Poster: A Two-Stage Encrypted Cryptomining Traffic Detection Mechanism in Campus Network

Ruisheng Shi✉, Haoran Sun, Lina Lan✉, Zhiyuan Peng, Chenfeng Wang
Beijing University of Posts and Telecommunications, Beijing, China
shiruisheng, lanlina@bupt.edu.cn

***Abstract*—Detecting cryptomining behavior within campus networks has always been an important issue. Network traffic-based detection schemes are widely used in campus networks. However, widely used proxies and private mining pools frequently change their domain names or IP addresses to evade detection, resulting in outdated blacklists and false positives or underreporting. In addition, most mining pools also support SSL/TLS protocols, while detection schemes based on deep packet inspection are usually unable to identify encrypted cryptomining traffic. Meanwhile, existing cryptomining behavior detection schemes often fail to provide fine-grained information about cryptomining behavior, and there is no automated solution to reduce false positives and underreporting. Therefore, we propose a two-stage encrypted cryptomining traffic detection mechanism. Unlike existing schemes, it provides both fine-grained detection results through machine learning and reduces false positives and underreporting from classifiers through active probing.**

I. Introduction

College campuses are the second biggest miners of cryptocurrencies behind the energy and utilities sector [1]. In campus networks, two main risks prevail: insider misuse and external cyber-attacks. Insiders might exploit equipment and electricity for active mining, which includes solo mining and pool mining. Meanwhile, external attackers often target these devices in campus, potentially commandeering them for cryptojacking operations. Cryptojacking is a malicious activity involving the unauthorized use of victims' device for cryptomining, which manifests in two forms: browser-based cryptojacking and binary-based cryptojacking.

Existing cryptomining behavior detection schemes primarily base on host behavior, browser behavior and network traffic [2], [3], [4]. Among them, the network traffic-based detection schemes are widely used in campus networks, with the advantages of easy deployment and no need for end-user cooperation. It relies on the creation and updating of blacklists that include results from open-source threat intelligence (e.g., underground forums, darknet marketplaces), plaintext mining traffic, or other detection schemes such as deep packet inspection. This detection scheme has proven effective for public pools, pools with plaintext traffic or distinctive domain names. However, the increasing use of proxy and private mining pools which often change their URLs leads to outdated blacklists and a high number of false positives [5]. Furthermore, it is difficult to timely add newly emerging pools to the blacklist, resulting in significant underreporting. In addition, most mining pools now offer SSL/TLS encryption services, while detection schemes based on deep packet inspection are usually unable to identify encrypted cryptomining traffic. Additionally, most existing encrypted cryptomining traffic detection schemes only provide binary classification results, which makes it difficult for supervisors to obtain fine-grained information about cryptomining behavior, such as the cryptocurrencies being mined. Meanwhile, existing schemes usually lack of solutions to reduce false positives, which undoubtedly reduces the usability of these schemes in real-world scenarios.

In this paper, we focus on the current state of network traffic-based mining detection and propose a two-stage encrypted cryptomining traffic detection mechanism to meet the regulatory needs of campus networks. For underreporting, we analyze the mining traffic under different configurations for mainstream cryptocurrencies and propose a fine-grained classification scheme of mining traffic based on machine learning. Meanwhile, through in-depth analysis of different configurations of mining pool communication protocol, we propose and evaluate an active probing scheme to reduce false positives caused by outdated blacklists and detection classifiers.

II. Methodology

A. Two-Stage Detection Mechanism

For traffic that passes through the campus gateway, the campus network usually uses blacklists, DPI-based or other detection schemes to quickly classify the traffic. They are identified as cryptomining traffic if they contain suspicious destination addresses or mining-related message content. However, there is no certainty that ”normal” encrypted traffic is innocent. Therefore, we can use our two-stage mechanism for further detection.

As shown in Figure 1, we extract features from ”normal” encrypted traffic and classify it with a trained classifier in step (1). The classifier is able to indicate whether the traffic is cryptomining traffic or not, and further gives information about the coins being mined. From the result of traffic classification, we can get a list of suspicious addresses which are the destination addresses and ports obtained from the traffic classified as cryptomining traffic. We use these suspicious addresses as inputs to the active probing module in step (2). The module will probe these addresses based on request construction and response parsing. Eventually, based on the probing results, we can efficiently identify encrypted cryptomining traffic and update the blacklists at the same time.

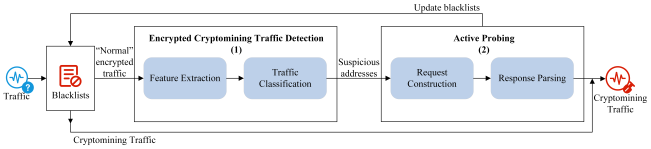


Fig. 1. The two-stage mechanism of encrypted cryptomining traffic detection

B. Fine-grained Detection Scheme

Nowadays, most miners and mining pools support SSL/TLS to improve communication security. However, the encryption of traffic has less impact on its time series features. In addition, different cryptocurrencies often differ in their mining algorithms and communication protocols, which are reflected in their time series features. Therefore, we can still use machine learning to learn the features and classify the traffic.

We captured the cryptomining traffic of 7 different cryptocurrencies from different devices, including traffic generated by active mining and cryptojacking behaviors. Meanwhile, we also captured different types of encrypted traffic generated by normal behaviors. Then, we extracted and selected 231 statistically significant time series features for binary classification experiments to ensure that we could correctly detect cryptomining traffic. Among the seven machine learning models we used, XGBoost performed better and achieved a recall of 0.99 and F1 score of 0.99. Further, we use the same feature extraction and selection methodology to perform multi-classification experiments on cryptomining traffic from 7 cryptocurrencies using XGBoost. By learning 182 features, the classifier achieved an mlogloss of 0.083 for the test set after 30 epochs and a 99.39% correct recognition rate after 50 epochs.

C. Active probing Scheme

Stratum protocol is widely adopted in Bitcoin, Monero, and Ethereum for communication. Thus, we focus on their different implementation of Stratum protocols and conducted a comprehensive investigation of related papers, source code, and collected cryptomining traffic datasets. Stratum protocol is mainly structured around four message formats: **miner subscription**, **miner authentication**, **mining job notification**, and **share submission**. Our active probing method focuses on different specific implementations of the first three messages with each corresponding success and error responses.

Our active probing method is based on subscription message construction and response parsing. For each target URL, we construct subscription messages for Bitcoin (Stratum-BTC) , Monero (including Stratum-XMR and Stratum-Webmine-XMR), and Ethereum (Stratum-ETH). Before sending the subscription message, we try to establish a TCP connection with the URL to confirming its availability. If it is confirmed to be alive, we send these messages using HTTP and HTTPS for Stratum-BTC, Stratum-XMR, and Stratum-ETH. Websocket was reserved exclusively for Stratum-Webmine-XMR, as it is predominantly used in browser-based mining, with Monero being a principal cryptocurrency.

Our analysis focused on specific fields in the success and error responses. We find that standard Stratum protocol responses typically include key fields (e.g., *id*, *result*, and *error*). Utilizing this pattern, we match fields within the responses to check for these common elements. The presence of these fields allows us to determine that the probed URL is offering a mining pool service.

To evaluate the effectiveness of our active probing method, we collect 149 mining pool service URLs from different types of pools, including public, proxy and private pools. According to the mining pool intelligence website, we registered and collected the URLs of the top 10 public pools of Bitcoin, Monero and Ethereum series (Ethereum Classic, Ethereum PoW) according to the total hashrate, which could account for more than 92% of the hashrate of their respective blockchain network. In addition, we collected proxy mining pool URLs from malicious scripts, and built the most representative proxy and private mining pools from public code repositories. Ultimately, our probe successfully received success or error responses from 147 different mining pool service URLs.

III. Conclusion

In this paper, we propose a two-stage encrypted cryptomining traffic detection mechanism to solve the problem of detecting encrypted cryptomining traffic and reduce the false positives and underreporting for the existing schemes in campus network. Then, we propose a fine-grained cryptomining traffic detection scheme based on time series features and machine learning. Furthermore, unlike the existing schemes, we propose and evaluate an active probing scheme to reduce the false positives of the classifiers.

Acknowledgment

This work was supported by the Beijing Natural Science Foundation under Grant M21037, 2022 Industrial Internet Public Service Platform - Industrial Internet Oriented Virtual Currency Mining Governance Public Service Platform Project by the Ministry of Industry and Information Technology of PRC (Grant No. TC220H054), Major Research and Application Project for the Supervision Platform of Virtual Currency Mining Behavior by the Ministry of Education of PRC.

References

[1] PCMag's report. https://www.pc-mag.com/news/college-kids-are-using-campus-electricity-to-mine-crypto.

[2] D. Tanana, Behavior-based detection of cryptojacking malware. In *2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, IEEE, pages 543-545, 2020.

[3] F. N. Naseem, A. Aris, L. Babun, E. Tekiner, and A. S. Uluagac. MINOS: A Lightweight Real-Time Cryptojacking Detection System. In *the Network and Distributed System Security Symposium (NDSS)*, pages 1-15, 2021.

[4] E. Tekiner, A. Abbas, and A. S. Uluagac. A lightweight IoT cryptojacking detection mechanism in heterogeneous smart home networks. In *Network and Distributed System Security Symposium (NDSS)*, 2022

[5] Z. Zhang, G. Hong, X. Li, Z. Fu, J. Zhang, M. Liu, et al. Under the dark: a systematical study of stealthy mining pools (ab) use in the wild. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS 23)*, pages 326-340, 2023.

Poster: A Two-Stage Encrypted Cryptomining Traffic Detection Mechanism in Campus Network

Ruisheng Shi, Haoran Sun, Lina Lan, Zhiyuan Peng, Chenfeng Wang
Beijing University of Posts and Telecommunications, Beijing, China
Email: shiruisheng, lanlina@bupt.edu.cn

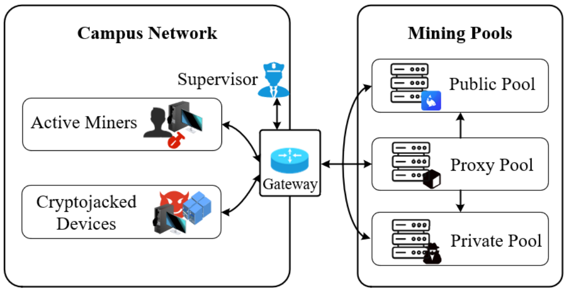
Problem and Motivation

Existing cryptomining behavior detection schemes in campus network:

- Blacklists
- Deep Packet Inspection
- Open-source threat intelligence (e.g., underground forums, darknet marketplaces)

Problems of existing schemes:

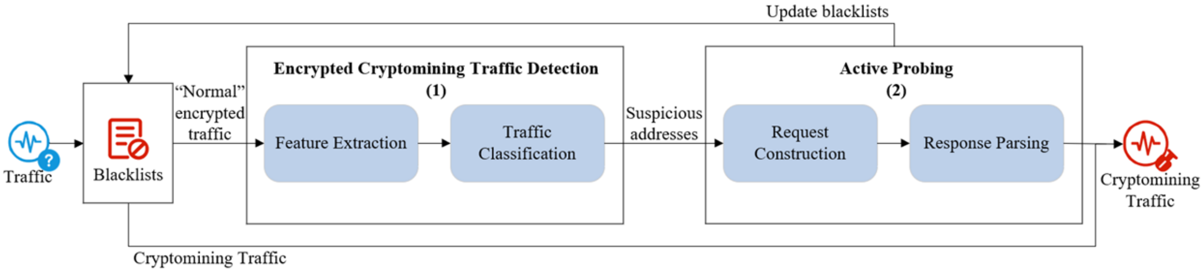
- **High FPR and underreporting without solution**: The increasing use of proxy and private mining pools, which often change their URLs, leads to outdated blacklists and false positives. It is difficult to add newly emerging pools to the blacklist in a timely manner, resulting in significant underreporting.
- **Plaintext traffic only**: Most mining pools also support SSL/TLS protocols, while detection schemes based on deep packet inspection are usually unable to identify encrypted cryptomining traffic.
- **Lack of fine-grained detection**: Supervisors are limited to binary classification results and lack of fine-grained information about cryptomining behavior, such as the coins being mined.



Threats within campus network:

- Active mining: solo mining & pool mining
- Cryptojacking: browser-based & binary-based

A Two-Stage Encrypted Cryptomining Traffic Detection Mechanism



a) We extract time series features from "normal" encrypted traffic and classify it with trained classifiers in step(1). The classifiers are able detect cryptomining traffic, and further gives information about the coins being mined.

b) From the result of traffic classification, we can get a list of suspicious addresses which are the destination addresses and ports obtained from the traffic classified as cryptomining traffic.

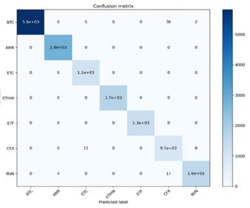
c) We use these suspicious addresses as inputs to the active probing module in step (2). The module will probe these addresses based on request construction and response parsing. Eventually, based on the probing results, we can efficiently identify encrypted cryptomining traffic and update the blacklists at the same time.

Fine-grained Detection

Different cryptocurrencies have some differences in mining algorithms and protocols, which obviously affect traffic features.

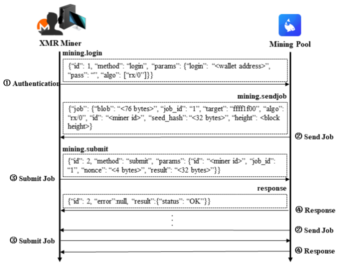
- We captured the cryptomining traffic of 7 different cryptocurrencies from different devices, including traffic generated by active mining and cryptojacking behaviours.
- Then, we extracted and selected 231 statistically significant (p-*value* ≤ 0.01) time series features for binary classification experiments and 182 for multi-classification.
- Among the classifiers we used, *XGBoost[2]* performed better:
  a) achieved a recall and F1 score of 0.99 in binary classification experiments.
  b) achieved an *mlogloss* of 0.083 for the test set after 30 epochs.
  c) achieved a 99.39% correct recognition rate after 50 epochs.

Classifier	Accuracy	Precision	Recall	F1 Score	Test ROC
Logreg	0.97	0.97	0.97	0.97	0.998
KNN	0.92	0.93	0.92	0.92	0.965
SVM	0.95	0.95	0.95	0.95	0.989
GNB	0.81	0.83	0.82	0.82	0.847
GBM	0.97	0.97	0.97	0.97	0.988
RF	0.98	0.99	0.98	0.99	0.998
XGB	**0.99**	**0.99**	**0.99**	**0.99**	**0.999**



Active probing

- Stratum protocol [3] is widely adopted in cryptomining. Thus, we focus on different implementation and conducted a comprehensive investigation of related papers, source code and collected traffic.
- Stratum protocol is mainly structured around four message formats: **miner subscription**, **authentication**, **job notification** and **share submission**.



mining.login

```
{"id": 1, "method": "login", "params": {"login": "<wallet address>", "pass": "", "algo": ["rx/0"]}}
```

mining.sendjob

```
{"job": {"blob": "<76 bytes>", "job_id": "1", "target": "ffff1f00", "algo": "rx/0", "id": "<miner id>", "seed_hash": "<32 bytes>", "height": <block height>}
```

mining.submit

```
{"id": 2, "method": "submit", "params": {"id": "<miner id>", "job_id": "1", "nonce": "<4 bytes>", "result": "<32 bytes>"}}
```

response

```
{"id": 2, "error":null, "result":{"status": "OK"}}
```

Evaluation: We collect 149 mining pool service URLs:
a) **Public pools**: the top 10 public pools of Bitcoin, Monero and Ethereum series by total hashrate, account for more than 92% of the hashrate of their network.
b) **Proxy pools**: public intelligence**,** malicious scripts, built from public code repositories.
c) **Private pools**: built from public code repositories.
Ultimately, our probe successfully received success or error responses from 147 different mining pool service URLs.

References

[1] Z. Zhang, G. Hong, X. Li, Z. Fu, J. Zhang, M. Liu, et al. Under the dark: a systematical study of stealthy mining pools (ab) use in the wild. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS 23), pages 326-340, 2023.
[2] T. Chen, C. Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pages 785-794, 2016.
[3] Stratum Protocol. https://zh.braiins.com/stratum-v1/docs.