LIRIS [1] INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON [1] orange [2]

07/02/2020

Smart device monitoring and behavior analysis

Ph.D. Student : Naji NAJARI[1,2]

Supervisors

Christophe GARCIA[1]

Stefan DUFFNER[1]

Grégoire LEFEBVRE[2]

Samuel BERLEMONT[2]

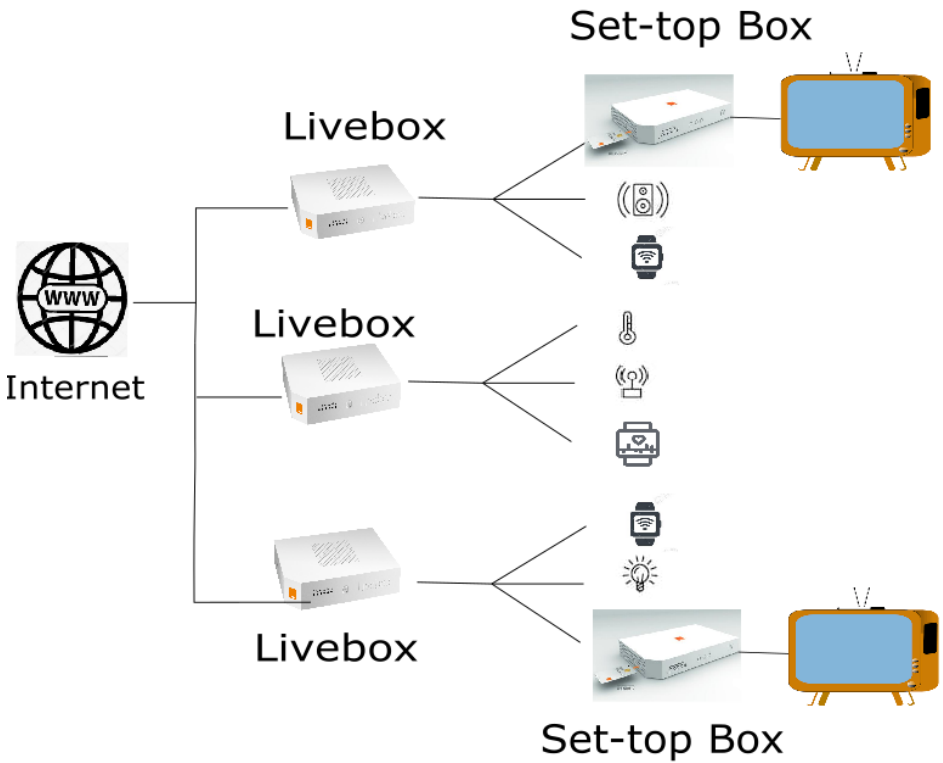
orange

Winter School

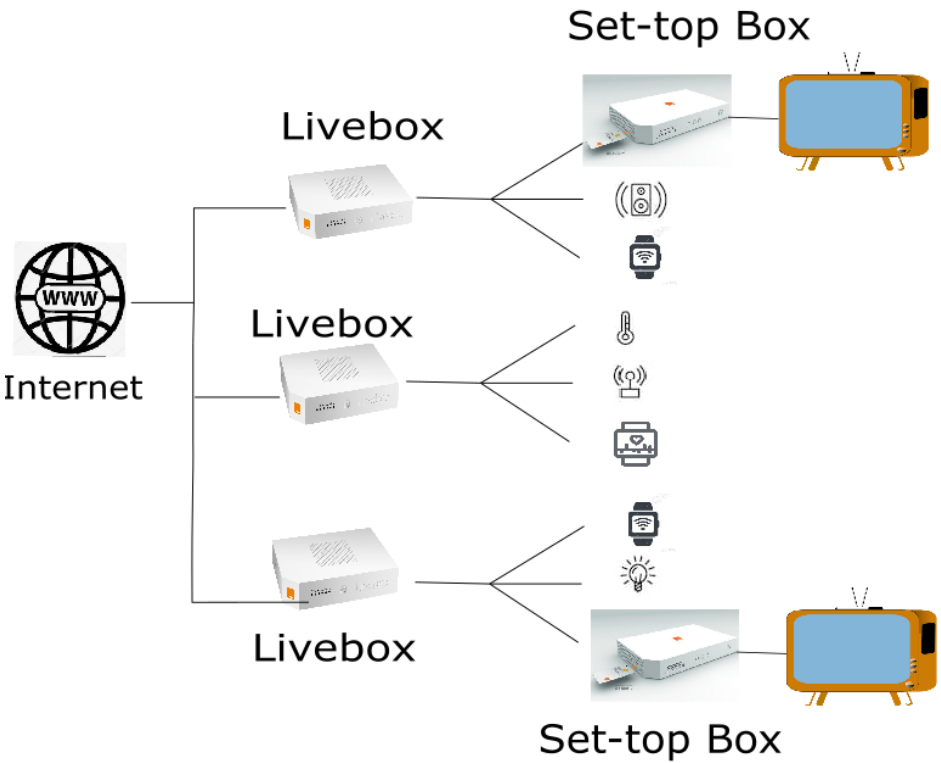
Plan

1. Context and Challenges
2. Problem Statement
3. Related Work
4. Contribution
5. Conclusion & Future Work

Context



Context



Device Management (DM)

Remote provisioning, maintenance, assistance and tracking of connected device in a secure environment.

DM Server — DM Protocol — DM Client

Diagnostics and monitoring

Challenges

Scalability

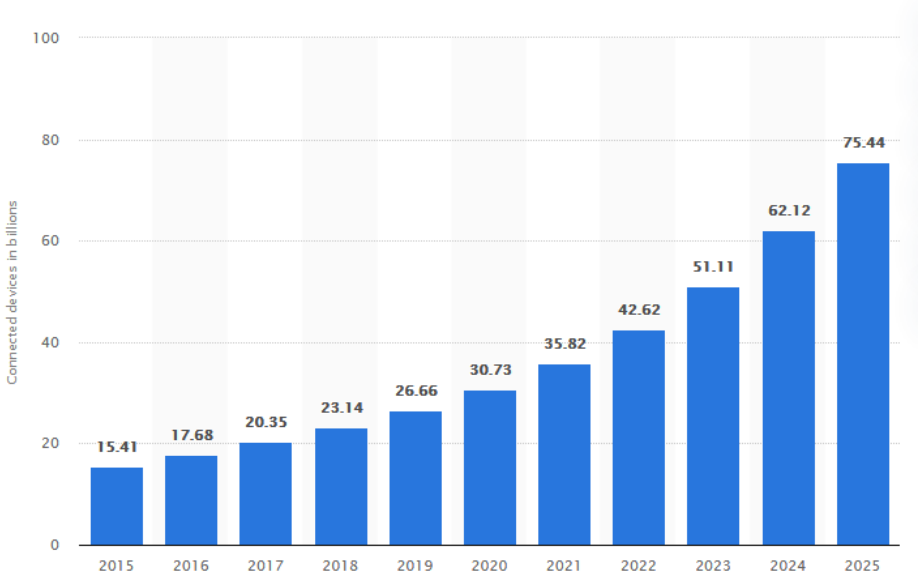


Figure 1 : number of connected devices worldwide 2015-2025 [1]

1 https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

2 https://wikibon.com/the-vital-role-of-edge-computing-in-the-internet-of-things/

Challenges

Scalability

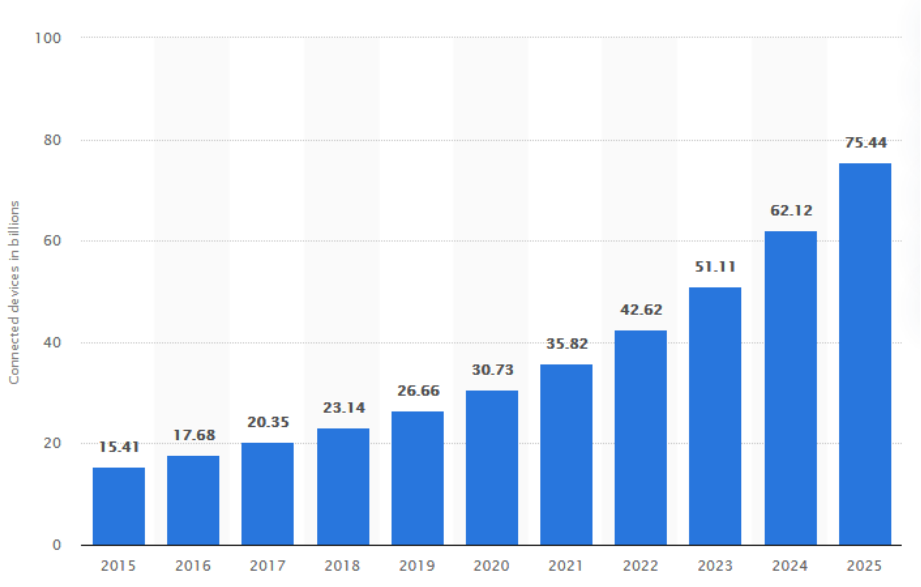
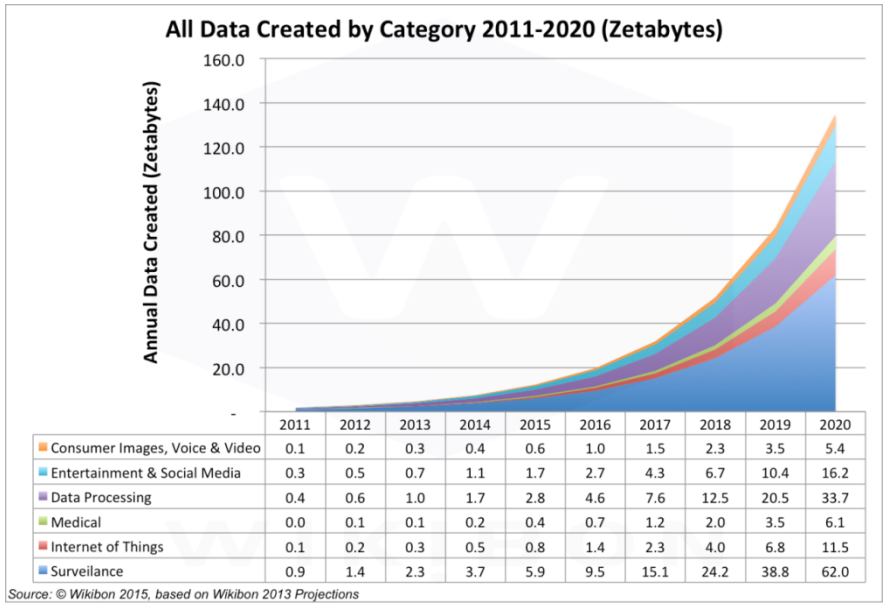


Figure 1 : number of connected devices worldwide 2015-2025 [1]

Data



	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Consumer Images, Voice & Video	0.1	0.2	0.3	0.4	0.6	1.0	1.5	2.3	3.5	5.4
Entertainment & Social Media	0.3	0.5	0.7	1.1	1.7	2.7	4.3	6.7	10.4	16.2
Data Processing	0.4	0.6	1.0	1.7	2.8	4.6	7.6	12.5	20.5	33.7
Medical	0.0	0.1	0.1	0.2	0.4	0.7	1.2	2.0	3.5	6.1
Internet of Things	0.1	0.2	0.3	0.5	0.8	1.4	2.3	4.0	6.8	11.5
Surveilance	0.9	1.4	2.3	3.7	5.9	9.5	15.1	24.2	38.8	62.0

Source: © Wikibon 2015, based on Wikibon 2013 Projections

Figure 2 : Data created between 2011-2020 [2]

[1] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[2] https://wikibon.com/the-vital-role-of-edge-computing-in-the-internet-of-things/

Challenges

Heterogeneity

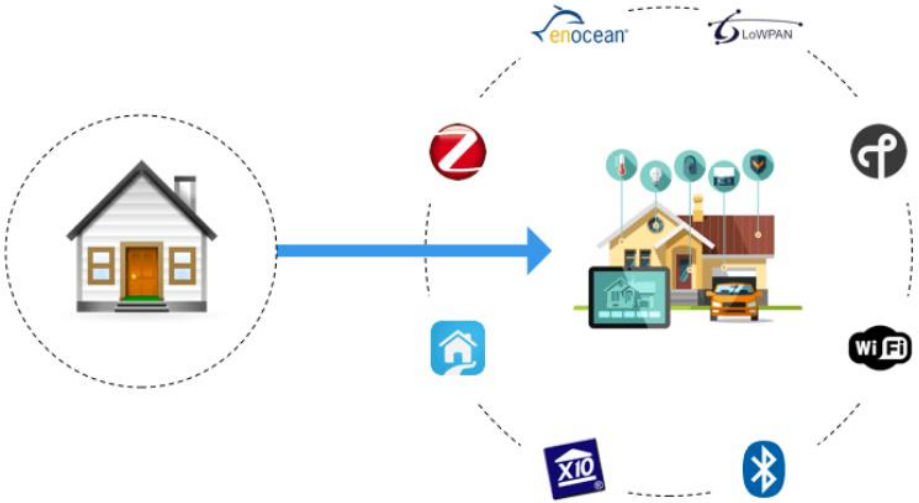


Figure 3 : Heterogeneity of IoT devices[3]

3 https://www.simform.com/iot-home-automation-protocols/

4 https://securityaffairs.co/wordpress/76348/malware/iot-devices-threat-landscape.html

Challenges

Heterogeneity

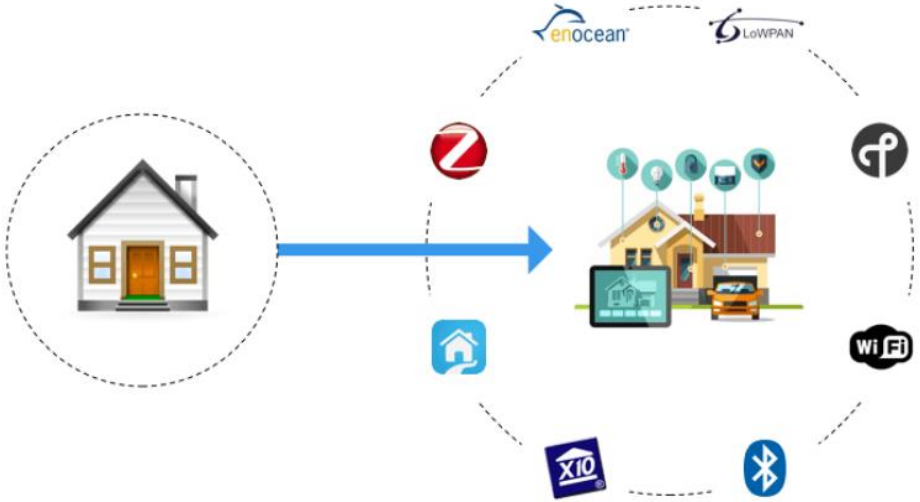


Figure 3 : Heterogeneity of IoT devices[3]

Security

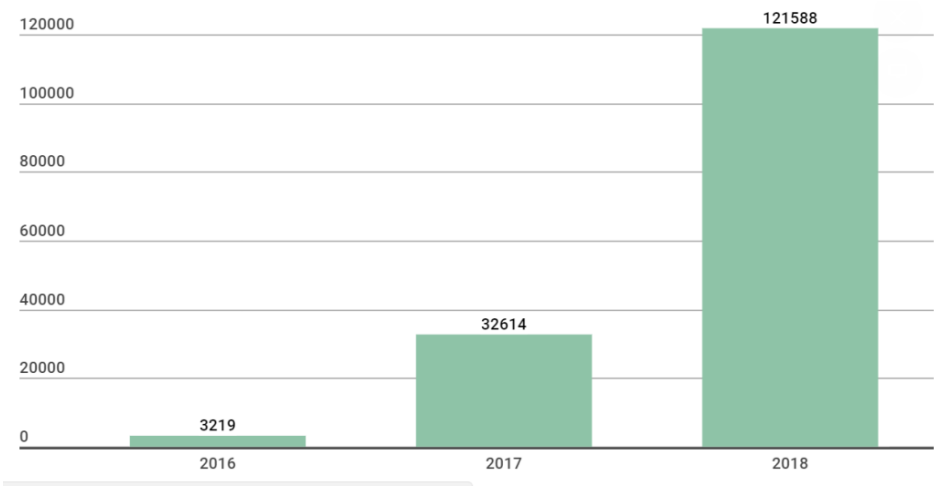


Figure 4 : Evolution of threats landscape for IoT devices[4]

3 https://www.simform.com/iot-home-automation-protocols/

4 https://securityaffairs.co/wordpress/76348/malware/iot-devices-threat-landscape.html

Motivation

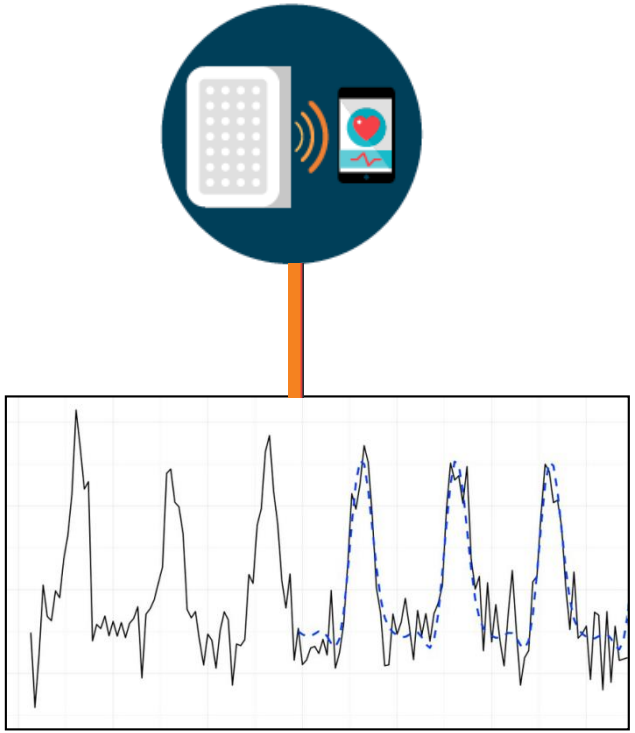
Problem statement

Monitor many devices in a Local Area Network (LAN)

1

- **Identify connected devices**

- **Authentication**
- **Information declared by devices**
  - **MAC or IP address, hostname**
  - **Not trusted**
- **Network traffic analysis**



Problem statement

Monitor many devices in a Local Area Network (LAN)

1 • Identify connected devices

- Authentication
- Information declared by devices
  - MAC or IP address, hostname
  - Not trusted
- Network traffic analysis

2 • Detect anomalies

- Continuously diagnose their state
- Trigger alerts in case of anomaly
  - Real-time and proactif system

Anomaly

"Something that deviates from what is standard, normal, or expected"
Oxford dictionary

Related work

Vertical Approach

Signature-Based Anomaly Detection

- Specific rule for each known anomaly
- Based on expert knowledge
- Threshold to detect anomalies
- Antivirus, IDS

Horizontal Approach

Behavior-Based Anomaly Detection

- Model a device network behavior (normal)
- Use (dis)similarity score to detect anomaly
- Metric learning

Related work

Vertical Approach

Signature-Based Anomaly Detection

- Specific rule for each known anomaly
- Based on expert knowledge
- Threshold to detect anomalies
- Antivirus, IDS


Precise in detecting known anomalies


Need experts to annotate data


Can not detect new anomalies


Scalability

Horizontal Approach

Behavior-Based Anomaly Detection

- Model a device network behavior (normal)
- Use (dis)similarity score to detect anomaly
- Metric learning


Detect known anomalies


Predict new anomalies


No data annotation


High false positive and negative rates

Related work

Vertical Approach

Signature-Based Anomaly Detection

- Specific rule for each known anomaly
- Based on expert knowledge
- Threshold to detect anomalies
- Antivirus, IDS

- Precise in detecting known anomalies
- Need experts to annotate data
- Can not detect new anomalies
- Scalability

Horizontal Approach

Behavior-Based Anomaly Detection

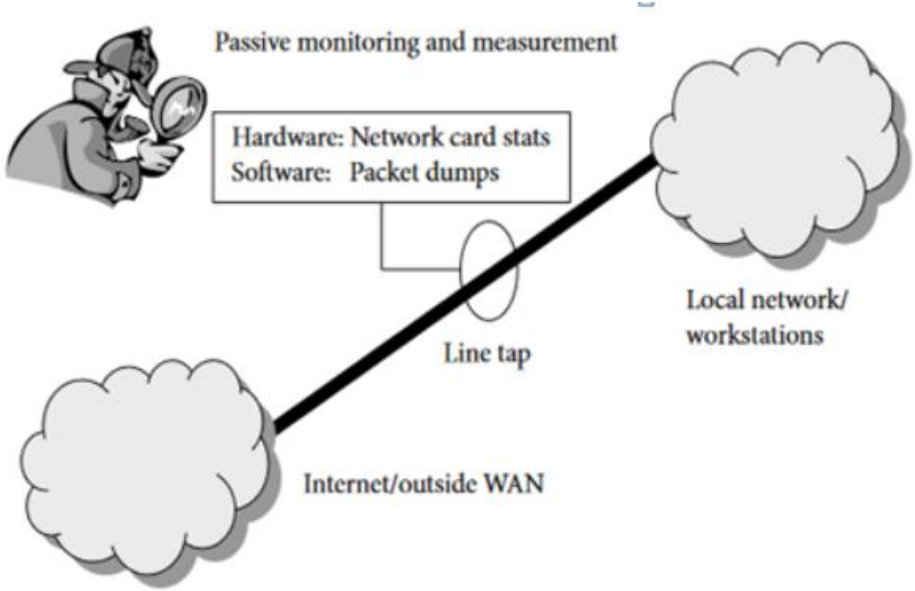
- Model a device network behavior (normal)
- Use (d[illegible] anomaly
- Metri[illegible]

- Detec[illegible]
- Predi[illegible]
- No data annotation
- High false positive and negative rates

Contribution

- Horizontal approach:
  - behavior-based approach
- Deep learning:
  - One model per device
- Network traffic analysis

Passive device fingerprinting [1]



[1] Network Traffic Analysis and SCADA Security

Evaluation

- Public and private datasets
- Cross Validation
  - Train , Validation, Test
- Evaluation metrics
  - Classification metrics
    - Accuracy, Precision, F1-score …
    - Confusion matrices
  - Metric learning
    - Mean Square Error
    - Cosine similarity

Metric	Formula
True positive rate, recall	$\frac{TP}{TP+FN}$
False positive rate	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-measure	$\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

Future Work

- State-of-the-art algorithms
- Contributions
- Communication
  - Conference
  - Research papers

Thank you

Naji NAJARI
TGI/OLS/HOME/VIBES/CARE

IoT Research Domain
Smart Object Management Project

Naji.najari@orange.com
Naji.najari@insa-lyon.fr
Naji.najari@grenoble-inp.org

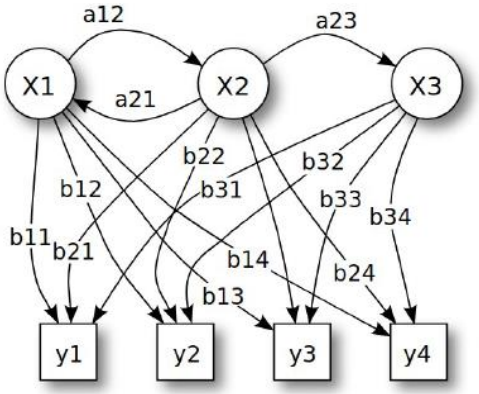
orange™

LIRIS

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

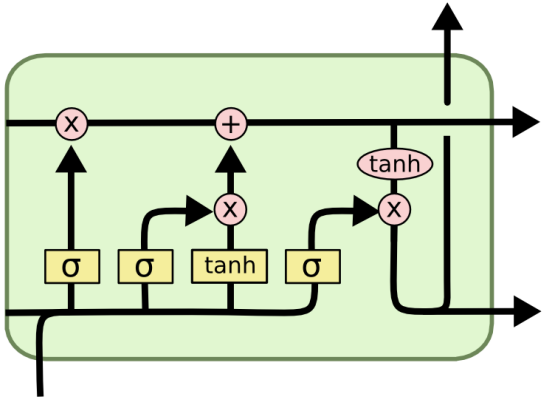
Appendix

Hidden Markov Model (HMM)



- BIC score to the identify number of hidden states
- EM algorithm

Long Short Term Memory (LSTM)



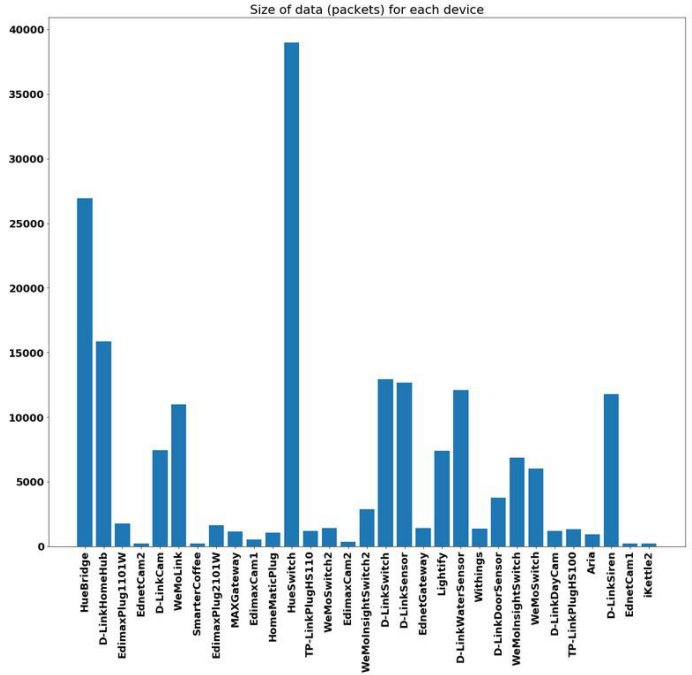
- Predict next step
- Back propagation through time
- Minimize the mean square error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2$$

Appendix

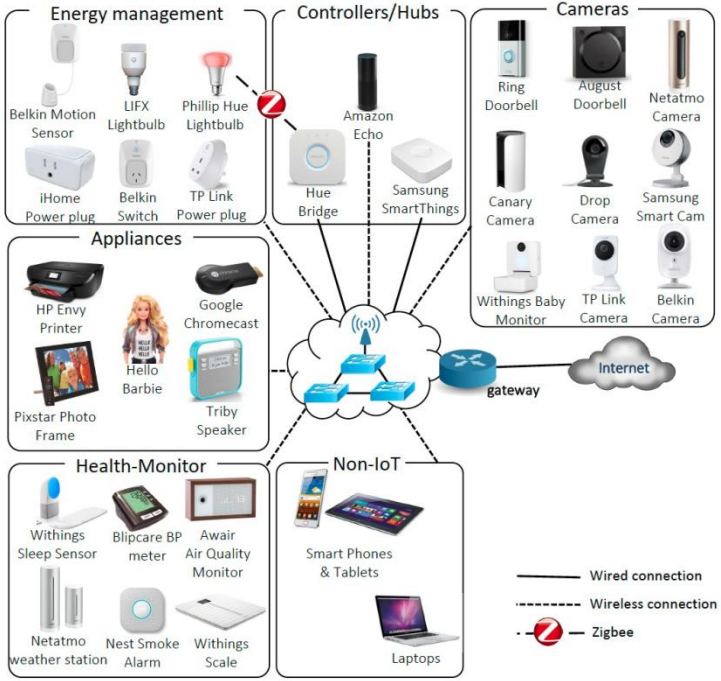
Dataset 1 : IoT Sentinel

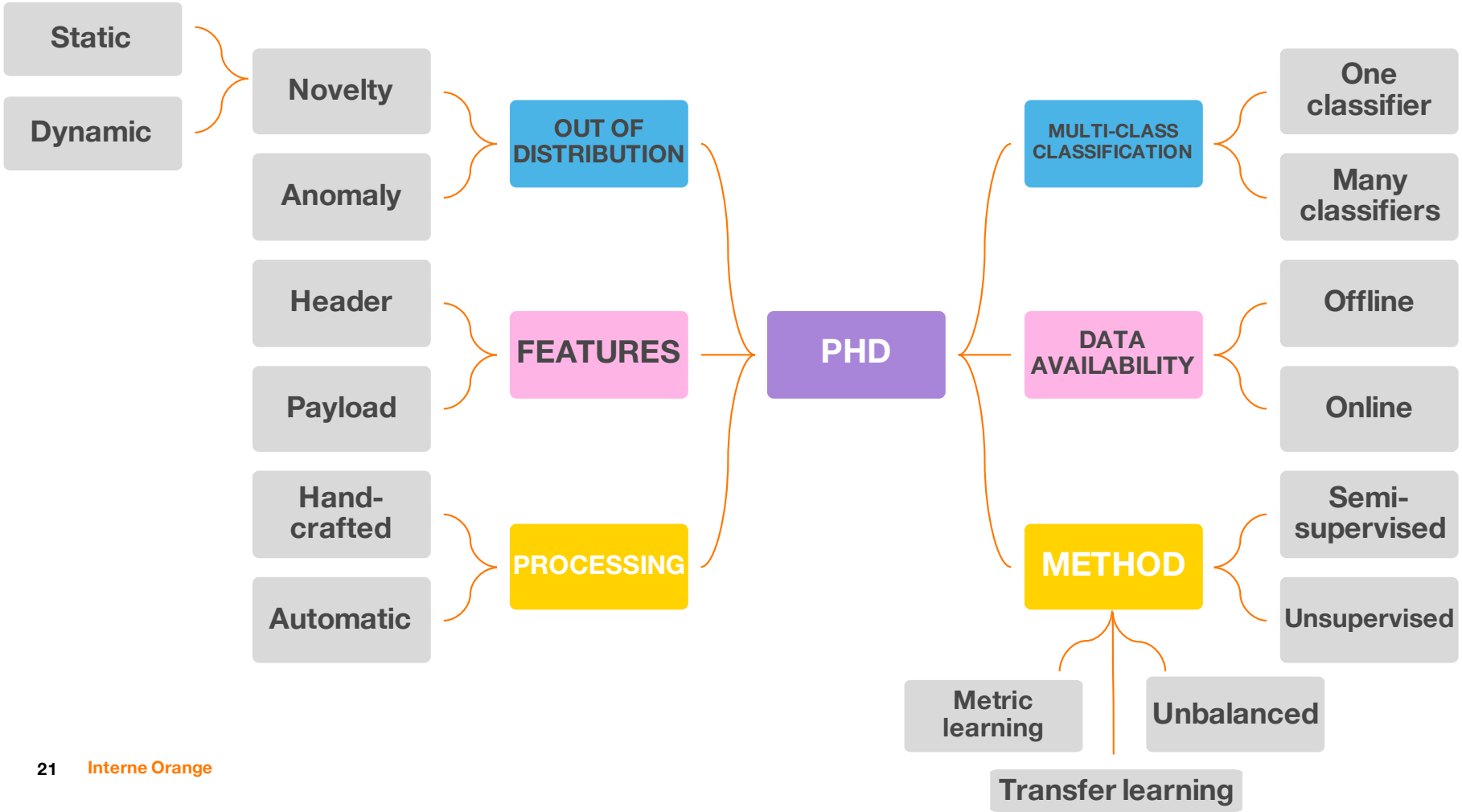
- Network traffic traces (pcap files)
- During setup
- 31 IoT devices



Dataset 2

- 28 IoT devices During setup
- During 20 days




Static
Dynamic
Novelty
Anomaly
Header
Payload
Hand-crafted
Automatic
OUT OF DISTRIBUTION
FEATURES
PROCESSING
PHD
MULTI-CLASS CLASSIFICATION
DATA AVAILABILITY
METHOD
One classifier
Many classifiers
Offline
Online
Semi-supervised
Unsupervised
Metric learning
Transfer learning
Unbalanced

Appendix

- Discriminative approaches
  - Random Forest
  - K-Nearest Neighbors
  - SVM

Appendix

- Discriminative approaches
  - Random Forest
  - K-Nearest Neighbors
  - SVM

Generative models: one model per device

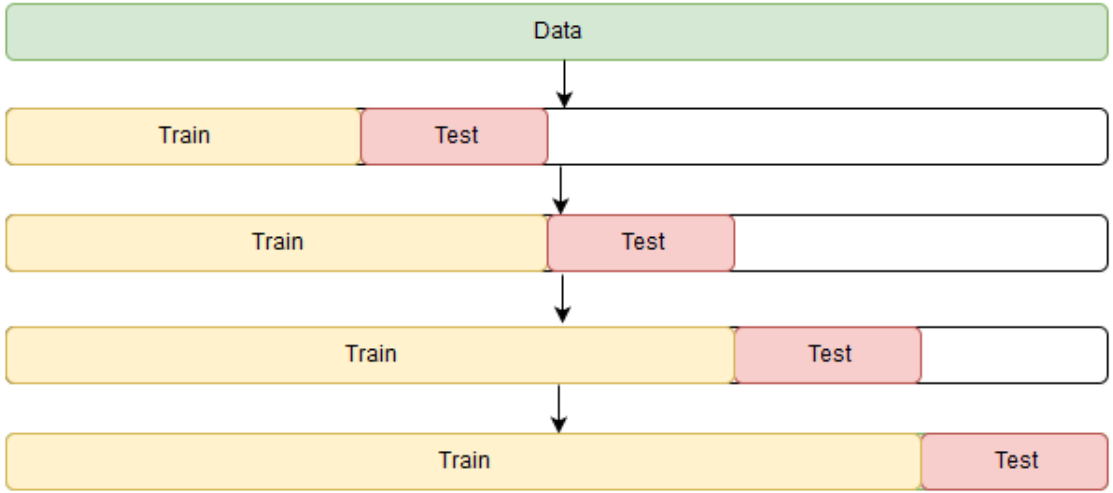
Appendix

- **Feature engineering**
  - **Header metadata**

Protocol Layer/Type	Features
Network	IP/ICMP/ICMPv6/EAPoL
Transport	TCP/UDP
Application	HTTP/HTTPS/DHCP/BOOTP/SSDP/DNS/MDNS/NTP
IP Options	*Padding/Router Alert*

  - **Payload metadata: Packet length and TCP-window size**

Evaluation



- **Publishing a paper**