Analyzing Websites and Hyperlinks for Malicious Codes

Ass.Lecturer **Mustafa Sabah Mustafa***

ABSTRACT

As we all know Internet is a very widely used network, containing different types of multimedia, documents, programs and websites, for that reason it is difficult to maintain it as a malicious-free environment, these malicious codes are embedded into an innocent looking website, these websites in turn are very wide spread sites. An important feature a malicious code depends on reaching to as many users as it can get.

This paper implemented generally to find and detect a malicious code in Internet websites, because of the wide spread of vandal users and malicious programs, the need for that tool has become a flat fact. A targeted website is analyzed for existing code and if that code is does not exist then its links are followed individually to check whether or not there is an indirect malicious code, this approach is useful in rapidly growing websites.

The application is run on the server and when the client makes its request to load a website this request is sent first to the server then the system follows that on line to internet by (URL) and checks the contents of its website and follows all its sub-links in that website then if the website is clean (malicious free) then the client is granted access to that site, otherwise the system prints a warning message to the user containing the type of the malicious code and the time required to detect the infection.

*** Al-Mansour University College/Computer science department/ Baghdad, Iraq.**

1-Introduction

The Internet has become the main method in exchanging cultures and transferring knowledge between people. The underlying language of the World Wide Web, HTML and its related scripting languages are widely exploited. Design web is not an easy task. To have an attractive web you need a skill, ideas and information. Therefore, it must be protected and secured the web site.

Security technologies are commonly used to establish identity (authentication, authorization, and access control) and ensure some degree of data integrity and confidentiality in a network. It provides solutions for securing network access and data transport mechanisms within the corporate network infrastructure [1]. Malicious code refers to viruses, worms, Trojan horses, logic bombs, and other "uninvited" software. Malicious code is sometimes mistakenly associated only with personal computers, but can also attack systems that are more sophisticated. However, actual costs attributed to the presence of malicious code have resulted primarily from system outages and staff time involved in repairing the systems [2, 3].

Hyperlink security is very important when you protect the web site from any attacks or any malicious code, therefore structure of web page must be analyzed and all links suspicion checked to detect his malicious code. Although pure HTML viruses have not given security experts a great reason to be

alarmed, there are several ways HTML can be maliciously used in [4, 5].

2- Types of Connections and Connectors

The Internet is decentralized; Internet communication is made possible by the transmission control protocol/internet protocol (TCP/IP) software on your computer figure (1) shows the multiple connections among hosts in the net [6].

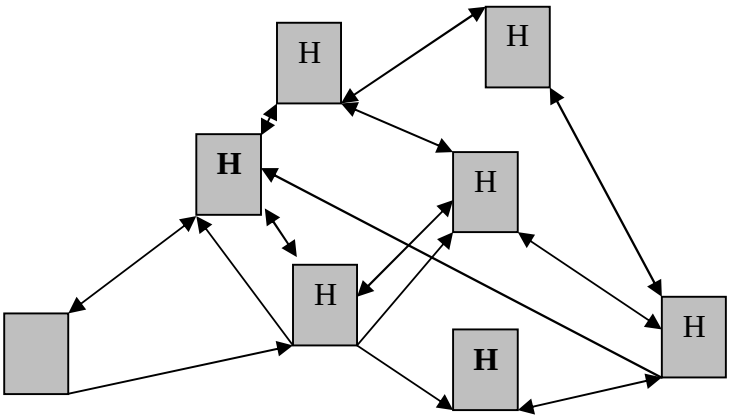

Figure (1) Multiple Connections among Hosts

There are two types of computer hosts connected to the Internet: server hosts and client hosts. The Server Host can be described as an "information provider" .This type of host contains some type of resource or data. The second type of host connected to the Internet is the Client Host, which can be described as an "information Retriever"; the client host will access resources and data located on the server hosts, but usually will not provide any resources Back to the server host; both server and client host computers can be connected to the Internet by various methods that offer different communication capabilities dependent on varied communications. There are three connections to the Internet:

2.1 Direct Internet Connections

A computer connected directly to the Internet via a network interface that allows the user the highest inter network functionality. Each computer connected in this manner must also have a unique Internet (IP) address. This type of connection is also the most expensive [6].

2.2 Serial Internet Connections

Another type of connection offering most communications capabilities is a SLIP (Serial Line Internet Protocol) or PPP (Point to Point Protocol) connection; full network and application capability over a serial (modem) line. Since this connection offers full TCP/IP each computer configured in this manner requires its own IP address. the ISP assigns an IP address at that point. It also means that the address for the dialer may change each and every time the system connects [8].

2.3 Host Access Connections

The most limited type of network access is available as a User account on a host, which is directly connected, to the Internet. The user will then use a terminal to access that host using a standard serial connection. This type of connection is usually the most inexpensive form of access. This type of connection is by far the most limiting, since the computer has no electrical connection to the Internet at all. This type of connection is the most secure

because there is no direct access to the user's computer by a hacker [8].

3- Aspect of Security [6]

Defining a security policy is complicated because each organization must decide which aspects of protection are most important, and often must compromise between security and ease of use. For example, an organization can consider:

1- **Data Integrity: Integrity refers to protection from change: it is the data that arrives at a receiver exactly the same as the data that was sent.**
2- **Data Availability: Availability refers to protection against disruption of service. Data remain accessible for legitimate uses.**
3- **Data Confidentiality and Privacy: Confidentiality and Privacy refer to protection against snooping or wiretapping. Data must be protected against unauthorized access.**

4- Network Security:

Network worms represent a serious threat to identity, integrity, an availability of computer resources on the Internet, The existing automated network-security solutions (anti-virus software, intrusion detection systems) and human-dependent counter measures (software patching, traffic blocking) have been deemed inadequate for timely detection and control of worm propagation.

The problem of Network worms is worsening every year despite increasing security measures [12].

Attacks employ Malicious Mobile-Code (MMC) a program designed to perform a malicious action, MMCs may be grouped into three classes: Trojan horses, computer viruses, and network worms [7].

5. A Hyperlink Network Analyze [10,11]

Hyperlink affiliation networks as a function of the credibility among web sites and the desire to strengthen certain dimensions of credibility.

A website perceived highly credible gets more links from others. The strength of links, in this case, the number of incoming hyperlinks, is an indicator of the web site's credibility. Thus, website position relative to other commercial web sites could be examined as a hyperlink network. They developed a sites-by-sites matrix of hyperlink existence upon which they conducted hierarchical cluster analysis.

The *Classification* process has two important responsibilities:

1. **Classify the links for the Downloading process to generate successors of the currently downloading page.**
2. **Classify the downloaded data to place it in the correct directory corresponding to its type.**

Hyperlink-Management there are two types

5.1 Common Hyperlink-Management

A link-management tool has to avoid broken links that are left on the web site. Each link has to be checked for its destination existence. Supply suggestions for new links between documents. The link-management tool has to display all documents that will contain broken links if the document is removed, If a document is only moved to another location (URL changes), the link-management system should offer support in repairing the links getting broken in subsequence.

5.2 Multilingual Hyperlink-management

Provides the same functionality as a system built for the management of a single language web site.

6. Malicious Mobile Code and security

Malicious mobile code (MMC) is any software program designed to move from computer to computer and network to network, in order to intentionally modify computer systems without the consent of the owner or the operator. MMC includes viruses, Trojan horses, worms; a script attacks, and rogues Internet code. The intentional part of the definition is important design flaws in the Microsoft windows operating system which is responsible for more data loss than all the malicious code put together, but

windows wasn't intentionally designed to destroy your data and crash your system. Today, add all harmful programs created with scripting languages and empowered by Internet technologies: macro viruses, HTML, java applets, ActiveX, VBScript, JavaScript, and Instant message [15, 16].

We can assume all end users will ignore or forget any advice about running entrusted code. We can assume that end users will visit malicious web sites, open any email, run any attachment, and use infected diskettes and programs. The truth is that end users shouldn't have to concern themselves with how to prevent malicious code. They just want to use their computer and visit the Web. If we want maximum malicious code protection, disable Internet access, uninstall any Internet browsers, remove email, and disable the floppy drive. If we only need reasonable protection the following recommendations, are the steps we should make to protect PCs under our control [17, 18].

1- Install an anti virus scanner

Installing a reliable up to date antivirus scanner is the single best thing we can do to prevent malicious code. The question is where to scan: desktop, file server, e-mail gateway or firewall.

2- Install the latest version of the software where possible. Install the latest versions of all known exploitable application and operating systems. This only not means Windows, Microsoft Office, and the Internet explorer, but all other applications.

3- Reveal hidden files and extensions. The user must make sure to unhide window defaults hidden files and a file extension. This means setting options under window explorer and editing the registry.

4- Rename dangerous executables although not considered an elegant way to prevent malicious code attacks, renaming common files used by malicious hackers is an easy way to prevent attacks. As an example, the following files can be renamed or deleted depending on the environment's potential use of them.

FORMAT.EXE

REGEDIT.EXE(OR REGED32.EXE)

DEBUG.EXE

WSCRIPT.EXE(AND CSCRIPT.EXE)

The plan should encompass all the policies and procedures needed to protect the PCs and networks under control. It must address the protection of PCs; include end user education, list the tools will use to fight malicious code. And establish how outbreaks will be handled, each personal computer under your control need to be modified to prevent malicious mobile code from attacking and from spreading further.

7. HSAW System

Because of the complexity, computers and computer networks have become a target of computer crime more and more often. Large theoretical and practical efforts are concentrated today on this problem. Nevertheless, a perfectly secure system is still a myth. Many modern computer system still lack properly implemented security services, contain a variety of vulnerabilities exploited by threats, analyzing website techniques are used to strengthen the system security and increase its resistance to internal and external attacks.

This paper aim to give details of the hyperlinks security analysis website (HSAW) system. HSAW detects malicious codes (viruses, Trojan, horse, and worms) in websites. It analyzing website tags (links) and computes the required time to detect the malicious code. When the user uses the Internet many results will be displayed by the Internet. The user chooses one of the displayed websites to download. The proposed system (HSAW) starts execution. It opens the HTML file (source code), structure analyze web page tags (links, and use the database that contains a List of malicious codes (see appendix A). The malicious code database can be modified (add new malicious code or detect an old one). HSAW compares the opened website with the known malicious codes in the database to detects if it is a malicious code or not. It checks the link if it is real or not. Also it computes the time required to detect the malicious code.

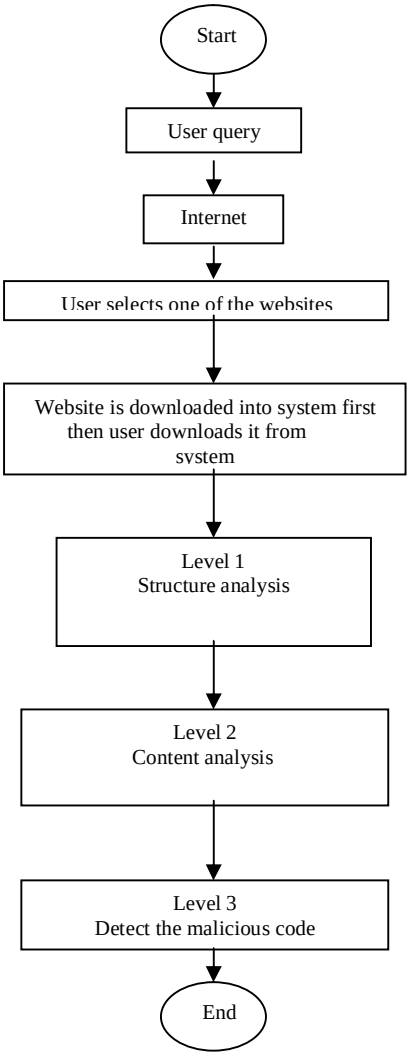

Figure (2) System Block Diagram of HSAW

Figure (2) is the block diagram of the HSAW. There are (3) levels of working in HSAW. They are level1, level2, and level3. Each level has some tasks to do. When it completes its tasks, the next level will be start working until the total levels of the system

is complete. Each level performs analysis and test cycle to satisfy the goal of the system.

The (3) levels are: -

1- Level1: Structure analysis.

2- Level2: Content analysis.

3- Level3: Detect malicious code.

Structure analysis opens HTML files (source code) and the database. It finds all tags (links) involved in the web page. It checks the hyperlinks in the web page. It also matches the tags with the malicious codes in the database.

Content analysis checks the links if they are real or not. if they are real then they are connected to important web page. This web page has to be checked for any suspicious code.

Detect malicious code by using the database, which contains many kinds of malicious codes. It matches tags (links) and different kinds of malicious code. It displays the result in the area of check result .the results are the file name and the type of malicious code.

The checking operation consists of two procedures that are called sequentially to check the following: -

1. **The database, which contains the names of the known suspicious codes.**
2. **The dangerous tags, this procedure searches for dangerous tags (links) that can call and execute a malicious code.**

After applying all the rules of the two procedures above, a new check is done on the HTML file. This new processing search in the HTML files for a link (source) to another file document, if any

link exists, the same two above mentioned procedures are applied to the linked file, otherwise the system will display the check results of the HTML file.

8. Experimental Result and Discussion

The following are some results, which are obtained from running HSAW on the server. The user can make his requests from the lost computes after connecting it to the Internet using URL. The user enters the website name and choose the go icon. HSAW system will display the website information and opens web pages to check all tags in the existing open files that belong to the same family as HTML. Figure (3) is the output of the above-mentioned operations.

Figure (3) Online connection to the Internet using URL

HSAW opens the HTML files and searches the database for theses files, which tags (links) involved in the web page. The purpose of this analysis is to check the hyperlinks in these web pages and to discover if malicious codes exist. These operations are presented in figure (4) and figure (5).

Figure (4) Result of Checking HTML file

Figure (5) Adding New Malicious code to the database

HSAW displays the results in the check result area. The results are the file name, its links, and the malicious code as shown in figure (6).

Figure (6) End Result of Checking the HTML File

9. Conclusion and future work

The following are some conclusions obtained from the implementation and execution of the HSAW:

1. **A good analysis of the website structure is very important step to discover malicious codes.**
2. **As long as the Internet websites are developed and grown, malicious code problems grow with it and become more difficult to detect.**
3. **The probability of a malicious code into a website is greater if that website contains many links to other web pages.**

4. **Building an updateable library is important to include all the variety of malicious codes and downloading updates for this library.**

Some suggestions for future related works are:

1. **Making the library automatically updateable from Internet by new malicious code signatures.**

2. **Building a miniature model of the system as a web crawler and sending it over the Internet to clean it up.**

3. **Running several copies over the client/server network each one of which at the client's computer to increase speed over a network.**

4. **Adding a checksum entry into the database, updateable only by the system to prevent other malicious code from manipulating the database.**

Refrence

1- Jim K. Mark, "*Understanding the World Wide Web*", University of Albany libraries, 1998.

2- Merike Kaeo,"*Designing Network Security*", Second Edition, Macmillan India Ltd.2004.

3- Mark C. John," *Searching the World Wide Web*", 1998.
http://www.neci.nec.com/~lawrence/papers/search-ic98/search-ic98.pdf

4- Terry Escamilla,"*Intrusion Detection*", John Wiley& Sons, Inc, 1998.

5- Doglas E.Comer,"*Computer Network and Internet*", second edition, Prentice Hall Inc, 1999.

6-John E. Howland, "*Introduction to Internet Security*", Department of Computer Science, MS.C. Trinity University, 2002

7- Zoran Nikoloski , Narsingh Deo, and Ludek Kucera, "*Correlation Model for Worm Propagation on Scale-free Graphs*", MS.C., Charles university,2005.

8- Choon Yang Quel, "Classification World Wide Web", MS.C, Carnegie Mellon University,2005.

9- GE Cormack,"*Website Structure*", Prentice Hall PTR.,1998

http://www.feccauton.org/website/websitetips/structure.html

10- James L. Mohler and John M. Duff," *Designing Interactive Websites*", Delmar,2000

11- MSDN Library, "*Website Planning*" 2006.

www.adptrain.com/pages/site-planning.html

12- TrendMicro, "Security *Testing, Firewall, and Checkers*", Proc., 2002

www.carnet.hr/cuc/cuc2000/radovi/prezentacije/trendmicro.com,

13- K.Smith,"*Securing an Internet Name Server*",2000

http://www.praxis.bond.edu.au/prism/papers/refereed/paper5.pdf

14- Jackson A. William, "*Assessing the structure of communication on the World Wide Web*", Journal of Computer-Mediated Communication, 2000.

15- CERT advisory, "*malicious code and application*", O'reilly &Associates Inc. 2000.

16- Roger A. Grimes, "*Malicious Mobile Code*", O'reilly &Associates Inc., 2001.

17- Merike Kaeo,, "*Design Network Security* ", Pearson Education (Singapore) Pte..Ltd., India, 2004.

18- Net Sam, "*Maximum Security*", Macmillan Computer publishing, 1997.

Appendix A : Examples of Malicious codes

1-

{Operating system DDL's: which contains all operating system routines}

SERVER_33.DLL

Windll.dll

Kernel.dll

WATCHING.DLL

MOVOKh_32.dll

wsock32.dll

2-

{Programs}

Bo2k.exe

Bg10.exe

windll.exe

exec.exe

systray.exe

server.exe

NBsvr.exe

nnsx.exe

patch.exe

server.exe

netbus.exe

DATA2.EXE

TINURAK.EXE

WINDOW.EXE

NODLL.EXE

Rundll16.exe

c:windowslinks.vbs

c:windowssystemrundll.vbs

SKA.EXE

wsock32.SKA

win32.SKA SKA

wsocks.SKA SKA.exe

K2PS.EXE

K2PS.CFG

agent.exe

files32.vxD

Promail.exe

Promail121.ZIP

Zipped_files.exe

worm explore.ZIP

win32.explore explore.ZIP

c:windows_setup.exe

c:windows_explore.exe

3-

{Windows script library: can run any windows rotine using windows script by running wscript.shell}

4-

{Java script malicious functions: examples}

java.applet

java.awt

java.awt.image

java.awt.peer

5-

{Class ID:example clsid:13709620-C279-11CE-A49E-444553540000}

sometimes attack deal with certain class by it's ID instead of it's Name whick makes it so hard to tracking (every component in windows environemt treated as class and has class ID)

6-

{Windows registry libraries}

regclosekey

regcreatekey

regcreatekeyex

regdeletekey

regdeletevalue

regenumkey

regenumkeyex

regenumvalue

regflushkey

regopenkey

regopenkeyex
regqueryinfokey
regqueryvalue
regqueryvalueex
regsetvalue
regsetvalueexwshshell.reg

By using these libraries by VBS OR Java script attacker can read, modify, and update the values of certain entries as shown in the following examples:

1- **love letter attack: once it executed the viruses check to change the following entries in windows registry:-**

HKEY_CURRENT_USER\software\windows\service\host\setting\ Timeout

And will create the following entries: -

HKEY_LOCAL_MACHINE\SOFTWARE\windows\currentversion\run\MSKernel32

HKEY_LOCAL_MACHINE\SOFTWARE\windows\currentversion\run\Win32DLL
Which means that on boot up the c:\windows\MSKernel32 and c:\windows\Win32DLL.VBS are executed.

2- **The binary executable part of the worm which it downloads from the net is a password stealing Trojan, sort of utility. The Trojan tries to find a hidden window named "BAROK". If it is present, the Trojan exist immediately, if not the main routine takes control the Trojan checks for the WinFat32 sub key in the following registry key:**
HKEY_LOCAL_MACHINE\software\Microsoft\Windows \Current Version\run
Then, the Trojan sets internet explorer startup page to about blank that the Trojan tries to find and delete the following keys:
Software\Microsoft\Windows\CurrentVersion\Policies\Network\HideSharePwds
Software\Microsoft\Windows\CurrentVersion\Policies\Network\HDisablePwdsCaching
DEFAULT\Software\Microsoft\Windows\CurrentVersion\Policies\Network\HideSharePwds
DEFAULT\Software\Microsoft\Windows\CurrentVersion\Policies\Network\DisablePwdCaching

تحليل المواقع والتوصيلات للبرامج الخبيثة

م. م. مصطفى صباح مصطفى*

المستخلص

كما نعلم بان الانترنيت اصبح شبكة واسعة الاستخدام ، تحتوي على انواع عديدة من الاوساط المتعددة كالوثائق و البرامج و المواقع ، لهذا السبب انه من الاصعب ادامتها كبيئة خالية من التخريب ، هذه البرامج الخبيثة تكون مخفية داخل مواقع ذات مظهر برئ ، هذه المواقع بدورها واسعة الانتشار ، وهي صفة مهمة للبرامج الخبيثة كي تعتمد عليها للوصول الى اكبر عدد من المستخدمين.

الهدف الرئيسي من هذا البحث هوايجاد وكشف البرامج الخبيثة في مواقع الانترنيت ، ولان هذه البرامج والمبرمجين المخربين منتشرون بكثرة فان الحاجة لهذه الاداة اصبحت حقيقة واضحة ، الموقع المستهدف يتم تحليله لكشف وجود هذه البرامجيات الخبيثة فاذا لم يجد هذه البرامجيات فيه فيتم تتبع الوصلات بصورة مفردة لفحص هل البرامج موجودة بصورة غير مباشرة ، هذه الطريقة مفيدة في المواقع ذات الانتشار السريع.

يتم تشغيل النظام المقترح على الخادم الرئيسي عندما يقوم المستخدم بطلب تحميل صفحة فهذا الطلب يتم ارساله اولا الى الخادم ثم يتبع هذا المسار ويتفحص المحتويات لذلك الموقع متتبعا كل الوصلات الثانوية فاذا كان الموقع نظيف (اي خالي من البرامج الخبيثة) فان الخادم يسمح باستخدام ذلك الموقع والافسيقوم بكشفها ويكشف نوعها ويهيئ تقرير للمستخدم عن نوع الاصابة كرسالة تحذيرية في الموقع الخاص بالفحص ويحسب الوقت المستغرق لإيجادها.

*قسم علوم الحاسبات/كلية المنصور الجامعة