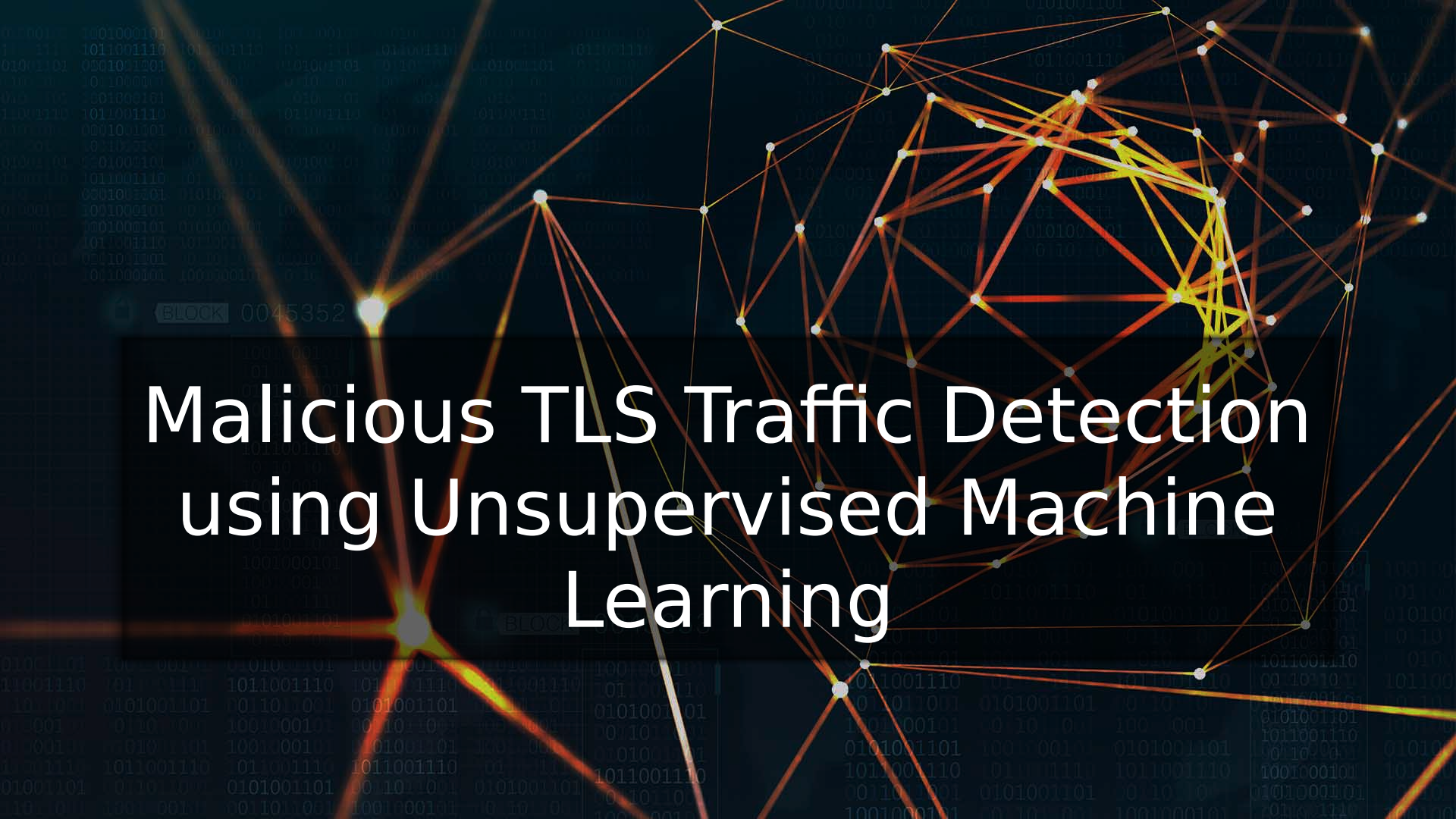


Malicious TLS Traffic Detection using Unsupervised Machine Learning

Encrypted C&C Channel

Malware coordinates through C&C:

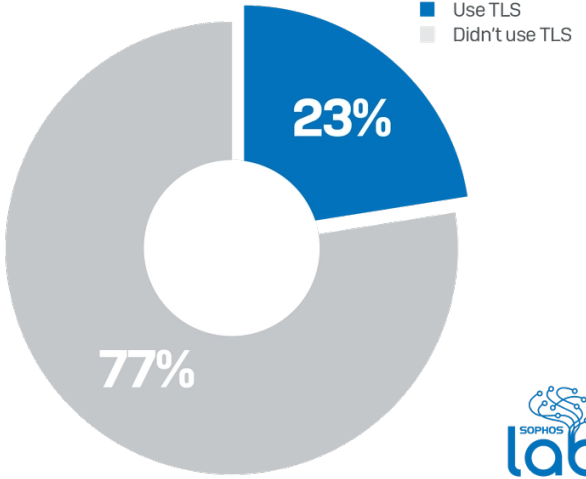
- IRC, XMPP, SMTP, HTTP
- Plain-text protocols

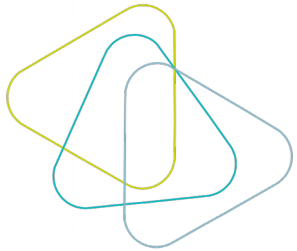
Next level: obfuscate communication

Malware + TLS

Malware usage of TLS:

- TLS is a standard protocol
- From 10% in 2016, to 23% in 2020
- HTTPS dominates





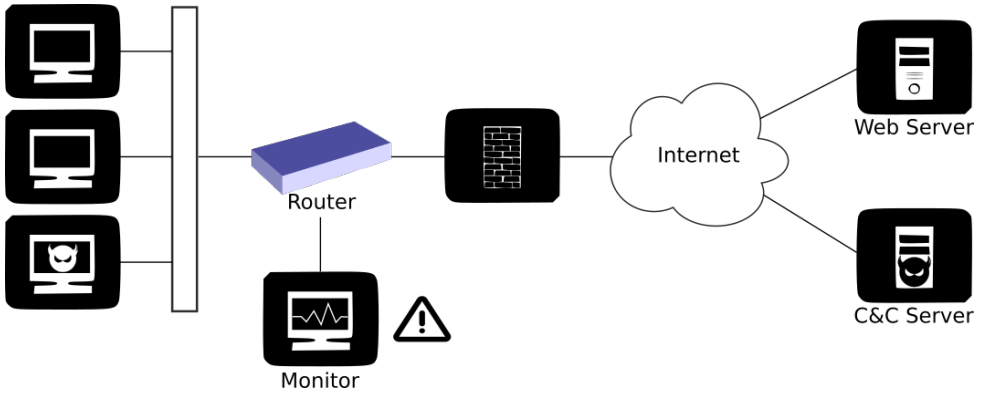
C&C Detection

Plain text:

- Content Signatures (CS)
- Deep Packet Inspection (DPI)

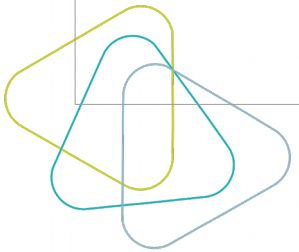
Encrypted:

- Man in the Middle (MITM)



State-of-the-art

Client Hello Fingerprints	Lots of FP	Unsupervised ML	Use of TLS Handshake + Payload features
Supervised ML	Binary very challenging		Different types of traffic
	Multi-class needs labeled data		No labels needed
	Features from unencrypted protocols undermine privacy		TLS only
	Not tested on TLS v1.3		Tested on TLS v1.3

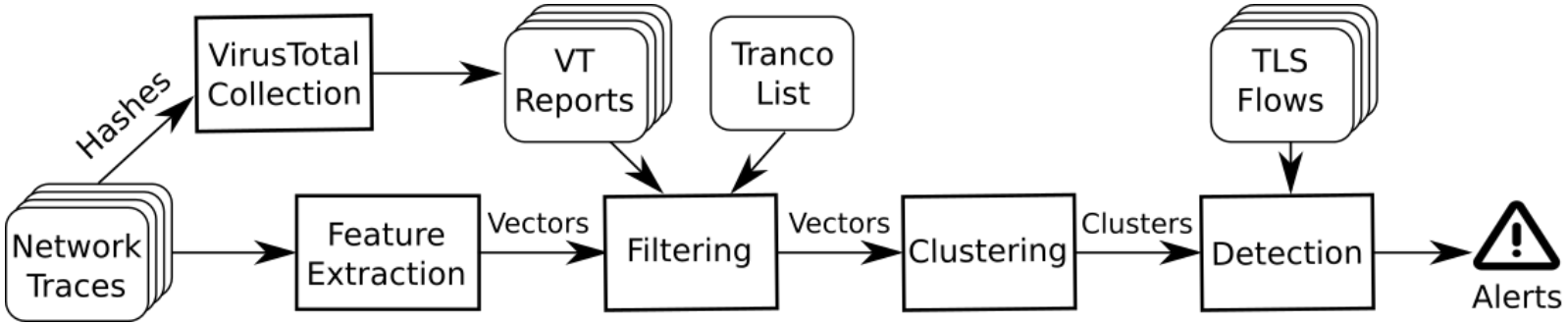


Contributions

- Unsupervised classifier
- Privacy aware
- Sandbox analysis from 972k samples
- A model with FDR of 0.03%
- TLS v1.3 clusters


Introduction
Approach
Evaluation

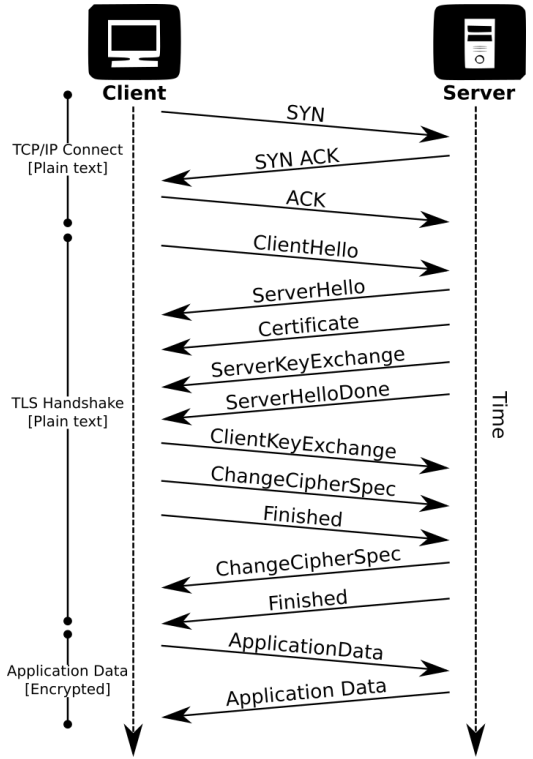
Architecture



Feature Extraction

91 TLS features (50 new):

- Client features (Client Hello)
- Server features (Server Hello)
- Certificate features (Certificate)
- Payload features (Encrypted Application Data)



Filtering

- Flows without encrypted data:
  - Non-established TLS flows
  - Flows without application data
- Benign traffic (VT, Tranco)
  - Not malware samples
  - Background traffic
  - Connectivity tests
- Vanilla Tor

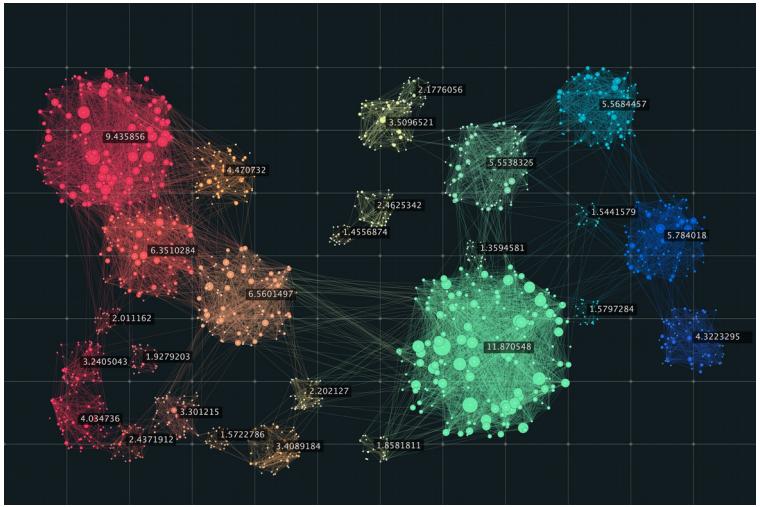
Clustering

Group similar feature vectors:

- Flows from different samples
- Same sample, different clusters

Algorithms:

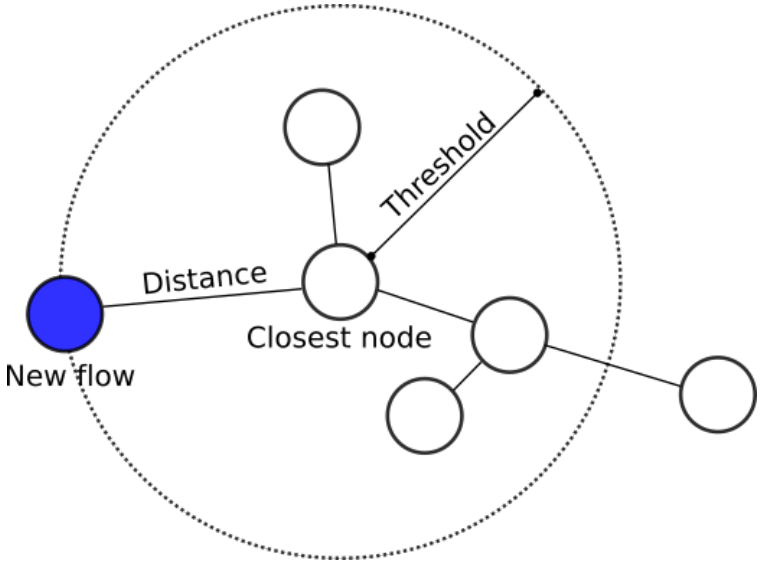
- MeanShift
- FISHDBC



Detection

Decide if a flow belongs to a cluster:

- Search its closest node
- Distance below threshold: malicious
- Otherwise, benign.




Introduction
Approach
Evaluation

Datasets

Malware traces:

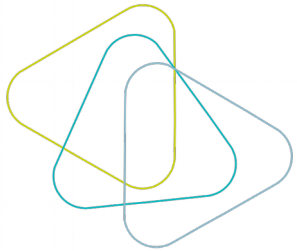
- Samples: 972.6K
- Flows: 12.9M
- 2017-2019

Ground truth:

- Manually labeled subset (29 clusters)
- 41k flows, ~28K samples

Benign traces:

- Flows: 34.4M
- 2019-2020

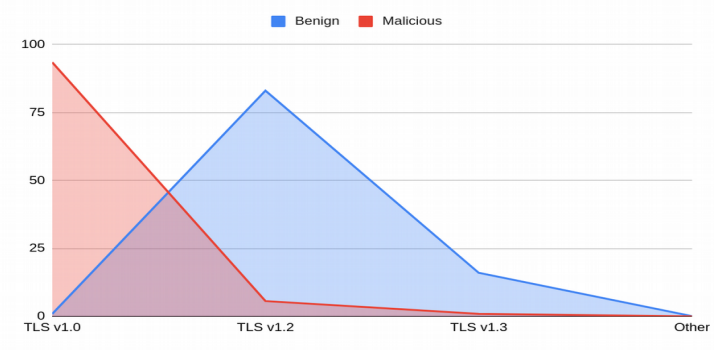


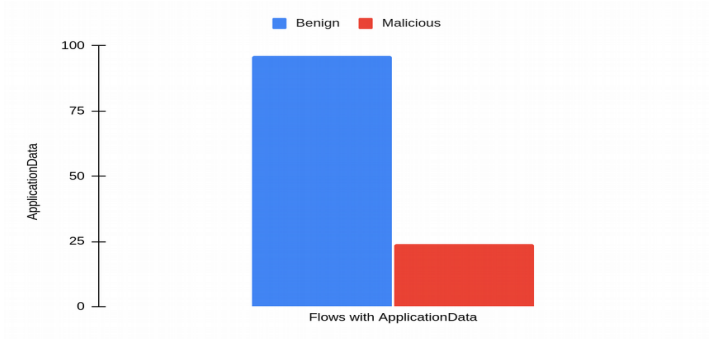
Data Analysis

Significant differences between both datasets:

- TLS version
- Number of flows with Application Data packets

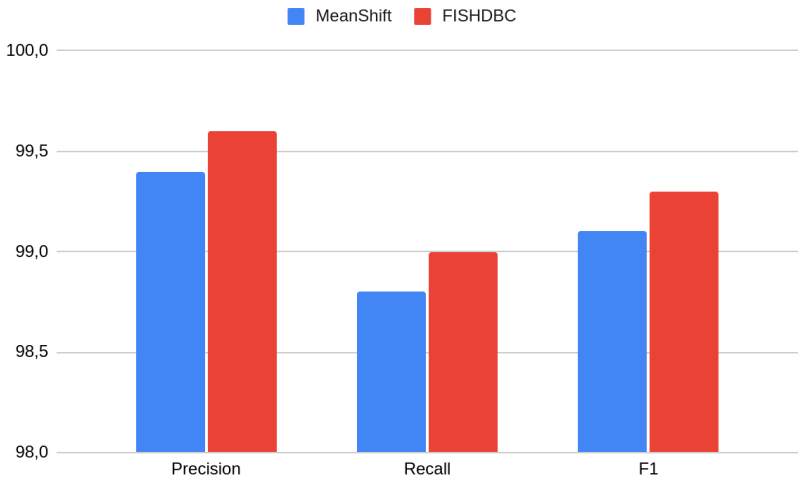
Differences rooted in the sandbox (Windows 7).





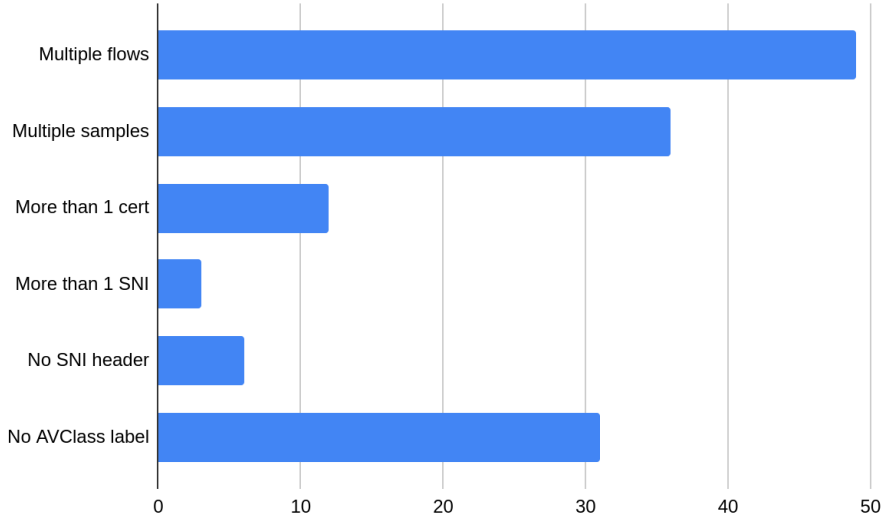
Clustering Results

- FISHDBC achieves the best results:
  - Precision: 99.6%
  - Recall: 99.0%
  - F1: 99.3%
- Server and Payload features provide most information.
- Certificate features are not useful.



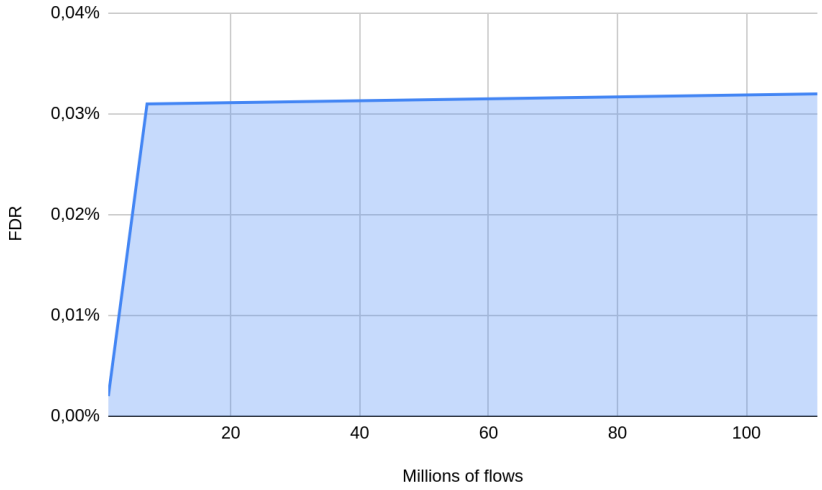
Cluster Analysis

- Multiple flows: 49%
- Multiple samples: 36%
- Certificate polymorphism: 12%
- Domain polymorphism: 3%
- Clusters without SNI: 6%
- Unlabeled clusters: 31%
- TLS 1.3: 50 clusters (~7K samples)



False Detection Rate

- One day (95K flows): 0.002%
- One week (13.2M flows): 0.031%
- Four months (24.8M flows): 0.032%



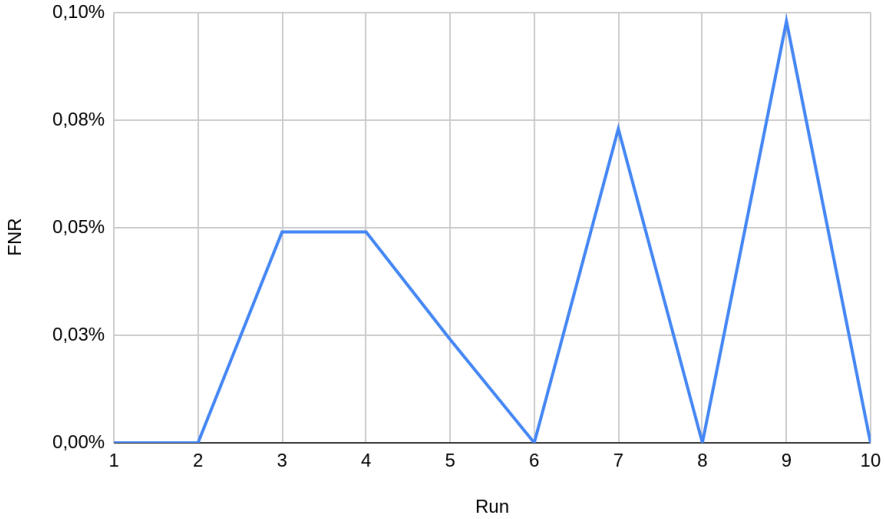
False Negative Rate

On the ground truth:

- FNR: 0.029%
- TPR: 99.97%

On malicious data:

- FNR: 0.054%
- TPR: 99.46%




Thanks For Your Time!