International Journal of Engineering Inventions
ISSN: 2278-7461, www.ijeijournal.com
Volume 1, Issue 8 (October2012) PP: 01-06

AUTOMATED WEB PATROLING WITH EXPLOIT DETECTION OF MALICIOUS SITES

Ms. Rajashree C. Sonawane[1], Prof. Mrs. S.M.Jaybhaye[2]
[1,2]Department of Information TechnologySinhgad College of Engineering, University of Pune, Pune

Abstract:—Client-side attacks are on the rise malicious websites that exploit vulnerabilities in the visitor's browser are posing a serious threat to client security, compromising innocent users who visit these sites without having a patched web browser. Malicious websites are websites which have any kind of content that could be a threat for the security of the clients requesting particular sites. Currently, there is neither a freely available comprehensive database of threats on the Web nor sufficient freely available tools to build such a database. The purpose of an exploit is almost always to introduce some piece of arbitrary code on the victim machine, as a way to achieve a larger attack goal. It is not uncommon to see a malicious web page attempting to exploit multiple browser vulnerabilities in order to maximize the chance of a successful attack. Hence Honey Monkeys, introduce the Monkey-Spider project. Utilizing it as a client honey pot, the portray the challenge in such an approach and evaluate system as a high-speed, Internet scale analysis tool to build a database of threats found in the wild. Furthermore, it evaluates the system by analysing different crawls performed.

I. INTRODUCTION

Internet attacks that use a malicious or hacked web site to exploit unpatched client-side vulnerabilities of visiting browsers are on the rise. These attacks allow web servers that host compromised URLs to install malcode on visiting client machines without requiring any user interaction beyond visitation. There have been several manual analyses of these events. Although these analyses provide very useful and detailed information about which vulnerabilities are exploited and which malware programs are installed, such efforts are not scalable, do not provide a comprehensive picture of the problem, and are generally ineffective at efficiently finding new malicious sites. To address these issues, developed a system that uses a pipeline of active, client-side honeypot, called Strider HoneyMonkeys, to perform large-scale, systematic and automated web patrol. The HoneyMonkey

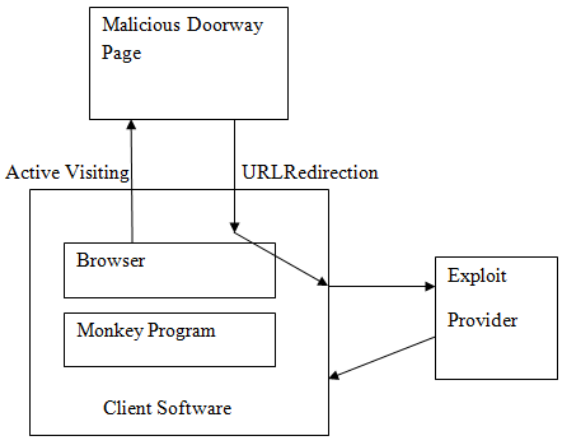


Figure 1: Exploit detection at client side

system uses monkey programs that run within with OS's of various patch levels to drive web browsers in an attempt to mimic human web browsing. This approach adopts a state-management methodology to cyber security: instead of directly detecting the acts of vulnerability exploits, the system uses the Strider Tracer to catch unauthorized file creations and configuration changes that are the result of a successful exploit. [1]

II. LITERATURE SURVEY

Web Patrol: This web page monitoring software makes following hundreds of web sites and blogs as simple and fast as checking your e-mail. It will visit web sites for you, as often as you like, and notify you when it finds a change. Web patrol is used to Checks single pages or entire web sites for changes on the particular page. It is used to Checks normal pages or frame based pages and Monitors PDF files for changes .This *web* page monitoring software makes monitors web sites,

blogs and text files for changes .It also improves your productivity by working silently and unattended .Web patrol Notifies you of changes by popup, sound and/or email and Shows pages with changes highlighted in the internal browser.

Different security issues related to web sites:

Website is a broad field, but most websites have common security issues that need to be addressed, regardless of the particular technologies used or functions deployed.

1. Validation of input and output data

All data written as output (displayed) needs to be safe to view in a browser, email client or other software and the integrity of any data that is returned must be checked..

2. Malicious file executions

Uploaded files or other data feeds may not be what they seem. Never allow user-supplied input to be used in any file name or path (e.g. URLs or file system references). Uploaded files may also contain a malicious payload so should not be stored in web accessible locations.

3. Authentication and session management

Websites rely on identifying users to provide access permissions to data and functions. [10]

Existing tools:

1. Online link Scanner: Online Link Scan - Virus, Trojan, Adware and Malware Scanner has come up with. It will save and ultimately protect the computer and the data of the user from getting unethically violated. So if you are worried about constant violation of your personal space while your surf or go to your chosen links by being redirected to harmful threats, at Online Link Scan - Virus, Trojan, Adware and Malware Scanner take the problem off you and offer safeguards like detection of hidden links that are not possible for you to notice. It detect an impending danger to your computer by allowing you to scan for suspicious links that might gets infected with viruses, trojanhorses, spyware and other malware. Although it accept no liability for the accuracy and integrity of all scan results but will do our best to help protect your system as to reduce the possibility of being invaded.[8]

2. Web site pulse: Website Pulse is a leading provider of global, independent and objective monitoring of the availability and performance of web sites, servers and network components, web applications and e-business transactions, web-based and e-mail systems. The advanced web monitoring service ensures reliable early problem and error detection with multi-stage verification process, followed by real-time multimedia alerts. Detailed reports and website diagnostic tools allow web system owners and operators to quickly locate and troubleshoot any problem, minimizing the downtime of their web systems. The Website Pulse monitoring agent - the proprietary platform of Website Pulse - is designed for synchronized 24/7 monitoring through an integrated global network of interconnected and redundant monitoring locations throughout North America, Europe, Asia, Australia, and New Zealand. It constantly measures website accessibility and web-based systems performance, immediately triggers alerts and troubleshooting notifications of detected problems, and provides its customers with real time data charts, graphs and raw data for detailed analysis.[9]

III. EXISTING METHODOLOGY FOR EXPLOIT DETECTION

(1) Executable files created or modified outside the browser sandbox folders:

This is the primary mechanism for exploit detection. It is implemented on top of the Strider Tracer, which uses a file-tracing driver to efficiently record every single file read/write operation. [1]

(2) Processes created:

Strider Tracer also tracks all child processes created by the browser process.

(3) Vulnerability exploited:

To provide additional information and to address limitations of the black-box approach, developed and incorporated a Vulnerability-specific detector. This is based on the *vulnerability* signature of the exploit, rather than on any particular piece of malcode.

(4) Redirect-URLs visited:

Since malcode is often laundered through other sites, this module allows us to track redirections to determine both the real source of the malcode and those involved in the distribution chain Upon detecting an exploit, the monkey saves its logs and notifies the Monkey Controller on the host machine to destroy the infected VM and re-spawn a clean Honey Monkey, which then continues to visit the remaining URL list. The Monkey Controller then passes the detected exploit-URL to the next monkey in the pipeline to further investigate the strength of the exploit. [1][2]

Redirection Analysis

Many exploit-URLs do not perform the actual exploits but instead act as front-end content providers that serve "interesting" content such as pornography in order to attract browser traffic. This traffic is then sold and redirected to back-end exploit providers, which specialize in exploiting clients and installing malware.URLs visited through traffic redirection can be tracked with a Browser Helper Object running within each browser process or by intercepting and analysing network packets. When the Honey Monkey runs in its "redirection analysis" mode, any automatically visited URLs are fed back to the system for further checking. [1]

IV. OVERVIEW OF MODULES

There is different vulnerability exploits Related to web browser are:

Web Site Vulnerability Exploitation:

Malicious activities performed by actual web sites exploiting browser vulnerabilities can be divided into four steps: code obfuscation, URL redirection, vulnerability exploitation, and malware installation.

MODULE 1 URL Redirection from list of URLs visited :

Most malicious web sites automatically redirect browser traffic to additional URLs. Specifically, when a browser visits a primary URL, the response from that URL instructs the browser to automatically visit one or more secondary URLs, which may or may not affect the content that is displayed to the user. Since redirection is commonly used by non-malicious sites to enrich content, simply eliminating redirection from a browser would present significant complications.

Redirections

Type	Details
Report ID :	81
Site :	http://www.google.com
Status :	Redirections Found..!
Location Redirection :	0
Refresh Redirection :	0
Windows Location Replace Redirection :	0
Windows Location Href Redirection :	1
TOTAL :	1

Figure 2 Report for Redirection analysis

MODULE 2 Vulnerability Exploitation due redirections and popups

It is not uncommon to see a malicious web page attempting to exploit multiple browser vulnerabilities in order to maximize the chance of a successful attack. HTML fragment that uses various primitives to load multiple files from different URLs on the same server to exploit three vulnerabilities fixed in Microsoft Security Bulletins.

popups

Type	Details
Report ID :	16
Site :	www.yahoo.com
Status :	Popups Not Found..!
Window Open Popups :	0
Window Show Modal Dialog Popups :	0
Message Box Popups :	0
Alert Popups :	0
TOTAL :	0

Figure 3 Report for popup

If any of the exploits succeeds, a Trojan downloader named win32.exe is downloaded and executed. Note that although Internet Explorer is the common target due to its popularity, other browsers can also be attacked.

MODULE 3 Malware Installations:

The purpose of an exploit is almost always to introduce some piece of arbitrary code on the victim machine, as a way to achieve a larger attack goal, including viruses that infect files, backdoors that open entry points for future unauthorized access, boot programs that allow the attacker to control a whole network of compromised systems, Trojan downloader's that connect to the Internet and download other programs, Trojan droppers that drop files from themselves without accessing the Internet, and Trojan proxies that redirect network traffic. Some spyware programs and even anti-spyware programs are also installed through exploits.[1]

MODULE 4 Result Analysis: Generated report for site Status:

Honey Monkey System will display the result of exploitation
by displaying result in the report format. Following is the report generated by system for the result of all the contents of URL http://www.google.com. Similarly different Redirections
and popup is detected during exploitation detection process are displayed in the results. Finally it will show the status of web page and automatically it will redirect to that web page.

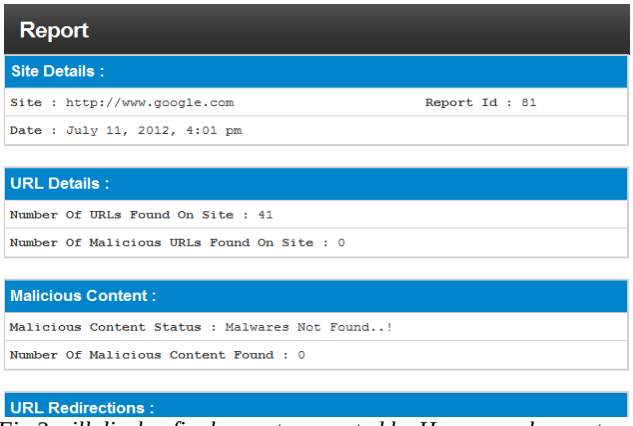


Fig 3 will display final report generated by Honey monkey system.

Figure 4: Report for exploit detection

V. PROPOSED SYSTEM

Exploit Detection:

Monkey program that launches a browser instance to visit each input URL and then waits for a few minutes to allow downloading of any code which may have a short time delay. It detects a group of persistent-state changes to signal exploitation.

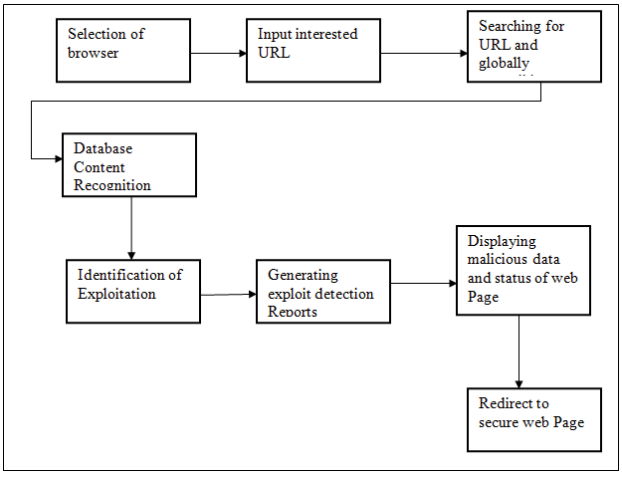


Figure 5: Flow diagram Of HoneyMonkey System

Since the monkey is not instructed to click on any dialog box to permit software installation, any executable files or registry entries created outside the browser sandbox indicate an exploit. This approach has the additional important advantage of allowing the detection of known-vulnerability exploits.

HoneyMonkey uses a black box system to detect exploits, i.e., it doesn't use a signature of browser exploits to detect exploits. A Monkey Program, a single instance of the HoneyMonkey project, launches Internet Explorer to visit a site. The monkey does not allow pop-ups, nor does it allow installation of software. Any read or write that happens out of Internet Explorer's temporary folder therefore must have used browser exploits. These are then analysed by malware detection programs and then manually analysed.

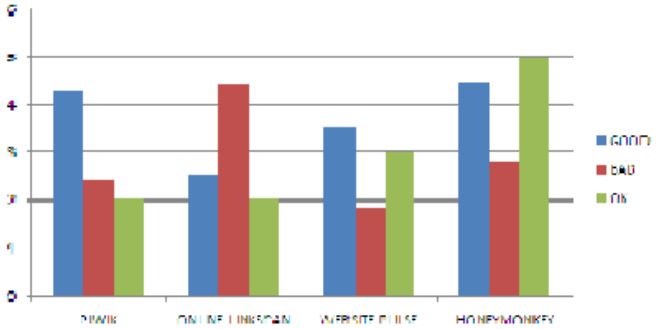
Site redirection is a browser concept that enables a site URL to be redirected from one site to one or more other sites. It is capable of providing modularity on the internet. The redirection may occur without the authorization, or even knowledge, of the user. Especially when the site redirection occurs without the knowledge of the user, redirection may be

more precisely termed a "cross-domain auto-visit". For most exploit scenarios, the page content stays at the URL to which the user originally directed a browser while the other URL(s) are being auto-visited by the browser under the cover.

Although there are legitimate purposes to redirection that benefit web site operators and browser users alike, site redirection can also be used for nefarious purposes. For example, it is often used by exploit providers to hide their identity and/or make exploit detection more difficult to track and prevent. The exploit "experts" pay provider sites to redirect clients to their exploitive sites. Malware development companies then pay exploiters to install their software.

VI. RESULT AND ANALYSYS

Malware classification of different websites using different existing tools and our system Honeymokey is done in given chart below:

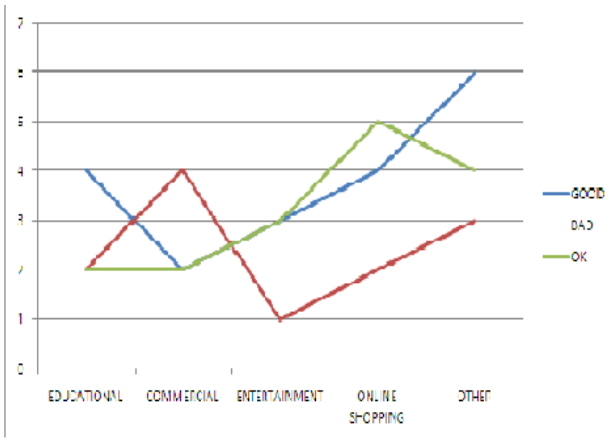


Graph 1: Malware classification

The above graph shows comparison between different tools used for scanning URLs and no of sites detected good, bad or ok are shown in graph up to 10 URLs are scanned using different tools and depend upon result graph is displayed. Comparison between different types of sites is shown by line chart, chart shows comparison depend on result given by Honey monkey system. 10 URLs of each type of websites are scanned and result of no. of sites comparison is shown by line chart Depend upon values stored in table.

Types of URLs scanned	Status displayed of suspicious lists	Excellent (100%)	Good (50-75%)	Bad (Below50%)
Educational Sites	Scanned upto10 URLs	5	4	1
Entertainment Sites	Scanned upto10 -15 URLs	6	3	1
Commercial Sites	Scanned upto 10 URLs	4	5	1
Online Shopping Sites	Scanned upto 10 URLs	4	2	2
others	Scanned upto 10-20 URLs	5	4	2

Table 1: Comparisons of different types of URLs using HoneyMonkey System



Graph 2: Comparison of different types of URLs

Above Line chart shows the URLs Scanned depend on their types such as Commercial, Entertainment, Shopping related sites. After scanned different results are displayed about exploitation of different sites depend on no of redirections, popup and malicious files found in database of Webpage**.**

VII. CONCLUSION

"Automated web patrolling and exploit detection of malicious sites" is systematic approach for the automated web patrol in finding for malicious web sites that exploits browser vulnerabilities at client side.

Honey monkey uses web crawler as application automatic web patrol. It is used as a general tool but modified application of web crawlers for web patrol. Design and Implementation of Strider Honey Monkeys as the first systematic method for automated web patrol to hunt for malicious web sites that exploit browser vulnerabilities.

Web sites can adopt Honey Monkeys detection which motivates to incorporate additional vulnerability-specific exploit detection. An easy way of avoiding the threat of malicious websites has to be the automatic avoidance of these dangerous parts of the Web in an easy to use and error-free manner. Monkey-Spider can also be used to collect information that can in turn be used to protect the Internet community. Analysis of four sets of data showed that the densities of malicious URLs such as for educational sites 50% excellent and 10% bad similarly for Entertainment sites 60% excellent 30% bad, for commercial sites 40% excellent,20% ok and 40% gives Bad status . It quantifies the nature of spyware web centric point of view.

REFERENCES

1. "Strider HoneyMonkeys: Active, Client-Side Honeypots for Finding Malicious Websites"Yi-Min Wang, Yuan Niu, Hao Chen, Doug Beck, Xuxian Jiang, Roussi Roussev,Chad Verbowski, Shuo Chen, and Sam King,2009
2. Strider Honey Monkey Exploit Detection, http://research.microsoft.com/HoneyMonkey.
3. T. Holz and F. Raynal, "Detecting Honeypots and other suspicious environments," in Proc. IEEE Workshop on Information Assurance and Security, 2005.
4. Yi-Min Wang, Doug Beck, Xuxian Jiang, and Roussi Roussev, "Automated Web Patrol with Strider HoneyMonkeys: Finding Web sites That Exploit Browser Vulnerabilities," Microsoft Research technical Report MSR-TR-2005-72, July 2008
5. S. Anderson and V. Abella,"Data execution Prevention changes to Functionality in Microsoft Windows XP Service Pack 2,"Memmory Protection Technology"Aug. 2006
6. Niels Provos ,"A virtual Honeypot Framework" in proc. USENIX security Symposium, Aug 2004.
7. **"**Monkey-Spider: Detecting Malicious Websites with Low-Interaction Honeyclients", Ali Ikinci Thorsten Holz Felix Freiling,University of Mannheim, Germany
8. http://www.onlinelinkscan.com
9. http://www.websitepulse.com/
10. https://www.watsonhall.com/resourses/dowmloads/to p10-wesite-security-issues.pdf