RESEARCH ARTICLE OPEN ACCESS

# Behavior Analysis Of Malicious Web Pages Through Client Honeypot For Detection Of Drive-By-Download Malwares

Supinder Kaur*, Harpreet Kaur**
*M.Tech Scholar (Computer Science and Engineering, Sant Baba Bhag Singh Institute of Engineering and Technology – Punjab Technology University)
** Assistant Professor (Computer Science and Engineering, Sant Baba Bhag Singh Institute of Engineering and Technology –Punjab Technology University)

**ABSTRACT**
Malwares which is also known as malicious software's is spreading through the exploiting the client side applications such as browsers, plug-ins etc. Attackers implant the malware codes in the user's computer through web pages; thereby they are also known malicious web pages. Here in the paper, we present the usefulness of controlled environment in the form of client honeypots in detection of malicious web pages through collections of malicious intent in web pages and then perform detailed analysis for validation and confirmation of malicious web pages. First phase is collection of malicious infections through high interaction client honeypot, second phase is validations of the malicious infections embedded into web pages through behavior based analysis. Malwares which infect the client side applications and drop the malwares into user's computers sometimes overrides the signature based detection techniques; thereby there is a need to study the behavior of the complete malicious web pages.

***Keywords*** – Malwares, Network Security, Intrusion Detection System, Client Honeypot

## I. INTRODUCTION

Business of the most of organizations is greatly affected by the internet today as business of most part of the world is shifted on internet. As we can see in today's current life of the human being is greatly affected by the social shopping sites such as flipcart. snapdeal etc. They are largely impact the human life in the form of daily living life. As we can see in our daily life internet plays the biggest role and there is a lots of impact of internet in our day to day life. Like shopping purposes, we can use the popular internet websites, but are these web pages are secured one so that we can blindly rely on these. There might be web pages which are infected by the malwares and which can infect the user's computer, but how to detect those malwares, what can be the useful and suitable mechanism for detection of these kinds of malwares.

When an organization or any innocent users of the internet make its resources on the internet such as web servers, at the same time, those resources is also being accessible to the malicious users. Those malicious users also known as hackers can get access of the complete resource of the organization in many ways. First of all, he will exploit the loopholes in the organization network also known as software bugs or vulnerabilities in the resources. After exploitation, he can get access of the organizational network.

In the organization network, the firewall, IDS/IPS are being placed for the security of the organizational network, but the protection given by these security products are limited as all these devices are relying on the signature based detection techniques, therefore no network can be hundred percent secure by putting these security devices. The firewall provides security by allowing only specific services through it. The firewall implements a policy for allowing or disallowing connections based on organizational security policy and business needs. The firewall also protects the organization from malicious attack from the Internet by dropping connections from unknown sources [1].

Honeypots plays the significant roles in terms of detection of known and unknown attacks spreading in the network. Honeypots are the decoy systems which are being placed in the network to capture the attacks spreading in organizational network. Honeypots are divided into two categories in terms of their detection capabilities known as low interaction and high interaction honeypots. And in terms of security attack detection capabilities, honeypots can be categorized as server honeypots and client honeypots. Here client honeypots plays a biggest role for detection of infections embedded into the web pages. When a normal user in any organization browse these malicious URLs which is not detected by signature based security devices, then those infection can damage the user's computers. There is a malicious infection into the malicious web pages which exploit the client side applications such as browser and can attack the user's computers. These kind of attacks are occurred on user's computer when he or she visit these malicious websites through these client side applications [2].

To detect and correspondingly to develop the defense mechanism, there is a need to study the behavior of the malicious servers in terms of how the attacks are spreading and how they are exploiting the client side applications to target the innocent users. The technology so called client honeypots is latest and emerging area of research to detect the malicious server and attacks spreading through these malicious servers. Client honeypots actively interact with these malicious servers and during this active browsing, it monitor the complete system changes for any unauthorized system changes which clearly indicate the malicious activities in the system [3]–[6]. Any unauthorized state changes detected by client honeypots can be studied and investigated for the complete cycle of the attacks spreading through malicious servers.

## NETWORK ATTACKS AND THEIR DETECTION

The security attacks can be classified into two categories known as server side and client side attacks. As the term itself signifies that server side attacks are those kinds of attacks which exploit the server side vulnerabilities and loopholes in servers. The attackers can exploit the vulnerabilities of the server and planted the shell code and other exploits into the server side [7].

The attacks which are being propagated through exploitation of client side applications known as client side attacks. In this type of attacks, an attacker uses client application vulnerability to take control of client system by malicious server. A typical target is web browser. However, these attacks can occur on any client/server pairs such as email, instant messaging, FTP, multimedia streaming, etc. One example of non-browser application vulnerability exploits is Adobe Reader v8.1.2 which is prone to stack-based buffer overflow vulnerability. In general, client side exploits require user-interaction such as enticing the user to click a link, open a document, or somehow to let her visit the malicious websites [8].

### 1.1 Intrusion Detection System

The security device known as Intrusion Detection System (IDS) monitors the network traffic and provides the analysis of that traffic for possible attacks in network streams based on the signature databases associated with it. The core detection capabilities of Intrusion Detection System are the signatures which are periodically updated by the security vendors.

With the inclusion of firewall to secure the network, the intrusion detection systems provide another layer of protection to the network by sensing the network traffic and protect the network from known attacks

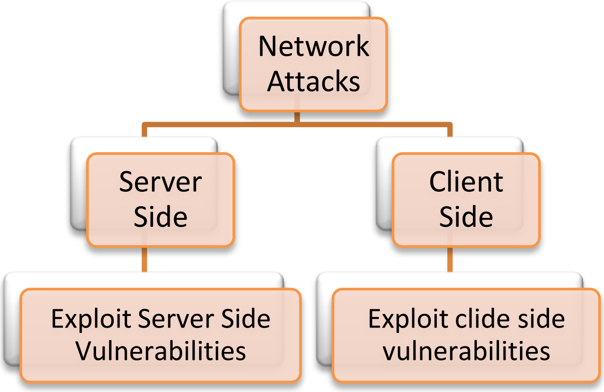

Network Attacks
Server Side
Client Side
Exploit Server Side Vulnerabilities
Exploit clide side vulnerabilities

Figure 1: Network Security Attacks

### 1.2 Classifications of Intrusion Detection System

- Host Based IDS
- Network Based IDS

### 1.2.1 Host Based IDS

Host based IDS consists of software or AGENT components, which exist on Server, Router, Switch or Network appliance. The agent versions must report to a console or can be run together on the same Host. This is NOT the preferred method though.

### 1.2.2 Network Based IDS

Network based IDS captures network traffic packets (TCP, UDP, IPX/SPX, etc.) and analyses the content against a set of rules or signatures to determine if a possible event took place. False positives are common when an IDS system is not configured or “tuned” to the environment traffic it is trying to analyse. Network Node is merely an extended model of the networked IDS systems adding aggregated and dedicated IDS servers on each NODE of a network in order to capture all the networked traffic not visible to other IDS servers.

### Significance of the research:

As internet is growing, more and more innocent users are connected through internet. When they browse any malicious web pages, they are not aware about what kind of infection can be occurred on their computer which can steal any useful and critical information. The attacks spreading through the exploitation of client side applications are the major security concern for many researchers and product oriented companies. But how to protect from these attacks, what is the technology which can be used and enable us to protect the end users from these security attacks. The answers of all these questions can be given in terms of client honeypots which is emerging area of research to defend these security attacks and to study the behaviour of the attackers to protect the end user computer.

## II. BACKGROUND AND RELATED WORK

Broadly the methods to detect the malicious web pages can be categorized into three categories known as signature based detection technique, state based detection and machine learning based approach for detection of malicious web pages.

### a) Signature Based Approach

In this technique, the detection is purely based on the signature database of the security devices. They use to detect the security attacks based on the known signatures. Signatures can be from some well-known Intrusion Detection Systems (IDS) or anti-virus applications.

The main drawback of this approach is that some classes of attacks can be missed by this technique as when there is no signatures into their databases which pertains to a new class of attacks for those detection methods, thereby it can miss the chances of detection of those kind of attacks. Moreover, unknown attacks are not covered by signatures so they are missed.

### b) State Based Approach

This approach is commonly used in case of high interaction client honeypot which is one of emerging research area in terms of detection of malicious web pages by complete monitoring the system changes. Any unauthorized state changes by the web pages is good indication of malicious infection which can be later determined by deep investigation by applying analysis techniques such static and dynamic analysis approached for malware analysis collected on high interaction client honeypots.

In the Strider HoneyMonkeys system, a monkey program loads a browser, instruct it to visit each URL and wait for a few minutes for downloading process. The state changes in the system are then detected against unauthorized creating executable files or registry entries in the system [9]. Moreover, to detect drive-by-download attack, Moshchuk, Bragin, Gribble and Levy used event triggers. They created some trigger conditions to track unauthorized activities in process creation, file system and registry system. The trigger conditions also include any event that makes browser or the system crash. During visitation, if an URL makes a trigger fire, it is classified as unsafe [10].

### c) Machine Learning Approaches

Seifert, Welch and Komisarczuk [11] proposed a novel classification mechanism to detect malicious web pages. Hou, Chang, Chen, Laih and Chen proposed a machine learning approach to detect malicious web content [12]. The key point in this research is the method used to choose features according to the usages of DHML (Dynamic Hypertext Markup Language) knowledge.

To detect malicious web pages, Bin, Jianjun, Fang, Dawei, Daxiang and Zhaohui [13] proposed the concept of abnormal visibilities. According to their studies, malicious web pages are usually changed in their display modes in order to be invisible or almost invisible.

Ma, Saul, Savage and Voelker [14] pinpointed a new approach to detect malicious web pages named lightweight URL classification.

Chia-Mei, Wan-Yi and Hsiao-Chung [15] proposed a model to detect malicious web pages based on unusual behaviour features such as encoding, sensitive key word splitting and encoding and some dangerous JavaScript functions.

## III. IMPLEMENTATIONS

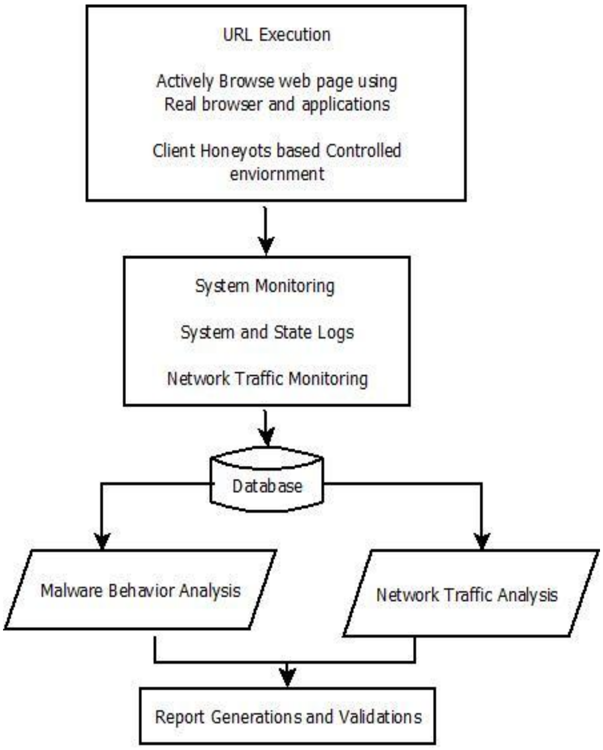

URL Execution
Actively Browse web page using Real browser and applications
Client Honeyots based Controlled enviornment
System Monitoring
System and State Logs
Network Traffic Monitoring
Database
Malware Behavior Analysis
Network Traffic Analysis
Report Generations and Validations

Figure 1: System Architecture of Implementation

As shown in the above figure, we execute the URL in a controlled environment known as high interaction client honeypot through which the execution of the URLs is performed using real browser and real client side applications. During the execution, the system behavior is being monitored to detect the abnormal system changes such at registry level or process level. Then malware is being extracted based on system state monitoring, we have seen that it is trying to steal the user's information after disabling the Anti-virus and firewall of the system. The complete experimental results are being discussed in the following section. Those collected malwares can further be studied with the help of various analysis techniques. Following are the building blocks of the system for analysis of web pages in controlled environments.

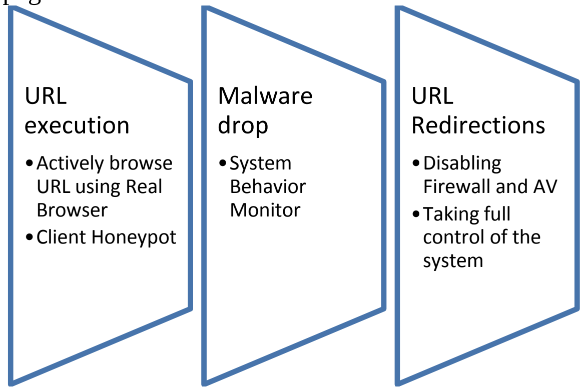

URL execution
• Actively browse URL using Real Browser
• Client Honeypot
Malware drop
• System Behavior Monitor
URL Redirections
• Disabling Firewall and AV
• Taking full control of the system

Figure 2: Process Flow of system

- **Client Honeypot**

High Interaction Client Honeypot: the ultimate goal of client honeypots is to identify the malicious infections in the web pages through the monitoring of URL during its live execution. Client Honeyots should be able to detect the known and known malicious threats in the malicious web pages which exploit the client side applications.

- **System Monitoring**

Here with the help of system monitoring, we are monitoring the complete profile of the system for any changes/modifications into the system. Those changes can be in terms of processes, registry, and network changes. Then complete monitored data is transferred from the visitor machine to base operating system for further analysis of the collected data. The collected data in the form of malwares are further analyzed for their behavior. Following the monitoring tools are used for monitoring the activities of the visitor machine.

| Tool used | Purpose | Output data |
|---|---|---|
| TCPDUMP [16] | Network monitoring | PCAP data |
| RegMon [17] | Registry monitoring | Registry changes |
| ProcMon[18] | Process Monitor | Processes |

Table 1.1 System Monitoring Tools

**Behavior Analysis:**

Broadly analysis techniques can be categorized into two categories known as static or code analysis and behavior or dynamic analysis technique. In static analysis technique the code of the malware is being studied and analyzed whereas in case of the behavior analysis the complete run time behavior of the malware is being observed. Behavior analysis is technique with the help of which we can analyze the run time behavior of the malwares. Here after the extraction of the malwares from the monitored data, we are performing the dynamic analysis of the malware for their modifications into the system. Behavior analysis. Also the first level analysis of network traffic is being performed for open source tools. Following the tools used during the dynamic analysis of the collected data.

| Tool Used | Purpose | Applied on which data set |
|---|---|---|
| WIRESHARK[19] | Packet capturing and analysis | PCAP |
| SNORT[20] | Intrusion Detection | PCAP |
| Anubis[21] | Malware analysis | Malwares |

Table 2: Tools for behavior analysis

**A. Working Flow of the System:**

Here we present the working flow of the implemented system. As shown in following flowchart, the web pages is being firstly executed through the controlled enviornment in the form of high interaction client honeypots using real browser and applications. For this puporse, we have used window XP service pack2 operating system and Internet explorer 6 for the URL executions. During the visitation of the web pages in the high interaction enviornment, the compelete system monitoring has been performed which logs all the activities performed by the web pages. The changes in the form of states of the virtual machine have been taken in thr form of snapshots of the virtual machine. These state changes are being monitoed in the high intection client honeypots. As discussed, any unathorized state changes by the web pages into the client honeypot machine is a indication of malicious infection itno it. The malware which is a malicious code is being extracted and further analysis techniques has been applied for deep analysis of malwares which tells us the kind of behavior of the malware. Once the identity of the malware is confirmed which determine the malicious code executed on the client machine by activerly visitation of those web page. The dropped malwares codes in the victim machine can be taken from file system changes or it can be extracted from the network traffic data. The dumps of the network traffic in the form of PCAP data is being saved into the base machine and further deep packet inspections algorithms have been applied on it to extract the inferential informations of the exploits. Once the inferential informations have been extracted which give us the proof of exploitations on the client machine of the innocent users. These kind of malwares are known as drive by download malwares which exploit the user’s system without his concern.

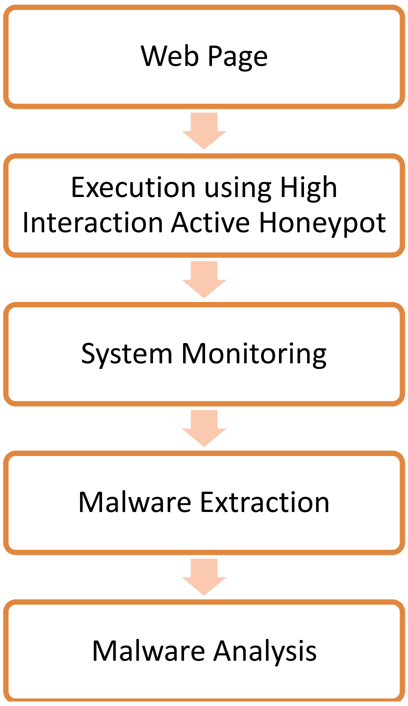


**Figure 3: Operation flow of the implementation**

**Phases of detection of malicious URL:**

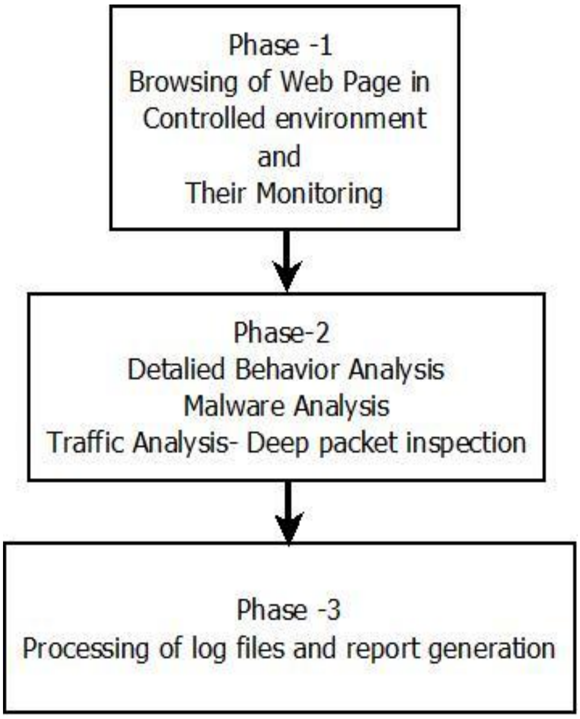

Phase -1
Browsing of Web Page in Controlled environment and Their Monitoring
Phase-2
Detalied Behavior Analysis
Malware Analysis
Traffic Analysis- Deep packet inspection
Phase -3
Processing of log files and report generation

**Figure 4: Different Phases of Execution**

Figure 3 depicts the functional modules and different phases of system implementation during which the determination of malicious has been performed. As shown, first phase is the execution and actively visitation of web pages through the client honeypot controlled environment. Here controlled environment we mean that the isolated environment which does not affect the other system in the network if there is any infections occurs. During the execution environment all the system changes have been monitored and logged as well as network traffic monitoring. Malware dropped on a victim machine which is a client honeypot machine is being analyzed with the help of analysis technique known as dynamic analysis or behavior analysis.

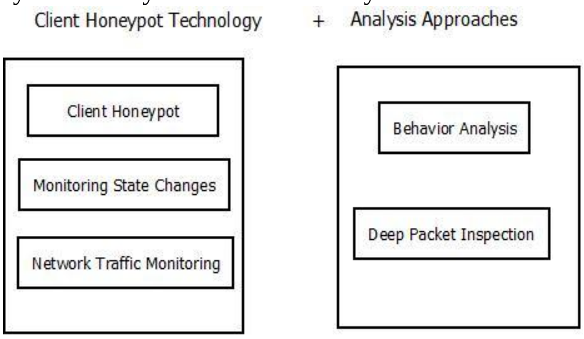


Figure 5: Client Honeypot with Behavior Analysis

**Observation**

Following observations were recorded while visiting website http://xxx.in,

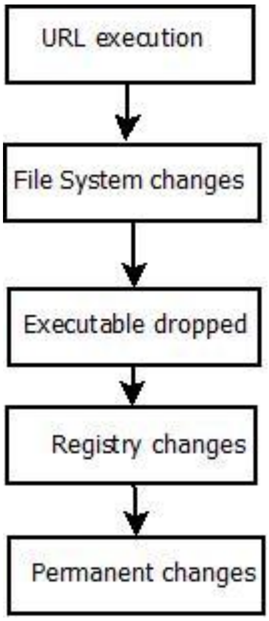
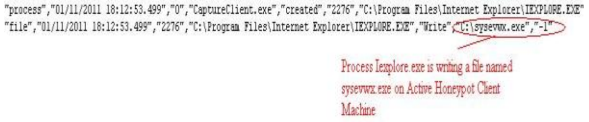


Figure 6: Observations during URL executions

Following listing displays the output of file system changes



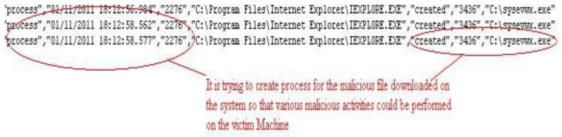
```
"process","01/11/2011 18:12:53.499","0","CaptureClient.exe","created","2276","C:\Program Files\Internet Explorer\IEXPLORE.EXE"
"file","01/11/2011 18:12:53.499","2276","C:\Program Files\Internet Explorer\IEXPLORE.EXE","Write","C:\sysevwx.exe","-1"
```

Process Iexplore.exe is writing a file named sysevwx.exe on Active Honeypot Client Machine

Figure 7: File system changes

The log line displayed in figure 8 show that the creation of a new executable files name sysevwx.exe.The first few lines show that iexplore.exe process.exe) Internet explorer application is writing to this file and the last line represents the final creation of this file.

In the next step the file sysevwx.exe is executed and more modifications to the client system are monitored.



```
'process","01/11/2011 18:12:56.984","2276","C:\Program Files\Internet Explorer\IEXPLORE.EXE","created","3436","C:\sysevwx.exe"
'process","01/11/2011 18:12:58.562","2276","C:\Program Files\Internet Explorer\IEXPLORE.EXE","created","3436","C:\sysevwx.exe"
'process","01/11/2011 18:12:58.577","2276","C:\Program Files\Internet Explorer\IEXPLORE.EXE","created","3436","C:\sysevwx.exe"
```

It is trying to create process for the malicious file downloaded on the system so that various malicious activities could be performed on the victim Machine

Figure 8: Executables dropped on system

Then changes were done to the Microsoft windows registry that honeyclient monitors as well, as soon as the malicious binary SYSEVWX.EXE is executed, it sets a number of registry entry for example disable firewall settings that otherwise would block further action taken by malware.

Then it makes changes to system.ini for making the above changes permanent after startup.

### D. Development of automated script for analysis of network traffic through signature based detection tools:

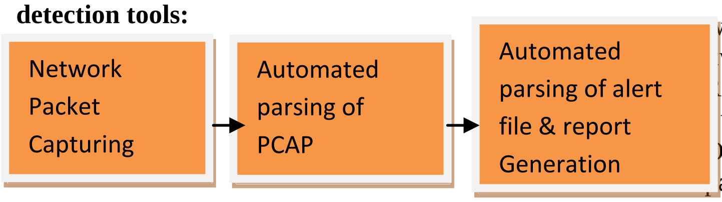

Network Packet Capturing
Automated parsing of PCAP
Automated parsing of alert file & report Generation

**Figure 9: Automated packet capturing, processing and report generation.**

**Research Tools used:**

- **NIDS (Network Intrusion Detection System)**

For the analysis of traffic captured on client honeypots, the open source network intrusion detection system known as SNORT is being used which is popularly used by the research community. Snort is a NIDS that employs both a signature and anomaly based detection system that is based on a rules database. The choice to use Snort is because it is freely available and most of the research community is using this tool. The Snort open source project is the most well known open source NIDS product available today.

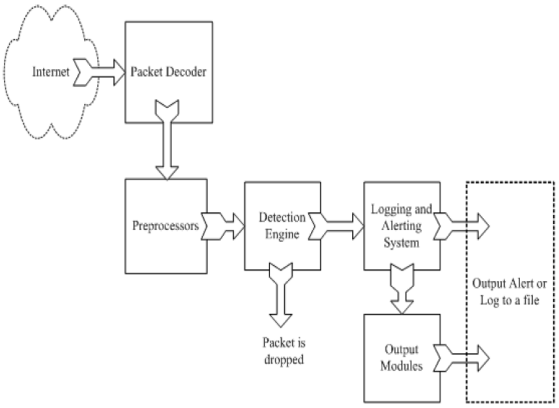


**Figure 10: SNORT Engine Components**

- **IPS ( Intrusion Prevention System)**

Snort, in addition to being IDS, can also be customized as an IPS. However, there are several issues with this option which we will solve during the research implementation of the solution.

- **SNORTALOG**

SnortALog is a script that summarizes snort logs making it easy to view any attacks against your network.

ve develop the script for automate analysis of work traffic collected during the execution of L. Following is the algorithmic steps followed the development:

- )nline packet capturing through standard acket capturing tools.
- Automated processing of PCAP through SNORT IDS and generation of alert file.
- Automated submission of alert files through snortalog for generation of report like top attack statistics.

Skelton of the code developed:

```
#!/bin/bash
#Attack Report Generation and categorization Code#
# Search the snort_full alert file in $1 directory which is given as 1st argument through command line mode. #
for f1 in `find $1 -name snort_full `; do
#Print the path of snort_full file#
	echo $f1
# To go into directory of snortalog.pl which is given as second command line argument#
		cd $2
#Process the file through snortalog parser and save the report in report directory#
	cat $f1 | perl snortalog.pl -r -g gif -o /root/Desktop/report/report.html -report
	#mkdir /root/Desktop/report-file
	#mv /tmp/*.html /root/Desktop/report-file/
Done
```

During the process of implementation of automated code, we submit all the captured traffic through it and generate the report. Our findings are that automated generation of attacks classification is very important and plays a major role for system and security engineers so that by looking at the report, they can take the remedial actions for their organizational network.

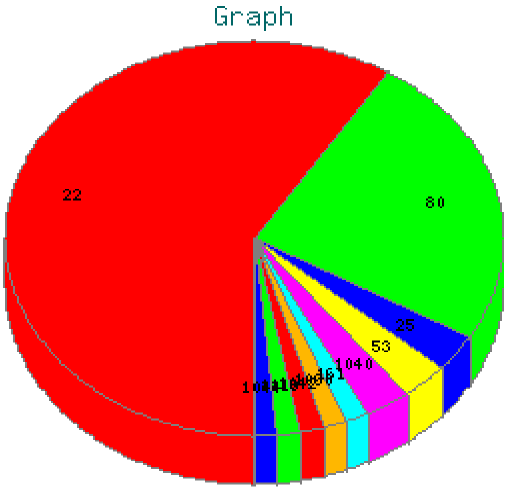


Figure 11: Port-wise distribution of attacks

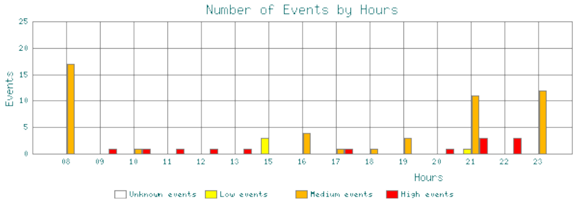


Figure 12: Hourly distribution of events

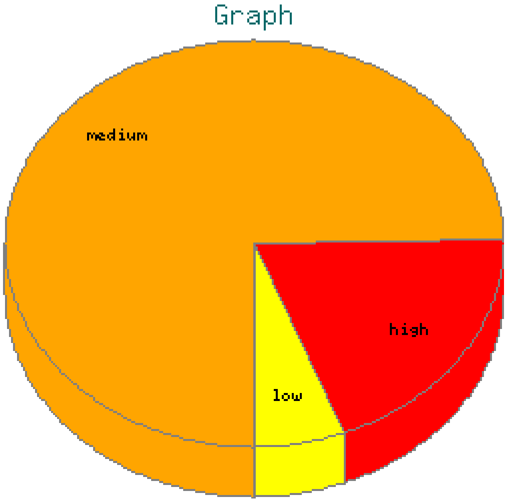


Figure 13: Severity based distribution of attacks.

## IV. Conclusion

In this paper, we try to solve the problem of malware detection embedded into malicious URL with the help of high interaction client honeypots and then try to present the behavior analysis of the URL which tries to override the signature based detection approaches such as firewalls and anti-virus of the client machine. The detected malware try to make some changes in the system level and gain access of the user's machine through automated implantation of the malwares into the system. We conclude that signature based detection approaches are not well suited for malware detection. At the end we try to generate the automated report of the network traffic captured during the active visitation of URLs. The collected network traffic in the form of PCAP data is being automated processed through the snort IDS engine and generate the alert file. Then we processed the alert file through popular tool such snortalog and generate the report which is well suited for any system and network administrators.

## V. Acknowledgements

I would like to acknowledge Ms. Harpreet Kaur for her support and guidance in writing this research paper

## References

[1] http://www.sans.org/reading-room/whitepapers/detection/intrusion-detection-systems-definition-challenges-343

[2] Van Lam Le, *Two-Stage Classification Model to Detect Malicious Web Pages*, 2011 International Conference on Advanced Information Networking and Applications.

[3] K. Wang, "*Honeyclient*," vol. 2007, no. 1/2/2007, 2005, p. available from http://www.honeyclient.org/trac; accessed on 2 Janurary 2007.

[4] Y.-M. Wang, D. Beck, X. Jiang, R. Roussev, C. Verbowski, S. Chen, and S. King, "*Automated web patrol with strider honeymonkeys: Findingweb sites that exploit browser vulnerabilities,*" in *13th Annual Networkand Distributed System Security Symposium*. San Diego: Internet Society, 2006.

[5] A. Moshchuk, T. Bragin, S. D. Gribble, and H. M. Levy, "*A crawlerbased study of spyware on the web,*" in *13th Annual Network and Distributed System Security Symposium*. San Diego: The Internet Society, 2006.

[6] C. Seifert and R. Steenson, "*Capture - honeypot client*," 2006, pp. available from https://www.client--honeynet.org/capture.html; accessed on 22 September 2007.

[7] Supinder Kaur et al, *Malware Identification Embedded into Malicious Websites Using Client Honeypot Based on Hybrid Detection*, International Journal of Computer Science and Communication Engineering Volume 3 issue 1(February 2014 issue) .

[8] Supinder Kaur et al , *Client Honeypot Based Malware Program Detection Embedded Into Web Pages*, *Int. Journal of Engineering Research and Applications* www.ijera.com *ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.849-854*

[9] Y.-M. Wang, D. Beck, X. Jiang and R. Roussev, *Automated WebPatrol with Strider HoneyMonkeys: Finding Web Sites that Exploit Browser Vulnerabilities*, IN NDSS (2006).

[10] E. Moshchuk, T. Bragin, S. D. Gribble and H. M. Levy, *A crawler based study of spyware on the Web*, (2006).

[11] http://www.spybye.org/

[12] B. Garrett, H. Travis, I. Micheal, P. Atul and B. Kevin, *Social networks and context-aware spam, Proceedings of the ACM 2008conference on Computer supported cooperative work*, ACM, SanDiego, CA, USA, 2008.

[13] D. Gollmann, *Securing Web applications*, Information Security Technical Report, 13 (2008), pp. 1-9.

[14] J. Ma, L. K. Saul, S. Savage and G. M. Voelker, *Beyond blacklists:learning to detect malicious web sites from suspicious URLs, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, Paris, France, 2009.

[15] C. Chia-Mei, T. Wan-Yi and L. Hsiao-Chung, *Anomaly Behavior Analysis for Web Page Inspection, Networks and Communications, 2009. NETCOM '09. First International Conference on*, 2009, pp. 358-363

[16] www.tcpdump.org/

[17] en.wikipedia.org/wiki/Process_Monitor

[18] en.wikipedia.org/wiki/Process_Monitor

[19] www.**wireshark**.org/

[20] www.snort.org

[21] https://**anubis**.iseclab.org/

[22] J. Mehdi, *Some Trends in Web Application Development, 2007 Futureof Software Engineering*, IEEE Computer Society, 2007.

[23] C. Chia-Mei, T. Wan-Yi and L. Hsiao-Chung, *Anomaly Behavior Analysis for Web Page Inspection, Networks and Communications, 2009. NETCOM '09. First International Conference on*, 2009, pp. 358-363.

**Journal Papers:**

[24] Masood Mansoori and Ray Hunt, *International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.5, Sep 2011, AN ISP BASED NOTIFICATION AND DETECTION SYSTEM TO MAXIMIZE EFFICIENCY OF CLIENT HONEYPOTS IN PROTECTION OF END users*

***Thesis:***

[25] *Yaser Alosefer, Analysing Web-based Malware Behaviour through Client Honeypots Cardiff University School of Computer Science & Informatics, Feb-2012*