

COMAR: Classification of Compromised versus Maliciously Registered Domains

Sourena Maroofi, Maciej Korczyński, Cristian Hesselman,
Benoît Ampeau, Andrzej Duda, Thymen Wabeke

Univ. Grenoble Alpes, SIDN Labs, AFNIC

maciej.korczynski@univ-grenoble-alpes.fr

26 May 2021

ICANN DNS Symposium

Plan

1. Motivation
2. Overview of COMAR
3. Results
4. Conclusions

Overview: DNS reputation systems

DNS reputation systems can detect malicious domains using different techniques and at different phases:

- at the registration time (e.g., PREDATOR¹)
- domain activity phase (e.g., EXPOSURE²)

¹ Hao, Shuang, et al. "PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration." ACM CCS 2016

² Bilge, Leyla, et al. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." NDSS 2011

Overview: DNS reputation systems

DNS reputation systems can detect malicious domains using different techniques and at different phases:

- at the registration time (e.g., PREDATOR¹)
- domain activity phase (e.g., EXPOSURE²)
- They classify domains as either **malicious** or **benign**.

¹ Hao, Shuang, et al. "PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration." ACM CCS 2016

² Bilge, Leyla, et al. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." NDSS 2011

Overview: DNS reputation systems

DNS reputation systems can detect malicious domains using different techniques and at different phases:

- at the registration time (e.g., PREDATOR¹)
- domain activity phase (e.g., EXPOSURE²)
- They classify domains as either **malicious** or **benign**.
- They do not consider **compromised** domains.

¹ Hao, Shuang, et al. "PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration." ACM CCS 2016

² Bilge, Leyla, et al. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." NDSS 2011

Motivation

Why we need to consider **compromised** domains (hacked websites)?

Compromised domains have also legitimate traffic we may not want to block

1. The mitigation action for compromised domains (websites) is different from malicious domains

- Should we block/hold/take down the domain?
- Should we notify the hosting provider?

Motivation

<https://user-paypal.oz4.top/LkQwxCf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>



Log In

[Forgot your email or password?](#)

Sign Up

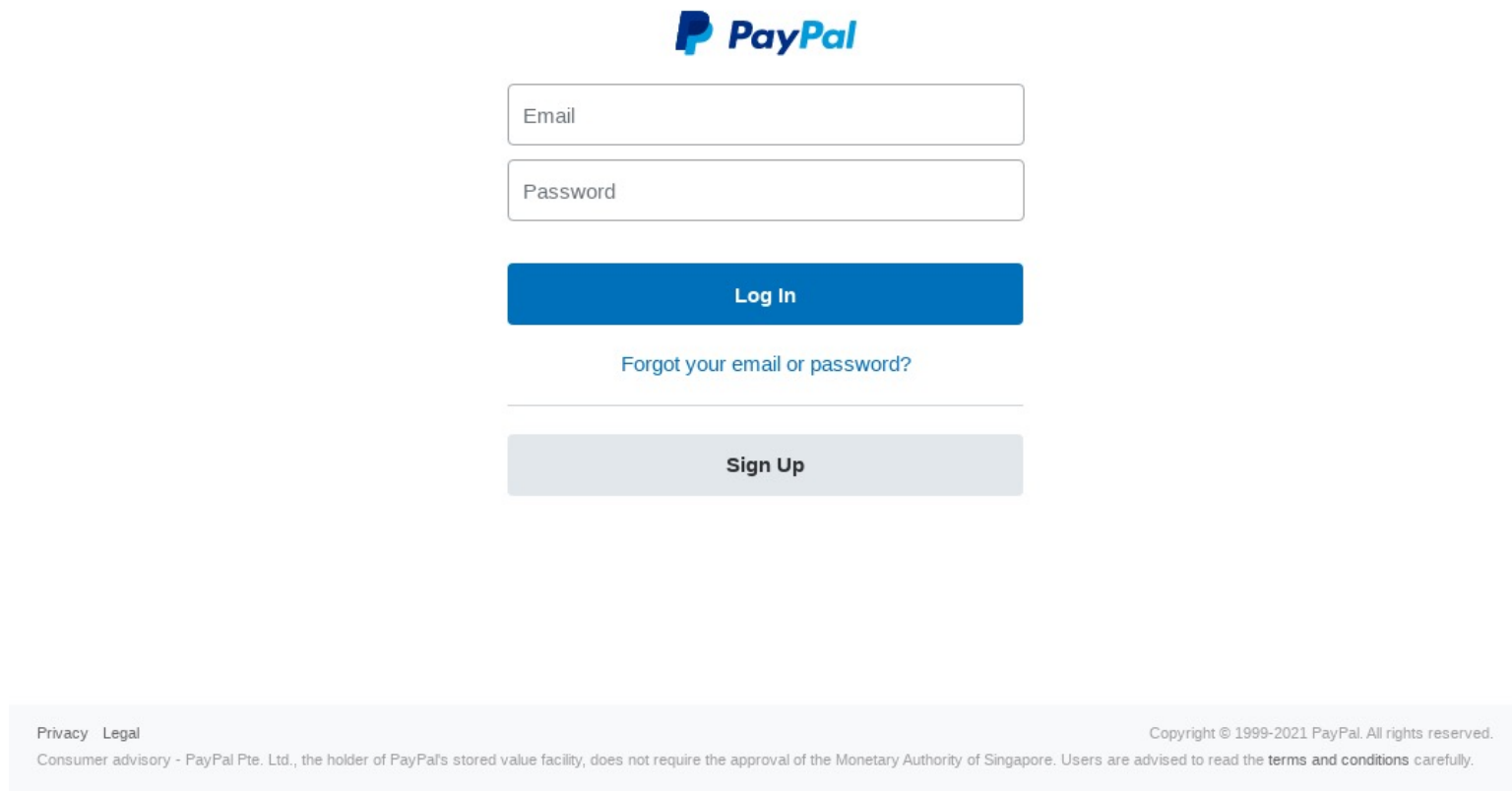
[Privacy](#) [Legal](#)

Copyright © 1999-2021 PayPal. All rights reserved.

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the [terms and conditions](#) carefully.

Motivation

<https://user-paypal.oz4.top/LkQwx Cf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>



The image shows a PayPal login page. At the top center is the PayPal logo. Below it are two input fields: one for 'Email' and one for 'Password'. Under the password field is a blue 'Log In' button. Below the button is a link that says 'Forgot your email or password?'. Below this is a horizontal line, and then a grey 'Sign Up' button. At the bottom of the page, there is a footer section with links for 'Privacy' and 'Legal' on the left, and copyright information on the right: 'Copyright © 1999-2021 PayPal. All rights reserved.' Below the copyright is a line of small text: 'Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the terms and conditions carefully.'

PayPal

Email

Password

Log In

[Forgot your email or password?](#)

Sign Up

Privacy Legal

Copyright © 1999-2021 PayPal. All rights reserved.

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the terms and conditions carefully.

What can be done to mitigate this abuse?

Motivation

<https://user-paypal.oz4.top/LkQwx Cf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>

<http://oz4.top>

Forbidden

You dont have permission to access / on this server.



Log In

[Forgot your email or password?](#)

Sign Up

Submission Date: 2021-01-15 18:00:05

WHOIS:

Updated Date: 2021-01-13T15:32:36Z

Creation Date: 2021-01-13T15:26:54Z

[Privacy](#) [Legal](#)

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the [terms and conditions](#) carefully.

Copyright © 1999-2021 PayPal. All rights reserved.

Motivation

<https://user-paypal.oz4.top/LkQwxCf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>

<http://oz4.top>

Forbidden

You dont have permission to access / on this server.



Log In

[Forgot your email or password?](#)

Sign Up

Submission Date: 2021-01-15 18:00:05

WHOIS:

Updated Date: 2021-01-13T15:32:36Z

Creation Date: 2021-01-13T15:26:54Z

[Privacy](#) [Legal](#)

Copyright © 1999-2021 PayPal. All rights reserved.

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the [terms and conditions](#) carefully.

Technical abuse (maliciously registered domain name) and **website content abuse** (illegal/abusive content)

Motivation

<https://user-paypal.oz4.top/LkQwxCf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>

<http://oz4.top>

Forbidden

You dont have permission to access / on this server.



Log In

[Forgot your email or password?](#)

Sign Up

Submission Date: 2021-01-15 18:00:05

WHOIS:

Updated Date: 2021-01-13T15:32:36Z

Creation Date: 2021-01-13T15:26:54Z

[Privacy](#) [Legal](#)

Copyright © 1999-2021 PayPal. All rights reserved.

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the [terms and conditions](#) carefully.

Technical abuse (maliciously registered domain name) and **website content abuse** (illegal/abusive content)

Action: Block/hold/take down domain name ...and clean the hosting content


Motivation

<https://user-paypal.oz4.top/LkQwxCf2/rfFDbZaPR9Ti/loiujYnPGh/ANWfgiB2vk8b/1>

<http://oz4.top>

Forbidden

You dont have permission to access / on this server.



[Forgot your email or password?](#)

Submission Date: 2021-01-15 18:00:05

WHOIS:

Updated Date: 2021-01-13T15:32:36Z

Creation Date: 2021-01-13T15:26:54Z

[Privacy](#) [Legal](#)

Consumer advisory - PayPal Pte. Ltd., the holder of PayPal's stored value facility, does not require the approval of the Monetary Authority of Singapore. Users are advised to read the [terms and conditions](#) carefully.

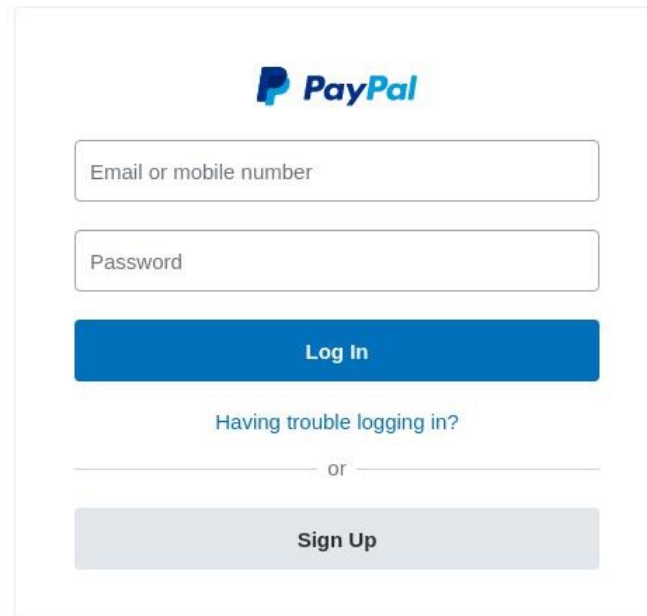
Copyright © 1999-2021 PayPal. All rights reserved.

Technical abuse (maliciously registered domain name) and **website content abuse** (illegal/abusive content)


Action: Block/hold/take down domain name ...and clean the hosting content
Intermediary: DNS service operator (registrar, registry) and hosting provider

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



The image shows a PayPal login interface. At the top is the PayPal logo. Below it are two input fields: "Email or mobile number" and "Password". A blue "Log In" button is positioned below the password field. Underneath the button is a link that says "Having trouble logging in?". Below this is a horizontal line with the word "or" in the center. At the bottom is a grey "Sign Up" button.



Email or mobile number

Password

Log In

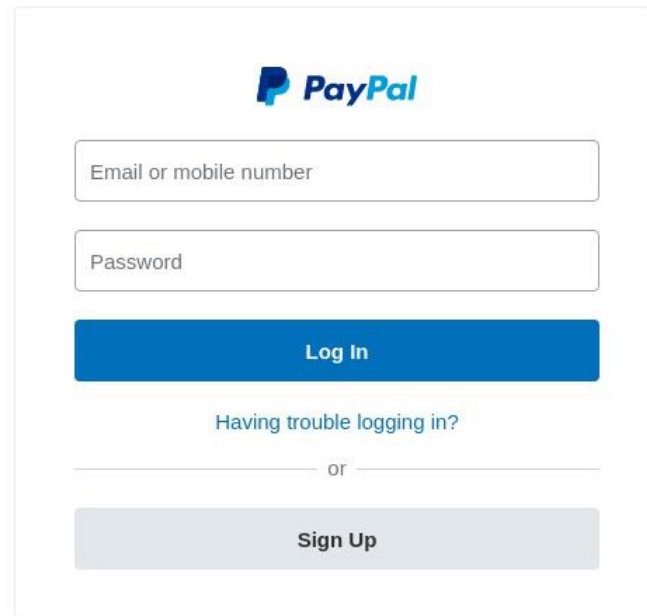
[Having trouble logging in?](#)

or

Sign Up

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



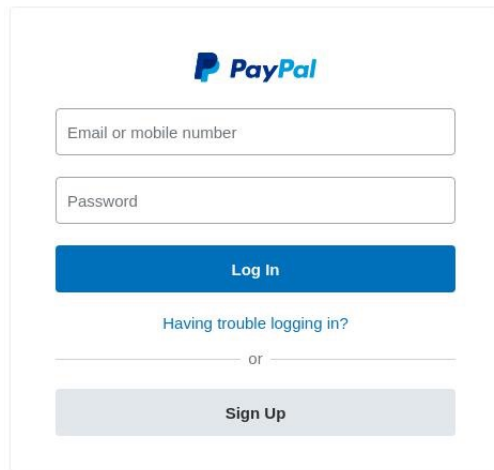
The image shows a PayPal login form. At the top is the PayPal logo. Below it are two input fields: "Email or mobile number" and "Password". A blue "Log In" button is positioned below the password field. Below the button is a link that says "Having trouble logging in?". A horizontal line with the word "or" in the center separates this from a grey "Sign Up" button at the bottom.

[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

What can be done to mitigate this abuse?

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



The image shows a PayPal login form. At the top is the PayPal logo. Below it are two input fields: "Email or mobile number" and "Password". A blue "Log In" button is positioned below the password field. Below the button is a link that says "Having trouble logging in?". A horizontal line with the word "OR" in the center separates this from a grey "Sign Up" button at the bottom.

[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

<https://nutribiocorp.com>



The screenshot shows the homepage of Nutri Bio Corp Pvt Ltd. The header includes the company logo and navigation links: Home, About Us, Company Profile, Our Products (with a dropdown arrow), and Contact Us. The main banner features a photograph of several chickens in a green field. Below the banner, the "About Us" section contains text about the company's focus on animal health and nutrition, its location in Jind, Haryana, India, and the services it provides. To the right of the text is an image of a white plastic bottle of "Gautherb" product. At the bottom of the "About Us" section, there is more text about health care solutions.

About Us

Nutri Bio Corp Pvt Ltd is one of the leading Animal Health and Nutrition Company headquartered in Jind, Haryana, India. The company is involved in providing solutions for different types of problems associated with poultry and animal industries.

The services include Healthcare solutions, Farm management, Disease management, Nutritional support and Technical support.

Health care solutions include a range of innovative products which are designed and developed based on the clear understanding of the customer needs. These includes Liver tonic

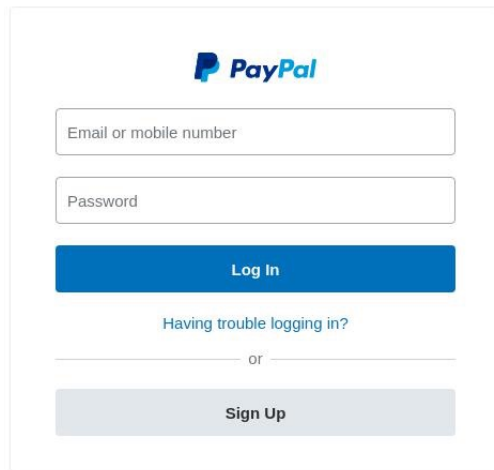
WHOIS:

Creation Date: 2014-03-04T00:00:00Z

Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



The image shows a PayPal login form. At the top is the PayPal logo. Below it are two input fields: "Email or mobile number" and "Password". A blue "Log In" button is positioned below the password field. Below the button is a link that says "Having trouble logging in?". At the bottom, there is a horizontal line with "OR" in the center, and a grey "Sign Up" button below it.

<https://nutribiocorp.com>



WHOIS:

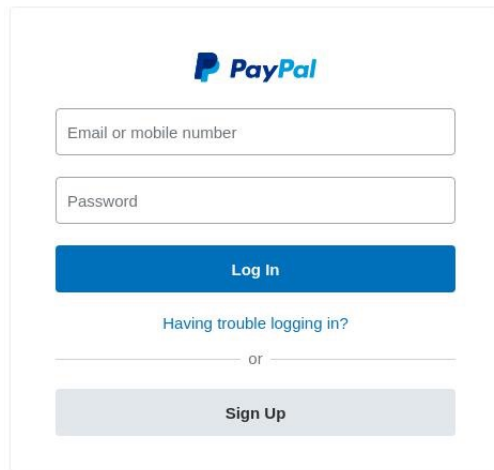
Creation Date: 2014-03-04T00:00:00Z

Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

Website content abuse (illegal/abusive content), **benign** domain but **compromised** website...

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



<https://nutribiocorp.com>



WHOIS:

Creation Date: 2014-03-04T00:00:00Z

Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

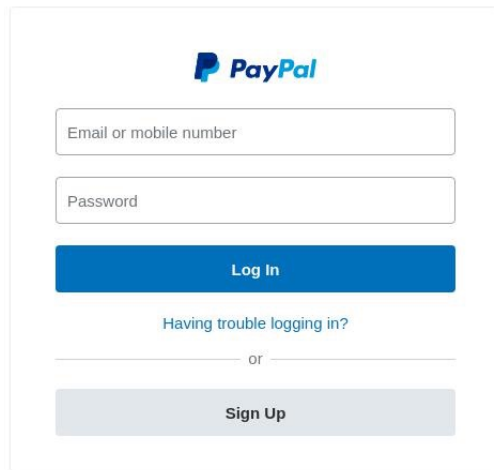
[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

Website content abuse (illegal/abusive content), **benign** domain but **compromised** website...

Action: Block/hold/take down domain name and clean the hosting content

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



<https://nutribiocorp.com>



WHOIS:

Creation Date: 2014-03-04T00:00:00Z

Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

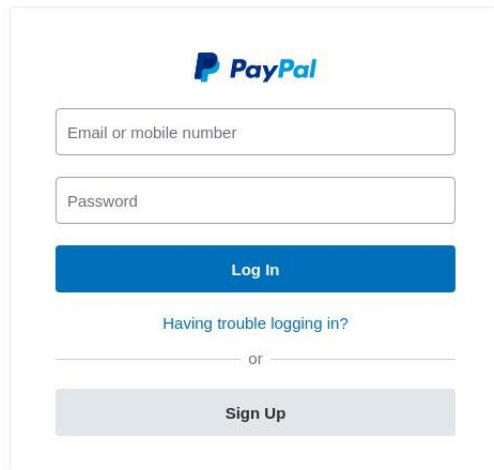
[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

Website content abuse (illegal/abusive content), **benign** domain but **compromised** website...

Action: ~~Block/hold/take down domain name and clean the hosting content~~

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



<https://nutribiocorp.com>



WHOIS:

Creation Date: 2014-03-04T00:00:00Z

Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

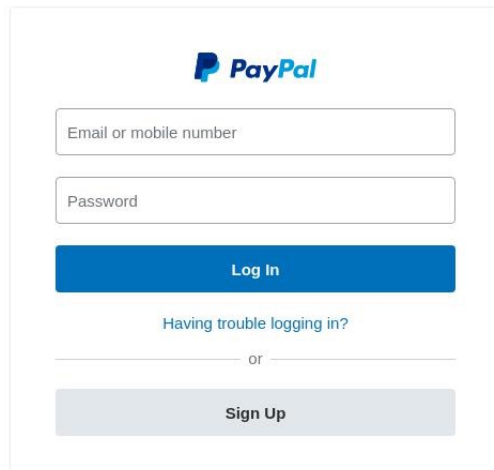
Website content abuse (illegal/abusive content), **benign** domain but **compromised** website...

Action: ~~Block/hold/take down domain name and~~ clean the hosting content

Intermediary: DNS service operator and hosting provider

Motivation

<https://nutribiocorp.com/wp-includes/paypal/paypal/login/update.account-PayPal/account-has-been-limited/logins.html>



<https://nutribiocorp.com>



WHOIS:

Creation Date: 2014-03-04T00:00:00Z
Registrar Registration Expiration Date:
2022-03-04T00:00:00Z

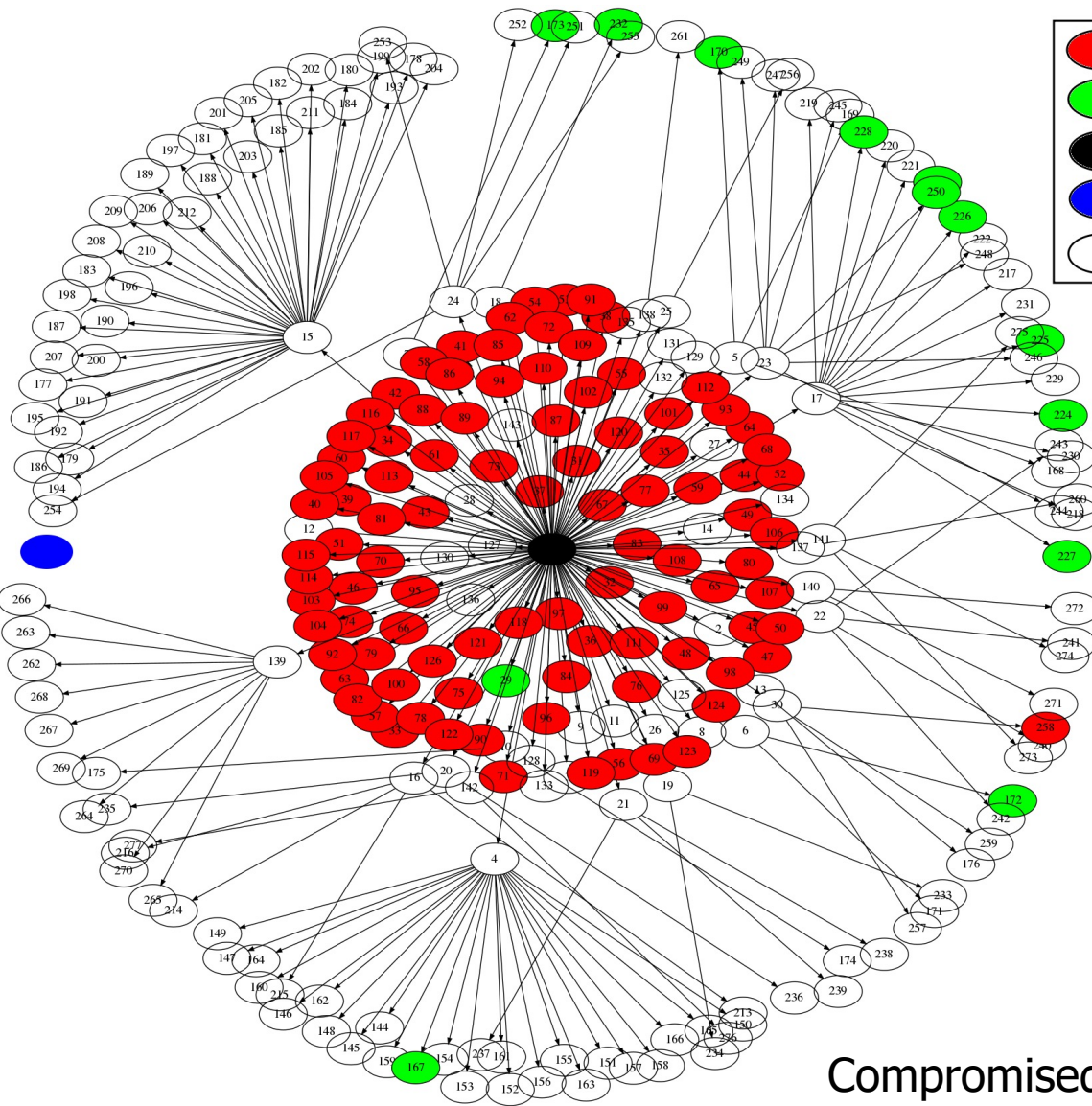
[Contact Us](#) [Privacy](#) [Legal](#) [Worldwide](#)

Website content abuse (illegal/abusive content), **benign** domain but **compromised** website...

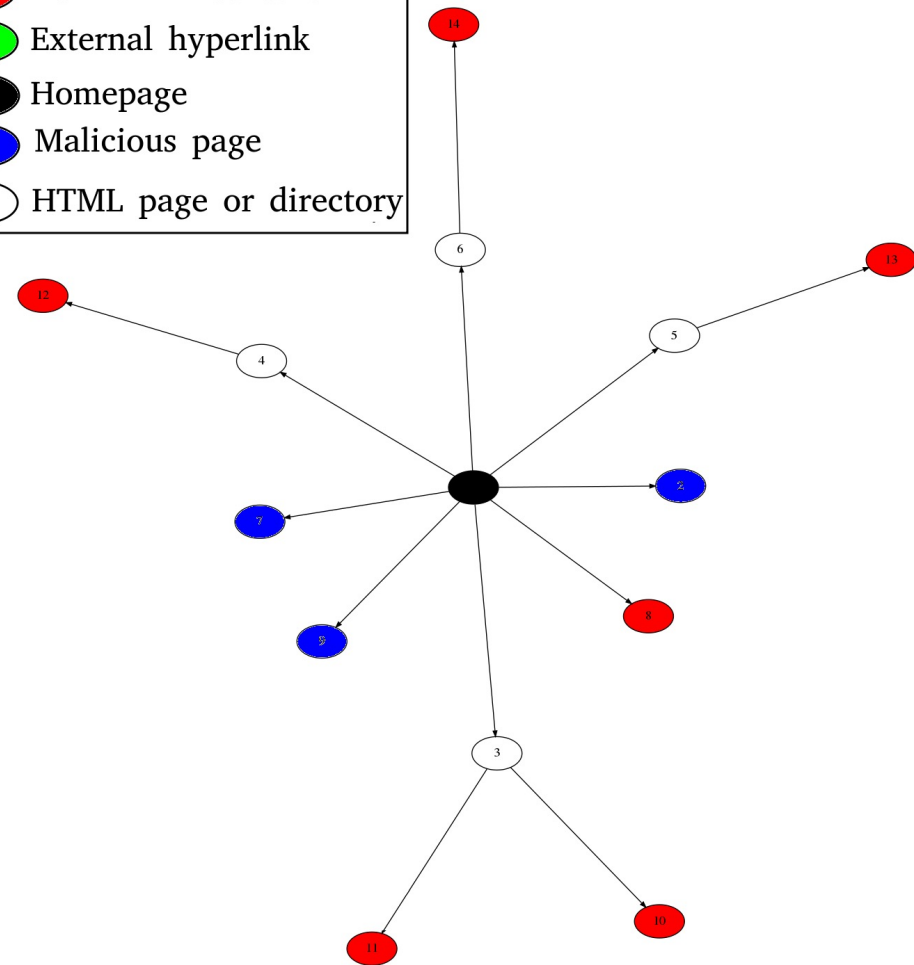
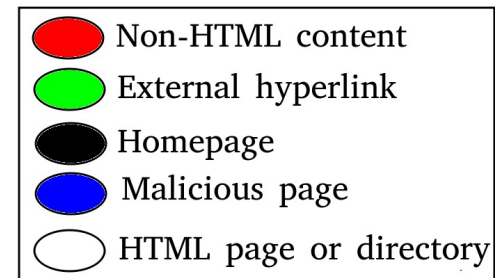
Action: ~~Block/hold/take down domain name and clean the hosting content~~

Intermediary: ~~DNS service operator and hosting provider ...and the website owner/administrator~~

Website structure of compromised vs. malicious



Compromised domain



Malicious domain

Motivation

Why we need to consider **compromised** domains?

Compromised domains have also legitimate traffic we may not want to block.

1. The mitigation action for compromised domains is different from malicious domains

- Should we block/hold/take down the domain?
- Should we notify the hosting provider?

Motivation

Why we need to consider **compromised** domains?

Compromised domains have also legitimate traffic we may not want to block.

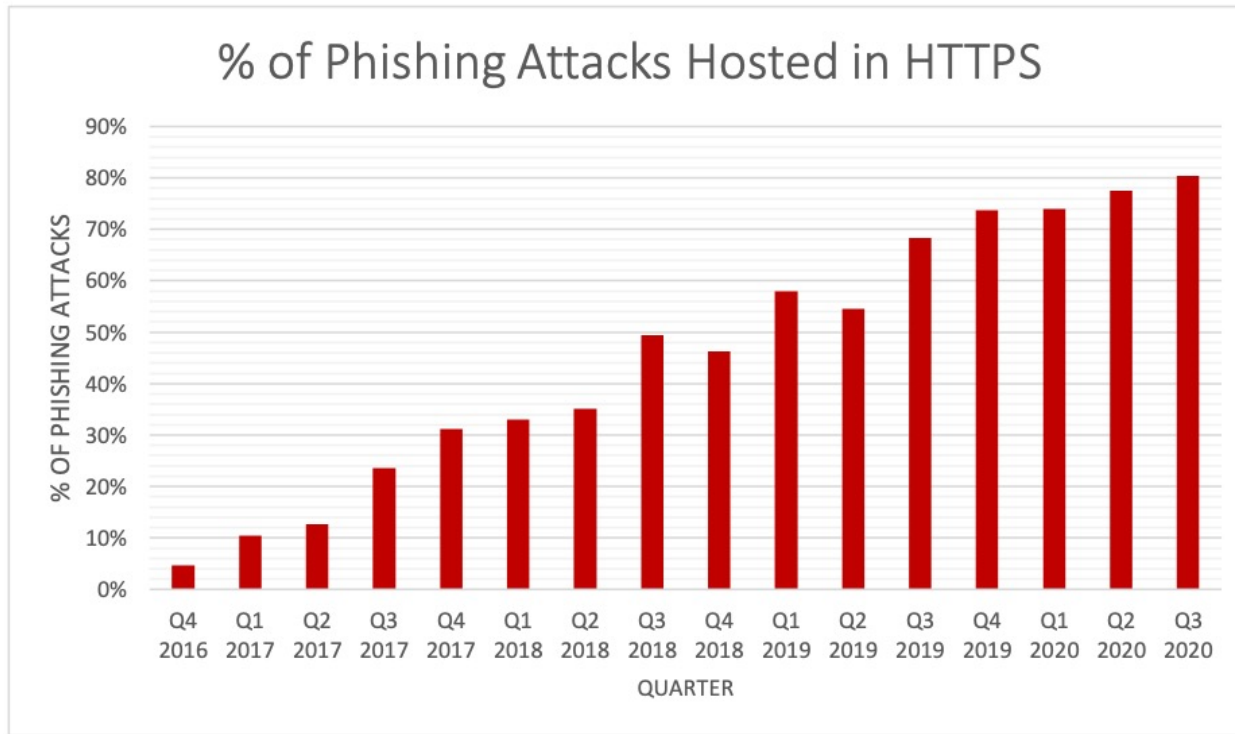
1. The mitigation action for compromised domains is different from malicious domains

- Should we block/hold/take down the domain?
- Should we notify the hosting provider?

2. Creating more effective domain blacklist feed → better insights into attackers' behavior

- Do current URL blacklists give us the correct insights?

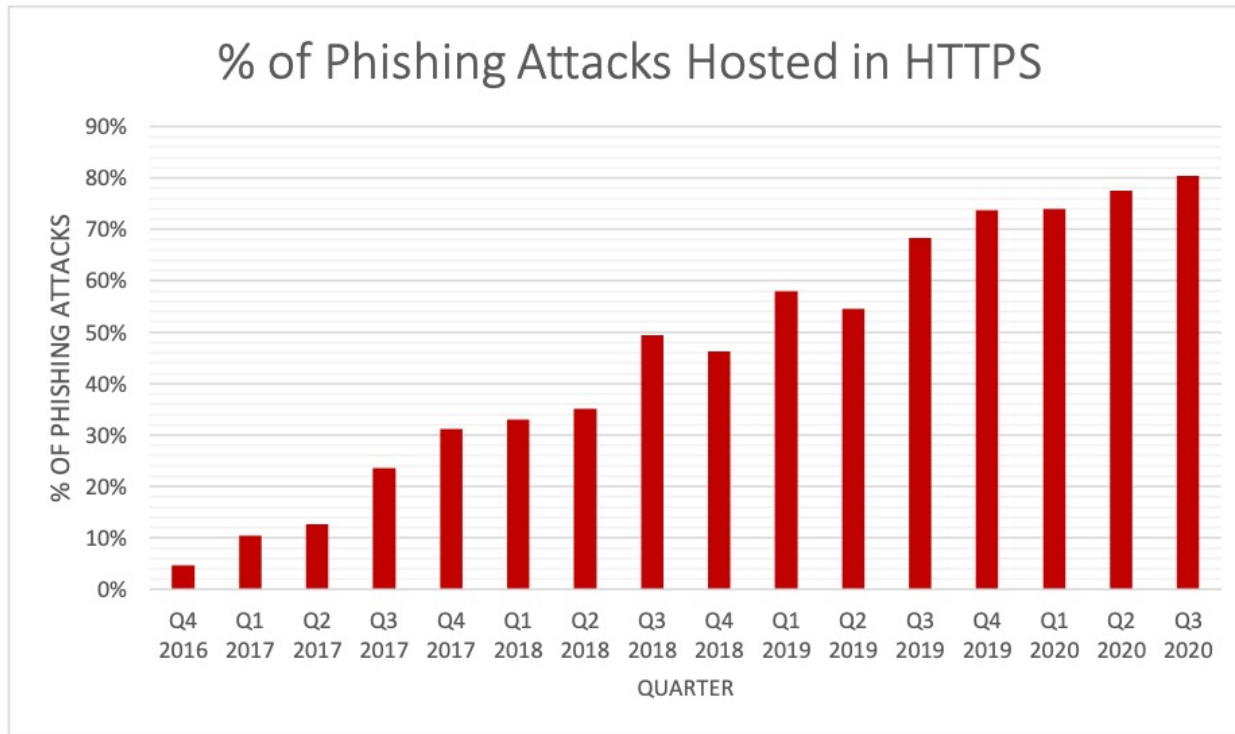
Motivation



“Eighty percent of phishing sites have SSL encryption enabled to fool victims.” [1]

“ (...) 8.6 percent were OV (Organization Validation) certs, and just 0.1% were Extended Validation (EV)” [1]

Motivation

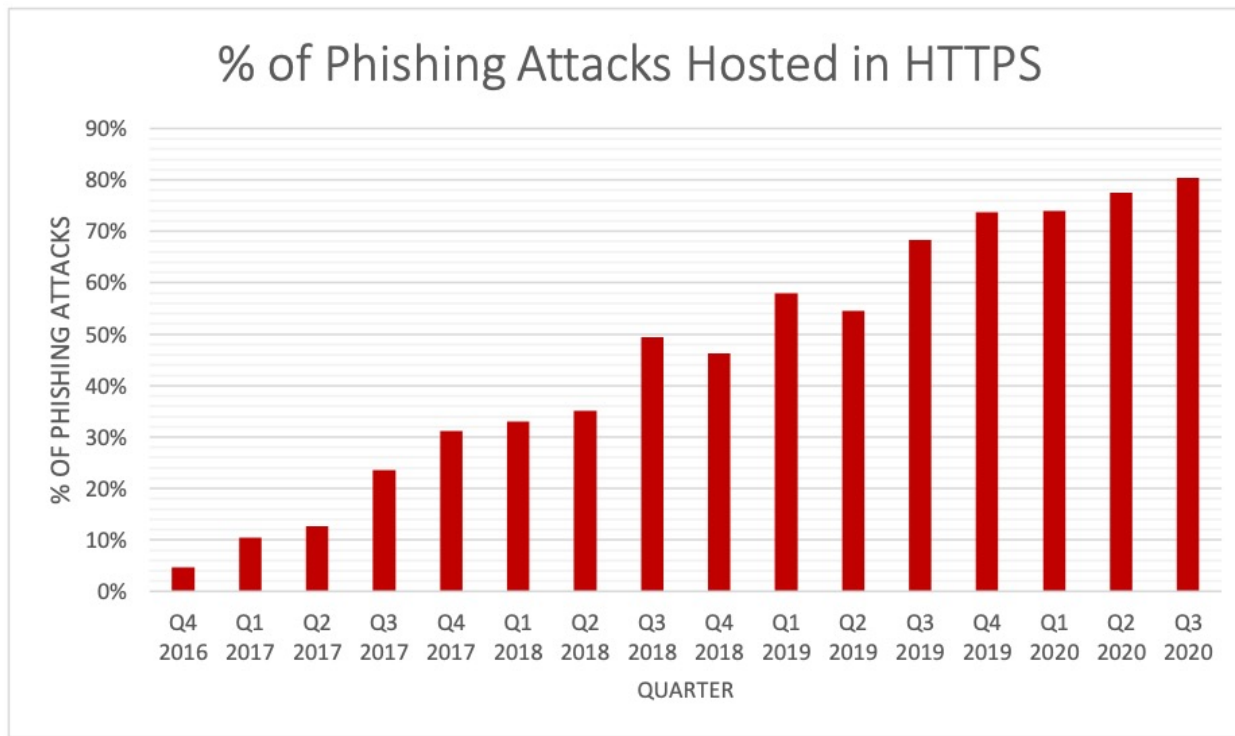


“Eighty percent of phishing sites have SSL encryption enabled to fool victims.” [1]

“ (...) 8.6 percent were OV (Organization Validation) certs, and just 0.1% were Extended Validation (EV)” [1]

How to interpret the results?

Motivation



“Eighty percent of phishing sites have SSL encryption enabled to fool victims.” [1]

“ (...) 8.6 percent were OV (Organization Validation) certs, and just 0.1% were Extended Validation (EV)” [1]

How to interpret the results?

Distinguishing between compromised and malicious domains can give better insights into attackers' behavior

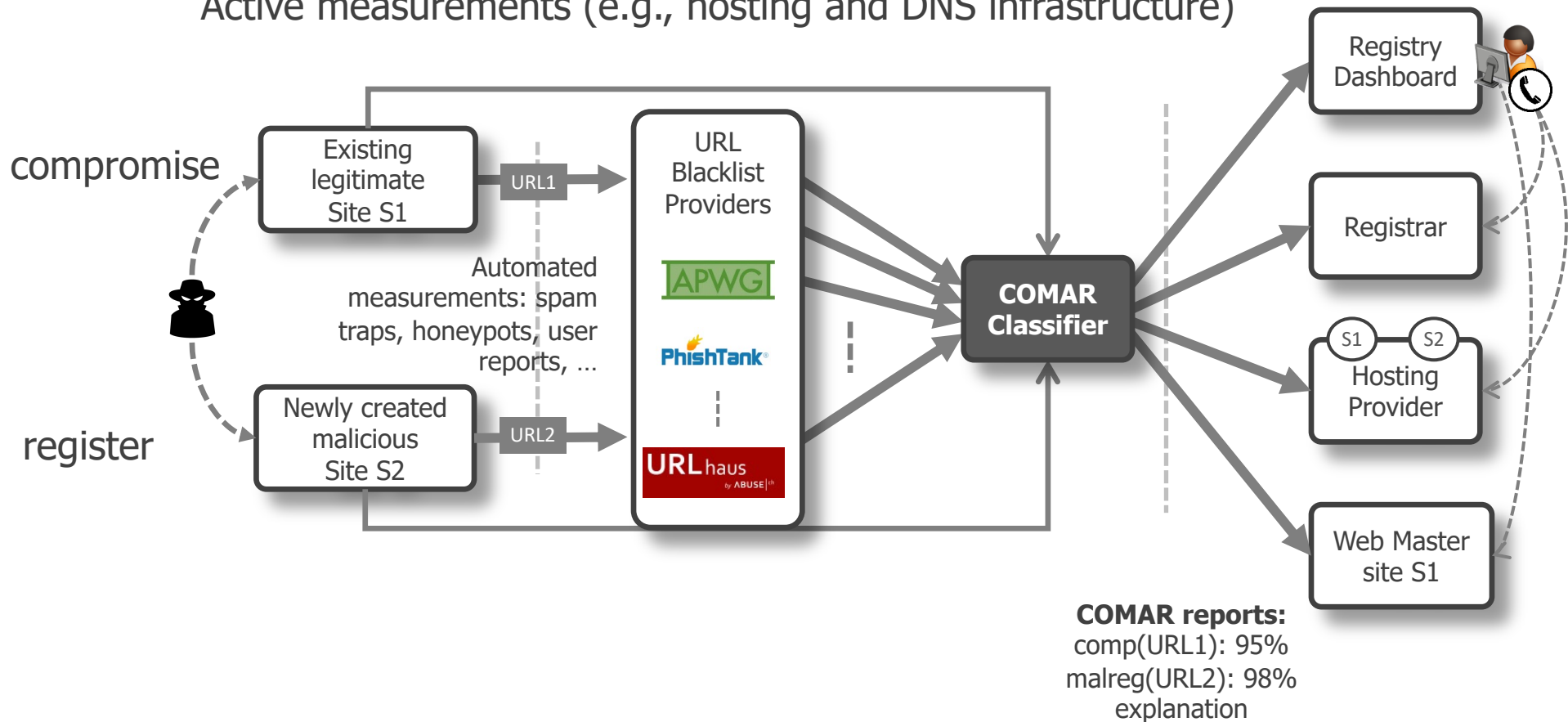
[1] Phishing Activity Trends Report APWG – 3rd Quarter 2020 – Published: 24, November 2020

Contributions

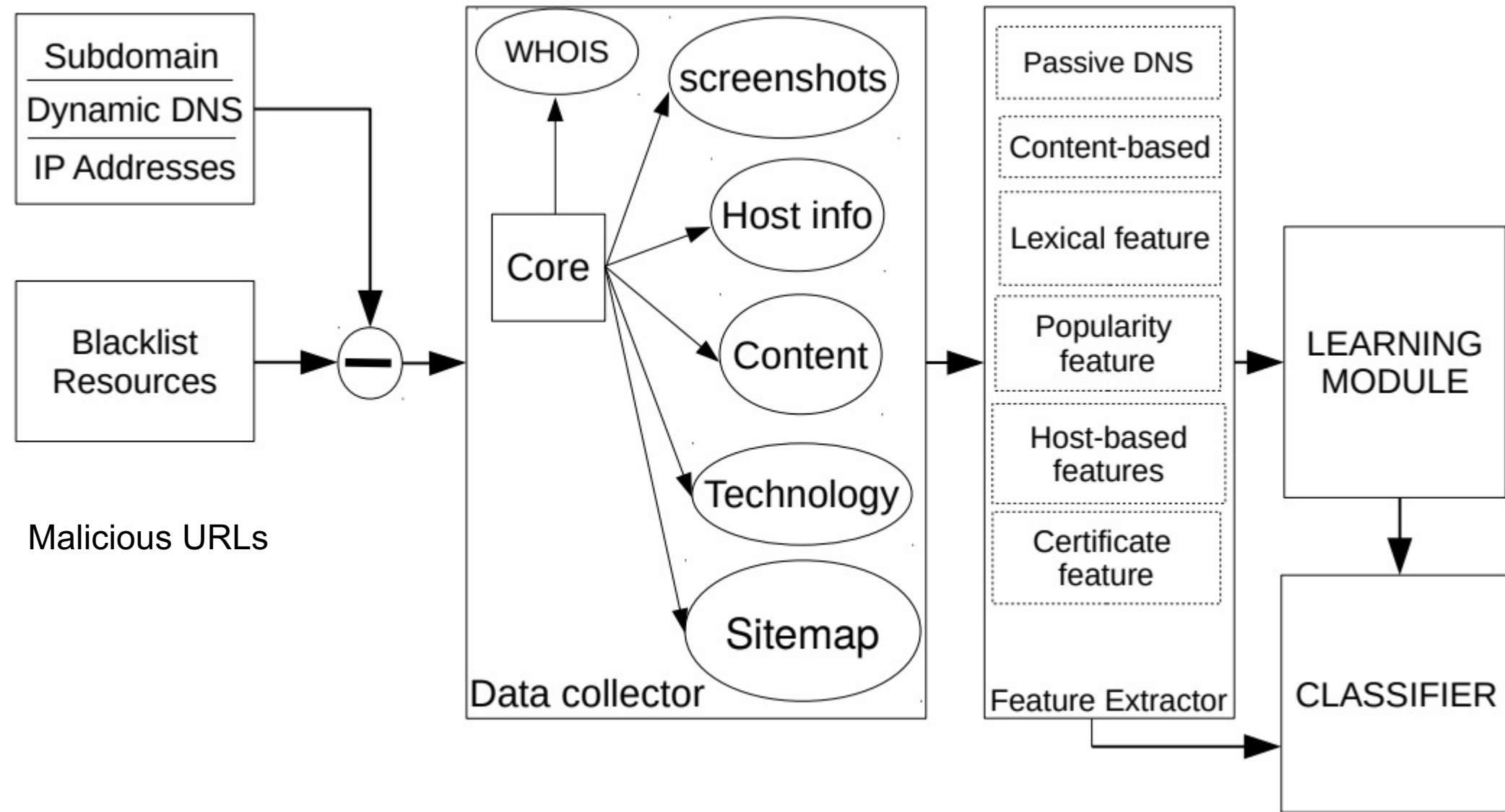
- We developed COMAR, a machine-learning system to classify domains exhibiting malicious behavior as either *compromised* or *maliciously registered* by only using **publicly available** and **readily accessible resources** and achieve 97% accuracy with 2.5% of false positives.
- We leverage 38 features to identify the state of a domain, 14 of which are new and have not been used in previous work.
- We introduce a new method to estimate the domain creation time in cases there is no access to WHOIS information, which outperforms standard statistical methods in filling missing values.
- We show that content-based features are the most important ones in representing the domain status.

COMAR Concept

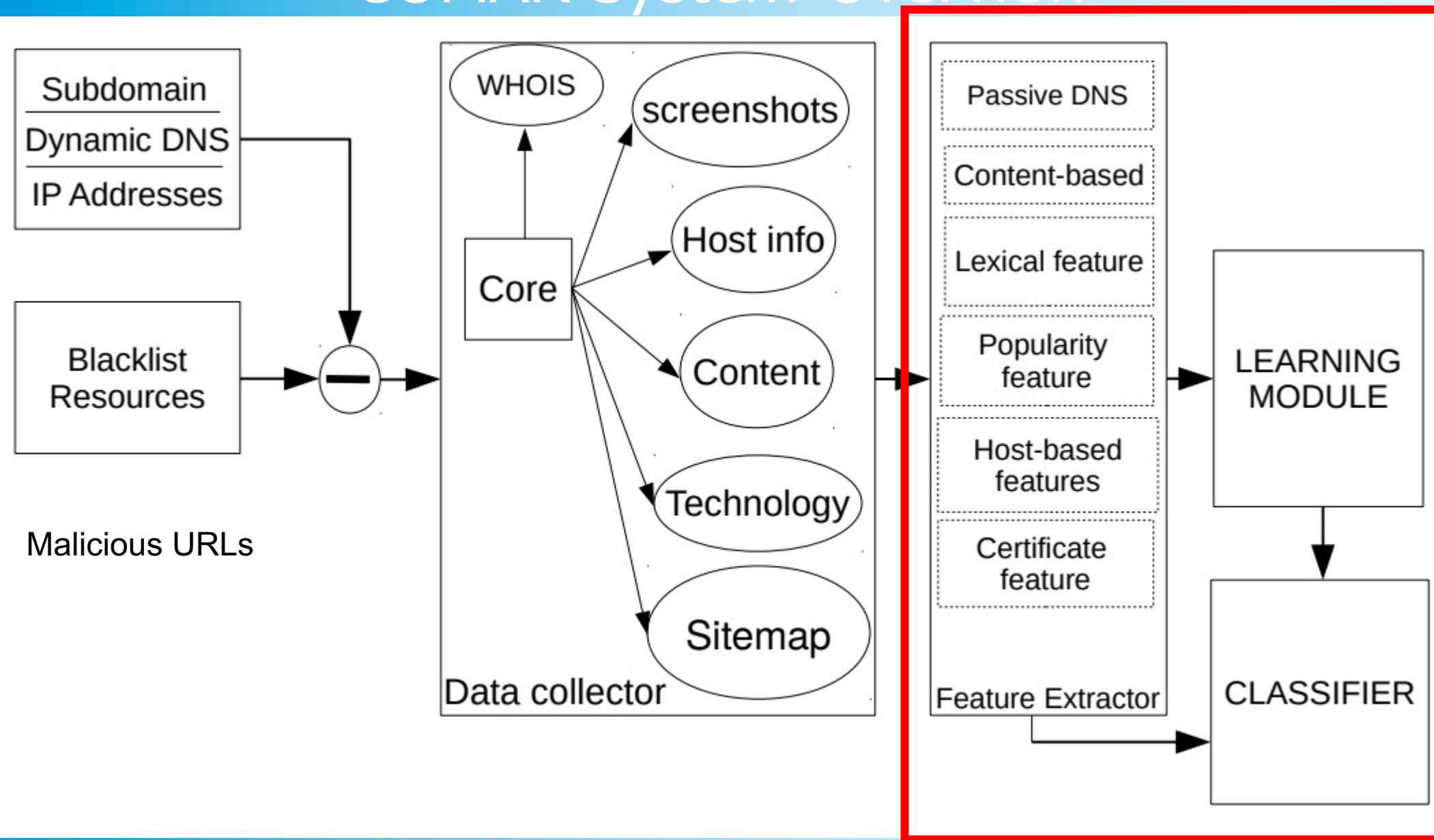
Passive information (e.g., lexical patterns of URLs)
Active measurements (e.g., hosting and DNS infrastructure)



COMAR System Overview



COMAR System Overview



Feature categories

In total we implemented **38** features (**14** of them new) in 7 categories:

1. Lexical features
2. Ranking system and popularity features
 - Internet archive, Alexa, Majestic, Umbrella, ...
3. ~~Passive DNS features (DNSDB)~~
4. Content-based features
 - Internal and external hyperlinks
 - Content length
 - Vulnerable technologies and number of technologies
 - Existence of home page (Is it default or not?)
 - Using redirection techniques
 - ...
5. WHOIS and TLD-based features
 - Spamhaus index
 - Domain age
 - ...
6. TLS certificate features
7. Active DNS features

To make COMAR practical, system uses only **publicly available** and **readily accessible resources**, we removed “passive DNS”.

Missing features

What if we can not collect data for some features?

In practice, there are always missing values when it comes to collecting features

If we can not collect data, we can not classify domains or even worse: misclassification

We need to fill the missing values appropriately. For example:

- WHOIS feature:
Some TLDs do not provide registration date (e.g., .de, .tk, .ml, ...)
- Content features:
We can not fetch the content for any reason (bot detection or host suspension)

Dataset & Machine Learning

Dataset → manually labeled 2,329 domains as maliciously registered, compromised, subdomain/free service, or false positives

1. Number of phishing domain labeled manually: 1,321 (OpenPhish, PhishTank, Anti-Phishing Working Group)

58% maliciously registered - 42% compromised

2. Number of malware domains from URLhaus labeled manually: 1,008

43% maliciously registered - 57% compromised

Dataset & Machine Learning

Machine learning

1. Logistic regression

- Parametric method known for its efficiency
- Low computational resources
- Interpretability

2. Random forest

- Non-parametric
- training a non-linear model
- No feature transformation

Results

Overall results for both classifiers and comparison with method used in Global Phishing Surveys.

Domain is malicious if:

- it is reported “very short time after registration”, or
- contains a brand name or misleading string, or
- registered in batch (challenging after GDPR)

Method	DB	Acc	Precision	Recall	F1	MCC
RF	Phish	97%	95%	97%	96%	0.93
LR	Phish	96.5%	96.59%	95%	95.7%	0.92
APWG	Phish	85%	82%	93%	88%	0.69
RF	Mal	96%	97%	96%	97%	0.92
LR	Mal	94.5%	95.6%	95.2%	95.4%	0.89

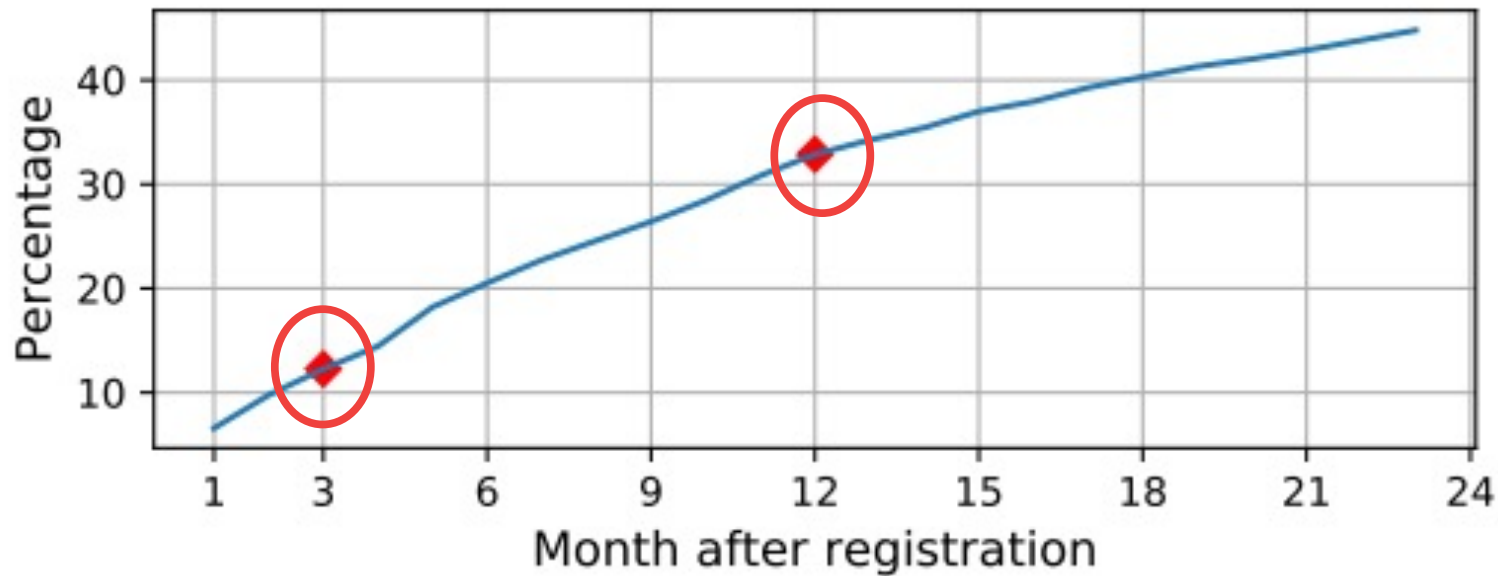
False positive rate: COMAR → **2.5%** APWG → **26%**

False positive: maliciously registered domains classified as compromised

Results

Risk of false negatives

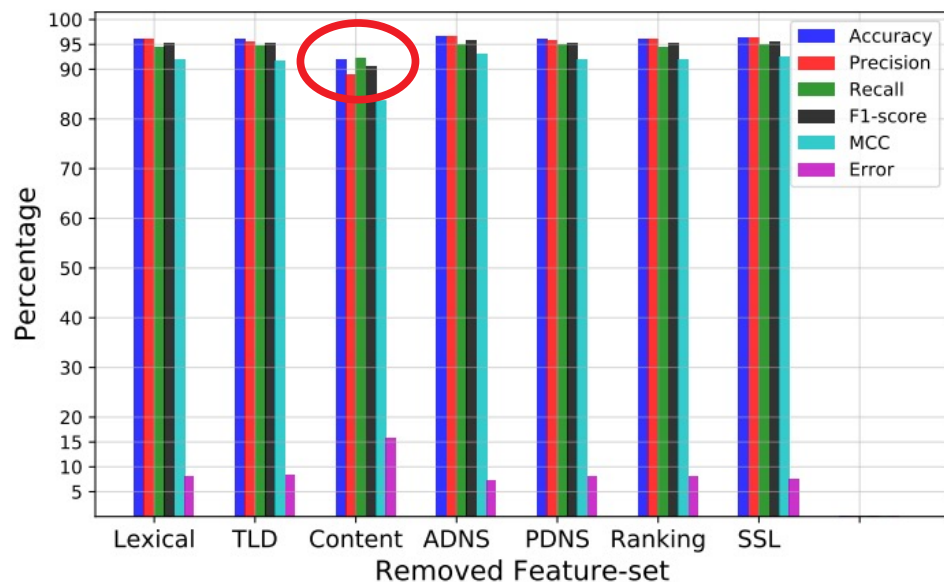
False negative: classifying compromised domains as maliciously registered



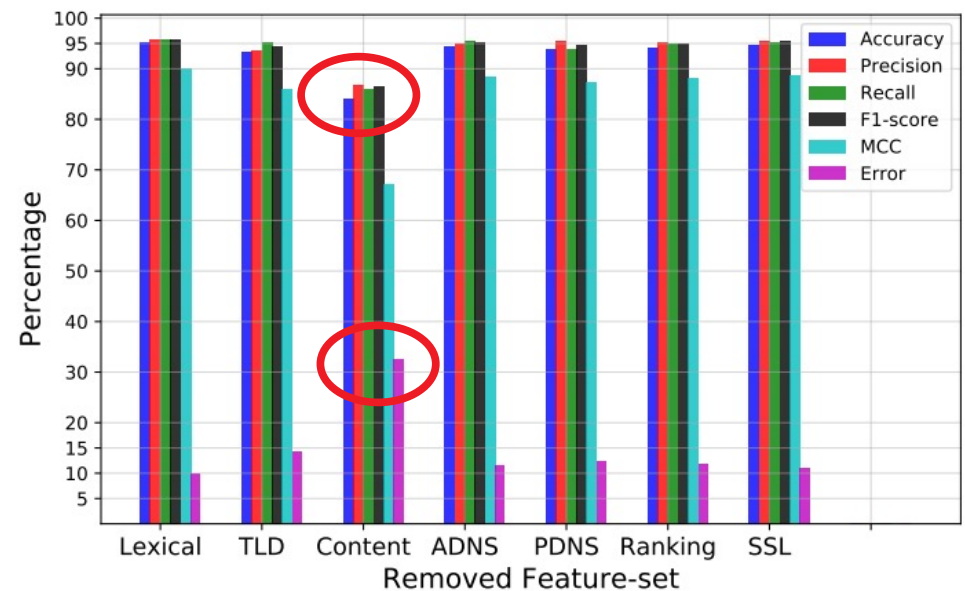
Partial cumulative distribution of the compromised domains after registration

Results

Applying logistic regression on the phishing and malware dataset by removing one feature-set at a time.



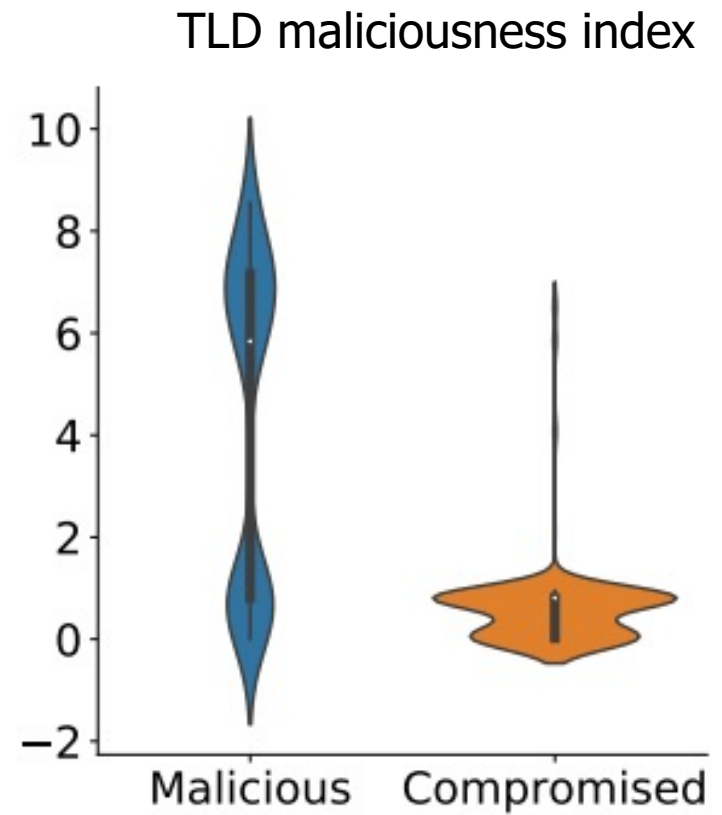
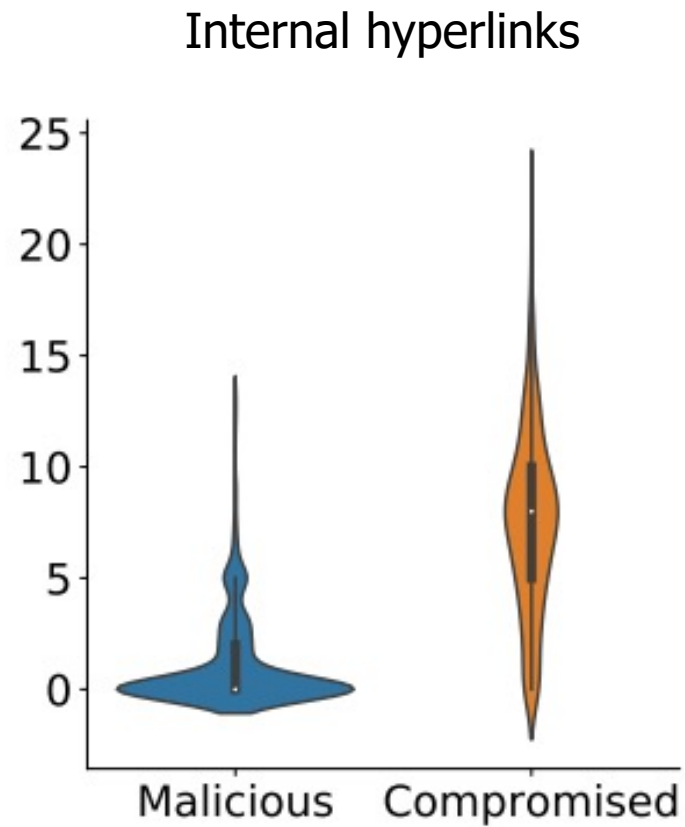
Phishing dataset



Malware dataset

Results

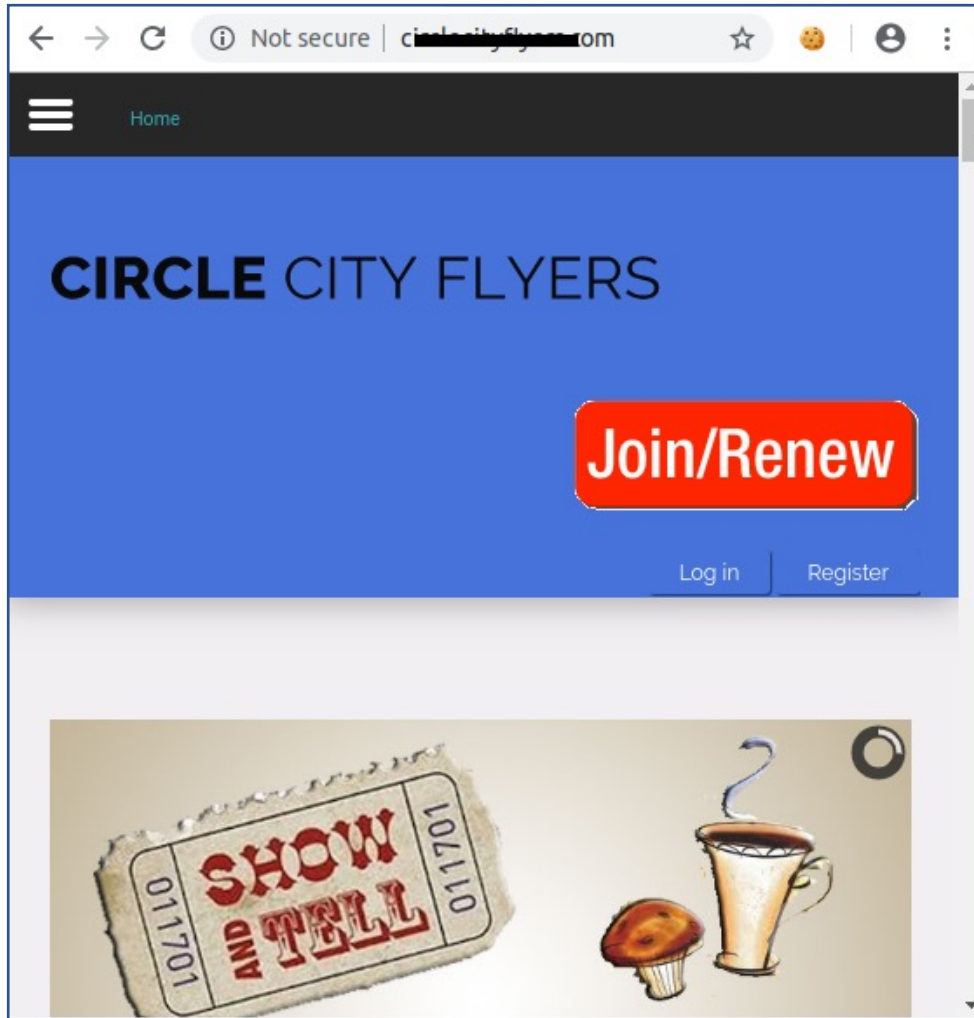
Distribution of the 'internal hyperlinks' (logarithmic scales) and 'TLD maliciousness index'¹



¹<https://www.spamhaus.org/statistics/tlds>

New anti-phishing evasion technique

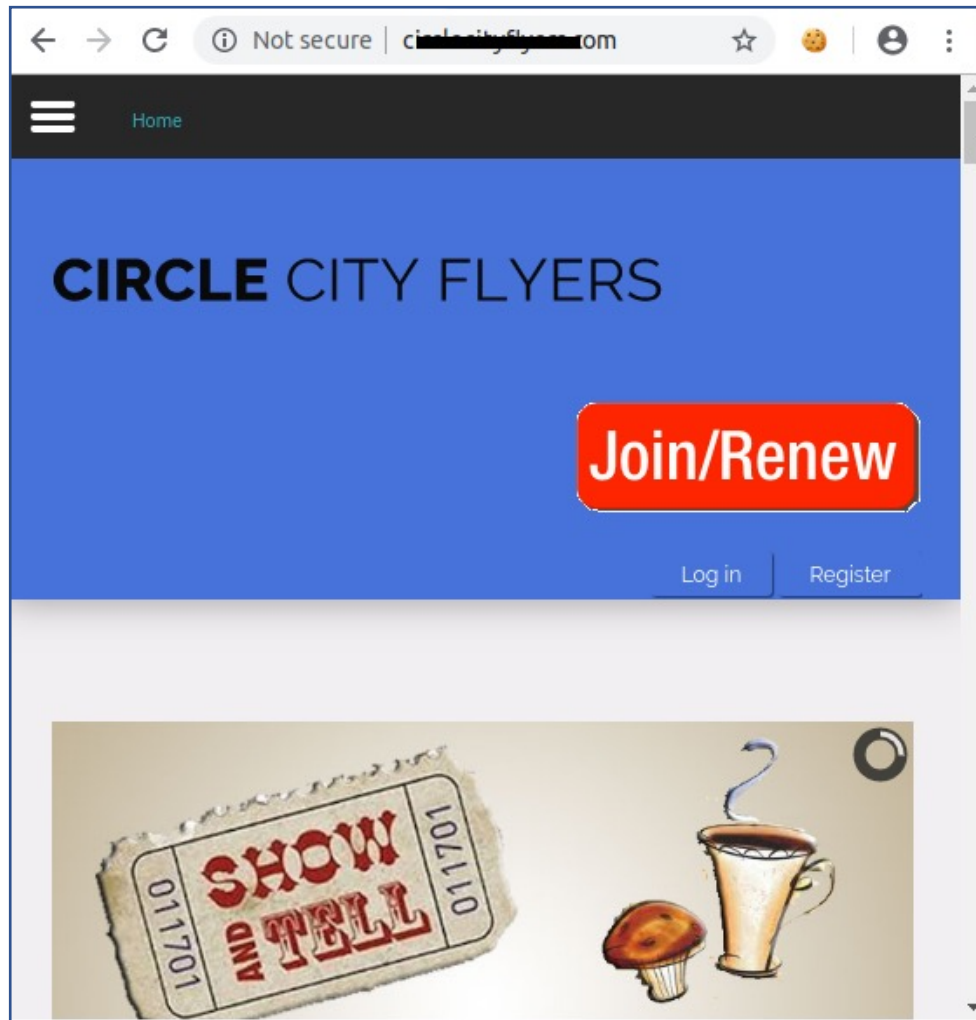
<http://circlecityflyers.com>



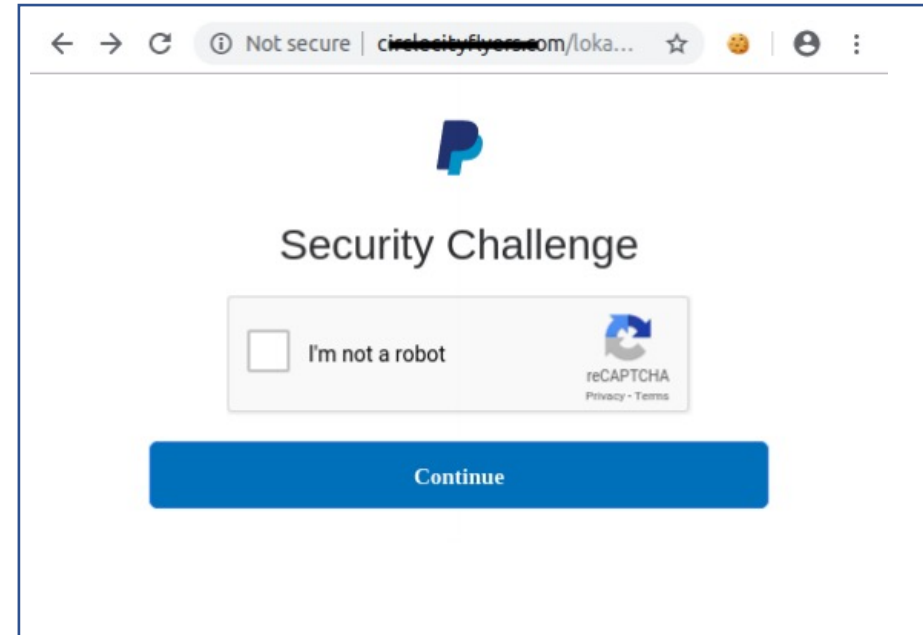
<http://circlecityflyers.com/lokako/newicn/cust0mers...>

New anti-phishing evasion technique

<http://circlecityflyers.com>

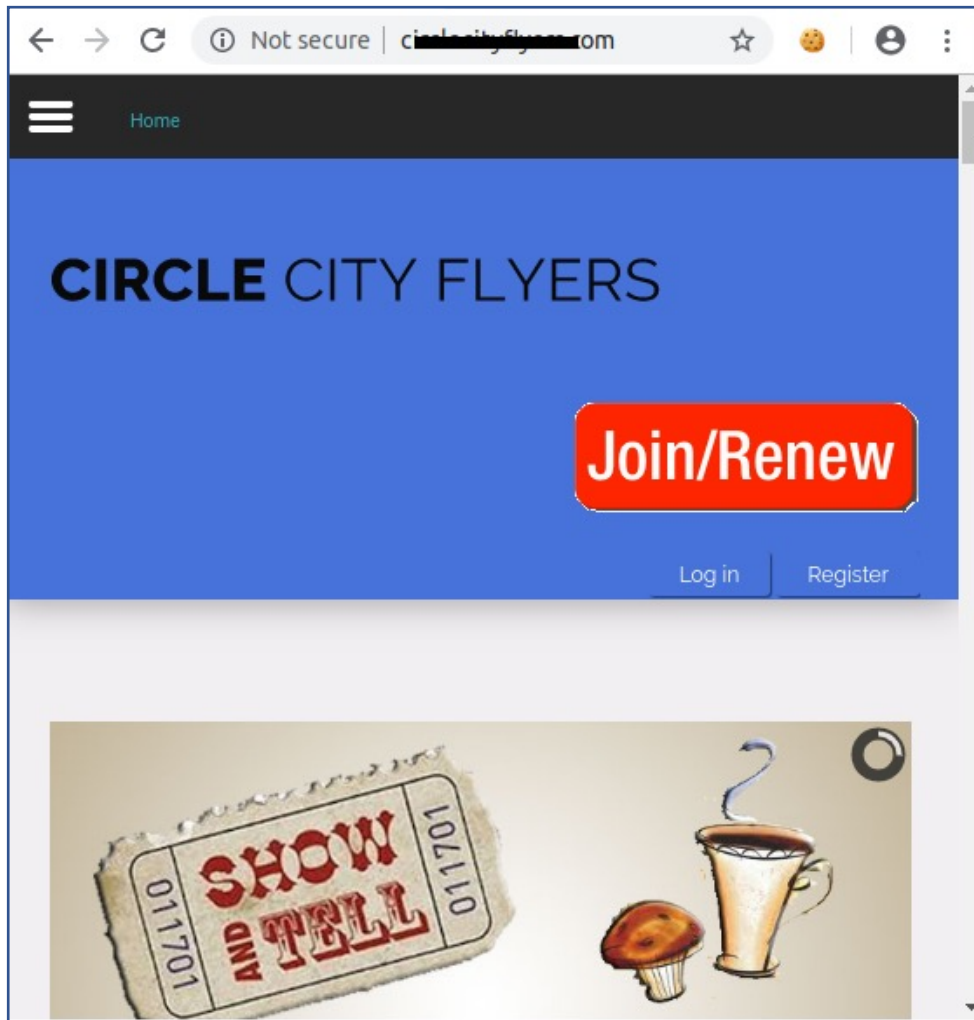


<http://circlecityflyers.com/lokako/newicn/cust0mers...>

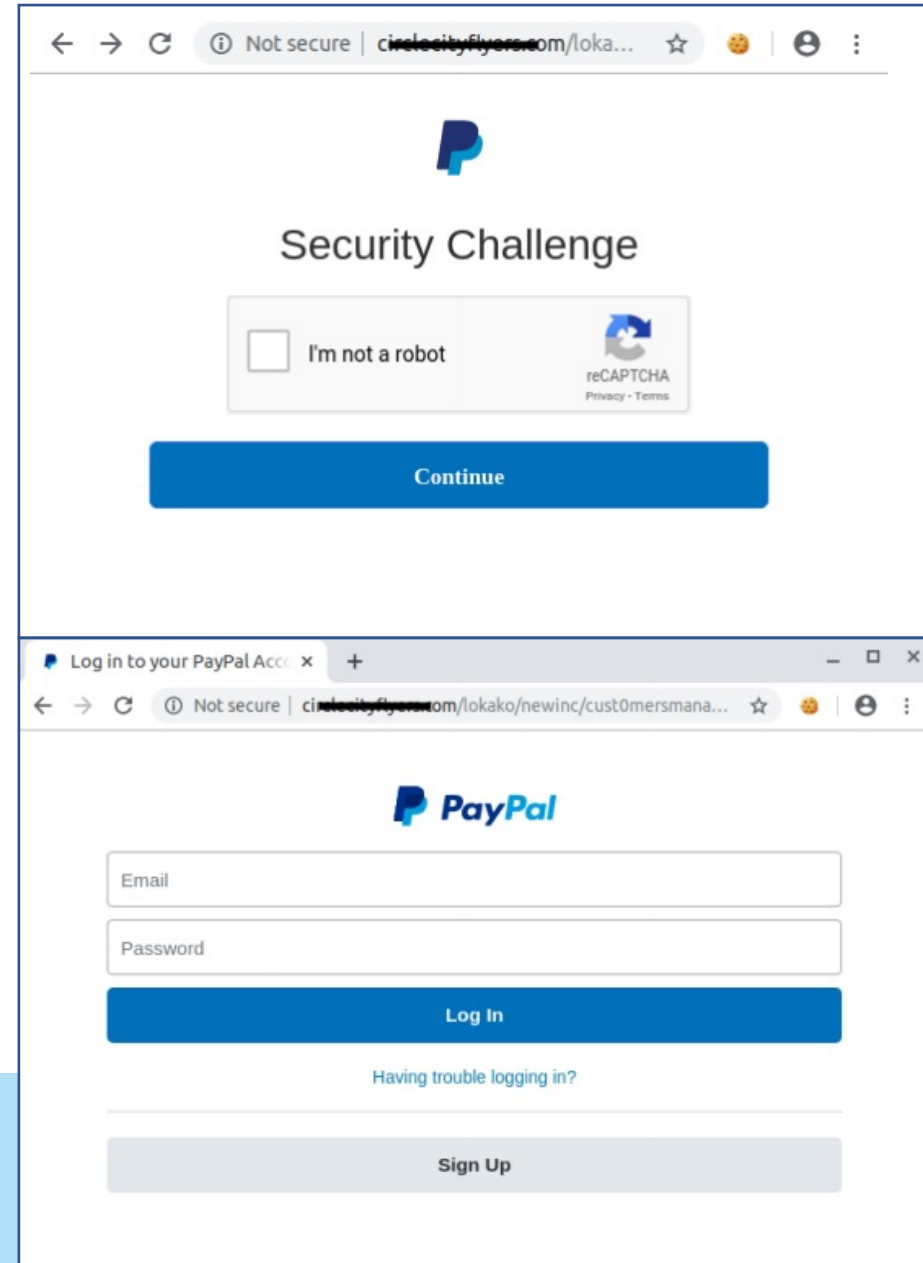


New anti-phishing evasion technique

<http://circlecityflyers.com>



<http://circlecityflyers.com/lokako/newicn/cust0mers...>



New anti-phishing evasion technique: Google CAPTCHA to hide the real content of the malicious page

Conclusion

COMAR leverages publicly available data and makes classification decisions based on the extracted features

Registries, registrars, and hosting providers can use it to decide on appropriate mitigation actions for each domain with malicious content

It can also serve as an effective tool for creating domain blacklists from the existing URL ones

We showed that the content-based features are the most effective in capturing the 'amount of beingness' of domains during their life cycles

Relatively hard to evade features

Conclusion

"COMAR: Classification of Compromised versus Maliciously Registered Domains", Sourena Maroofi, Maciej Korczynski, Cristian Hesselman, Benoit Ampeau and Andrzej Duda, IEEE European Symposium on Security and Privacy (EuroS&P 2020), September 2020 (Acceptance rate: 14,6%)

"Are You Human? Resilience of Phishing Detection to Evasion Techniques Based on Human Verification", Sourena Maroofi, Maciej Korczynski, and Andrzej Duda, ACM IMC 2020, Pittsburgh, PA, USA, October 2020 (Acceptance rate: 24,5%)

<https://comar-project.univ-grenoble-alpes.fr/>

Acknowledgments

This work has been carried out in the framework of the COMAR project funded by SIDN, the .NL Registry and AFNIC, the .FR Registry.

We thank: Anti-Phishing Working Group, OpenPhish, PhishTank, URLhaus for providing access to their URL blacklists;

Farsight Security for sharing DNSDB, and the DNSDB data contributors

This work was partially supported by the ANR projects: the Grenoble Alpes Cybersecurity Institute CYBER@ALPS under contract ANR-15-IDEX-02, PERSYVAL-Lab under contract ANR-11-LABX-0025-01, and DiNS under contract ANR-19-CE25-0009-01.

Thank you for your attention!

Contact:

maciej.korczynski@univ-grenoble-alpes.fr