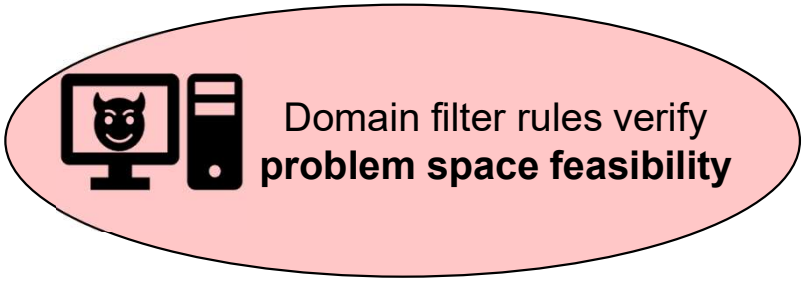
# Design

**Evasive attack framework**



***Robustness***: ProvNinja

# Design : Framework

**ProvNinja:**
**Evasive Attack Framework**

Identify Conspicuous Events

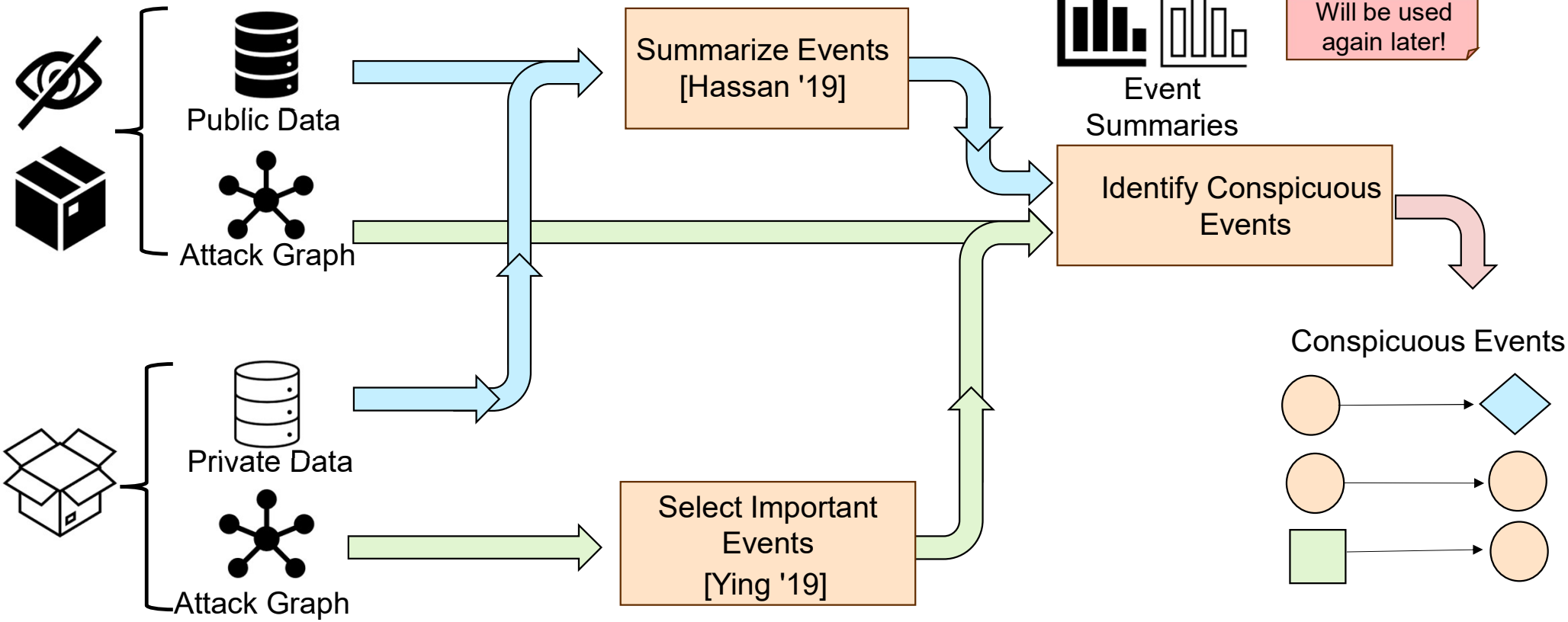
Replace with Common Events

Camouflage Processes

Realize the Evasion

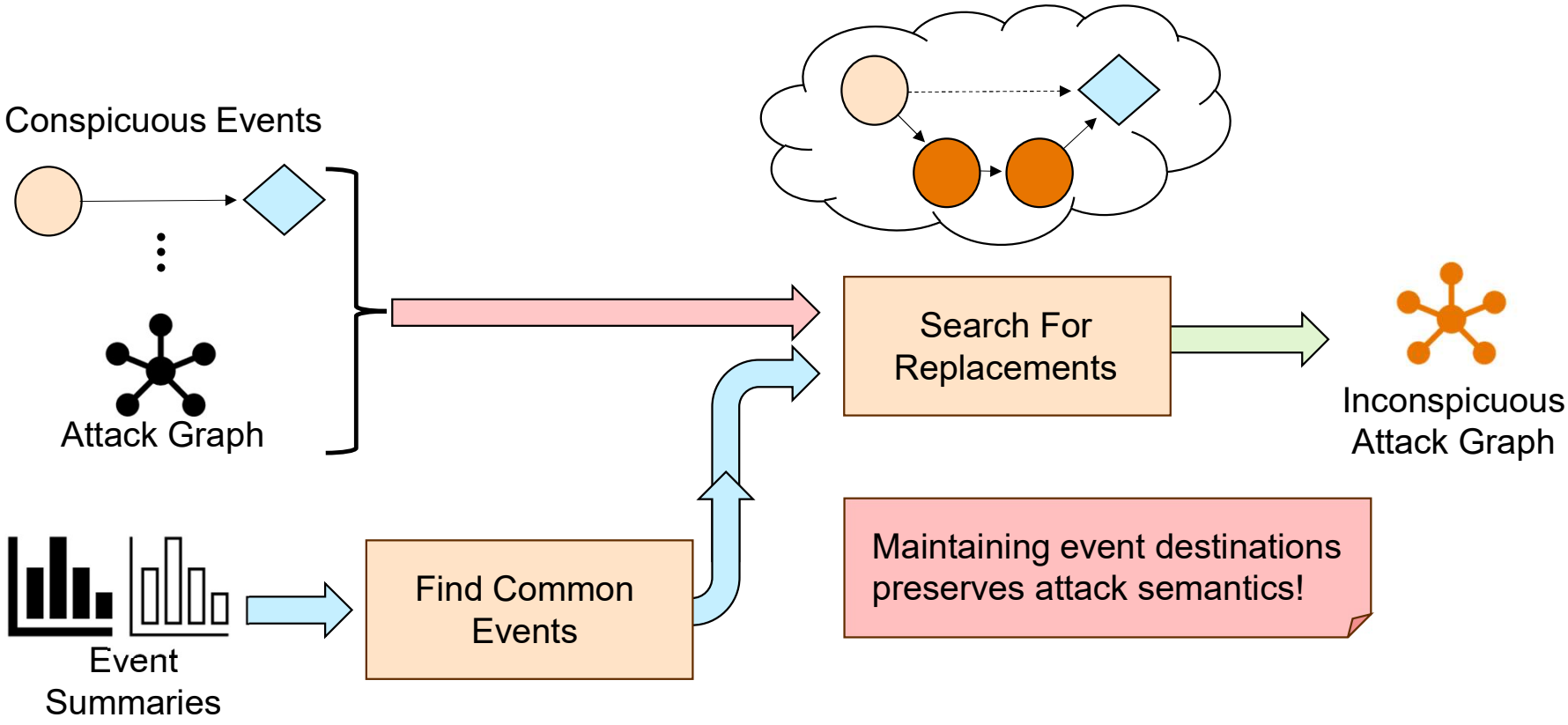
***Robustness***: ProvNinja

# Design : *Identify Conspicuous Events*



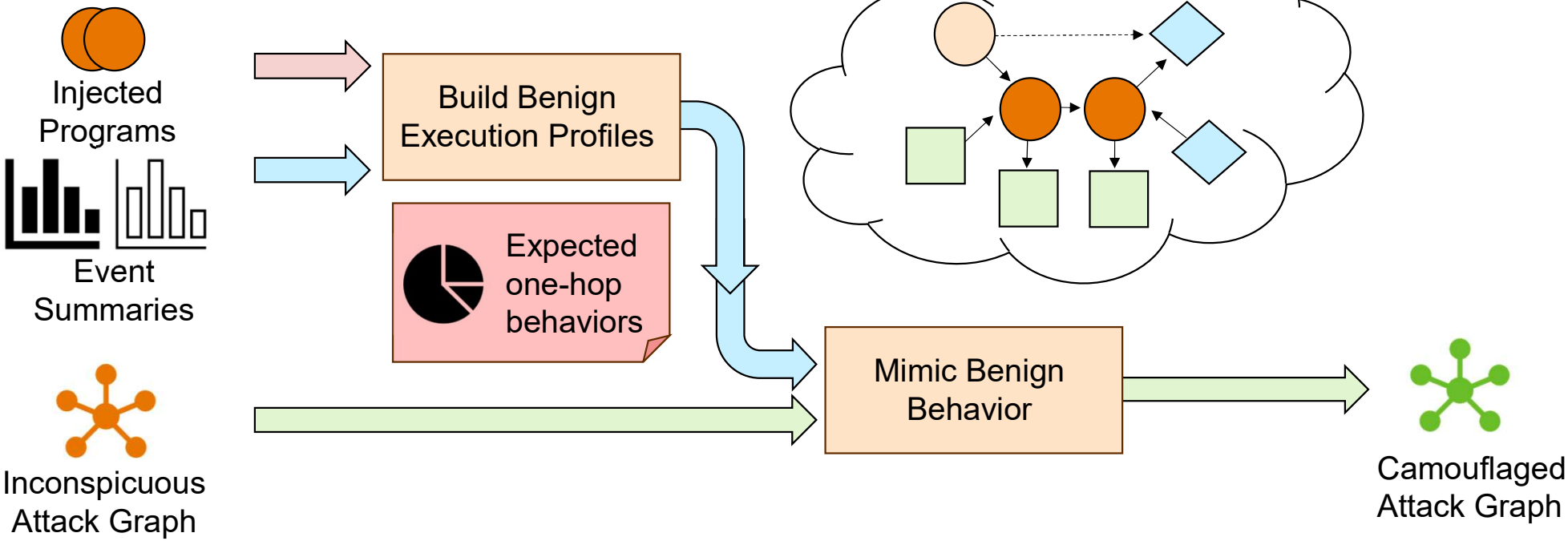
***Robustness***: ProvNinja

# Design : *Replace with Common Events*



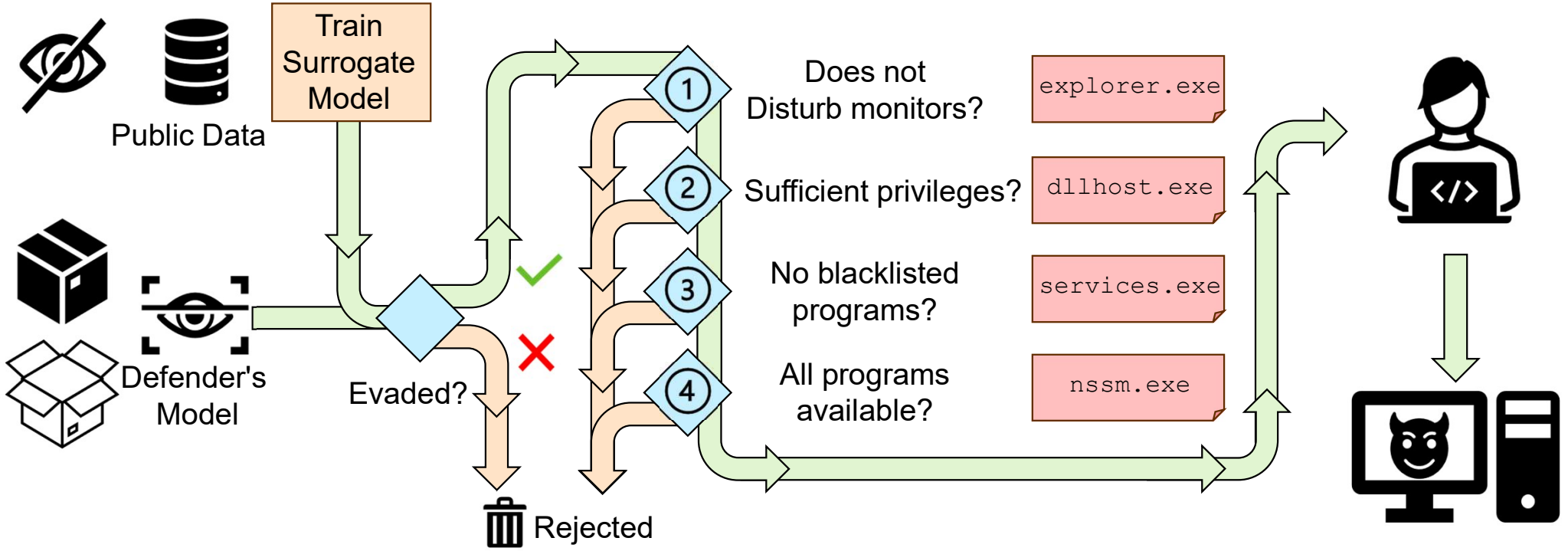
***Robustness***: ProvNinja

# Design : *Camouflage Processes*



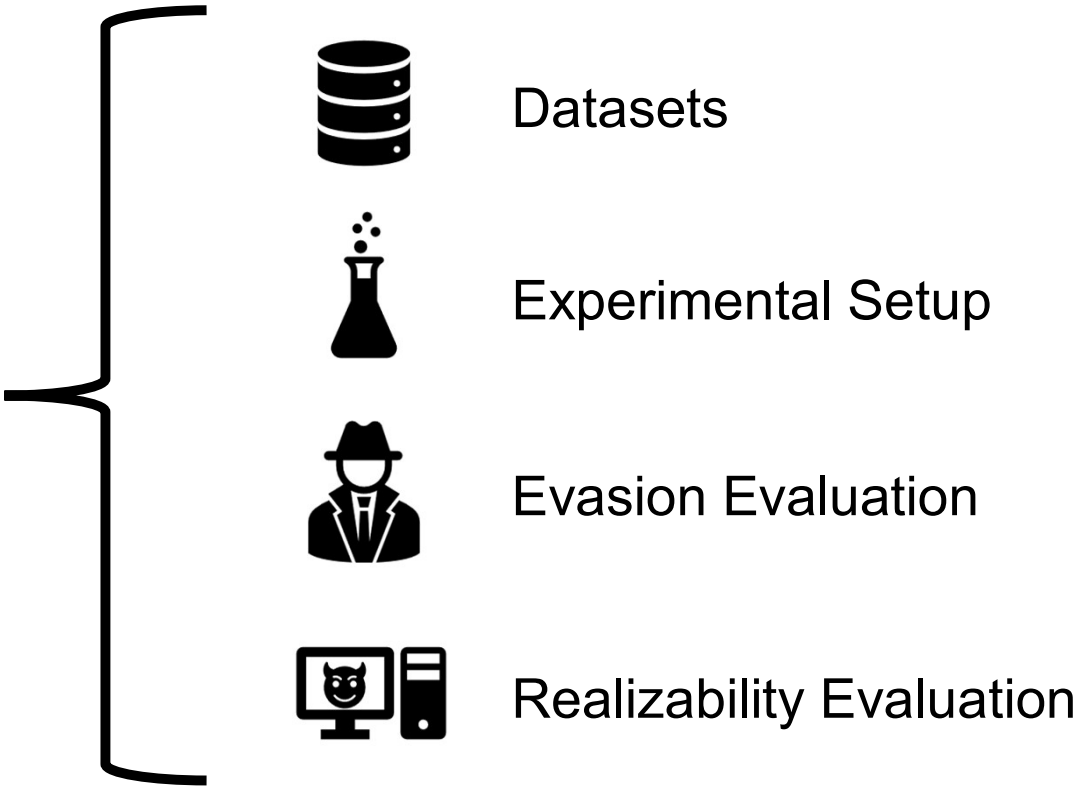
***Robustness***: ProvNinja

# Design : *Realize Evasion*



***Robustness***: ProvNinja
# Evaluation

**Evaluation**



***Robustness***: ProvNinja

# Evaluation : *Datasets*

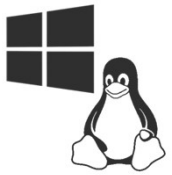
## Benign Datasets

| | DARPA (public) | In-House (private) |
|---|---|---|
| | Scripted | Real Users |
| | 8 Hosts | 86 Hosts |
| | 12 Days | 13 Months |

## Malicious Datasets

Enterprise

**1,779** Graphs



Supply Chain

**1,091** Graphs

Fileless Malware
[Barr-Smith '21]

**1,206** Graphs

***Robustness***: ProvNinja

# Evaluation : *Experimental Setup*

## Threat Models

Blind: Public data only

Black-box: Public data + model queries

White-box: Private data + model weights

## Provenance-based IDS

[Veličković '17]

[Wang '20, Han '21]

[Zeng '22]

## Dataset Allocation

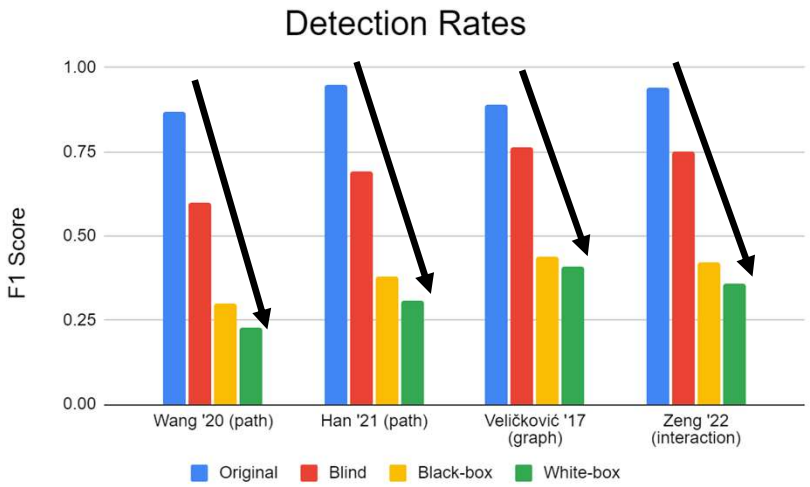
Public

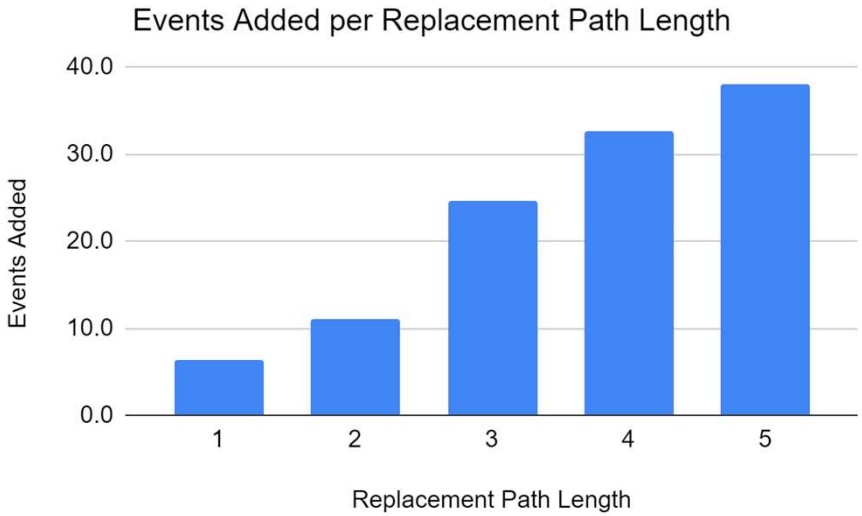
Private

DARPA

***Robustness***: ProvNinja

# Evaluation : *Evasion Result*



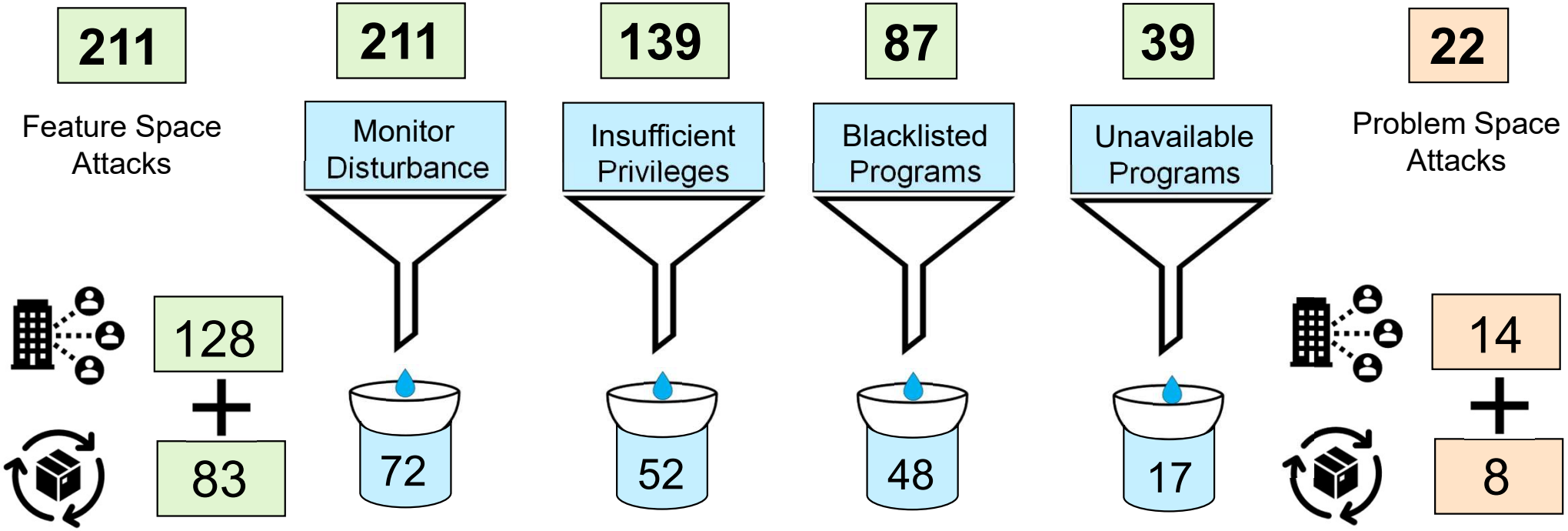

Reduces detection rates against SOTA Provenance-based IDS

Scales to threat model

Each replacement adds fewer than 40 events

***Robustness***: ProvNinja

# Evaluation : *Attack Realizability*



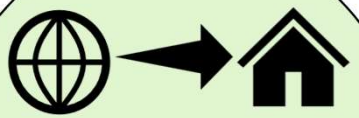
***Robustness***: ProvNinja

# Project Summary

ProvNinja **systematically challenges** Provenance-based IDS

**57%**
Average detection rate reduction

**<150**
Average events added per attack



Transfers behavioral insights across environments

Supports adversarial testing and verification

Inspiring the development of **robust** IDS with **realistic** adversarial examples

# Agenda

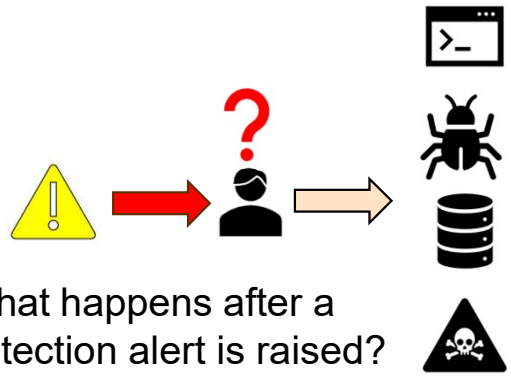
1. Background
2. Motivation
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. **Explainability: *ProvExplainer***
6. Research Contribution
7. Future Work
8. Conclusion

**Kunal Mukherjee**, Feng Chen, Murat Kantarcioglu, Kangkook Jee, et.al. "Interpreting gnn-based ids detections using provenance graph structural features," arXiv:2306.00934, 2023.

**Kunal Mukherjee**, Joshua Wiedemeier, Tianhao Wang, Muhyun Kim, Feng Chen, Murat Kantarcioglu, and Kangkook Jee. "Interpreting gnn-based ids detections using provenance graph structural features," arXiv preprint arXiv:2306.00934, 2023.

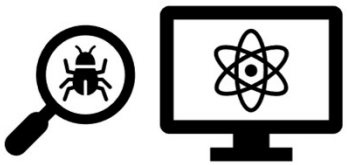
***Explainability***: ProvExplainer

# Motivation

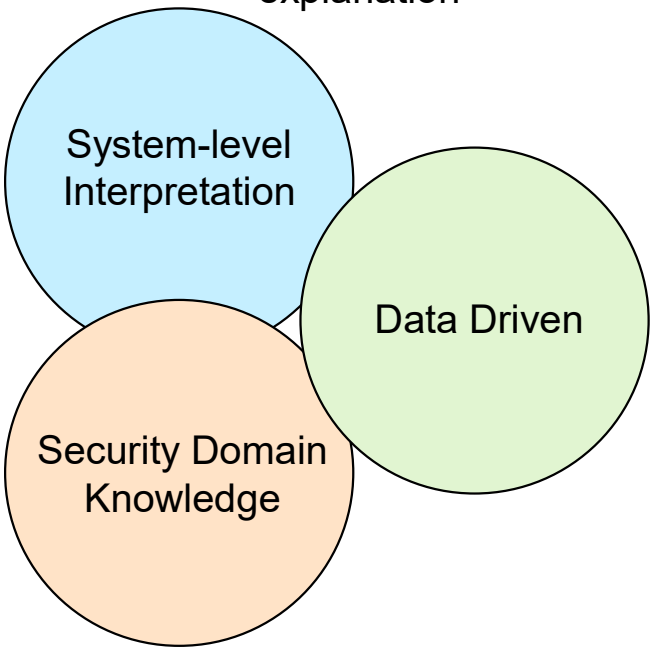


What happens after a detection alert is raised?

Security-aware explanations to **interpret** and bring **context** to alerts.



**Desired** qualities of security-aware explanation



***Explainability***: ProvExplainer

# Motivation : *GNN-based IDS*

SOTA provenance–based IDS [Cheng ‘24, Rehman ’24] use GNN graph embeddings for detection

**Explanation** verification is hindered due to **black-box** nature of GNN

General purpose GNN Explainers are **not** security-aware

***Explainability***: ProvExplainer

# Design

**ProvExplainer:**
**Security Aware**
**Explanation Framework**

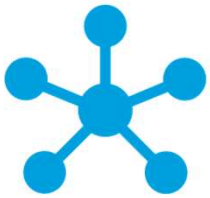
Extract Security Aware Features

Train Surrogate DT

Interpret Surrogate DT

***Explainability***: ProvExplainer

# Design



Instance-level

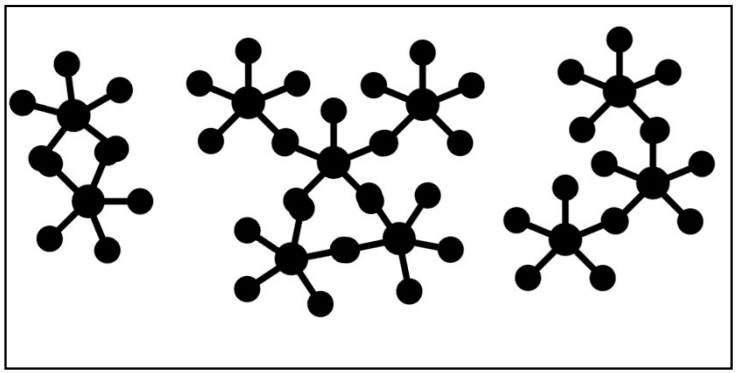
Black box

Structure only

***Explainability***: ProvExplainer

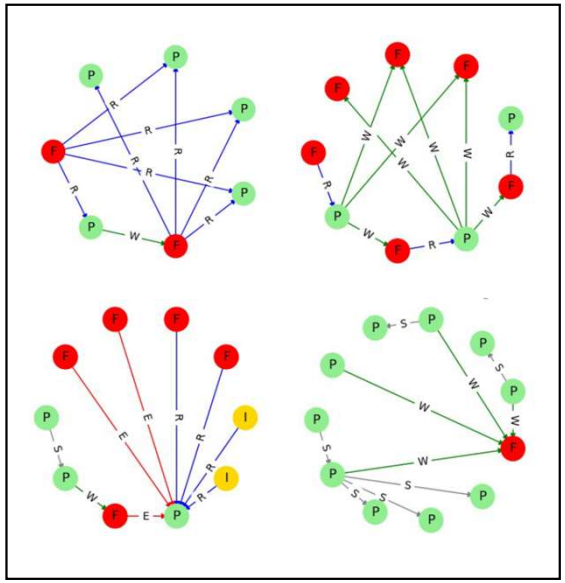
# Design : *Security-Aware Graph Structural Features*

Anomalous Graphs



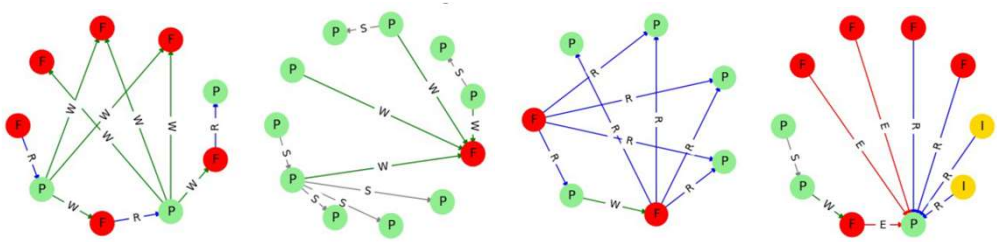
Subgraph Pattern Mining [Graph Evolutionary Rule Miner]

Subgraph Patterns



***Explainability***: ProvExplainer

# Design : *Security-Aware Graph Structural Features*



Pattern **Size** – total number of nodes

Pattern **Diversity** – average "importance score" of node types

Pattern **Support** – average number of occurrences across *all* graphs in the dataset

Pattern **Coverage** – total number of occurrences across *anomalous* graphs in the dataset

Discriminant Pattern Score = $\alpha$ * **size** + $\beta$ * **diversity** + $\gamma$ * **support** + $\eta$ * **coverage**

***Explainability***: ProvExplainer

# Design : *Security-Aware Graph Structural Features*

Discriminant Pattern Score = $\alpha$ * **size** + $\beta$ * **diversity** + $\gamma$ * **support** + $\eta$ * **coverage**

Hyperparameter: $\alpha$=0.1, $\beta$=0.2, $\gamma$=0.4,and $\eta$= 0.3

Prioritize **support** and **coverage** as the primary differentiators between patterns followed by **diversity** and pattern **size**

Select the patterns that have the **highest** Discriminant Pattern Score

***Explainability***: ProvExplainer

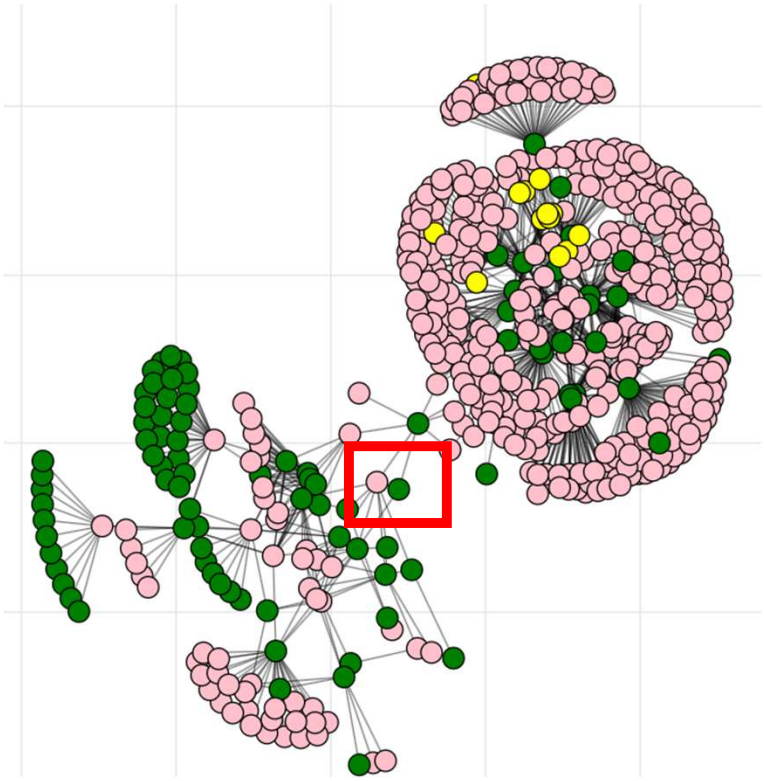
# Design : *Graph Structural Features*

| **Clustering Measures** |
| --- |
| Clustering Coefficient |
| Triangles |

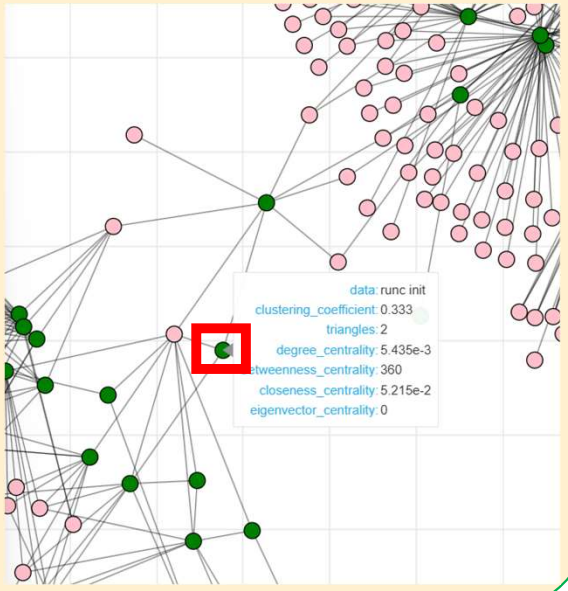
| **Centrality Measures** | |
| --- | --- |
| Degree | Closeness |
| Betweenness | Eigenvector |

***Explainability***: ProvExplainer

# Design : *Graph Structural Features*

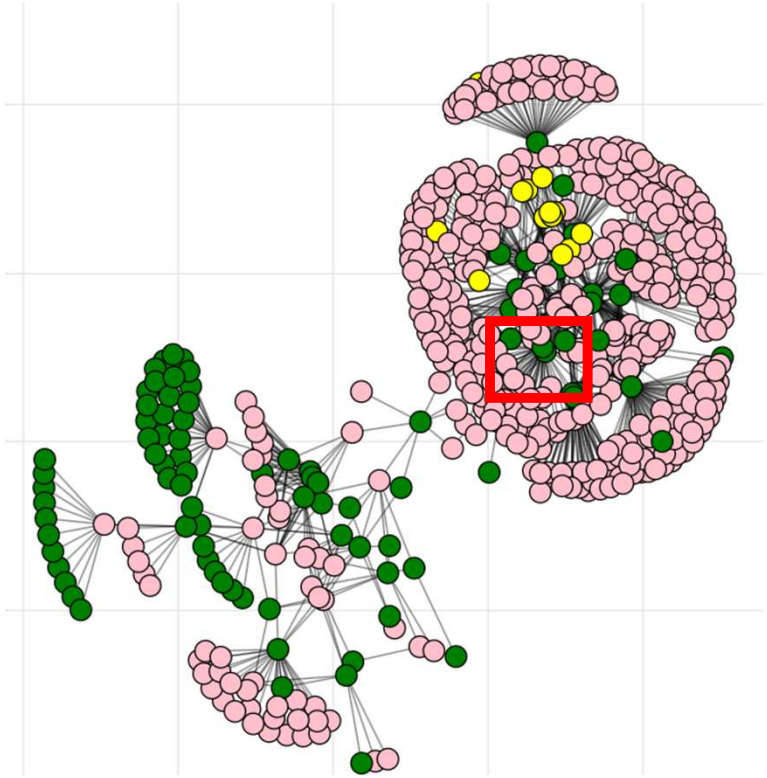


**Clustering Coefficient: 0.333**

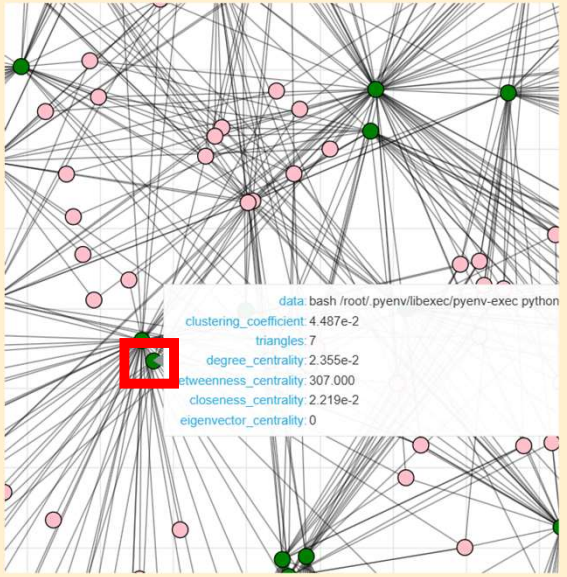


***Explainability***: ProvExplainer

# Design : *Graph Structural Features*

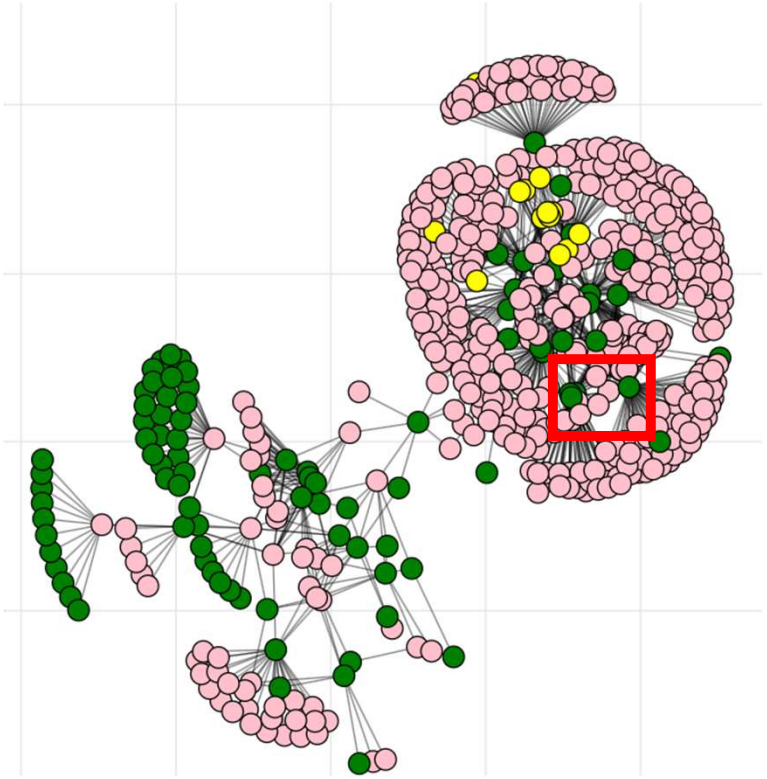


**Clustering Triangles: 7**

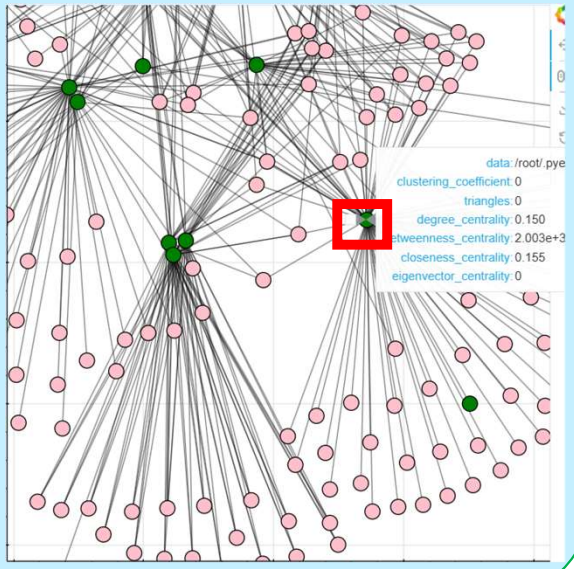


***Explainability***: ProvExplainer

# Design : *Graph Structural Features*

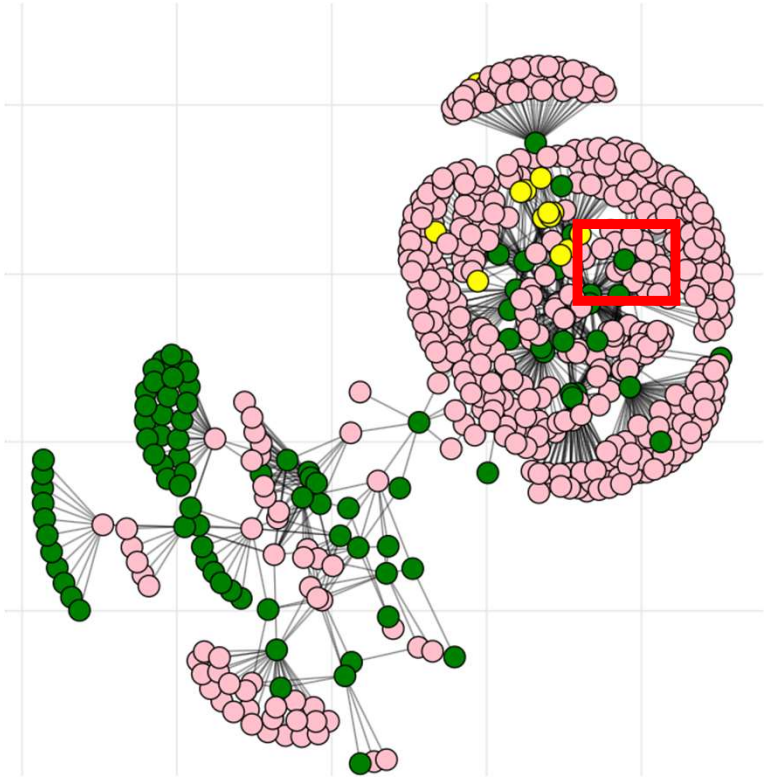


**Degree Centrality: 0.150**

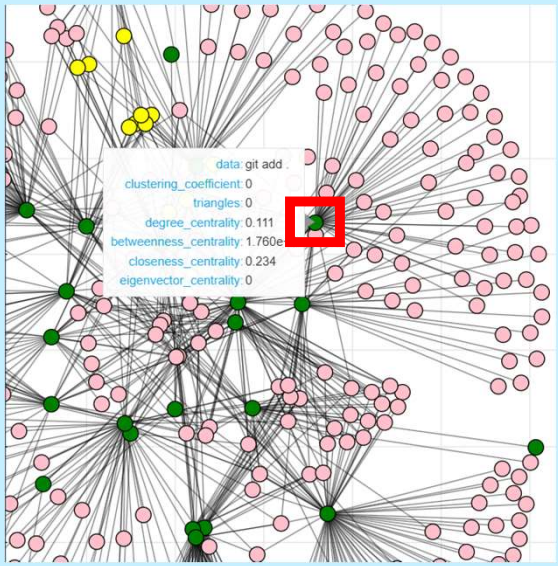


***Explainability***: ProvExplainer

# Design : *Graph Structural Features*

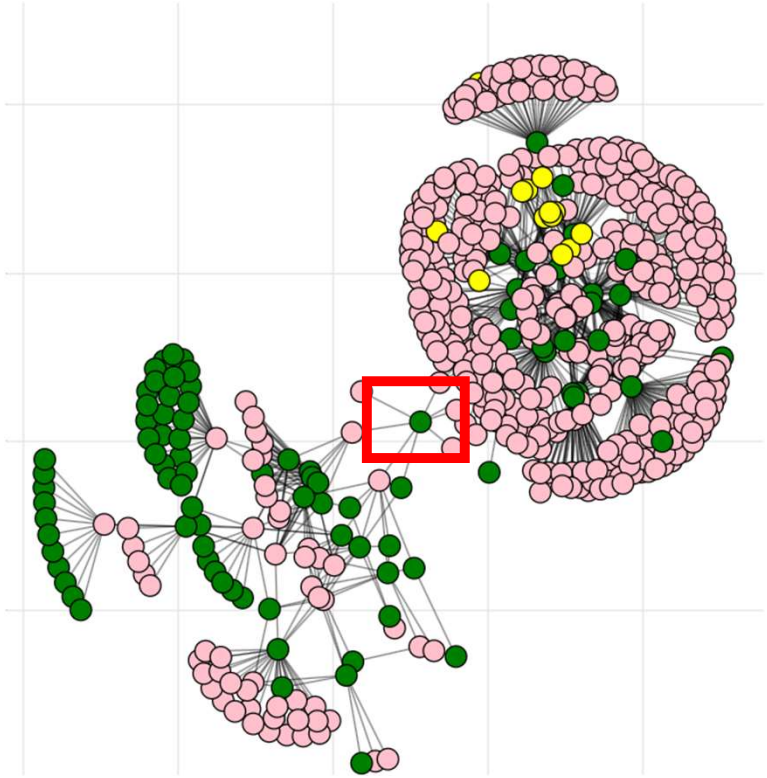


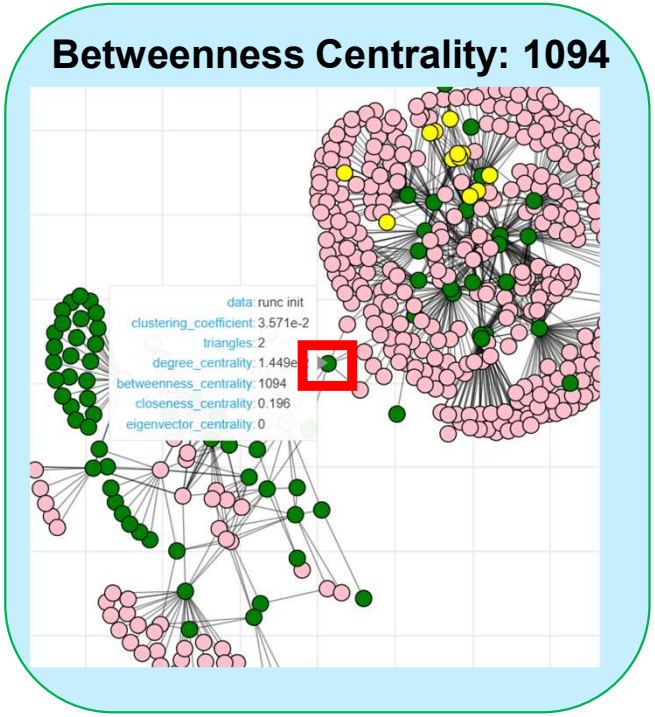
**Closeness Centrality: 0.234**



***Explainability***: ProvExplainer

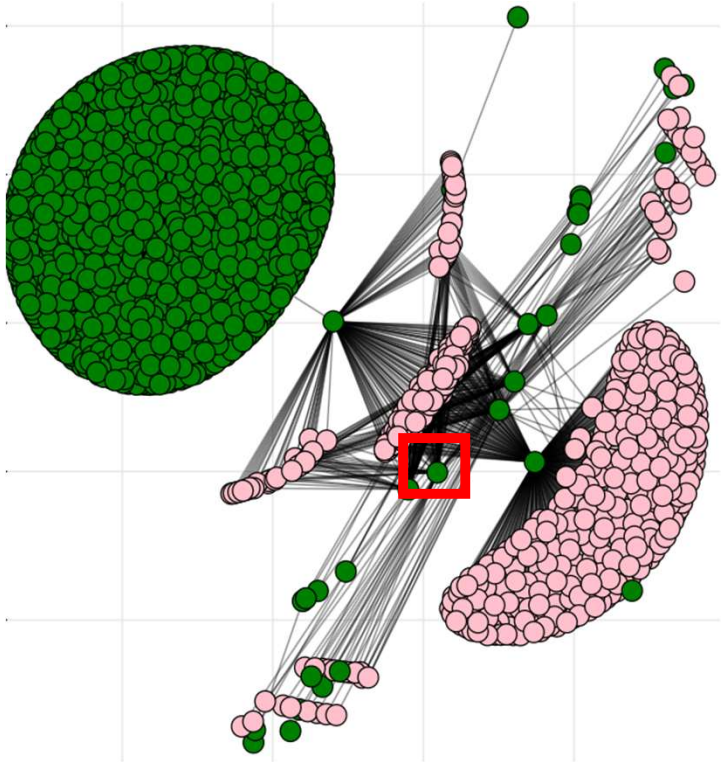
# Design : *Graph Structural Features*



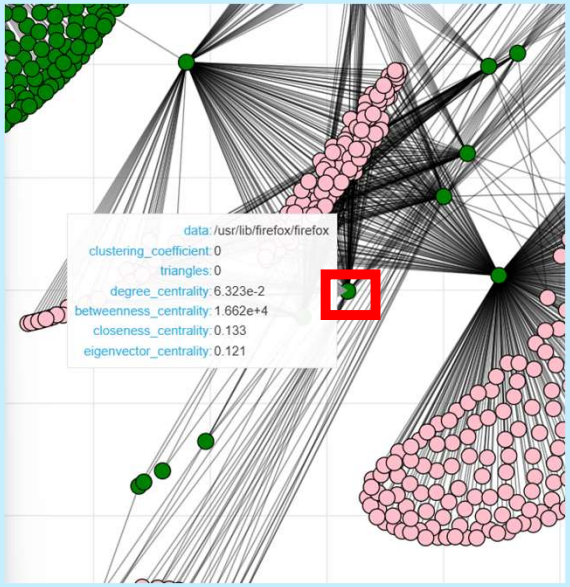


***Explainability***: ProvExplainer

# Design : *Graph Structural Features*



**Eigenvector Centrality: 0.121**



***Explainability***: ProvExplainer

# Design : *Using the Graph Structural Features*

Train a surrogate DT using the count of discriminant subgraph patterns and avg. of graph structural features

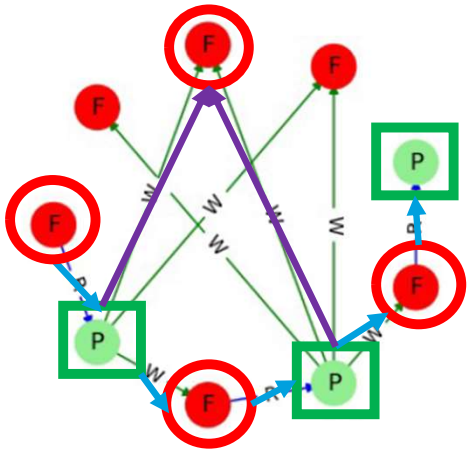
Use the surrogate DT to classify graphs

Extract node information from patterns along the decision path

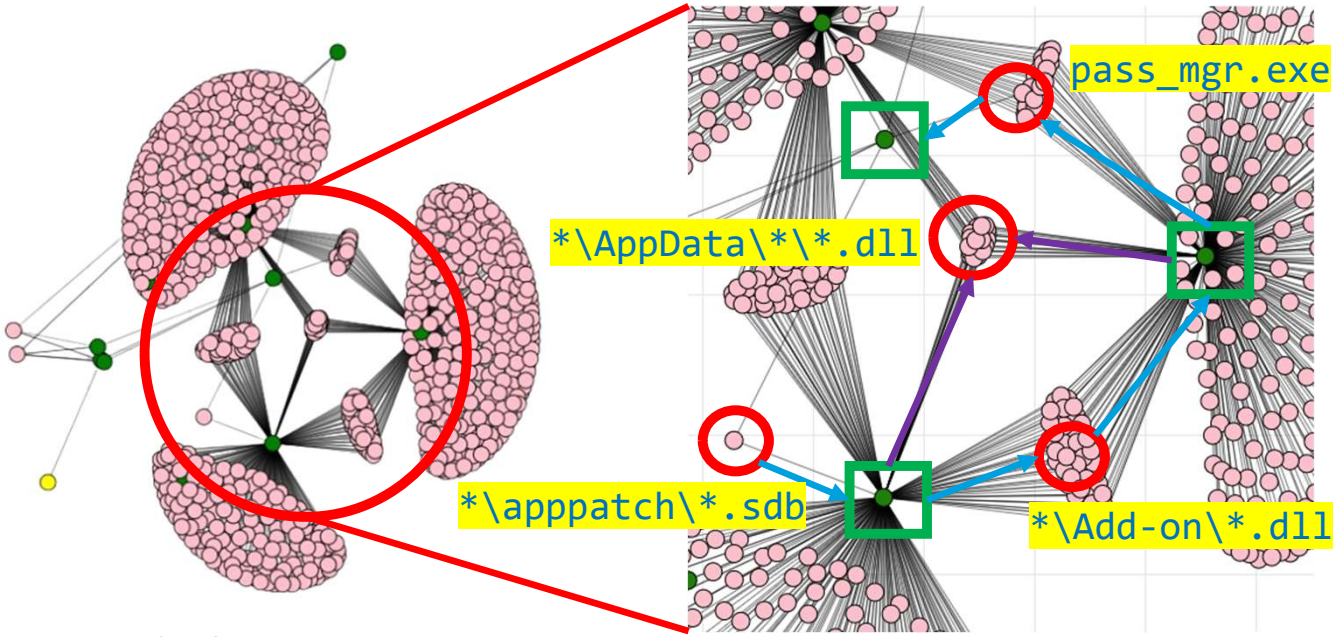
Patterns that appear high in the decision path are more important

***Explainability***: ProvExplainer

# Design : *Example of Explanations*




Example of patterns selected for explanation

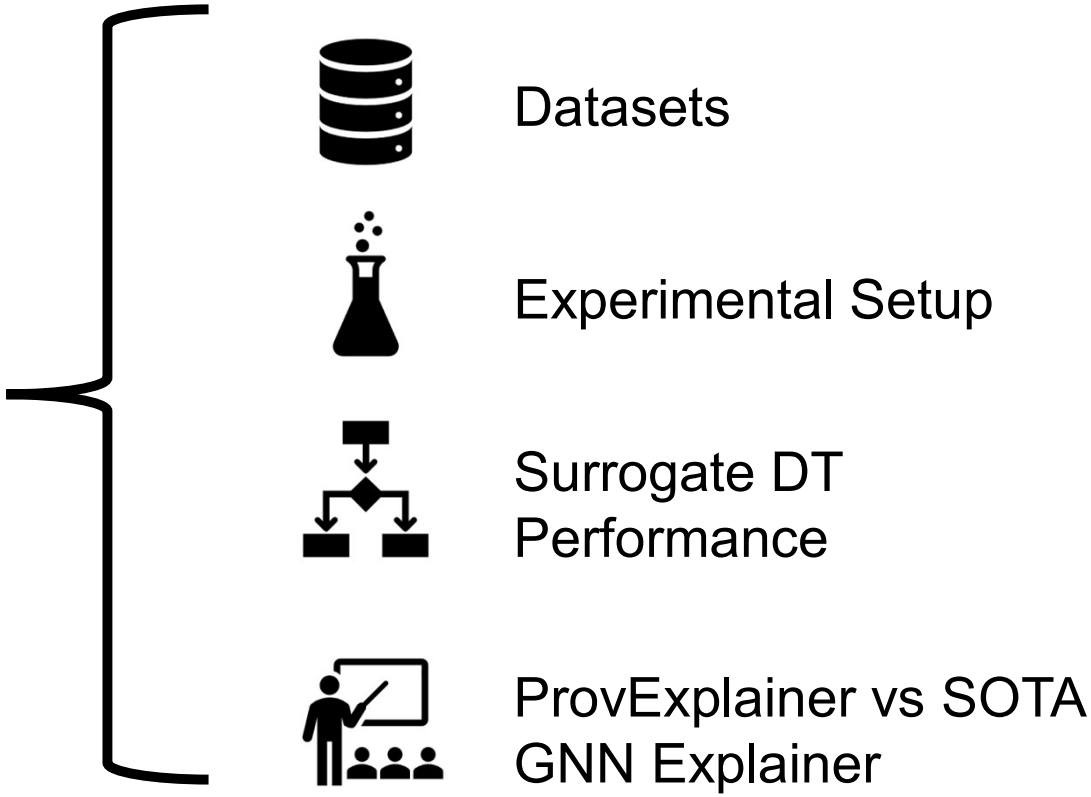



Exploits *firefox.exe* by downloading a malicious password manager, enabling it to read and store sensitive system files.

***Explainability***: ProvExplainer

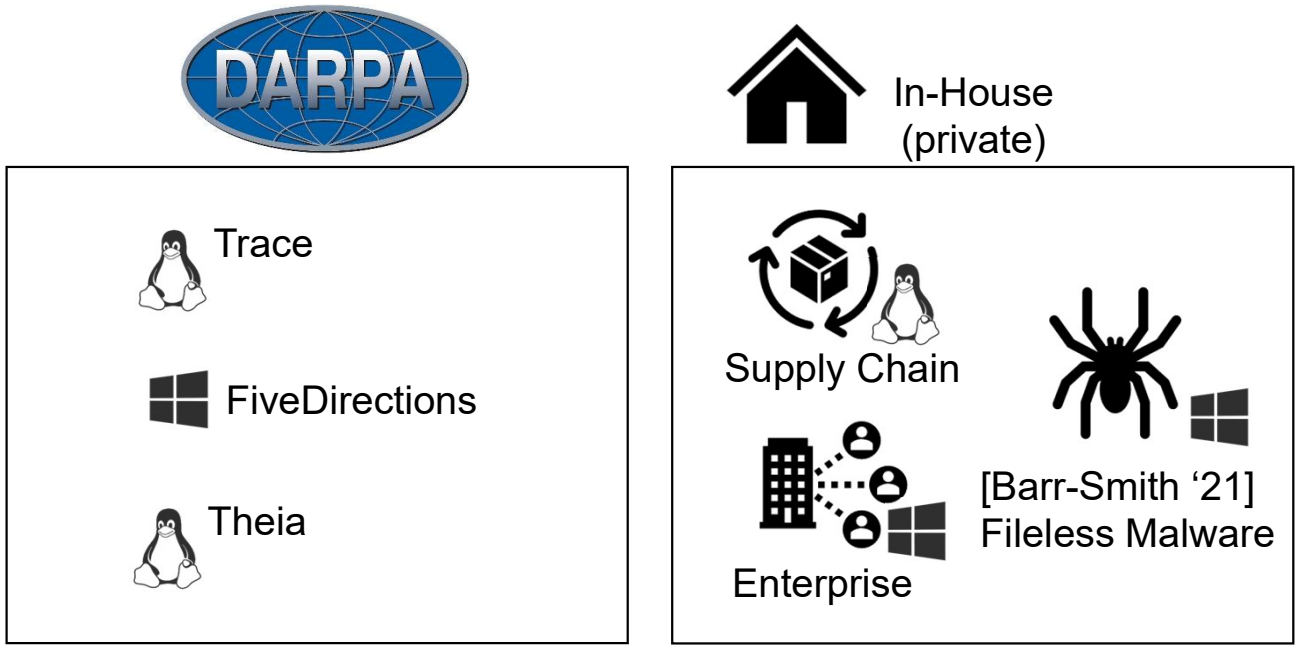
# Evaluation

**Evaluation**



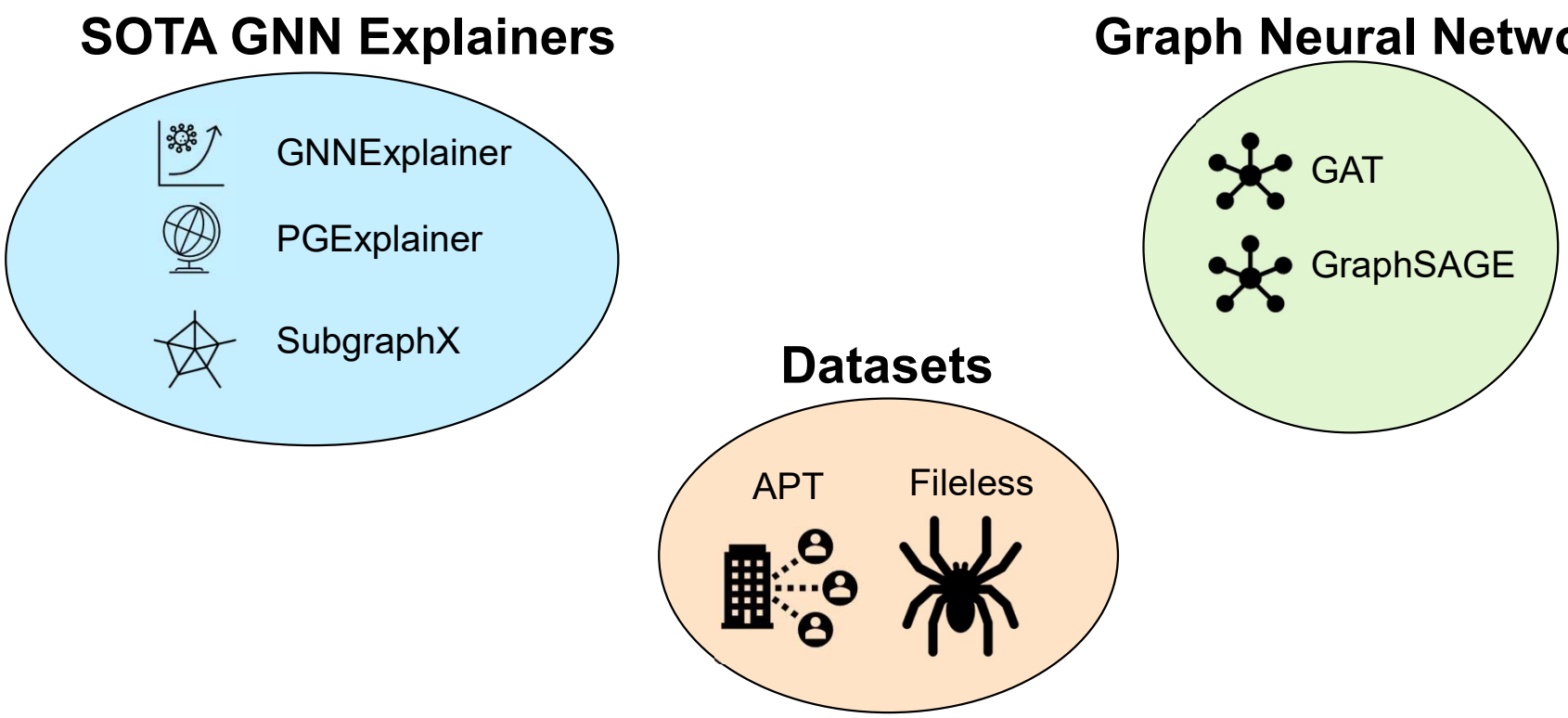
***Explainability***: ProvExplainer

# Evaluation : *Anomaly Detection Datasets*



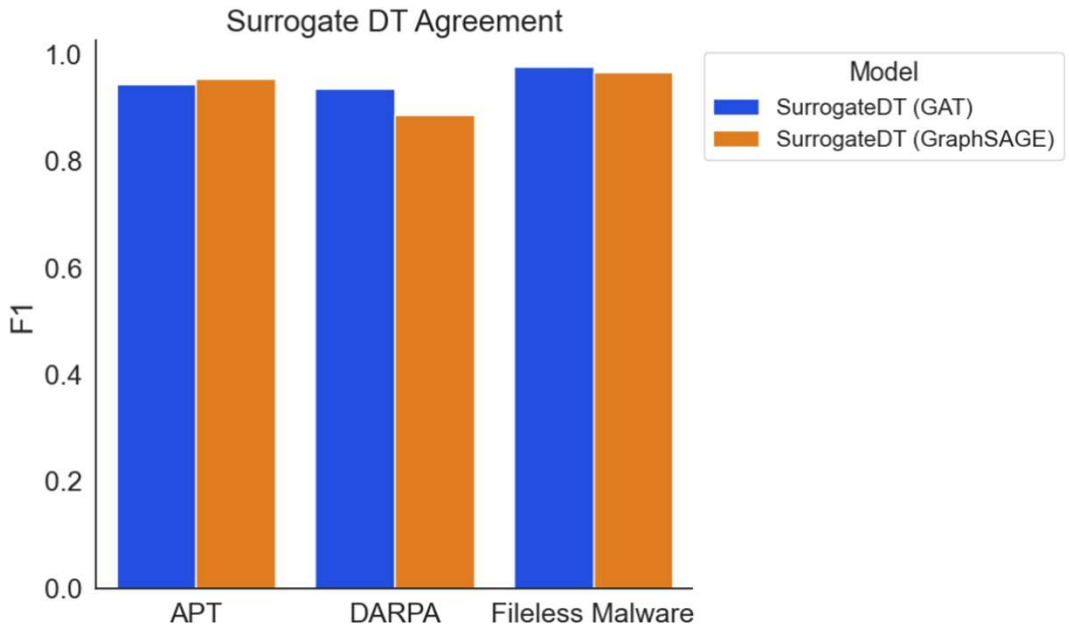
***Explainability***: ProvExplainer

# Evaluation : *Experimental Setup*



***Explainability***: ProvExplainer

# Evaluation : *Surrogate DT Performance*



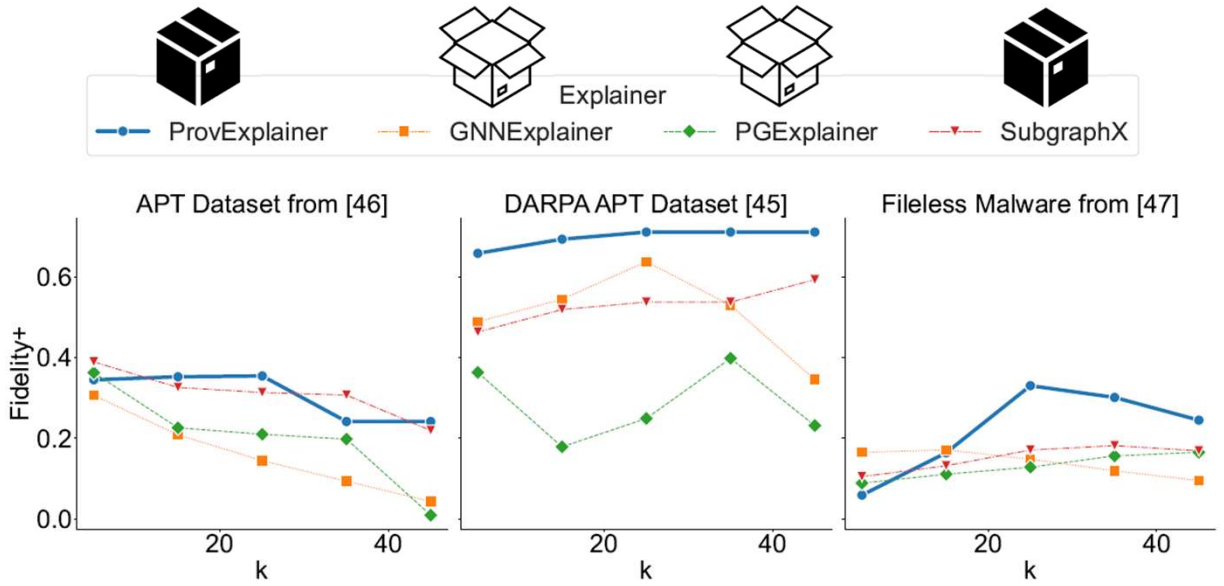

Surrogate DTs have high agreement with GNNs

***Explainability***: ProvExplainer

# Evaluation : *SOTA Comparison – Fidelity+*

**Fidelity+** = the difference in prediction after removing **k important** nodes from the graph

(**higher** is better)

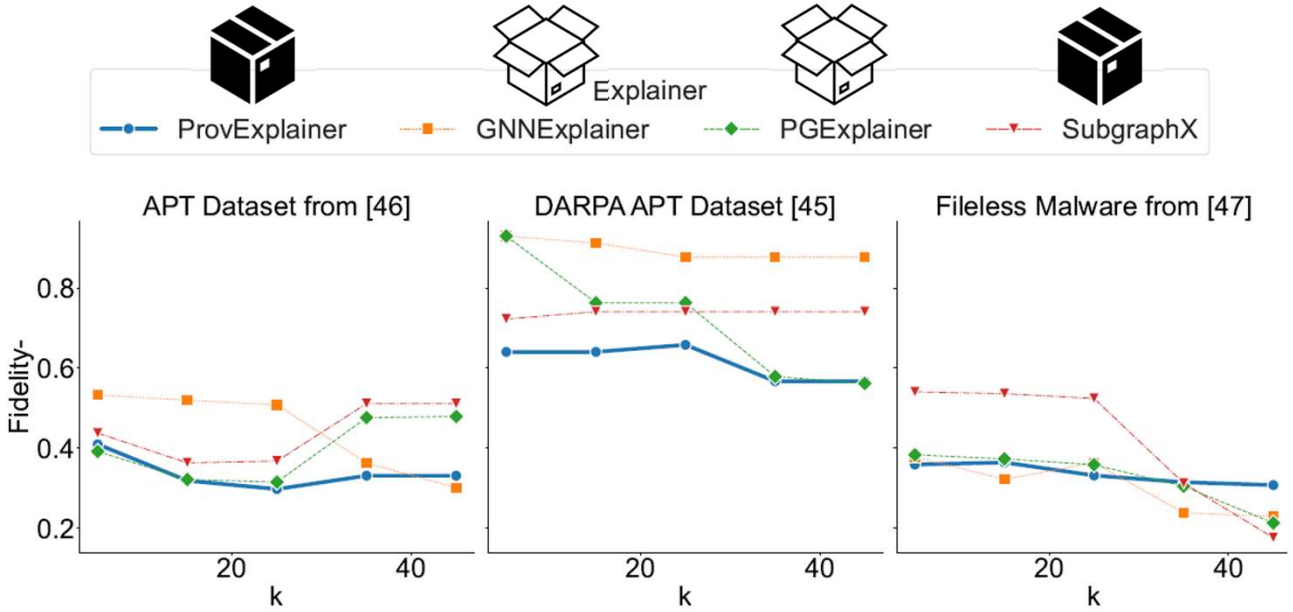


***Explainability***: ProvExplainer

# Evaluation : *SOTA Comparison – Fidelity-*

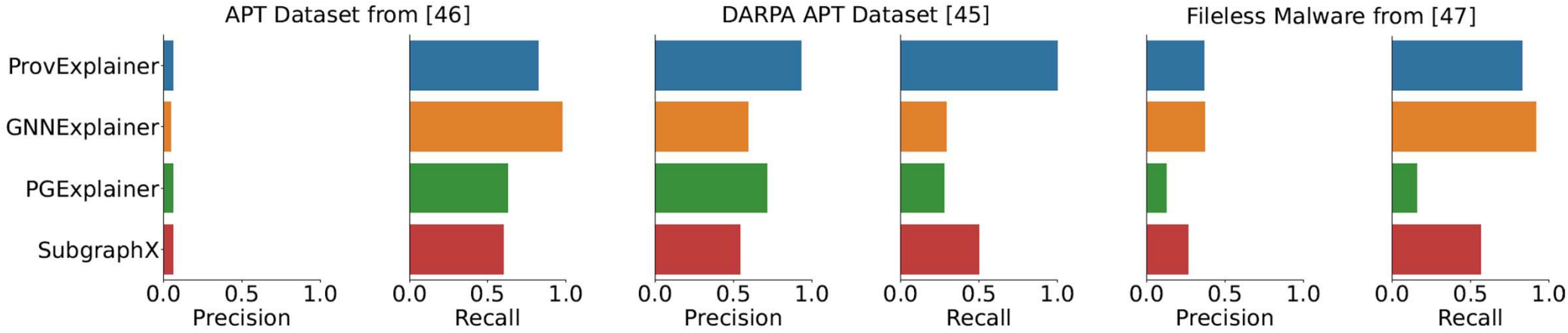
**Fidelity-** = the difference in prediction after removing **k unimportant** nodes from the graph

(**lower** is better)



***Explainability***: ProvExplainer

# Evaluation : *SOTA Comparison – Precision/Recall*




Competitive performance compared to SOTA

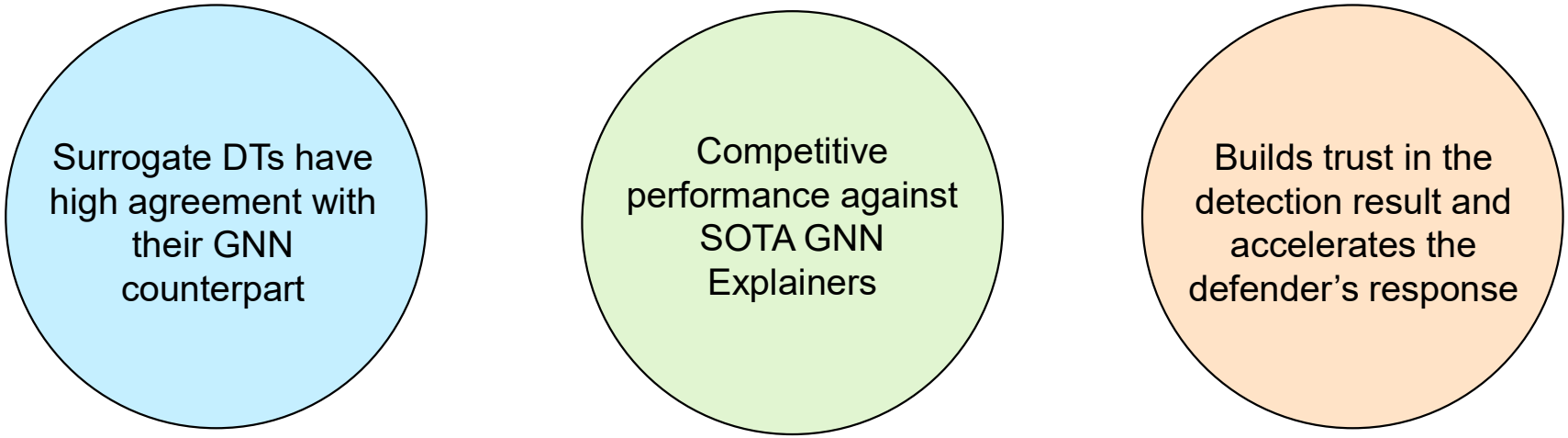
Extracts important system entities

Superior performance in DARPA dataset

***Explainability***: ProvExplainer

# Project Summary

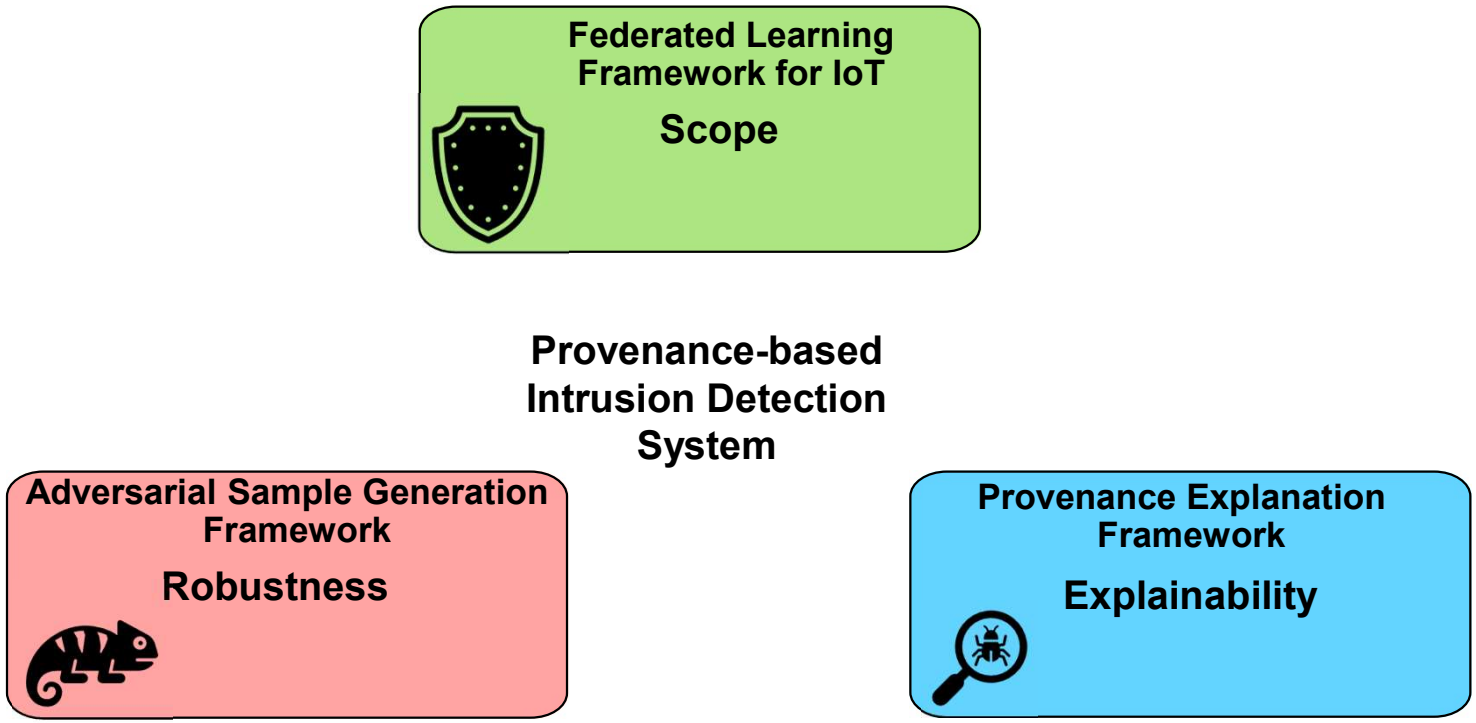
ProvExplainer generates **security-aware explanations**



# Agenda

1. Background
2. Motivation
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. **Research Contribution**
7. Future Work
8. Conclusion

# Research Contribution



# Research Contribution

1. ***Federated Learning Framework for IoT***. Proposed ProvIoT a provenance-based security detection approach in the context of IoT that counters stealthy attacks using federated learning and on-device detection. [ACNS '24]

2. ***Adversarial Sample Generation Framework***. Designed ProvNinja, a data driven approach to construct evasive attack vectors with minimal human oversight and realistic system constraints. [USENIX '23]

3. ***Ground-truth aware Explanation Generation Framework***. Created ProvExplainer to examine graph structural features that are interpretable and can be directly mapped to the system provenance problem space (e.g., system actions), making the explanations human understandable.

# Agenda

1. Background
2. Motivation
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. Research Contribution
7. **Future Work**
8. Conclusion

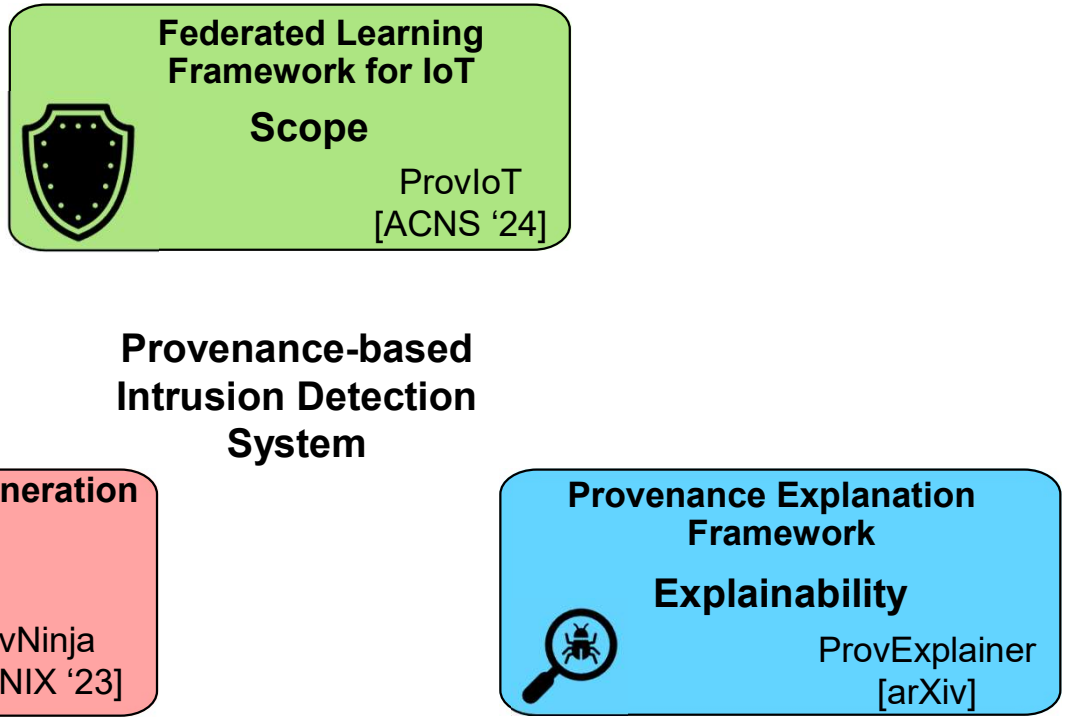
# Future Works

1. **Real-time detection in streaming audit logs for IoT domain**.
2. **LLM guided adversarial example generation framework**.
3. **LLM guided explanation framework.**

# Agenda

1. Background
2. Motivation
3. Scope: ProvIoT
4. Robustness: ProvNinja
5. Explainability: ProvExplainer
6. Research Contribution
7. Future Work
8. **Conclusion**

# IoT Integration, Adversarial Attacks, and Threat Explanations in Provenance-Based Intrusion Detection Systems



# Published Research Papers

[1] **Kunal Mukherjee**, Joshua Wiedemeier, Tianhao Wang, James Wei, Feng Chen, Muhyun Kim, Murat Kantarcioglu, and Kangkook Jee. "Evading Provenance-Based ML Detectors withAdversarial System Actions," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023,pp. 1199–1216.2.

[2] **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. "ProvIoT:Detecting Stealthy Attacks in IoT through Federated Edge-Cloud Security," in *International Conference on Applied Cryptography and Network Security*, Springer, 2024, pp. 241–268.

# Completed Manuscript

[3] **Kunal Mukherjee**, Joshua Wiedemeier, Tianhao Wang, Muhyun Kim, Feng Chen, Murat Kantarcioglu, and Kangkook Jee. "Interpreting gnn-based ids detections using provenance graph structural features," arXiv preprint arXiv:2306.00934, 2023.

# THANK YOU

Looking forward to your questions and comments

Scan the QR code to access the paper

**KUNAL MUKHERJEE**

**KXM180046@utdallas.edu**

**www.KUNMUKH.com**

ProvIoT
[ACNS ‘24]

ProvNinja
[USENIX ‘23]

ProvExplainer
[arXiv]

# References


Wagner & Soto ’02 - Wagner, David, and Paolo Soto. "Mimicry attacks on host-based intrusion detection systems." *Proceedings of the 9th ACM Conference on Computer and Communications Security*. 2002.

Tan & Maxion ’03 - Tan, Kymie MC, and Roy A. Maxion. "Determining the operational limits of an anomaly-based intrusion detector." *IEEE Journal on selected areas in communications* 21.1 (2003): 96-110.

Velickovic ’17 - Veličković, Petar, et al. "Graph attention networks." *arXiv preprint arXiv:1710.10903* (2017).

Hassan ’19 - Hassan, Wajih Ul, et al. "Nodoze: Combatting threat alert fatigue with automated provenance triage." *network and distributed systems security symposium*. 2019.

Dang ’19 - Dang, Fan, et al. "Understanding fileless attacks on linux-based iot devices with honeycloud." *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 2019.

Ying ’19 - Ying, Zhitao, et al. "Gnnexplainer: Generating explanations for graph neural networks." *Advances in neural information processing systems* 32 (2019).

Wang ’20 - Wang, Qi, et al. "You Are What You Do: Hunting Stealthy Malware via Data Provenance Analysis." *NDSS*. 2020.

Han ‘21 - Han, Xueyuan, et al. "{SIGL}: Securing Software Installations Through Deep Graph Learning." *30th USENIX Security Symposium (USENIX Security 21)*. 2021.

Barr-Smith ’21 - Barr-Smith, Frederick, et al. "Survivalism: Systematic analysis of windows malware living-off-the-land." *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021.

Zeng ‘22 - Zeng, Jun, et al. "Shadewatcher: Recommendation-guided cyber threat analysis using system audit records." *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022.

Goyal ’23 - Goyal, Akul, et al. "Sometimes, You Aren’t What You Do: Mimicry Attacks against Provenance Graph Host Intrusion Detection Systems." *30th ISOC Network and Distributed System Security Symposium (NDSS’23), San Diego, CA, USA*. 2023.

Colonial – Easterly, Jen “The Attack on Colonial Pipeline: What We’ve Learned & What We’ve Done over the Past Two Years: CISA.” Cybersecurity and Infrastructure Security Agency CISA, 8 Aug. 2023, www.cisa.gov/news-events/news/attack-colonial-pipeline-what-weve-learned-what-weve-done-over-past-two-years.

SolarWinds - “The Solarwinds Cyber-Attack: What You Need to Know.” *CIS*, 9 Nov. 2021, www.cisecurity.org/solarwinds.

# 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.