Designing A Filter System For Web Security

George Eskender Ejaam
Department of Computer Science, Babylon University

Abstract

This approach designed for checking the fake websites as an attempt to prevent Phishing attempts which includes criminally and fraudulently operations to acquire sensitive information, such as usernames, passwords and credit card details information.

The system will combine two techniques in order to notify Phishing one is based on the semantic web contents which might contain malicious intents (i.e. semantically checking verification of the integrity of the websites using a cryptographic one way hash function).

The second technique is supporting user reporting to the system's database to keep track of the user's experiences with fake websites.

Similar techniques still based on the manual monitoring and users reports while the proposed system implementation as plug-Ins as well as web based will assist to block any website might contain suspected contents in its behavior beside reports to ensure more security while surfing.

Key words: Web Security, Anti-Phishing, Cryptographic Hash function for Web Security.

الخلاصة

الطريقة المقترحة تستخدم لفحص المواقع الالكترونية الوهمية كمحاولة لمنع الاحتيال Phishing والذي يتضمن المحاولات الاجرامية والمخادعة للحصول على معلومات حساسة مثل اسم مستخدم ، كلمة السر الخاصة به ومعلومات بطاقات الائتمان.

النظام سيتكون من تقنيتين للابلاغ عن الاحتيال الاولى تعتمد على محتوى الموقع المعنوي والذي يتضمن نية الاذى (اي بمعنى آخر يتم فحص والتأكد من كمالية محتوى موقع الويب بإستخدام دالة تشفير ذات إتجاه واحد).

التقنية الثانية هي دعم المستخدم للابلاغ عن الحالات لقاعدة بيانات النظام لتتبع تجربة المستخدم مع المواقع الوهمية. تقنيات مشابهه لازالت تعتمد على المراقبة اليدوية وتقارير المستخدمين في حين ان النظام المقترح كتنفيذ يمكنه ان يكون كجزء من برنامج مساعد لمتصفح الويب او كتطبيق مباشر على الويب للمساعدة في منع اي موقع من الممكن ان يحتوي على سلوك مشكوك به بالاضافة الى اسلوب التقارير لضمان امنية اكثر عند التصفح.

Introduction

Phishing is a type of social engineering with the goal of obtaining personal information, credentials, credit card number, or financial data. The attackers lure, or fish, for sensitive data through various different methods (*Shon Harris 2008*) it can be carried out by email or instant messaging (*Tan, Koon 2006)*, and often directs users to enter details at a website, although phone contact has also been used (*Skoudis 2006)*.

Microsoft with its Internet Explorer (IE v.7), Firefox, Opera and many other companies started integrating their browsers' with a check mechanism to verify each site visit. The concept of their contribution is to get information from user's own experiences and manual reports then manually verifying the suspected websites.

We will use a novel approach to avoid web forgery Phishing techniques, the new use of cryptographic MD5 algorithm idea will assist to verify the pretending websites by comparing the MD-5 digest for the URL address, main website name and the type of service with the actual values of the corresponding MD5-digest to those sites within the system which will be pre-initialized within a database, this method deal with small amount of data which will be easily transferred over Internet and can detect forgery even when a slight different are made.

The reminder of this paper is organized as follows. Section 1 reviews Phishing and Anti-Phishing techniques. Section 2 describes the proposed system for preventing forgery websites. Section 3 reports results. Section 4 shows discussion. Finally section 5 presents the conclusions.

1- Phishing and Anti-Phishing techniques

Recent Phishing attempts have targeted the customers of banks and online payment services, E-mail providers supposedly from the customer services or from the web master asking for more information or just confirming previous information (*Internet irs.gov*), below we'll review the most common Phishing and Anti-Phishing current techniques:

1-1 Phishing using Link Manipulation

Most methods of Phishing use some form of technical deception designed to make a link in an email (and the spoofed website it leads to) appear to belong to the spoofed organization. Misspelled URLs or the use of sub-domains are common tricks used by phishers, such as this example URL,
http://www.yourbank.com.example.com/.

Another common trick is to make the anchor text for a link appear to be valid, when the link actually goes to the Phishers' site. An old method of spoofing used links containing the '@' symbol, originally intended as a way to include a username and password (contrary to the standard). (Berners-Lee & Tim 2008)
For example, the link
http://www.google.com@members.tripod.com/ might deceive a casual observer into believing that it will open a page on www.google.com, whereas it actually directs the browser to a page on members.tripod.com, using a username of www.google.com: the page opens normally, regardless of the username supplied. Such URLs were disabled in Internet Explorer, (*Internet Microsft 2005*) while the Mozilla (*Fisher, Darin* 2005) and Opera web browsers opted to present a warning message and give the option of continuing to the site or canceling.

A further problem with URLs has been found in the handling of internationalized domain names (IDN) in web browsers, which might allow visually identical web addresses to lead to different, possibly malicious, websites. Despite the publicity surrounding the flaw, known as IDN spoofing (*Johanson, Eric 2005*) or a homograph attack,(Evgeniy G. & Alex G. 2002) no known phishing attacks have yet taken advantage of it. Phishers have taken advantage of a similar risk, using open URL redirectors on the websites of trusted organizations to disguise malicious URLs with a trusted domain. (Leyden, John 2006)(*Levine, Jason 2006*) (*Leyden, John 2007*)

1-2 Phishing using Filter evasion

Phishers have used images instead of text to make it harder for anti-phishing filters to detect text commonly used in phishing emails.(*Mutton, Paul 2006*)

1-3 Phishing using Website Forgery

Once the victim visits the website the deception is not over.(*Mutton, Paul Dec 2006*) Some phishing scams use JavaScript commands in order to alter the address bar. This is done either by placing a picture of a legitimate URL over the address bar, or by closing the original address bar and opening a new one with the legitimate URL. (*Internet BBC news 2004*)

An attacker can even use flaws in a trusted website's own scripts against the victim. (*Krebs, Brian 2006*) These types of attacks (known as cross-site scripting) are particularly problematic, because they direct the user to sign in at their bank or service's own web page, where everything from the web address to the security certificates appears correct. In reality, the link to the website is crafted to carry out the attack, although it is very difficult to spot without specialist knowledge. Just such a flaw was used in 2006 against PayPal.(*Mutton, Paul 2006*)

A Universal Man-in-the-middle Phishing Kit, discovered by RSA Security, provides a simple-to-use interface that allows a phisher to convincingly reproduce websites and capture log-in details entered at the fake site.(*Hoffman, Patrick 2007*)

To avoid anti-phishing techniques that scan websites for phishing-related text, phishers have begun to use Flash-based websites. These look much like the real website, but hide the text in a multimedia object.(*Miller,Rich 2007*)

1-4 Phishing using Phone

Not all phishing attacks require a fake website. Messages that claimed to be from a bank told users to dial a phone number regarding problems with their bank accounts.(*Gonsalves, Antone 2006*) Once the phone number (owned by the phisher, and provided by a Voice over IP service) was dialed, prompts told users to enter their account numbers and PIN. Voice phishing sometimes uses fake caller-ID data to give the appearance that calls come from a trusted organization.(*Internet Slicon.com 2005)*

1-5 Anti-Phishing

Current techniques to combat phishing, includes legislation, social responses and manual monitoring. Many countries started to take legal responses as well, some taken action to help identifying legitimate sites and the current technology created specifically to protect against phishing still depends on manual reports by users as well as automatically blocking some bad websites and displays the website address on the toolbar (Brandt, Andrew 2006).

2- Proposed Anti-Phishing Filter system

A cryptographic hash algorithm MD5 will be employed in our proposed system in order to generate a message digest and for verification purposes. The filter system will be on two parts (a) Filter system using cryptographic MD5 hash algorithm. (b) Filter system depends on user experiences. We will brief the MD5 then further information about the system's components will be illustrated.

2-1 MD-5

Ronald Rivest is the father of the MD (for Message Digest) family issued a more conservative replacement, for MD4, in 1991 called MD5. MD5 uses a 512-bit block size (only 448 are from the actual message, however; the rest is internal padding) and produces a 128-bit hash. More details and a sample C implementation can be found in RFC 1321. No collisions have been found yet (*Nick Galbreath 2002*). An abstract class from which all implementations of the MD5 hash algorithm inherit within the Dot Net environment at:

Namespace (System.Security.Cryptography)

This gave programmers an easy way to generate MD5 digest (*Internet Microsoft 2008*). MD5 employed in HTTP basic authentication scheme recently which implements password-based authentication to protect and to control access to the resources of a server. (*Rolf Oppliger 2003*)

2-2 The Filter System Using MD5.

A filter system shell will be designed using three stages a) MD5- Digest Generation Phase. b) Check Phase. c) Response Phase. See figure (1)

In MD5- digest generation phase some trusted websites might be entered initially, certain items about each websites will be taken like: URL page address, main website address, and the type of service (i.e. mail, bank or other). The items will be processed by the MD5 to get a message digest with only 128 bit long for each website items. On self check phase those information can be obtained to generate their corresponding MD5 digest and then to store it in a safe remote hosted database. The database has been chosen to be remote in order to keep it for both the availability and to serve users who use the same system which will gradually contains valuable information. The database will keep track of the bad websites with forgery attempts and thought number of the detections happened amongst all remote users can be stored as well for surveys and future reports purposes.

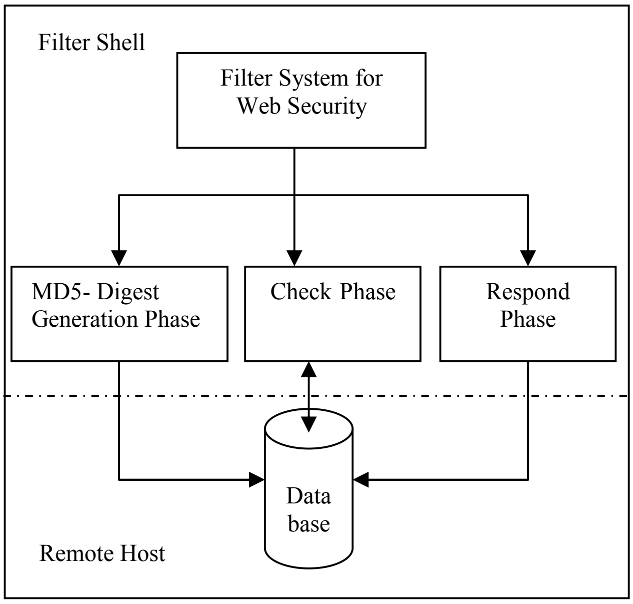

Figure (1) Architecture of the Filter System

The check and response phase are integrated together. When a user try to visit a website, its information will be checked with the database before opening it, and if there is a record about such website the check mechanism will make sure of the integrity of the current MD5- digest generated for the website information comparing it with the stored one in the system's database. If an attempt to visit a new website occurs, the system will generate an MD5 digest for it and will check for the multimedia contents weather it has been found with same host or not, since so many Phishing new techniques use known companies' logos with their forgery attempts. The finite state automata which has been designed to validate the URL address to prevent URL manipulation will manage to check and detect most of the link manipulation Phishing techniques, see figure(2).

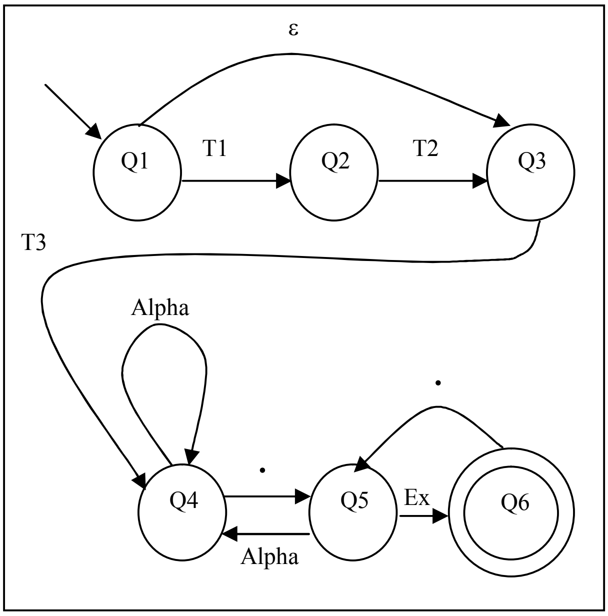

Figure (2) Uniform Resource Locator - Finite State Automata checker (URL-FA)

There are six states in figure (3) and the arcs which represent the probable terminals are in BNF- notation as follows in table(1):

<T1>::= http | https | ftp
<T2>::= ://
<T3>::= www | ε
<Ex>::= com | edu | gov | info |org | iq | ch | us |..|ir
<Alpha>::= <Alphabet> | <Alphabet> <Number>| <Number> |<Number><Alphabet>
<Alphabet>::= A|..|Z|a..|z| A <Alphabet>|..|Z <Alphabet> | a < Alphabet >|…|z < Alphabet >
<Number>::= 0|..|9 |0 <Number> | .. | 9 <Number>

Table (1) BNF- Notation for FSA-URL

The result and the information relates to it will be stored with its status for future use. An algorithm in the table (2) below states a full description of the system.

```
Algorithm Web_Filter_of_Phishing
Input:  URL page address (URL_pg_addr), Main website Address (web_addr), Type
of Service (ToS such as e-mail, Bank, other)
Output: Notification of a bloacked Fake Website or Normal
Steps:
Open database Connection(DB)
A ← MD5(URL,web_addr,ToS)
If (URL and web_addr ) exists in the DB then
B ← get_from_DB(MD5(URL,web_addr,ToS)
If (A = B) then
Status ←  “Safe”
Else
Status ←  “Fake”
Action ← “Block Access”
Add_to_DB(URL,web_addr,ToS, Status, Action)
End if
Else
Check Multi-Media(MM) contents
If (MM In host) then
Status ←  “Safe”
Else
Status ←  “Fake”
End if
Call URL-FA(URL)
If (URL-FA ← “False”) then
 Status ←  “Fake”
Action ← “Block Access”
else
Status ←  “Safe”
Action ← “Allow Access”
End if
Add_to_DB(URL,web_addr,ToS, Status, Action)
End if
End Algorithm
```

Table (2) Algorithm of the Proposed Filter System

2-3 User Dependent System

User dependent system gives the user the ability to add websites to the blocked list when they experience forgery, and a threshold value of the mean number of complains could give the decision of trusting or blocking websites more privilege, or the user may adjust the suspicions by trusting websites on his own decision which will not effect the system since there is a threshold value depends on the overall number of participations viewpoints.

3- Results

- The proposed system empowered by MD5- cryptographic algorithm has zero false negative rate when check performed on websites visited. (i.e. the system blocked any website with forgery intent by 1) URL-Manipulation. 2) Fake websites uses trusted companies logos and pictures. 3) Fake websites with suspected multimedia contents.
- Some false positive states happened trying to alarm the user and acknowledge for caution while visiting trusted websites.
- Fast performance for the system and the efficiency has been increased with more users' experiences.

4- Discussion

The main principle idea behind the Filter System for Web Security is to take advantage of a fast- safe cryptographic algorithm (MD5) for the sake of integrity purpose for the websites. Fake websites uses different methods to deceive the user and though using a fast algorithm for the integrity that only transfer 128 bit for each site is good since we use also low amount of data to be validated. The database should be safe and remotely use for saving information and results as well as a block capability supported by user experiences which added more power to the system as a confident measure as long as an accepted threshold has been passed to take user's viewpoints. The system as a plug-Ins allows the generalization where it can be used with any web explorer. The user can add entries via a web-based application which ensure the safety and the validity of the information linked as well as with the plug-Ins options.

5- Conclusions

The system is fast, reliable and can be implemented easily. Its importance came from the fact that if any slight change happens to a website name, address, type of service, and multimedia contents, the MD5- digest will get different and we will be able to prevent the user from visiting such site. The system has zero false negative rates with low false positive. The system is incapable of detecting fake websites that uses the cross-site scripting and the flash based action scripts contents.

References

Shon Harris, "CISSP® All-in-One Exam Guide", Fourth Edition, McGraw-Hill, © 2008.

Tan, Koon. "Phishing and Spamming via IM (SPIM)". Internet Storm Center. Retrieved on December 5, 2006 http://isc.sans.org/diary.php?storyid=1905

Skoudis, Ed. "Phone phishing: The role of VoIP in phishing attacks", searchSecurity, http://searchsecurity.techtarget.com/tip/0,289483,sid14_gci1193304,00.html, June 13, 2006

"Suspicious e-Mails and Identity Theft." Internal Revenue Service. http://www.irs.gov/newsroom/article/0,,id=155682,00.html Retrieved on July 5, 2006.

Berners-Lee, Tim. "Uniform Resource Locators (URL)", IETF Network Working Group. http://www.w3.org/Addressing/rfc1738.txt Retrieved on April 2, 2008.

Microsoft, "A security update is available that modifies the default behavior of Internet Explorer for handling user information in HTTP and in HTTPS URLs". Microsoft Knowledgebase. http://support.microsoft.com/kb/834489, Retrieved on August 28, 2005.

Fisher, Darin. "Warn when HTTP URL auth information isn't necessary or when it's provided." Bugzilla. Retrieved on August 28, 2005.

Johanson, Eric. "The State of Homograph Attacks Rev1.1.", The Shmoo Group. Retrieved on August 11, 2005.

Evgeniy Gabrilovich and Alex Gontmakher. "The Homograph Attack". Communications of the ACM 45(2): 128 (February 2002).

Leyden, John. "Barclays scripting SNAFU exploited by phishers", The Register, August 15, 2006.

Levine, Jason. Goin' "phishing with eBay. Q Daily News." Retrieved on December 14, 2006.

Leyden, John. "Cybercrooks lurk in shadows of big-name websites", The Register, December 12, 2007.

Mutton, Paul. "Fraudsters seek to make phishing sites undetectable by content filters." Netcraft. Retrieved on July 10, 2006.

Mutton, Paul. "Phishing Web Site Methods." FraudWatch International. Retrieved on December 14, 2006.

"Phishing con hijacks browser bar", BBC News, April 8, 2004.

Krebs, Brian. "Flaws in Financial Sites Aid Scammers", Security Fix. Retrieved on June 28, 2006.

Mutton, Paul. "PayPal Security Flaw allows Identity Theft", Netcraft. Retrieved on June 19, 2006.

Hoffman, Patrick. "RSA Catches Financial Phishing Kit", eWeek, January 10, 2007.

Miller,Rich. "Phishing Attacks Continue to Grow in Sophistication", Netcraft. Retrieved on December 19, 2007.

Gonsalves, Antone. "Phishers Snare Victims With VoIP", Techweb, April 25, 2006.

"Identity thieves take advantage of VoIP", Silicon.com, March 21, 2005.

Brandt, Andrew. "Privacy Watch: Protect Yourself With an Anti-phishing Toolbar", PC World - Privacy Watch. http://www.pcworld.com/article/125739-1/article.html, Retrieved on September 25, 2006.

Nick Galbreath, "Cryptography for Internet and Database Applications", Wiley Publishing, Inc., 2002.

Microsoft Support, "Microsoft Visual Studio 2008 Documentation", www.msdn.com/visualstudio2008, 2008.

Rolf Oppliger, Security Technologies for the World Wide Web, Second Edition, Artech House©2003.