# An Effective Approach for Prevention of Phishing Attack

**S. Thiruvenkatasamy** [1], **C. Sathyapriya** [2]

[1] Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, othakuthirai, Gobi-638408, India
Email: samysdotcom@yahoo.com

[2] Department of IT, Shree Venkateshwara Hi-Tech Engineering College, othakuthirai, Gobi-638408, India
Email: c.sathyapriya@gmail.com

***Abstract-*** An effective approach for phishing Web page detection is proposed, which uses Earth Mover's Distance (EMD) to measure Web page visual similarity, and gives digital water marking approach for more authentication purpose. First convert the involved Web pages into low resolution images and then use color and coordinate features to represent the image signatures. This paper proposes to use EMD to calculate the signature distances of the images of the Web pages. To train an EMD threshold vector for classifying a Web page as a phishing or a normal one. Digital watermarking approach is used to protect the original Web pages.

***Index Terms***- phishing**,** anti-phishing, visual assessment, earth mover's distance

## I. INTRODUCTION

Phishing Web pages are forged Web pages that are created by malicious people to mimic Web pages of real Web sites. Most of these kinds of Web pages have high visual similarities to scam their victims. Some of these kinds of Web pages look exactly like the real ones. Unwary Internet users may be easily deceived by this kind of scam. Victims of phishing Web pages may expose their bank account, password, Credit card number or other important information to the phishing web page owners.

Phishing is a relatively new Internet crime in comparison with other forms, e.g., virus and hacking. More and more phishing Web pages have been found in recent years in an accelerative way. A report from the Anti-Phishing Working Group [10] shows that the number of phishing Web pages is increasing each month by 50 percent and usually 5 percent of the phishing e-mail receivers will respond to the scams. Also, there were 15,050 phishing cases reported simply in one month in June 2005 [10]. This problem has drawn high attention from both Industry and the academic research domain since it is a severe security and privacy problem and has caused huge negative impacts on the Internet World. It is threatening people's confidence to use the Web to conduct online finance-related activities.

In this paper, we propose an effective approach for detecting phishing Web pages, which employs the Earth Mover's Distance (EMD) [6] to calculate the visual similarity of Web pages. The most important reason that Internet users could become phishing victims is that phishing Web pages always have high visual similarity with the real Web pages, such as visually similar block layouts, dominant colors, images, and fonts, etc.

We follow the antiphishing strategy to detect phishing and providing more security to the original web pages.

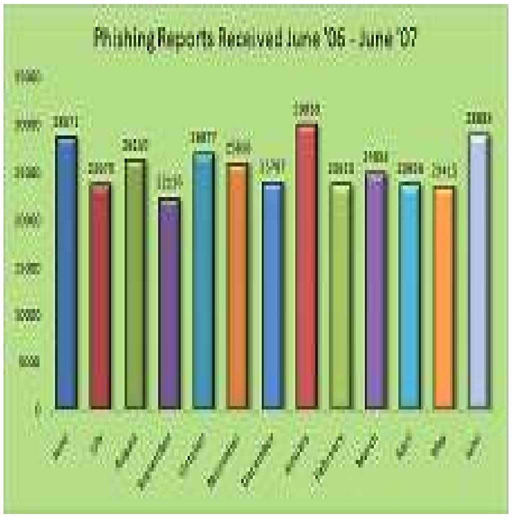

Figure 1: Phished Report from APWG [10]

## II. BACKGROUND

In [2] M.Wu described about Web Wallet which was a browser sidebar. In this, users have to submit their sensitive information online. It detects phishing attacks. It suggests an alternative safe path. It asks security questions to user's workflow so it cannot be ignored by the user.

In [4] R. Dhamija described about interaction between user and browser. Here personal image is selected by user. Information is maintained only by browser. Abstract image is generated by server. Visual hash is generated between browser and server. Browser has to verify and authenticate it. In [3] W. Liu described about the basic methodology of visual similarity assessment. It deals with only visual calculations.

In [5] Y. Chen proposed about resized images. Mobile devices have already been widely used to access the Web. However, because most available web pages are designed for desktop PC in mind, it is inconvenient to browse these large web pages on a mobile device with a small screen. In this paper, they propose a new browsing convention to facilitate navigation and reading on a small-form-factor device. In [6] Grauman proposed EMD which deals matching between two shapes and their features which reveals how similar the shapes are.

## III. PROPOSED SYSTEM

It works well on visual similarities. Digital watermark approach which is embedded in an image of a Web site which is used to protect against phishing attacks. The hidden watermark could be used to identify a legitimate Web site, distinguishing it from a bogus phishing site used for stealing credentials.

In this proposed system first retrieve the suspected Web pages and protected Web pages from the Web and generate their signatures. The task of our Web page preprocessing approach contains three procedures: 1) obtain the image of a Web page from its URL, 2) perform normalization, and 3) represent the Web page image into a Web page visual signature (consists of color and coordinate features), which is used to evaluate the visual similarity of a pair of Web pages. The process of displaying a Web page in a Web browser on the screen from HTML and accessory files (including pictures, Flash movies, ActiveX plug-ins, Java Applets, etc.) is the Web page rendering process. We use GDI (graphic device interface) API provided by the Microsoft IE browser to get Web page images (in jpeg format). The images of the original sizes are processed into images with normalized size (e.g., 100 * 100). The Lanczos algorithm [10] is used to calculate the resized image because the Lanczos algorithm has very strong antialiasing properties in Fourier domain, and it is also easy to be computed in spatial domain. Moreover, sharp images can be generated with the Lanczos algorithm as intuitively, the sharp images could provide better signature for identification from the others. www.bbb.org is an example of a square-like image, www.banktechnews.com is an example of a longer image, and www.bankofcyprus.com is an example of a wider image.

## IV. METHODOLOGY

### *A. EMD (Earth Mover's Distance) Method*

The minimum cost of matching features from one shape to the features of another often reveals how similar the two shapes are. The cost of matching two features may be defined as how dissimilar they are in spatial location, appearance, curvature, or orientation; the minimal weight matching is the correspondence field between the two sets of features that requires the least summed cost. The proposed system present a contour matching algorithm that quickly computes the minimum weight matching between sets of descriptive local features using a recently introduced low-distortion embedding of the Earth Mover's Distance (EMD)[6] into a normed space. Our method EMD is to calculate the similarity of two Web pages based on their signatures as follows:

Step 1: The distance matrix

$D = [dij]$ $(1 \leq i \leq m, 1 \leq j \leq n)$

First calculate the normalized Euclidian distance of the degraded ARGB colors, and then calculate the normalized Euclidian distance of centroids. The two distances are added up with weights p and q, respectively, to form the feature distance, where $p + q = 1$.

Step 2: The color distance

Feature $\varphi i = <dci, C_{dci}>$, where $dci = < dA_i, dR_i, dG_i, dB_i >$.

Feature $\varphi j = <dcj, C_{dcj}>$, where $dcj = < dA_j, dR_j, dG_j, dB_j >$.

Step 3: The maximum color distance

$MD_{color} = \| <MaxA - 0, MaxR - 0, MaxG - 0, MaxB - 0> \|$

Where MaxA, MaxR, MaxG, MaxB are the maximum numbers of the four components of ARGB.

Step 4: Normalized color distance

NDcolor is defined as

$ND_{color} (dc_{i,} dc_{,j}) = \sqrt{}(dci - dcj) \times (dci–dcj)^T / MD_{color.}$

is article guides a stepwise walkthrough by Experts for writing a successful journal or a research paper starting from i

## V. DESIGN OF THE PROPOSED SYSTEM

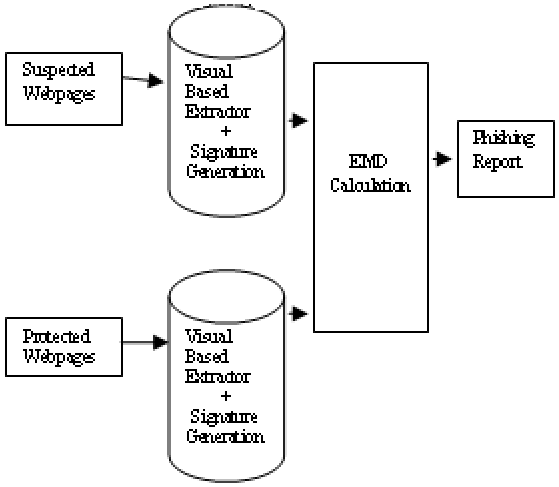
Suspected Webpages
Visual Based Extractor + Signature Generation
EMD Calculation
Phishing Report
Protected Webpages
Visual Based Extractor + Signature Generation

**Figure 2: Design Diagram**

1. Store the original web ages and perform web page preprocessing.
2. Using Digital water marking which gives more security to original web pages and their database.
3. Store the phished web pages and perform web page preprocessing.
4. Generate the signatures for both web pages.
5. Using EMD method Calculate visual similarity between the pages.
6. The difference value is finally compared with threshold value.
7. If the value is greater than threshold then report the page is phished one.

### *A. Similarity assessment*

The Visual Similarity Assessment Module measures the visual similarity between two web pages in three aspects: block level similarity, layout similarity, and overall style similarity. All these three aspects are defined based on web page segmentation. Next, our method defines three similarity metrics in the following three subsections respectively.

### *B. Block Level Similarity*

The block level similarity [3] measures the visual similarity of two pages at the level of individual blocks. It is defined as the weighted average of the visual similarities of all matched block pairs between two pages. Basically, the content of a block can be categorized as either text or image. We use different features to represent text blocks and image blocks. The features for text blocks include colors, border style and alignment, etc., and the features for image blocks include alternative text, dominant color, and image size, etc. We first calculate their similarity in terms of each feature in the feature set and then use a weighted sum of the individual feature similarities as the total similarity of the two blocks. The weight of each feature means its importance to the total similarity and can be assigned empirically. In our implementation, we focus more on color related features. Two blocks are considered as matched if their similarity is higher than a threshold. After this method obtain the similarity values of all pairs of possible matching blocks, our method finds a matching scheme between the two web pages blocks. This is actually a bipartite graph matching problem and a globally optimal solution can be obtained.

### *C. Layout Similarity*

Usually, it takes many efforts to make a brand new web page mimicking a true web page. A convenient way is to copy the source file of the true one and modify it a little bit for this purpose. In this case, the main web page structure is kept. Hence, our method defines the layout similarity as the ratio of the weighted number of matched blocks to the total number of blocks in the true web page. We employ the method in [3] for layout matching. Two blocks are considered matched if they both exhibit high visual similarity and satisfy the same constraints with corresponding matched blocks. We then define the layout similarity of two web pages as the ratio of the weighted number of matched blocks to the total number of blocks in the true web page, and the weight of each block is assigned differently according to its importance to the whole web page.

### *D. Overall Style Similarity*

In addition to the web page content, the style consistency is another important feature which can easily cheat the victims' eyes. Generally, all web pages owned by one company would keep the style consistent. The overall style similarity focuses on the visual style of a web page, which can be represented by several format definitions, e.g., the font family, background color, text alignment, and line spacing. We first obtain the histogram of the style feature values for each web page.

## VI. IMPLEMENTATION OUTLINE

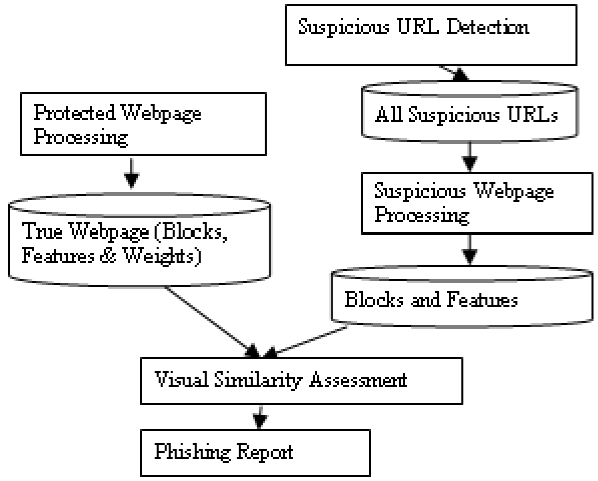
Suspicious URL Detection
Protected Webpage Processing
All Suspicious URLs
Suspicious Webpage Processing
True Webpage (Blocks, Features & Weights)
Blocks and Features
Visual Similarity Assessment
Phishing Report

Figure 3 Implementation Diagram

The proposed system has been implemented in Java Swing. Here Web capture is used for capturing the online web pages. Image Magick is used for compressing the web page images. The first module is focusing on preprocessing process. In the second module is for EMD calculation and the protection of original web pages. In the last module which will be focusing on efficiency of the method and digital water marking approach.

## VII. CONCLUSION

Phishing has becoming a serious network security problem, causing finical lose of billions of dollars to both consumers and e-commerce companies. Our proposed method of EMD to calculate visual similarity between phished and original web pages. The implementation of the proposed system is in progress and it is expected that the implementation results will detect phished web pages and gives more security to the original web pages using digital watermarking approach.

## REFERENCES

[1] M. Wu, R.C. Miller, and G. Little, "Web Wallet: Preventing Phishing Attacks by Revealing User Intentions," Proc. Symp. Usable Privacy and Security, 2006.

[2] 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.

[3] R. Dhamija and J.D. Tygar, "The Battle against Phishing: Dynamic Security Skins," Proc. Symp. Usable Privacy and Security, 2005.

[4] 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.

[5] K. Grauman and T. Darrell, "Fast Contour Matching Using Approximate Earth Mover's Distance," Proc. 2004 IEEE CS Conf. Computer Vision and attern Recognition, vol. 1, pp. 220-227, 2004.

[6] A.Y. Fu, X. Deng, and W. Liu, "A Potential IRI Based Phishing Strategy," Proc. Sixth Int'l Conf. Web Information Systems Eng. (WISE '05), pp. 618-619, Nov. 2005.

[7] Y. Chen, W.Y. Ma, and H.J. Zhang, "Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices," Proc. 12th Int'l Conf. World Wide Web, pp. 225-233, 2003.

**First Author** – S.Thiruvenkatasamy, received his Post Graduate Degree in Master of Engineering in Computer Science, from Karpagam University, Coimbatore, Tamil Nadu, India. Currently he is working as Assistant Professor in the Department of Computer Science and Engineering in Shree Venkateshwara Hi-Tech Engineering College, Gobichettipalayam, Tamil Nadu, India. He published a book "An Excellent Guide for Visual C#.Net" and had presented 6 Papers in Various National Conferences. His interest includes Data mining, computer networks and Network security. He has Published 1 paper in National Journal and 1 paper in International journal.

**Second Author** – C.Sathyapriya, received her Post Graduate Degree in Master of Engineering in Computer Science, from Velalar College Of Engineering and Technology, Erode, Tamil Nadu, India. Currently she is working as Assistant Professor in the Department of Information Technology in Shree Venkateshwara Hi-Tech Engineering College, Gobichettipalayam, Tamil Nadu, India. She had presented 9 papers in Various National Conference and also she presented 1 International Conference on "A Test Generation Method to Find Errors in HTML Language" in VIT University, Vellore, Tamil Nadu, India on 21st April 2011. Her interest includes Software Engineering, Computer Networks and Data Mining. She has Published 1 paper in National Journal and 1 paper in International journal.