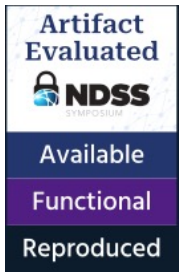


MineShark: Cryptomining Traffic Detection at Scale

Shaoke Xi, Tianyi Fu, Kai Bu, Chunling Yang, Zhihua Chang, Wenzhi Chen, Zhou Ma, Chongjie Chen, Yongsheng Shen, Kui Ren

The Rapid Rise of Cryptojacking Attacks

LILY HAY NEWMAN SECURITY FEB 20, 2018 5:06 PM

Hack Brief: Hackers Enlisted Tesla's Public Cloud to Mine Cryptocurrency

The recent rash of cryptojacking attacks has hit a Tesla database that contained potentially sensitive information.

New Cryptojacking Worm Found in Docker Containers

A new cryptojacking worm, named Graboid, has been spread into more than 2,000 Docker hosts, according to the Unit 42 researchers from Palo Alto Networks. This is the first time such a piece of malware has spread via containers within the Docker Engine (specifically docker-ce).

Oct 22nd, 2019 1:48pm by Jack Wallen

Detect large-scale cryptocurrency mining attack against Kubernetes clusters

By Yossi Weizman, Security Research Software Engineer, Azure Security Center

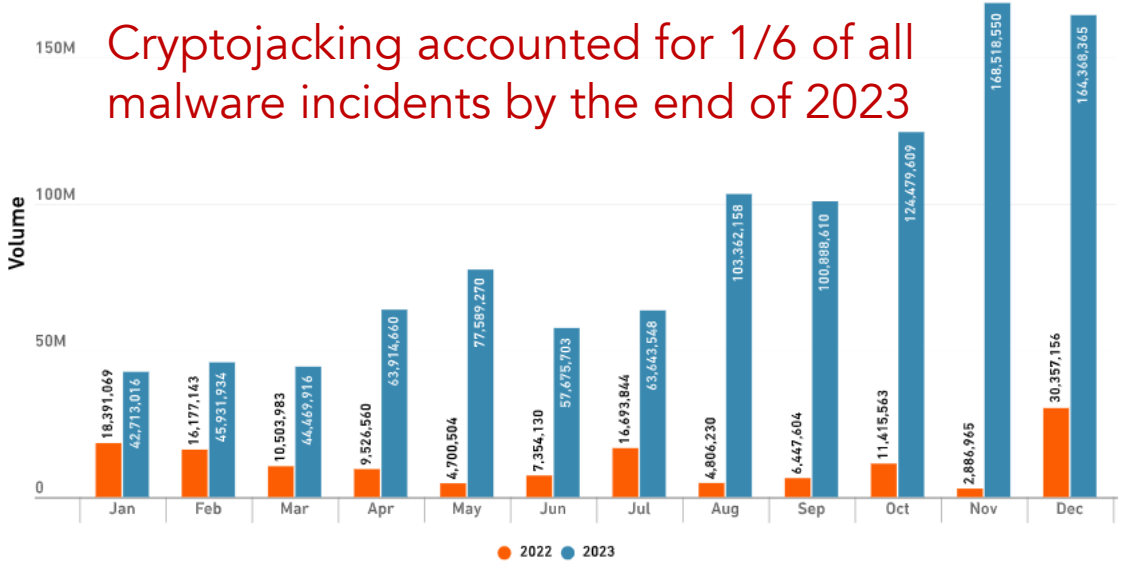
Docker Images Containing Cryptojacking Malware Distributed Hub

Jun 25, 2020 Ravie Lakshmanan

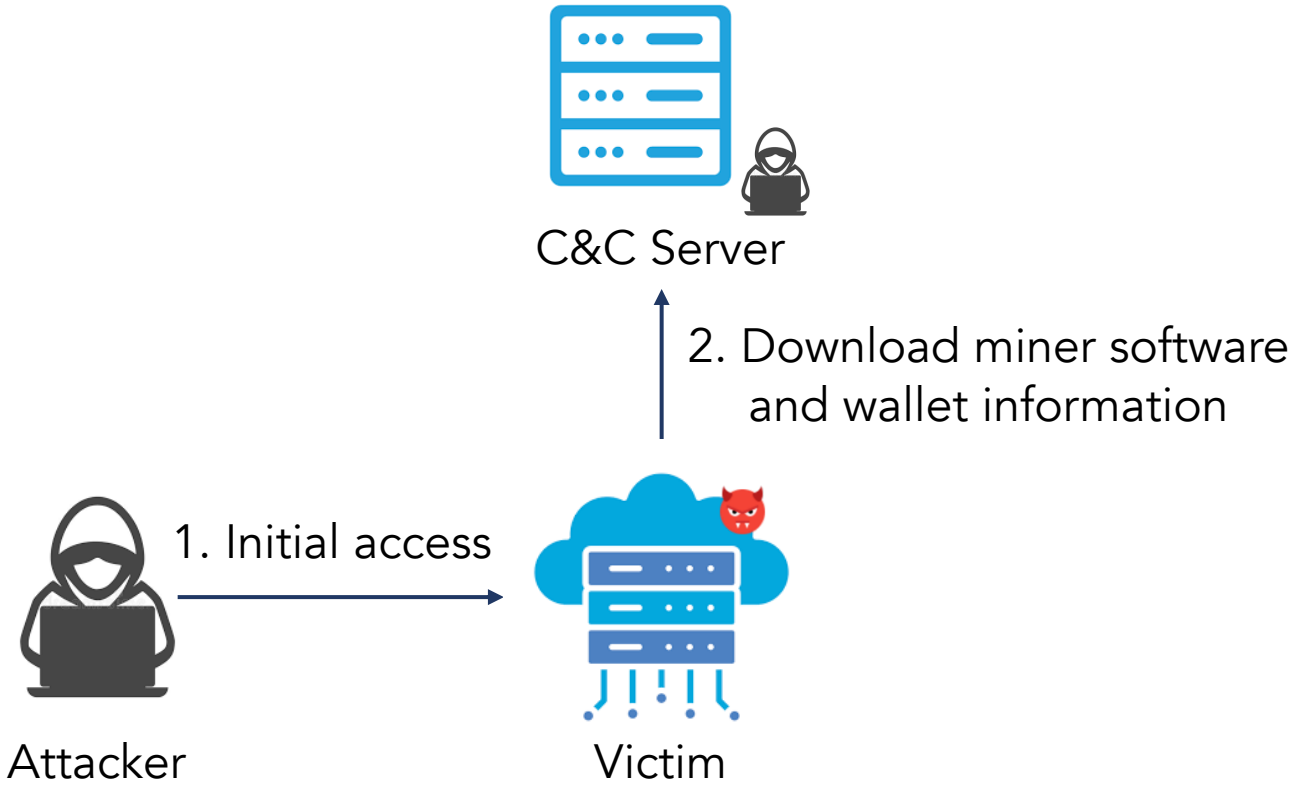
Cryptojacking via CVE-2023-22527: Dissecting a Full-Scale Cryptomining Ecosystem

A technical analysis on how CVE-2023-22527 can be exploited by malicious actors for cryptojacking attacks that can spread across the victim'

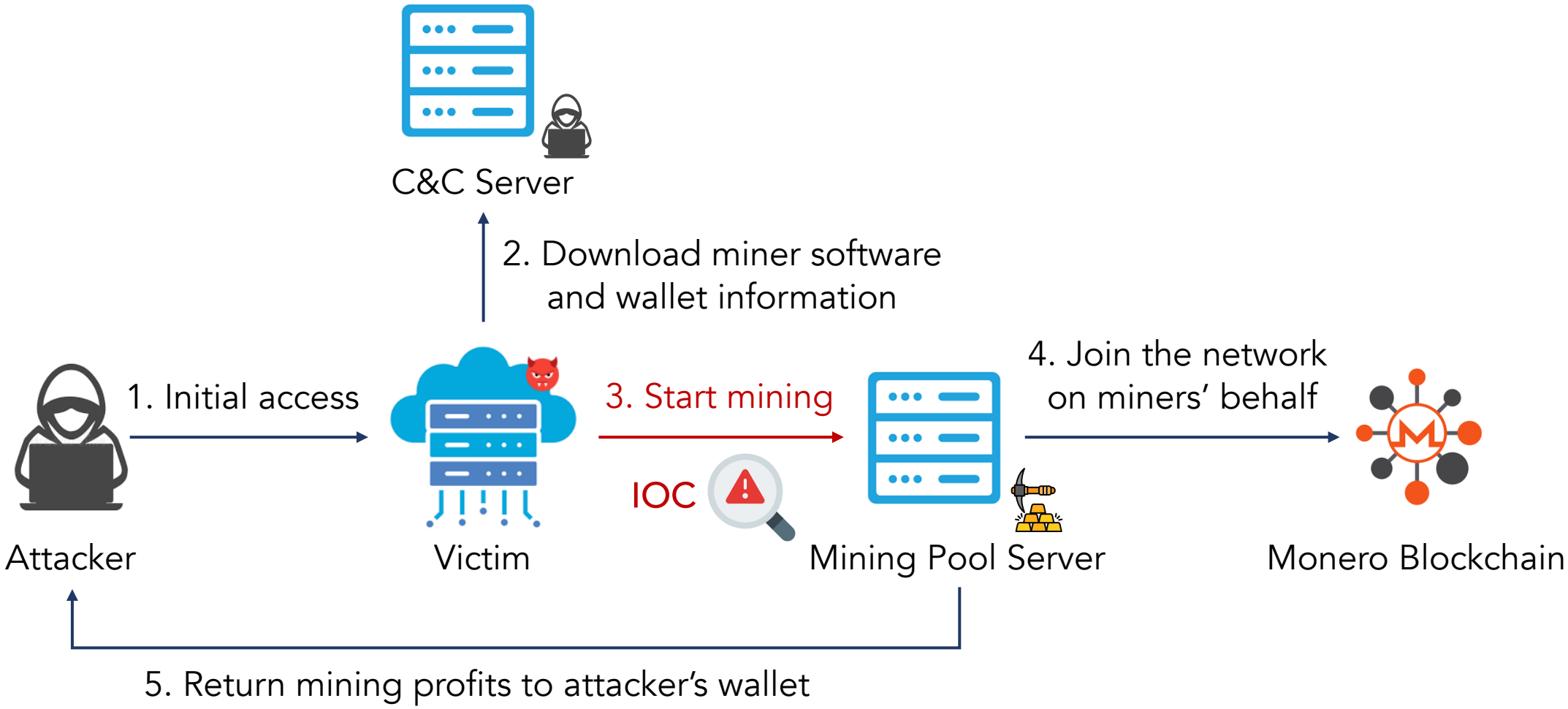
Cryptojacking accounted for 1/6 of all malware incidents by the end of 2023



Malware Adopts Pool Mining for Profit



Malware Adopts Pool Mining for Profit



IOC: Indicators of Compromise

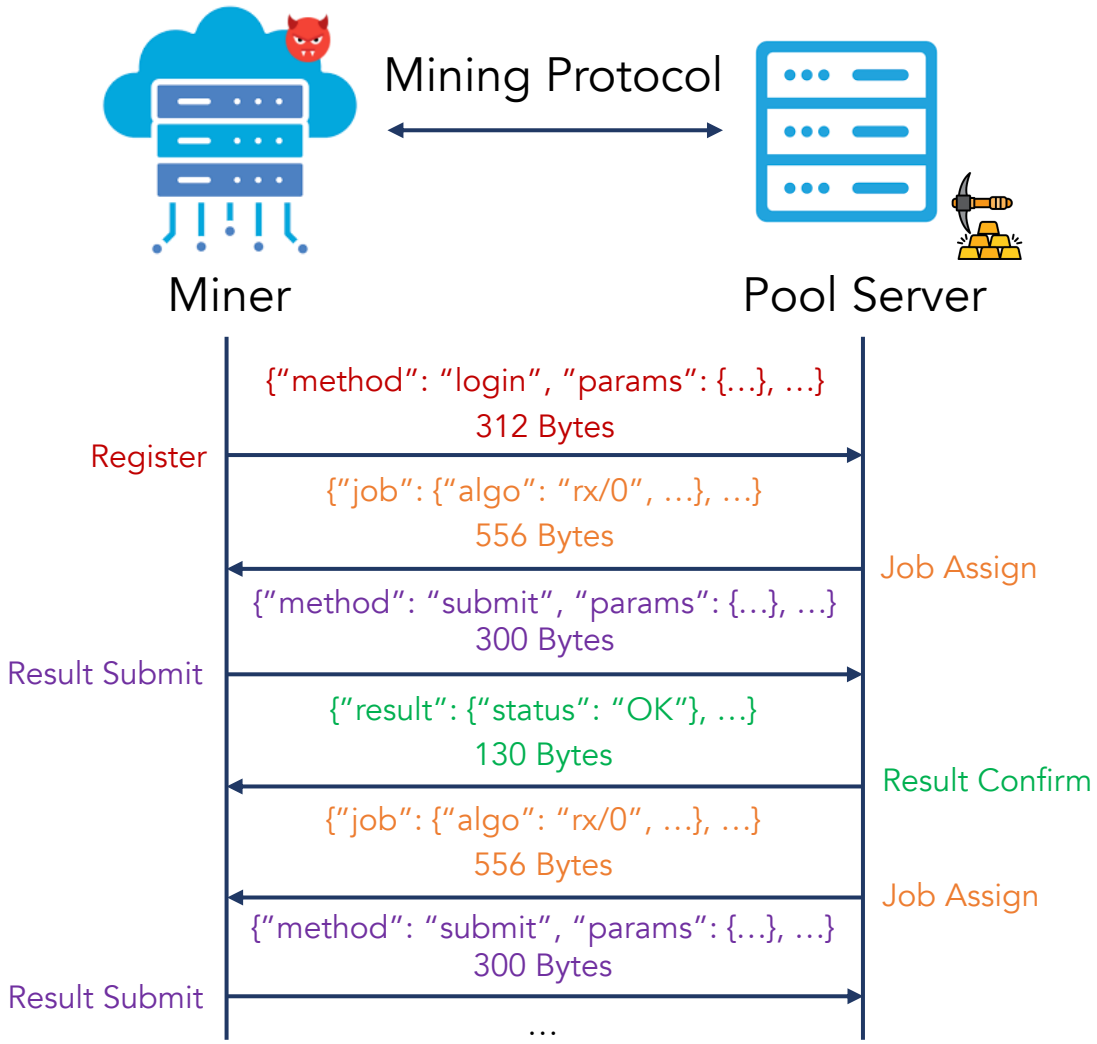
Stealthy Pool Mining Behaviors

- 51.5% of mining pools use TLS encryption.[1]
  - Helps avoid detection by deep packet inspection (DPI) methods.

- Nearly half of mining pool domains expire within 10 days.[1]
  - Among those, 33% expire within just 1 day.
  - Makes firewall deny lists quickly ineffective.

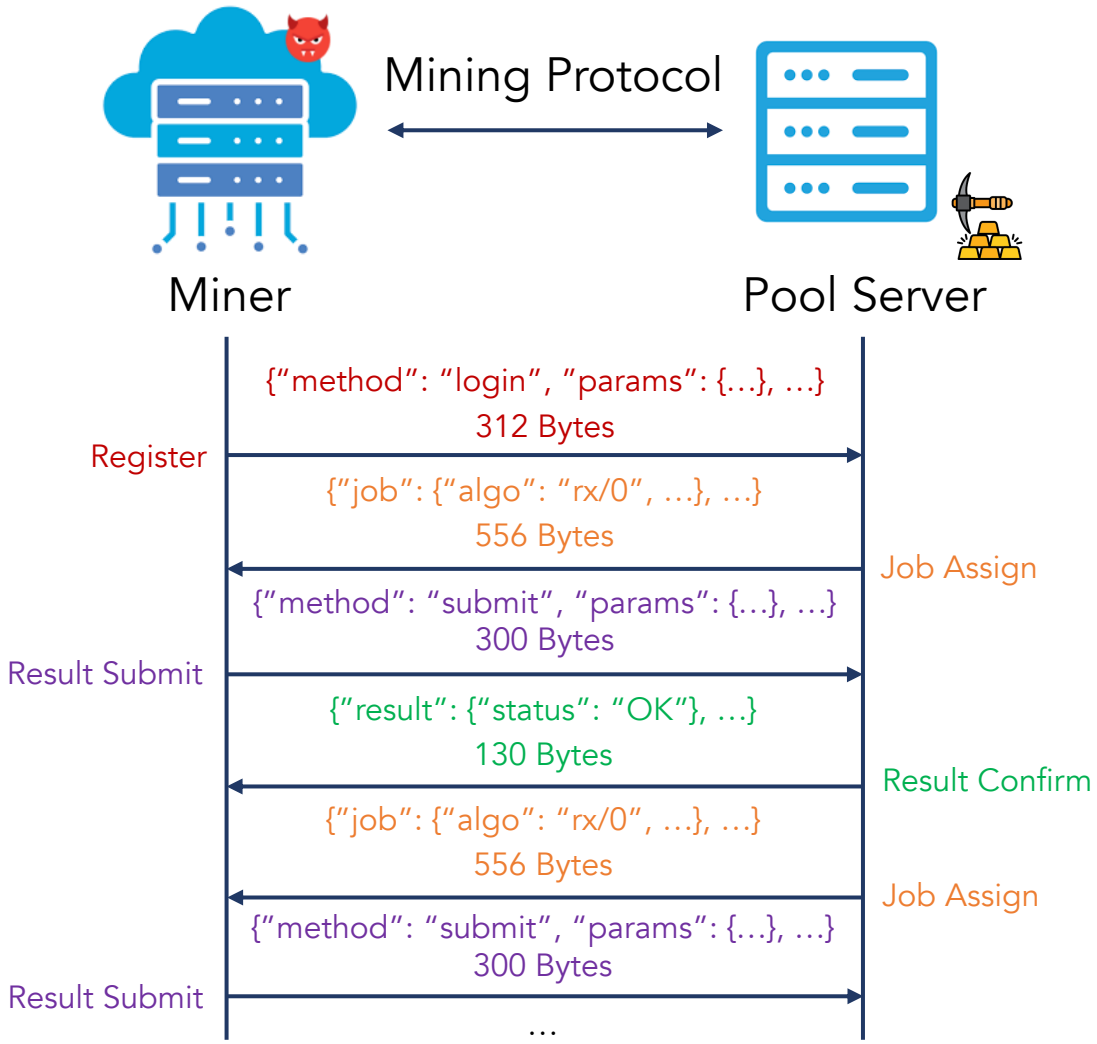
[1] Under the Dark: A Systematical Study of Stealthy Mining Pools (Ab)use in the Wild (CCS'23)

Using Machine Learning for Detection



- Identificable Traffic Features

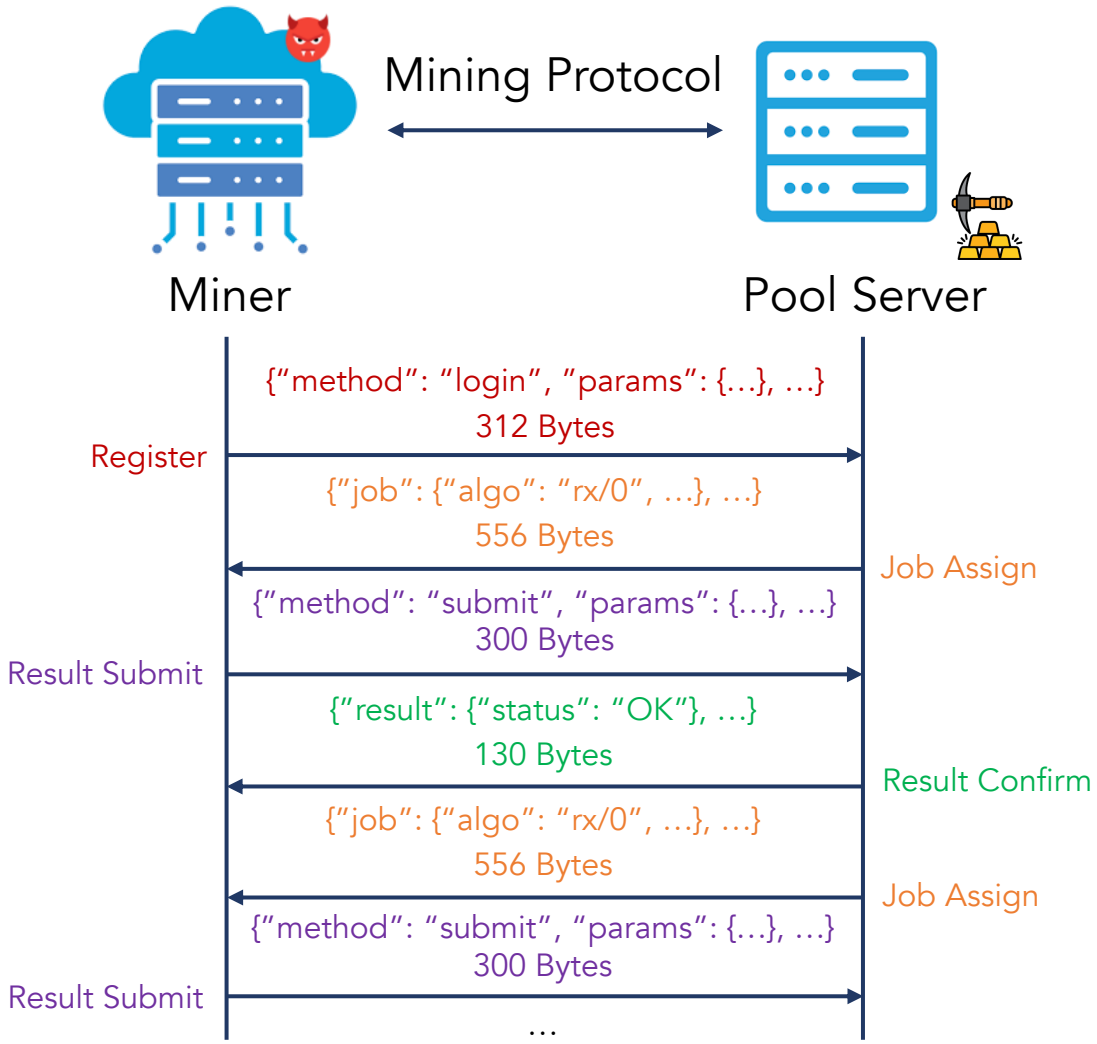
Using Machine Learning for Detection



- Identificable Traffic Features
  - Unique Packet Sizes

Different semantic messages exhibit unique sizes.

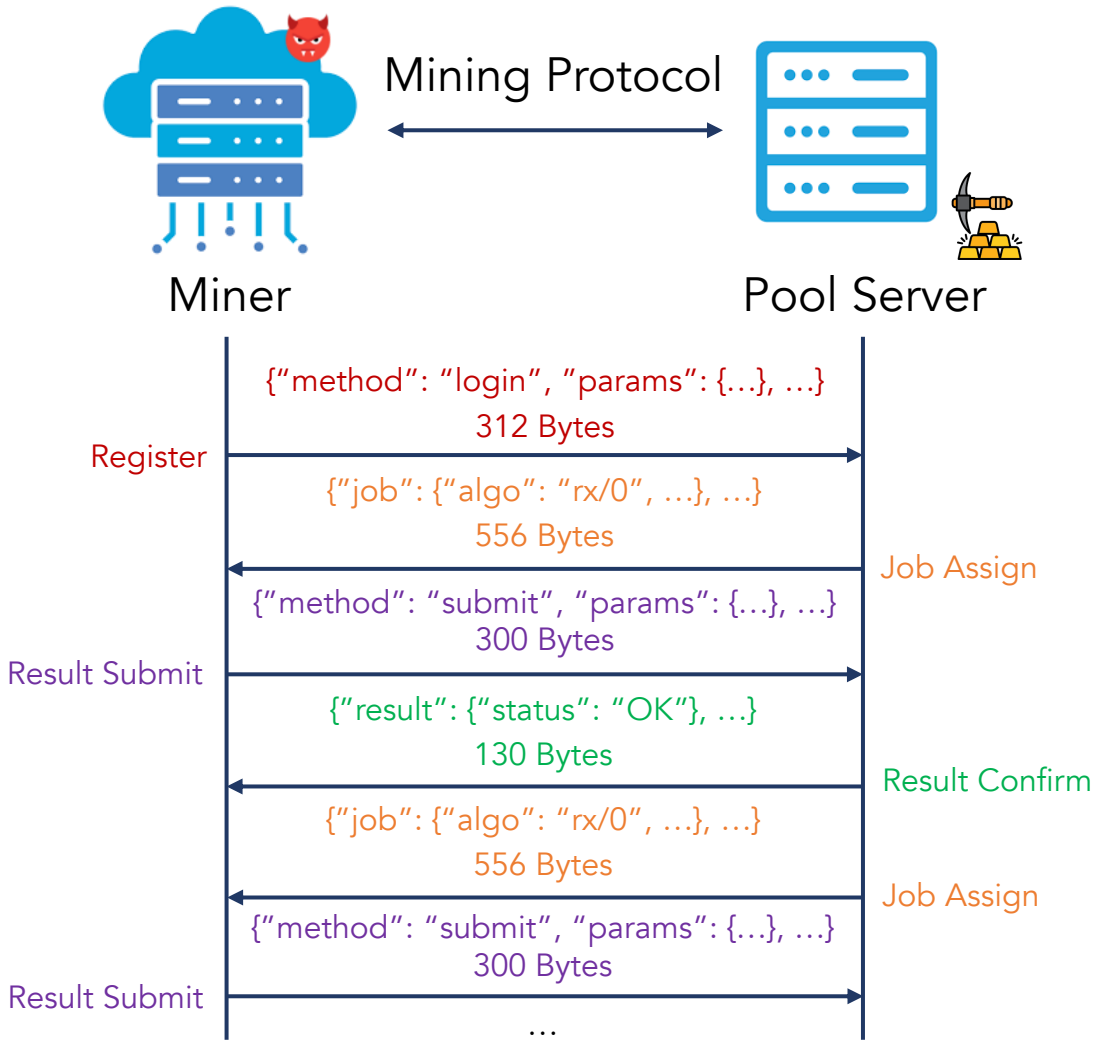
Using Machine Learning for Detection



- Identificable Traffic Features
  - Unique Packet Sizes
  - Regular Inter-packet Delays

Message frequency is consistent based on selected mining hardware and target cryptocurrency.

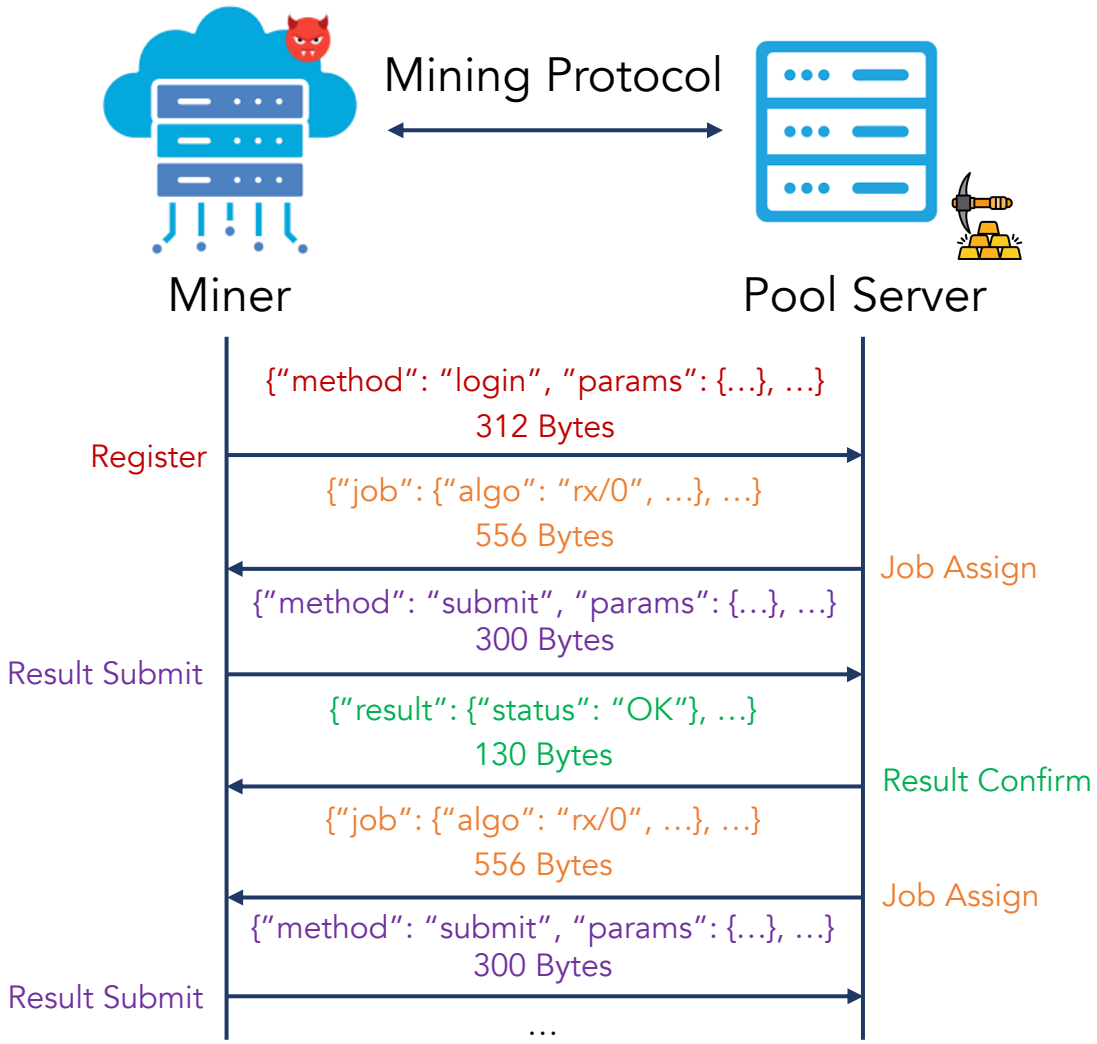
Using Machine Learning for Detection



- Identificable Traffic Features
  - Unique Packet Sizes
  - Regular Inter-packet Delays
  - Repetitive Packet Orders

The mining pool continuously assigns new jobs after confirming the miner's calculation results.

Using Machine Learning for Detection



- Identificable Traffic Features
  - Unique Packet Sizes
  - Regular Inter-packet Delays
  - Repetitive Packet Orders
- Long Connection Behavior

Mining profits increase with time.

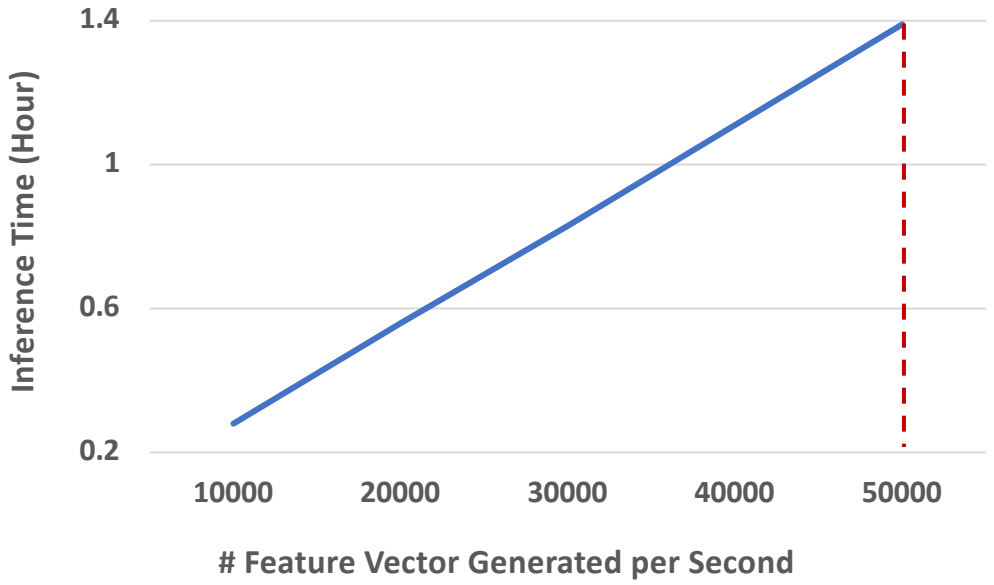
State-of-the-Arts

ML-based System	Features	Throughput	Investigation
MineHunter (ACSAC'21)	Packet timing	2.8 Gbps	Manual
CJ-Sniffer (RAID'22)	Packet timing	10 Gbps	Manual
Tekiner et al. (NDSS'22)	Packet size and timing	Unknown	Manual
MineShark (This work)	Packet size and timing	92 Gbps	Automatic

- Existing work shows promising results on labeled data sets.
- However, they are not applicable for large-scale deployment:
  - Network IDSes are commonly deployed at 10 Gbps gateways.
  - Limited time budget is available for investigating detection results.

Chanllege 1: Insufficient Throughput

- High traffic speeds lead to inference bottlenecks.



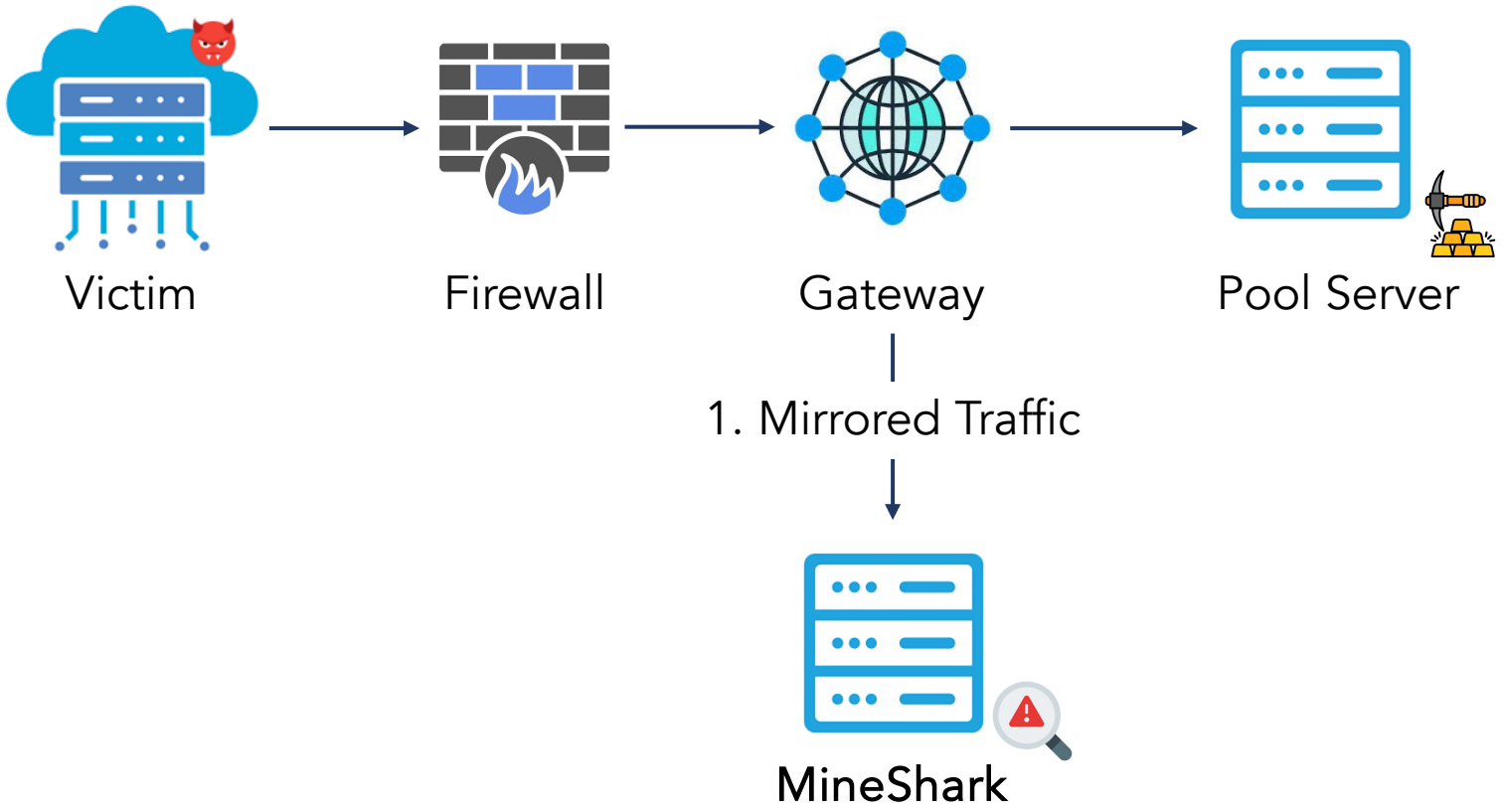
Inference time required to process all per-second feature vectors
using the Tekiner et al. (NDSS'22) model on a 10 Gbps campus gateway

Chanllege 2: Operational Overhead

- Large traffic volumes exacerbate false alarms.

Number of alarms generated each day by the most accurate model on live traffic from a 10 Gbps campus gateway

MineShark: Overview



MineShark: Overview

Online Filtration

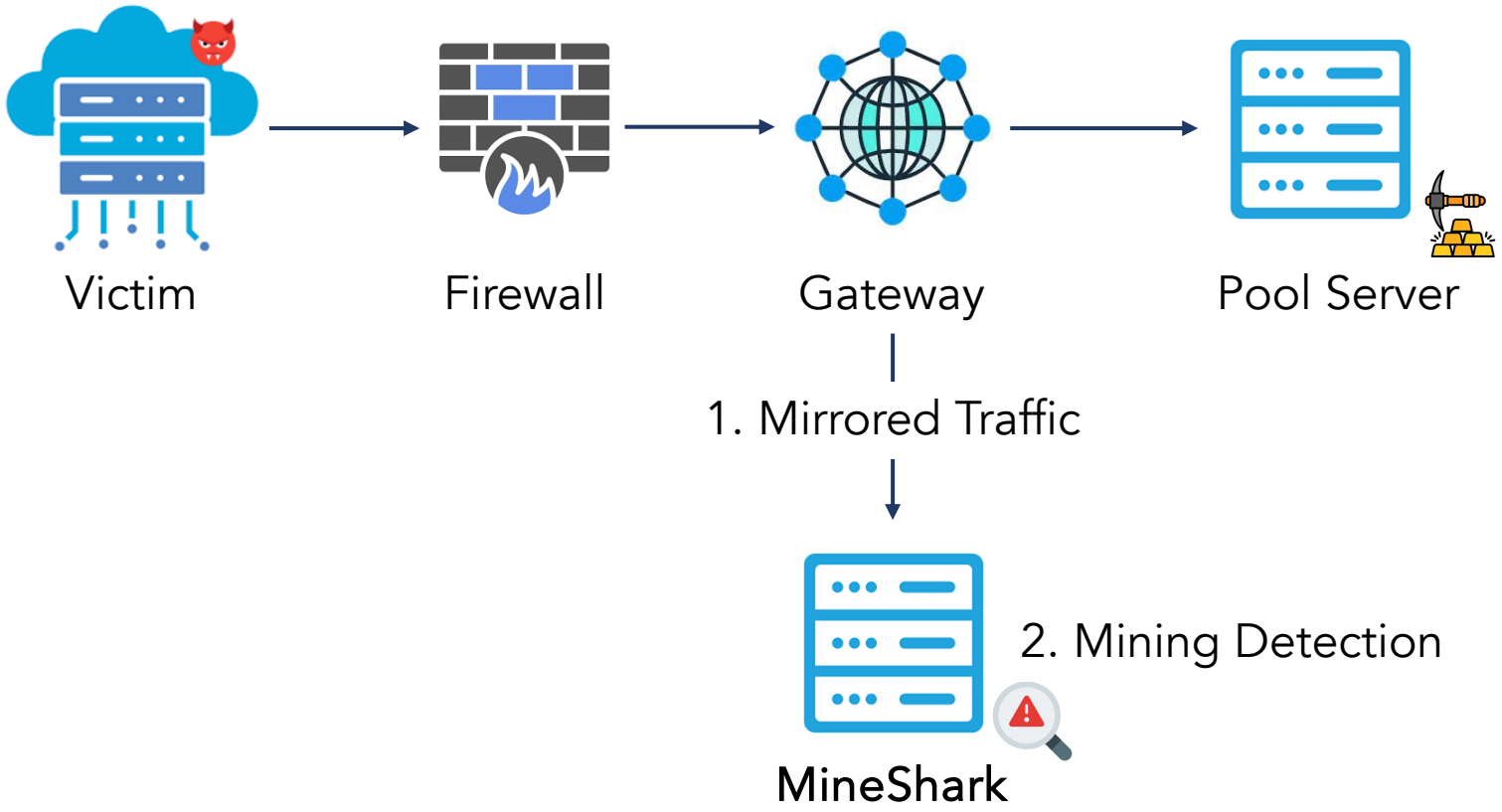
Filter mining threats without causing an inference bottleneck.

Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.



MineShark: Overview

Online Filtration

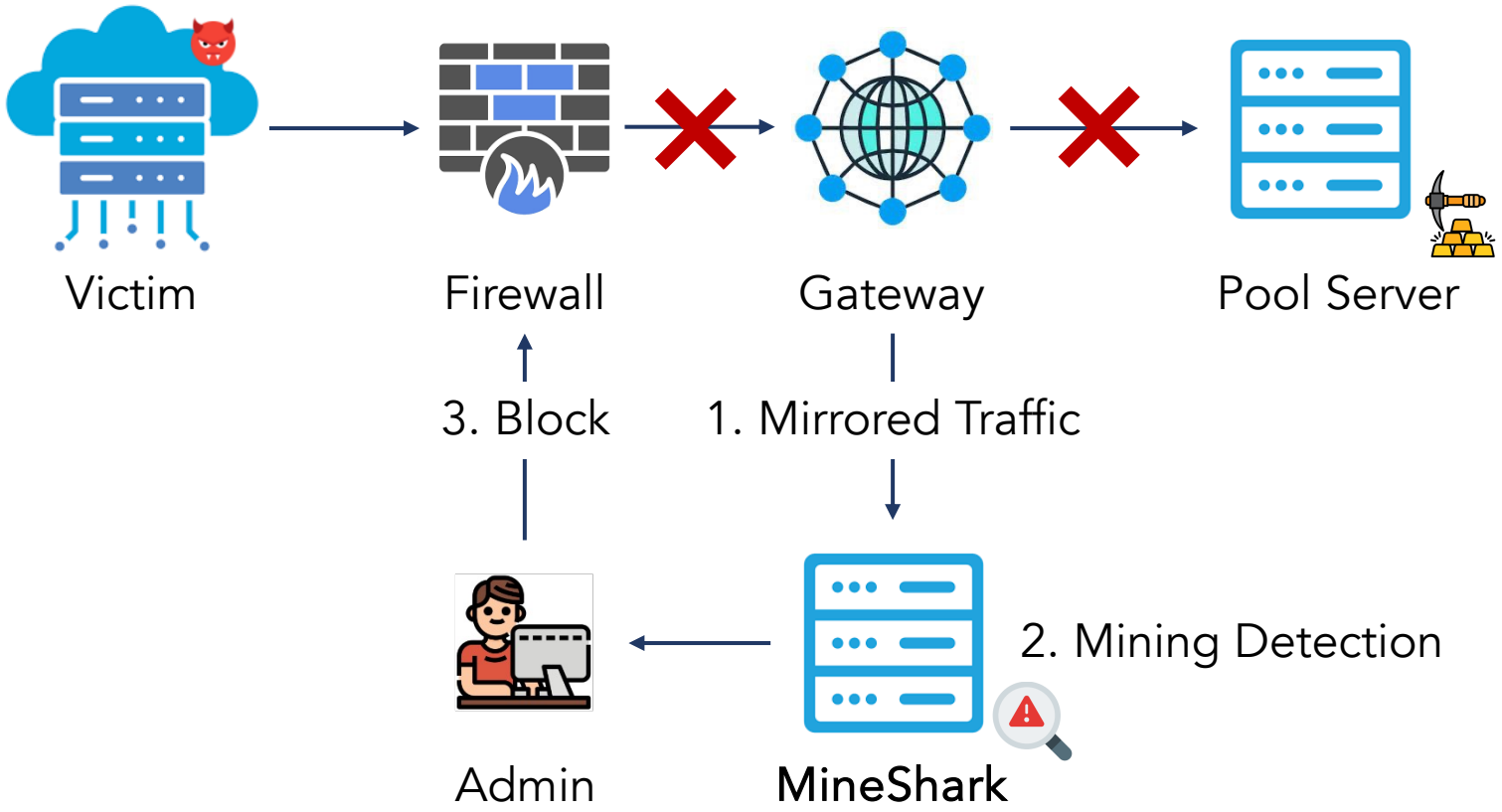
Filter mining threats without causing an inference bottleneck.

Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.



MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

↓

Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

↓

Proactive Defense

Defend against future attacks using historical attack data.

Throughput-first: Select lightweight models with high recall, allowing for temporary misclassification of some benign flows.

Model	Precision	Recall	Throughput (Mpps)
SVM	94.5%	76.2%	2.5
👍 CNN	99.1%	97.3%	1.5
LSTM	99.5%	98.0%	1.0

MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.

Throughput-first: Select lightweight models with high recall, allowing for temporary misclassification of some benign flows.

Model	Precision	Recall	Throughput (Mpps)
SVM	94.5%	76.2%	2.5
👍 CNN	99.1%	97.3%	1.5
LSTM	99.5%	98.0%	1.0

Inference pipeline: Conduct parallel computation across CPUs and GPUs.
Increase alarm count (i.e., # Alarm) upon detecting a positive feature.
Flag non-zero alarm flows as suspicious and send them to the next stage.

Flow ID	# Alarm	# Feature	First Seen	Last Seen	Lable
IPs-Ports	10	15	T_1	T_2	~~Normal~~

Suspicious

MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

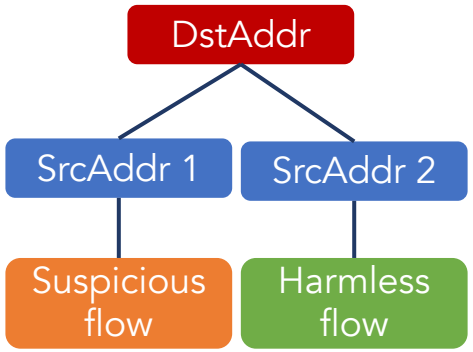
Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

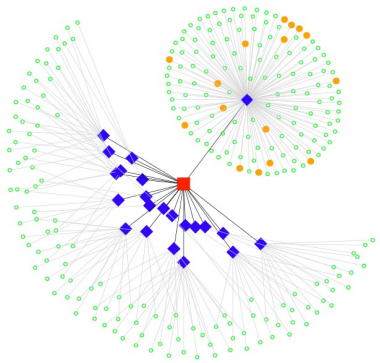
Proactive Defense

Defend against future attacks using historical attack data.

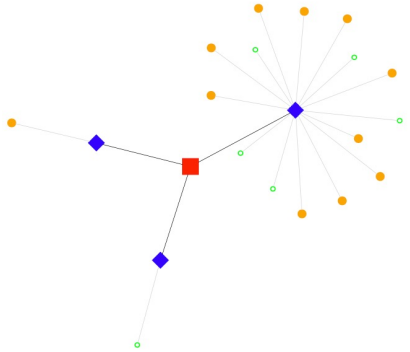
Correlation Graph: Aggregate suspicious flows by destination address to form long-term visit patterns.



Correlation Graph



Website



Mining Pool

Harmless flow: A suspicious flow whose alarm count is one.

False Alarm Removal: Non-mining addresses are distinguished by setting a threshold for the harmless flow ratio (e.g., 90%).

MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

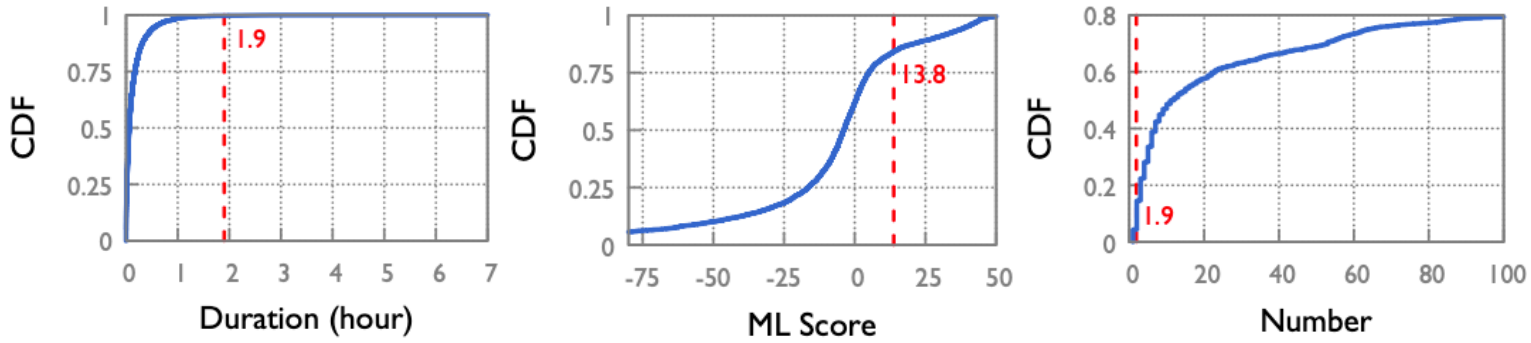
Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.

Address Ranking: Rank remaining addresses based on three features that capture additional mining behavioral patterns.



Features:

- **Flow Duration:** Time elapsed since the flow started.
- **ML Score:** Weighted score of all flows associated with an address.
- **Parallel Visit:** Average number of unique source addresses connected to the destination in a two-hour time window.

MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

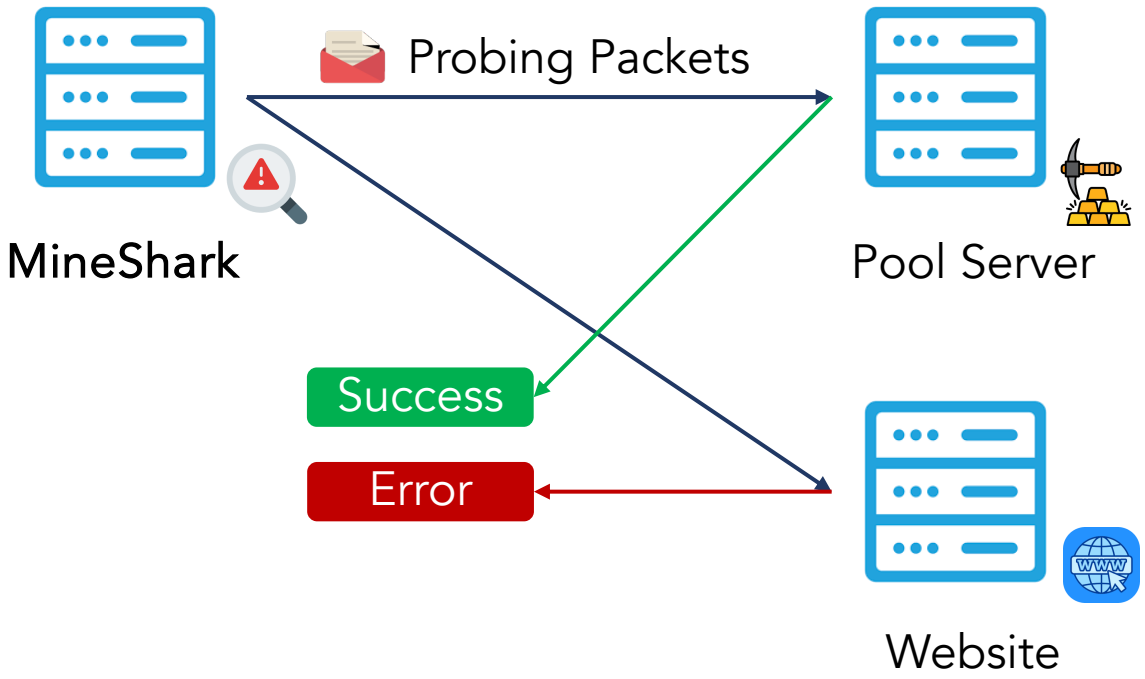
Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.

Active Probing: Simulate mining software to request mining services from ranked suspicious mining pool addresses.



MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

Auto-investigation

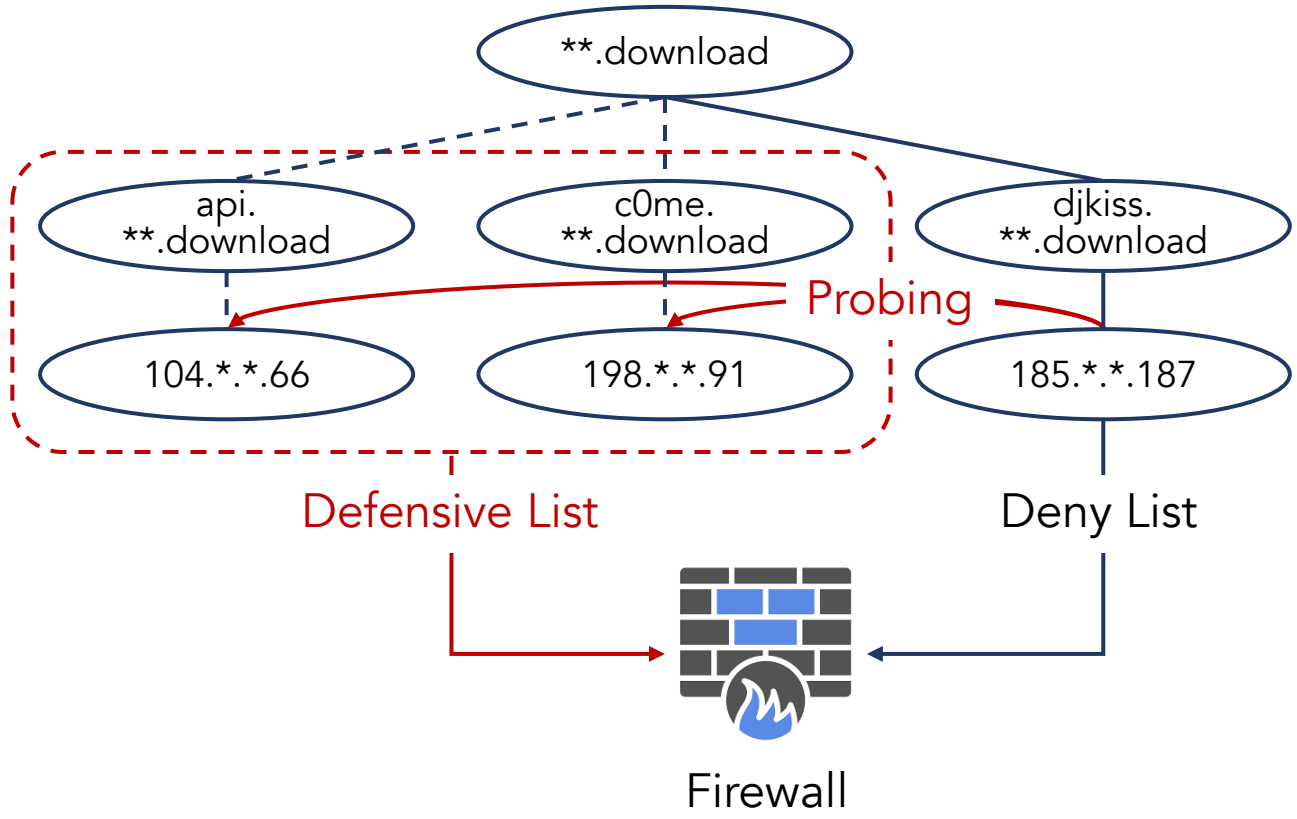
Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.

Domain Correlation: Search for addresses associated with detected mining domains. Identify mining pools with probing in advance.

Explanation: Attackers often correlate backup addresses by domains.



MineShark: Design

Online Filtration

Filter mining threats without causing an inference bottleneck.

Auto-investigation

Confirm mining activities with active probing, based on a ranked address list.

Proactive Defense

Defend against future attacks using historical attack data.

Port Fingerprinting: Fingerprint mining service ports to improve the efficiency of probing suspicious addresses.

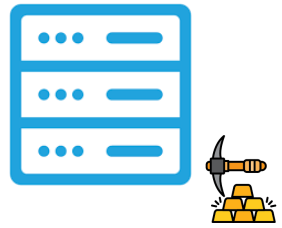
Explanation: Malware families often configure same service ports.

Pool Records	Port	Features
192.*.*.178 192.*.*.113 192.*.*.114	10083	Plain, Low rate
	28888	Encrypt, Low rate
	33337	Encrypt, High rate
103.*.*.178	43211	Encrypt, Low rate

Matched Ports

Probing

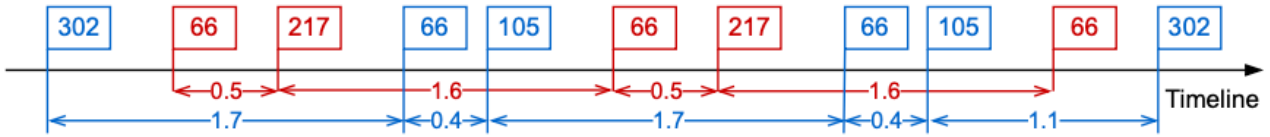
Suspicious Pools	Port
192.*.*.218	10083



Pool Server

Deployment Setup

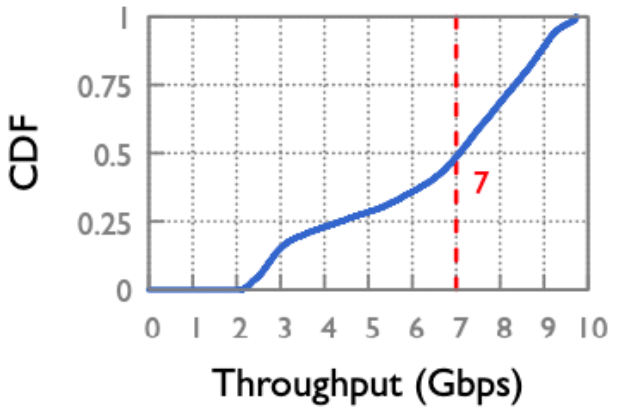
- Software Implementation
  - Model: CNN
  - Traffic processing: DPDK
  - Inference: TensorFlow



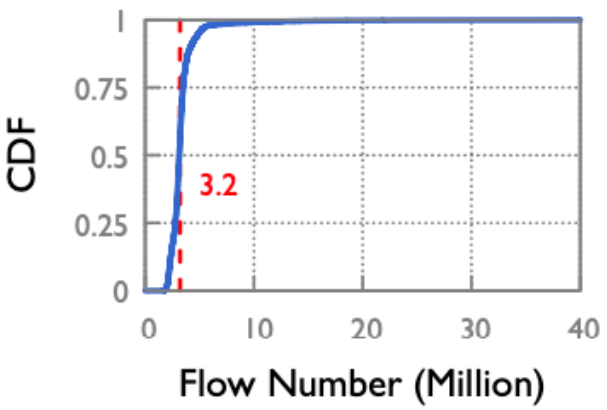
Inbound size	302	66	105	66	105	302	66
Outbound size	66	217	66	217	66	66	217
Inbound delay	1.7	0.4	1.7	0.4	1.1	1.7	0.4
Outbound delay	0.5	1.6	0.5	1.6	1.1	0.5	1.6

Deployment Setup

- Hardware
  - Mirrored traffic: 10 Gbps
  - CPU: 4 cores
  - GPU: 1 NVIDIA RTX-2060
  - Memory: 16 GB

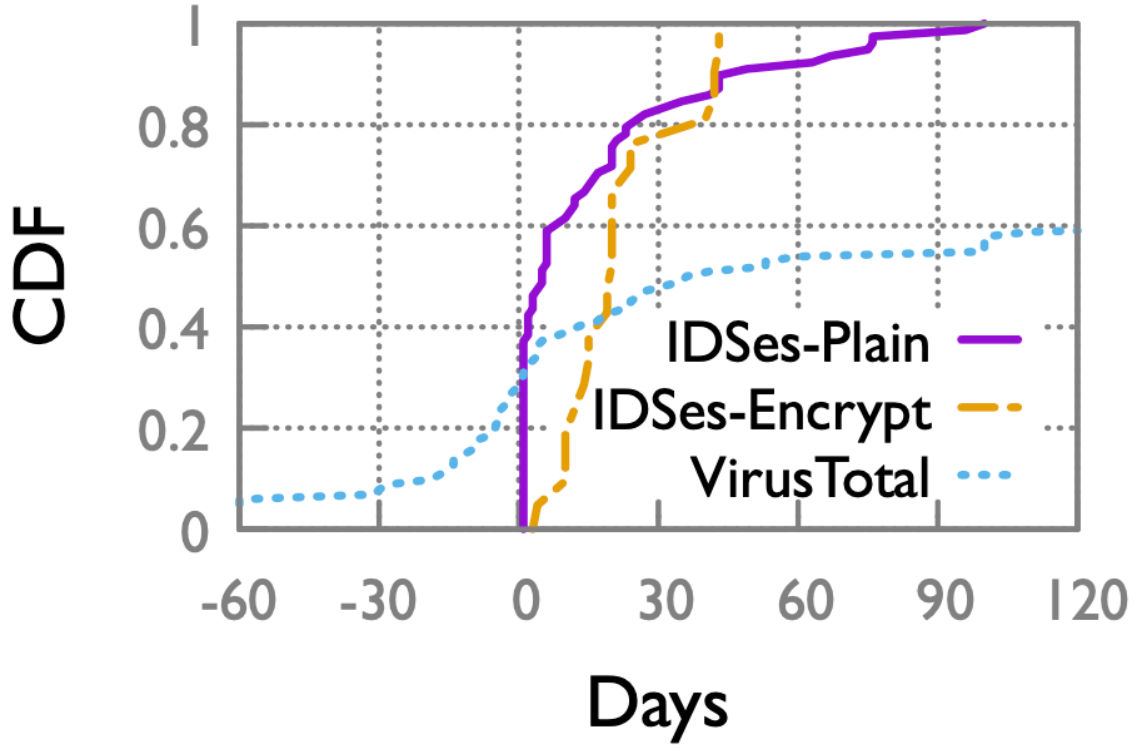


Input Traffic Speed



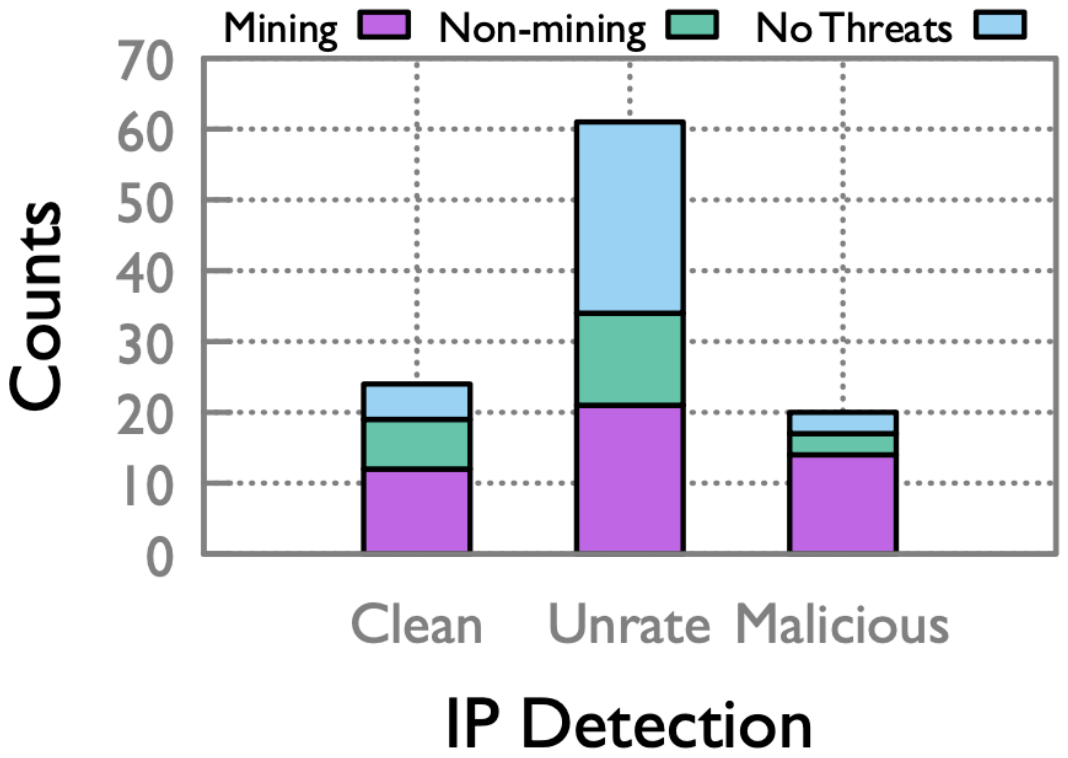
Flow Concurrency

Evaluation: Timeliness Improvement



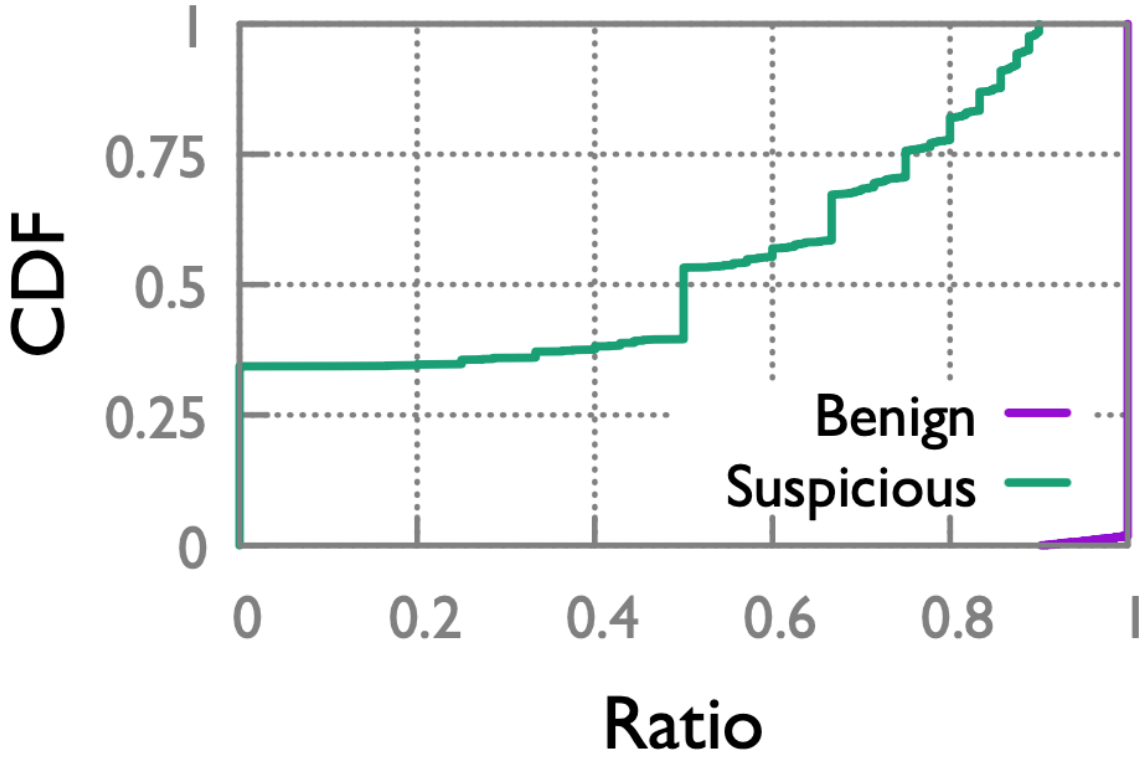
MineShark detected cryptojacking attacks before commercial IDSes and VirusTotal.

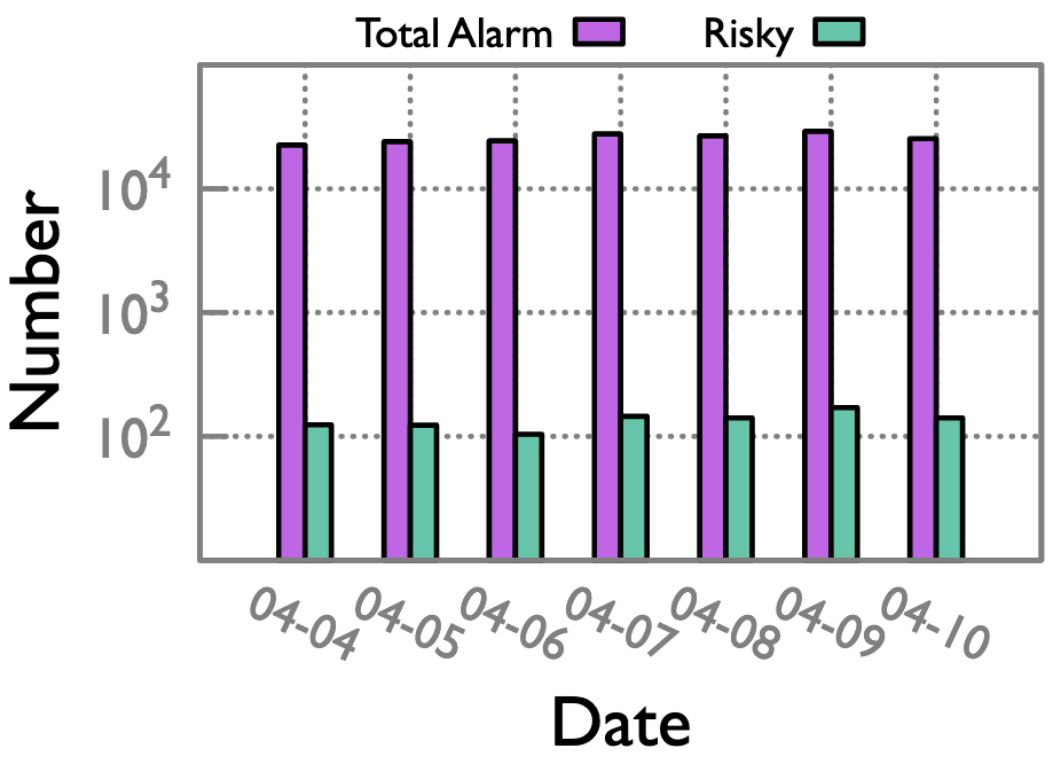
Evaluation: Accuracy Improvement



MineShark can improve the accuracy of VirusTotal's intelligence.

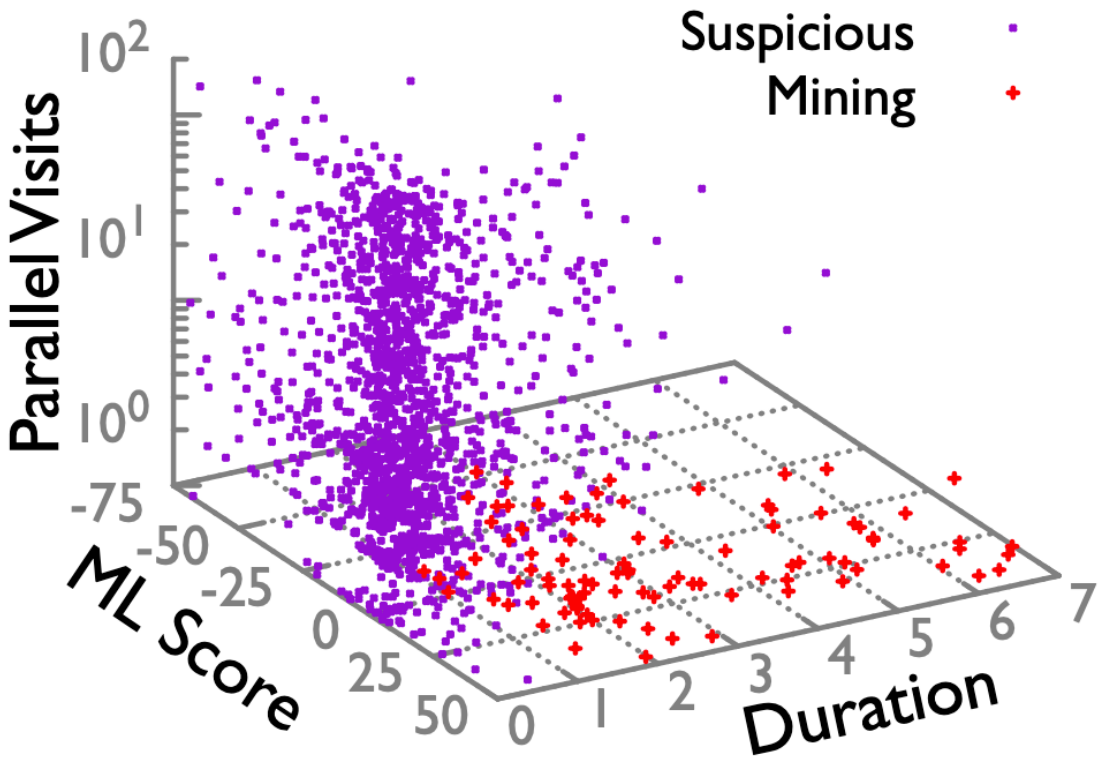
Evaluation: False Alarm Removal





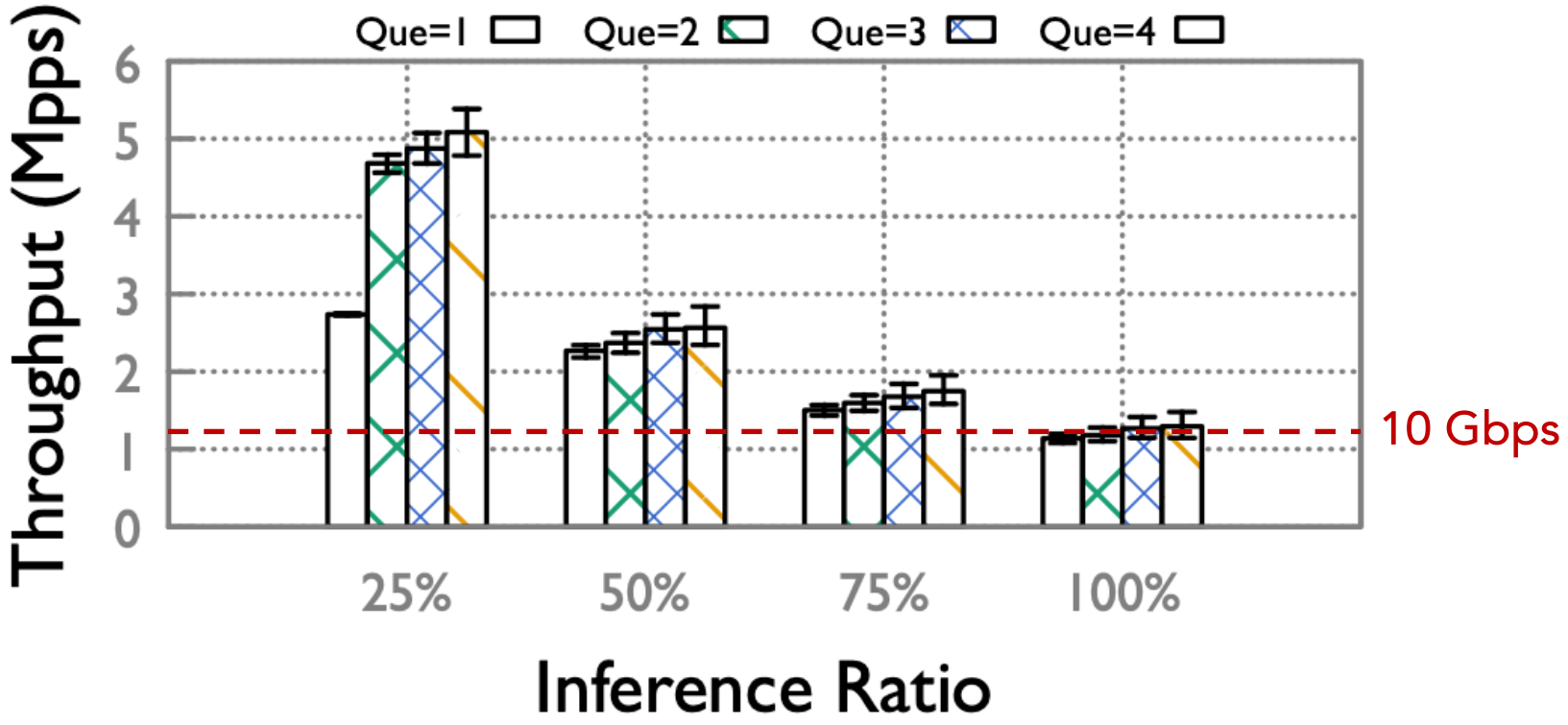
MineShark reduced false alarms by two orders of magnitude.

Evaluation: Address Ranking



MineShark ranked mining pool addresses at the top of the probing list.

Evaluation: System Efficiency



MineShark achieved line-rate processing on a 10 Gbps network gateway.

Conclusion

- MineShark is an ML-based system for online cryptomining traffic detection.
- MineShark incorporates a line-rate inference pipeline to drain high-speed traffic, an auto-investigation module to provide reliable detection results, and conducts proactive defense.
- MineShark can improve detection timeliness and accuracy in real-world scenarios.

Thanks!

- MineShark's artifact is available for result reproduction.
  - https://doi.org/10.5281/zenodo.13624057

- If you have any questions about this paper,

feel free to contact shaokexi@zju.edu.cn