

# Internet of Things Threat Detection Using Process Conformance and Machine Learning Classification

**ASPAI26, Pohang, South Korea**

**Adam Burke<sup>1,2</sup>, Mashal Ashraf<sup>1</sup>, Marimbay Kadirov<sup>1</sup> and Andrzej Janusz<sup>1,2</sup>**

(1) School of Information Systems, QUT

(2) Centre for Data Science, QUT



Technology > Gadgets

# Mum's nightmare as baby monitor gets hacked while son sleeps

A parent's nightmare became reality after a creep hacked into a baby monitor while a little boy was sleeping.



Brooke Rolfe

@rolfeb 2 min read November 2, 2023 - 3:31PM news.com.au

6 Comments

Advertisement



TEQSA Provider ID PRV12079 Australian University | CRICOS No.00213J



## Mirai-Based xlabs\_v1 Botnet Exploits ADB to Hijack IoT Devices for DDoS Attacks

Ravie Lakshmanan May 06, 2026

IoT Security / Malware



Cybersecurity researchers have exposed a new [Mirai](#)-derived botnet that self-identifies as **xlabs\_v1** and targets internet-exposed devices running Android Debug Bridge (ADB) to enlist them in a network capable of carrying out distributed denial-of-service (DDoS) attacks.

Hunt.io, which [detailed](#) the malware, said it made the discovery after identifying an exposed



# Attack Detection

- Many vulnerable IoT devices
- High volume
- Attacks follow attack *patterns*
- Attacks inherently *sequential* in time
  - Eg, Mirai malware
    1. Scans for vulnerable devices
    2. Infects vulnerable devices
    3. Stays dormant
    4. When launching attack, instructs device to perform Denial-of-Service on target



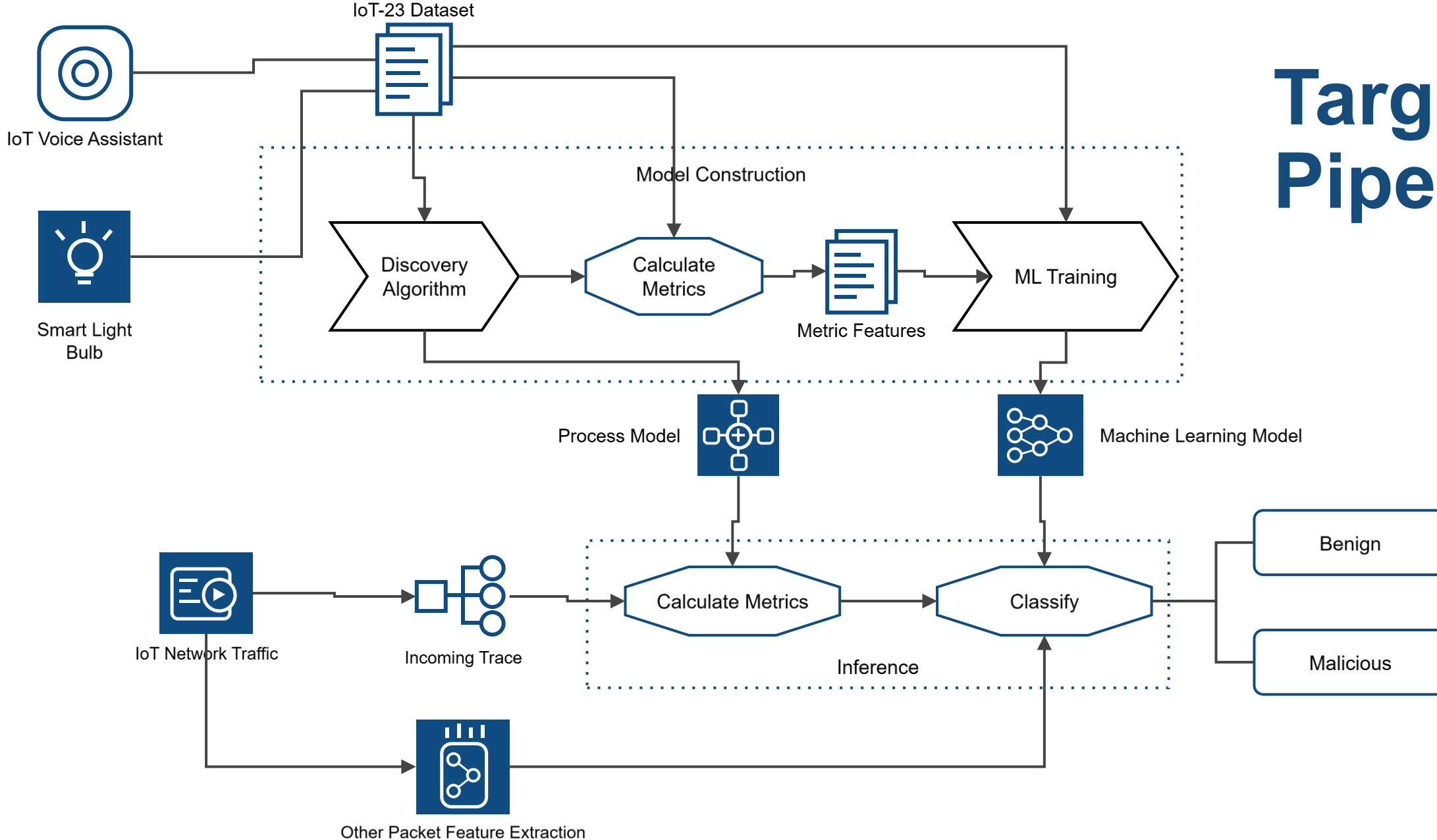
# Problem and Solution Space

- High Volume
- Time Sensitive
- Sequential Modal Data
- Pattern Recognition



- Automation
- Meaningful Alerting
- Process Mining
- Machine Learning

# Target Pipeline



# Training and Evaluation Data

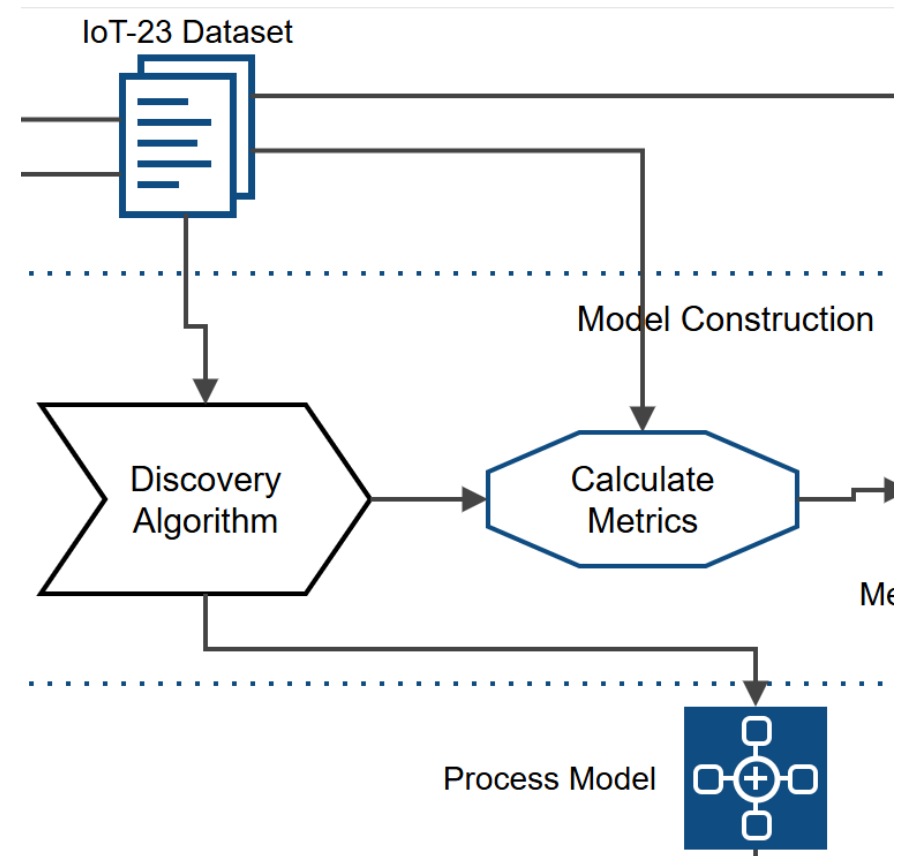
- IoT-23 dataset
- Recorded network traffic from 23 IoT devices in laboratory setting
- Several infected with malware
- TCP, UDP and control protocols
- Traffic labelled at connection level as benign or malicious

## IoT-23 Dataset

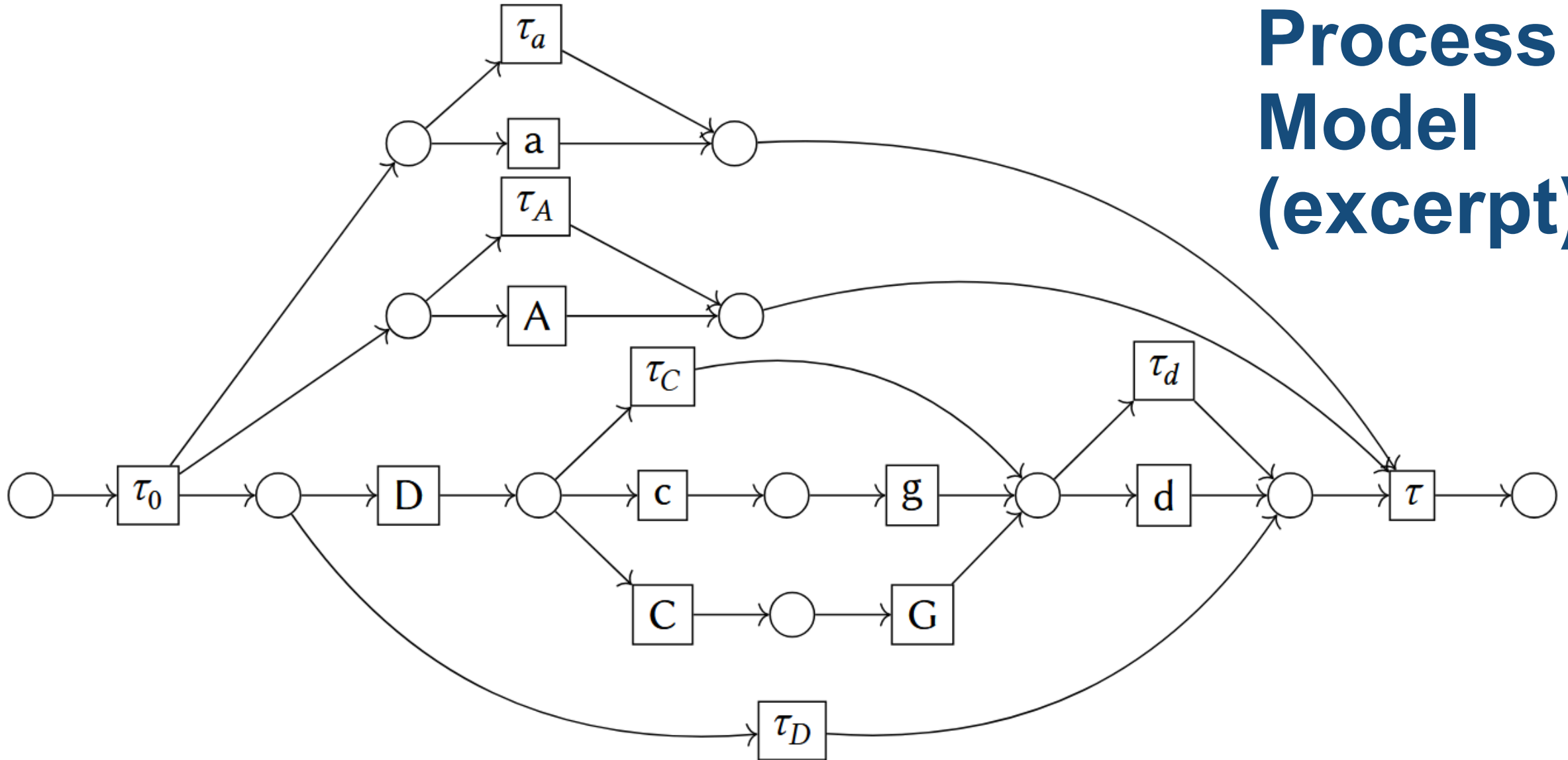


# Process Mining Component

- Case notion: network connection
- Activity notion: packet type + direction
  - Client-to-server or server-to-client
- Model discovery (Inductive Miner + Alignment Estimator)
- Conformance Metrics relative to single trace
  - Chi-Squared
  - Earth Movers Stochastic Conformance
  - Entropic Relevance
  - Trace Probability

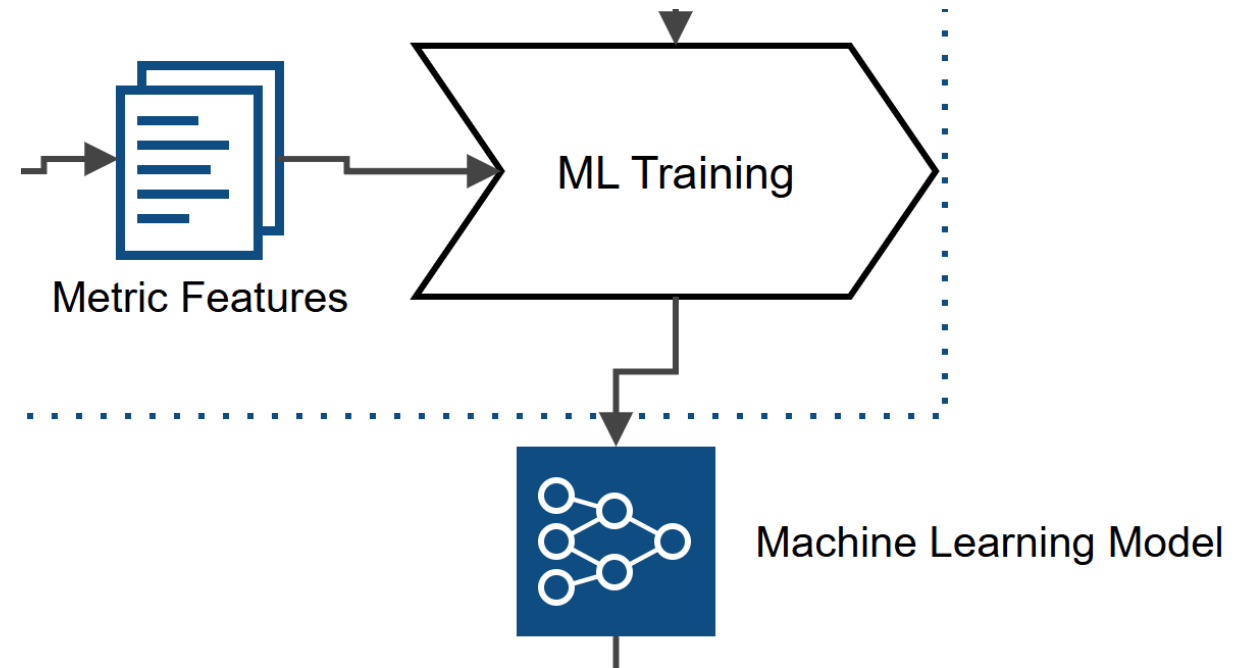


# Process Model (excerpt)

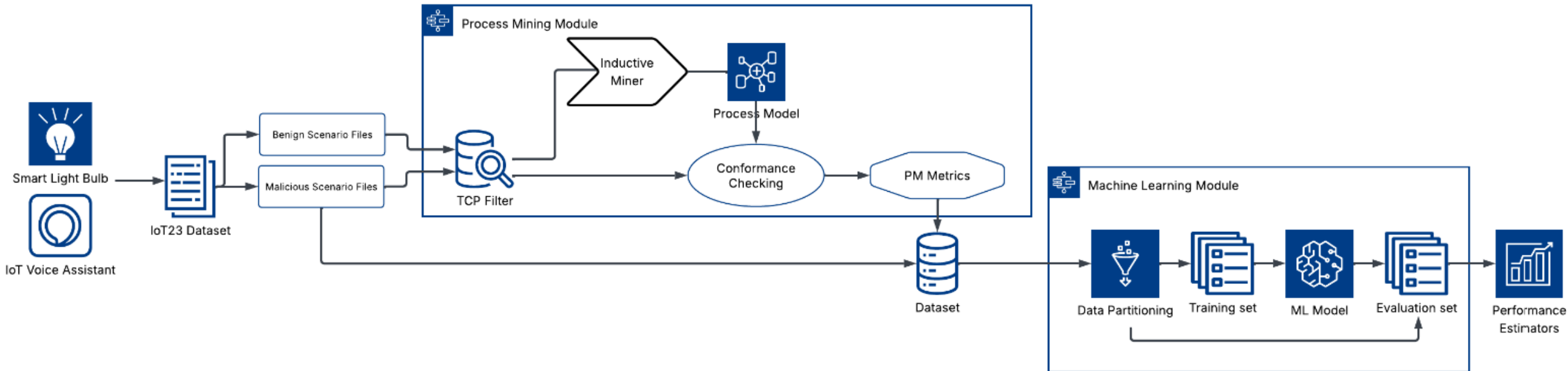


# Machine Learning Training

- Machine Learning inputs
  - Data features
    - Source port, destination port, source bytes, duration, ...
  - Conformance metric features
    - Chi-Squared
    - Earth Movers Stochastic Conformance
    - Entropic Relevance
    - Trace Probability



# Experiment Design



- Training (i.e., process discovery) on benign files only
- ML model: Random Forest
- Leave One File Out (LOFO) evaluation
- Use subset of 15/23 files (scaling and benign partition) – **IoT-15-Light**

# Results



| Feature            | Feature Importance |
|--------------------|--------------------|
| Earth Movers       | 0.3651 ± 0.0522    |
| Chi Square         | 0.3567 ± 0.0472    |
| Entropic Relevance | 0.2050 ± 0.0136    |
| Trace Probability  | 0.0003 ± 0.0004    |

|                                        | Accuracy               | Precision              | Recall                 | F1                     |
|----------------------------------------|------------------------|------------------------|------------------------|------------------------|
| LOFO Benchmark IoT-15-Light            | 0.7748 ± 0.3004        | 0.6459 ± 0.4222        | 0.8481 ± 0.2896        | 0.6401 ± 0.3876        |
| <b>LOFO w PM Features IoT-15-Light</b> | <b>0.9559 ± 0.1232</b> | <b>0.8973 ± 0.2712</b> | <b>0.9993 ± 0.0021</b> | <b>0.9067 ± 0.2720</b> |

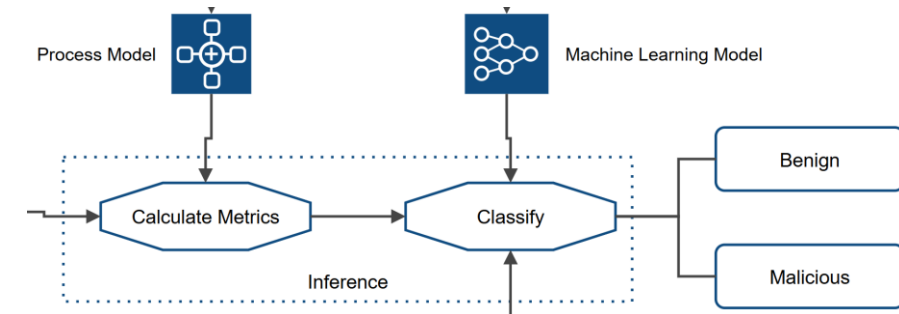
# Summary

## Conclusion

- Cyberattacks on IoT devices inherently sequential
- Process mining can be used to communicate this sequential information to ML models
- Improved benchmark performance
  - For subset of attack types

## Extension

- Full IoT-23 dataset, other datasets
- Expanded benchmarks
- Model discovery variation
- ML model variation
- Hardware footprint



# Questions welcome

Acknowledgement:  
We would like to thank the **QUT Centre for Data Science**  
for supporting this research through a travel grant.

# Related literature (selected)

1. Garcia, A. Parmisano, M. J. Erquiaga, lot-23: A labeled dataset with malicious and benign IoT net-work traffic, 2020. doi:10.5281/zenodo.4743746
2. A. Burke, S. J. J. Leemans, M. T. Wynn, Stochastic Process Discovery by Weight Estimation, in: Process Mining Workshops, ICPM 2020
3. S. J. J. Leemans, W. M. P. van der Aalst, T. Brockhoff, A. Polyvyanyy, Stochastic process mining: Earth movers' stochastic conformance, Information Systems 102 (2021) 101724.
4. Li, S. J. J. Leemans, A. Polyvyanyy, Applying statistical distances for stochastic conformance checking (2026). Under review.
5. Alkhamash, A. Polyvyanyy, A. Moffat, L. García-Bañuelos, Entropic relevance: A mechanism for measuring stochastic process models discovered from event data, Information Systems 107(2022)