DNS based In-Browser Cryptojacking Detection

Rohit Kumar Sachan, Rachit Agarwal, and Sandeep Kumar Shukla
IIT Kanpur, India

Supported by:

What is CryptoJacking?

- Distributed crypto mining approach
- Uses the victim's computing power without their consent
- Aim is to gain profits with out sharing
- Approaches:
    - Install malware
    - Execute scripts through the web application
        - In-Browser CryptoJacking

Procedure of In-Browser CryptoJacking

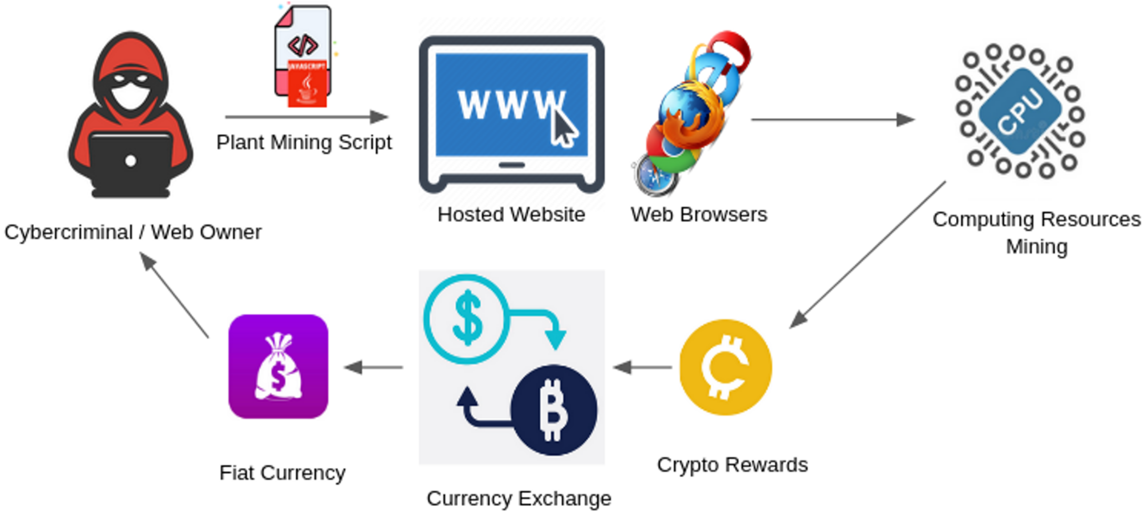

Fig. Procedure of In-Browser CryptoJacking

Cryptojacking: 415,000 Routers Infected with Cryptocurrency Mining Malware Globally

December 6, 2018 at 7:00 pm by Ogwu Osaemezu Emmanuel ALTCOINS BITCOIN BLOCKCHAIN

Google Sues to Shutter Cryptojacking Botnet That Infected 1M+ Computers

www.coindesk.com • 07 December 2021 22:41, UTC

Microsoft warns cryptojacking is still a major threat, despite crypto winter

By Sead Fadilpašić published about 22 hours ago

Bitcoin may be down, but cryptojackers are still flying high

Latest Report Shows Cryptojacking Increased By 30% During The Crypto Slump

www.newsbtc.com • 30 July 2022 18:30 UTC

Cryptojacking on the rise despite market slump 1

www.cryptopolitan.com • 27 July 2022 09:08, UTC

'Cryptojacking' Attacks on Financial Firms Surged in First Half

By Tanzeel Akhtar
July 26, 2022 at 3:35 PM GMT+5:30

26 Jul 2022 08:52 PM GMT+5:30 · 4 min read

World Economic Outlook, Crypto Tax Exemption, Cryptojacking Rising + More News

Hackers mined a fortune from Indian websites

Cryptojacking turns AP govt sites, among hundreds of others into mining platforms.

17 Sep, 2018, 08.2

MARTIN YOUNG

'Cryptojacking' rises 30% to record highs despite crypto slump: Report

Approaches to detect Cryptojacking

- Signatures/keywords crawling
- Analysis of computational resource utilization
- Analysis of scripting code
- Opcode analysis
- Trace network packets
- Analysing the hash function of mining script

Evasion techniques are used to evade from these detection approaches.

(CPU limiting, Code obfuscation, Payload hiding, and Changes in script code)

Motivation

- Websites have a unique signature on their metadata like,
  - Domain Name (DN) and
  - Domain Name System (DNS) records

- *Can these metadata help to detect websites performing/involved in in-browser cryptojacking?*

In this work

- Similarity analysis between cryptojacking DNs and other malicious DNs
- Measure the effectiveness of the DN-based approach [1] for identifying cryptojacked DNs
- Analysis of Indian Government websites

[1]. Sachan, R. K., Agarwal, R., & Shukla, S. K. (2021). Identifying malicious accounts in Blockchains using Domain Names and associated temporal properties. arXiv preprint arXiv:2106.13420.

Literature Study

Technique	Ref.	S	P	M	D	N	C	O	H	DNS	Oth	Method	Datasets: Source	Datasets: Size	Performance / Results	Limitation
Static	[16]	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	Crawling	Alexa	1.2M	901 TLDs	Unable to handle obfuscation techniques and Memory overhead
Dynamic	[7]	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	Threshold-based	Alexa	853K	2770 TLDs	Detects only hash modeled signatures
	[12]	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	RF	VirusShare OpenDNS	1K	Acc=>99.0% Recall=99.2% Precision=99.2% TPR=99.2% FPR=0.9%	Performance validated on limited data
	[19]	✗	✓	✓	✗	✓	✗	✗	✗	✗	✓	Crawling	Alexa BlackLists, PublicWWW, CoinHive, CryptoLoot, JSEcoin, CoinHave	200K	Profit$\approx$5.5$\times\downarrow$ CPU$\approx$59$\times\uparrow$ Temp$\approx$52.8$\times\uparrow$ Power$\approx$2.0$\times\uparrow$	Performance and Time overhead
	[11]	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓	CNN	Alexa	47K	Acc=98.7% TPR=97.87% FPR=0.74%	Address exclusively browser-based mining
	[20]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	TLC, SMO, MISVM, Random SubSpace	Alexa	1.2K	1837 TLDs Precision=1.0% Recall=1.0%	Performance validated on limited data
	[21]	✗	✓	✓	✗	✓	✗	✗	✗	✗	✗	K-Means DBSCAN Agglomerative	Hybrid dataset, CIC-IDS2018	-	Precision, Recall, F1-Score= >92.0	Limited mining samples
	[22]	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	RF	Self Generated	-	F1-Score=96.0% AUC=99.0%	Solely relying on the network traffic
	[13]	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	CNN	PublicWWW	-	Acc=98.97% Precision=93.07% F1-Score=95.04%	Considers only WASM modules and does not support JS modules
Hybrid	[8]	✓	✓	✓	✗	✓	✓	✗	✗	✗	✗	Crawling	Alexa	1M	-	Detect only CryptoNight miners, Do not support JS miners
	[14] [15]	✓	✓	✗	✗	✓	✓	✗	✗	✗	✓	FCM SVM RF	Pixalate Netlab360	5.7K	Acc=96.4% FPR=3.3% FNR=3.7%	Scalability issue, Code obfuscation and WASM are not considered
	[10]	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	CNN	Self Generated	1.8K	DR=87.0% DR=99.0% (after 11 sec.)	Address exclusively browser-based mining
	[9]	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	Crawling	Alexa, Majestic, PublicWWW, [23]	1.8M 48.9M	204 Campaigns 1136 TLDs	Exclusively depends on vulnerabilities of CMS providers- such as WordPress

• **Based on**: ^{S} Signature, ^{P} Processor / CPU, ^{M} Memory, ^{D} Disk, ^{N} Network Analysis, ^{C} Code Analysis, ^{O} Op-code, ^{H} Hashing Algorithm, ^{DNS} Domain Name

Methedology

- Analyzes the DNS traffic records
- Identifies 48 temporal and non-temporal properties/features
- Over the 2 hour *(2H)* and complete data granularity *(ALL)*
- Applies both supervised and unsupervised ML models to detect cryptojacked DNs

[1]. Sachan, R. K., Agarwal, R., & Shukla, S. K. (2021). Identifying malicious accounts in Blockchains using Domain Names and associated temporal properties. arXiv preprint arXiv:2106.13420.

Features

- Non-Temporal features:
  - String-based features
  - DNS Query-based features
- Temporal features:
  - Burst-based features:
    - Query frequency burst
    - Query Inter-Event Burst
  - DNS graph-based features:
    - Degree
    - Diameter

Datasets

Dataset	Cisco Umbrella top 1 million dataset (January 2020) [2]
Total DNS queries	335 Million
Unique DNS queries	1771626 ≈ 1.77 Million
Malicious tag	42002 DNS queries
Cryptojacked Dataset	29777 DNs/TLDs (from public sources)
Cryptojacked in Umbrella	1188 cryptojacked DNs, 21743 DNS queries
Unmarked cryptojacked in Umbrella	9681 DNS queries

[2]. OpenINTEL Consortium, “Cisco umbrella 1m,” 01 2019. Accessed: 02/10/2020.

Results

- Minimal divergence between temporal features of mDNs and cDNs.
- Unsupervised ML:
  - 9339 DNs > 1% probability to be involve in cryptojacking
  - 228 DNs > 99.0% probability to be involve in cryptojacking
  - Effective to detect cryptojacked DNs.
- Supervised ML:

Cryptojacking DNs in Dataset		Classifier	Results in (%)			
Train	Test		BAcc	Pre	Rec	F1
-	100%	DT†	67.56	86.0	35.64	50.0
80%	20%	DT‡	72.02	85.0	44.45	58.0
Total		1771626				

  - A low Recall on the cDN class signifies the need of improvement.

Case Study: Analysis of Indian Government websites

- 8669 Indian GOI web URLs [3]

Approach	Results/Findings
Signature crawling	66 Cryptojacking signatures None-of-the Indian webpages contains cryptojacking signatures
Resource utilization	19 resource measure (November to December 2021) 10 DNs have different properties These should be monitored
Analysis of DNS records	DNS graph using IP and NS addresses 7 connected components in the DNS plot 21 unique countries • DNs of 6728 webpages are hosted in India, • DNs of 48 webpages are hosted in the USA, and • DNs of 10 webpages are hosted in Estonia

[3]. IGOD, “Integrated Government Online Directory.” Accessed: 05/08/2021.

Future Work

- Like to improve the metadata-based approach and test it in a large dataset to detect in-browser cryptojacking.
- Like to develop temporal data of Indian Government websites, which will be helpful for the metadata-based approach in the future.

Thank you

sachan.rohit100@gmail.com