NDSS Symposium 2025

Revealing the Black Box of Device Search Engine: Scanning Assets, Strategies, and Ethical Consideration

Mengying Wu[†*], Geng Hong[†*], Jinsong Chen[†], Qi Liu[†],
Shujun Tang[‡§], Youhao Li[‡], Baojun Liu[§], Haixin Duan[§¶], Min Yang[†]

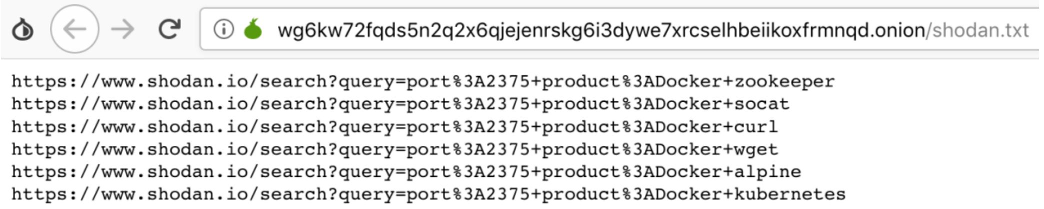
[†]Fudan University, [‡]QI-ANXIN Technology Research Institute [§]Tsinghua University, [¶]Quancheng Laboratory

2025/02/25

Motivation

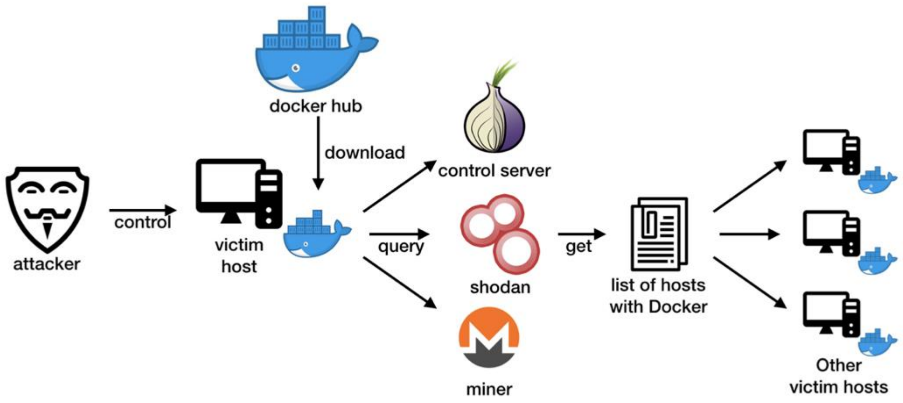
Why is my living room on the Internet?

Xulu Botnet Propagation



wg6kw72fqds5n2q2x6qjejenrskg6i3dywe7xrcselhbeiikoxfrmnqd.onion/shodan.txt

```
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+zookeeper
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+socat
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+curl
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+wget
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+alpine
https://www.shodan.io/search?query=port%3A2375+product%3ADocker+kubernetes
```



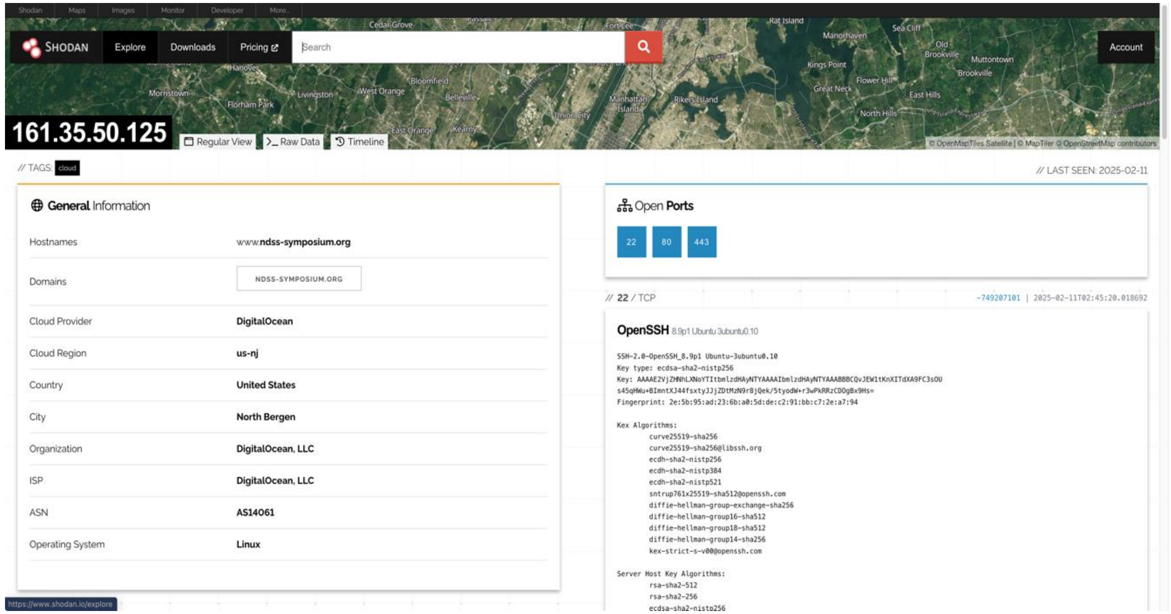
How do attackers find the victims?

From the powerful Device Search Engines

Device Search Engine

censys SHODAN Computer Search Engine FOFA ZoomEye

Search engine for Internet-connected devices



161.35.50.125

// TAGS: cloud

// LAST SEEN: 2025-02-11

General Information

Hostnames	www.**ndss-symposium.org**
Domains	NDSS-SYMPOSIUM.ORG
Cloud Provider	DigitalOcean
Cloud Region	us-nj
Country	United States
City	North Bergen
Organization	DigitalOcean, LLC
ISP	DigitalOcean, LLC
ASN	AS14061
Operating System	Linux

Open **Ports**

22 80 443

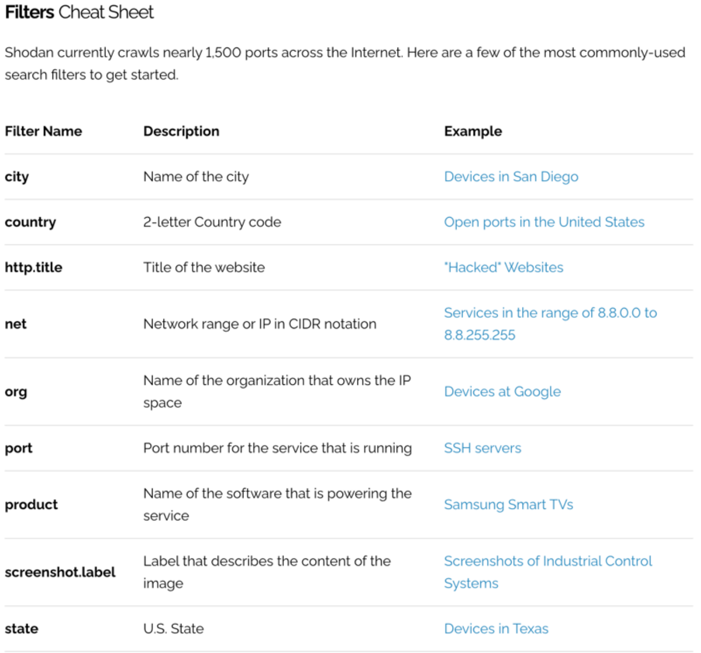
// **22** / TCP -749207101 | 2025-02-11T02:45:20.018692

OpenSSH 8.9p1 Ubuntu 3ubuntu0.10

```
SSH-2.0-OpenSSH_8.9p1 Ubuntu-3ubuntu0.10
Key type: ecdsa-sha2-nistp256
Key: AAAAE2VjZHNhLXNoYTItbmlzdHAyNTYAAAAIbmlzdHAyNTYAAABBBCQvJEW1tKnXITdXA9FC3sOU
s45qHWu+BImntXJ44fsxtyJJjZDtMzN9r8jQek/5tyodW+r3wPkRRzCDOgBx9Hs=
Fingerprint: 2e:5b:95:ad:23:6b:a0:5d:de:c2:91:bb:c7:2e:a7:94

Kex Algorithms:
        curve25519-sha256
        curve25519-sha256@libssh.org
        ecdh-sha2-nistp256
        ecdh-sha2-nistp384
        ecdh-sha2-nistp521
        sntrup761x25519-sha512@openssh.com
        diffie-hellman-group-exchange-sha256
        diffie-hellman-group16-sha512
        diffie-hellman-group18-sha512
        diffie-hellman-group14-sha256
        kex-strict-s-v00@openssh.com

Server Host Key Algorithms:
        rsa-sha2-512
        rsa-sha2-256
        ecdsa-sha2-nistp256
```



Filters Cheat Sheet

Shodan currently crawls nearly 1,500 ports across the Internet. Here are a few of the most commonly-used search filters to get started.

Filter Name	Description	Example
city	Name of the city	Devices in San Diego
country	2-letter Country code	Open ports in the United States
http.title	Title of the website	"Hacked" Websites
net	Network range or IP in CIDR notation	Services in the range of 8.8.0.0 to 8.8.255.255
org	Name of the organization that owns the IP space	Devices at Google
port	Port number for the service that is running	SSH servers
product	Name of the software that is powering the service	Samsung Smart TVs
screenshot.label	Label that describes the content of the image	Screenshots of Industrial Control Systems
state	U.S. State	Devices in Texas

Port 22 of NDSS website's IP run a OpenSSH_8.9p1

Search for device in city/port/product/screenshot...

Do they always play as white hat? What if they do something bad?🤔

Our Paper

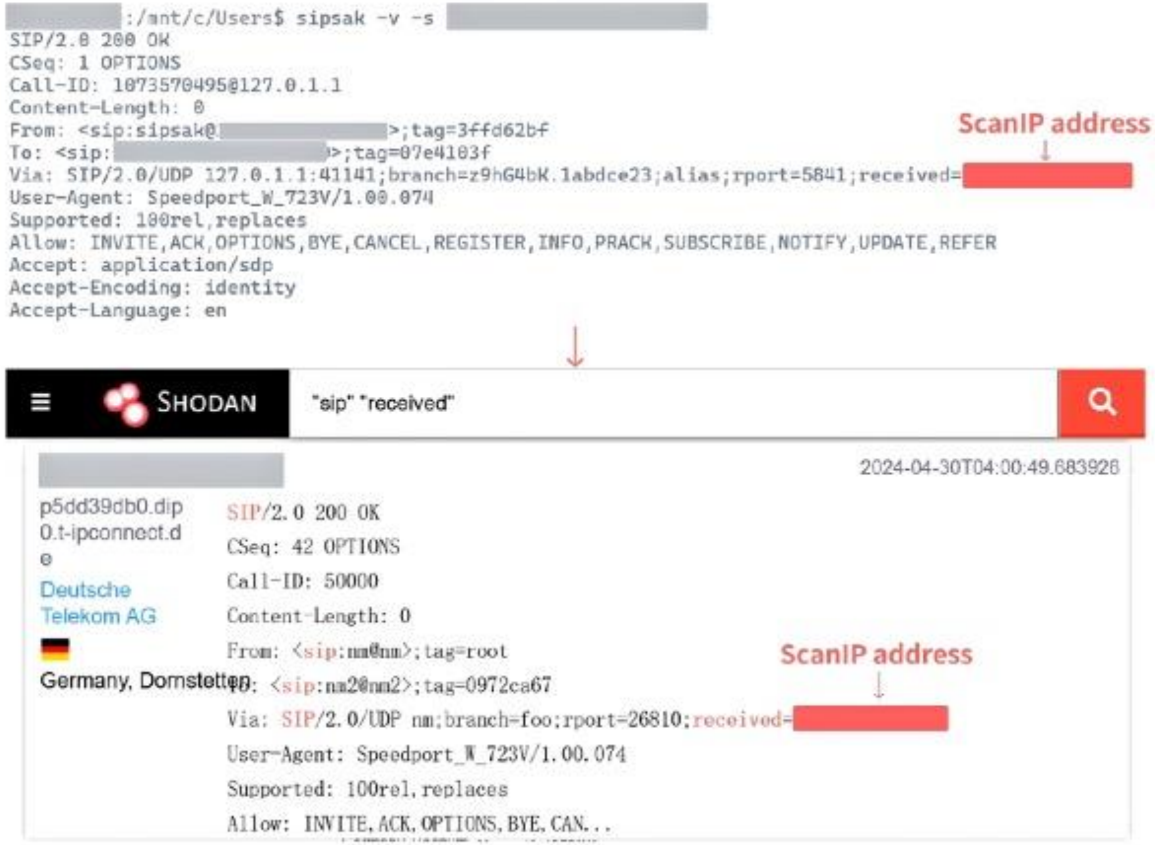
The first measurement study on
the working strategies of device search engines

01 What **scanning strategy** do device search engines apply?

02 How do device search engines **identify services** on ports?

03 Will the scanning of the device search engine introduce any **security or privacy concerns** to the services being scanned?

IP Mirror Service



```
[redacted]:/mnt/c/Users$ sipsak -v -s [redacted]
SIP/2.0 200 OK
CSeq: 1 OPTIONS
Call-ID: 1073570495@127.0.1.1
Content-Length: 0
From: <sip:sipsak@[redacted]>;tag=3ffd62bf
To: <sip:[redacted]>;tag=07e4103f
Via: SIP/2.0/UDP 127.0.1.1:41141;branch=z9hG4bK.1abdce23;alias;rport=5841;received=[redacted]
User-Agent: Speedport_W_723V/1.00.074
Supported: 100rel,replaces
Allow: INVITE,ACK,OPTIONS,BYE,CANCEL,REGISTER,INFO,PRACK,SUBSCRIBE,NOTIFY,UPDATE,REFER
Accept: application/sdp
Accept-Encoding: identity
Accept-Language: en
```

"sip" "received"

[redacted]

2024-04-30T04:00:49.683926

p5dd39db0.dip0.t-ipconnect.de
Deutsche Telekom AG
Germany, Domstetten

```
SIP/2.0 200 OK
CSeq: 42 OPTIONS
Call-ID: 50000
Content-Length: 0
From: <sip:nm@nm>;tag=root
To: <sip:nm2@nm2>;tag=0972ca67
Via: SIP/2.0/UDP nm;branch=foo;rport=26810;received=[redacted]
User-Agent: Speedport_W_723V/1.00.074
Supported: 100rel,replaces
Allow: INVITE,ACK,OPTIONS,BYE,CAN...
```

Interacting with a SIP server will reply the sender's IP

Main Challenge

Differentiate device search engine scanning activities from others

Insight

- Network services may include the visitor's IP for debugging, error prompting, or log metadata purposes.
- When device search engines scan those services, their IP addresses (ScanIP) are inevitably logged.

IP Mirror Service

IP Mirror Services are widely scanned and logged by device search engines

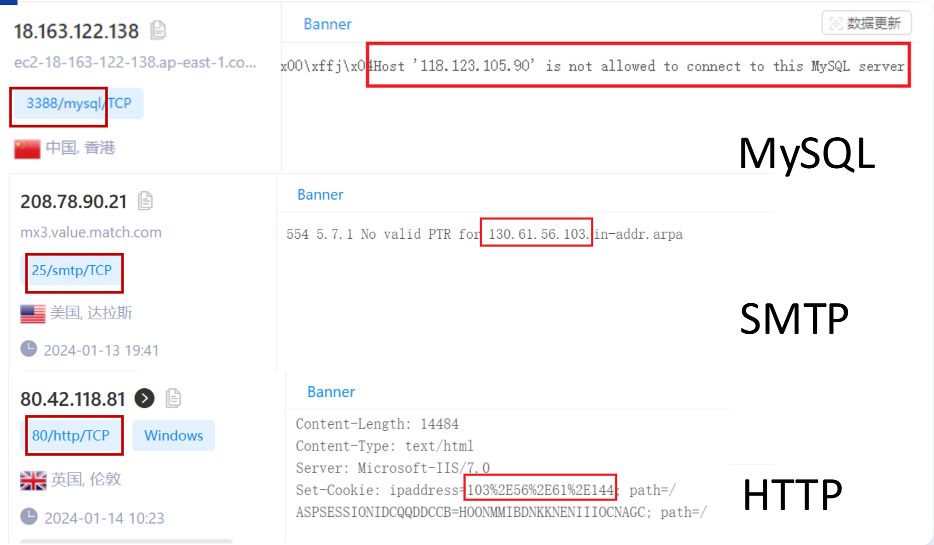
Engine	Country	Year	HTTP	MySql	SIP	SMTP	HTTP
			X-Forward-For	ERR_HOST	Received	No Valid PTR	Location
Shodan[2]	USA	2009	●	○	○	◑	◐
ZoomEye[15]	China	2013	●	●	●	◑	◐
Censys[13]	USA	2015	●	●	-	◑	◐
FOFA[14]	China	2015	●	○	○	◑	◐
BinaryEdge[27]	Switzerland	2015	●	●	●	◑	◐
Netlas[28]	Armenia	2021	●	●	-	-	-
Hunter [29]	China	2021	●	●	●	◑	-

Formats of IPs

Standard IP

Reverse IP

URL Encoding IP



```
x00\xffj\x04Host '118.123.105.90' is not allowed to connect to this MySQL server
```

MySQL

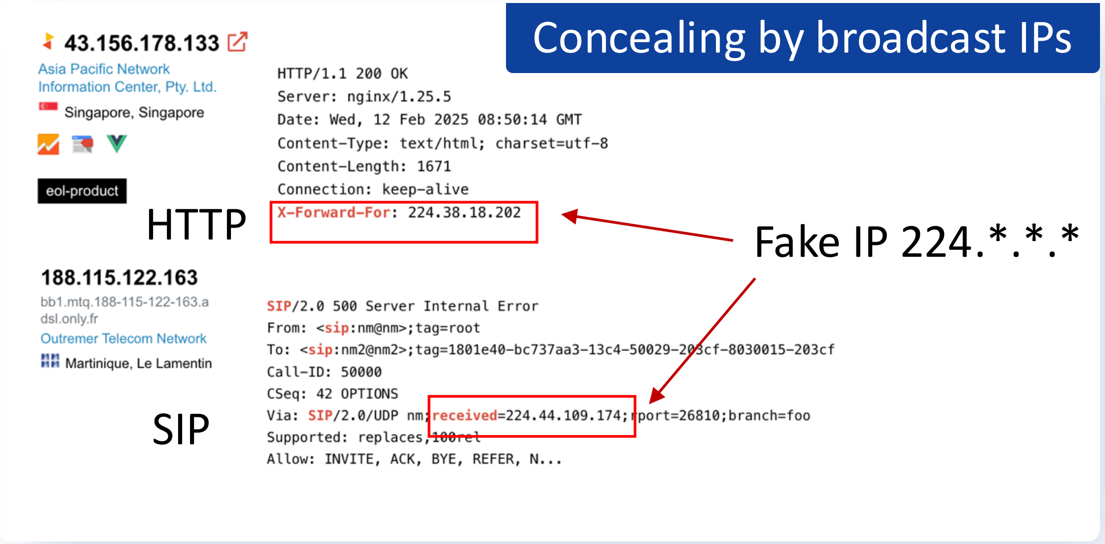
```
554 5.7.1 No valid PTR for 130.61.56.103.in-addr.arpa
```

SMTP

```
Content-Length: 14484
Content-Type: text/html
Server: Microsoft-IIS/7.0
Set-Cookie: ipaddress=103%2E56%2E61%2E144; path=/
ASPSESSIONIDCQQDDCCB=HOONMMIBDNKKNENIIIOCNAGC; path=/
```

HTTP

Concealing by broadcast IPs



HTTP

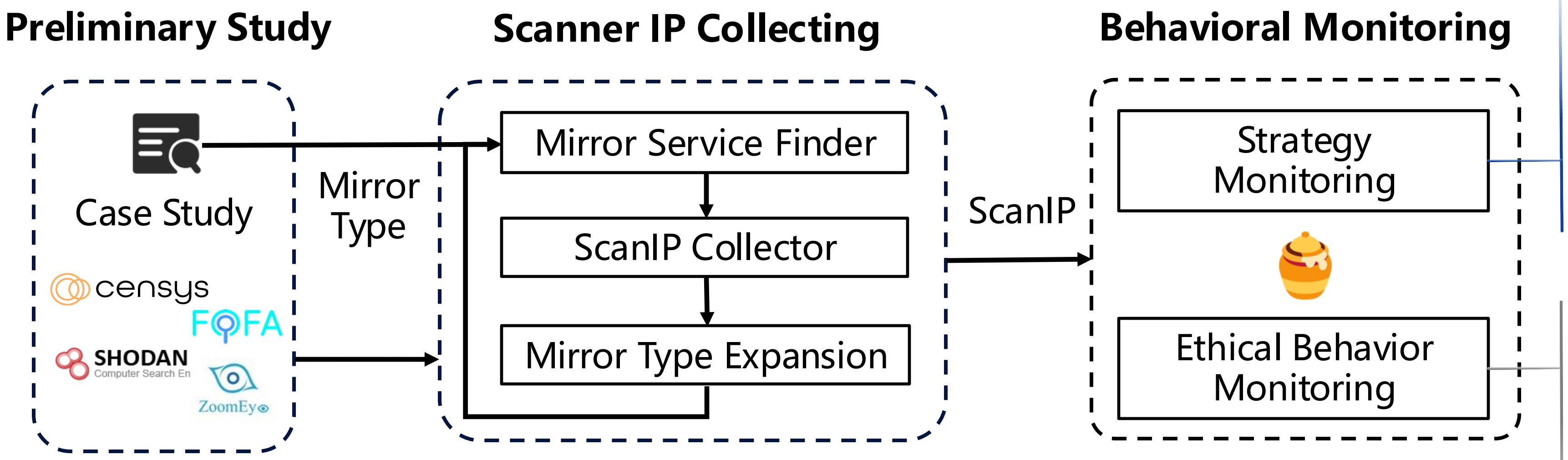
```
HTTP/1.1 200 OK
Server: nginx/1.25.5
Date: Wed, 12 Feb 2025 08:50:14 GMT
Content-Type: text/html; charset=utf-8
Content-Length: 1671
Connection: keep-alive
X-Forward-For: 224.38.18.202
```

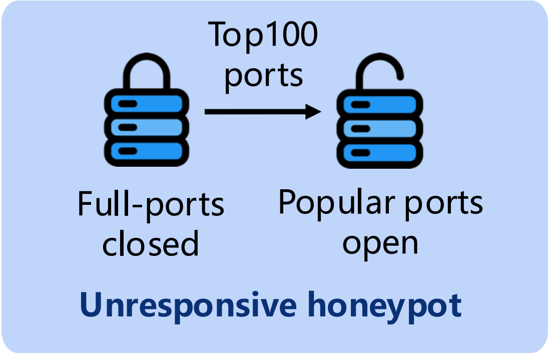
SIP

```
SIP/2.0 500 Server Internal Error
From: <sip:nm@nm>;tag=root
To: <sip:nm2@nm2>;tag=1801e40-bc737aa3-13c4-50029-203cf-8030015-203cf
Call-ID: 50000
CSeq: 42 OPTIONS
Via: SIP/2.0/UDP nm;received=224.44.109.174;rport=26810;branch=foo
Supported: replaces,100rel
Allow: INVITE, ACK, BYE, REFER, N...
```

Fake IP 224.*.*.*

Methodology





Unresponsive honeypot

Web honeypot

- Customized default pages for IoT devices
- Decoy sensitive paths with real-time snapshots
- Dynamic trackable links

79 **Mirror Types**

106,132 **Mirror Services**

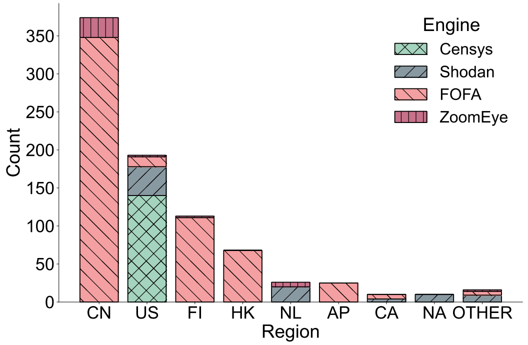
1,407 **Scanner IPs**

Censys	Shodan	FOFA	ZoomEye
481	91	668	167

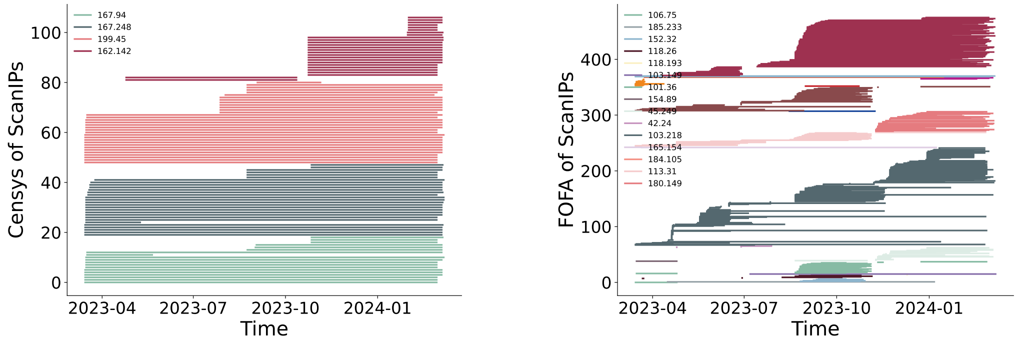
839 **IPs found in Honeypot**

Scan Strategy

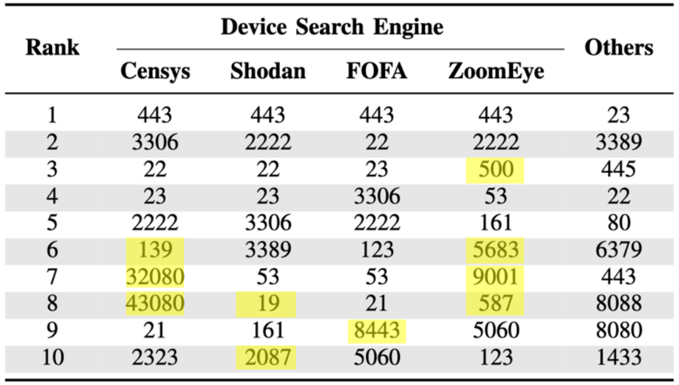
- Landscape
  - FOFA and ZoomEye do not use fixed scanning assets
    - Users can hardly avoid being scanned by blocklisting device search engine IPs
  - 665 ScanIPs have been labeled in AbuseIPDB by users
    - Tags: Port Scan Hacking Brute-Force



ScanIP Region Distribution



Lifespan of ScanIPs in Censys and FOFA



Rank	Device Search Engine				Others
	Censys	Shodan	FOFA	ZoomEye	
1	443	443	443	443	23
2	3306	2222	22	2222	3389
3	22	22	23	500	445
4	23	23	3306	53	22
5	2222	3306	2222	161	80
6	139	3389	123	5683	6379
7	32080	53	53	9001	443
8	43080	19	21	587	8088
9	21	161	8443	5060	8080
10	2323	2087	5060	123	1433

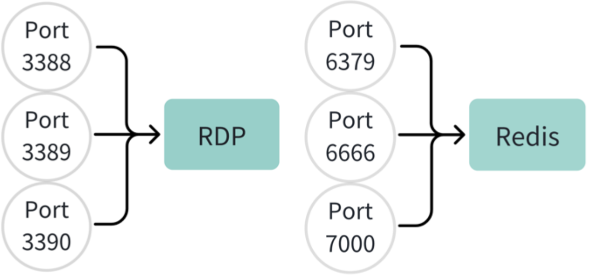
Top 10 ports scanned

Protocol Identification Strategy

Besides send protocol-specific probes to protocol default port, what other strategy do engines apply?

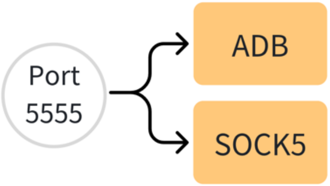
Neighbor Strategy

- Probe services on neighbor ports
- Users cannot evade scans by migrating ports of services they wish to hide!



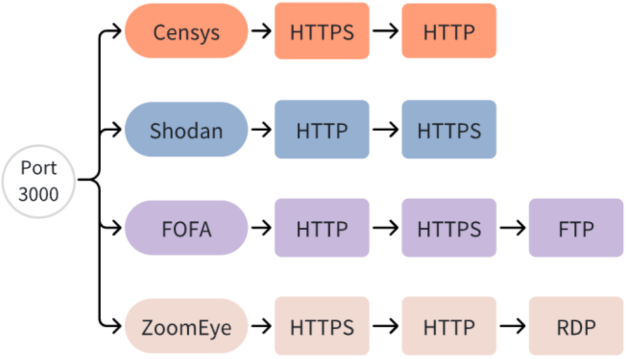
Share Strategy

- Multiple probes from various potential protocols to the same port

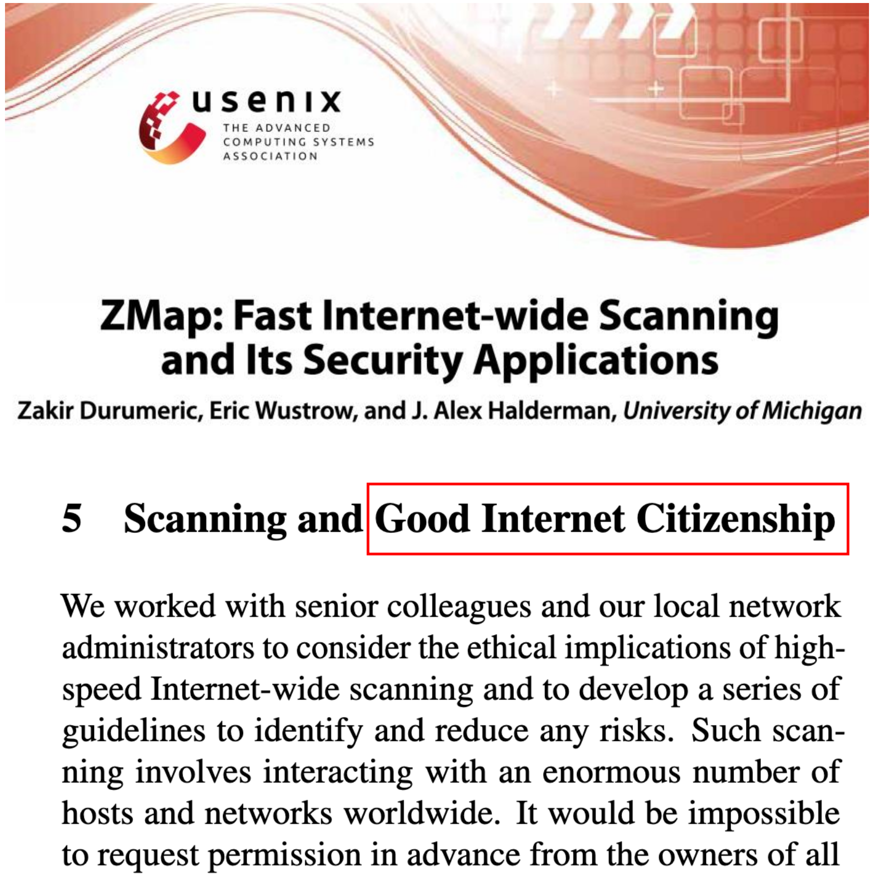


Fallback Strategy

- When fail to identify protocol on specific ports, they guess …



Ethical Scanning



usenix
THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

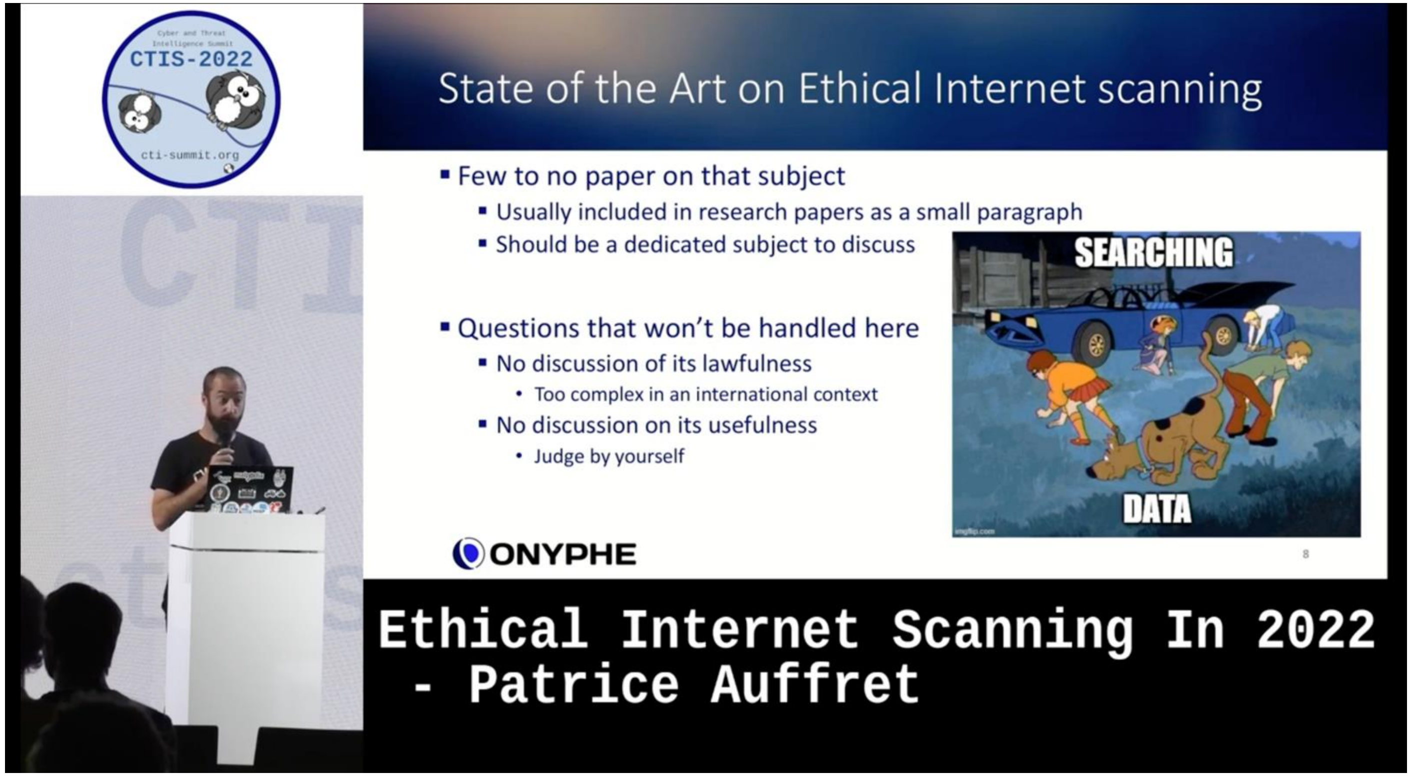
ZMap: Fast Internet-wide Scanning and Its Security Applications

Zakir Durumeric, Eric Wustrow, and J. Alex Halderman, *University of Michigan*

5 Scanning and Good Internet Citizenship

We worked with senior colleagues and our local network administrators to consider the ethical implications of high-speed Internet-wide scanning and to develop a series of guidelines to identify and reduce any risks. Such scanning involves interacting with an enormous number of hosts and networks worldwide. It would be impossible to request permission in advance from the owners of all

Two engines have already considered ethical things



State of the Art on Ethical Internet scanning

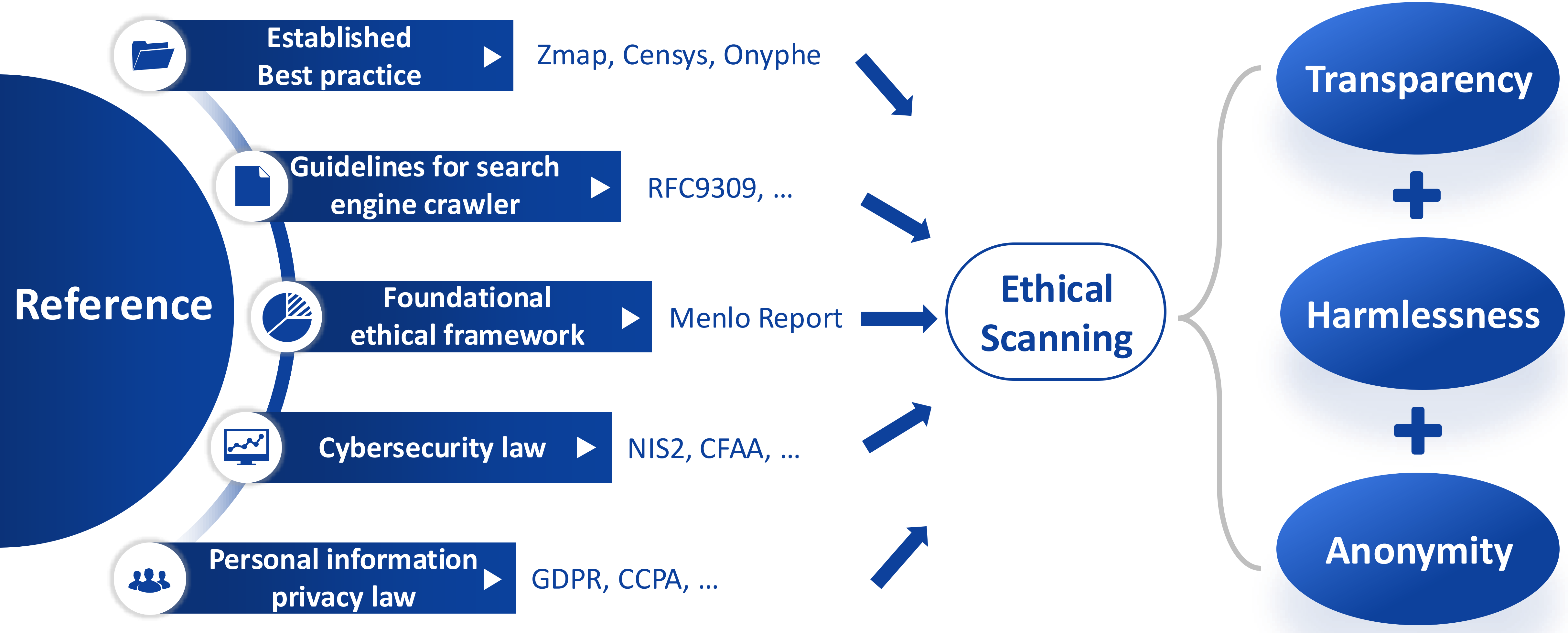
- Few to no paper on that subject
  - Usually included in research papers as a small paragraph
  - Should be a dedicated subject to discuss
- Questions that won't be handled here
  - No discussion of its lawfulness
    - Too complex in an international context
  - No discussion on its usefulness
    - Judge by yourself

ONYPHE

Ethical Internet Scanning In 2022
- Patrice Auffret

Our goal: To find an Ethical Way of Internet Scanning

Ethical Scanning



Reference

Established Best practice — Zmap, Censys, Onyphe

Guidelines for search engine crawler — RFC9309, …

Foundational ethical framework — Menlo Report

Cybersecurity law — NIS2, CFAA, …

Personal information privacy law — GDPR, CCPA, …

Ethical Scanning

Transparency + Harmlessness + Anonymity

Transparency

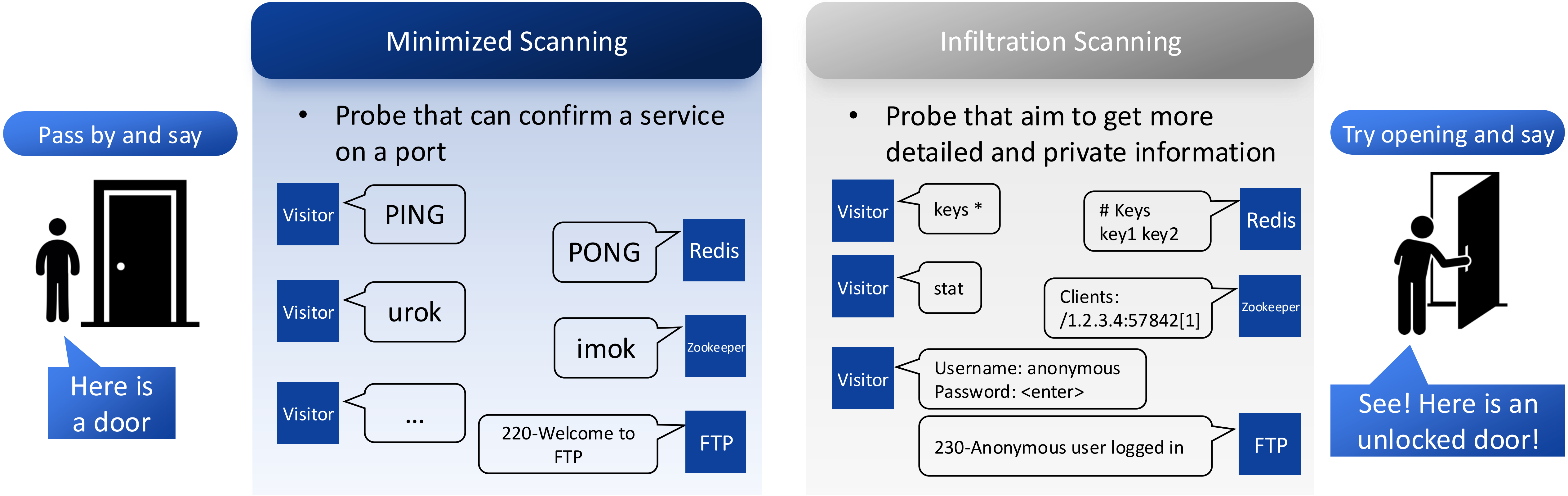
Action	Censys	Shodan	FOFA	ZoomEye
Explain the purpose on every probe	😐	😈	😈	😈
Publish probes IP address list for opt-out	🙂	😈	😈	😈
Use fixed IP addresses instead of trashable ones	🙂	🙂	😈	😈
Set whois records with organization and abuse email	🙂	😐	😈	😈
Reverse DNS pointing to the company	😐	😐	😈	😈

Censys proposed it
But not implemented

Users cannot identify whether the scans originate from FOFA or ZoomEye

Harmlessness

- Unauthorized access
  - Attempt to access paths requiring authentication but are left insecure
  - Engines do not adhere to data minimization principles during scanning



Pass by and say

Here is a door

Minimized Scanning

- Probe that can confirm a service on a port

Infiltration Scanning

- Probe that aim to get more detailed and private information

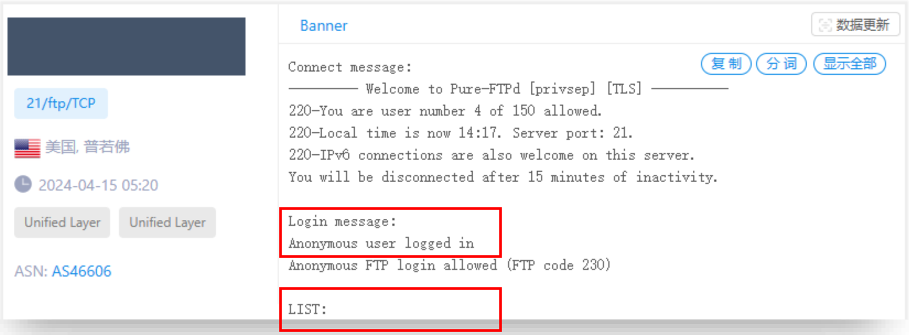
Try opening and say

See! Here is an unlocked door!

Harmlessness

- Engines are infiltrating database, nodes, configurations, file lists, …
- Successful infiltrations exposed weakly protected hosts lacking authentication
  - 74.97% (59,725/79,664) of Redis hosts listed on Shodan and 182,137 hosts on Fofa are vulnerable to arbitrary access
  - 99.91% Zookeeper hosts are vulnerable to unauthenticated access
- Shodan attempted to access and retain 25 sensitive paths for IP camera configuration details and real-time feeds
- The probe used for RDP (except Censys) is exploiting a vulnerability (CVSS3 score: 9.8)

Harmlessness



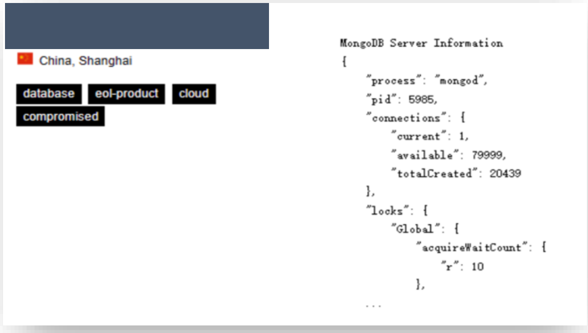
[redacted]

```
Connect message:
———————— Welcome to Pure-FTPd [privsep] [TLS] ————————
220-You are user number 4 of 150 allowed.
220-Local time is now 14:17. Server port: 21.
220-IPv6 connections are also welcome on this server.
You will be disconnected after 15 minutes of inactivity.

Login message:
Anonymous user logged in
Anonymous FTP login allowed (FTP code 230)

LIST:
```

ZoomEye login and list file on FTP



[redacted]

```
MongoDB Server Information
{
    "process": "mongod",
    "pid": 5985,
    "connections": {
        "current": 1,
        "available": 79999,
        "totalCreated": 20439
    },
    "locks": {
        "Global": {
            "acquireWaitCount": {
                "r": 10
            },
    ...
```

MongoDB server info

Type	Action		Censys	Shodan	FOFA	ZoomEye
Harmlessness[2]	Malformed requests		🙂	🙂	🙂	😈
	Unauthorized Access Service	**Minimized Probe**				
	FTP	Null Probe	🙂	😈	😈	😈
	Redis	Command: ping	😈	😈	😈	😈
	ZooKeeper	Command: ruok	😈	😈	😈	😈
	ElasticSearch	Path: /	😈	😈	😈	🙂
	MongoDB	Command: mongo	😈	😈	😈	😈
	RDP	RDP Handshake	🙂	😈	😈	😈
	LDAP	LDAP Handshake	😈	😈	🙂	🙂
	Memcached	Command: stats	🙂	😈	🙂	🙂
	CouchDB	Path: /	🙂	😈	😈	🙂
	IP Camera(Web Service)	Path: /	🙂	😈	🙂	🙂
	OpenWrt Router(Web Service)	Path: /	🙂	🙂	🙂	😈
	Prometheus(Web Service)	Path: /	😈	🙂	🙂	🙂

Engine	Type	Path
Censys	Web(Prometheus)	/api/v1/label/goversion/values /api/v1/label/goversion/values /api/v1/query /api/v1/labels /api/v1/label/__name__/values /api/v1/targets /api/v1/label/version/values /api/v1/status/config /tr064dev.xml /api/json
Shodan	IoT(IP Camera)	/cgi-bin/authLogin.cgi /filestation/wfm2Login.cgi /photo /video /snapshot.cgi /cgi-bin/viewer/video.jpg /cgi-bin/snapshot.cgi /snapshot.jpg /tmpfs/auto.jpg /cgi-bin/view/image /axis-cgi/jpg/image.cgi /ipcam/jpeg.cgi /ISAPI/Streaming/channels/101/picture /jpg/image.jpg /Streaming/channels/1/picture /Streaming/channels/101 /image/jpeg.cgi /img/snapshot.cgi /-wvhttp-01-/GetLiveImage /-wvhttp-01-/GetOneShot /videostream.cgi /get_status.cgi /videostream.asf /cgi-bin/video_snapshot.cgi /snap.jpg
FOFA	Web(Elasticsearch)	/_cat/indices
ZoomEye	IoT(OpenWrt Router)	/cgi-bin/luci/ /studio/index.html

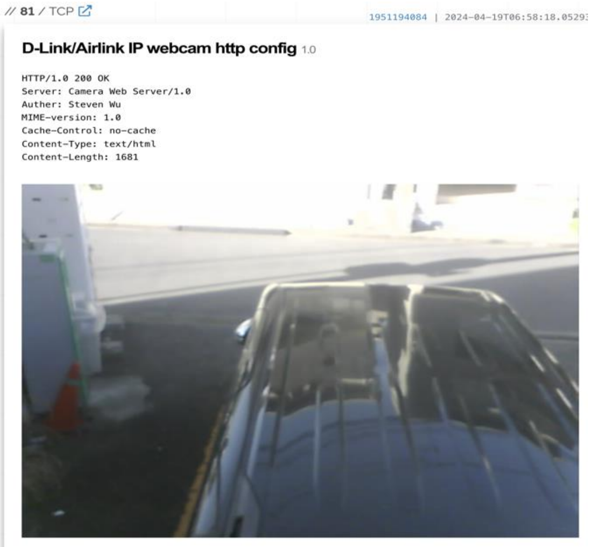
Sensitive path access caught by honeypots

Anonymity

- Failure to anonymize the privacy before displaying on search results can lead to privacy leakage risks.

Type	Action	Censys	Shodan	FOFA	ZoomEye
Anonymity[3]	FTP	😐	😐	😐	😐
	Redis	😐	😈	😐	😐
	ZooKeeper	😐	😐	😐	😐
	ElasticSearch	😐	😈	😈	😐
	MongoDB	😐	😈	😈	😐
	RDP	🙂	😈	😈	😈
	LDAP	😈	😈	🙂	🙂
	Memcached	😐	😐	😐	😐
	CouchDB	😐	😈	😈	😐
	IP Camera	🙂	😈	🙂	🙂

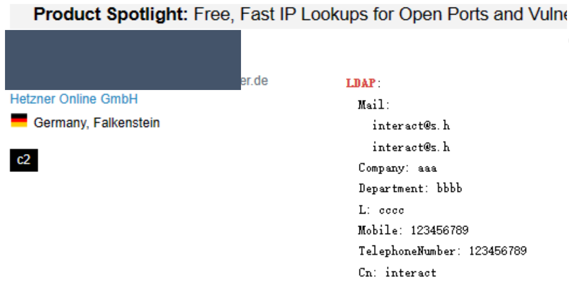
😐 means *version* is leaked


// 81 / TCP

1951194084 | 2024-04-19T06:58:18.0529

D-Link/Airlink IP webcam http config 1.0

```
HTTP/1.0 200 OK
Server: Camera Web Server/1.0
Auther: Steven Wu
MIME-version: 1.0
Cache-Control: no-cache
Content-Type: text/html
Content-Length: 1681
```


Product Spotlight: Free, Fast IP Lookups for Open Ports and Vulne

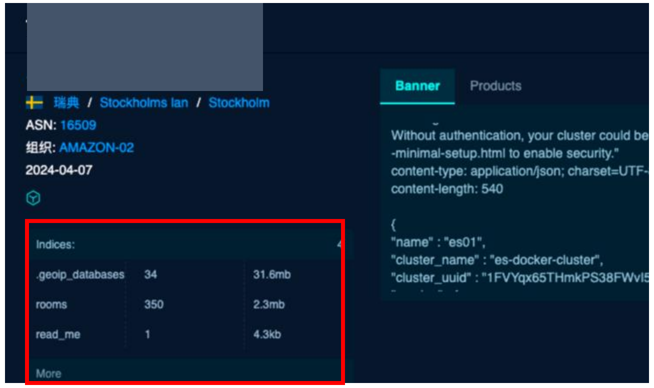
[redacted]er.de

Hetzner Online GmbH

Germany, Falkenstein

c2

```
LDAP:
  Mail:
    interact@s.h
    interact@s.h
  Company: aaa
  Department: bbbb
  L: cccc
  Mobile: 123456789
  TelephoneNumber: 123456789
  Cn: interact
```


[redacted]

瑞典 / Stockholms lan / Stockholm

ASN: 16509

组织: AMAZON-02

2024-04-07

```
Without authentication, your cluster could be
-minimal-setup.html to enable security."
content-type: application/json; charset=UTF-
content-length: 540

{
"name" : "es01",
"cluster_name" : "es-docker-cluster",
"cluster_uuid" : "1FVYqx65THmkPS38FWvl5
```

```
Indices:
.geoip_databases   34    31.6mb
rooms              350   2.3mb
read_me            1     4.3kb
More
```

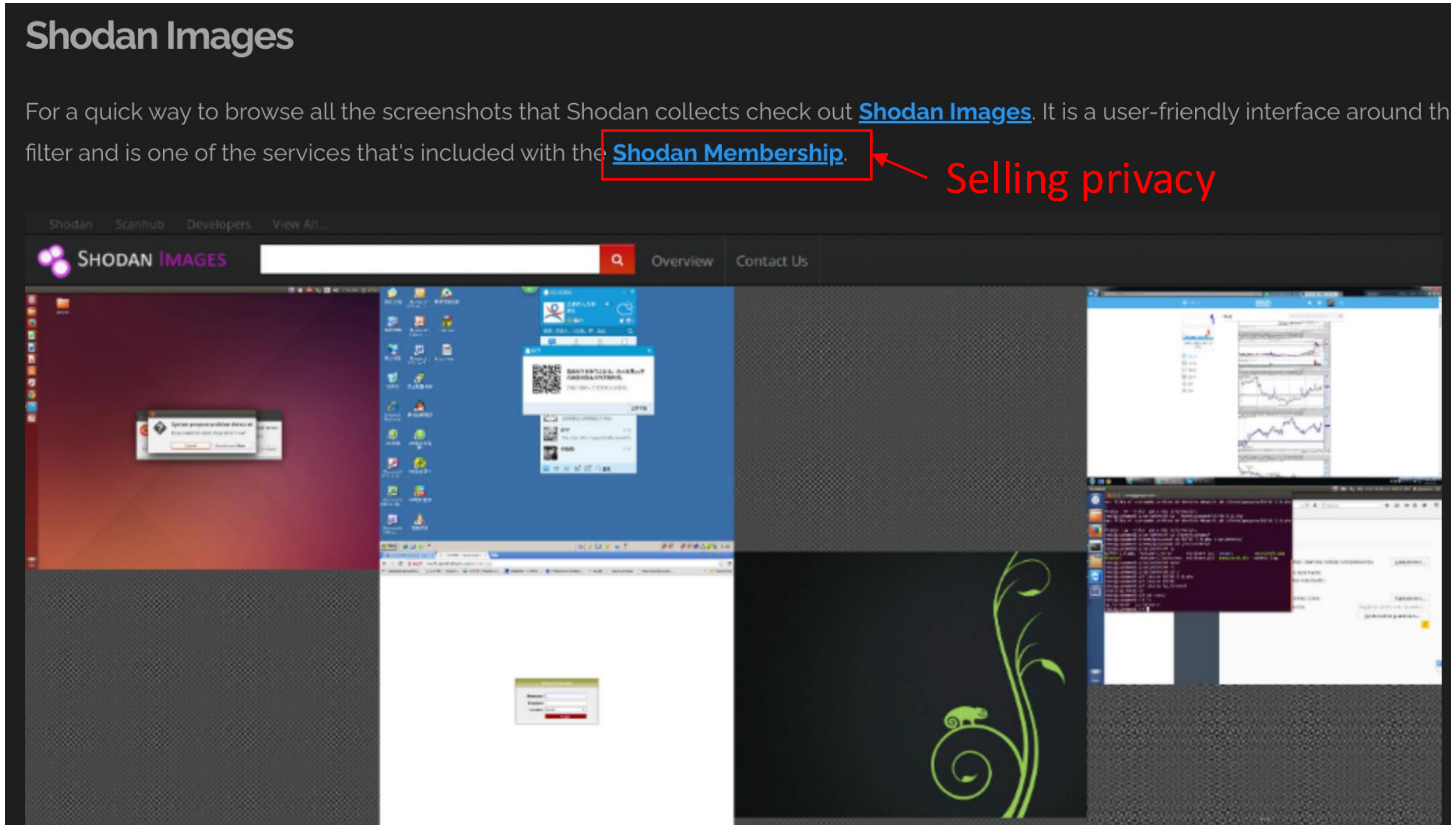
735 Phone number from LDAP

326,495 Database index and entries

68,543 Redis keys

Anonymity: images.shodan.com

Introduction by Shodan: a quick way to browse all the screenshots



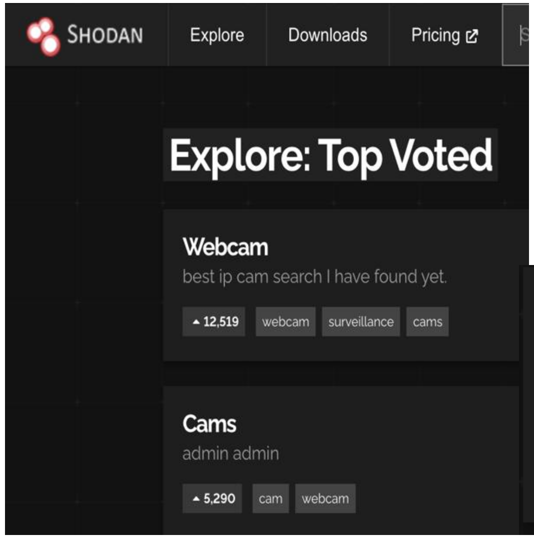
Shodan Images

For a quick way to browse all the screenshots that Shodan collects check out Shodan Images. It is a user-friendly interface around th filter and is one of the services that's included with the Shodan Membership.

Selling privacy

65,042 **Camera snapshots**

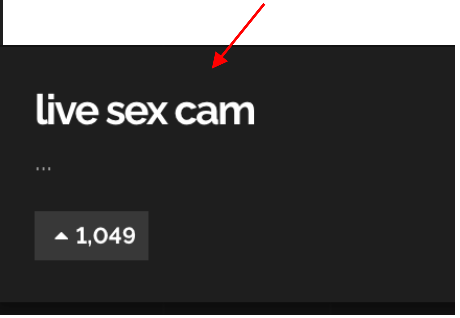
791,333 **Remote desktop screenshot**



Explore: Top Voted

Webcam
best ip cam search I have found yet.
▲ 12,519 webcam surveillance cams

Cams
admin admin
▲ 5,290 cam webcam



live sex cam
...
▲ 1,049

“Explore” module: shared queries

Rank 7

Attackers abuse it for illicit camera spying and exacerbate the sale of voyeuristic content

Take Away

- Discover *Mirror Services* that can reflect scanner IPs of device search engines and uncover 1,407 scanner IPs.
- The first comprehensive analysis of the scan strategy of device search engines, proving that users cannot evade scans by blocklisting scanner IPs.
- Unveil how device search engines identify protocol on ports, offering insights into how users can hide their services.
- First ethical scanning analysis, uncovering instances where engines conceal their identities, engage in unauthorized access, and expose user camera interfaces.

Thank you for your Audience!

For more details, welcome to follow our paper.

Revealing the Black Box of Device Search Engine: Scanning Assets, Strategies, and Ethical Consideration

Mengying Wu[†*], Geng Hong[†*], Jinsong Chen[†], Qi Liu[†], Shujun Tang[‡§], Youhao Li[‡], Baojun Liu[§],
Haixin Duan[§¶] and Min Yang[†]
[†]Fudan University, China, {wumy21,jschen23,qiliu21}@m.fudan.edu.cn, {ghong,m_yang}@fudan.edu.cn
[‡]QI-ANXIN Technology Research Institute, China, liyouhao@qianxin.com
[§]Tsinghua University, China, tsj23@mails.tsinghua.edu.cn, {lbj, duanhx}@tsinghua.edu.cn
[¶]Quancheng Laboratory, China

***Abstract*—In the digital age, device search engines such as Censys and Shodan play crucial roles by scanning the internet to catalog online devices, aiding in the understanding and mitigation of network security risks. While previous research has used these tools to detect devices and assess vulnerabilities, there remains uncertainty regarding the assets they scan, the strategies they employ, and whether they adhere to ethical guidelines.**

This study presents the first comprehensive examination of these engines' operational and ethical dimensions. We developed a novel framework to trace the IP addresses utilized by these engines and collected 1,407 scanner IPs. By uncovering their IPs, we gain deep insights into the actions of device search engines for the first time and gain original findings. By employing 28 honeypots to monitor their scanning activities extensively in one year, we demonstrate that users can hardly evade scans by blocklisting scanner IPs or migrating service ports. Our

employed these engines to collect data on resident IP addresses [3], electric vehicle charging management systems [4], and insecure industrial control systems (ICS) [5].

Attackers can abuse the powerful scanning capabilities of such engines to identify vulnerable devices and establish zombie networks for malicious activities like cryptocurrency mining [6]. It is estimated that the over-collection of data by Shodan-like services led to a loss of approximately $3.86 million in 2020 alone [7]. Moreover, it remains uncertain whether these engines consider ethical implications while striving to provide competitive network assessment reports. Users who care about security and privacy have started to take action, including reporting abusive scanning IPs to AbuseIPDB [8], a public IP blocklist, and moving services from default ports to other ports. To the best of our knowledge, there has been