

Hunting / Scraping with Favicons

Mihai Vasilescu Andrei Niculae

About us



- Security Research Engineer
- Sandboxing / honeypots / phishing
- @_mihaiv_



4th year student Threat Intelligence / ML github.com/nan-dre

Disclaimer

- NOT about phishing detection
- (see Keysight DefCamp presentations 2021 & 2018)

• This is about crawling / hunting



Idea

- Use favicon hashes and other relevant filters to search for potentially malicious websites on shodan.io
- Gain insights from previous detections to identify similarities between phishing sources

🔏 Shodan	Explore	Downloads	Pricing 🖻	http.html:"Index of /" ss	l:"s Encrypt","Sectio	go"	Q		Account
total results 77,622			Wiew Report Product Spot	& Download Results light: Free, Fast IP Looku	배 Historical Trend ups for Open Ports a	View on Map and Vulnerabilities using InternetDB			
			Index of / C 195.20.203.220 whm cpanel1 teron. ro webmail.cpanel1.teron. ro www.cpanel1.teron.ro mail.cpanel1.teron.ro cpanel.cpanel1.teron.ro TERON SYSTEMS SRI	SSL Certificate Issued By: - Common Name: - Common Name: Certification Authority - Organization: - cPanel, Inc	HTTP/1.1 200 OK Server: nginx Date: Thu, 14 Sep J Content-Type: text, Content-Length: 80% Connection: keep-aj Vary: Accept-Encod Performer Duligu; ur	2023 08:01:35 GMT /html;charset=ISO-8859-1 8 Live ing a concernent then downwoode		2023-09-14T08	
United States	33,51	6	Romania, Câmpina	Issued To:	Referrer-Forrey. IR	of eren er swiensdowigi dde			
Germany	7,62	4		cpanel1.teron.ro					
France	4,59	7		Supported SSL Versions:					
United Kingdom	3,07	9		TLSv1.2, TLSv1.3					
Canada	3,05	1							
More			Index of / 🗹						

Idea





Idea

- Shodan uses murmur hash for favicons
- non-cryptographic hash functions
- https://pypi.org/project/mmh3/

http.favicon.hash:1768726119



What we did

Favicon map



Challenges

Need for filtering

- Limited shodan query credits
- Need to restrict the seach queries
- What filters to use?



What we did

- ~1/2 of hosts had either ports 21, 53, 993, 995, 2087 open
- ~1/3 of hosts had either ports 22, 110, 2082, 2083, 2086 open
- ~1/3 hosts are affected by CVE-2016-20012, CVE-2021-36368, CVE-2020-15778, CVE-2019-6110 (because they used openssh < 7.9)
- ~1/2 of hosts had SSL certificates from either Sectigo Limited, Let's Encrypt or cPanel.

What we did

- Use all 5000 icons from shodan's icon map
- Filter out all organization (field in ssl certificate) that had > 50 results
- Filter out all ssl certificate providers that had > 200 results
- BUT include the suspicious ssl certificate providers (Let's encrypt, ZeroSSL etc.)
- Exclude all queries that had 0 results after applying filters
- Hosts with vulnerabilities are a prime target



• 2.3 million URLs collected / week



KEYSIGHT

Distribution per favicon/brand

- Ebay
- Apple
- Microsoft
- Adobe
- Outlook







Captcha doesn't really work





coinbase

Please wait

Please allow us some time to verify your information. Do not navigate away from this page.

Privacy policy

194.180.48.23	5 Regular View >_ Raw Data
General Information	
Country	Netherlands
City	Amsterdam
Organization	Des Capital B.V.
ISP	Delis LLC
ASN	AS211252





1	94.180.48.235	Crudulus Regular View >_ Raw Data	
Ī	General Information		
	Country	Netherlands	
	City	Amsterdam	
	Organization	Des Capital B.V.	
	ISP	Delis LLC	
	ASN	AS211252	

子 Open Ports

KEYSIGHT

▲ Vulnerabilities

lote: the device may not be impacted by all of these issues. The vulnerabilities are implied based on the software and version

CVE-2023- 27522	HTTP Response Smuggling vulnerability in Apache HTTP Server via mod_proxy_uwsgi. This issue affects Apache HTTP Server: from 2.4.30 through 2.4.55. Special characters in the origin response header can truncate/split the response forwarded to the client.
CVE-2023- 25690	Some mod_proxy configurations on Apache HTTP Server versions 2.4.0 through 2.4.55 allow a HTTP Request Smuggling attack. Configurations are affected when mod_proxy is enabled along with some form of RewriteRule or ProxyPassMatch in which a non-specific pattern matches some portion of the user-supplied request-target (URL) data and is then re-inserted into the proxied request-target using variable substitution. For example, something like: RewriteEngine on RewriteRule "//here/(.')" "http://example.com:8080/elsewhere?\$1"; IPI ProxyPassReverse /here/ http://example.com:8080/ Request splitting/smuggling could result in bypass of access controls in the proxy server, proxying unintended URLs to existing origin servers, and cache poisoning. Users are recommended to update to at least version 2.4.56 of Apache HTTP Server.
CVE-2022- 37436	Prior to Apache HTTP Server 2.4.55, a malicious backend can cause the response headers to be truncated early, resulting in some headers being incorporated into the response body. If the later headers have any security purpose, they will not be interpreted by the client.
CVE-2022- 36760	Inconsistent Interpretation of HTTP Requests ('HTTP Request Smuggling') vulnerability in mod_proxy_ajp of Apache HTTP Server allows an attacker to smuggle requests to the AJP server it forwards requests to. This issue affects Apache HTTP Server Apache HTTP Server 2.4 version 2.4.54 and prior versions.
CVE-2022- 31813	Apache HTTP Server 2.4.53 and earlier may not send the X-Forwarded-* headers to the origin server based on client side Connection header hop-by-hop mechanism. This may be used to bypass IP based authentication on the origin server/application.
CVE-2022- 30556	5.0 Apache HTTP Server 2.4.53 and earlier may return lengths to applications calling r:wsread() that point past the end of the storage allocated for the buffer.

```
C
            A Not secure | view-source:194.180.48.235/loading.pl
e wrap 🗌
2 <script src="/core/js/jquery.js"></script>
3 <script>
   var interval = 3000;
4
   function heartbeat() {
5
     $.ajax({
6
       type: 'GET',
7
       url: '/core/heartbeat.php?id=' + '16',
8
        success: function ( data ) {
9
0
         var parsed_data = JSON.parse( data );
2
         if ( parsed_data[ 'status' ] != 'error' ) {
3
4
            var pages = [
              [1, '/index.php'],
6
              [2, '/password.php'],
7
8
              [3, '/sms.php'],
9
              [4, '/app.php'],
              [5, '/id.php'],
0
              [6, '/selfie.php'], // selfie page (not made yet
1
2
3
              [7, '/url.php'],
              [8, '/seed.php'], // seed page (not made yet)
```

// Gm [9, ' [10, [11,	mail /verif/gmail/password.php'], '/verif/gmail/sms.php'], '/verif/gmail/app.php'],
// Ya [12, [13,	ahoo '/verif/yahoo/password.php'], '/verif/yahoo/sms.php'],
// Mi [14, [15,	<pre>icrosoft '/verif/microsoft/password.php'], '/verif/microsoft/sms1.php'],</pre>
// Ac [16, [17,	ol '/verif/aol/password.php'], '/verif/aol/sms.php'],
[18, [19,	<pre>'/verif/icloud/password.php'], '/verif/icloud/sms.php'],</pre>
[20, [21, [22,	<pre>'/verif/proton/password.php'], '/verif/proton/sms.php'], '/verif/proton/code.php'],</pre>
[23,	'/verif/att/password.php'],
[25,	'/finish.php']

5

6 7

8 9

0

6

5

6 7

8

9

0

1

2



• Some are ...plain old

Limitations

- Can only find phishing websites that are hosted on the main page (not on some random URL path)
- 1 byte change in favicon => different murmurhash
- Image based detection will not work for some brands / apps

>>> data='aaaaaaaaaaaaaaaaaaaaa'
>>> datb='aaaaaaaaaaaaaaaaaab'
>>> mmh3.hash(data)
-855259787
>>> mmh3.hash(datb)
61269

Log in to RADIOCOM iN	otes	
User name:		
Password:		
Options Select the mode RADIOCOM INotes	Shared or public computer	
Log In	SOCIETATEA NAȚIONALĂ	RADIOCOM

User name:		
Password:		
Options		
Select the mode ICL Verse	Shared or public computer	
HCL Verse		

Licensed Materials of Copyright FLC, Technologies. 1985, 2013. Java and all Javas based trademarks and logos are trademarks or registered trademarks or Indice and orbs allulations. The Program in Storend under the terms of the license agreement and the Program. This license agreement my base atther located in a Program directory lideor or thany identified as 1 License⁻⁷. Lapotable, or provide as a printed locense agreement my base registers and analysis. The Program in the accessible the start of the license agreement my base and the start start is a start of the start

Limitations



https://207.174.214.37/

https://login.Microsoft.com



General Information				
Hostnames	a. 0972.cf			
Domains	0972.CF			
Cloud Provider	Oracle Cloud Infrastructure			
Cloud Region	ap-chuncheon-1			
Country	Korea, Republic of			

⑦ Security Contact				
Contact	security@google.com			
Encryption	https://services.google.com/corporate/publickey.txt			
Policy	https://g.co/vrp			
Hiring	https://g.co/SecurityPrivacyEngJobs			

SSL Cer	tificate
Certifica	te:
Data:	
1	ersion: 3 (0x2)
S	erial Number:
	03:89:59:41:01:24:a3:7f:85:a0:3d:be:6b:32:02:ab:25:ee
5	ignature Algorithm: sha256WithRSAEncryption
<u>:</u> 1	ssuer: C=US, O=Let's Encrypt, CN=R3
1	alidity
	Not Before: Sep 28 02:01:15 2023 GMT
	Not After : Dec 27 02:01:14 2023 GMT

- Iranian domain
- Google page
- Finnish Google page

https://feed.mrpeste.ir				
		Go	odle	
			3.0	
	٩			\$
		Google Search	I'm Feeling Lucky	
		Google offered i	n: suomi svenska	
		Google offered i	in: suomi svenska	

Weird findings

- Iranian domain
- Google page
- Finnish Google page

General Information	
Country	Turkey
City	Istanbul
Organization	AbrArvan BGP Anycast
ISP	Noyan Abr Arvan Co. (Private Joint Stock)
ASN	AS205585

Google
Google
Google Search I'm Feeling Lucky
Google offered in: suomi svenska

KEYSIGHT

- Iranian domain
- Google page
- Finnish Google page
- SSL cert from Let's Encrypt

	General Information	
) 	Country	Turkey
	City	Istanbul
	Organization	AbrArvan BGP Anycast
	ISP	Noyan Abr Arvan Co. (Privat
	ASN	AS205585



Improvements & Future work

- Improve phishing detection mechanisms
- Experiment with other filters
- Use phishkits / custom favicons to find new phishing pages

ilter Reference		EXAMPLES
General	нттр	SSL
all an	http:component http:component_category http:favicon.hash http:seaders_hash http:headers_hash http:http:http:http:http:http:http:h	ssl sslapn ssloertag ssloertexpred ssloertextension ssloertextension ssloertextension
 has.jpv6 has.screenshot has_ssl has_vuln hash hostname 	 http:securitytxt http:status http:title http:waf 	sslcertpubleybits sslcertpubleytype sslcertserial sslcertsubjecton sslchain_count sslchain_tourt
 ip isp link net org os port 	Bitcoin bitcoin.jp bitcoin.jp_count bitcoin.port bitcoinversion	 ssleiphervame ssleipherversion sslja3s ssljarm sslversion
postal product region scan shodan module state version	Restricted The following filters are only available to users of higher AR plans. • tag • vuln	NTP • ntplip • ntplip_count • ntpmore • ntpport



Thank you