Silent Whispers of Malware: Unveiling Hidden Threats in Legitimate Network Traffic

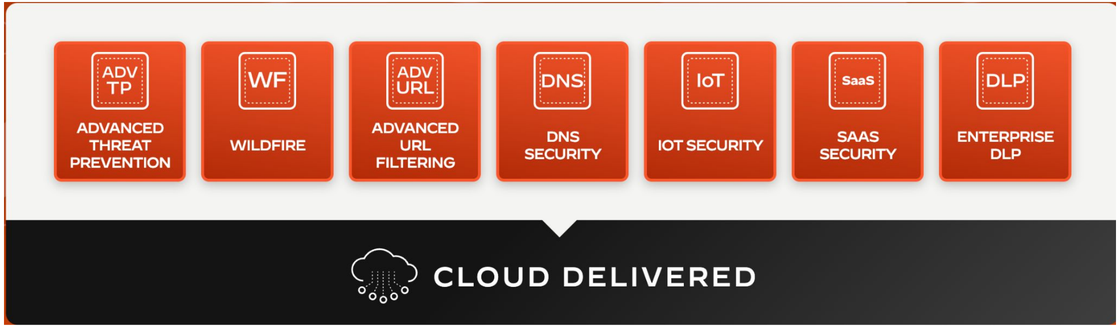
Zhanhao Chen, Chao Lei, Royce Lu, Daiping Liu

Palo Alto Networks

Virus Bulletin 2023 London

About US

CDSS (Cloud-Delivered Security Services)



ADV TP ADVANCED THREAT PREVENTION

WF WILDFIRE

ADV URL ADVANCED URL FILTERING

DNS DNS SECURITY

IoT IOT SECURITY

SaaS SAAS SECURITY

DLP ENTERPRISE DLP

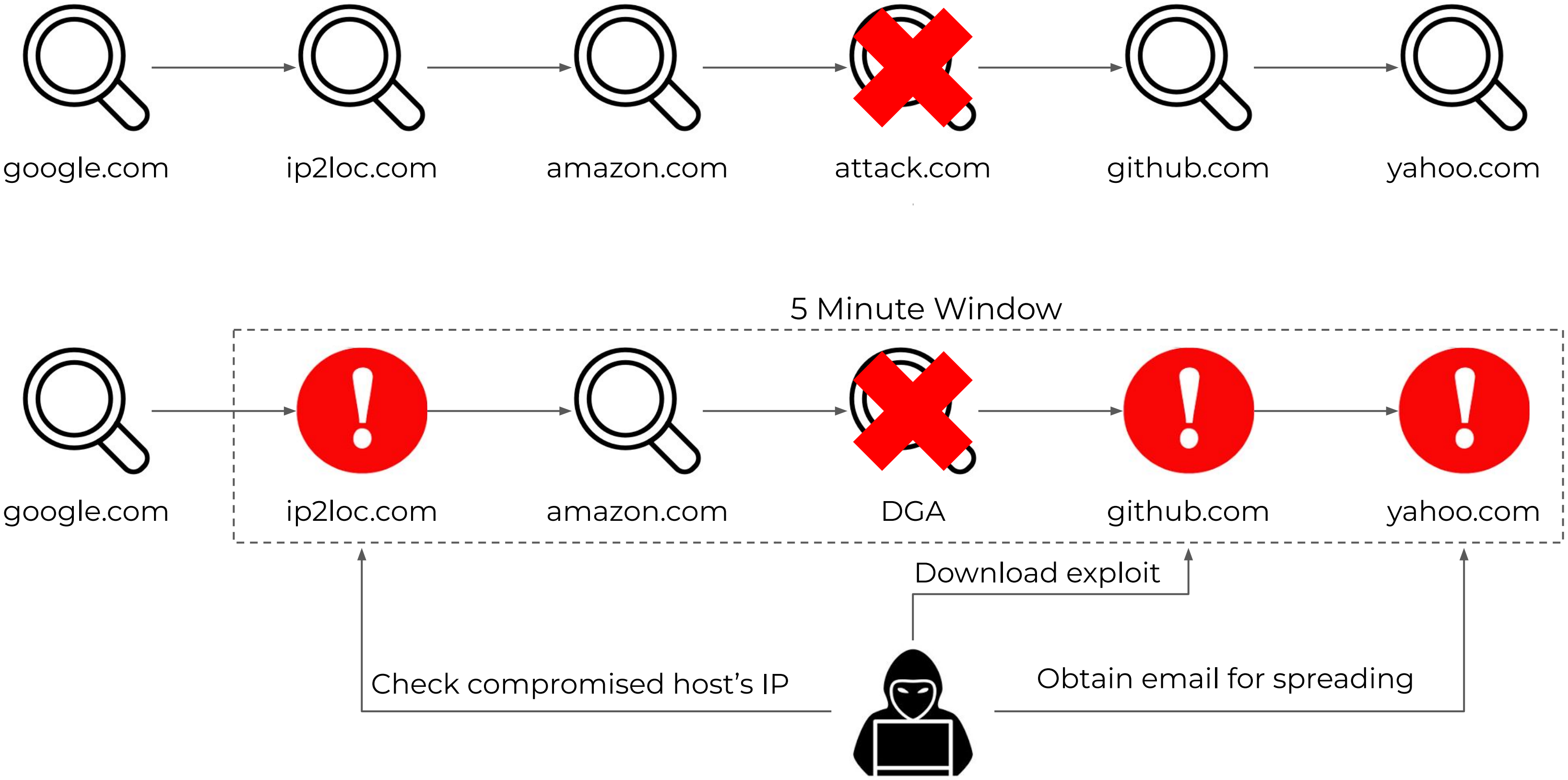
CLOUD DELIVERED

- DNS security
- Malware Analysis - Wildfire (Sandbox)
- Network Security - Advanced Threat Prevention

Agenda

- Problem
- Algorithm
- Evaluation
- Use cases

Motivation


google.com
ip2loc.com
amazon.com
attack.com
github.com
yahoo.com
5 Minute Window
google.com
ip2loc.com
amazon.com
DGA
github.com
yahoo.com
Download exploit
Check compromised host's IP
Obtain email for spreading

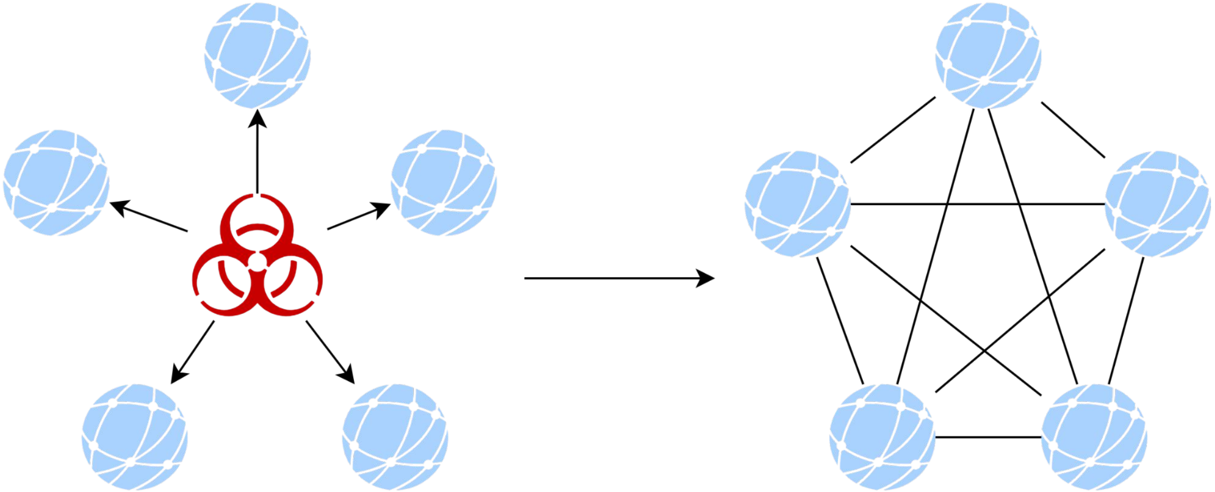
Key Questions

- Data Source: What data can reveal the abuse for legitimate networks services ?
  - A: Sandbox analysis pcap, reverse engineering reports, security blogs…
- Signature Generation: How to select the most indicative combinations of URLs ?
  - A: Use greedy graph expansion algorithm to efficiently generate candidate signatures based on their abuse levels.
- Quality Control: How to ensure the signatures won’t false alert legitimate activities?
  - A: Cross check the candidate signatures against real-world network traffic
- Integration: Where could we deploy the signatures?
  - A: Endpoint Security, Malware Sandbox, DNS Security, URL Filtering, Intrusion Detection System…etc

Algorithm Design for signature generation

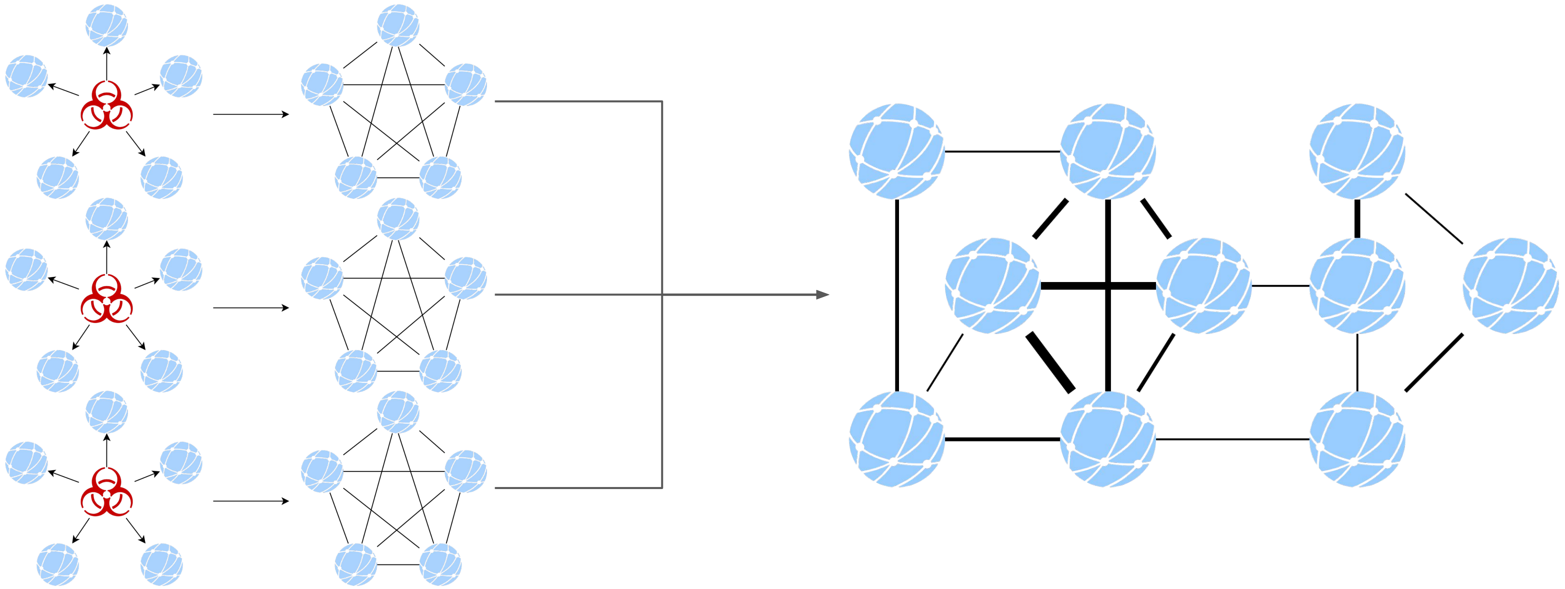
- For malicious URL, one single connection is enough to make decision.
- For one benign URL connection, even it has high malware rate, it is not enough to make decision.
- How about we start from one URL, and then combine other benign URLs that also has high malware rate ?
- If we look at the URLs as nodes of a graph, then we transfer the sequence pattern generation question into the problem to find the subgraph with the highest abuse level

Methodology - Data Ingestion



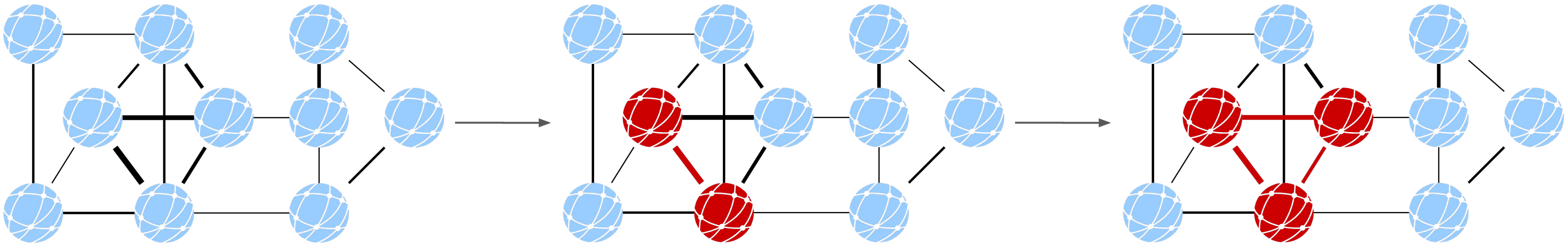
- Extract attacking network traffic from the sandbox analysis reports
- Adding relations (edge) among the benign URLs (node) abused by malware
  - If sample only connect to node A or only node B, no edge
  - If sample connect to node A and node B, add one edge
  - Because the graph is extracted from the analysis result, it will be a fully connected graph
- If a URL already has bad reputation, skip adding the node

Methodology - URL Graph Construction



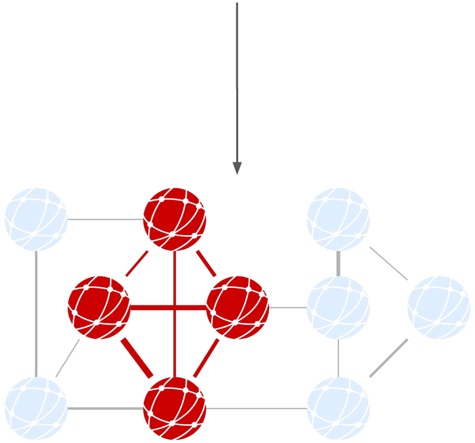
- Aggregate the relations into a graph whose edges are weighted by abuse level
  - Aggregated graph may not be a fully connected graph, unless every Node was visited by all samples.
- Accumulate malware rate (# malware, # benign) for network services relations (edge)
- Filter out the relations that are related to a lot benign samples as they could cause FP detection.

Methodology - Graph Expansion based Signature Generation

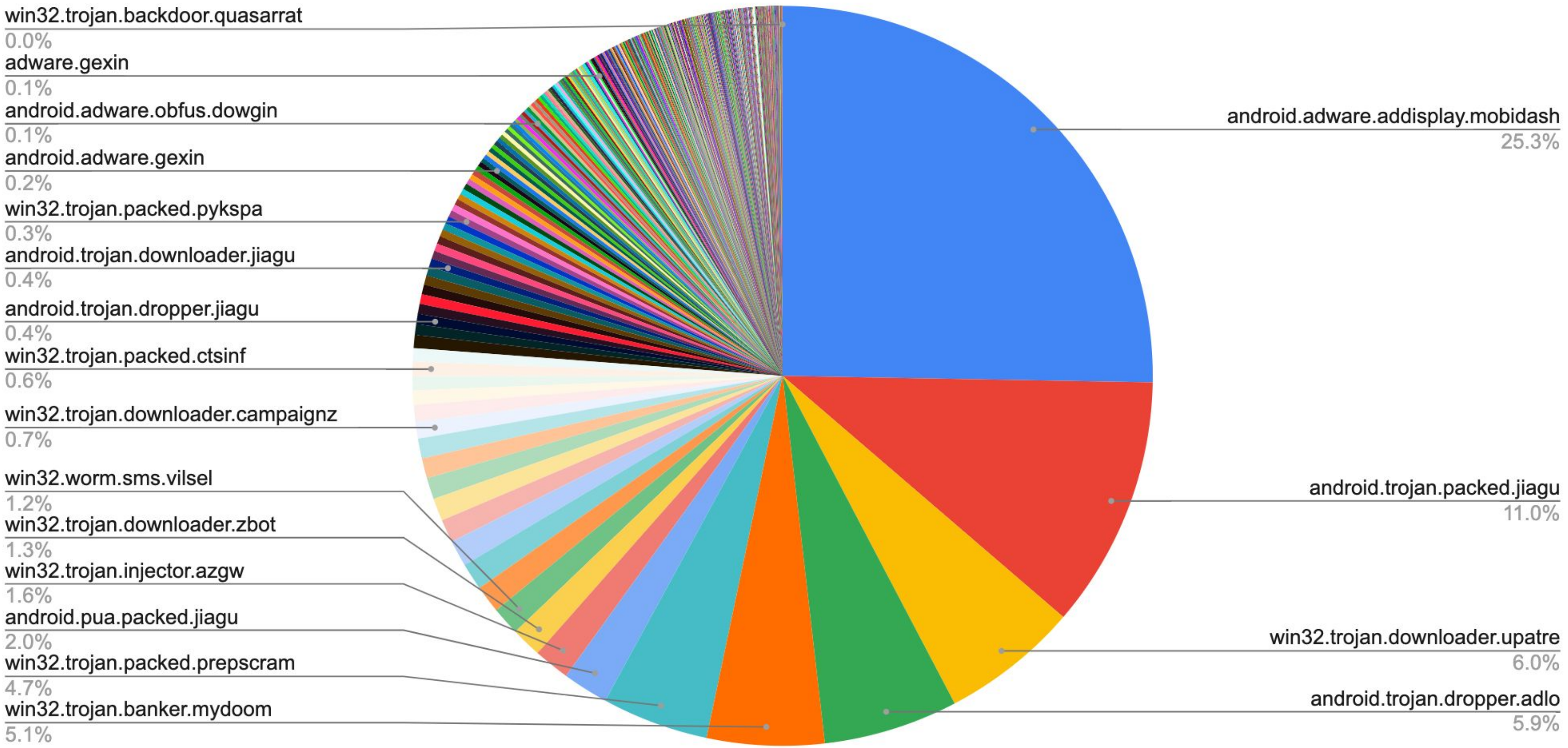


1. Graph expansion algorithm: Add node that connect all existing nodes
2. Greedy expansion: Add the node with highest abuse level

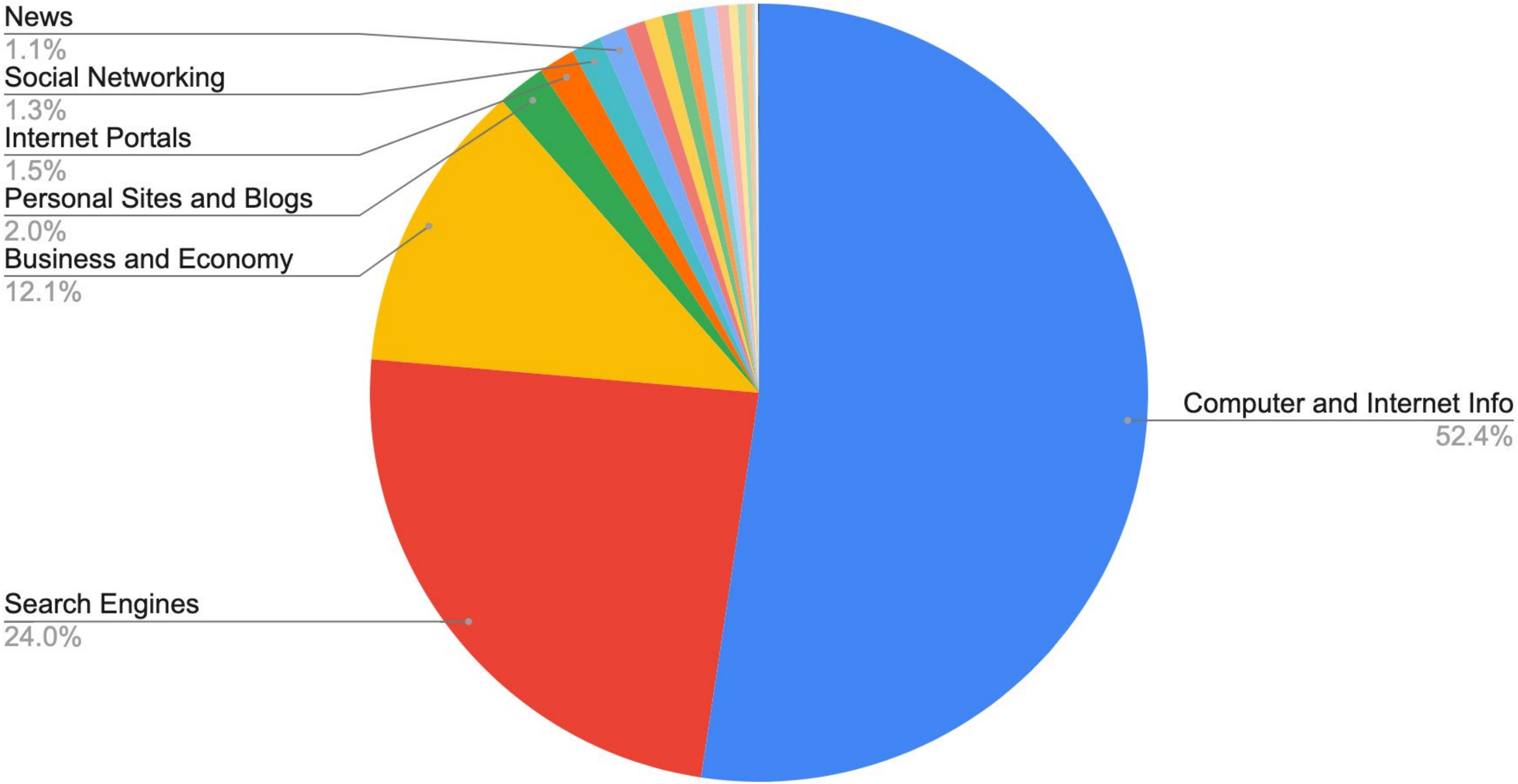
```
PickNextNode()
for each node that connect to all the red nodes.
        if sum(node.edges.abuse_level) > current_max
                MaxNode = node
                current_max = sum(node.edges.abuse_level)
graph.add(MaxNode)
```



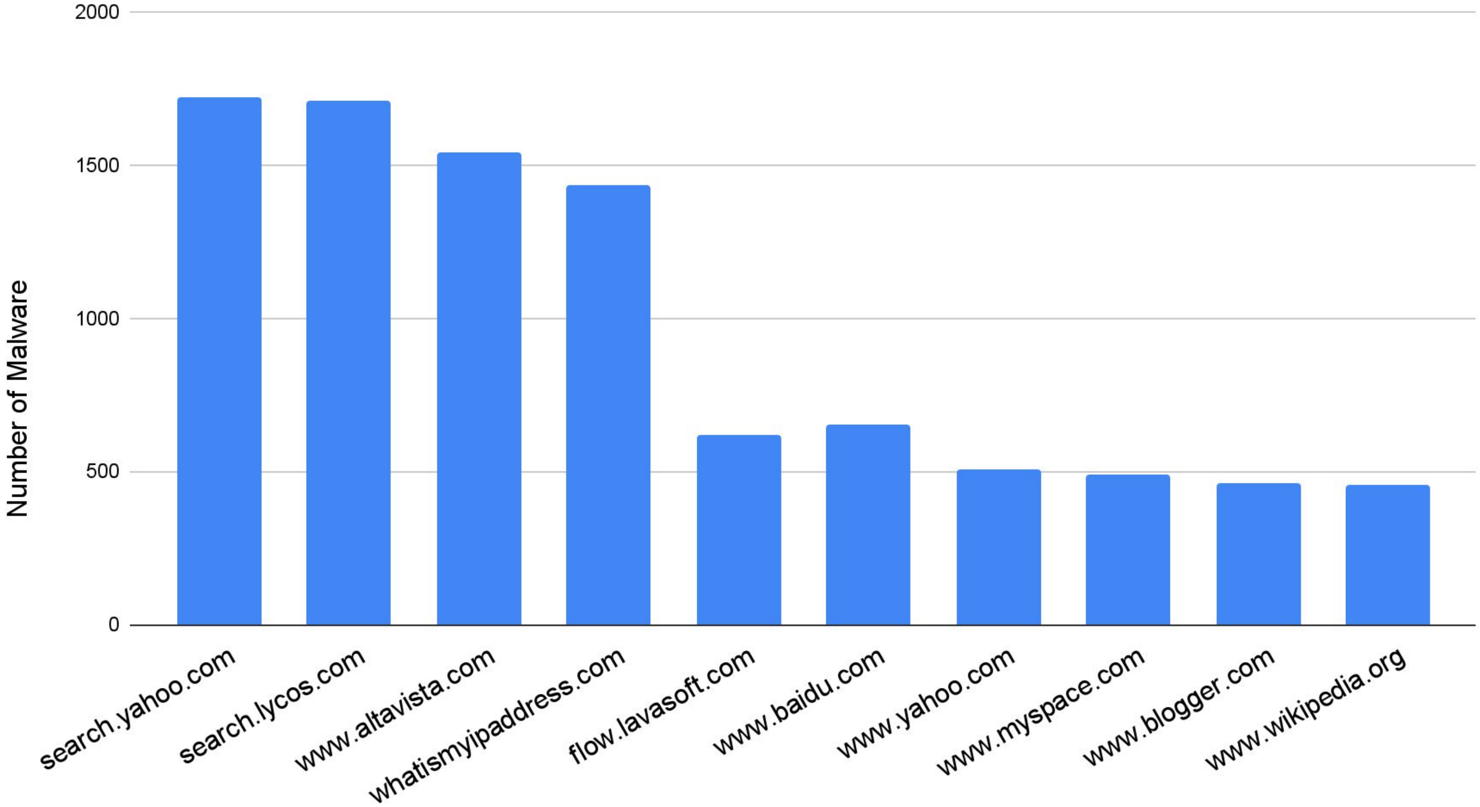
Statistics - Malware Families



Statistics - Most Abused Services Categories

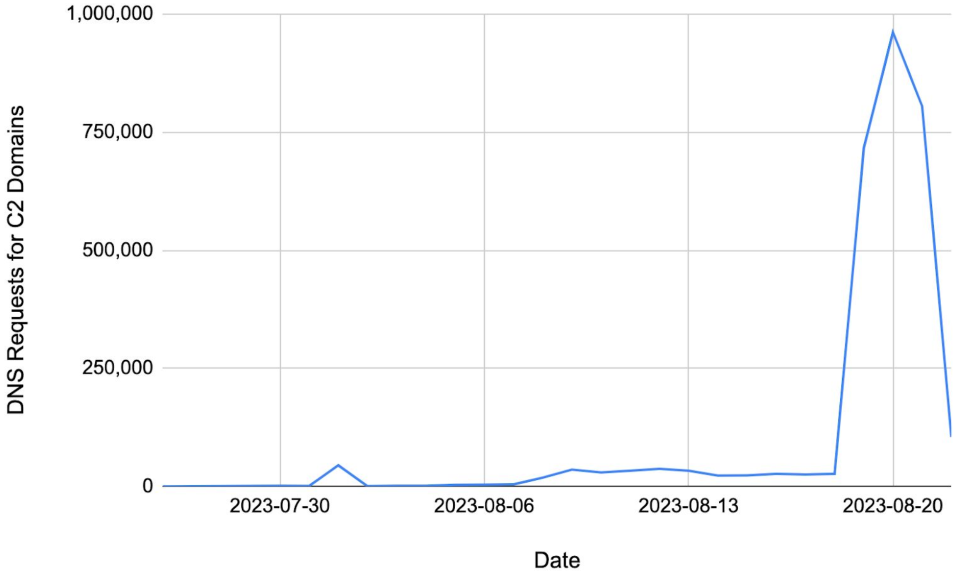


Statistics - Most Abused Services

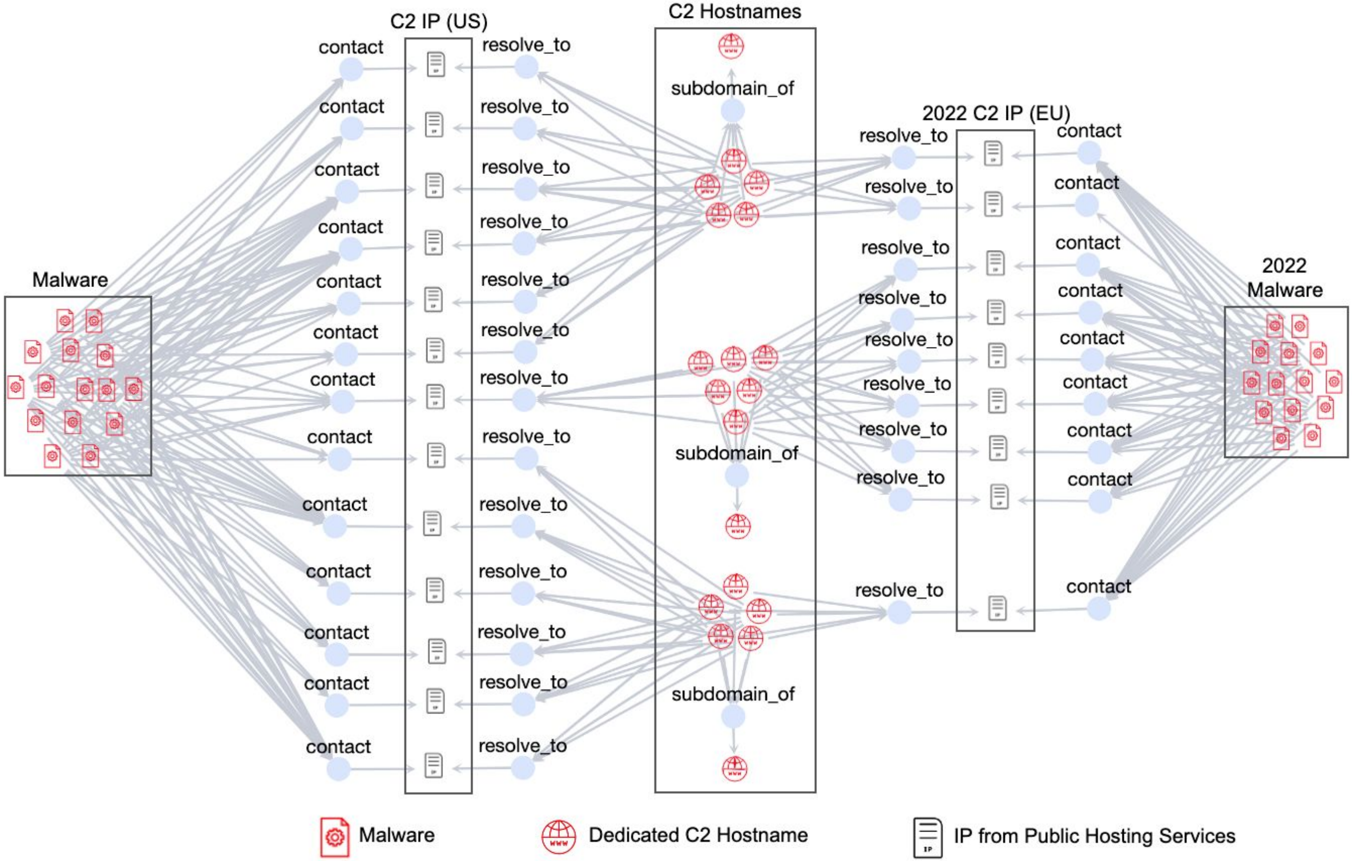


Case Study 1: XorDDoS Trojan

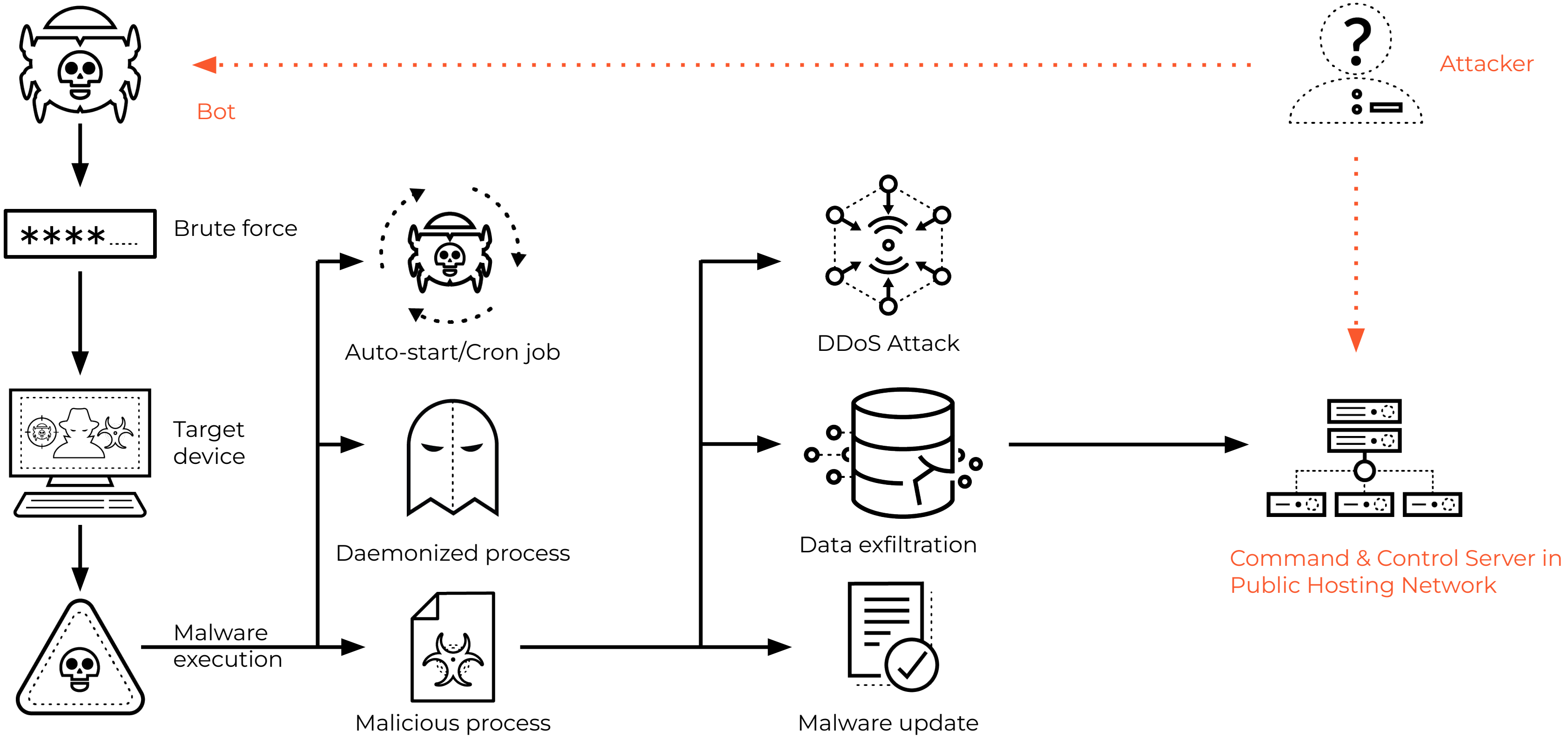
- Signature Combination: 142.0.138[.]41, 142.0.138[.]42, 142.4.106[.]75, 192.74.236[.]33, 192.74.236[.]34, 192.74.236[.]36
- Malware Count: 84
- Malware Family: DDoS Trojan/Rootkit



Case Study 1: XorDDoS Trojan



Case Study 1: XorDDoS Trojan

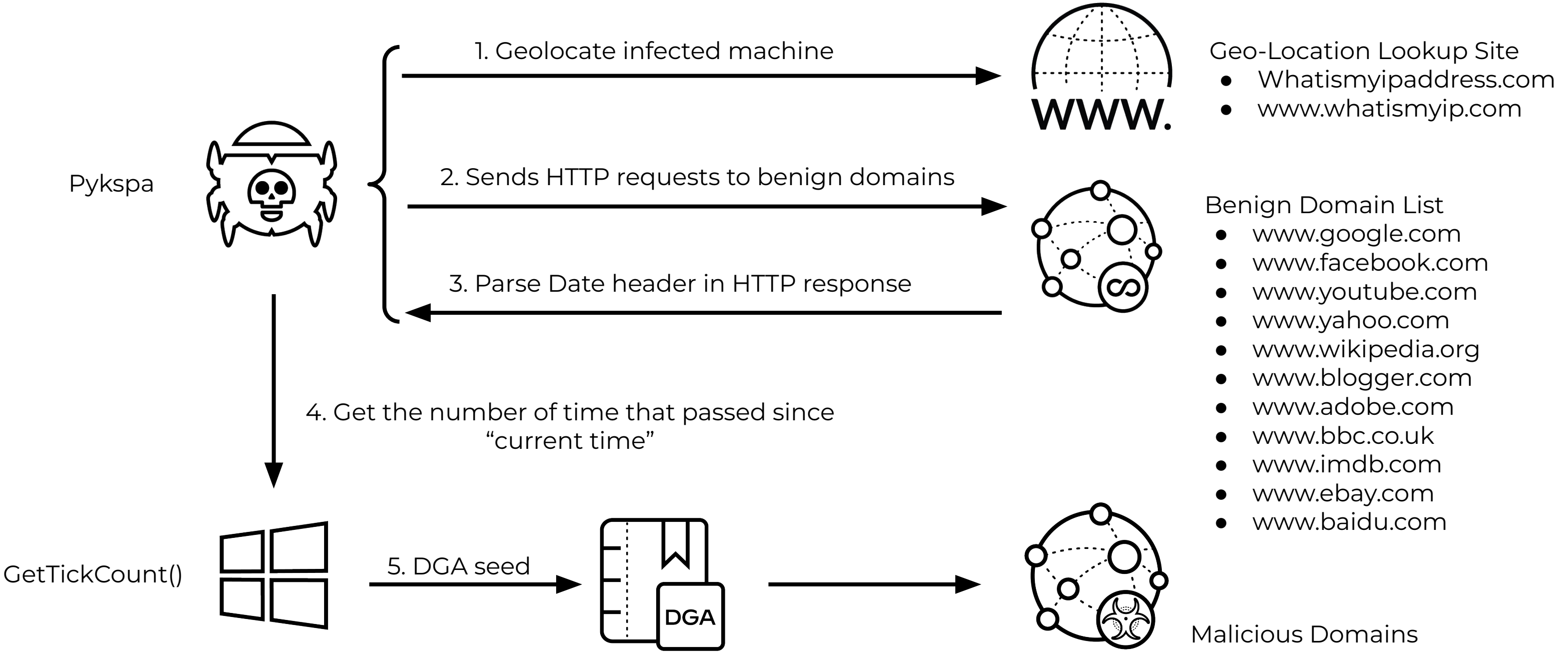


Reference: https://www.microsoft.com/en-us/security/blog/2022/05/19/rise-in-xorddos-a-deeper-look-at-the-stealthy-ddos-malware-targeting-linux-devices/

Case Study 2: Pykspa

- Signature: whatismyipaddress[.]com, www.whatismyip[.]com, www.wikipedia[.]org
- Malware Count: 729
- Harvest skype contact list
- Spreading through Skpy message
- Malware Family: worm

Case Study 2: Pykspa



Pykspa

1. Geolocate infected machine

Geo-Location Lookup Site

- Whatismyipaddress.com
- www.whatismyip.com

2. Sends HTTP requests to benign domains

Benign Domain List

- www.google.com
- www.facebook.com
- www.youtube.com
- www.yahoo.com
- www.wikipedia.org
- www.blogger.com
- www.adobe.com
- www.bbc.co.uk
- www.imdb.com
- www.ebay.com
- www.baidu.com

3. Parse Date header in HTTP response

4. Get the number of time that passed since “current time”

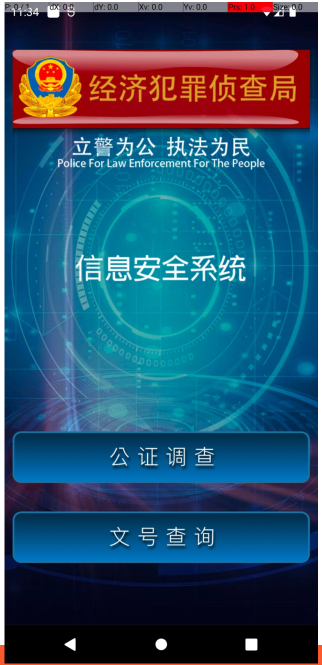
GetTickCount()

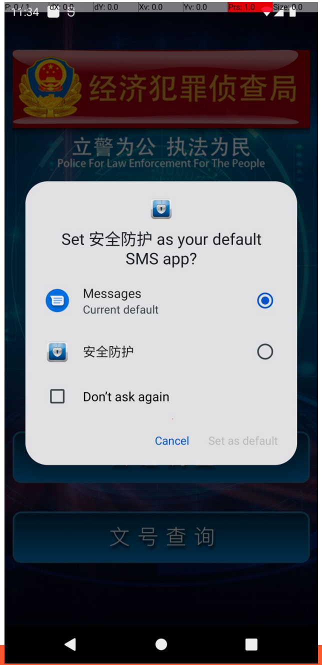
5. DGA seed

Malicious Domains

Case Study 3: Financial Fraud APK

- Hostname Combination: www.xunfei[.]com, soft.tbs.imtt.qq[.]com, dev.voicecloud[.]cn
- Malware Count: 49
- Malware Family: Financial Fraud Application





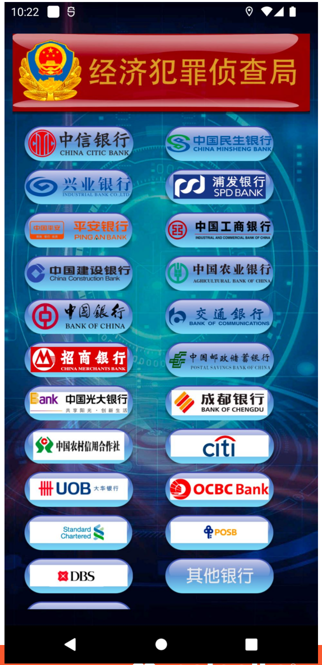
Set 安全防护 as your default SMS app?

Messages
Current default ●

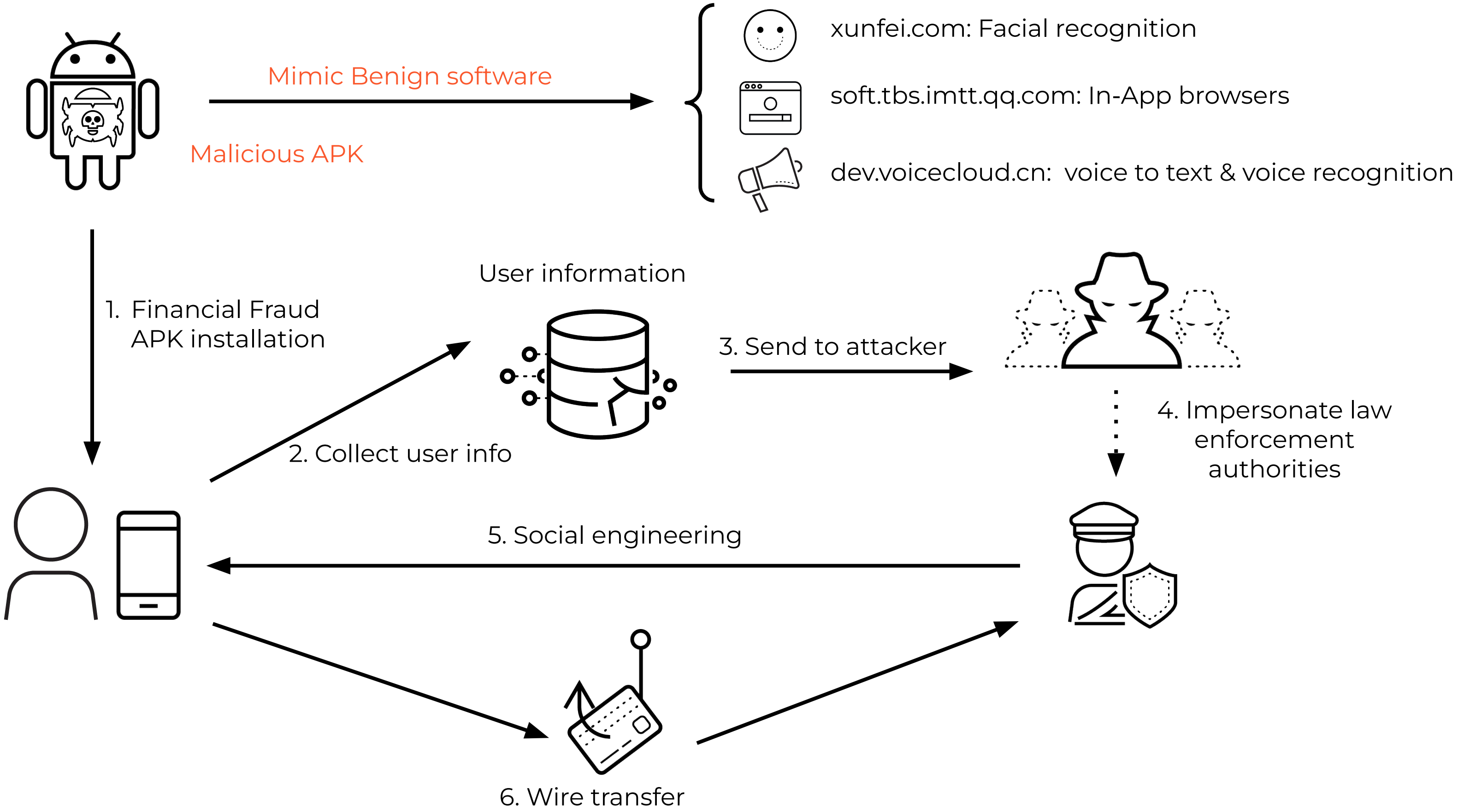
安全防护 ○

□ Don't ask again

Cancel Set as default

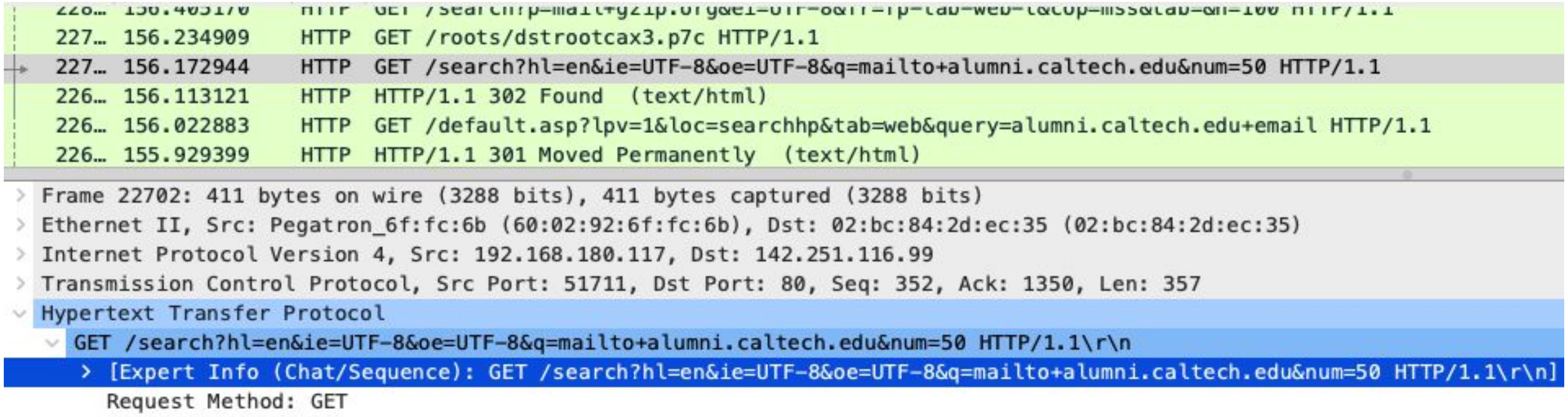


Case Study 3: Financial Fraud


Mimic Benign software
Malicious APK
xunfei.com: Facial recognition
soft.tbs.imtt.qq.com: In-App browsers
dev.voicecloud.cn: voice to text & voice recognition
1. Financial Fraud APK installation
User information
2. Collect user info
3. Send to attacker
4. Impersonate law enforcement authorities
5. Social engineering
6. Wire transfer

Case Study 4: MyDoom

- Hostname Combination: search.lycos[.]com, search.yahoo[.]com, www.altavista[.]com, www.google[.]com/search
- Malware Count: 1724
- Malware Family: Worm
- Analysis:
  - Retrieve Email lists from search engine for spreading.



```
228… 156.405170   HTTP  GET /search?p=mail+gzip.org&ei=UTF-8&fr=fp-tab-web-t&cop=mss&tab=&n=100 HTTP/1.1
227… 156.234909   HTTP  GET /roots/dstrootcax3.p7c HTTP/1.1
227… 156.172944   HTTP  GET /search?hl=en&ie=UTF-8&oe=UTF-8&q=mailto+alumni.caltech.edu&num=50 HTTP/1.1
226… 156.113121   HTTP  HTTP/1.1 302 Found  (text/html)
226… 156.022883   HTTP  GET /default.asp?lpv=1&loc=searchhp&tab=web&query=alumni.caltech.edu+email HTTP/1.1
226… 155.929399   HTTP  HTTP/1.1 301 Moved Permanently  (text/html)

> Frame 22702: 411 bytes on wire (3288 bits), 411 bytes captured (3288 bits)
> Ethernet II, Src: Pegatron_6f:fc:6b (60:02:92:6f:fc:6b), Dst: 02:bc:84:2d:ec:35 (02:bc:84:2d:ec:35)
> Internet Protocol Version 4, Src: 192.168.180.117, Dst: 142.251.116.99
> Transmission Control Protocol, Src Port: 51711, Dst Port: 80, Seq: 352, Ack: 1350, Len: 357
v Hypertext Transfer Protocol
  v GET /search?hl=en&ie=UTF-8&oe=UTF-8&q=mailto+alumni.caltech.edu&num=50 HTTP/1.1\r\n
    > [Expert Info (Chat/Sequence): GET /search?hl=en&ie=UTF-8&oe=UTF-8&q=mailto+alumni.caltech.edu&num=50 HTTP/1.1\r\n]
      Request Method: GET
```

Takeaways

- Besides blocking traffic to known malicious URLs, monitoring network traffic directed to legitimate services is crucial for comprehensive network security.
- We build an automatic pipeline to extract combination of legitimate network entities from sandbox analysis pcaps as indicators of compromise (IOC) that can efficiently detect malware traffic.
- These IOC can be integrated to various security platform to identify sophisticated that exploit legitimate network services.

Q & A