Catching Transparent Phish: Analyzing and Detecting MITM Phishing Toolkits

Brian Kondracki, **Babak Amin Azad**, Oleksii Starov and Nick Nikiforakis

PragSec Lab

paloalto NETWORKS

STONY BROOK UNIVERSITY

The Value of Stolen Data

Spotify Account
$2.75

Hulu Account
$2.75

Netflix Account
$1.00 - $3.00

Driver's License
$20.00

Credit Card
$8.00 - $22.00

Email Address & Password
$0.70 - $2.30

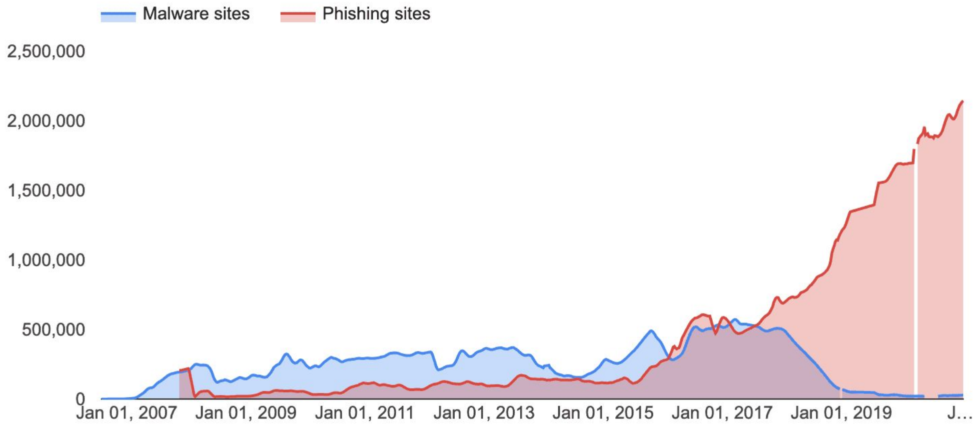
PayPal Credentials
$1.50

Social Security Number
$1.00

Medical Record from Large Scale Attack
$1.50 - $10.00

Complete Medical Record
Up to $1000.00

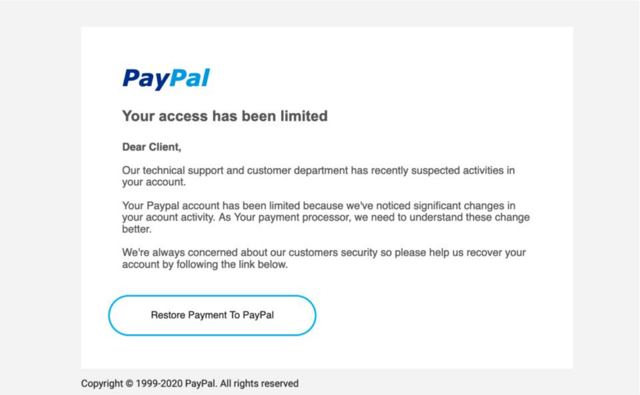
Phishing vs. Malware



https://www.keepersecurity.com/how-much-is-my-information-worth-to-hacker-dark-web.html

Anatomy of a Traditional Phishing Attack

- Attackers manually copy/recreate web content from target website
- Phishing content served from attacker-owned web server
  - Or a compromised web server
- Links to phishing webpages dispatched to victims through email or SMS



PayPal

Your access has been limited

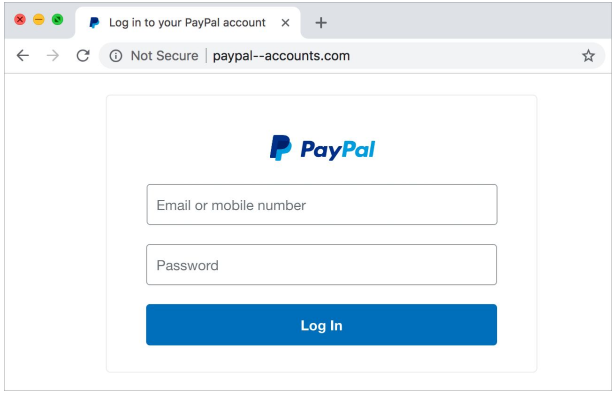
Dear Client,

Our technical support and customer department has recently suspected activities in your account.

Your Paypal account has been limited because we've noticed significant changes in your acount activity. As Your payment processor, we need to understand these change better.

We're always concerned about our customers security so please help us recover your account by following the link below.

Copyright © 1999-2020 PayPal. All rights reserved

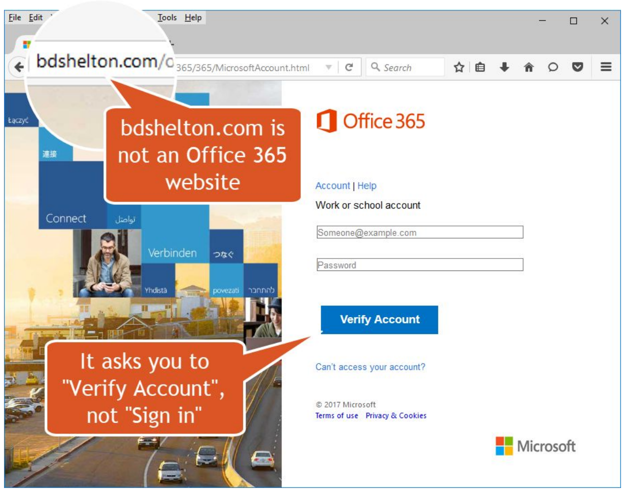


Log in to your PayPal account

Not Secure | paypal--accounts.com

Limitations of Traditional Phishing

- Implementation errors can lead to detection
- Webpages update at increasing speeds
- Detection by anti-phishing scanners leads to immediate blocklisting



bdshelton.com is not an Office 365 website

Account | Help

Work or school account

Someone@example.com

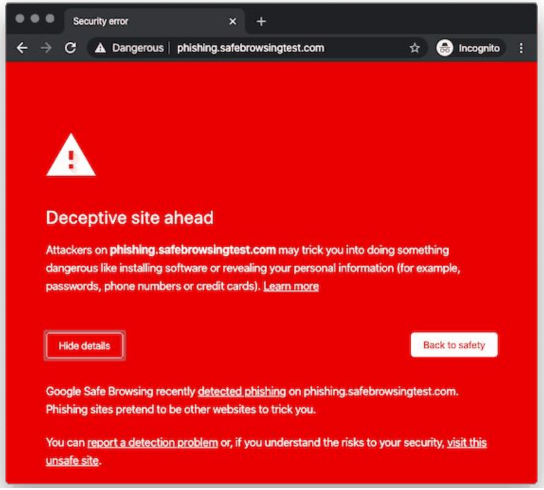
Password

Can't access your account?

© 2017 Microsoft

Terms of use Privacy & Cookies

It asks you to "Verify Account", not "Sign in"



Deceptive site ahead

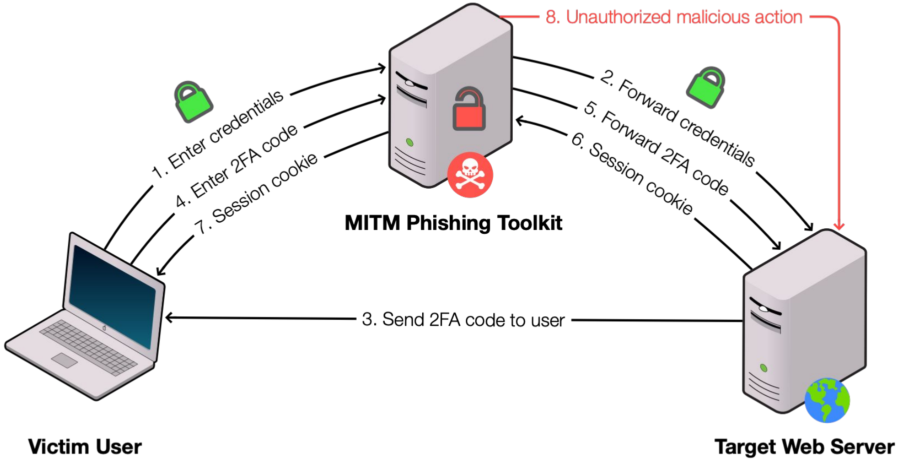
Attackers on **phishing.safebrowsingtest.com** may trick you into doing something dangerous like installing software or revealing your personal information (for example, passwords, phone numbers or credit cards). Learn more

Google Safe Browsing recently detected phishing on phishing.safebrowsingtest.com. Phishing sites pretend to be other websites to trick you.

You can report a detection problem or, if you understand the risks to your security, visit this unsafe site.

Man-in-the-Middle (MITM) Phishing Toolkits

- Malicious reverse proxy servers
  - Victims see live content from target website
  - Credentials stolen in transit
- Popular MITM phishing toolkits today:
  - Evilginx
  - Muraena
  - Modlishka



MITM Phishing Toolkit Demo

MITM Phishing Toolkit Threat Model

- Attackers control *all* application layer content
- Cloaking restricts access to phishing content
- Detection cannot rely on integrity of application layer content

MITM Phishing Toolkit Threat Model

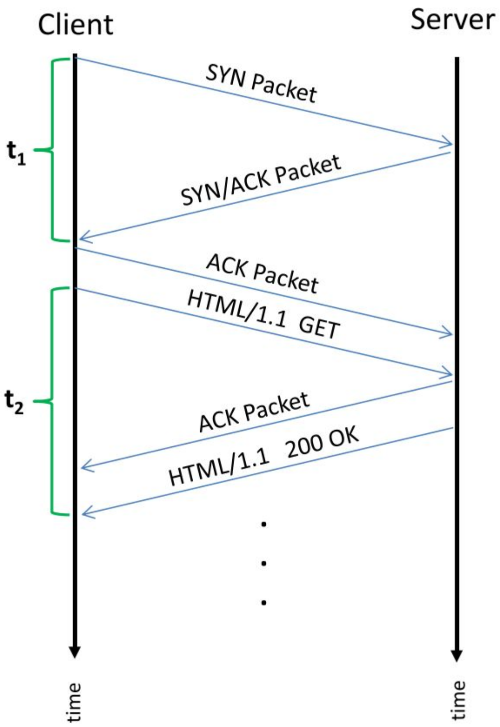
- Attackers control *all* application layer content

Fingerprint the server, not the content

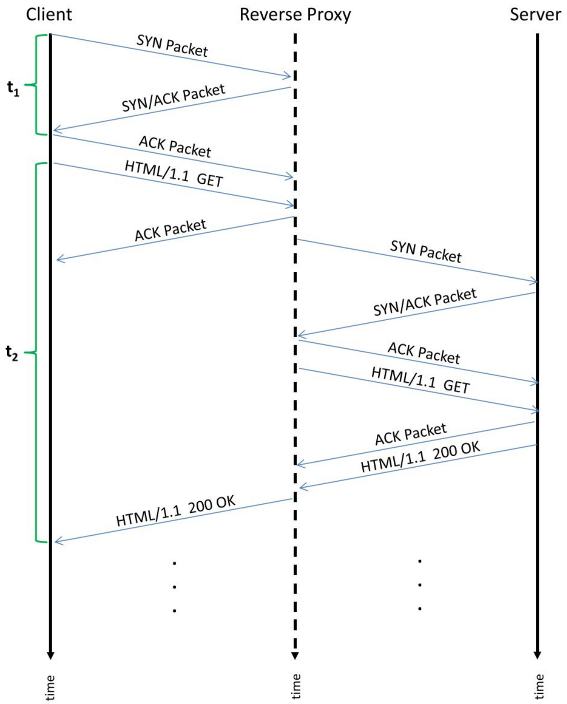
Network-Level Phishing Detection

- Network architecture can be leveraged to discover presence of toolkits
    - Network timing analysis
    - TLS fingerprinting
- Fingerprinting possible from both ends of the communication channel

Network Timing Analysis



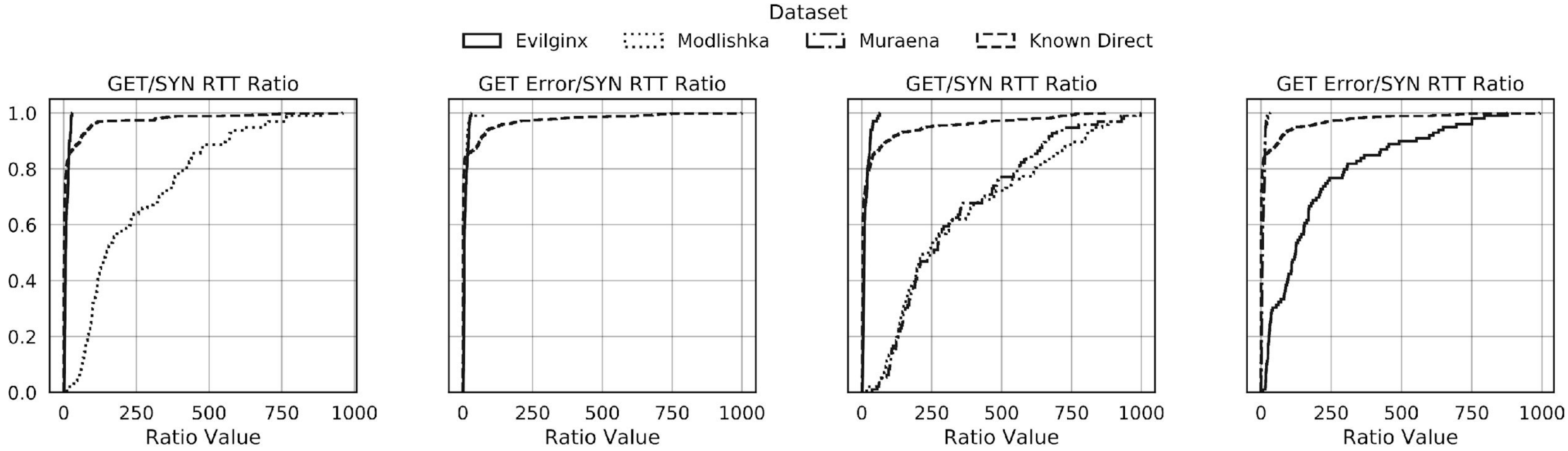
Direct Connection



Connection Through
Reverse Proxy Server

Figure Taken From: Daniel Alexander, “Inferring the Presence of Reverse Proxies Through Timing Analysis” (2015)

Network Timing Analysis

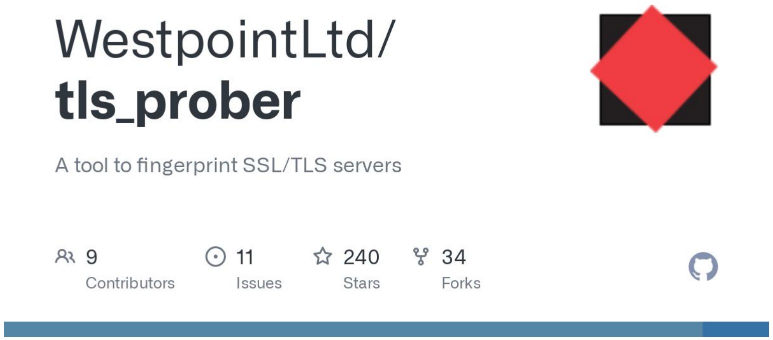


HTTP

HTTPS

TLS Fingerprinting

- MITM phishing toolkits utilize unusual TLS stacks
  - TLS versions supported
  - TLS libraries[1]



WestpointLtd/
tls_prober

A tool to fingerprint SSL/TLS servers

9 Contributors

11 Issues

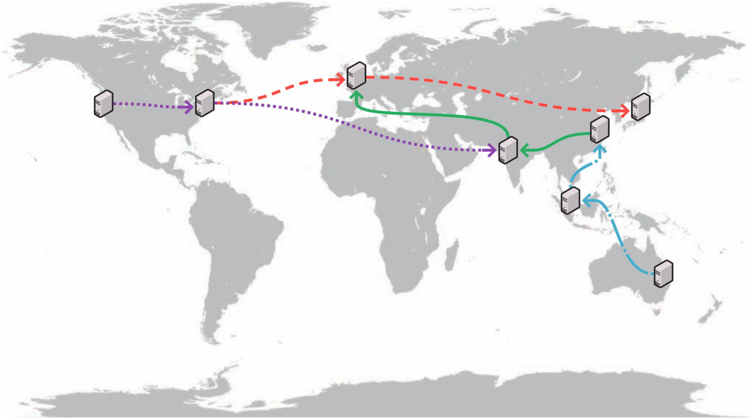
240 Stars

34 Forks

[1] https://github.com/WestpointLtd/tls_prober

MITM Phishing Toolkit Groundtruth

- We are the first to conduct a comprehensive study on MITM phishing toolkits
  - No groundtruth dataset on MITM phishing toolkit behavior
- Collected network-level data from 30 globally-distributed nodes
  - Recorded all permutations of client → MITM phishing toolkit → webserver
  - 146,160 data points in total

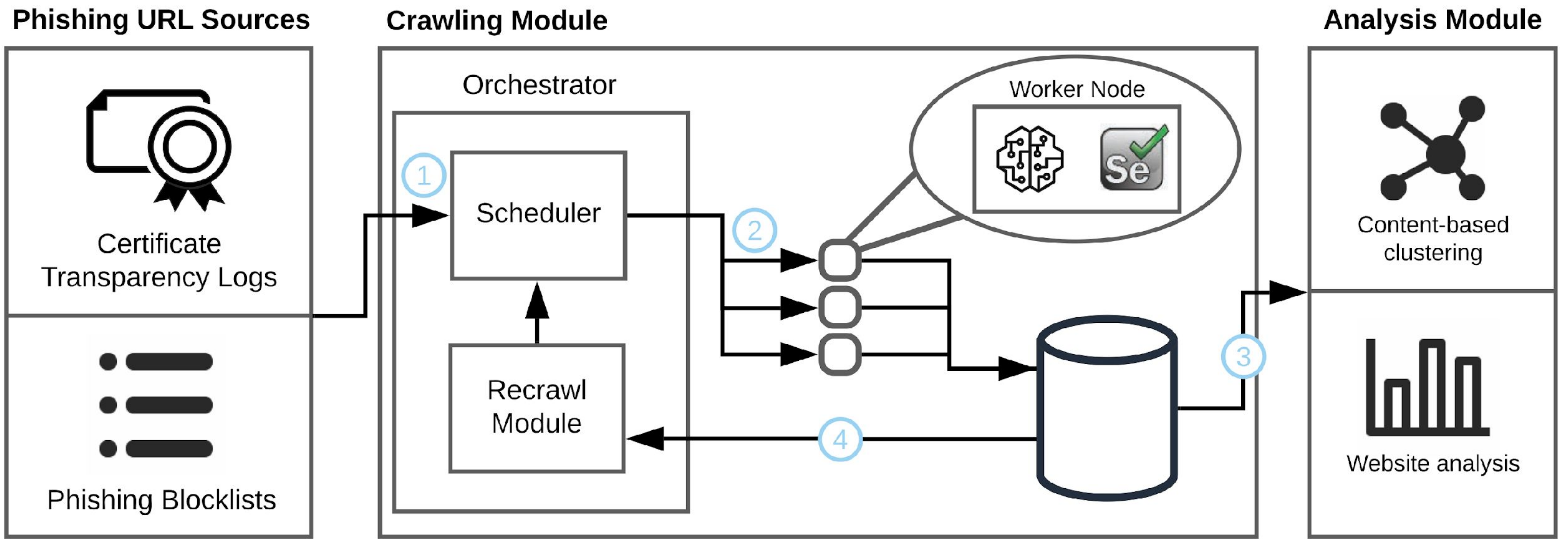


- Random forest classifier
  - Achieved 99.9% accuracy and five-fold cross validation score of 99.9%

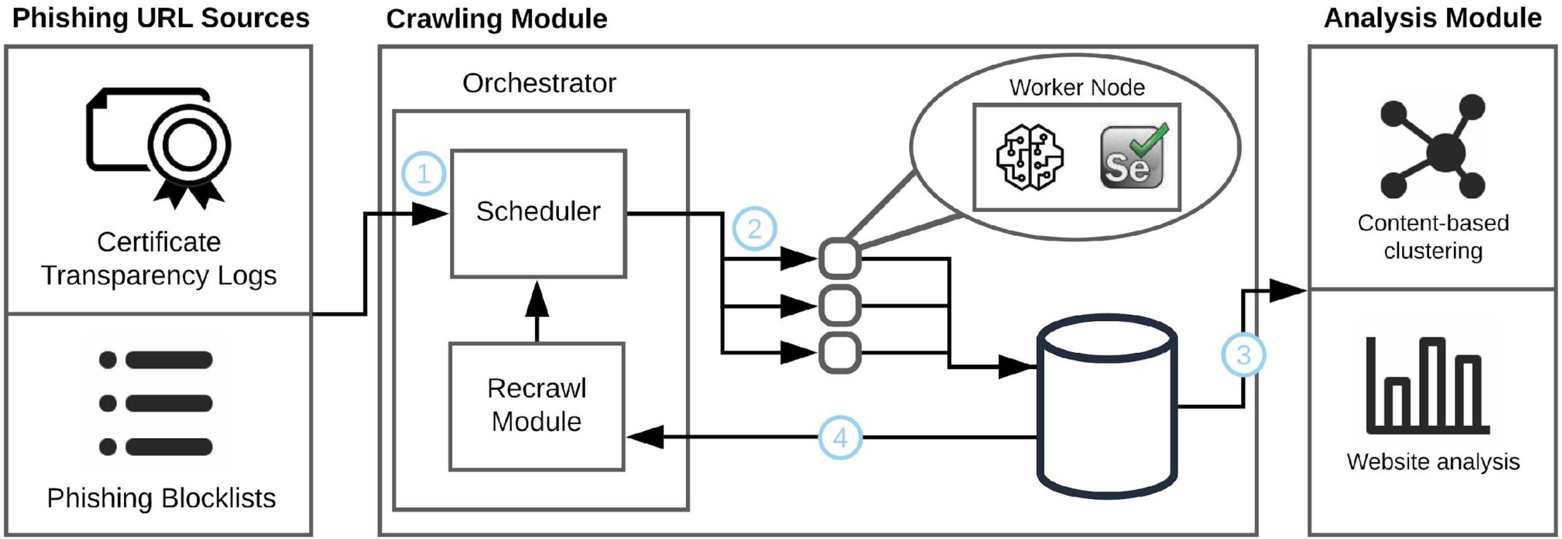
PHOCA: MITM Phishing Website Detector

- Framework to collect network-level data on, and detect MITM phishing websites
- Named after the Latin word for seal
    - Known to use vibrations in water to detect otherwise hidden prey

Phishing Website Crawling Infrastructure

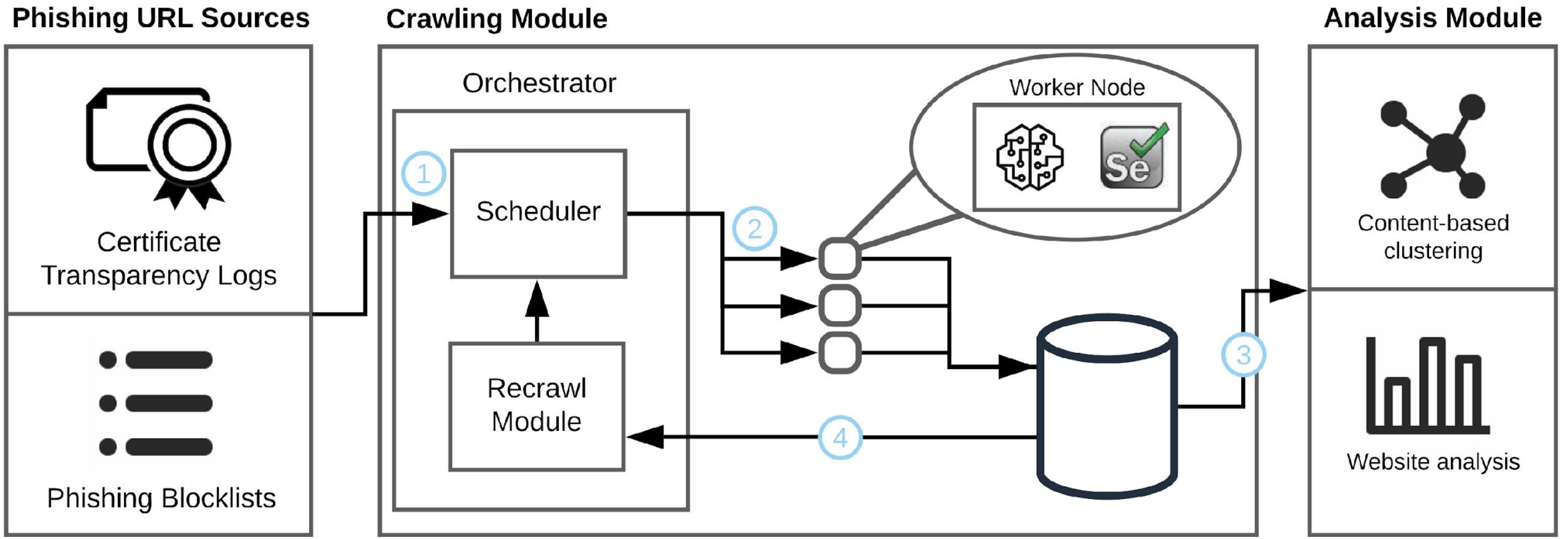


Phishing Website Crawling Infrastructure



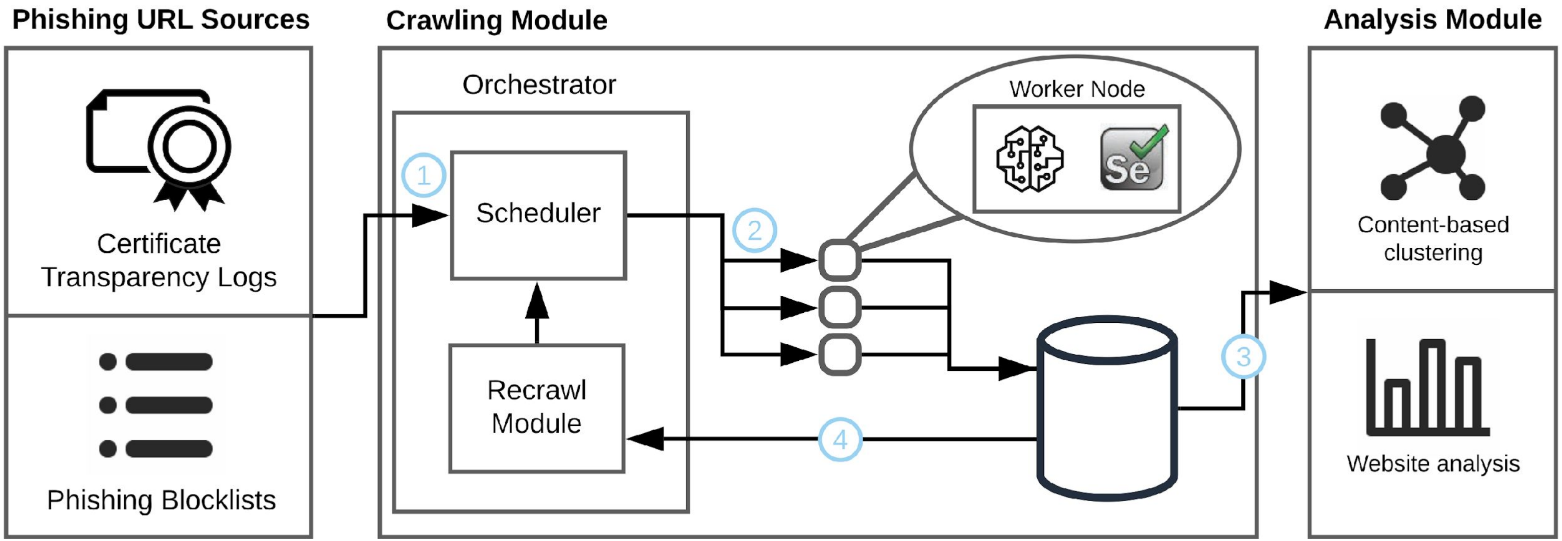
1. Candidate domains sourced from Certificate Transparency Logs and anti-phishing blocklists

Phishing Website Crawling Infrastructure



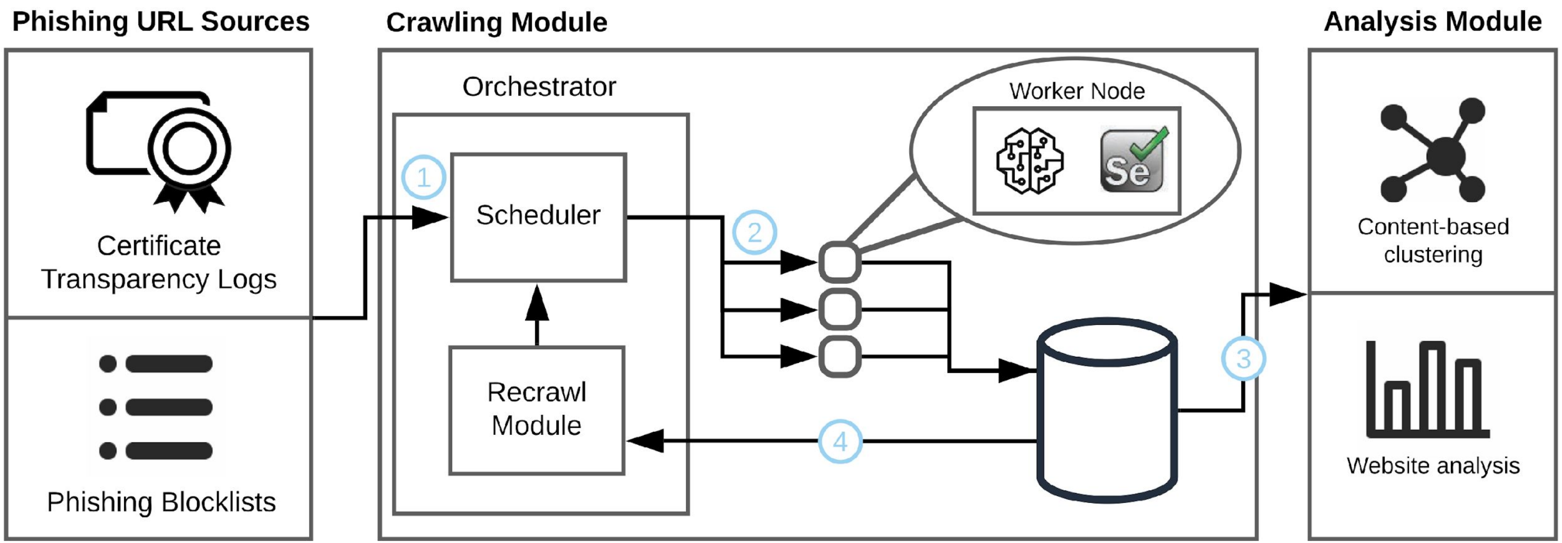
2. Scheduler module dispatches worker nodes to retrieve classification from PHOCA, and screenshot/HTML code using Selenium

Phishing Website Crawling Infrastructure



3. Collected data fed into analysis module for further processing

Phishing Website Crawling Infrastructure

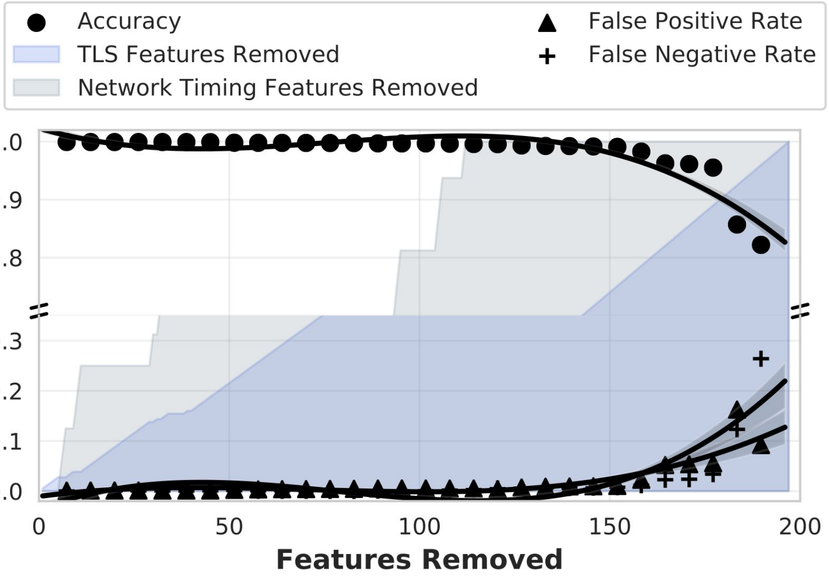


4. Recrawling module periodically revisits websites of interest

MITM Phishing Toolkit Classifier

- Trained random forest classifier on data from real websites and MITM phishing toolkits

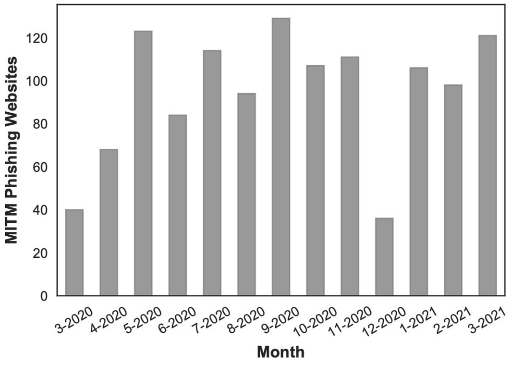
- Achieved 99.9% accuracy and five-fold cross validation score of 99.9%

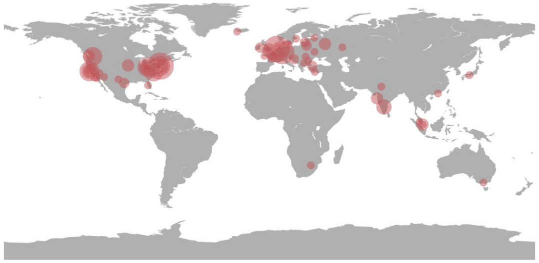


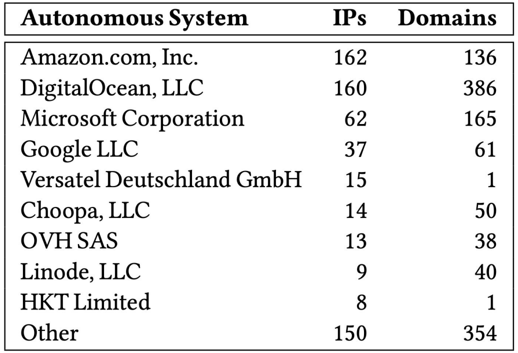
PHOCA Demo

MITM Phishing Toolkits on the Web

- Data collection period from March 25th, 2020 to March 25th, 2021
  - 841,711 web pages analyzed
  - 1,220 MITM phishing toolkits identified

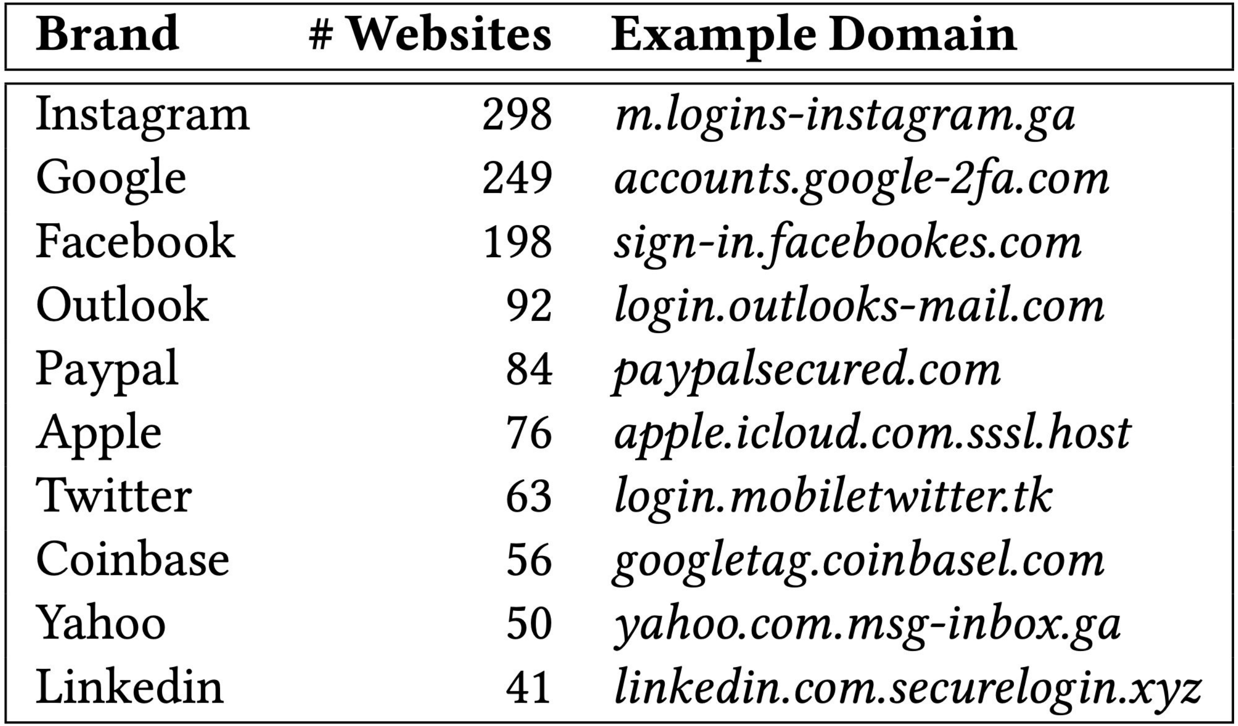






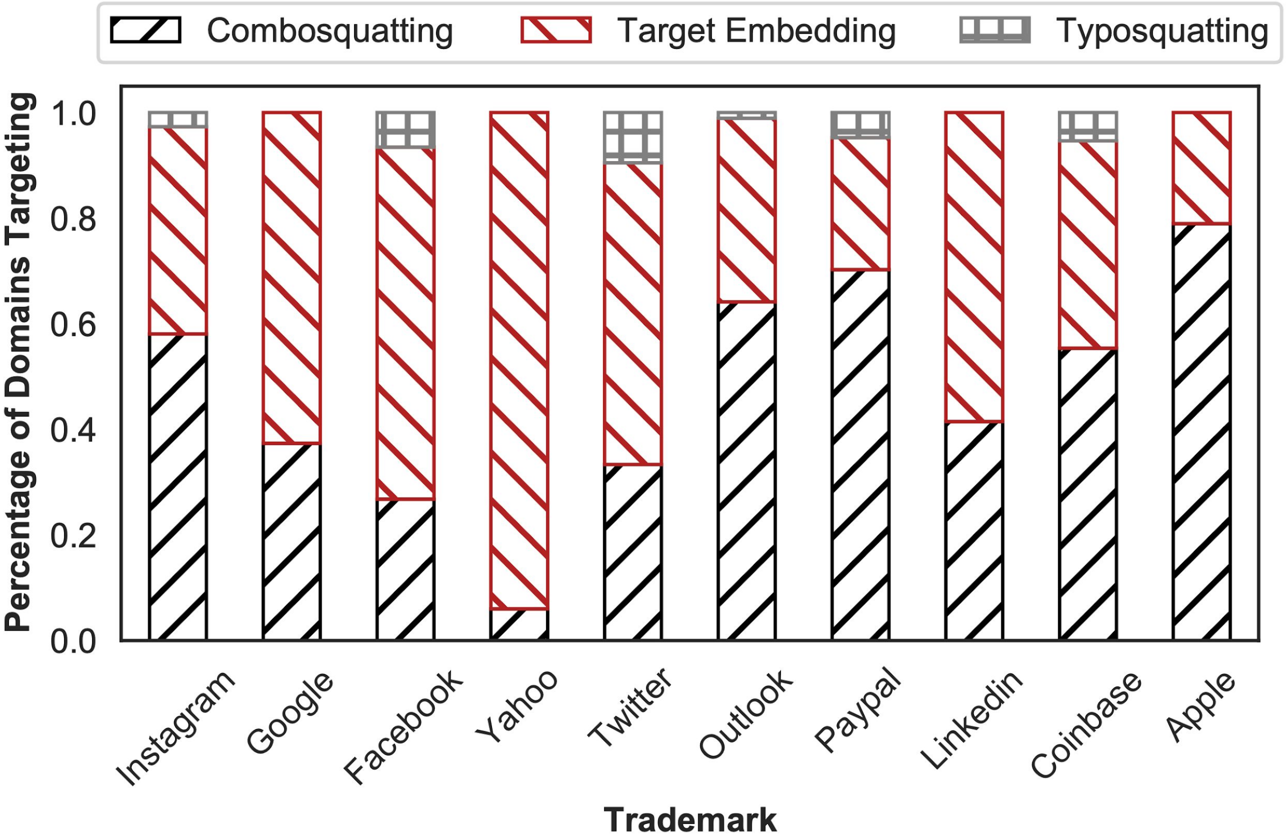
Autonomous System	IPs	Domains
Amazon.com, Inc.	162	136
DigitalOcean, LLC	160	386
Microsoft Corporation	62	165
Google LLC	37	61
Versatel Deutschland GmbH	15	1
Choopa, LLC	14	50
OVH SAS	13	38
Linode, LLC	9	40
HKT Limited	8	1
Other	150	354

MITM Phishing Website Targets

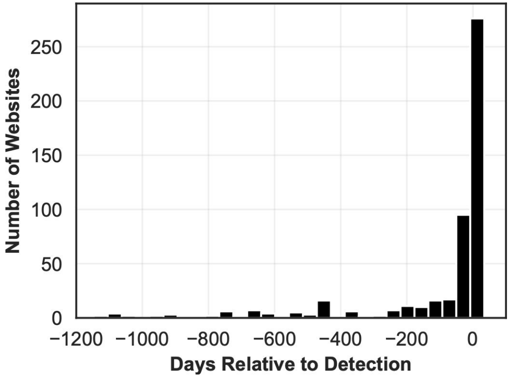


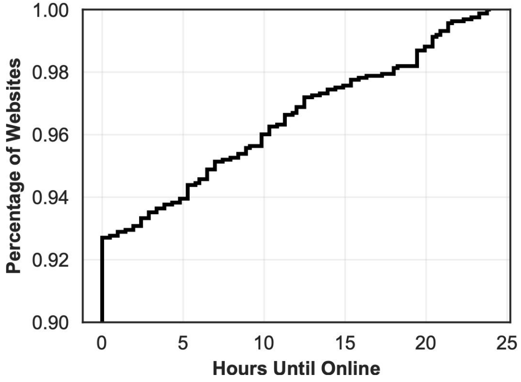
Brand	# Websites	Example Domain
Instagram	298	*m.logins-instagram.ga*
Google	249	*accounts.google-2fa.com*
Facebook	198	*sign-in.facebookes.com*
Outlook	92	*login.outlooks-mail.com*
Paypal	84	*paypalsecured.com*
Apple	76	*apple.icloud.com.sssl.host*
Twitter	63	*login.mobiletwitter.tk*
Coinbase	56	*googletag.coinbasel.com*
Yahoo	50	*yahoo.com.msg-inbox.ga*
Linkedin	41	*linkedin.com.securelogin.xyz*

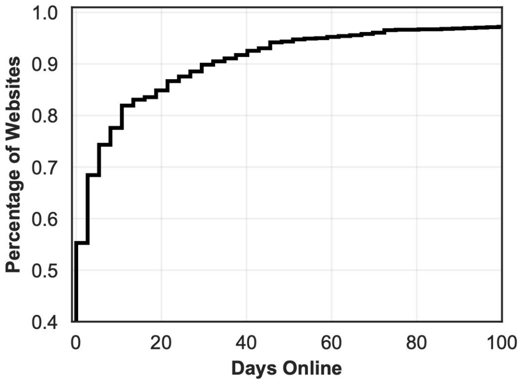
MITM Phishing Domain Types



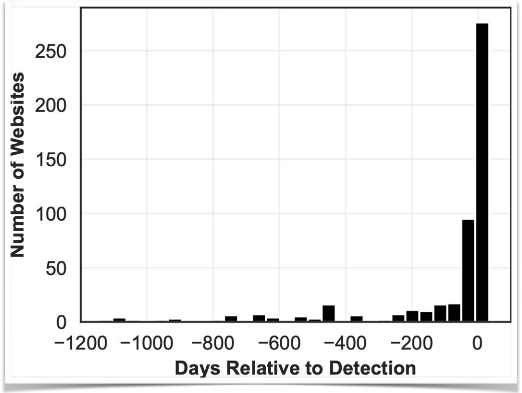
MITM Phishing Website Lifecycle

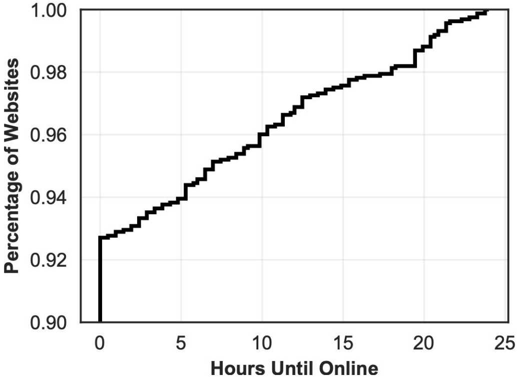


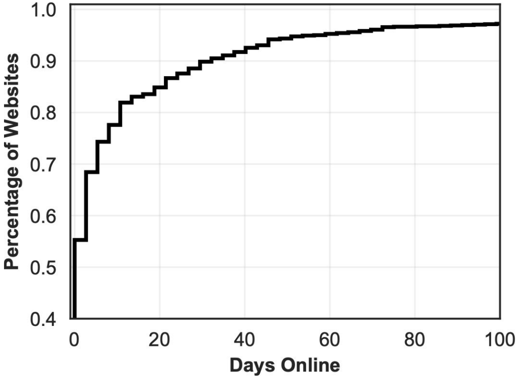




MITM Phishing Website Lifecycle

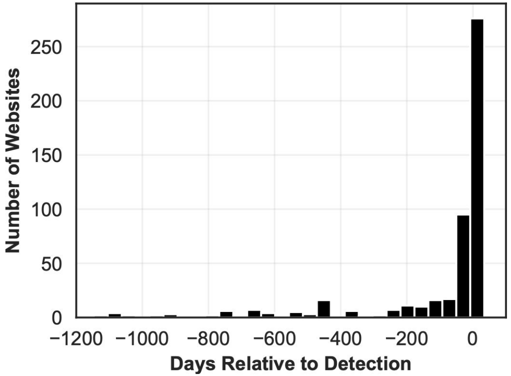


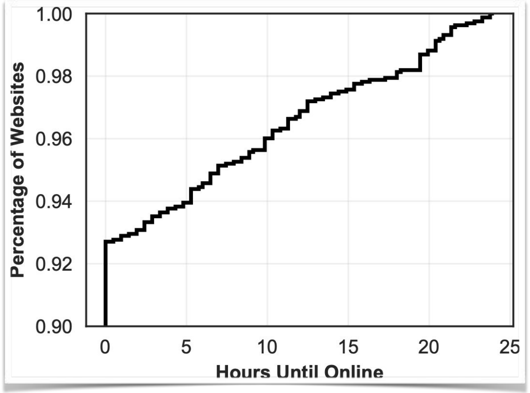


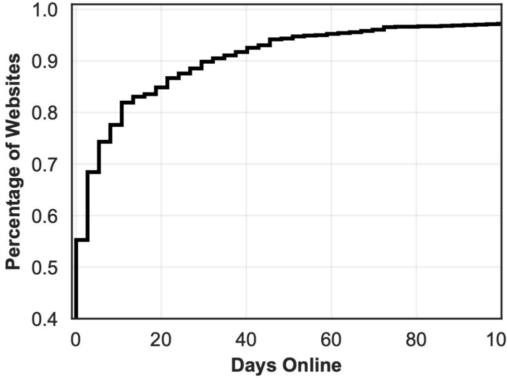


MITM phishing use freshly registered domains

MITM Phishing Website Lifecycle

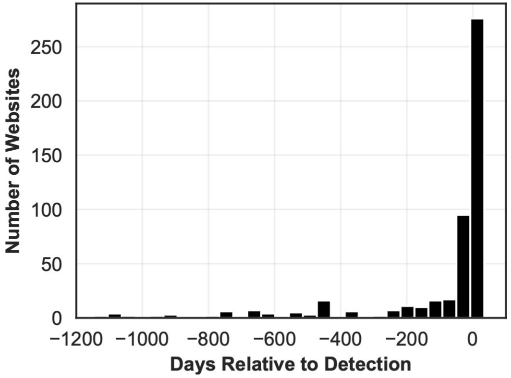


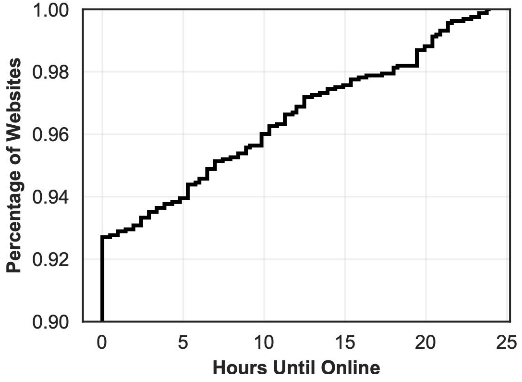


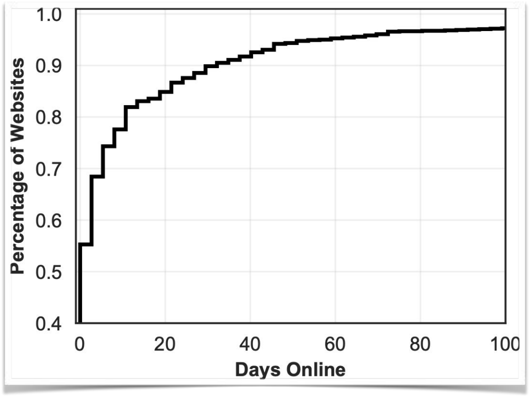


MITM phishing websites are weaponized immediately
after TLS certificate creation

MITM Phishing Website Lifecycle

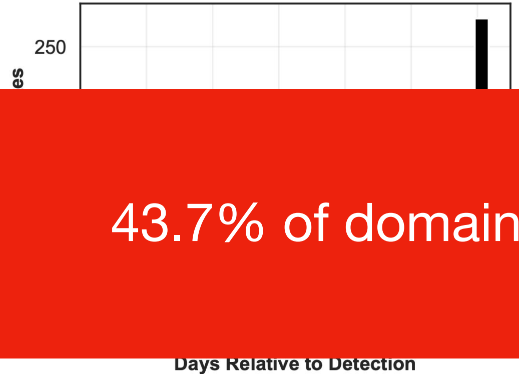


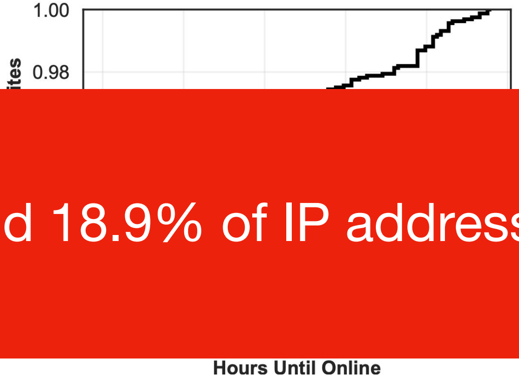


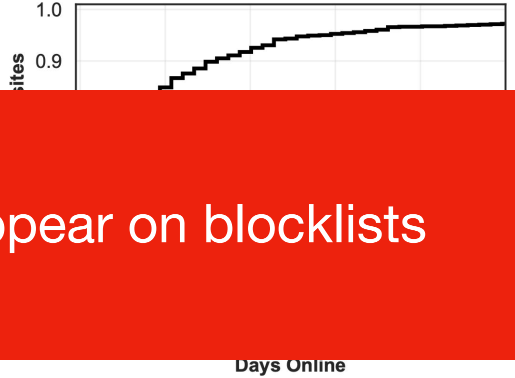


20% of MITM phishing websites remain active for longer than 10 days

MITM Phishing Website Lifecycle







43.7% of domains and 18.9% of IP addresses appear on blocklists

Case Study: Palo Alto Networks

- 56.7% of MITM phishing domains labeled as malicious by PAN in-line scanners
    - 15.1% received label at least one week after our initial discovery
- 6,403 customer requests directed towards 260 phishing websites over six months
    - Originating from 368 distinct firewall devices

Server-side TLS Fingerprinting

- MITM phishing toolkits do not utilize common web client TLS stacks
    - Forwarded HTTP User-Agent strings do not match TLS fingerprints
- JA3 TLS fingerprinting[1] utilized to identify unique TLS implementations
- Purchased 13,000 advertising impressions from a popular advertising service
    - Collected 163 unique TLS fingerprints from 4,311 distinct HTTP User-Agents
- TLS fingerprints of MITM phishing toolkits unique in this dataset

[1] https://github.com/salesforce/ja3

Countermeasures

- Users:
  - Analyze the primary domain of any suspicious URL encountered
  - Use U2F to secure online accounts
- Online Services/Anti-phishing Entities:
  - Look for discrepancies in client TLS fingerprints
  - Utilize network-level detection techniques when searching for phishing websites

Conclusion

- MITM phishing toolkits allow attackers to launch highly effective phishing attacks
- Unique architecture allows for fingerprinting at the network layer
- We found 1,220 MITM phishing toolkits operating in the wild, targeting real users
- Anti-phishing ecosystem does not effectively capture MITM phishing toolkits

Code and data: https://catching-transparent-phish.github.io

Thank you for your time! Any questions?