Botnet Fingerprinting: a Frequency Distributions scheme for Lightweight Bot Detection

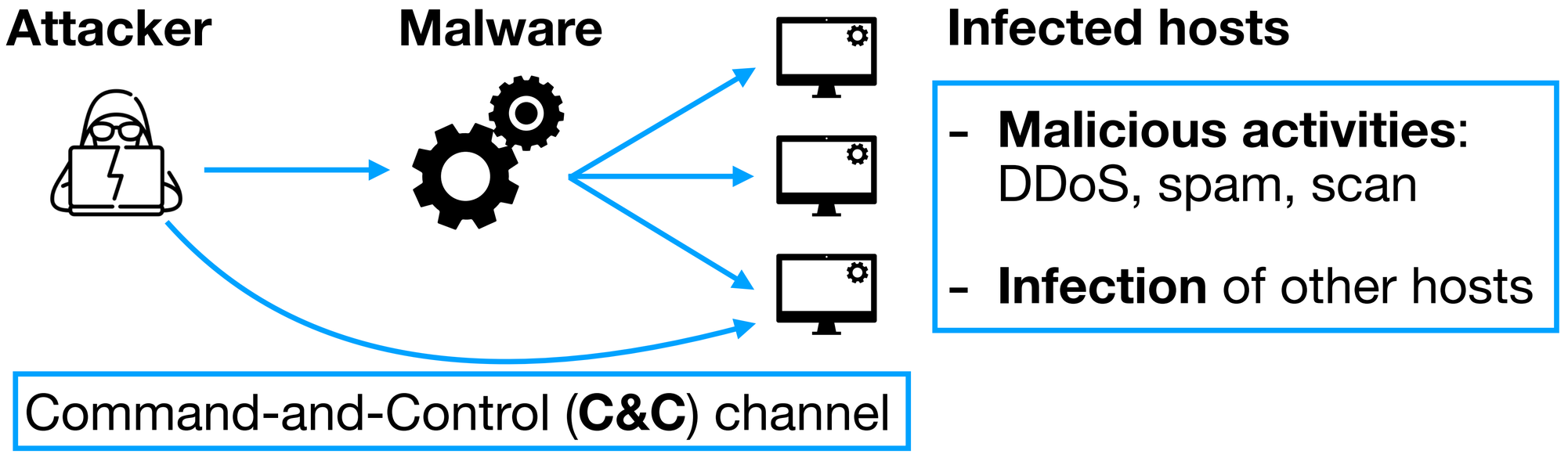
Agathe Blaise, Mathieu Bouet, Vania Conan, Stefano Secci

Journées RESSI - 16 décembre 2020
Session thèses

THALES

SORBONNE UNIVERSITÉ

Botnet architecture



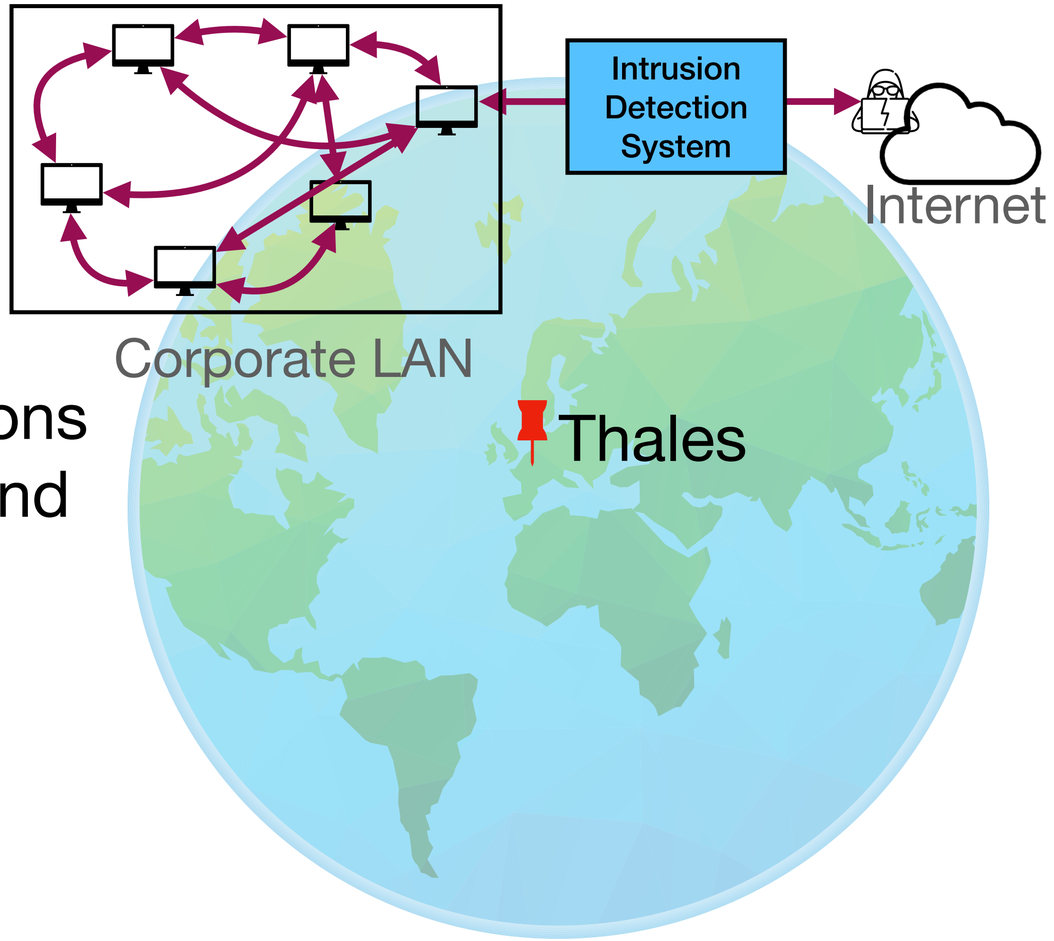
→ Need to identify **communication patterns specific to a bot.**

BotFingerPrinting

Challenge: **botnet detection** within LAN

- Flow-based approaches: miss communications patterns
- Graph-based approaches: not scaling

Our approach: simplify the communications graphs through histograms about hosts and services contacted



Our contributions:

- Very high accuracy compared to SOTA
- Lightweight compared to graph-based approaches

CTU-13 dataset (2011)

13 botnet scenarios: training and test (*) sets

Id	#bots	Malware	Activity
1*	1	Neris	IRC, SPAM, CF
2*	1	Neris	IRC, SPAM, CF
3	1	Rbot	IRC, PS
4	1	Rbot	IRC, DDoS
5	1	Virut	SPAM, PS
6*	1	Menti	PS
7	1	Sogou	HTTP
8*	1	Murlo	PS
9*	10	Neris	IRC, SPAM, CF, PS
10	10	Rbot	IRC, DDoS
11	3	Rbot	IRC, DDoS
12	3	NSIS.ay	IRC, P2P
13	1	Virut	HTTP, SPAM, PS

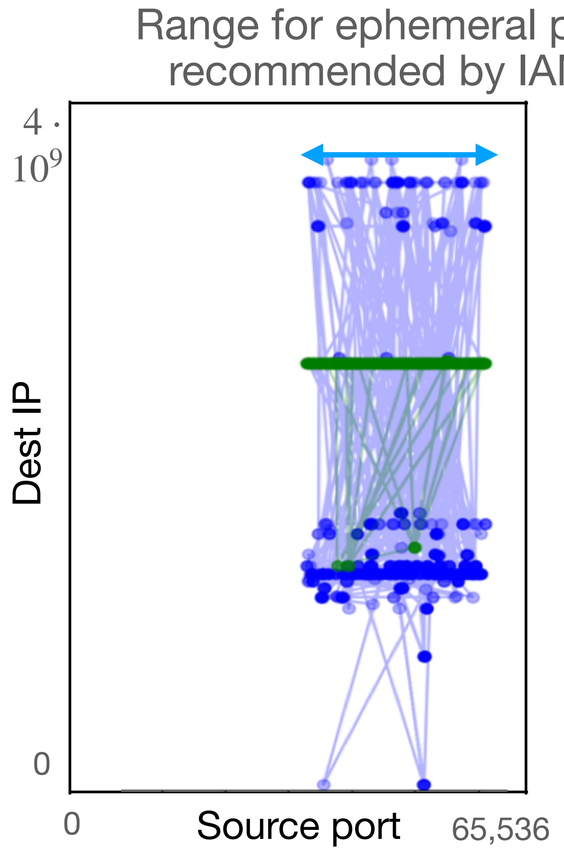
- **C&C channels**: IRC, HTTP, P2P
- **Malicious activities**: DDoS, port scan, spam, click fraud

→ **Objective**: learn from training set and perform the detection on test set.

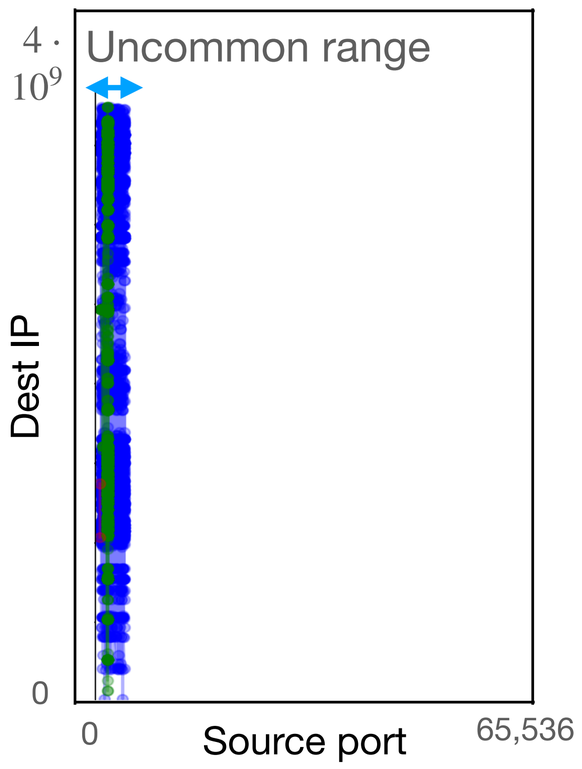
First observations on CTU-13

Inspecting the communications of two different hosts

Benign host



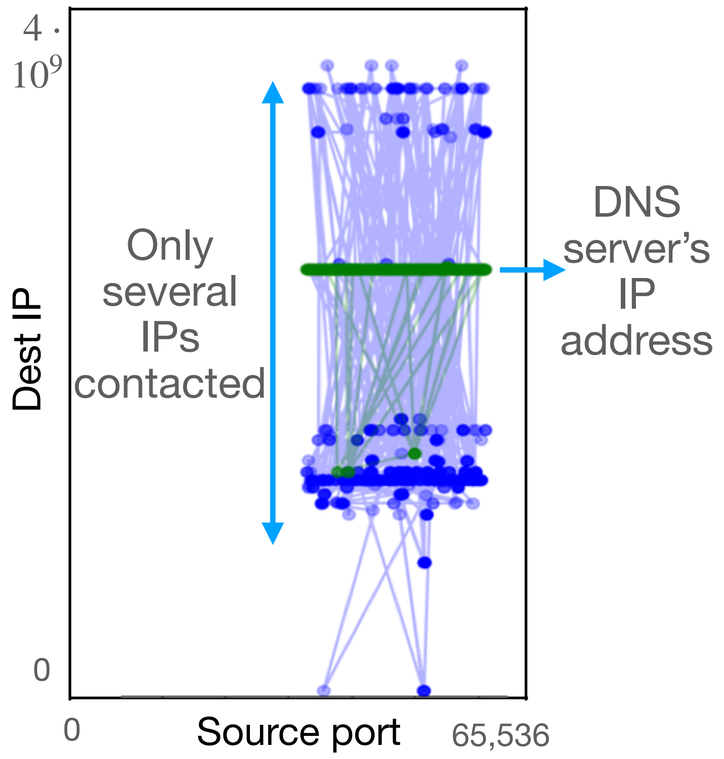
Infected host (bot)



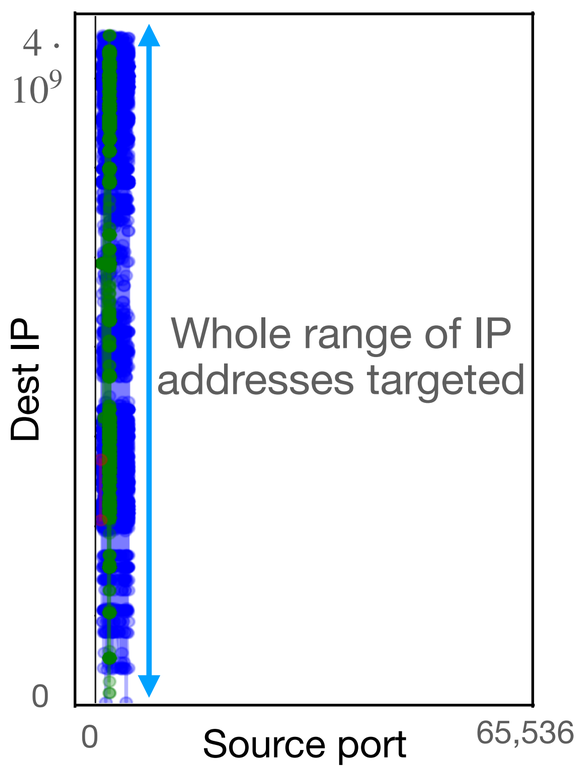
First observations on CTU-13

Inspecting the communications of two different hosts

Benign host



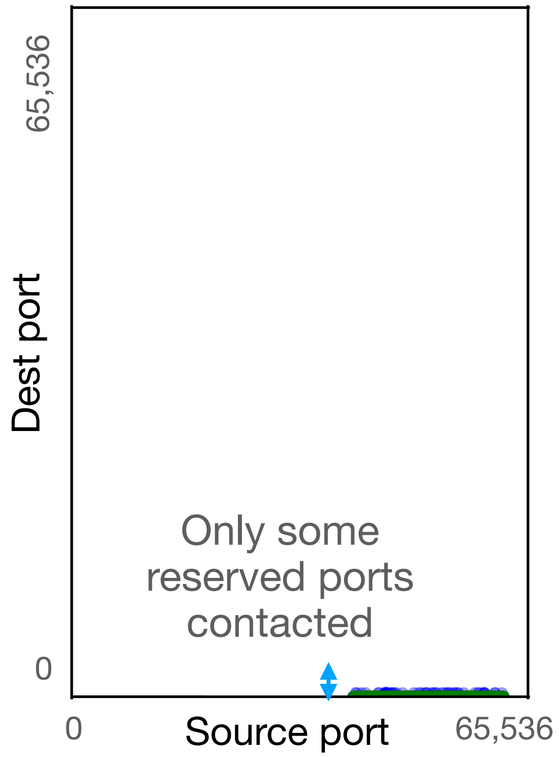
Infected host (bot)



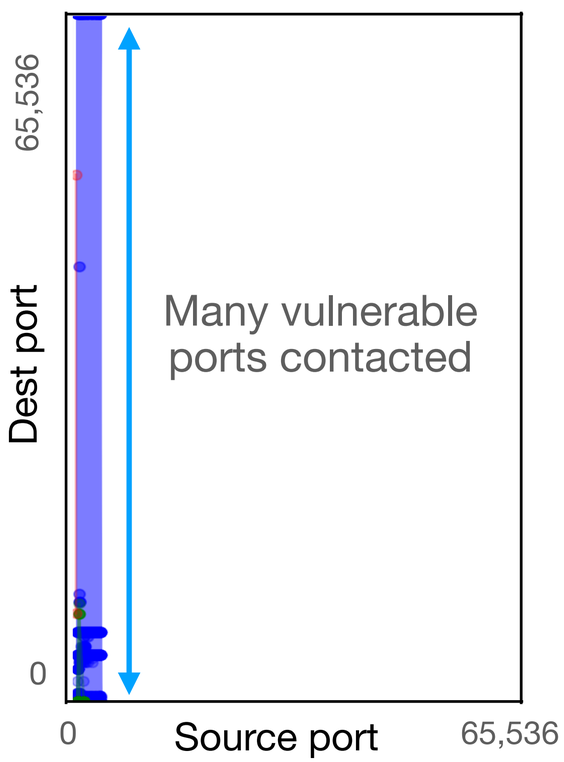
First observations on CTU-13

Inspecting the communications of two different hosts

Benign host



Infected host (bot)



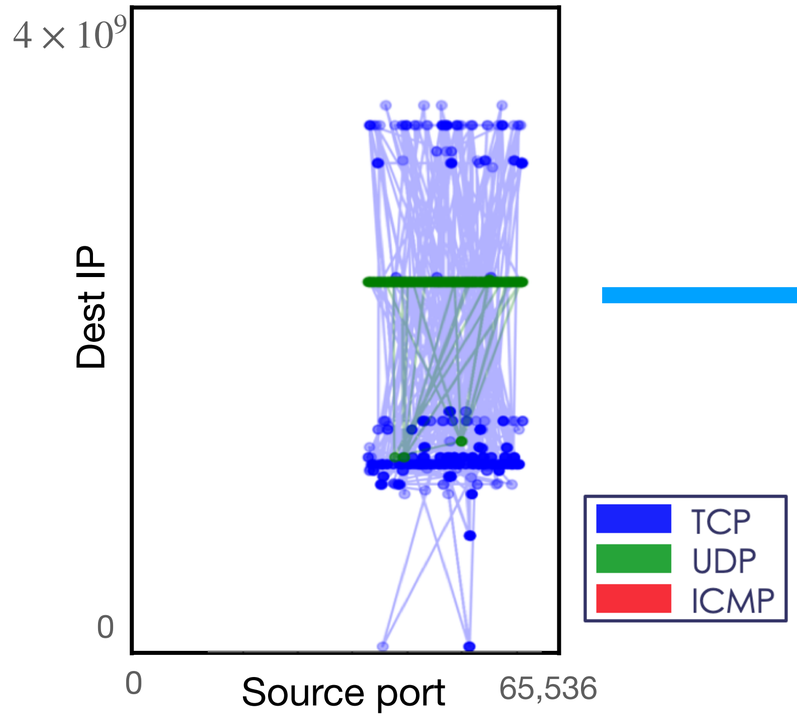
Frequency distribution of protocol uses

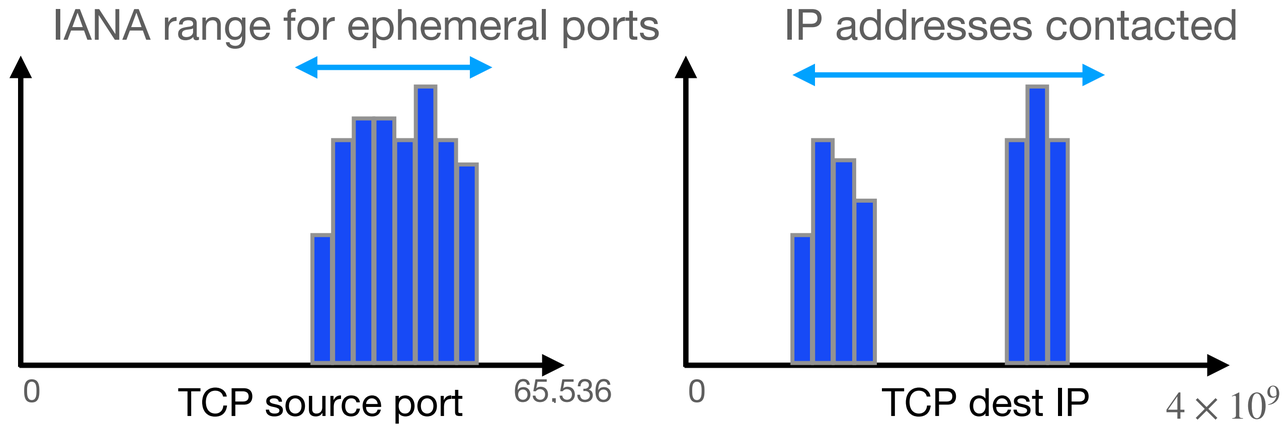
Host signature: **concatenation** of the **frequency distributions** of the 9 features:

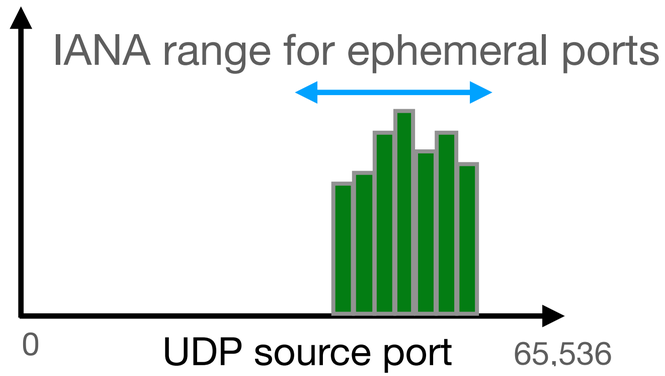
9 features from the combination of:

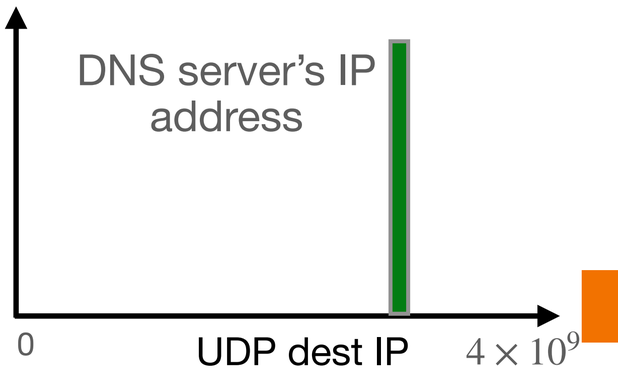
- TCP
- UDP
- ICMP
- Source port
- Destination port
- Destination IP address

Example for a **benign host:**







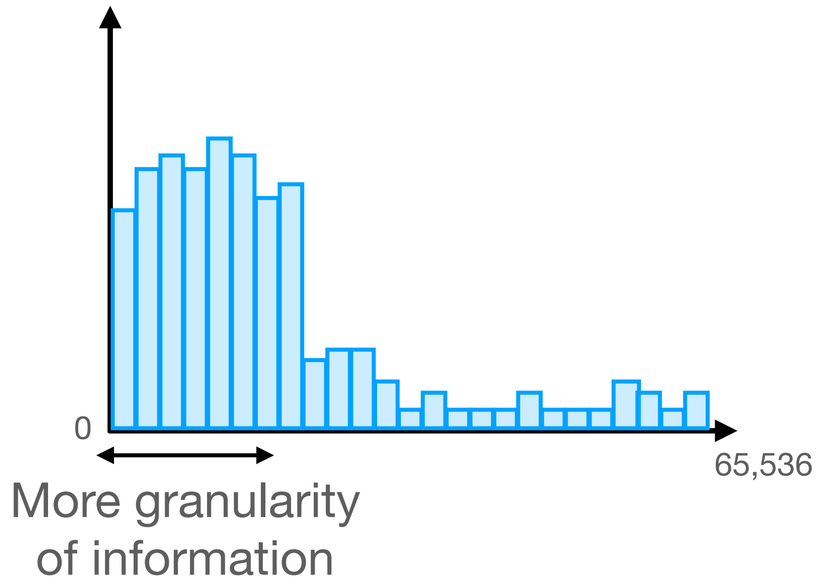


Quantisation bin

Example for a **Destination port TCP:**

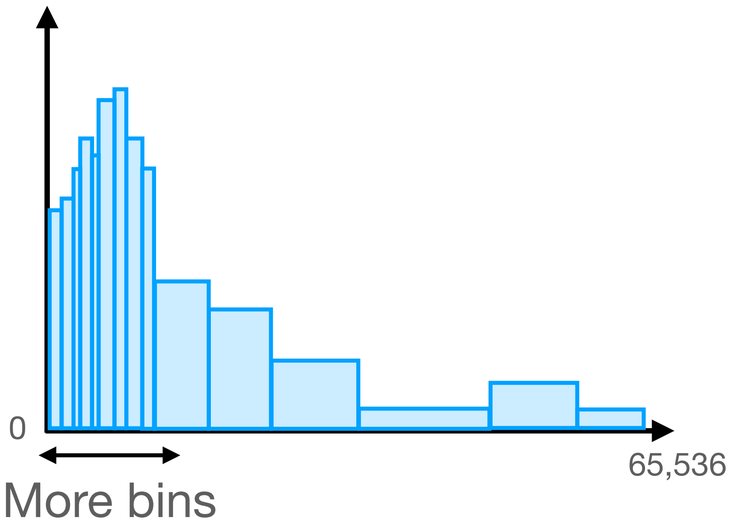
Regular bins

Bins of equal width

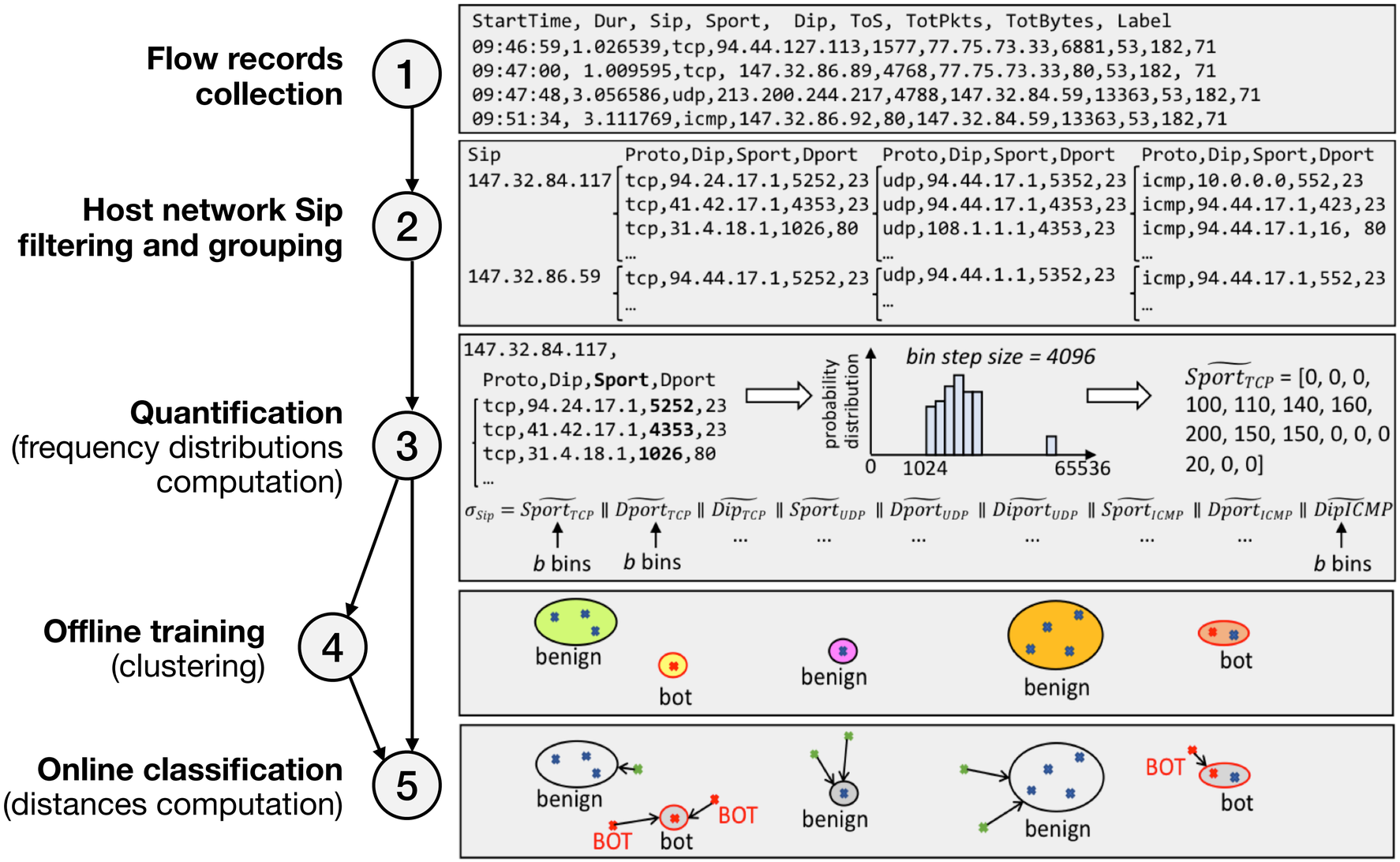


Adaptive bins

Bins width adapted to the density of information



Our general approach: BotFingerPrinting



Evaluation

- Tuning depending on the objective(s) to favour
  - Maximising the true bot detection
  - Minimising the false positive rate
  - Minimising the memory usage

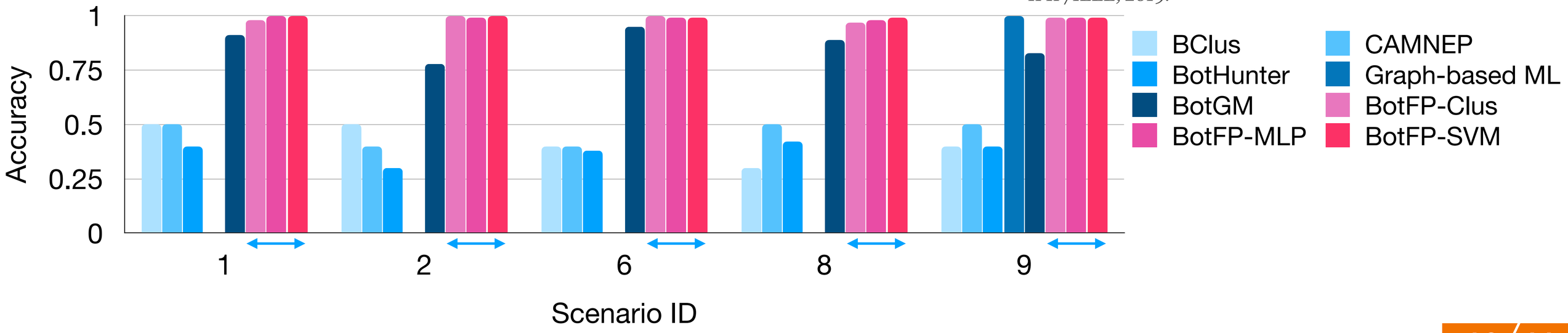
"An empirical comparison of botnet detection methods," Computers & Security, 2014.

"BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation," Usenix Security Symposium, 2007.

"BotGM: Unsupervised Graph Mining to Detect Botnets in Traffic Flows," CSNet, 2019.

"A Graph-Based Machine Learning Approach for Bot Detection," IFIP/IEEE, 2019.

Accuracy of state-of-the-art techniques and BotFP



→ BotFP accuracy **between 97 and 100%**

Conclusion

- **Histograms approximate the relations between hosts**
- Far **more lightweight** and **more efficient** than graph-based approaches
  - Very high accuracy (from 97 to 100%), outperforming other state-of-the-art techniques
  - Nearly all bots detected with very few false positives

Perspectives

- Explore unsupervised learning techniques

Thank you