UNIVERSITY OF CAMBRIDGE

Scalable Device Identification for IoT Networks using Binary Classification Models at the Edge

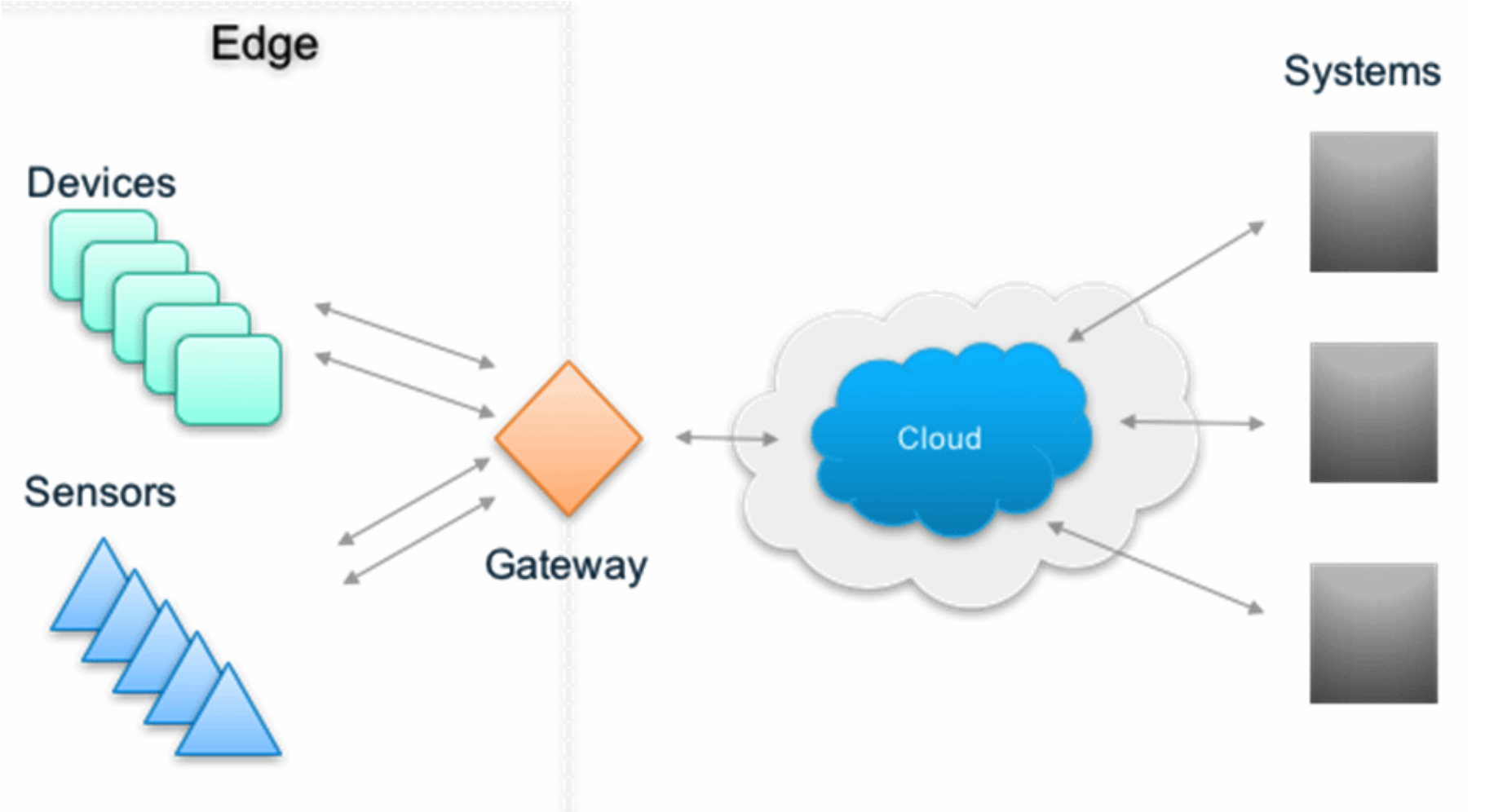
Roman Kolcun, Richard Mortier[1]

[1] University of Cambridge, UK

Why IoT?

- Billions of devices worldwide
- Deployed at our homes
- Always listening/watching
- Can leak information about our lives
- Device can be hacked or misconfigured
- In order to apply different rules, devices need to be identified
- Signatures could be used to identify when a device is misbehaving
- An unknown device can be identified depending on its network signature

Architecture



How Do We Differ From Other Solutions?

- Scalability
  - No two households are the same
  - Vast majority of other approaches use single model to classify all devices
- Responsivness
  - We want to achieve near real-time response
  - Many other approaches collect data for extended period of time (e.g. hours)

Dataset

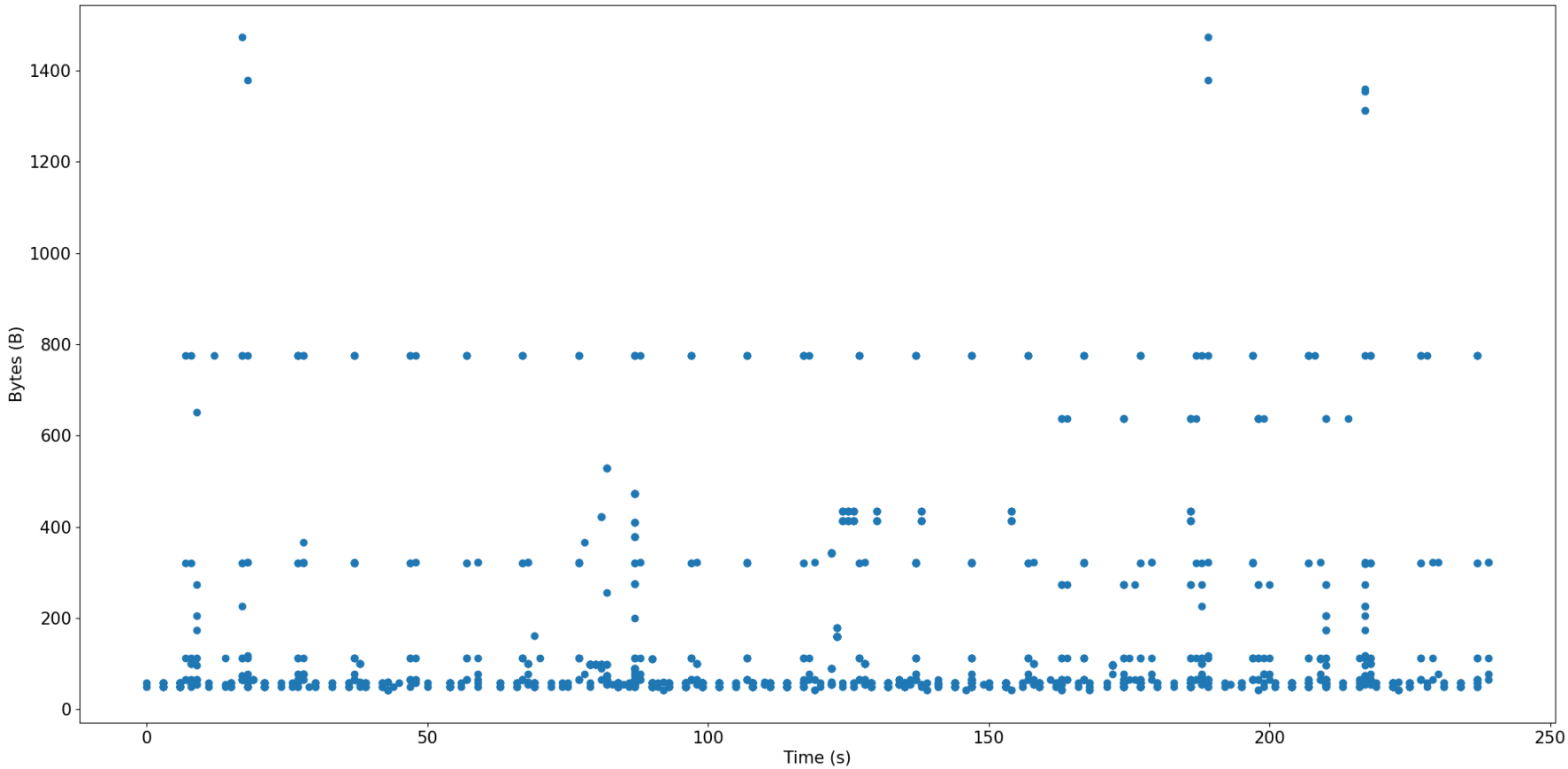
- 40 IoT devices in a testbed
- Contain active and idle periods
- Devices might have been disconnected for arbitrary lengths of periods
- Therefore data are split into 4 parts depending on percentage not time

Source: Kolcun, R., Popescu, D.A., Safronov, V., Yadav, P., Mandalari, A.M., Mortier, R. and Haddadi, H., 2021. Revisiting IoT device identification., TMA 2021

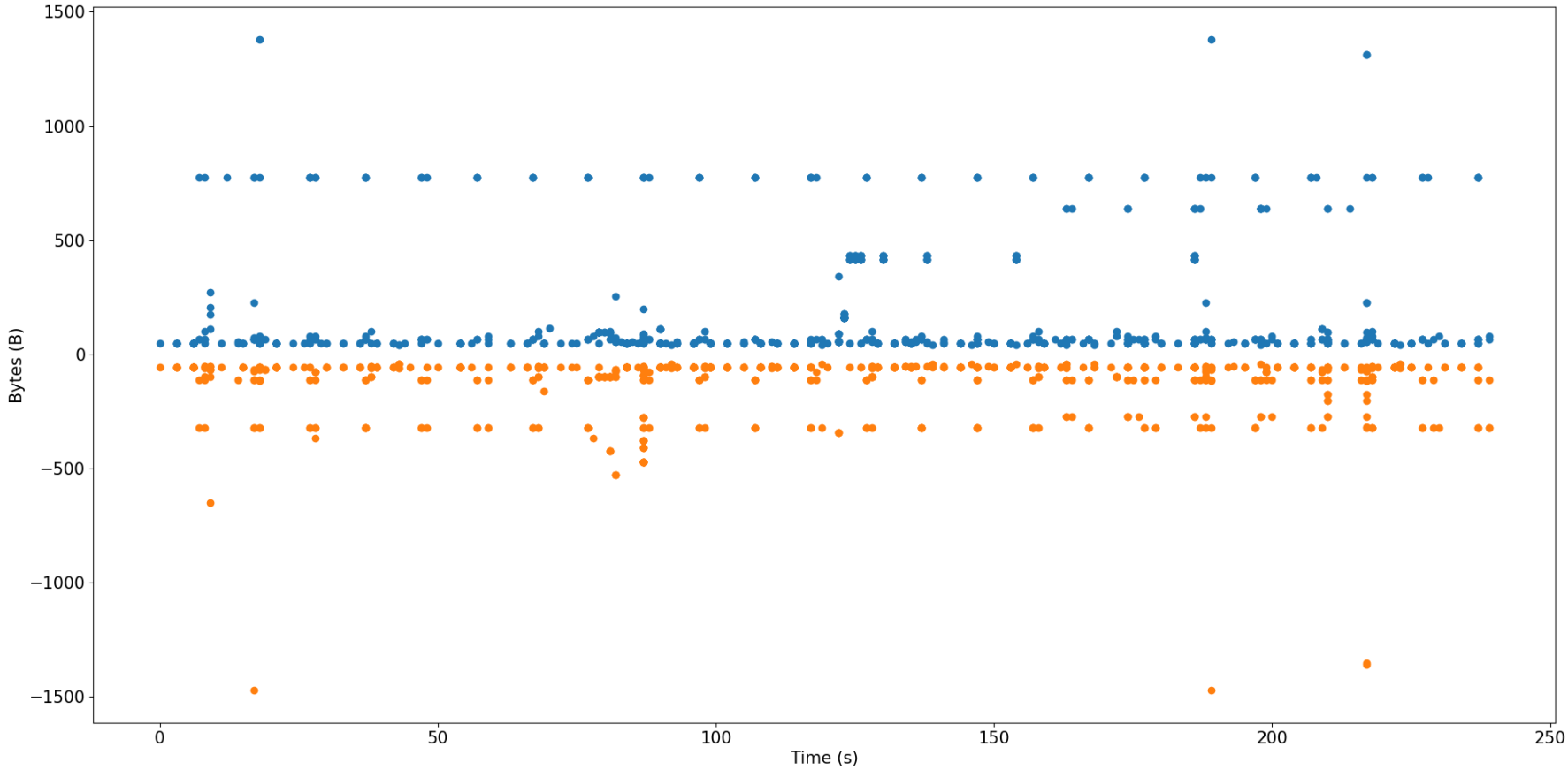
An Example - Echo Spot

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	52.95.121.5	192.168.20.130	TCP	60	443 → 35745 [ACK] Seq=1 Ack=1 Win=2804 Len=0
2	0.000084	52.95.121.5	192.168.20.130	TLSv1.2	100	Application Data
3	4.703030	155.198.142.7	192.168.20.130	DNS	146	Standard query response 0x5215 A 2.android.pool.ntp.org A 176.58.109.199 A 213.138.110.176 A 35.178.171.140 A 217.155.2.22
4	4.753077	176.58.109.199	192.168.20.130	NTP	90	NTP Version 3, server
5	4.793956	155.198.142.7	192.168.20.130	DNS	99	Standard query response 0xdd6d A arcus-uswest.amazon.com A 52.119.164.214
6	4.974124	52.119.164.214	192.168.20.130	TCP	66	443 → 57355 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
7	5.056173	155.198.142.7	192.168.20.130	DNS	90	Standard query response 0xab3f A api.amazon.com A 52.94.240.242
8	5.135543	52.94.240.242	192.168.20.130	TCP	66	443 → 60576 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
9	5.158207	52.119.164.214	192.168.20.130	TCP	60	[TCP Window Update] 443 → 57355 [ACK] Seq=1 Ack=1 Win=27136 Len=0
10	5.158916	52.119.164.214	192.168.20.130	TCP	66	443 → 58819 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
11	5.159312	52.119.164.214	192.168.20.130	TCP	66	443 → 33711 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
12	5.160786	52.119.164.214	192.168.20.130	TCP	60	443 → 57355 [ACK] Seq=1 Ack=231 Win=28160 Len=0
13	5.160860	52.119.164.214	192.168.20.130	TLSv1.2	146	Server Hello
14	5.160882	52.119.164.214	192.168.20.130	TLSv1.2	4901	Certificate
15	5.161730	52.119.164.214	192.168.20.130	TLSv1.2	392	Server Key Exchange
16	5.161792	52.119.164.214	192.168.20.130	TLSv1.2	63	Server Hello Done
17	5.186614	52.119.164.214	192.168.20.130	TCP	63	[TCP Retransmission] 443 → 57355 [PSH, ACK] Seq=5278 Ack=231 Win=28160 Len=9
18	5.199284	155.198.142.7	192.168.20.130	DNS	104	Standard query response 0x571c A device-metrics-us.amazon.com A 54.239.31.37
19	5.213925	52.94.240.242	192.168.20.130	TCP	60	[TCP Window Update] 443 → 60576 [ACK] Seq=1 Ack=1 Win=27136 Len=0
20	5.220990	52.94.240.242	192.168.20.130	TCP	60	443 → 60576 [ACK] Seq=1 Ack=254 Win=28160 Len=0
21	5.222174	52.94.240.242	192.168.20.130	TLSv1.2	1514	Server Hello
22	5.222197	52.94.240.242	192.168.20.130	TLSv1.2	3090	Certificate, Server Key Exchange, Server Hello Done
23	5.251443	52.94.240.242	192.168.20.130	TCP	170	[TCP Retransmission] 443 → 60576 [PSH, ACK] Seq=4381 Ack=254 Win=28160 Len=116
24	5.281859	54.239.31.37	192.168.20.130	TCP	66	443 → 42928 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
25	5.287284	54.239.31.37	192.168.20.130	TCP	66	443 → 41642 [SYN, ACK] Seq=0 Ack=1 Win=8190 Len=0 MSS=1460 WS=64 SACK_PERM=1
26	5.303406	52.119.164.214	192.168.20.130	TCP	60	[TCP Window Update] 443 → 58819 [ACK] Seq=1 Ack=1 Win=27136 Len=0
27	5.303463	52.119.164.214	192.168.20.130	TCP	60	[TCP Window Update] 443 → 33711 [ACK] Seq=1 Ack=1 Win=27136 Len=0
28	5.303770	52.119.164.214	192.168.20.130	TCP	60	443 → 58819 [ACK] Seq=1 Ack=231 Win=28160 Len=0
29	5.303794	52.119.164.214	192.168.20.130	TLSv1.2	146	Server Hello
30	5.303969	52.119.164.214	192.168.20.130	TCP	1514	443 → 58819 [ACK] Seq=93 Ack=231 Win=28160 Len=1460 [TCP segment of a reassembled PDU]
31	5.303994	52.119.164.214	192.168.20.130	TLSv1.2	3441	Certificate
32	5.305008	52.119.164.214	192.168.20.130	TCP	60	443 → 33711 [ACK] Seq=1 Ack=231 Win=28160 Len=0
33	5.305030	52.119.164.214	192.168.20.130	TLSv1.2	392	Server Key Exchange
34	5.305038	52.119.164.214	192.168.20.130	TLSv1.2	63	Server Hello Done
35	5.305046	52.119.164.214	192.168.20.130	TLSv1.2	146	Server Hello
36	5.305056	52.119.164.214	192.168.20.130	TCP	1514	443 → 33711 [ACK] Seq=93 Ack=231 Win=28160 Len=1460 [TCP segment of a reassembled PDU]
37	5.305099	52.119.164.214	192.168.20.130	TLSv1.2	3441	Certificate
38	5.305973	52.119.164.214	192.168.20.130	TLSv1.2	392	Server Key Exchange
39	5.305985	52.119.164.214	192.168.20.130	TLSv1.2	63	Server Hello Done

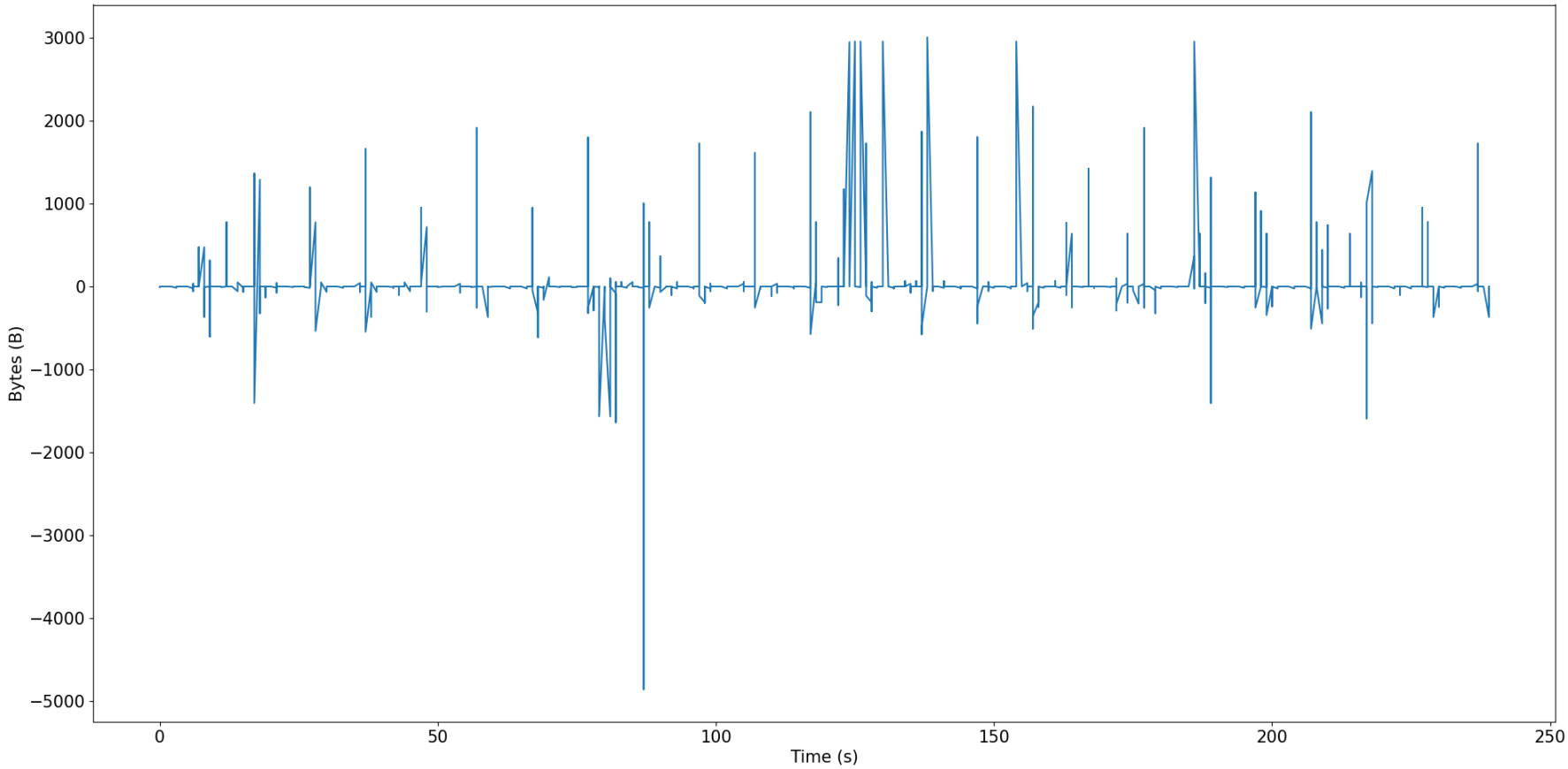
An Example - Echo Spot



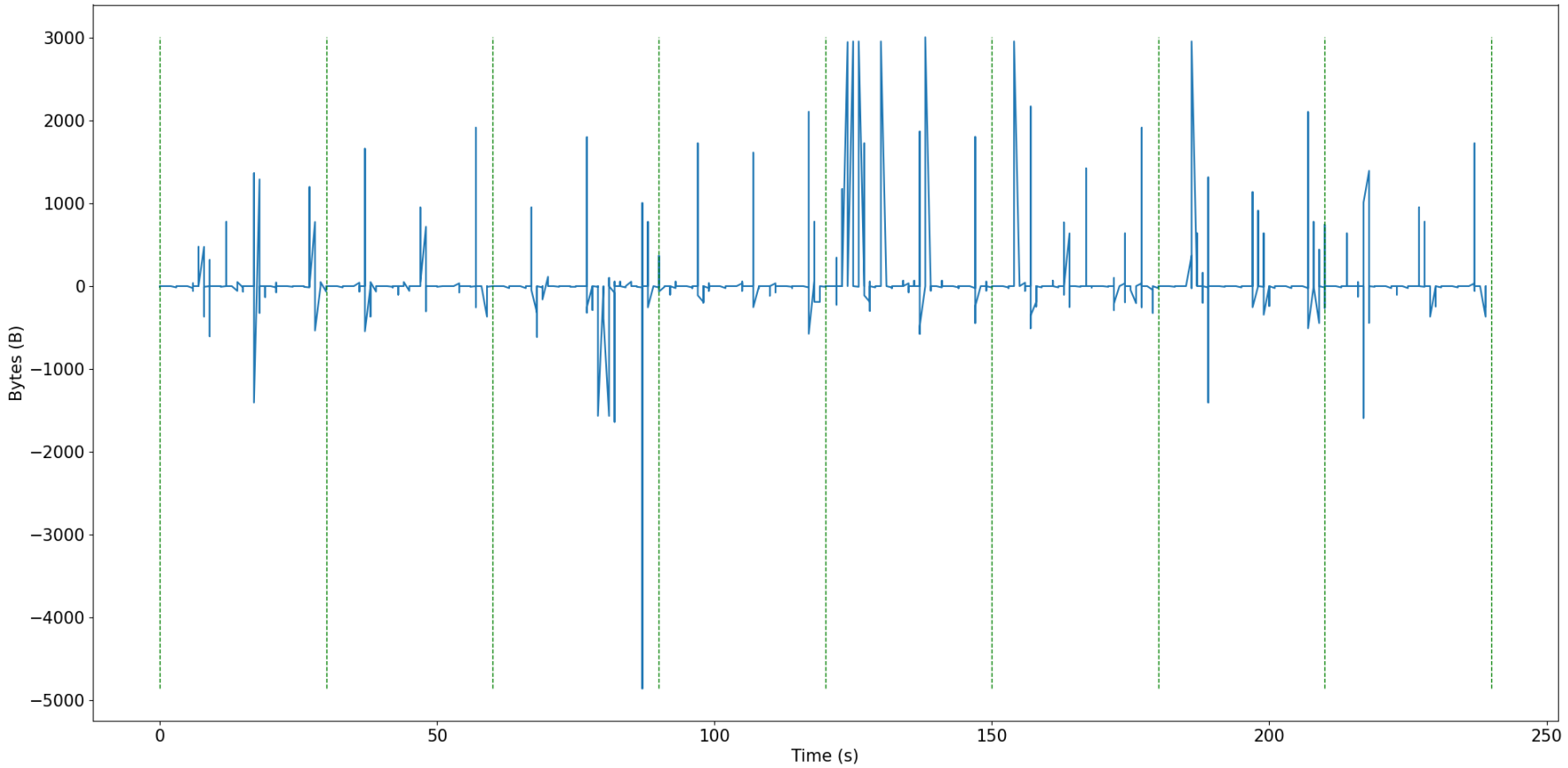
An Example - Echo Spot



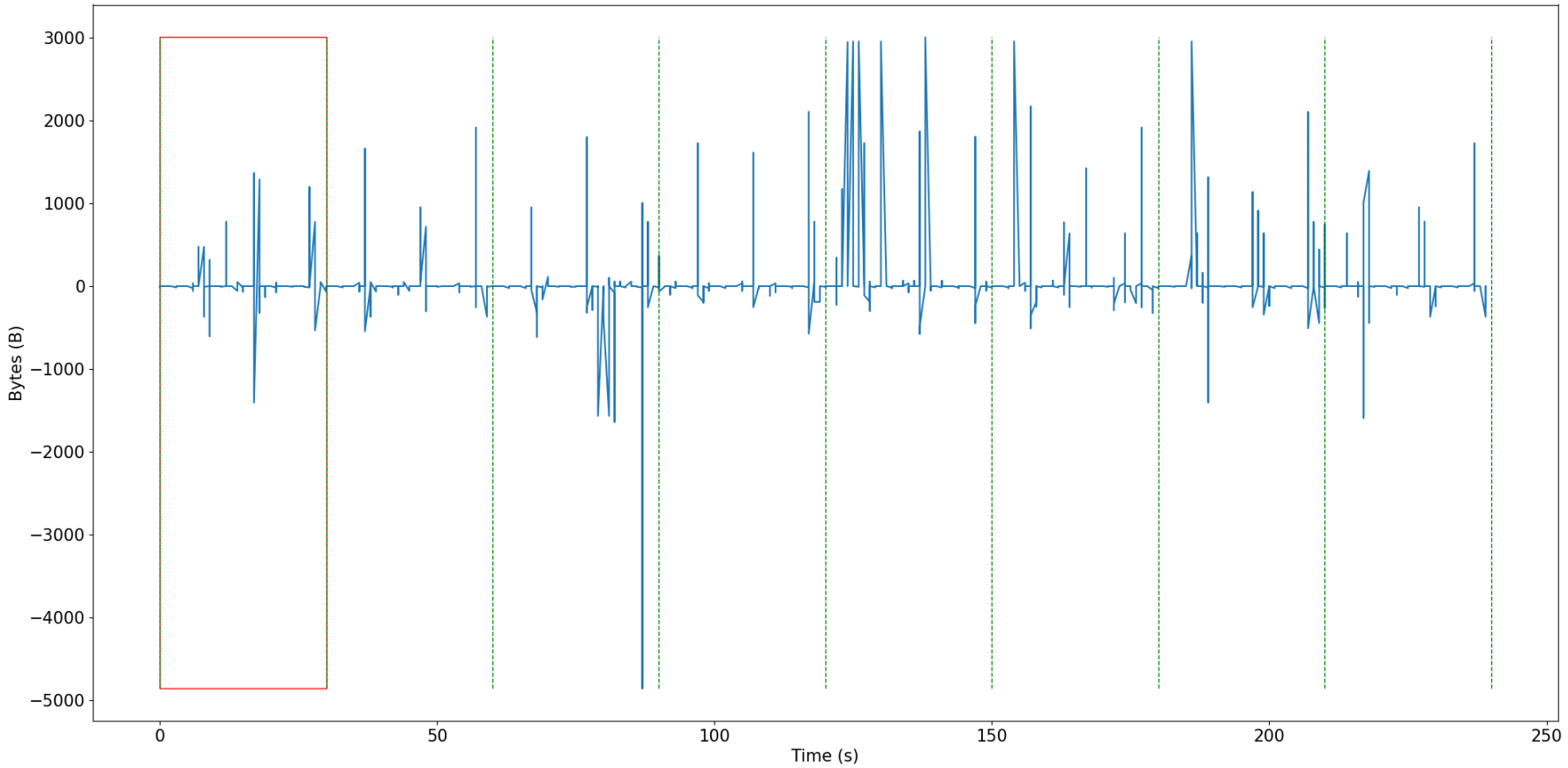
An Example - Echo Spot



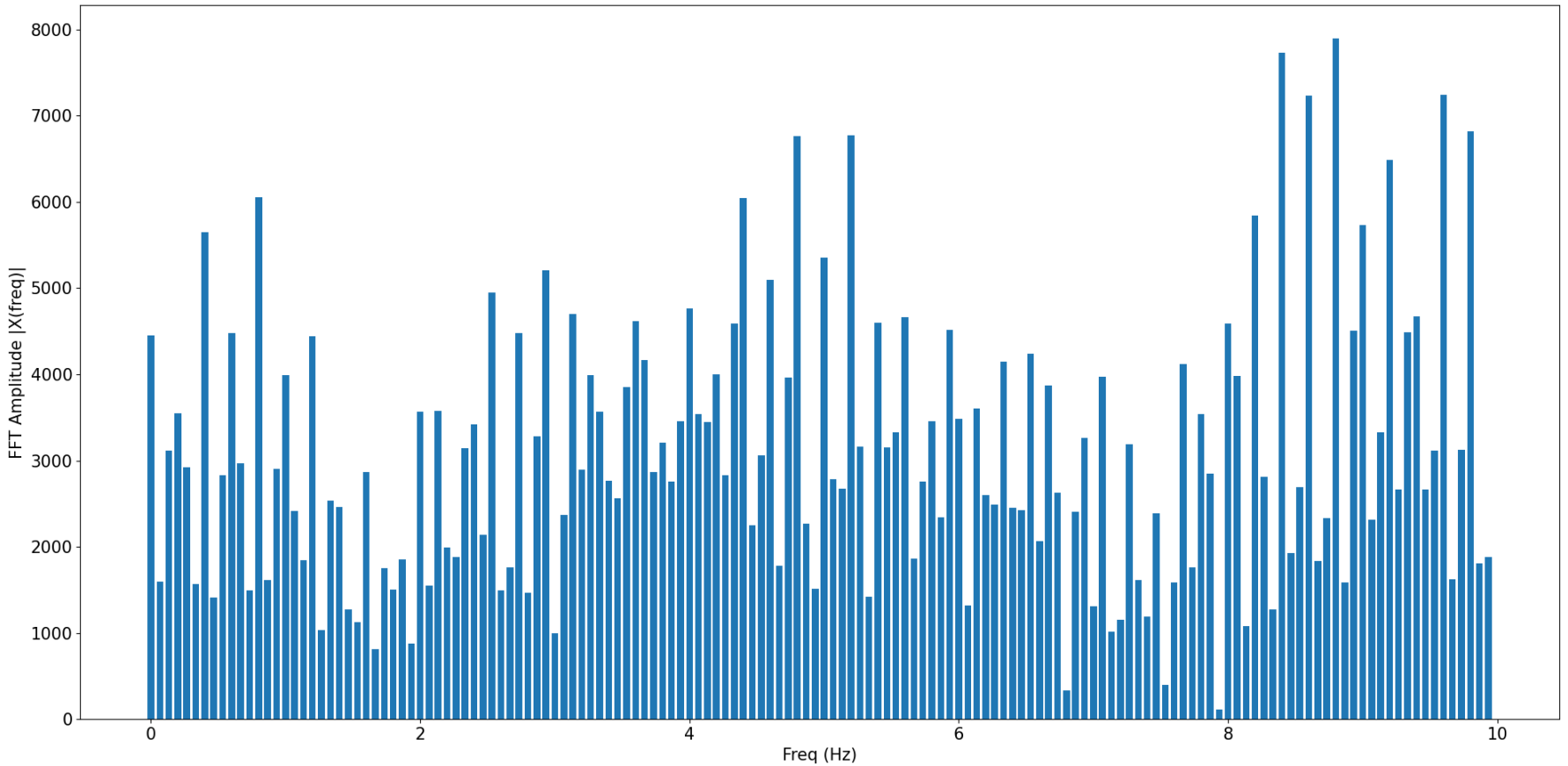
An Example - Echo Spot



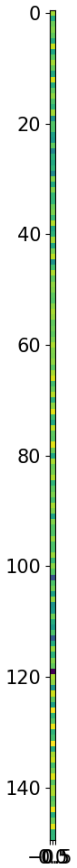
An Example - Echo Spot



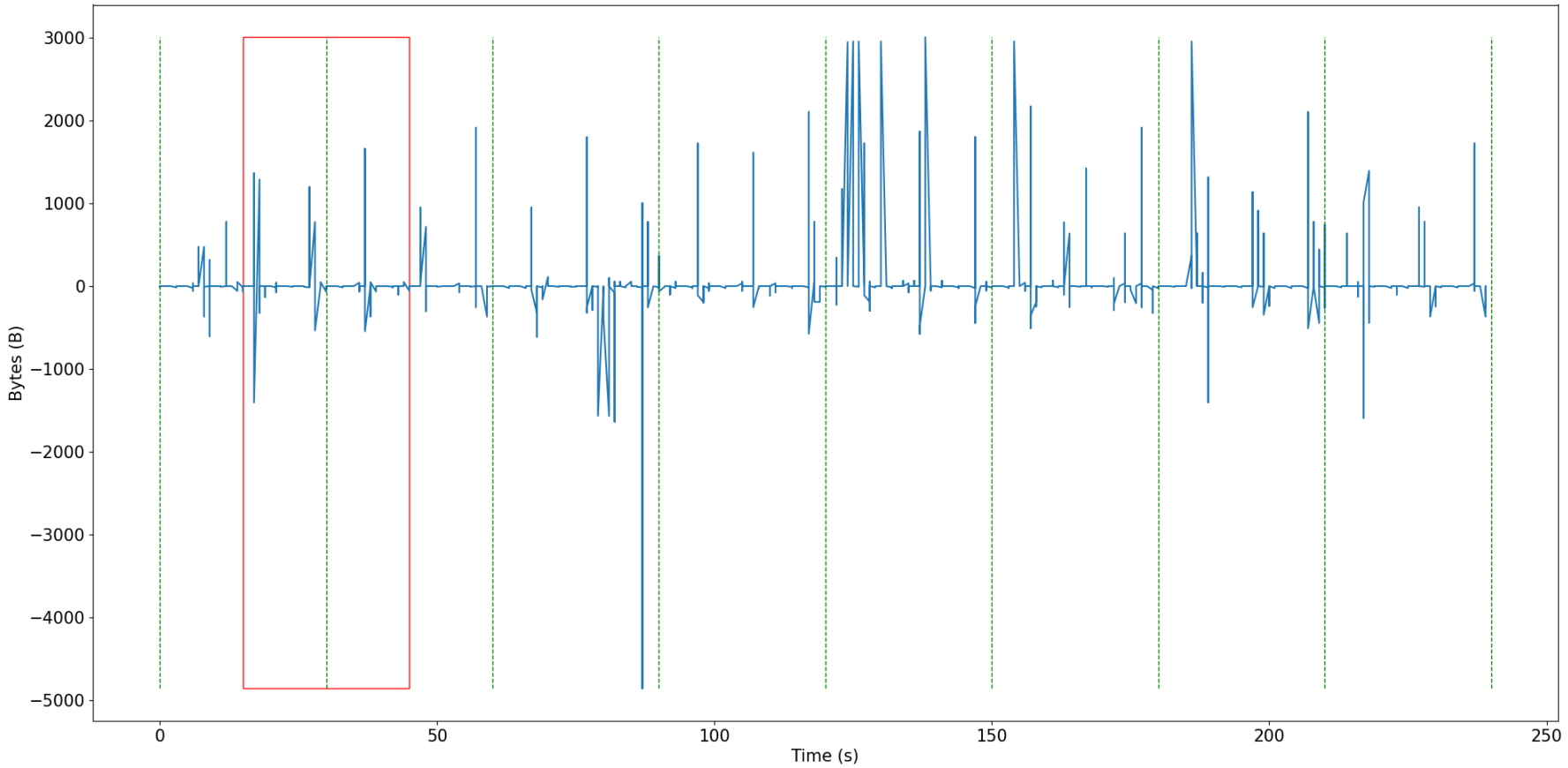
An Example - Echo Spot



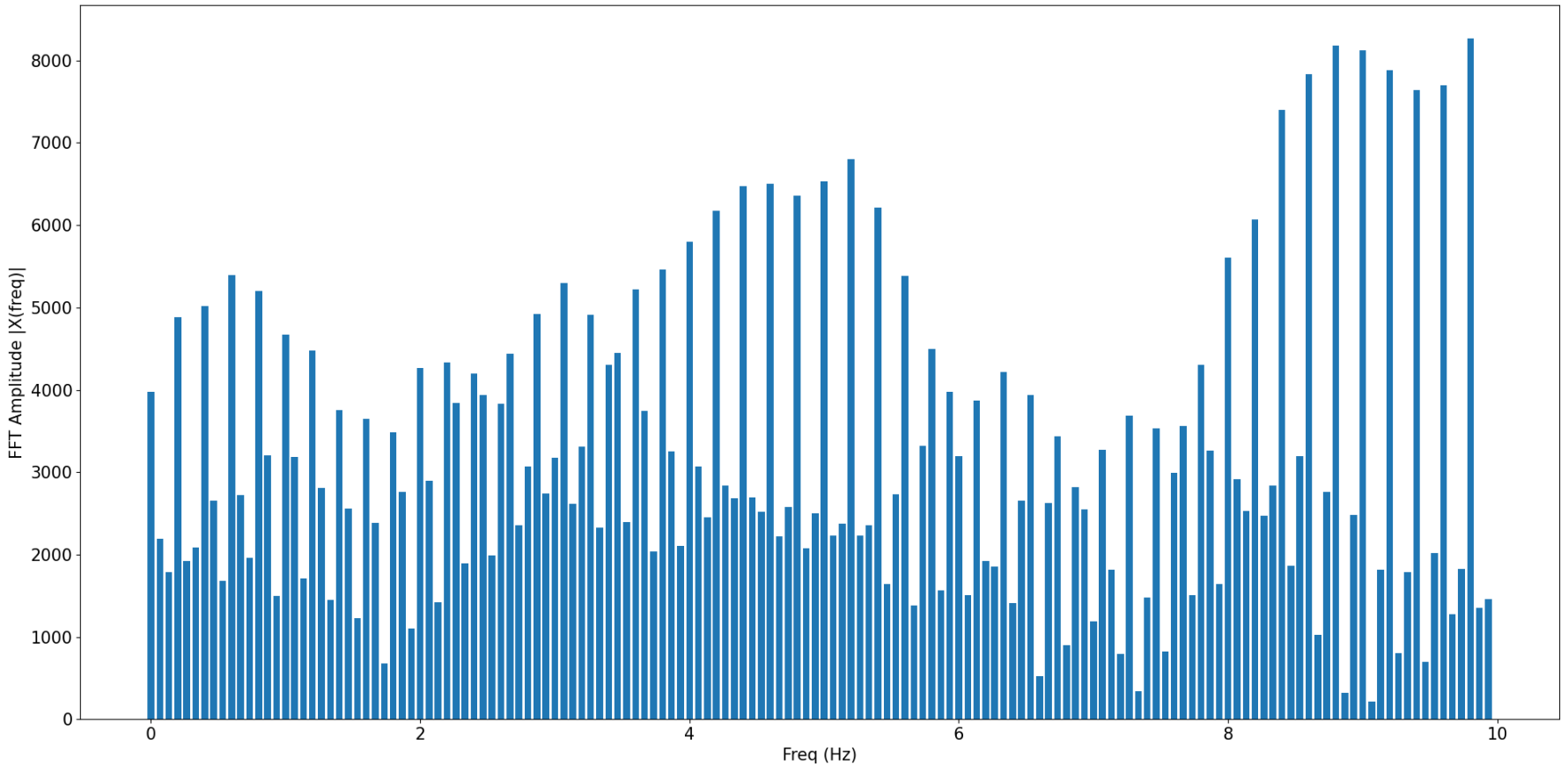
An Example - Echo Spot



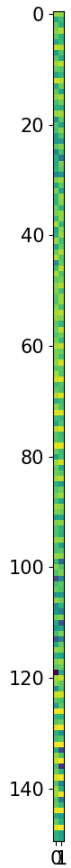
An Example - Echo Spot



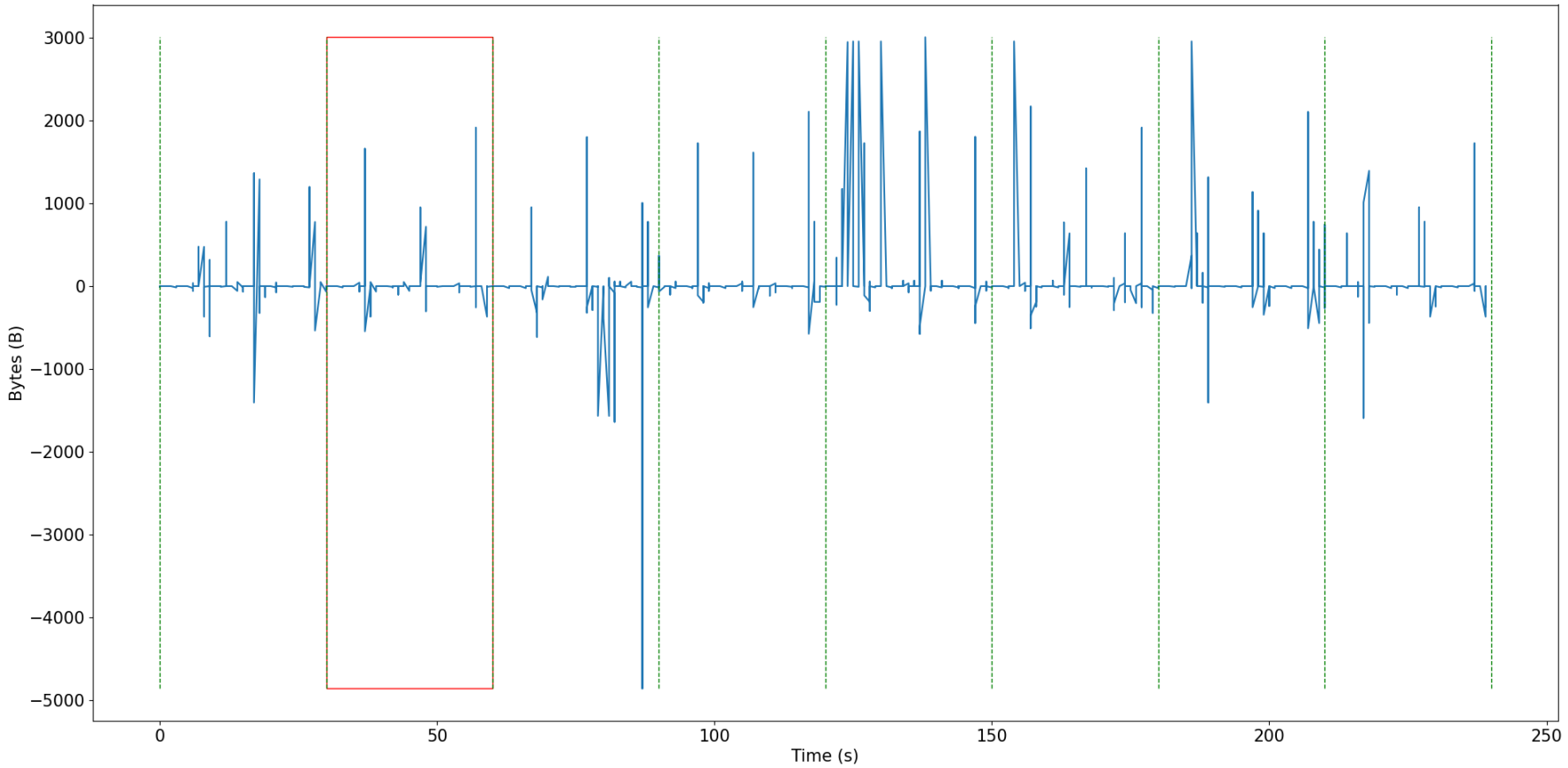
An Example - Echo Spot



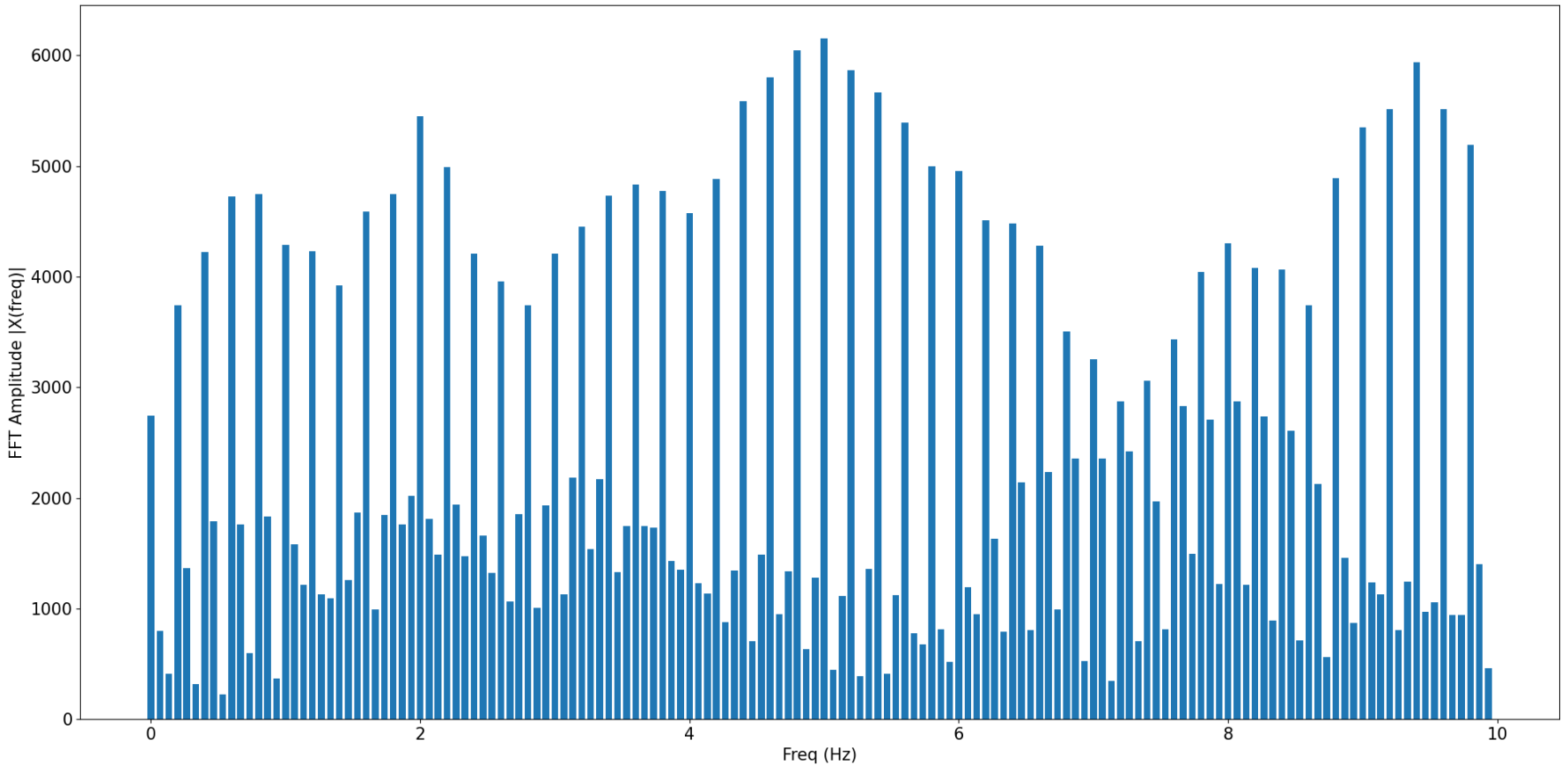
An Example - Echo Spot



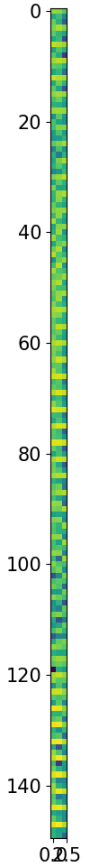
An Example - Echo Spot



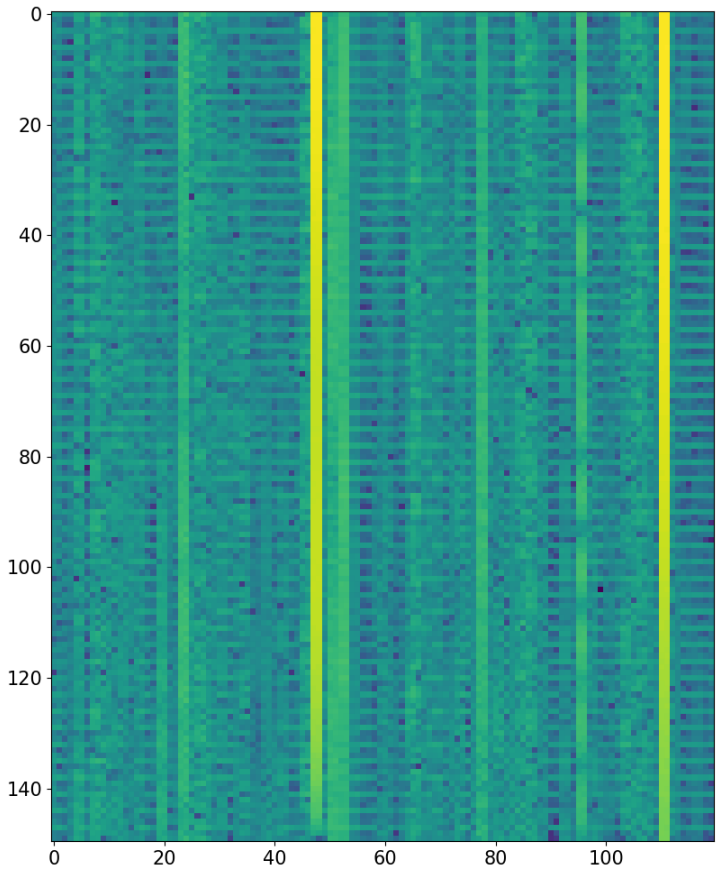
An Example - Echo Spot



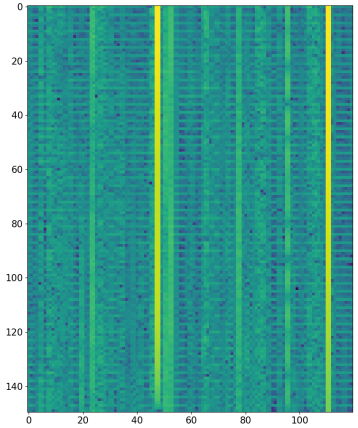
An Example - Echo Spot



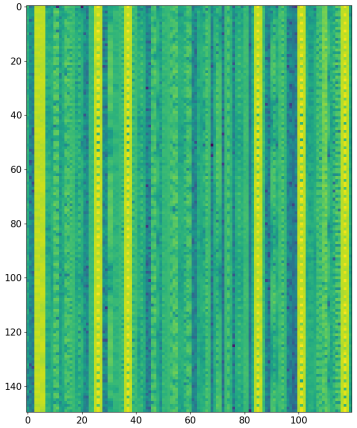
An Example - Echo Spot



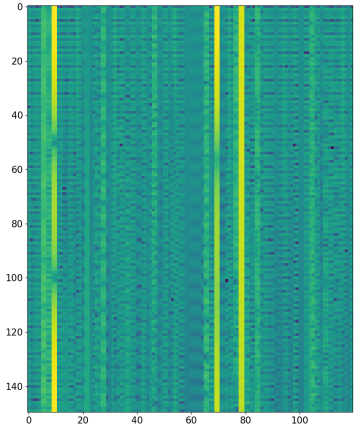
Other Examples



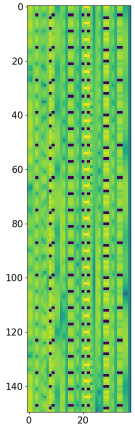
(a) Echo Spot



(b) Google Home



(c) Fire TV



(d) Ring Doorbell

Constants

We want to achieve near on-line device identification.

- Sampling frequency: 10 Hz and 2 Hz
- Number of samples: 300 and 150 (height of the input)
- Number of FFT parameters: 120 and 60 (width of the input)

Buffer time for input 150×120 is 30 minutes and new input arrives every 15 seconds.
Buffer time for input 75×60 is 7.5 minutes and new input arrives every 7.5 seconds.

Results - The Good, the Bad and the Ugly

- *Hihghest confidence (hc)* - positive classification if the model is the most confident amnogst all models.
- *Confident enough (ce)* - positive classification if the model is at least 50% confident.

Table: F_1 score of various classifiers. *hc* shows the *highest confidence* scenario while *ce* shows the *confident enough* scenario.

Input Size	75×60		75×120		150×60		150×120	
Prediction Window	hc	ce	hc	ce	hc	ce	hc	ce
0	0.90	0.94	0.91	0.95	0.91	0.94	0.93	0.95
1	0.72	0.78	0.75	0.80	0.74	0.78	0.76	0.80
2	0.47	0.55	0.50	0.56	0.50	0.56	0.54	0.59

Thank you

Contact:

roman.kolcun@cl.cam.ac.uk