# Analyzing Phishing Webpage Features From Three Perspectives

## Final Report

Yiwen Li

## Abstract

*Phishing websites, malicious sites that impersonate a trusted third party to gain access to sensitive private data, continue to be a major threat to Internet users. In this project, we extract, analyze, and evaluate features of phishing webpages from three perspectives, 1) Web Content, 2) Network Features, and 3) Visual Similarity. We collect data from PhishTank and use the predefined features to conduct experiments and analysis. Our results and observations may provide useful help for future work such as building efficient phishing URL detection tools and building phishing webpage classifier.*


## 1. Introduction

Phishing is a social engineering crime generally defined as impersonating a trusted third party to gain access to private data. For example, an adversary might send the victim an email directing him to a fraudulent website that looks like a page belonging to a bank. The adversary can use any information the victim enters into the phishing page to drain the victim's bank account or steal the victim's identity. Despite increasing public awareness, phishing continues to be a major threat to Internet users. According to the Gartner survey, approximately 109 million U.S. adults have received phishing e-mail attacks in 2006, up from 57 million U.S. adults in 2004. The average loss per victim has grown from $257 to $1,244 per victim in 2006 [21]. In 2007, 3.6 million adults lost $3.2 billion due to phishing attacks [22]. Report from Gartner states that the number of phishing incidents rose 39.8% with an average loss per incident of $351, over 5 million people affected in US in 2008 [23]. Report from Anti-Phishing Working Group also shows that phishing attack remains a serious problem far from solved. Table1 [19, 20] below show the trend of phishing is still growing.

| | 2008 (Jan-Jun) | 2008 (Jul-Dec) | 2009 (Jan-Jun) | 2009 (Jul-Dec) | 2010 (Jan-Jun) | 2010 (Jul-Dec) |
|---|---|---|---|---|---|---|
| Phishing Domain Names | 26678 | 30454 | 30131 | 28775 | 28646 | 42624 |
| Phishing Attacks | 47324 | 56959 | 55698 | 126697 | 48244 | 67677 |

**Table 1**

Many related works have been conducted, trying to detect and prevent phishing. Researchers have used different features extracted from web pages, URLs, and visual images to detect phishing websites. Apparently, not all methods are working very well since phishing is still growing strongly. We want to evaluate various features of phishing webpages to explore which features are more important and efficient.

In this project, we propose an approach to extract, analyze and evaluate features of phishing webpages from three perspectives: 1) Web Content, 2) Network Features, and 3) Visual Similarity.

## 2. Background

### 2.1 Definition of Phishing

Our work focuses on phishing webpages. So first, we would like to try to describe what falls into the category of phishing.

PhishTank defines phishing as "a fraudulent attempt, usually made through email, to steal ... personal information" [2]. More generally, Google defines a phishing page as any web page that, without permission, alleges to act on behalf of a third party with the intention of confusing viewers into performing an action with which the viewer would only trust a true agent of the third party [1]. Note that those actions include, but are not limited to, submitting personal information to the page. In a sense, this definition of phishing is closer to "web forgery". In our project, we adopt Google's broader definition of "phishing". This definition certainly covers the typical case of phishing pages displaying graphics relating to a financial company and requesting a viewer's login credentials. This definition also covers phishing pages, which display a trusted company's logos and request that the viewer download and execute an unknown binary [1]. Note that if one of these sites is sanctioned by the third party, then it would be properly authorized and therefore not a phishing page.

### 2.2 Phishing Attack Model

Anirudh Ramachandran, Nick Feamster, Balachander Krishnamurthy, Oliver Spatscheck, and Jacobus Van der Merwe described a basic phishing attack model and enumerate the possible data sources that could be used to detect phishing in the network. Based on this model, network operators can develop techniques for detecting phishing attacks from various types of network data available to them [3].

Figure 1 shows the sequence of steps involved in a phishing attack. There are six typical steps:

**Step 1: Domain registration.** Domain registration zone file logs can expose suspicious patterns of registering and unregistering domains. Moreover, domain names can be tested for similarity to phishing domain names using regular expressions. Recent measurement studies have indicated that it is indeed possible to discover phishing domain registrations and de-registrations using this source [3][12].

**Step 2: DNS update for phishing domain.** Recent reports indicate that phishing sites exhibit "fast flux" behavior, where the mappings of both names to IP addresses and names to authoritative nameservers are continually changing. Monitoring DNS records for a domain over time could also allow a detection system to determine whether a particular domain was likely being used for a phishing attack. A detection algorithm could also query the DNS to look for authoritative nameservers hosted on broadband or dialup machines (or on other short-lived connections) [3].

**Step 3: Phishing email.** The next step in the phishing process is "fishing" for vulnerable users by sending out a phishing email (or messages on web-based message boards and instant messages [13][14]), with a URL that directs users to the newly registered domain. Emails that contain phishing URLs could be detected either at the mail server or at an appliance that examines network traffic. An advantage to detecting phishing attempts at this stage is that phishing URLs in emails can be proactively detected, before the attack even reaches a user's inbox [3].

**Step 4: DNS lookup for the phishing domain.** After a phishing email reaches a user's inbox, the user may click on a URL contained in that email. If the URL involves a DNS lookup, monitoring these lookups provides another opportunity to detect phishing attacks. At this stage, the monitor could then look for evidence of fast-flux behavior [15], including low time-to-live (TTL) values for DNS A or NS records, or changes to the records themselves

[16].

**Step 5: HTTP request for phishing Web site.** A detection mechanism could examine several features that might indicate that a client was making a request to a suspect server. First, recent studies suggest that HTTP scam and phishing sites are typically short-lived [16, 17, 18], which suggests that HTTP requests to Web servers with short uptimes may be cause for suspicion. Second, the HTTP requests themselves may provide an indication that a user is making a request for a URL that is a likely phishing attack, because many phishing URLs disguise the path (i.e., GET) portion of the URL to be similar to a target organization's URL. Third, looking for HTTP requests to servers that are on IP blacklists, known bots or compromised hosts, or otherwise "suspect" IP ad- dresses might also prove to be helpful indicators. Finally, request patterns such as the presence of HTTP redirects may indicate an attempt to send a client to an ephemeral site (e.g., a phishing site hosted by a botnet) [3].

**Step 6: HTTP response (including content).** Once the request is issued and the client follows a sequence of redirects, a Web server returns the content to the client. This content is also visible in the network stream. In some cases, sophisticated analysis may be able to reconstruct the returned content and compare it to that of a legitimate Web site. An alternative method would be to look at how the page is constructed: for example, a phishing site might construct a page by assembling remotely hosted content (possibly even from the legitimate site itself) [3].

Each step has possible opportunity of detection and many related works have been conducted toward different steps. The work in our project tried to leverage information from step1, step 5 and step 6.

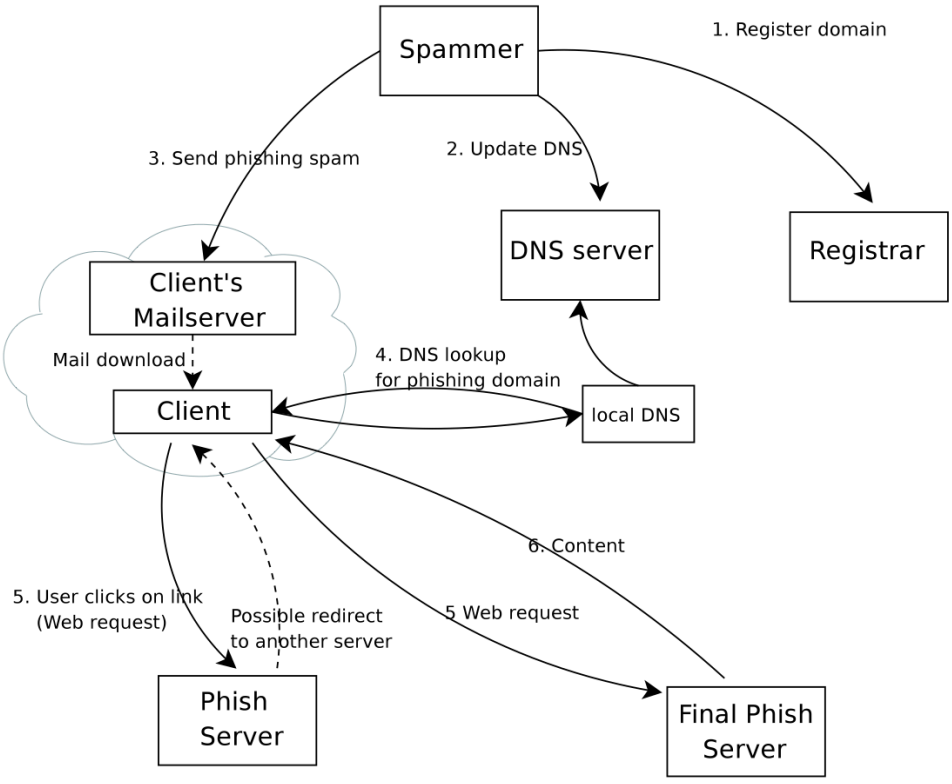


**Figure 1: Model of a phishing attack. Each step has a possible detection opportunity.**

## 2.3 Related Works

Our work focused on analyzing features of phishing websites from three perspectives, web content, network features and visual similarity. There are many useful related works that shed lights and provide guidance for us. We also modify and utilize some existing techniques from these works.

### 2.3.1 Webpage-feature-based approaches

Some works have utilized features extracted from web pages as well as URLs to do analysis. In Large-Scale Automatic Classification of Phishing Pages [1], Colin Whittaker, Brian Ryner, Marria Nazif described the design and performance characteristics of a scalable machine learning classifier they developed to detect phishing websites. Their classifier analyzes millions of pages a day, examining the URL and the contents of a page to determine whether or not a page is phishing. They conduct URL feature extraction and Page feature extraction to perform a page classification. Christian Ludl, Sean McAllister, Engin Kirda, and Christopher Kruegel explored the existence of page properties that can be used to identify phishing pages by analyzing a large number of phishing pages. They studied some important features extracted from the HTML source of a page and the page's URL. Using those extracted features, they applied the J48 algorithm to extract a decision tree that can classify pages as legitimate or phishing.

### 2.3.2 Visual-similarity-based approaches

Another interesting angle towards the problem is to detect phishing web pages through visual similarity. Phishers could use images to represent the Web page content rather than HTML text. Flash, Movie, ActiveX, Java Applet, and various types of pictures can be embedded into the Web pages instead of HTML text. In other words, phishers can easily create Web pages that look exactly the same as the real Web page but use completely different background coding (Text, Flash, ActiveX, Java Applet, and Various Pictures). A real Web page can correspond to countless fake Web pages with different code. That is the major motivation for investigating the phishing detection method through visual similarity. In [6], [7], [8], the DOM-based [9] visual similarity of Web pages is oriented, and the concept of visual approach to phishing detection was first introduced. Through this approach, a phishing Web page can be detected and reported in an automatic way rather than involving too many human efforts. Their method first decomposes the Web pages (in HTML) into salient (visually distinguishable) block regions. The visual similarity between two Web pages is then evaluated in three metrics: block level similarity, layout similarity, and overall style similarity, which are based on the matching of the salient block regions. In [11], the solution proposed is to extract DOM-Tree for each web page and take a Layout-Similarity-Based approach to detect phishing pages. In [5], they used Earth Mover's Distance (EMD) to measure Web page visual similarity. They first convert the involved Web pages into low-resolution images and then use color and coordinate features to represent the image signatures. They used EMD to calculate the signature distances of the images of the Web pages and trained an EMD threshold vector for classifying a Web page as a phishing or a legitimate one. Eric Medvet, Engin Kirda, and Christopher Kruegel proposed a technique to visually compare a suspected phishing page with the legitimate one. They identify and consider three page features that play a key role in making a phishing page look similar to a legitimate one. These features are text pieces and their style, images embedded in the page, and the overall visual appearance of the page as rendered by the browser [10].

# 3. Design and Implementation

## 3.1 Design

### 3.1.1 Overall System Design

Our project is mainly focused on measurement and analysis. Our approach is to extract, analyze and evaluate features from three perspectives: 1) Web Content, 2) Network features, and 3) Visual Similarity.

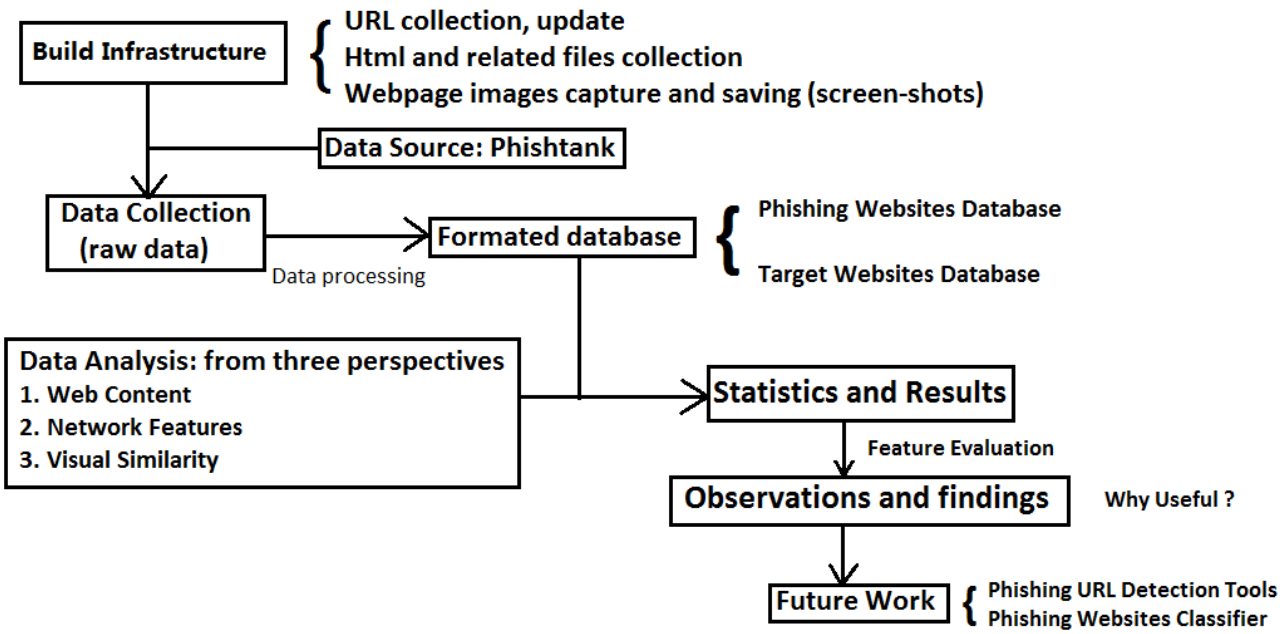


**Figure 2: Frame of Our Project**

Figure 2 shows the frame of our project. Our work consists of two main parts: Data Collection and Data Analysis.

Firstly, we build up our infrastructure to collect raw data of phishing and target websites. Then we transform the raw data into formatted database through data processing. Next, using the features predefined, we extract these features from the database and do some analysis. Based on the statistic results we get, some observations and conclusions can be made which will facilitate future work and further investigation.

### 3.1.2 Building up the Infrastructure for Data Collection

A main part of our work in this project is to build up the infrastructure for data collection. The infrastructure design is based on the goals of our project. Our goal is to extract and analyze features from three perspectives. So we need the infrastructure to collect and maintain relevant data. The structure of our infrastructure is shown in Figure 3.

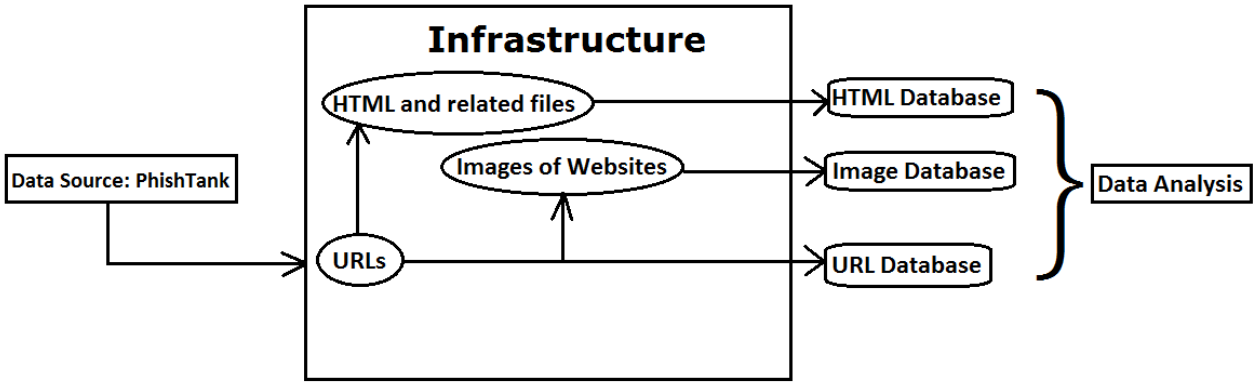


**Figure 3: Infrastructure for Data Collection**

The first step is to collect and maintain a URL database. Our infrastructure downloads URL list from PhishTank. PhishTank updates its phishing URL list every hour and the infrastructure we build can update the URL database as well. Once obtaining the Phishing URL list, the infrastructure will try to fetch HTML and other related files (embedded images) to construct a HTML database for later analysis. Meanwhile, the infrastructure will also try to capture screen-shot images of the phishing websites according to the URLs and construct an Image database. These three databases are essential for our data analysis and are well maintained by our infrastructure.

### 3.1.3 Defining Features for Data Analysis

Our approach is to extract, analyze and evaluate features from three perspectives: 1) Web Content, 2) Network Features, and 3) Visual Similarity. Here, we define and explain the features we used in our project.

#### 1) Web Content Features

**a. Links:** An important, general characteristic of every web page is its link structure. This not only takes into account links to other web pages, but also includes links to embedded images. Interestingly, many phishing pages contain links to the site they spoof, often to include original page elements from the victim page. To recognize such pages, we include properties that count the number of internal links to resources located in the page's domain as well as external links to resources stored on other sites. These links are extracted from a page by looking for <a> tags in the HTML source. By scanning for <img> tags, we extract links to internal images and external images. To underline the important difference of external links and internal links for detecting phishing pages, we count the total number of external links and internal links of each page separately.

**b. Script tags:** To distinguish between websites that make ample use of JavaScript and plain text pages, we count the number of JavaScript tags on a page and store it in the script tags property.

**c. CSS tags:** To distinguish websites that make ample use of CSS to manage its page structure, we count the number of CSS tags on a page and store it in the CSS tags property.

**d. Login Section:** Since the purpose of phishing is to try to steal sensitive information from the victims, phishing websites are very likely to contain login sections. We count the

number of login sections in each webpage, using character string "log", "sign", and "register".

**e. Suspicious Title:** In order to deceive victims, many phishing websites may contain character strings related to target websites in their titles. We first define some character strings for each target websites and then count the number of character strings appearing in the phishing webpage title.

**f. Suspicious URL:** Similar to suspicious Title, many phishing websites may contain character strings related to target websites in their URLs. We first define some character strings for each target websites and then count the number of character strings appearing in the phishing URL.

**g. URL Length:** From the URL list we obtained, we observe that many phishing URLs have peculiar structures and could be very long and obscure. To do further investigation, we compute and record the length of each given URL.

**h. Using SSL or not:** Another characteristic that was analyzed for each page is whether it is accessed over SSL (https) or not. In our preliminary studies, we observed that not many phishing sites make use of a secure server. One explanation could be that it is not straightforward to obtain a trustworthy certificate. Hence, not having such a certificate could indicate the potential of being a phishing website. We check for each given URL if the connection is using SSL.

#### 2) Network Features

**a. Domain name:** Can be used to query Whois Domain Registration data.

**b. Host Name**

**c. Host IP**

**d. Geographic Location:** Country

**e. ASN**

**f. Whois Domain Registration Information:**

creation_date, expiration_date, updated_date, domain_name, emails, name_servers, referral_url, registrar, status, and whois_server

#### 3) Visual Similarity Feature

For visual similarity, we only have one feature which is the overall visual similarity score. This score is ranged from 0 to 1. Score 0 means totally different and Score 1 means identical. The detail of how to compute this visual similarity score is shown in section 3.2.

### 3.2 Implementation

We implemented our system using Python. Our system consists of different modules, which can be viewed as independent APIs.

We use BeautifulSoup, an open source Python library to parse the HTML file.

We use Pywhois, an open source Python library to get the Whois Domain Registration data from Whois servers.

For visual similarity, we use an algorithm to compute the overall visual similarity. The input of the algorithm are two images of websites, the output will be a visual similarity score. The algorithm we adopt consists of three steps: (Shown in Figure 4)

1) Resize the images size into 1024*1024. Since the input images may have different sizes, it is necessary to resize them into uniform size before comparison.
2) Cut each image into blocks with size 128*128. Use color histogram to compute the distance between corresponding blocks of the two images.
3) Compute the average distance of all block pairs between the two images to generate the overall visual similarity score.

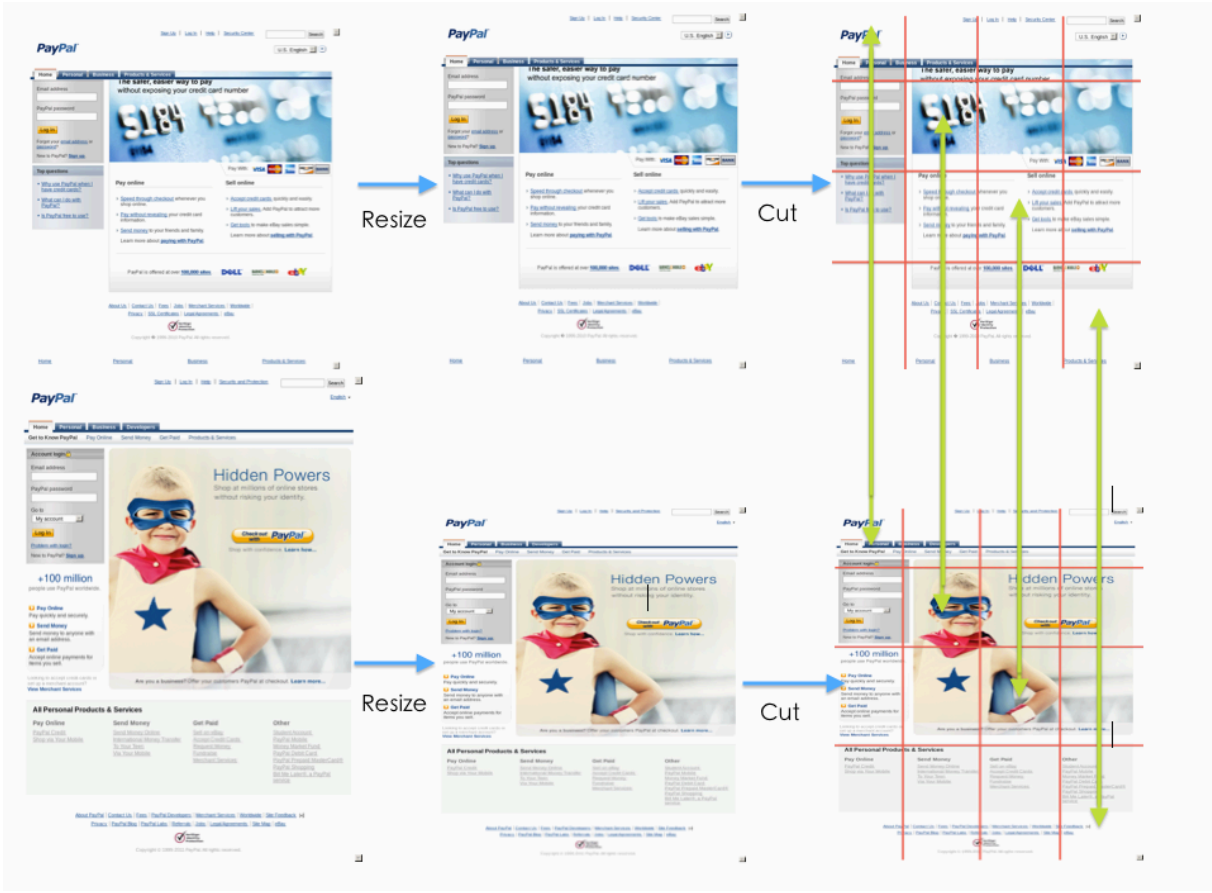


**Figure 4: Visual Similarity Algorithm**

## 4. Experiments and Results

In this section, we introduce the experiments we conducted in our project. And we show our experimental results and discuss some interesting observations we make.

### 4.1 Data Set

Our URL data source is PhishTank. We have collected data from Nov 28th, 2011 to Dec 09th, 2011. Totally 9157 phishing URLs are collected. Totally 6258 valid webpages are collected. We use these collected data to conduct our experiments and get relevant results.

### 4.2 Methodology

To evaluate the features we extracted, we conduct two kinds of comparison analysis.

Our data set could be classified into two big categories. One is about phishing websites, the other is about legitimate target websites. First, we compare different features across these two categories, which could give us information about the overall characteristics of malicious websites and benign websites. Secondly, phishing websites could be classified into different groups based on their target websites. Then we have groups targeting specific targets. Each group of phishing websites may have distinct characteristics related with their target website. We conduct comparison analysis across different groups to explore the group characteristics. We select phishing groups targeting PayPal, Facebook, and Orkut to study. Group characteristics could be interesting and useful because they could: 1) characterize target websites and provide information about how to improve the security of target websites. 2) help determine the target of given phishing websites 3) help classify phishing websites.

In the next section, we will show the results and some observations from the comparison experiments.

## 4.3 Results and Observations

### 4.3.1 Most Popular Targets

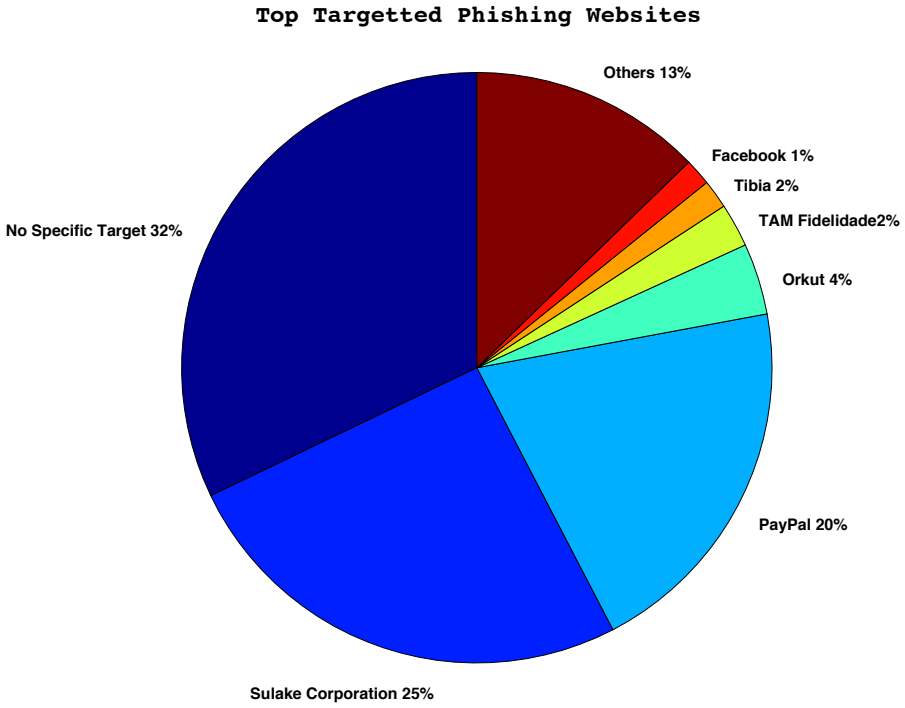


**Figure 5: Most Popular Targets**

Figure 5 shows the most popular targets from our data set. One interesting finding is that among all phishing webpages, about one third of them do not have specific target, which means these phishing webpages do not try to imitate specific legitimate websites. We find that those phishing webpages are mostly spam webpages disguising themselves as advertisements or surveys. This group of phishing webpages has distinct characteristics from other groups because their purpose is not to imitate specific legitimate target websites.

We observed that online banks, online transaction websites and social networks dominate the target list.

### 4.3.2 Web Content Features

Table 2 shows the results of analyzing web content features. Several observations could be made from the results.

**Observation 1:** Phishing webpages tend to have more external links than internal links while legitimate webpages tend to have more internal links than external links.

The reason behind observation 1 is that phishing websites usually contain external links to the target websites, which could make the phishing websites look more similar to the target websites and could potentially draw more victims.

**Observation 2:** Login sections and character strings have different impacts towards different target groups. These features could be utilized to characterize different phishing webpage groups.

**Observation 3:** Malicious phishing websites tend to have much longer URLs than legitimate websites, as shown in Figure 6, 7, and 8. This could be a very effective feature for phishing detection.

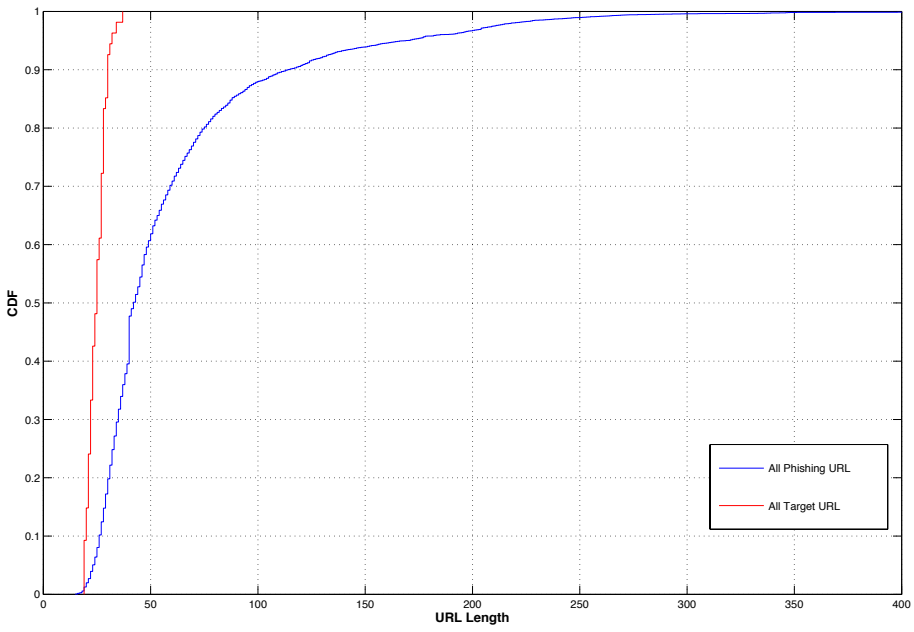


**Figure 6: URL Length for All Phishing Websites**

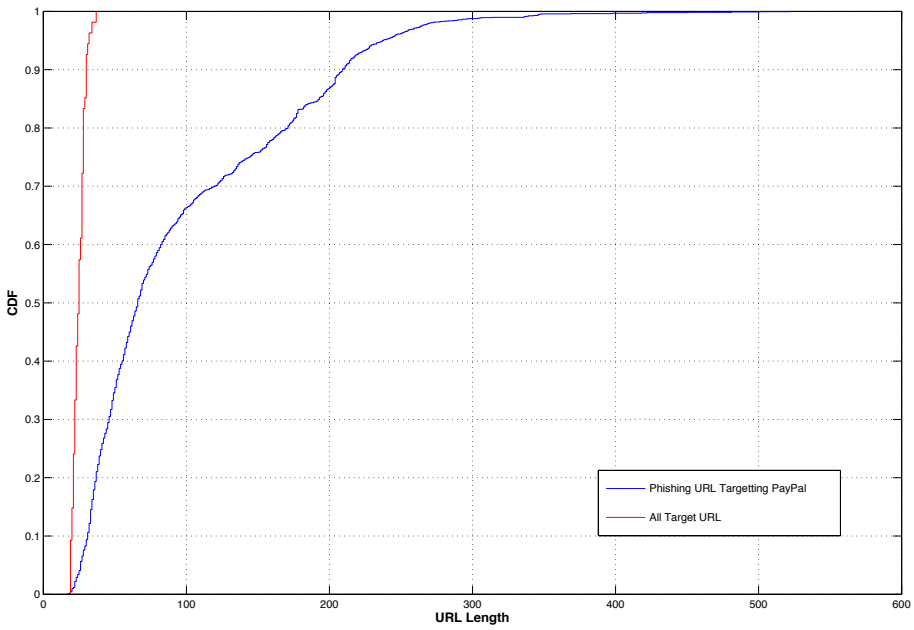


**Figure 7: URL Length for Phishing Websites Targeting PayPal**

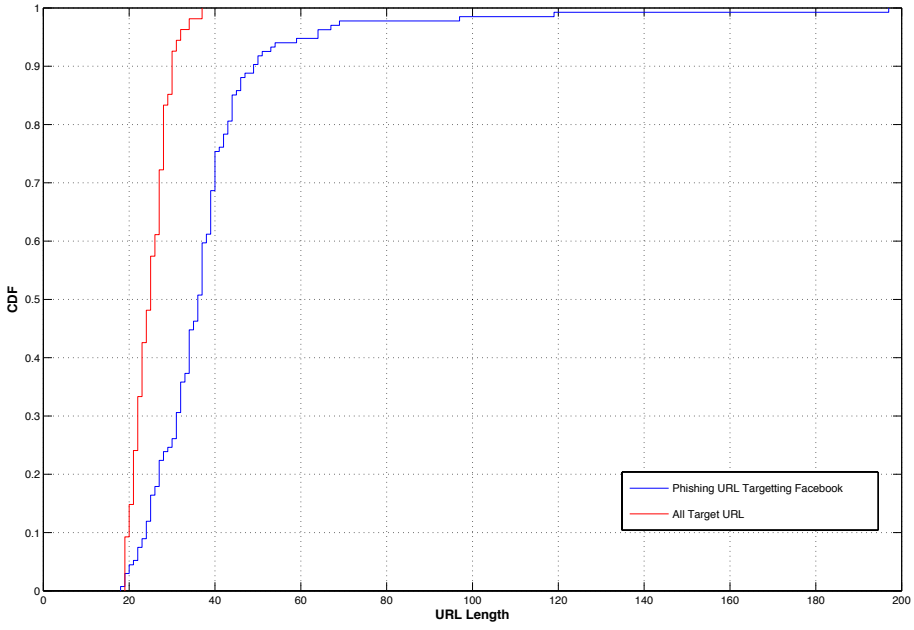


**Figure 8: URL Length for Phishing Websites Targeting Facebook**

**Observation 4:** Phishing websites rarely use SSL while legitimate websites related with online banking, transaction and account login use SSL frequently. This is also a very useful feature to detect suspicious and potentially malicious websites.

| | All Phishing Sites | All Target Sites | PayPal | Facebook | Orkut |
|---|---|---|---|---|---|
| Average external link number | 18 | 59 | 36 | 20 | 2 |
| Average internal link number | 12 | 77 | 12 | 13 | 0 |
| Average external javascript Number | 1 | 2 | 3 | 0 | 0 |
| Average external css link number | 2 | 2 | 3 | 1 | 0 |
| Webpages containing login | 1154/6258 (18.4%) | 9/50 (18%) | 245/366 (66.9%) | 13/25 (52%) | 6/243 (2.5%) |
| Webpages containing target title | 1005/6258 (16.1%) | 9/50 (18%) | 297/366 (81.1%) | 21/25 (84%) | 33/243 (13.6) |
| | | | | | |
| Average url length | 60 | 24 | 97 | 38 | 34 |
| URLs containing character string | | | 619/1858 (33.3%) | 47/134 (35.1) | 62/358 (17.3%) |
| Using SSL (https) | 101/8000 (1.3%) | 11/54 (20.4%) | | | |

**Table 2: Web Content Features: The first column shows web content feature statistics of all phishing webpages. The second column shows web content feature statistics of all target webpages. The third, fourth and fifth column show web content feature statistics of phishing webpages targeting PayPal, Facebook, and Orkut, respectively.**

### 4.3.3 Network Features

Two main network features we evaluate are country location and ASN.

From Figure 9 and 10, it seems difficult to draw any concrete conclusion from these two features. However, through further analysis and investigation, we find that these network features have potential benefits and may provide help for phishing detection.

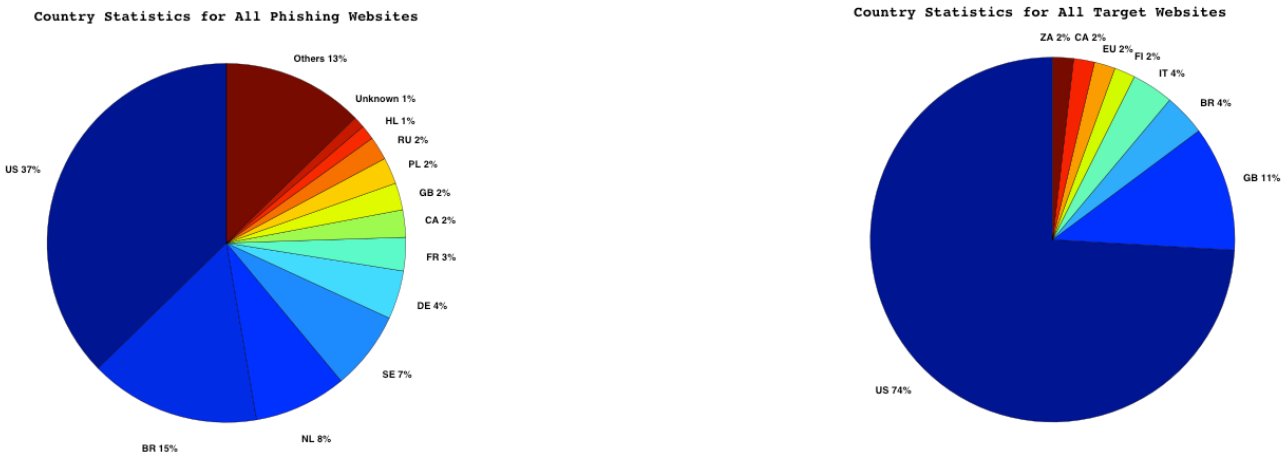


**Figure 9: Country Statistics Comparison between All Phishing Websites and All Target Websites**

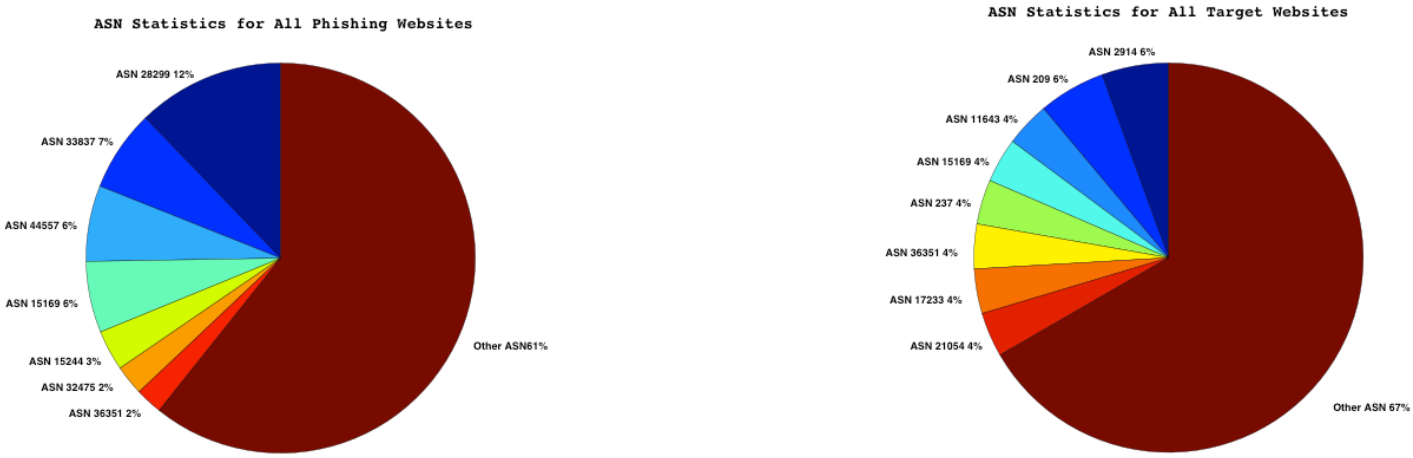


**Figure 10: ASN Statistics Comparison between All Phishing Websites and All Target Websites**

We analyzed the network features of phishing webpages targeting PayPal and Facebook. We find that there are some aggregations for both country location and ASN. And those two phishing webpage groups have different aggregations behavior, as shown in Figure 11 and 12.

**Observation 5:** Country and ASN aggregation could be used to group phishing websites. Phishing websites sharing the same country and ASN features tend to come from the same attacker or botnet machines. These features could be utilized to detect and filter phishing websites or to build blacklists. To improve accuracy, we can also use the Domain Registration Information collected. For phishing websites coming from the same country and ASN, we check if they also share the same Domain Registration registrar and emails. This allows a more accurate aggregation result.

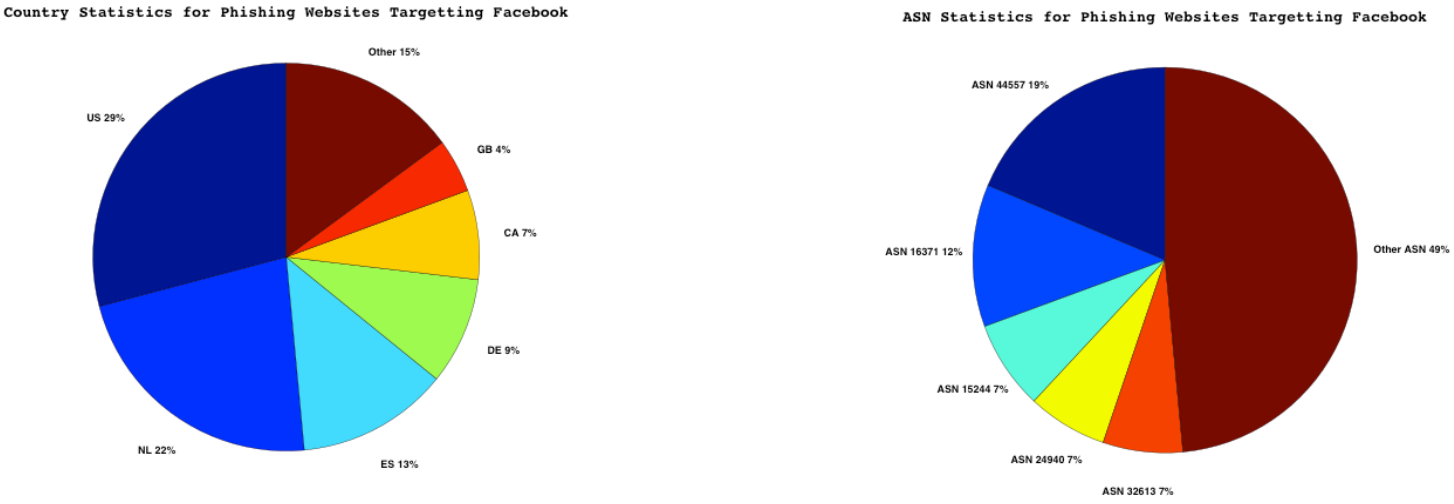


**Figure 11: Network Features Statistics for Phishing Websites Targeting Facebook**

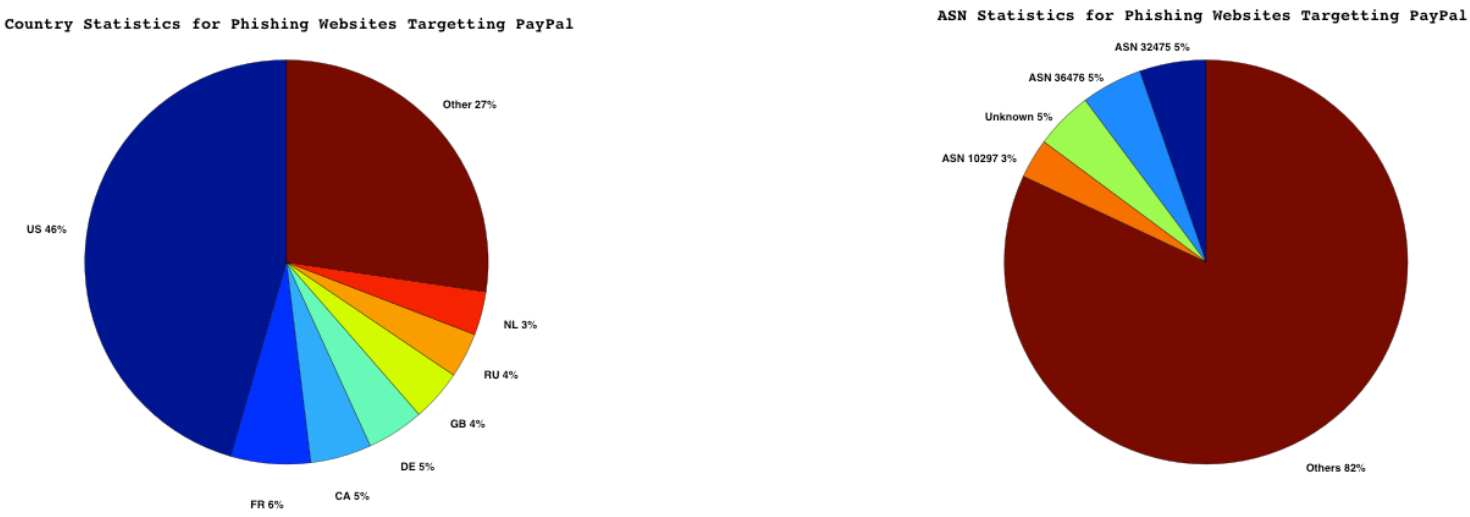


**Figure 12: Network Features Statistics for Phishing Websites Targeting PayPal**

#### 4.3.4 Visual Similarity Feature

We use our algorithm described in Section 3.2 to compute the visual similarity score between phishing websites and legitimate target websites. The results are shown in Figure 13 and 14.

**Observation 6:** Surprisingly, the overall similarity from visual sense between phishing websites and target websites are not necessarily high. This indicates that using image similarity alone to detect phishing websites may not be good enough. The reason behind this observation is that many phishing websites tend to have different images and background than the targets. But as long as they have some signature logo or title of the target websites, they are still able to deceive the victims. So a better solution is to use features extracted from the HTML file to facilitate the visual similarity algorithm.

In our experiments, the purpose is to first see what's the statistics of overall visual similarity between phishing websites and legitimate target websites. Through analysis, we propose possible reasons behind the results and solutions for improvement, as stated in Observation 6. We leave the improvement of the algorithm to future work.

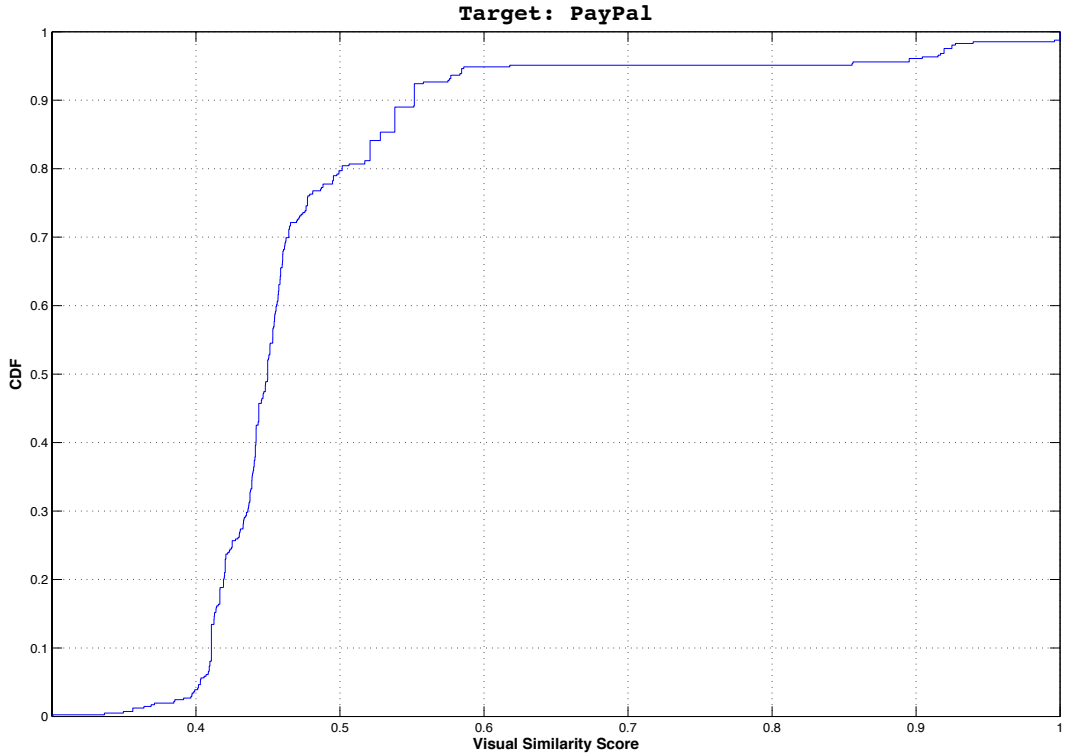


**Figure 13: Visual Similarity for Phishing Websites Targeting PayPal**

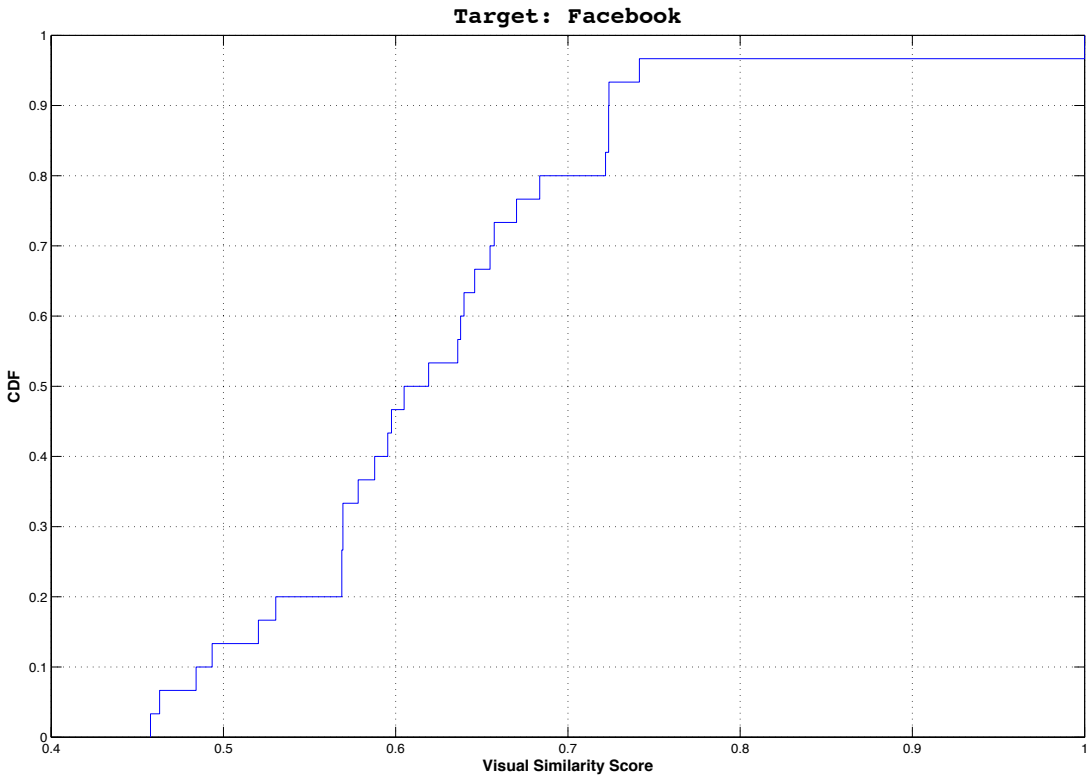


**Figure 14: Visual Similarity for Phishing Websites Targeting Facebook**

## 5. Future Work

In this project, our work is focused on extracting, analyzing and evaluating features of phishing webpages. The results and observations of our work could provide help to further investigation and construction of practical tools. Based on the evaluation results from our project, we can select effective features to build decision tree, using Machine Learning algorithm on training data. This could result in a practical phishing URL detection tool, which may be developed as a web browser plug-in. Another possible extension of our work is to build a phishing websites classifier. Using the features we extracted, we could use a training data set to generate characteristics of different phishing groups divided according to their targets.

Currently, our feature analysis and evaluation is done off-line. But the contents of both target websites and phishing websites may change from time to time. Not considering the changes and updates of websites may make the results less accurate and this is a limitation of our current work. A possible improvement is to make our system be able to cope with real-time phishing detection. There would be many challenges for real-time detection. One major challenge is how to maintain the latest database efficiently since the data will change constantly in real-time situation.

## 6. Conclusion

In this project, we focused on extracting, analyzing and evaluating features of phishing webpages. We define essential features for evaluation from three perspectives: 1) Web Content, 2) Network Features, and 3) Visual Similarity. We build up infrastructure to collect data from PhishTank. Using the predefined features, we analyzed and evaluated the data. We make several observations from the statistic results. Our results and observations could provide help for further investigation and future work.

## 7. Acknowledgements

We would like to thank Dr. Michael Bailey for discussing the topic with us.

We give special thanks to Jing Zhang, who helped us develop our ideas and provided useful guidance.

## References:


[1] Colin Whittaker, Brian Ryner, Marria Nazif. Large-Scale Automatic Classification of Phishing Pages. In *NDSS'10: Proceedings of the 17th Annual Network and Distributed System Security Symposium,* 2010.

[2] OpenDNS. Phishtank. http://www.phishtank.com/, 2009.

[3] Anirudh Ramachandran, Nick Feamster, Balachander Krishnamurthy, Oliver Spatscheck, Jacobus Van der Merwe. Fishing for Phishing from the Network Stream. Georgia Tech CSS Technical Report GT-CS-08-08.

[4] Christian Ludl, Sean McAllister, Engin Kirda, Christopher Kruegel. On the Effectiveness of Techniques to Detect Phishing Sites

[5] Anthony Y. Fu, Liu Wenyin, and Xiaotie Deng. Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover's Distance (EMD). *IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING*, VOL. 3, NO. 4, OCTOBER-DECEMBER 2006.

[6] W. Liu, X. Deng, G. Huang, and A.Y. Fu, "An Anti-Phishing Strategy Based on Visual Similarity Assessment," *IEEE Internet Computing*, vol. 10, no. 2, pp. 58-65, 2006.

[7] W. Liu, G. Huang, X. Liu, M. Zhang, and X. Deng, "Detection of Phishing Web Pages Based on Visual Similarity," *Proc. 14th Int'l World Wide Web Conf.*, pp. 1060-1061, 2005.

[8] W. Liu, G. Huang, X. Liu, M. Zhang, and X. Deng, "Phishing Web Page Detection," *Proc. Eighth Int'l Conf. Documents Analysis and Recognition*, pp. 560-564, 2005.

[9] L. Wood, Document Object Model Level 1 Specification, http:// www.w3.org, 2005.

[10] Eric Medvet, Engin Kirda, and Christopher Kruegel. Visual-Similarity-Based Phishing Detection.

[11] Angelo P. E. Rosiello, Engin Kirda, Christopher Kruegel, and Fabrizio Ferrandi. A Layout-Similarity-Based Approach for Detecting Phishing Pages.

[12] D. K. McGrath and M. Gupta. Behind Phishing: An Examination of Phisher Modi Operandi. In *First Usenix Workshop On Large-Scale Exploits and Emergent Threats (LEET)*, 2008.

[13] eweek.com. Phishing Dips into Yahoo IM. http://www.eweek. com/article2/0,1759,1779798,00.asp, 2005.

[14] PCWorld. Phishing Scam Takes Aim at Myspace.com. http://www.pcworld.com/article/id,125956-page,1/article.html?RSS=RSS, 2006.

[15] Honeynet Project & Research Alliance. Know Your Enemy: Fast-Flux Service Networks. http: //www.honeynet.org/papers/ff/fast-flux.html, 2007.

[16] T.Moore and R.Clayton. Examining the Impact of Website Take-down on Phishing. In *APWG eCrime Researchers Summit*, Pittsburgh, PA, 2007.

[17] D. S. Anderson, C. Fleizach, S. Savage, and G. M. Voelker. Spamscatter: Characterizing Internet Scam Hosting Infrastructure. In *Proc. 16th USENIX Security Symposium*, Boston, MA, Aug. 2007.

[18] Anti Phishing Working Group. Global Phishing Survey: Trends and Domain Name Use in 1H2011. http://www.antiphishing.org/reports/APWG_GlobalPhishingSurvey_1H2011.pdf, Nov. 2011.

[19] Anti Phishing Working Group. Global Phishing Survey: Trends and Domain Name Use in 1H2010, 2010.

[20] Anti Phishing Working Group. Global Phishing Survey: Trends and Domain Name Use in 2H2010, 2010.

[21] Gartner Inc. Gartner Says Number of Phishing E-Mails Sent to U.S. Adults Nearly Doubles in Just Two Years. http://www.gartner.com/it/page.jsp?id=498245

[22] Gartner Inc. Gartner puts phishing tab at $3.2 billion. http://www.zdnet.com/blog/security/gartner-puts-phishing- tab-at-32-billion/755

[23] Gartner Inc. Gartner Says Number of Phishing Attacks on U.S. Consumers Increased 40 Percent in 2008. http://www.smartbrief.com/news/aaaa/industryBW-detail.jsp?id=C87DC2AF-1BA9-4A16-9717-A07ACBB7BCDF