Classification of Anti-Phishing Solutions

S. Chanti
Research Scholar

Pondicherry University,
Puducherry

March 4, 2019

Phishing

- Phishing is a fraudulent activity in which the attacker tries to achieve illegal financial gain either (i) *by stealing and spoofing user identity/ credentials or* (ii) *by usurping control of access to user information*

- Phishing can also be achieved through
  - Identity theft
  - Unauthorised access
  - Installation of Malware or spyware

Anti-Phishing

- Anti-phishing is a method through which the *phishing scams are detected and prevented*.
- Anti-phishing browser extensions / toolbars are of two types:
  - Content based
  - Non Content based

Formulation of Research Questions

From the study, we formulated three research questions

1. What are the areas that current Anti-Phishing solutions address?
2. Do the Existing Anti-Phishing toolbars cover all the phishing attacks?
3. What are the current Research gaps in Anti-Phishing?

Anti-Phishing Solutions

- In Content-based Phishing detection, the phishing attack is detected by analyzing the content of email, website, and social media.
- Non-Content based approaches focus on the features other than content. Blacklist, based on user rating, popularity of the domain and so on.

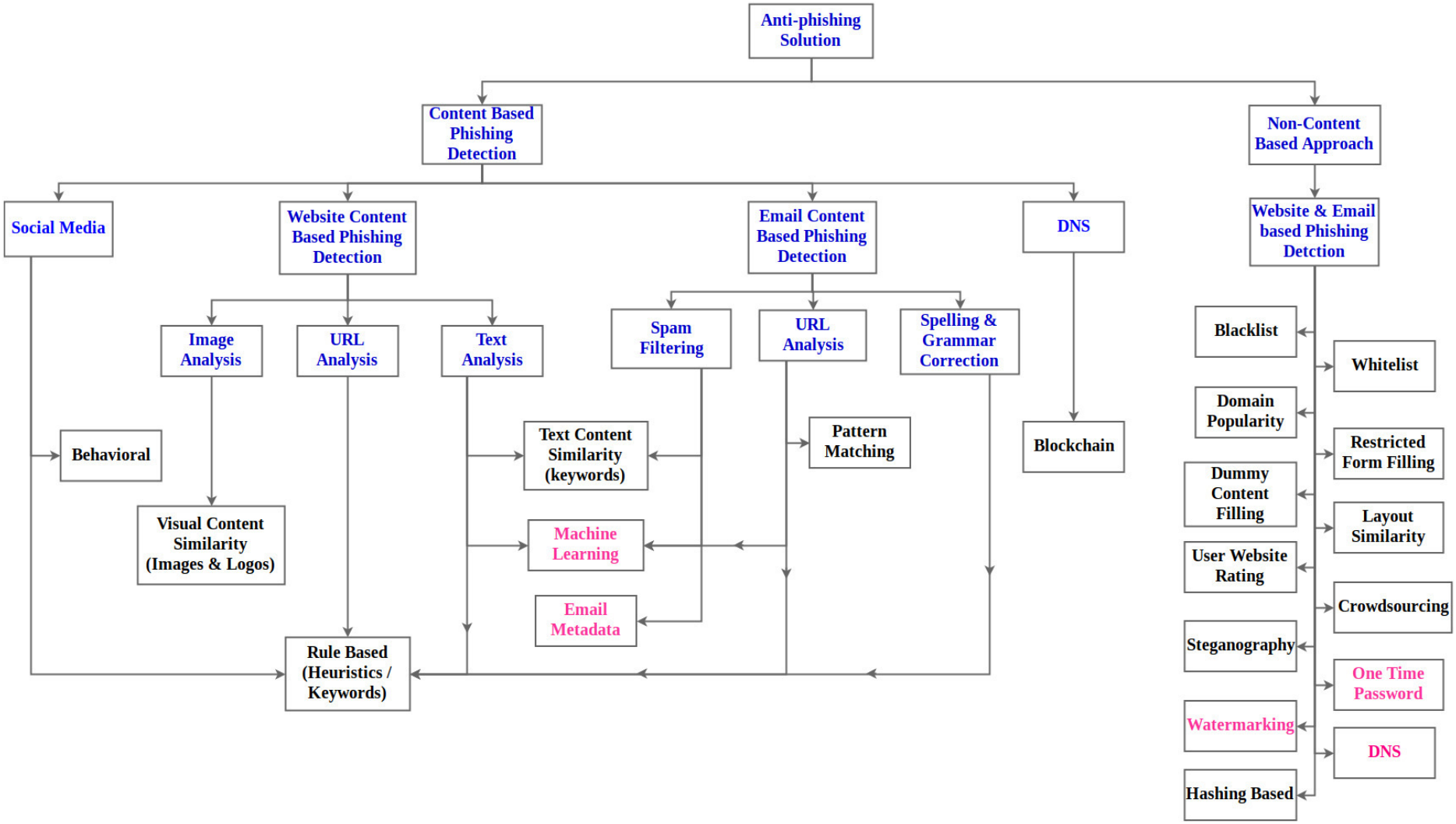
Content Based Phishing Detection

- Social Media
- Website Content
- Email Content
- DNS

Non-Content Based Phishing Detection

- Email and
- Website based Phishing Detection

Classification of Anti-phishing Solutions for Phishing Detection



List of Phishing Detection Features at Different Levels

<table>
<tr><th colspan="2">Email Features</th><th>Website Features</th></tr>
<tr><td colspan="2">Header Features</td><td>Address bar Features</td></tr>
<tr><td colspan="2">URL Feature in Email</td><td>Abnormal web Features</td></tr>
<tr><td colspan="2">Word List Feature</td><td>HTML and JavaScript</td></tr>
<tr><td colspan="2">Structural Features</td><td>Domain Features</td></tr>
<tr><td colspan="2">HTML Content</td><td>Graphical Features</td></tr>
<tr><td colspan="2">Email Body Features</td><td>Country-code & TLD</td></tr>
<tr><th colspan="3">URL Features</th></tr>
<tr><td rowspan="3">Social Media Features</td><td>Twitter</td><td>Facebook</td></tr>
<tr><td>Account Specific Features</td><td>Account Specific Features</td></tr>
<tr><td>Object Specific Features</td><td>Object Specific Features</td></tr>
</table>

Existing Anti-Phishing Approaches

Content Based Anti-Phishing Approaches

- Behavioral Based
- Visual Content Similarity Based
- Rule Based (Heuristics)
- Text Content Similarity Based
- Machine Learning Based
- Email Metadata Based
- Pattern Matching Based
- Blockchain Based

Existing Anti-Phishing Approaches

Non-Content Based Anti-Phishing Approaches

- Blacklist Based
- Whitelist Based
- Domain Popularity
- Restricted Form Filling
- Dummy Content Filling
- Layout Similarity
- User Website Rating
- Crowdsourcing
- Steganography
- One Time Password
- Watermarking
- DNS
- Hashing based

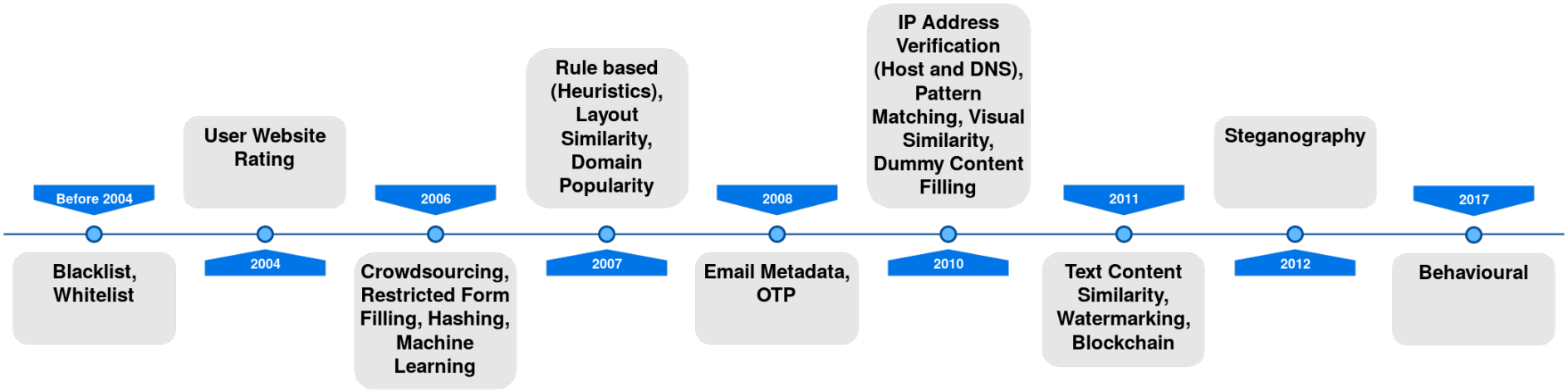
Popular Anti-Phishing algorithms used in Phishing Detection

Research papers			[63]	[35]	[5]	[73]	[1]	[68]	[38]	[11]	[55]
Dataset	Dataset Source	Phishing	APWG Archives	PhishTank	Manual	PhishTank	PhishTank	World Wide Web	WestPac	PhishTank	PIRT Report
		Legitimate	-	Google Whitelist	Manual	Alexa, Yahoo	Web Crawler	World Wide Web	WestPac	Common Crawl	Google Search
	Dataset Size	Phishing	203 Archives	200 web-sites	600 emails	100 URLs	3611 web-sites	279 web-sites	613048 emails	1 million emails	30 samples
		Legitimage	-	200 web-sites	400 emails	100 URLs	1638 web-sites	100 web-sites	4625 emails	1 million emails	500 sam-ples
Features	Email				*				*	*	
	Website		*	*			*	*			
	URL					*	*			*	*
	Social Media										
	DNS										
Approach Used			Rule Based, Pattern Matching	Machine Learning	Machine Learning	Rule Based	Machine Learning	Machine Learning	Machine Learning	Machine Learning	Blacklist
Algorithm Used			LinkGuard	TSVM	Natural Language pro-cessing, Wordnet	TF-IDF	Google Page Rank	Support Vector Machine (SVM)	Decision Trees	Random forest, LSTM	Blacklist Genera-tor

Popular Anti-Phishing algorithms used in Phishing Detection

Research papers		[63]	[35]	[5]	[73]	[1]	[68]	[38]	[11]	[55]
Performance in %	FPR	-	-	0.02	1	-	-	-	-	9
	FNR	-	-	0.04	-	-	-	-	-	-
	Precision	-	96.4	99.6	-	-	-	-	0.986	-
	Recall	-	90.7	99.3	-	-	-	-	0.989	-
	Accuracy	96	95.5	99.4	90	-	84	99.8	98.7	-
	F-Measure	-	-	-	-	-	-	-	0.987	-
Limitations		LinkGuard may result in false positives, since using dotted decimal IP address instead of domain names may be desirable in some special circumstances.	Major limitation of TSVM is that it involves an expensive matrix inverse operation when solving the dual problem.	The dataset size is small. The machine learning classifier needs more data for training the model to get good results.	It fails if the phisher uses a different language other than English. It is a time taking process as it choirs google each time. It also fails in the following cases. a. Using images in place of text, b. Using invisible text, c. Changing the words to confuse the system.	Google page rank algorithm cant classify phishing attacks correctly if it is a newly registered domain.	A smaller number of mislabeled examples can drastically affect DNS phishing attacks.	They considered only one part of features and they didnt address DNS phishing attacks.	The inner works are not easy to interpret easily in LSTM. The random forest required expert knowledge for feature selection.	Accuracy in detecting new phishing attacks is based on the updates received. It has a high false positive rate.

Evolution Roadmap of Anti-Phishing Solutions



Existing Anti-phishing Browser Extensions/Toolbars

- Most of the anti-phishing solutions are available as a browser extensions /toolbars.
- When the users install any anti-phishing toolbar /browser extension, it keeps monitoring the user activities and alerts them.
- There are few approaches that still at the research level, which is not fully evolved as a browser extension.
- The existing Anti-Phishing browser extensions/ toolbars are analyzed in terms of maturity level, mode of operation, pros, and cons.

Existing Anti-phishing Browser Extensions/Toolbars

Maturity Level

- Anti-Phishing Approaches that are fully explored as Browser extensions.
- Anti-Phishing Approaches that are still at research level.

Mode of operation

- Stand-alone
- From Server
- From Third Party

Existing Anti-phishing Browser Extensions/Toolbars

S.No.	Name of the Toolbar	Approach Used	Mode of Operation	PROS	CONS
1.	AntiPhish[47]	Restricted Form Filling	Stand-alone	AntiPhish detects phishing attacks correctly if it is purely an HTML webpage	It requires manual interaction of the user. Generates False alarms
2.	B-APT[37]	Machine Learning	Stand-alone	It uses machine learning approach with DOM analyzer for phishing detection.	B-APT is vulnerable to Website spoofing attack.
3.	BogusBitter[69]	Dummy Content Filling	Stand-alone	It feeds a large number of bogus credentials to protect the user credentials from the phisher.	The Phisher uses filtering techniques to collect the credentials
4.	DOM AntiPhish[51]	Layout Similarity	Stand-alone	The browser automatically stores the user password by hashing it. If the password is reused it will give an alert to the users.	Spoofed web pages with similar images and visual looks of the legitimate site to fool the user.
5.	Dynamic Security Skin[15]	Visual Similarity	Server	The user has to remember a image and a image to authenticate oneself to the server. To authenticate, the user has to perform a visual matching	There is a chance of leaking the verifier, leak of images, visual contents can be spoofed by the phisher.
6.	eBayAccount Guard[21]	Heuristic, Blacklist	Server	It allows users to submit the suspected sites to eBay which can be added to the their Blacklist.	Only applicable to eBay and PayPal sites and Denial of Service attacks are possible.
7.	FirePhish[60]	Open Database	Server	It maintains its own database to store the phishing site for better detecting the attacks.	They have to maintain their own safe and phishing sites.

Existing Anti-phishing Browser Extensions/Toolbars

S.No.	Name of the Toolbar	Approach Used	Mode of Operation	PROS	CONS
8.	GoldPhish [18]	Visual Similarity	Third Party	Protects from zero-day phishing.	Delays the rendering of a web page. Google PageRank algorithm is vulnerable to new phishing attacks.
9.	iTrustPage [50]	Blacklist, Whitelist	Third party	It is effective and easy to use.	Phishing pages should be discovered quickly and added to a blacklist. The Blacklist alone can't be a better solution for phishing detection.
10.	LinkGuard [63]	Blacklist, Whiltelist, Pattern Matching	Third Party	It detects known and unknown attacks with an accuracy of 96%. There is no false positive and false negatives for category 1.	False positives can possible in category 2 solution in the case of IP address verification in the place of Domain name.
11.	McAfee Site Advisor [57]	Rating the site with their own tests.	Server	McAfee maintains their own database that uses automatic crawlers that search the sites and perform tests and includes in the database.	It is vulnerable to detect phishing sites with embedded objects.
12.	Microsoft Smart Screen Filter [40]	Blacklist, Heuristics	Server	It provides additional security at the network level. It also protects from malicious attachments like keyloggers.	It may be vulnerable to newly created phishing attacks if the Blacklist not regularly updated.
13.	Netcraft [44]	Blacklist, Heuristics, User Rating	Standalone	It allows phishing site feed, provides phishing alerts, mapping of current phishing attacks.	The information like site rank, IP address, web server, net-block owner, and last changes made can help the phisher in many ways.

Existing Anti-phishing Browser Extensions/Toolbars

S.No.	Name of the Toolbar	Approach Used	Mode of Operation	PROS	CONS
14.	Passpet [67]	Restricted Form Filling	Server	Allows the user to remember only password to log in with multiple systems.	Vulnerable to Pharming attack. The phisher can steal the credentials of non-SSL protected sites by hijacking. It is also vulnerable to offline dictionary attacks.
15.	PhishProof [70]	Blacklist, Whitelist, Heuristics	Server	PhishProof uses three levels of security. It alerts the users on phishing sites. User input is not required. User can also report phishing sites.	It cannot protect the users from malware.
16.	PhishTank Site Checker [62]	Open Database	Server	It blocks the users for the sites which are already reported as phishing in their Open database.	New phishing attacks become difficult to detect unless the database is updated frequently. It is slow because the users have to report the site as phishing.
17.	PhishZoo [4]	Content Similarity	Server	PhishZoo creates their own trusted profiles with legitimate sites using a fuzzy hashing technique to detect phishing.	PhishZoo is vulnerable to website spoofing attack.
18.	Pixastic [61]	Steganography based	server	Robust Message based Image Stegnography algorithm is used to hide the secret image and protect the users not to enter the personal credentials in phishing websites.	Vulnerable to DNS spoofing attack, Brute force attack, and Print screen is also possible
19.	SpoofGuard [14]	Heuristics	Standalone	The advantage of this toolbar is stoping the outgoing data to phishing sites by performing image check and password check.	It shows a false alarm when the user visits the legitimate site for the first time.

Existing Anti-phishing Browser Extensions/Toolbars

S.No.	Name of the Toolbar	Approach Used	Mode of Operation	PROS	CONS
20.	SpoofStick [39]	-	Stand-alone	The user can change the appearance of the toolbar because of its user-friendliness and they address the graphics property.	Vulnerable to iframes attack if the user opens multiple windows while surfing.
21.	The Earthlink Toolbar [20]	Heuristics, User Rating	Server	It relays on the combination of heuristics, user ratings and manual verification. Toolbar displays a thumb to indicate whether the site is phishing or not.	No alert message is displayed for users. User ratings produce more false alarms.
22.	TrustWatch [26]	Blacklist	Server	TrustWatch provides a personal security ID to prevent the toolbar spoofing. It is easy to use.	Vulnerable to newly created phishing attacks if the database is not updated regularly.
23.	Verisign EV Green Bar Extension [23]	Domain Popularity	Server	It detects the phishing sites by verifying the SSL certificates of the site.	It only identifies SSL certificates given by VeriSign, not the other valid SSL certificates.
24.	Virtual Browser Extension [46]	Blacklist, Heuristics, Visual Similarity	Third Party	Alerts the users if the site is not present in the Whitelist they are maintaining.	Vulnerable to key-loggers, screen loggers, and client-side scripting attack.
25.	Web of Trust (WOT) [76]	Blacklist, Crowd-sourcing	Third Party	The reputation of the site is shown next to the search results. Very user-friendly.	A single rating from a person can make the site unsafe because it depends on user ratings.

Summary

The answers for the formulated research questions are as follows:

1. What are the areas that current Anti-Phishing solutions address?
   - When compared to Non-content based approaches, Content based approaches are better in detecting phishing.
   - Content based approaches like Rule based, Machine learning based approaches are good in detection.
   - Blockchain based aprroaches are good in protecting DNS level attacks.
2. Do the Existing Anti-Phishing toolbars cover all the phishing attacks?
   - Most of the Anti-Phishing toolbars work on any specific type of attacks.
3. What are the current Research gaps in Anti-Phishing?
   - Mobile Phishing, Voice Phishing, Social Media Phishing are the areas where more research is required.

Thank you

Reference