Detecting Data Exfiltration from BYOD and IoT Devices in Enterprise Networks Using Supervised Learning on Network Flow Metadata

Kai Chong Chio | C00315932 | MSc Cybersecurity, Privacy and Trust

Can edge flow metadata reveal exfiltration without payload inspection?

Ollscoil Teicneolaíochta an Oirdheiscirt
South East Technological University

1 Background

Why this topic matters

- BYOD and IoT add personal laptops, mobiles, cameras and sensors to enterprise networks.
- These endpoints widen the attack surface and are not managed as consistently as corporate hosts [2][7].
- Flow metadata is lower-friction to monitor than payload inspection and is more privacy-aware [1][3].

BYOD
IoT
Exfiltration risk

→ Edge monitoring needs metadata-driven detection

Normal outbound traffic can mask small or bursty uploads to external destinations.

2 Motivation and gap

Current challenge

- Enterprise teams may only see router, firewall or NetFlow records at the edge.
- Mixed-device behavior makes benign uploads harder to separate from true exfiltration.

Research gap

- Most prior work studies managed enterprise hosts, a single protocol, or one attack family.
- Less evidence exists for mixed BYOD and IoT environments using flow metadata only [1][3][6].

This study positions

Mixed BYOD + IoT	Host and flow views
Edge flow metadata only	Controlled exfiltration strategies

3 Research focus and contribution

Research questions

1. To what extent can supervised machine learning models trained on network flow metadata distinguish data exfiltration traffic from benign traffic in a mixed BYOD/IoT enterprise environment?
2. What flow-level and Host-level feature characteristics were the most significant contributors to successfully detecting Exfiltration in this Model?
3. How do various Exfiltration Strategies Affect Detection Model Effectiveness and False Positives?
4. What practical and ethical considerations arise when deploying flow-based exfiltration detection focused on BYOD and IoT devices in real organisations?

Contribution in one view

Mixed enterprise-style testbed | Host-level bins | Flow-level records | Supervised learning

Exfiltration strategies tested

Bulk transfer | IoT-based exfiltration | Slow-drip transfer | Hybrid

The design compares both data granularity and attack style in one controlled lab.

4 Use case and methodology

Experimental design

A controlled virtual enterprise testbed generates benign and exfiltration traffic. The edge gateway exports network flow metadata, which is transformed into host-level and flow-level features for supervised learning.

BYOD (Windows 11)
Admin (Ubuntu)
IoT Sensor (Alpine Linux)
IoT camera (Ubuntu)

→ Edge gateway NetFlow export

→ **Feature engineering**
Flow-level: bytes, packets, port, flags
Host-level: out ratios, counts, timing

→ **Model training and evaluation**
Logistic Regression | Random Forest | Gradient Boosting

Metrics:
- *precision*
- *Recall*
- *F1*
- *ROC-AUC*
- *PR-AUC*
- *FPR*

Scenario set used in the lab

Benign profiles
- B0 baseline state (included B3 sensor VM telemetry and B4 camera VM traffic)
- B1 browsing + tiny uploads
- B2 burst downloads
- B5 infrastructure noise
- B6 high-throughput download
- B7 streaming
- B8 admin + camera access

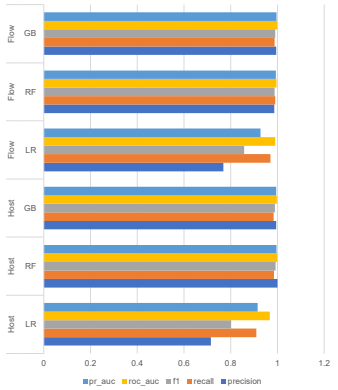
Exfiltration profiles
- E1 bulk upload from BYOD
- E2 slow-drip parallel exfiltration
- E3 IoT mixed TCP + UDP leakage
- E4 hidden exfiltration in camera-like traffic
- E5 hybrid exploratory bursts on common ports

Data recorded so far

36	370	12	205	165	274
batches	scenario runs	scenario types	benign labels	exfil labels	runs passed full integrity checks

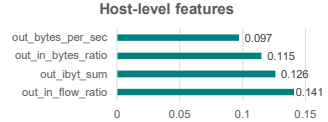
6 Early results, expected outcome, next steps

Model performance

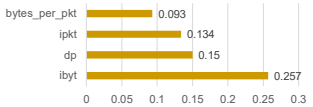


Key signal features

Host-level features



Flow-level features

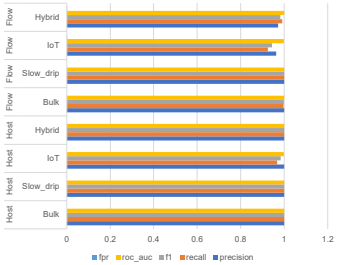


Flow-level Gradient Boosting

	Predicted Benign	Predicted Exfiltration
Actual Benign	89.09%	0.06%
Actual Exfiltration	0.12%	10.73%

F1 0.9915 | Precision 0.9943 | Recall 0.9887 | ROC-AUC 0.9997 | PR-AUC 0.997

Strategy-specific performance



Expected outcome and next steps

- Preliminary lab results show strong separation using host and flow metadata.
- IoT-style leakage is the hardest case at flow level and needs wider device diversity.
- Next steps include higher false-positive stress tests and stronger privacy and ethics analysis.

6 References

[1] Sabir et al., ACM Comput. Surv., 2021.
[2] NCSC, Bring Your Own Device, 2023.
[3] Willems et al., Electronics, 2023..
[4] Marques et al., Digit. Commun. Netw., 2020
[5] Zhan et al., Compute. Networks, 2022
[6] Rahman et al., Survey on IDS in IoT, 2025.
[7] Palo Alto Networks, Data Exfiltration, 2025.
[8] Ivanti / TechRadar Pro, BYOD and edge devices, 2025.