WiFinger: Fingerprinting Noisy IoT Event Traffic Using Packet-level Sequence Matching

Ronghua Li[1], Shinan Liu[2], Haibo Hu[1], Qingqing Ye[1], Nick Feamster[3]

1. The Hong Kong Polytechnic University 2. The University of Hong Kong 3. University of Chicago

Background

Traffic Fingerprint

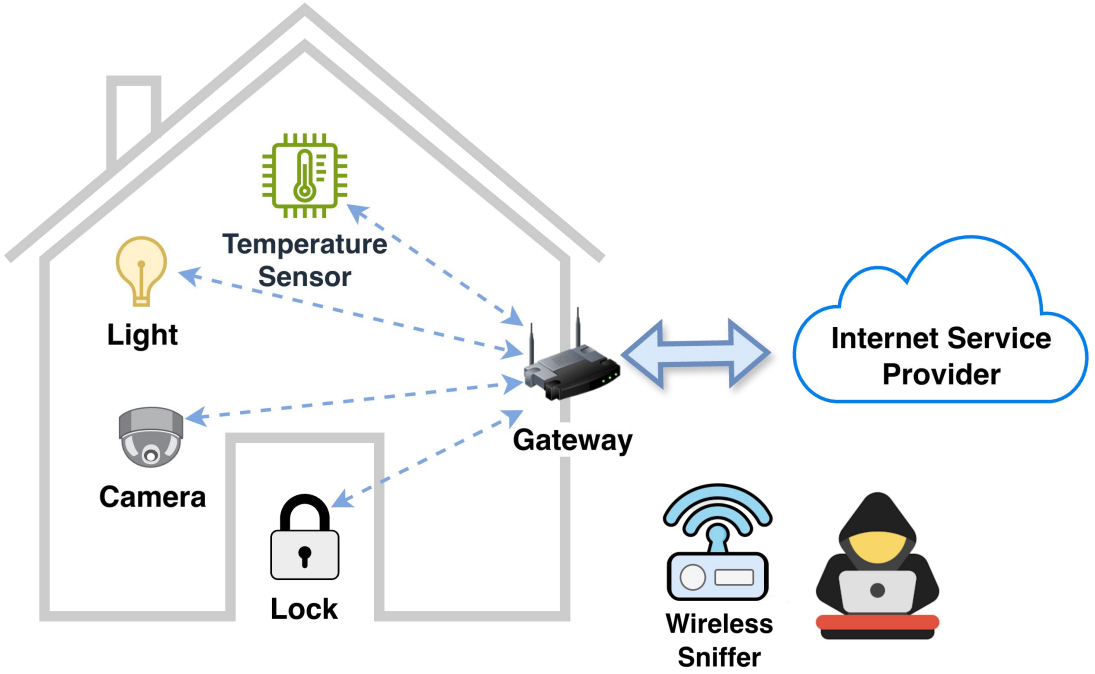
- Website
- Mobile App
- IoT Events
- Human Activities

Background

Traffic Fingerprint

- Website
- Mobile App
- IoT Events
- Human Activities

Door unlocked?
Security camera shutdown?
Empty houses?
.....



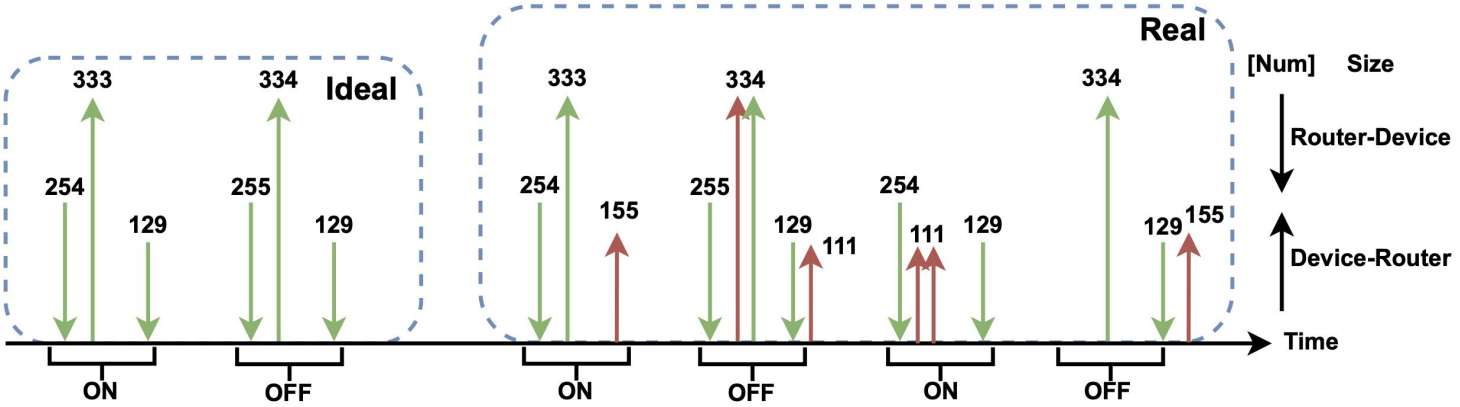
Problem

Traffic Fingerprint Approaches:

- ML: Random Forest, XGBoost, KNN
- DL: CNN, LSTM
- Heuristic: Size Matching

Question: What's the difference? Transport/Network (TCP/IP) v.s. Link (Wi-Fi)

Problem: Layer Difference




```
375  QoS Data, SN=3252, FN=0, Flags=.p....F.
447  QoS Data, SN=3254, FN=0, Flags=.p....F.
439  QoS Data, SN=3255, FN=0, Flags=.p....F.
384  QoS Data, SN=3256, FN=0, Flags=.p....F.
455  QoS Data, SN=3258, FN=0, Flags=.p....F.
```

TAKEAWAY: Wi-Fi IoT events are **short but versertile.** They have **sniffing loss** and **traffic obfuscation**

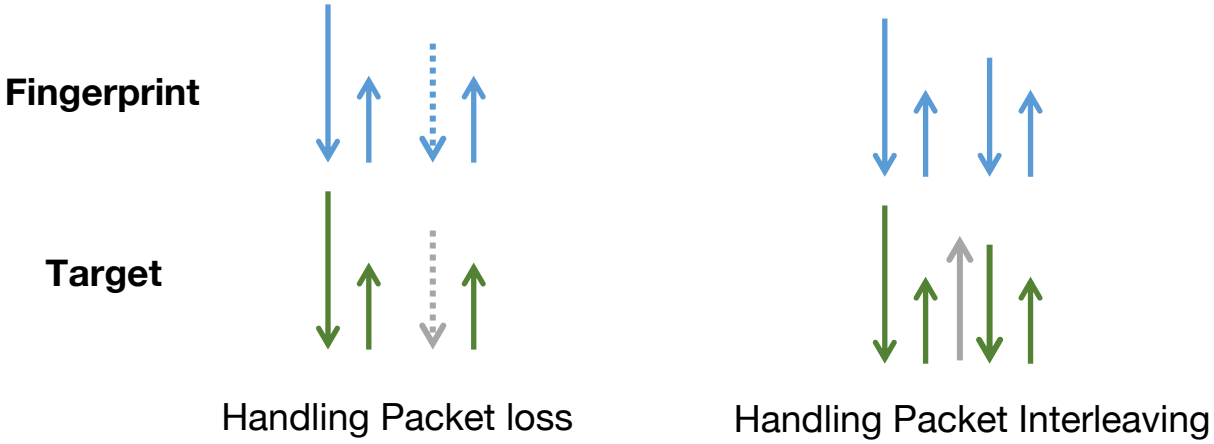
Problem: Trade-off

Methods	Features	Data Volume	Label Acc	Parameter Tuning	Traffic Completeness
ML	Traffic Statistics	Medium	High	High	Medium
DL	Raw Bytes	High	High	Low	Medium
Heuristic	Packet Directions & Sizes	Low	Low	Low	High

TAKEAWAY: Wi-Fi IoT events are **short but versertile.** They have **sniffing loss** and **traffic obfuscation.**

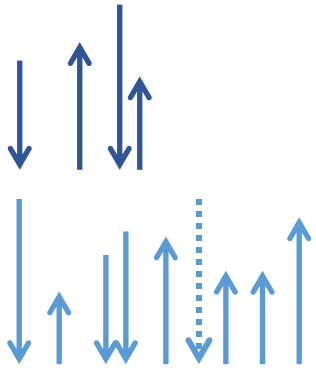
Addressing Loss and Obfuscation

- Packet Loss
- Packet Obfuscation



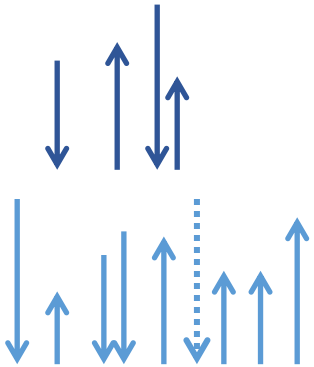
Network Traffic Longest Common Subsequence (NT-LCS)

FMLCS: Fuzzily Matched Longest Common Subsequence



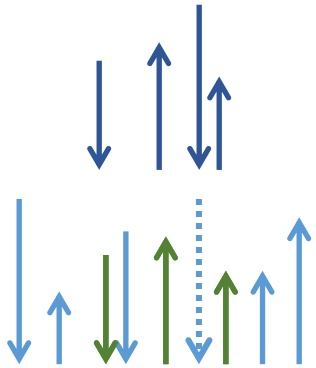
FMLCS: Fuzzily Matched Longest Common Subsequence

FMLCS: Fuzzily Matched Longest Common Subsequence



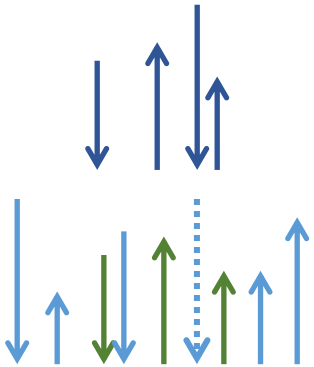
FMLCS: Fuzzily Matched Longest Common Subsequence

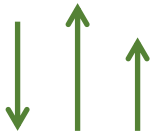
FMLCS: Fuzzily Matched Longest Common Subsequence



FMLCS: Fuzzily Matched Longest Common Subsequence

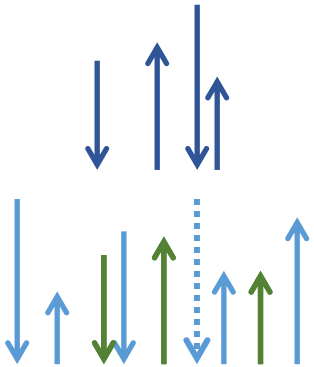
FMLCS: Fuzzily Matched Longest Common Subsequence

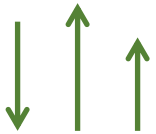




FMLCS: Fuzzily Matched Longest Common Subsequence

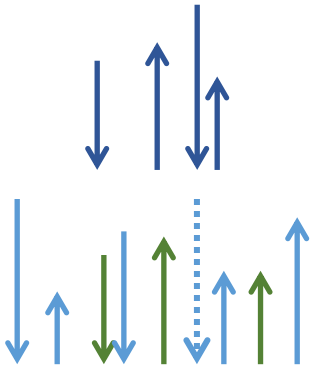
FMLCS: Fuzzily Matched Longest Common Subsequence





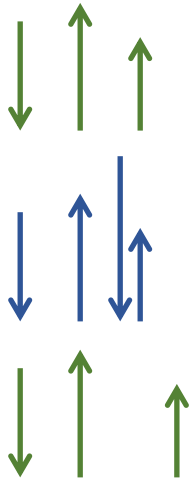
FMLCS: Fuzzily Matched Longest Common Subsequence

FMLCS: Fuzzily Matched Longest Common Subsequence



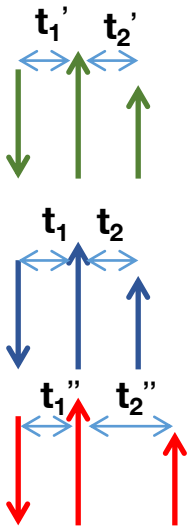
FMLCS: Fuzzily Matched Longest Common Subsequence

FMLCS: Fuzzily Matched Longest Common Subsequence



FMLCS: Fuzzily Matched Longest Common Subsequence

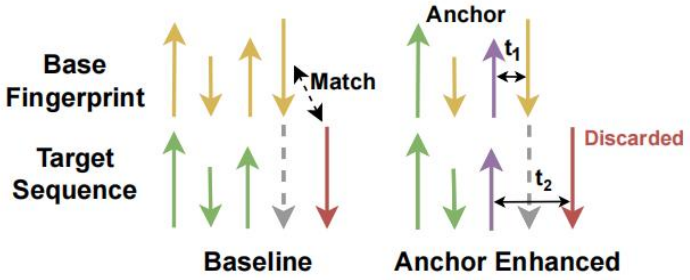
FMLCS: Fuzzily Matched Longest Common Subsequence



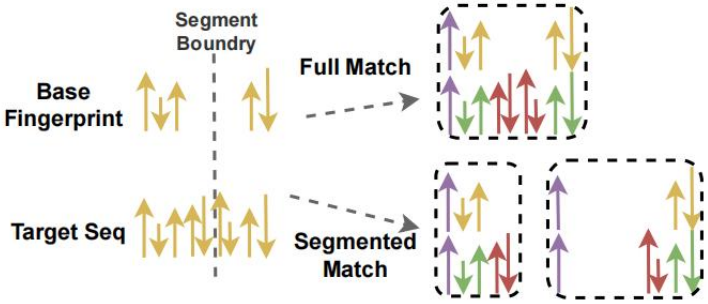
Challenges:

1. High time complexity
2. Real-environment adaptation

Accuracy & Efficiency



Anchor Reference



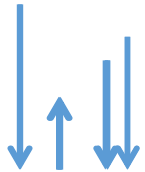
Split and Merge

Challenges:

1. High time complexity
2. Real-environment adaptation

Process Time(s)	
FMLCS	AFMLCS
0.13	0.13
0.17	0.07
2.34	0.15
7.97	0.85
39.2	1.3
0.14	0.15
0.34	0.1

Environment Adaptability



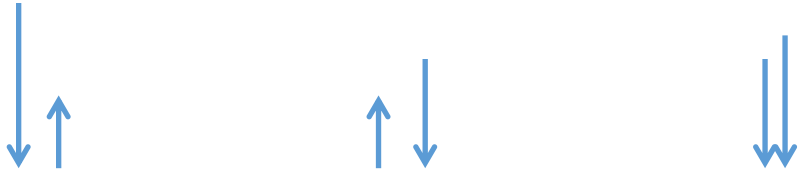
Challenges:

1. High time complexity
2. Real-environment adaptation

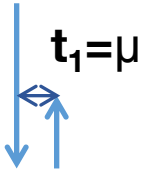
- Packet intervals may slightly change in various environments

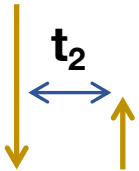
Environment Adaptability

Pair Interval Distribution



- Mean - μ
- Std - σ





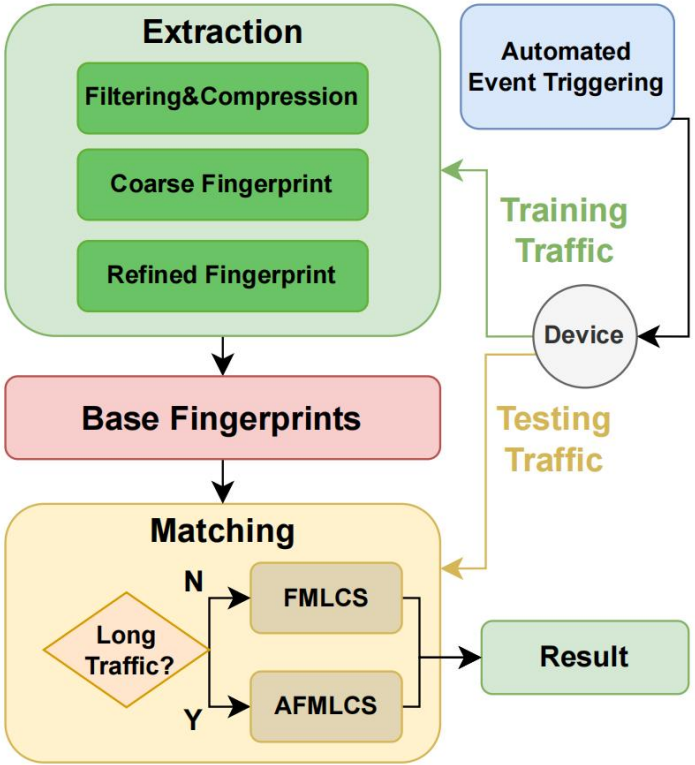
Interval Calibration

if $|t_2 - t_1| < \mu \pm 3\sigma$:
$t2 = \mu \pm \sigma$

Challenges:

1. High time complexity
2. Real-environment adaptation

Fingerprint Matching



- 30 training samples
- 20 testing samples

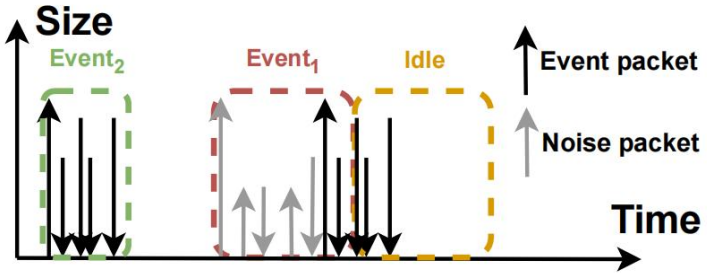
Experiments

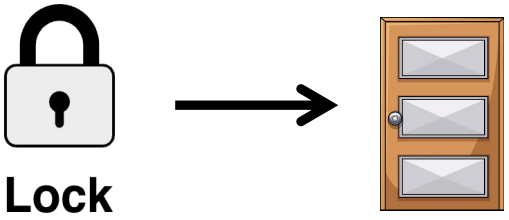
ID	Device Type	Device	Related Events
E1	Agentic Controller	Alexa Echo Dot	DND/UnDND
E2 (H)			Q1: What time is it?
E3 (H)			Q2: What's the price of eggs?
E4 (H)			Q3: What's the weather now?
E5 (H)			Q4: What's the weather like in X?
E6			Volume Up/Down
E7 (H)		Google Home	Volume Up
E8 (H)			Volume Down
E9	Smart Peripherals	Wiz Hue Light	On
E10			Off
E11		TP-Link Plug	On/Off
E12		ICX-RF A/C Controller	On/Off
E13		Gosund Plug	On/Off
E14		WAH Plug	On
E15			Off
E16	Integrated Smart Actuator	Mi Sweeper	On/Off
E17			Mode Silent/Standard/Strong
E18		Midea Dishwasher	On/Off
E19		Midea Dish Sterilizer	On/Off
E20		Xiaomi Humidifier	On/Off
E21			Continuous humidification/Close
E22		Wi-Fi Sprinkler	On/Off
E23		TuYa Thermostat	On
E24			Off
E25			Temperature INC
E26			Temperature DEC
E27		HW Thermostat	On
E28			Off
E29			Temperature INC/DEC
E30			Mode Comfort
E31			Mode Non-Frozen
E32		Ring Alarm	Mute
E33			Ring

- 15 Devices
- 43 Events
- Real-time event tracking

Evaluation - Precision v.s. Recall

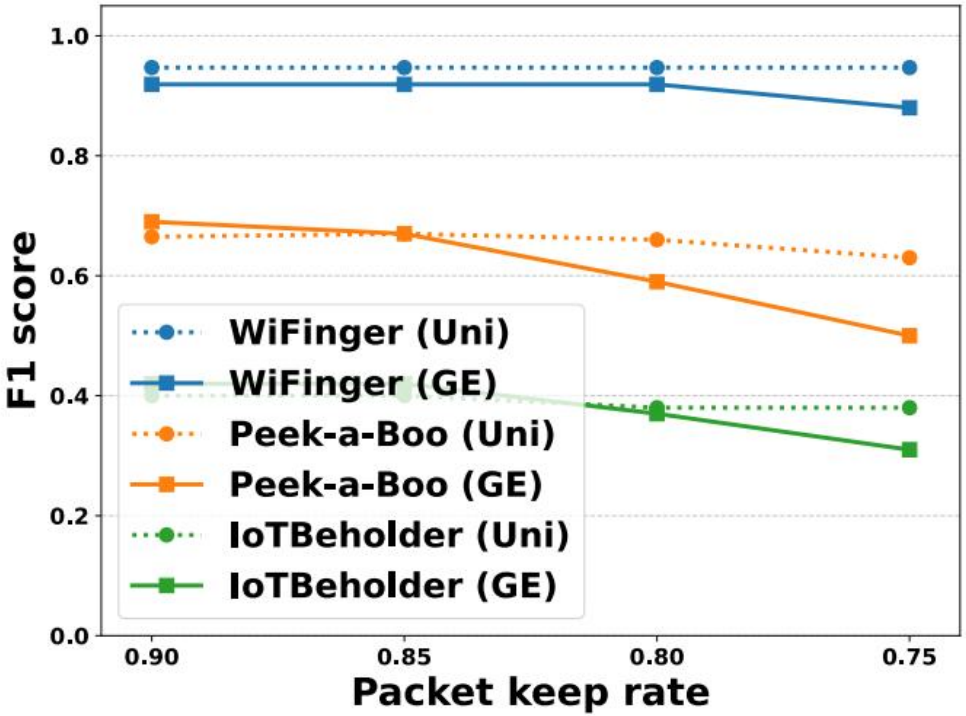
Precision: 96%; Recall: 89%





Precision is more important than Recall

Robustness against Various Loss Rates



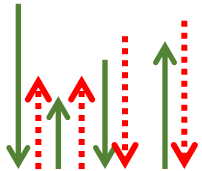
Loss distributions:

- Uniform
- Gilbert-Elliot

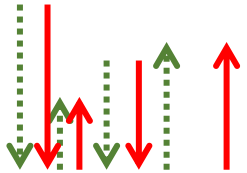
Defenses

delay, shaping, padding defenses

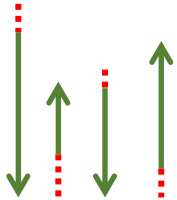
Shaping

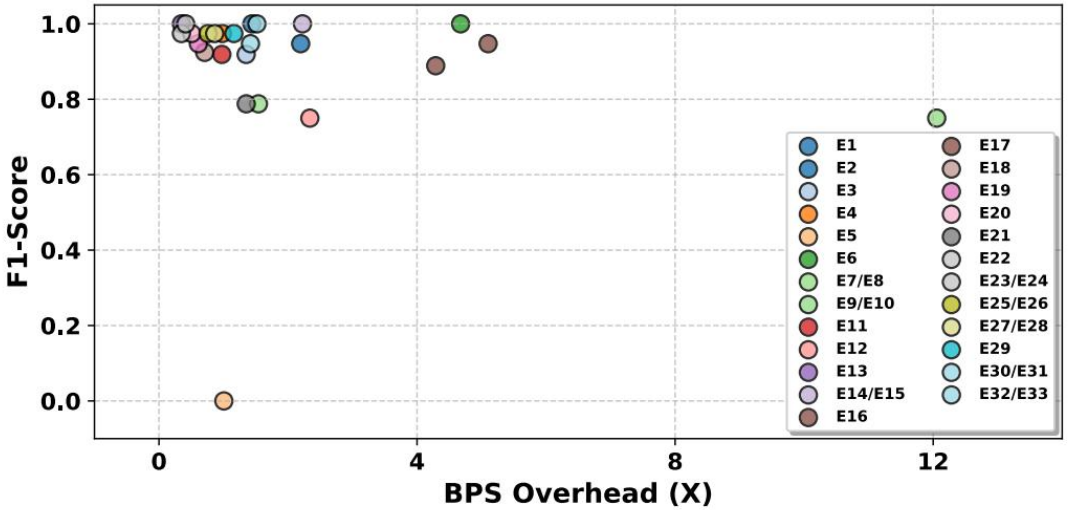


Delay



Padding

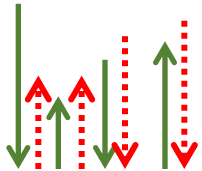




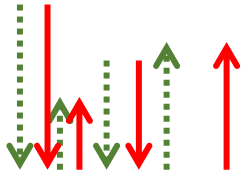
Defenses

delay, shaping, padding defenses

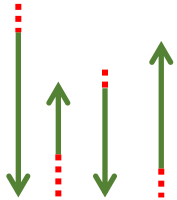
Shaping

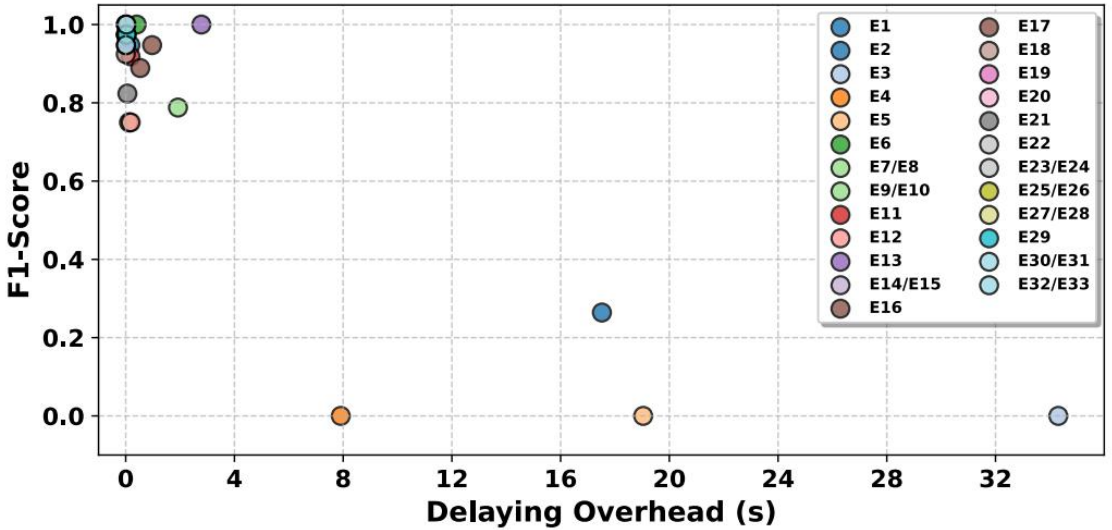


Delay



Padding

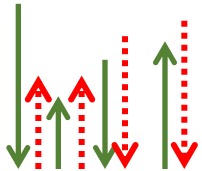




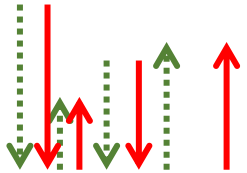
Defenses

delay, shaping, padding defenses

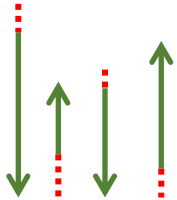
Shaping

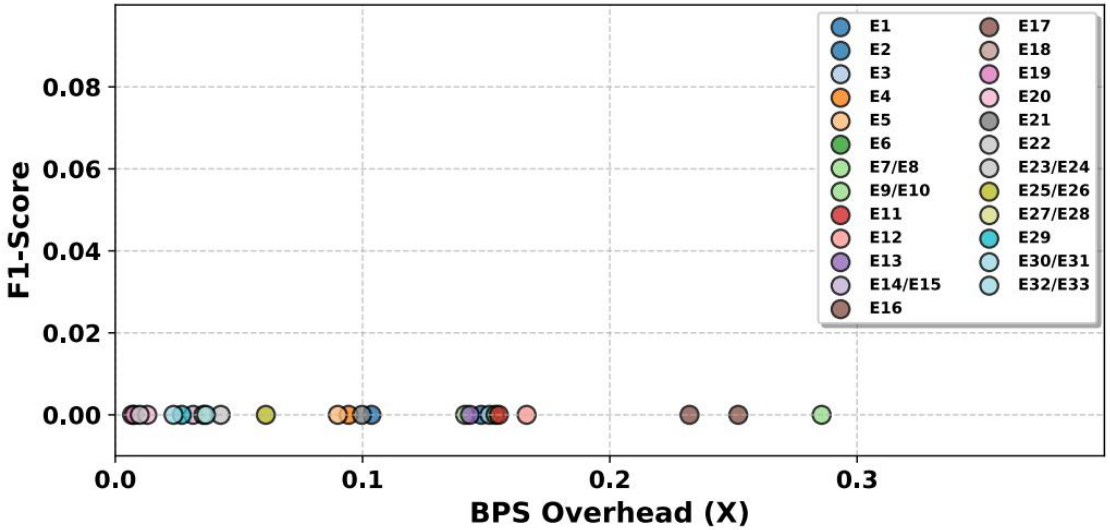


Delay



Padding





WiFinger: Fingerprinting Noisy IoT Event Traffic Using Packet-level Sequence Matching

Ronghua Li[1], Shinan Liu[2], Haibo Hu[1], Qingqing Ye[1], Nick Feamster[3]

1. The Hong Kong Polytechnic University 2. The University of Hong Kong 3. University of Chicago

ASTAPLE

Applied Security, Trust and Privacy Lab for Enterprise

https://www.astaple.com/

Paper

Question Form