Domain Classifier: Compromised Machines *vs* Malicious Registrations

SOPHIE LE PAGE, UNIVERSITY OF OTTAWA

What are Phishing Attacks?

Internet attack

Website impersonates brand/organization

Steals end user’s sensitive information

Email common medium to reach users

URL in email leads to phishing website

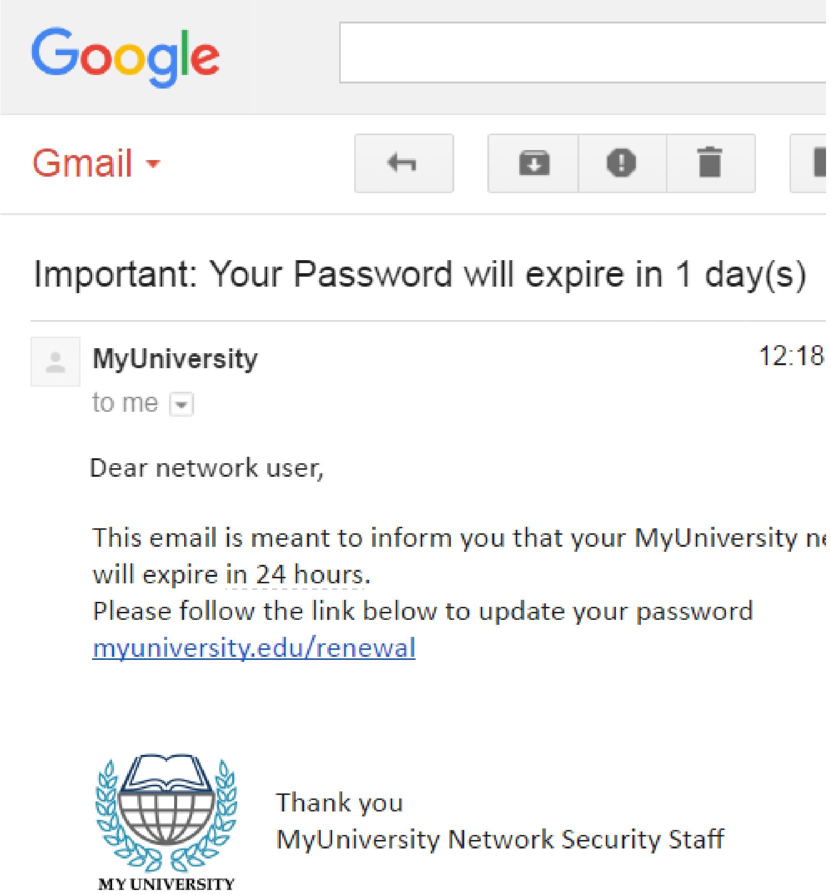
URL: www.**domain**.com/path

domain could lead to legit server hacked by attacker

- Compromised (comp): hacked, victim

domain could be owned by attacker

- Malicious (mal): owned, attacker



Important: Your Password will expire in 1 day(s)

MyUniversity 12:18
to me

Dear network user,

This email is meant to inform you that your MyUniversity n
will expire in 24 hours.
Please follow the link below to update your password
myuniversity.edu/renewal

Thank you
MyUniversity Network Security Staff

Problem and Motivation

Binary Classification Problem

Classify **known** phishing website hosted on **comp** or **mal** domain

(As opposed to classifying whether any site is a phishing website)

Motivation

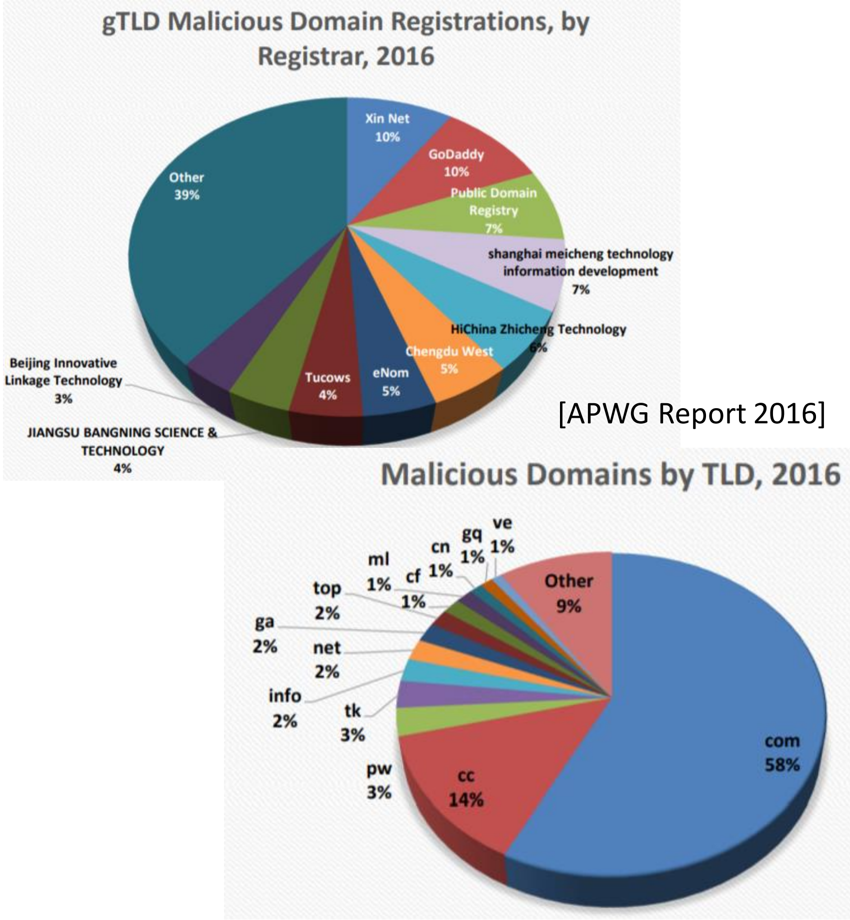
Understand how attackers commit their crime

Assess crucial information about server

Present different takedown strategies

Indicate which registrars allow more malicious registrations

Facilitate research focused on comp or mal cases



[APWG Report 2016]

Existing Solution

Anti-Phishing Work Group (APWG) reports[1]

- Report **49% mal** while rest are **comp**

Strategy

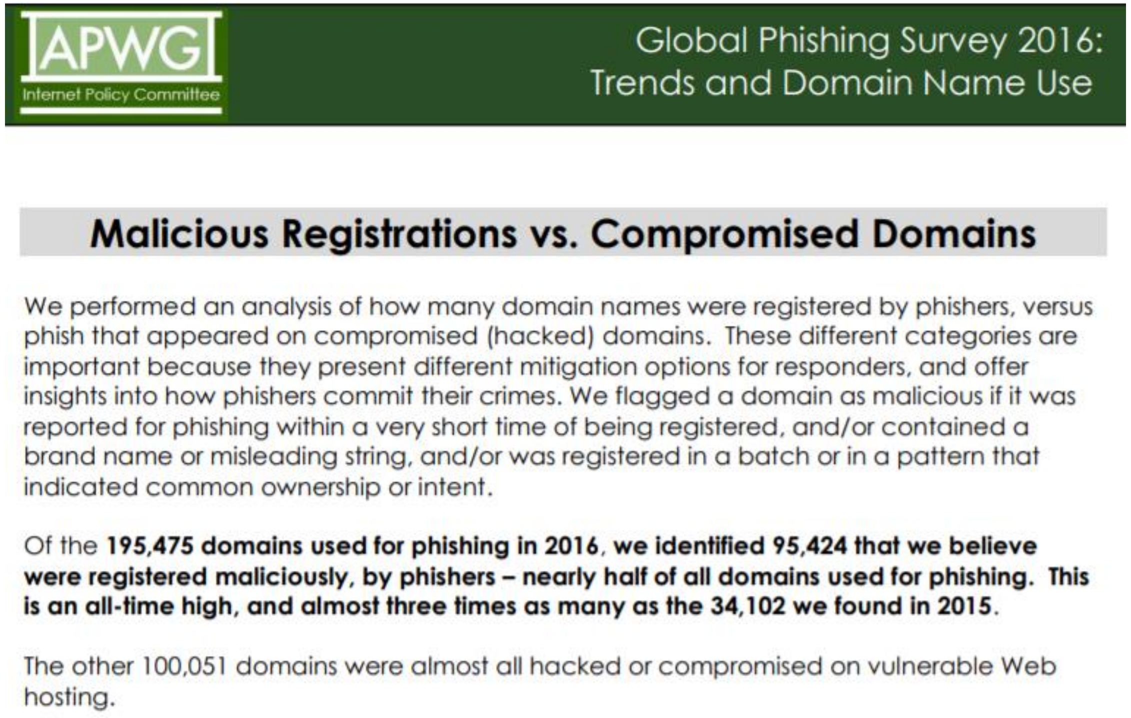
1) short timeframe from domain reg to report

2) brand/misleading string in domain name

3) batch domain names reg

Problem

- Only considers malicious criteria
- Not enough detail to reproduce method
- Domain registration not publicly available
- Registration info no longer available due to the General Data Protection Regulation (GDPR)



APWG Internet Policy Committee

Global Phishing Survey 2016: Trends and Domain Name Use

Malicious Registrations vs. Compromised Domains

We performed an analysis of how many domain names were registered by phishers, versus phish that appeared on compromised (hacked) domains. These different categories are important because they present different mitigation options for responders, and offer insights into how phishers commit their crimes. We flagged a domain as malicious if it was reported for phishing within a very short time of being registered, and/or contained a brand name or misleading string, and/or was registered in a batch or in a pattern that indicated common ownership or intent.

Of the **195,475 domains used for phishing in 2016**, **we identified 95,424 that we believe were registered maliciously, by phishers – nearly half of all domains used for phishing. This is an all-time high, and almost three times as many as the 34,102 we found in 2015**.

The other 100,051 domains were almost all hacked or compromised on vulnerable Web hosting.

[1] APWG Global Phishing Survey 2016: Trends and Domain Name Use

Our Widely Usable Solution

Apply machine learning (ML) algorithms with reproducible steps

… On labeled data from industry and research

… Weighing both comp and mal criteria

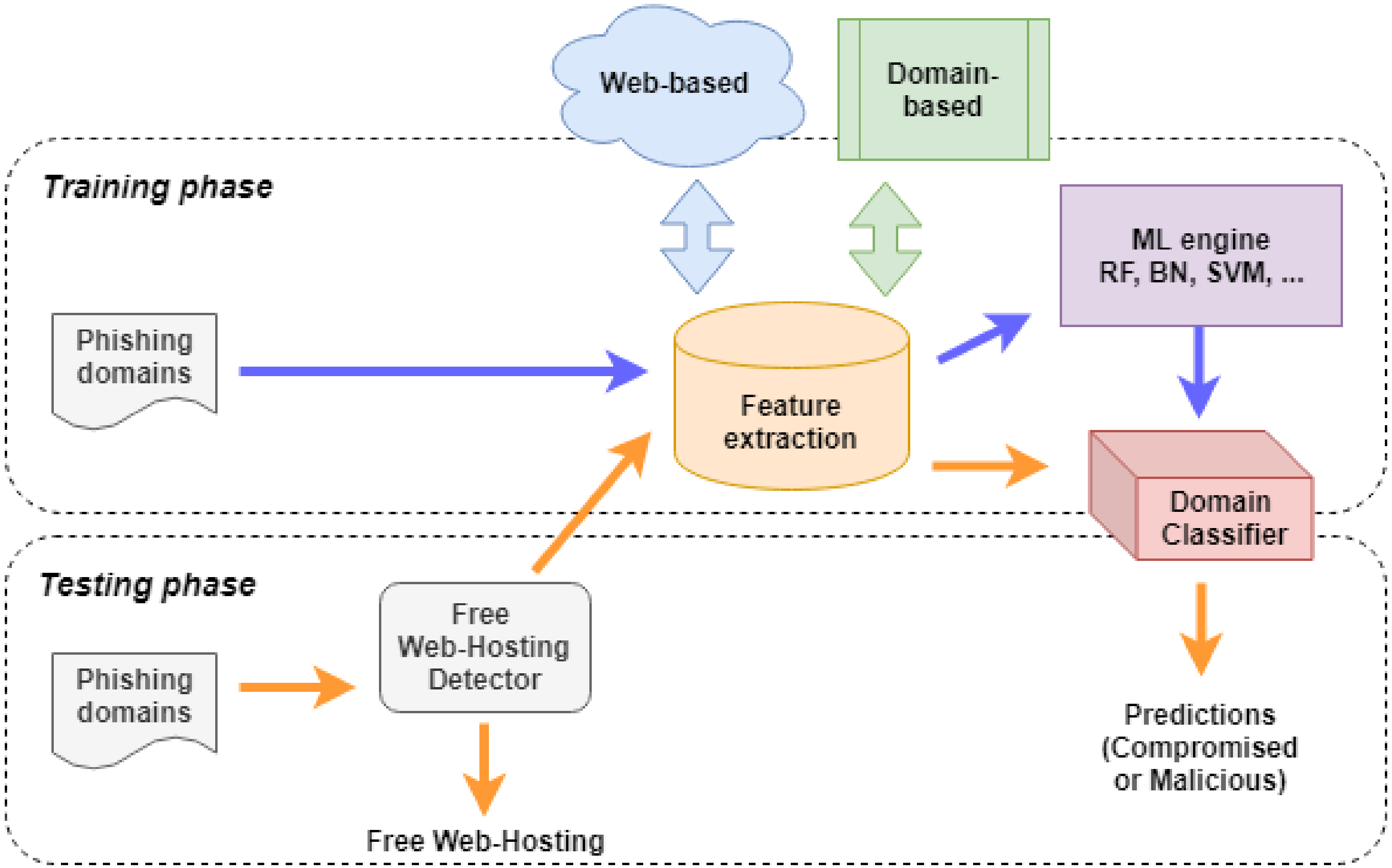
… Extracted from freely available data

INTERNET ARCHIVE

All resources for this work can be found at http://ssrg.site.uottawa.ca/icwe2019/

Experimental Setup

System Architecture



Labeled Datasets and Phish Feed

Labeled datasets

- 10,000 malicious phishing domains
- 1,400 compromised phishing domains
- PhishLabs, DeltaPhish

PHISHLABS
Prevent. Defend. Fight back.

DeltaPhish
Detecting Phishing Webpages in Compromised Websites

Research group unlabeled phish feed

- 69,000 phishing domains spanning 3 years
- PhishTank, IBM X-Force, OpenPhish

PhishTank

IBM
X FORCE

OpenPhish

~80,000 phishing domains in total

Fetching 15 features for ~80,000 domains

9 novel Web-based features

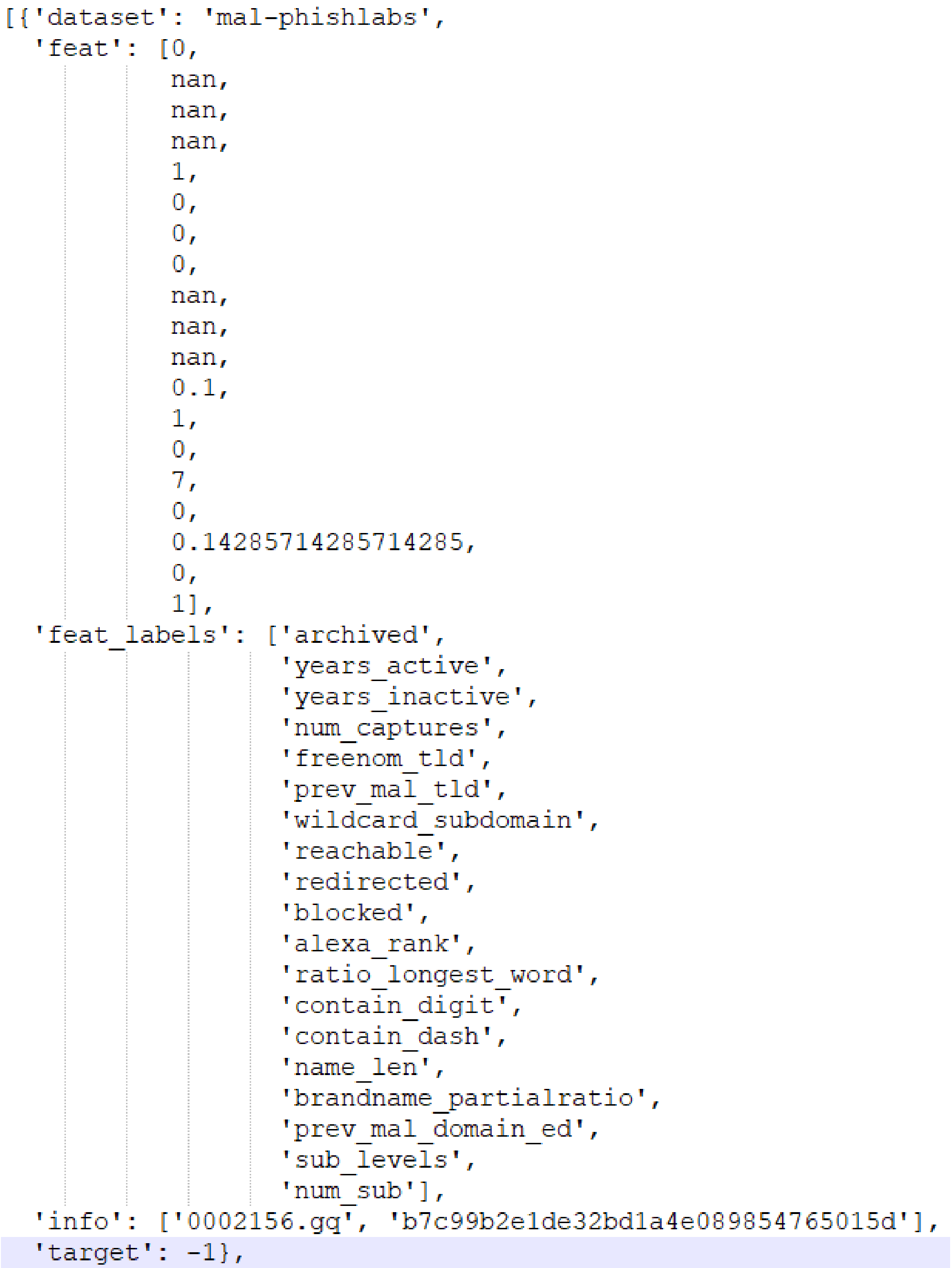
- Archived features **(domain history)**
- Reachable features **(domain presence)**
- Rank features **(domain reputation)**

6 Domain name based features[1,2]

- Contain digits/dashes
- Contain misleading string

25 sec per domain feature extraction

Feature extraction can be done in parallel



```
[{'dataset': 'mal-phishlabs',
  'feat': [0,
           nan,
           nan,
           nan,
           1,
           0,
           0,
           0,
           nan,
           nan,
           nan,
           0.1,
           1,
           0,
           7,
           0,
           0.14285714285714285,
           0,
           1],
  'feat_labels': ['archived',
                  'years_active',
                  'years_inactive',
                  'num_captures',
                  'freenom_tld',
                  'prev_mal_tld',
                  'wildcard_subdomain',
                  'reachable',
                  'redirected',
                  'blocked',
                  'alexa_rank',
                  'ratio_longest_word',
                  'contain_digit',
                  'contain_dash',
                  'name_len',
                  'brandname_partialratio',
                  'prev_mal_domain_ed',
                  'sub_levels',
                  'num_sub'],
  'info': ['0002156.gq', 'b7c99b2e1de32bd1a4e089854765015d'],
  'target': -1},
```

[1] APWG Global Phishing Survey 2016: Trends and Domain Name Use
[2] Predator: Domain abuse at time-of-registration, Hao et al., CCS 2016

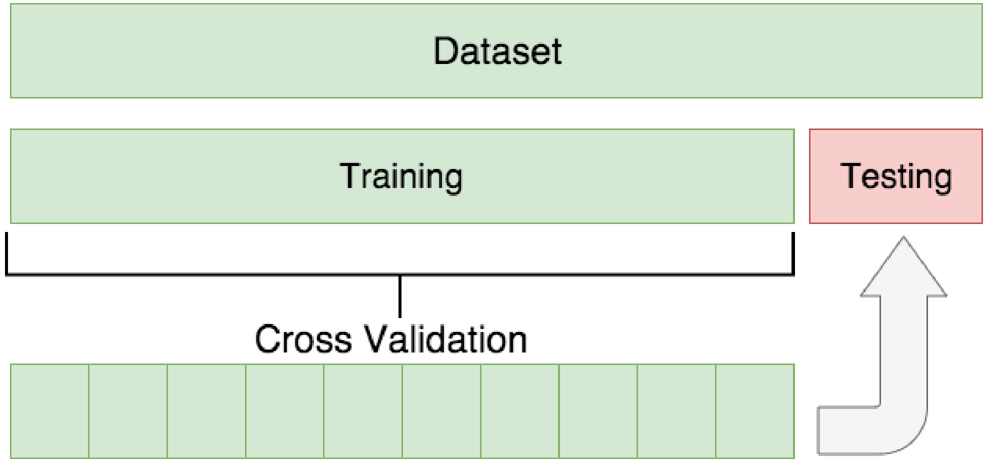
Evaluation Method

Randomized sampling with train / test split

Train on **balanced** dataset, test on the rest

- In real world volume of mal / comp cases are similar[1]
- Avoid ML algorithm favoring majority class

Use 5-fold cross validation on training dataset



Train on PhishLabs comp domains (less accurate)

- “my-apple-id.cf”, “doocs.gq”, “outlook-livesl.cf”.
- Take advantage of noisy labels to avoid overfitting data

Test on Deltaphish comp domains (more accurate, manually verified)

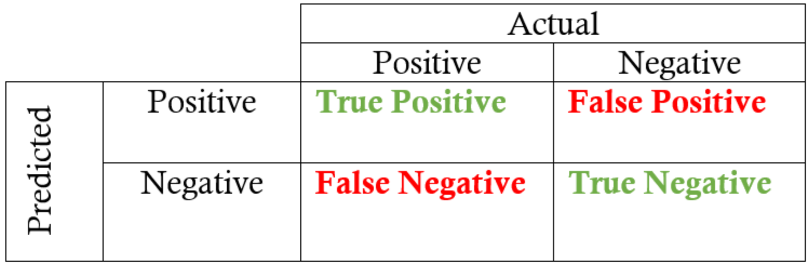
- Better indication of accuracy

[1] APWG Global Phishing Survey 2016: Trends and Domain Name Use

Evaluation Criteria

Compromised cases as positive class (+)

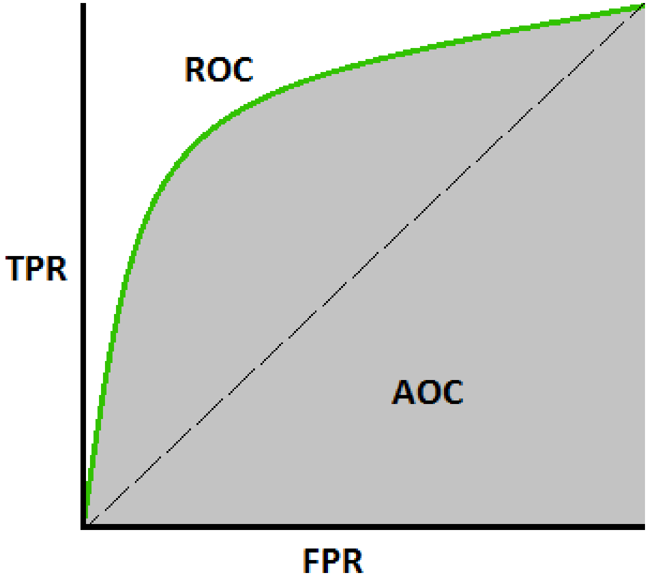
Malicious cases as negative class (-)



		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Evaluate classifier

- True Positive (TP) rate
- False Positive (FP) rate
- Accuracy



Tuning model (try different model parameters)

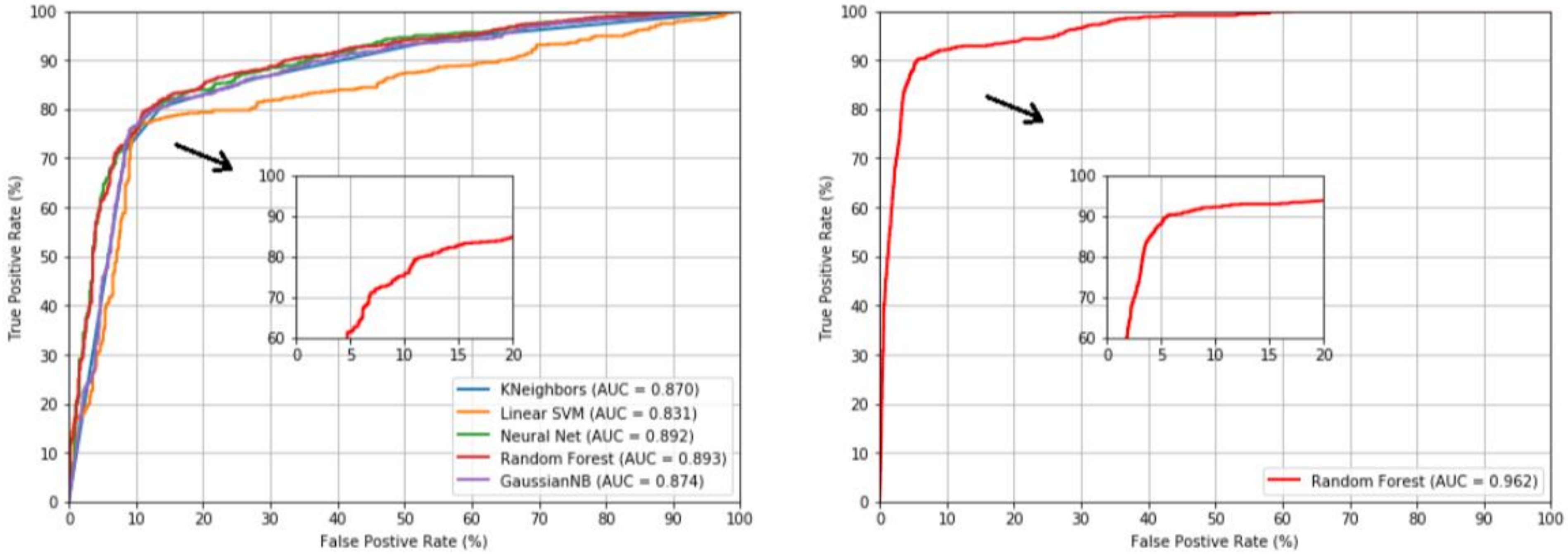
- Receiver operator curve (ROC), Area under the curve (AUC)
- Trade-off between TP and FP
- Summary statistic for model comparison

Experimental Results

Experimental Evaluation: Model Comparison

ROC curve (5-run cross validation average) and AUC of domain classifier using 5 ML algorithms

Random Forest (RF) classifier was the top performing algorithm



(a) Train set

(b) Test set

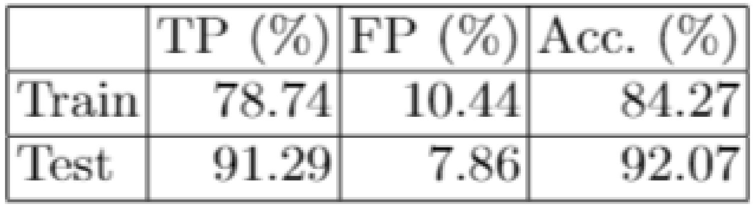
Experimental Evaluation: Random Forest Classifier

Performance of our RF domain classifier on test set

- High Accuracy of 92%
- High TP of 91%, reasonably low FP of 8%

Model performs better on test set than train set

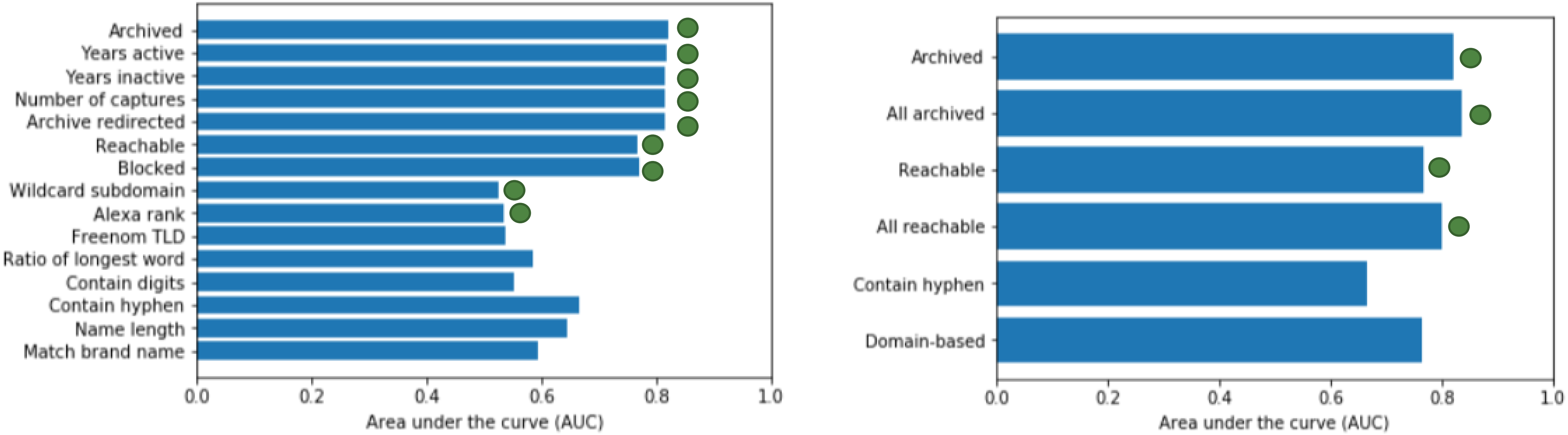
- Noisy compromised cases from PhishLabs used for training
- Manually verified compromised cases from DeltaPhish used for testing
- Model generalizes well to unseen data



	TP (%)	FP (%)	Acc. (%)
Train	78.74	10.44	84.27
Test	91.29	7.86	92.07

Learning with Individual Features

Our novel web-based features (●) stand out as high contributors to the classification.



(a) Individual features.

(b) Grouped features

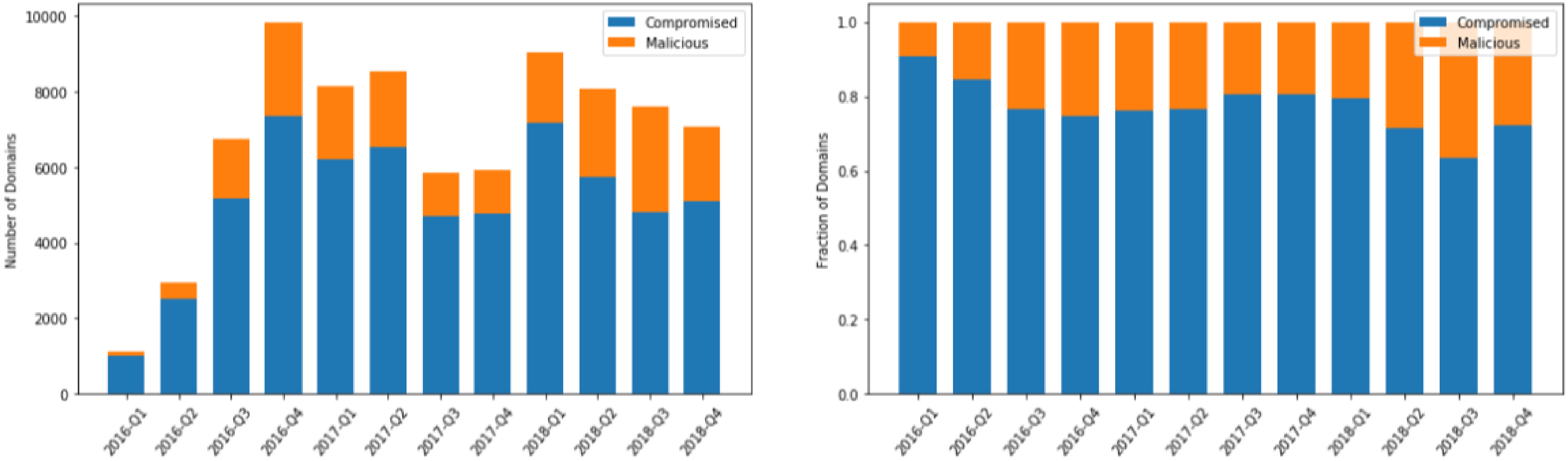
Predict Research Group Phish Feed

Predict 69,000 domains over 3 years: **73%** comp vs. **27%** mal (similar to other findings[1,2]).

Increase in mal cases toward **2018** indicates attackers exploiting free domain registrar services.

freenom

.tk .ga .cf .ml .gq



(a) Volume proportions.

(b) Fraction proportions.

[1] DeltaPhish: Detecting phishing webpages in compromised websites, Corona et al., ESORICS 2017

[2] Evil searching: Compromise and recompromise of hosts for phishing, Moore et al., Financial Cryptography 2009

Runtime Performance

Training phase can be done periodically offline.

Testing phase is conducted online as each phishing domain arrives.

Experiments done on standard computer with 2.10GHz processor and 14.7GB RAM.

Feature extractor module experiences critical time

- Crawler waits 10 seconds for page to load for Archive features
- Avg runtime 25 seconds per domain

Improve time performance

- Use 5 threads to process 5 domains in parallel (1 every 5 seconds)
- Cashing query and result of web-based features

Limitations

Internet Archive is internationally biased[1]

- E.g. less likely to have captured Chinese websites

Attackers may intentionally register a domain with history

- Limits attacker from registering domains in batches
- Domain with history may be more expensive, less scalable
- Prevents attacker from choosing domain name that is misleading

Hard to acquire accurate labeled data

- May need to use semi-supervised learning, or active learning

[1] A fair history of the web examining country balance in the internet archive, Thelwall et al. LISR 2004

Conclusion and Future work

Proposed widely useable domain classifier

- Uses only freely available data
- Fast feature extraction

Novel domain history features are good indicators of compromise

- Helps balance compromised and malicious criteria

Compromised cases are more common than malicious cases

- Still see a large percentage of malicious cases
- Free domain and free web hosting services may make malicious cases more popular

Combine domain predictions with click analysis

- Determine whether compromised cases receive more clicks and last longer than malicious cases

Thanks!

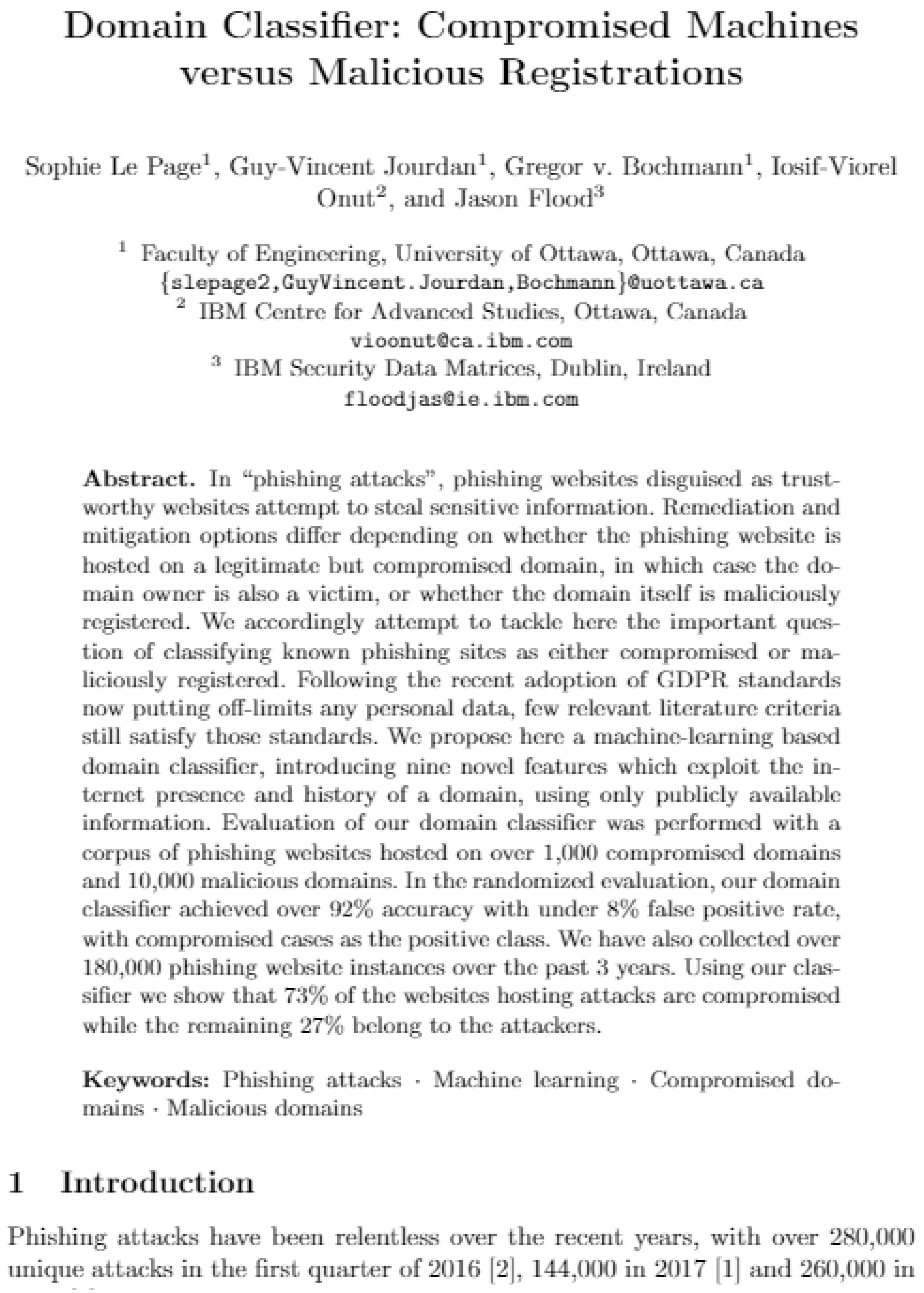
Any questions?

You can find me at:

slepage2@uottawa.ca

Or access our website:

http://ssrg.site.uottawa.ca/



Domain Classifier: Compromised Machines versus Malicious Registrations

Sophie Le Page[1], Guy-Vincent Jourdan[1], Gregor v. Bochmann[1], Iosif-Viorel Onut[2], and Jason Flood[3]

[1] Faculty of Engineering, University of Ottawa, Ottawa, Canada
{slepage2,GuyVincent.Jourdan,Bochmann}@uottawa.ca
[2] IBM Centre for Advanced Studies, Ottawa, Canada
vioonut@ca.ibm.com
[3] IBM Security Data Matrices, Dublin, Ireland
floodjas@ie.ibm.com

Abstract. In “phishing attacks”, phishing websites disguised as trustworthy websites attempt to steal sensitive information. Remediation and mitigation options differ depending on whether the phishing website is hosted on a legitimate but compromised domain, in which case the domain owner is also a victim, or whether the domain itself is maliciously registered. We accordingly attempt to tackle here the important question of classifying known phishing sites as either compromised or maliciously registered. Following the recent adoption of GDPR standards now putting off-limits any personal data, few relevant literature criteria still satisfy those standards. We propose here a machine-learning based domain classifier, introducing nine novel features which exploit the internet presence and history of a domain, using only publicly available information. Evaluation of our domain classifier was performed with a corpus of phishing websites hosted on over 1,000 compromised domains and 10,000 malicious domains. In the randomized evaluation, our domain classifier achieved over 92% accuracy with under 8% false positive rate, with compromised cases as the positive class. We have also collected over 180,000 phishing website instances over the past 3 years. Using our classifier we show that 73% of the websites hosting attacks are compromised while the remaining 27% belong to the attackers.

Keywords: Phishing attacks · Machine learning · Compromised domains · Malicious domains

1 Introduction

Phishing attacks have been relentless over the recent years, with over 280,000 unique attacks in the first quarter of 2016 [2], 144,000 in 2017 [1] and 260,000 in