Malware Behavioral Analysis System: TWMAN

Hsien-De Huang
National Center for High-Performance Computing
National Applied Research Laboratories
Tainan, Taiwan
TonTon@nchc.narl.org.tw

Chang-Shing Lee
Dept. Computer Science and Information Engineering
National University of Tainan
Tainan, Taiwan
leecs@mail.nutn.edu.tw

Hung-Yu Kao
Dept. Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan
hykuo@mail.ncku.edu.tw

Yi-Lang Tsai Jee-Gong Chang
National Center for High-Performance Computing
National Applied Research Laboratories
Tainan, Taiwan
{yilang, changjg}@nchc.narl.org.tw

***Abstract*—Malware is an important topic of security threat research. In this paper, a behavioral malware analysis system TWMAN was presented. This study focuses on using real operation system (OS) environment to analysis malware behavioral. Many researchers try to use virtual machine (VM) system to monitor the malware behaviors. These malware samples will only compromise the virtual operating system or virtual machine, which cannot reflect in the real operating system or real environment. Therefore, some malware researchers don't want their systems to be analyzed in VM environment, because the analyzer cannot much useful information in VM environment. There are many Anti-VM techniques which are used to ward off the collection, analysis, and reverse engineering features of the VM based malware analysis platform. There are differences between these two behaviors: malware behavior in real environment and in virtual environment. Therefore, malware researcher would get inaccurate analysis results from VM based malware analysis platform. In order to retrieve correct malware behavioral information, we need flexible, adaptable, and quickly analysis environment, which could discovery malware behavioral in real operation system environment, and which can quickly restore clear operation system to analysis another malware sample. For this reason, this study developed Taiwan Malware Analysis Net（TWMAN）, a real operation system environment for malware behavioral analysis and analysis report. We believe this system would be helpful to improve the correctness of malware analysis result and reduce the loss rate of malware analysis.**

Keywords-malware behavior; behavior analysis; real os environment; TWMAN

I. INTRODUCTION

In recent years, network security events were occurred frequently. They created disasters all around the world, including internet fraud activities, and data theft, etc... Malware was the key culprit. Therefore, how to detect Malware is a very important issue for network security. Malware has the potential to harm the machine, which designed to infiltrate or damage a computer system without the owner's informed consent (e.g., viruses, backdoors, spyware, Trojans and worms) [1]. There are many security incidents arisen by botnet, which has causes series dangers recently. Botnet is not a specific malware but a method, that possibly comprised of thousands or millions hosts controlled by hackers.

In the past years, botnet-based attacks become popular and dangerous. In order to facilitate observation of botnets, many researchers have proposed separate detection schemes and detection mechanism for monitoring and defending against them. Security expert Joe Stewart revealed that in late 2007, the operators of the botnet began to further decentralize their operations, in possible plans to sell portions of the botnet to other operators [2].

Some reports as of late 2007 indicated the botnet to be in decline, but many security experts reported that they expect the botnet to remain a major security risk online [3], and the United States Federal Bureau of Investigation considers that the botnet is the major risk to increased bank fraud, identity theft, and other cybercrimes [2]. As a result, the further research on the advanced botnet designed by the attackers becomes important. It's necessary to conduct such research, so as to deal with the threat of botnet we are facing today. Otherwise, our internet will frequently be attacked by the malware in the future. This paper developed a malware behavioral analysis tool, Taiwan Malware Analysis Net (TWMAN), which can analysis the varietal malware and output analysis report. The result scan support anti-virus to detect the known or unknown malware.

This paper contains the five sections. Section II describes the literature survey of malware behavioral analysis. Section III presents the structure of Taiwan Malware Analysis Net (TWMAN). Section IV discusses the experimental results of this study. Finally, section V gives the conclusions.

II. MALWARE ANALYSIS

In this section, we begin with an introduction to the challenges we address, our technical architecture, and describe the main components in our architecture.

The proliferation of malware continues to grow up at a staggering rate. It is estimated that 250 new variants of malware introduced into the world every day [4].Malwares are used to compromise and steal the users private data by the vulnerabilities of exploiting software. In the last several years, Internet malware attacks have grown up rapidly. Especially in 2008, the malware attack becomes more and more serious [5]. Up to the present, there are only two methods for malware behavioral analysis. One is the static analysis (code analysis). The other one is dynamic analysis (malware behavioral analysis), which can analyzes the network traffic of malware behavior and monitors the infected system to find out the changed files or registers. This technique focused on obtaining reliable and accurate information from the execution of malicious programs previously. Table I concludes the two methods.

978-1-61284-060-4/11/$26.00 ©2011 IEEE

TABLE I. TWO METHODS OF THE MALWARE ANALYSIS [6].

Malware Analysis Method	
Static analysis (Code analysis)	***Dynamic analysis (Behavioral analysis)***
file and rule signature	monitor process
black and white list	monitor file changes

Malicious software is an automatically software, which damage the software programs of computer, executes the unwanted actions, , and communicates itself to the internet. It is designed to infiltrate or damage a computer system without the owner's informed consent. Common examples of malware include Trojans (which usually disguise itself in a useful or popular package, but in fact it carries a malicious payload that the victim may never be aware of), worms (on the other hand are the variants of malware that can propagate on their own. They contain built-in functionalities that exploit computer networks and file transfer mechanisms to allow themselves to self-copy and infect other machines), rootkit (a rootkit is a malicious bundle of software designed to modify the underlying operating system of an infected computer to hide other malicious programs from the user of the system), spyware (A large fraction of the malware in the wild today is designed specifically for commercial use) and botnets (which is a collection of computers, connected to the internet, that interact to accomplish some distributed task, The compromised machines are referred to as drones or zombies, the malicious software running on them as 'bot'), a botnet is a network of compromised machines under the influence of malware (bot) code. In addition, some of the malware has been found that they exhibit the similar behavioral patterns, such as the usage of specific rules or modifications of particular system files [7, 8]. According to the Symantec Internet Security Threat Report, , there were 4,696,903 active botnet computers through the first six months of 2006, which has been becoming one of the most serious threats to Internet security.

Malware behavioral analysis can determine the behavior of malware. Although this technique have become more and more popular, the anti-detection technique of malware still grows up rapidly [9]. Behavioral analysis technique can be applied to monitor the behavior of the malware that infects your computer system by network traffic [10]. Malware behavioral analysis techniques have focused on obtaining reliable and accurate information on execution of malicious programs previously [11].

Although, many malware behavioral analysis have been developed by the software companies, such as the Norman Sandbox, Virus Total and Threat Expert [12, 7, 13, 8], some malware behavior still cannot be detected for fractional exceptional malware. The reason is that those malware can distinguish that the environment they stay in is a virtual or real environment. If they find out they stay in the virtual environment, they will try to obfuscate the monitor, and this mechanism will make the analysis result to be a fault report. Making the virtual machine to crash and detecting the existence of virtual environment are two main techniques to evade the analysis of VM based analysis.

In this work, we developed a real operation system (OS) environment to analysis malware behavioral, named Taiwan Malware Analysis Net (TWMAN). In the following, we will focus on how to use this real OS environment to analysis malware behavioral and describe the system structure of TWMAN briefly. In order to verify the analysis result obtained from TWMAN is more correct, it is compared with that from sandboxs, which are VM-based and Real OS analysis technique with CWSandbox of Sunbelt Software.

III. SYSTEM STRUCTURE AND DESIGN

Figure 1. TWMAN Logo

Fig. 1 shows TWMAN logo, and Fig. 2 shows the internal structure of TWMAN, which base on real OS environment around Joe Stewart's Truman [12, 4, 14, 13]. TWMAN is an open source (GPL v2) to dynamically analyze the behavior of malware, and which is an automated behavioral malware analysis environment to analyze the malware. TWMAN is a client-server architecture and configured to run the analysis automatically. Note that the server works on the linux, while the client is Microsoft Windows, which use PXE-Boot in a real environment.

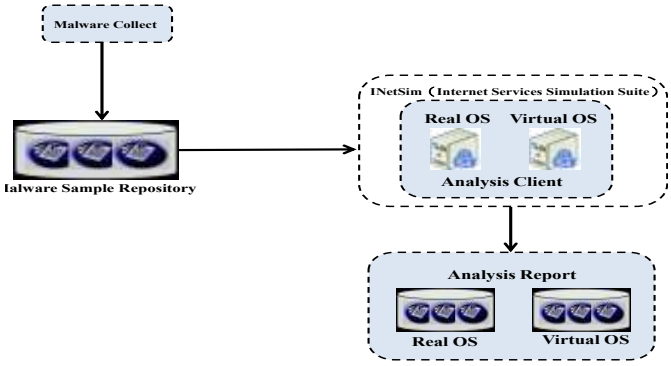

Figure 2. Internal structure of TWMAN

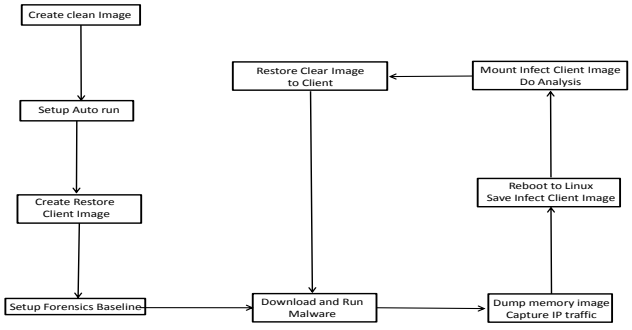

Figure 3. TWMAN flowchart

Figure 4. TWMAN boot screenshot

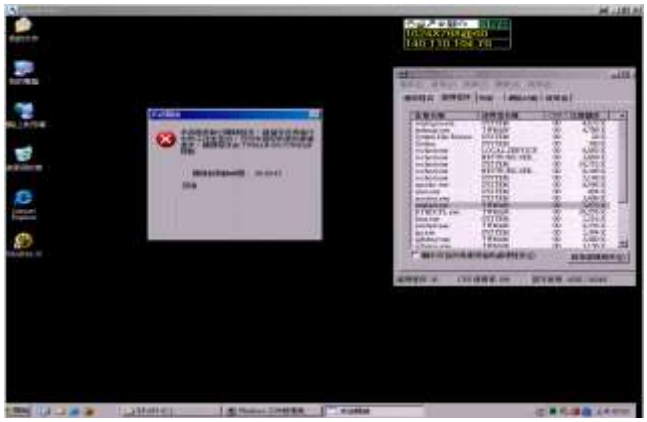

Figure 5. Wait 10 mintutes and reboot

Fig. 3 shows the flowchart for analyzing the malware behavior by TWMAN. Fig. 4 shows the screenshot of TWMAN. In the pre-work, we have to create a clear restore image of client and setup it as a forensics baseline. After the pre-work finished, the client begin to downloads malware sample from our repository of linux server and runs n client. A analysis procedure of a malware have to take about 10 minutes or more in client as can be seen in Fig. 5, which collects information about the network connect, files changes and registers changes (e.g. dump memory image, capture network connect traffic). After the procedure have finished, the client have to reboot in floppy Linux, and save the infect Microsoft Windows system as a image file in server as can be seen in Fig. 6. After that, the infect system image will mount into the server for analysis. At the same time, the infected client use the quick restore solution, as so-called open source clone system (OCS) or Clonezilla as can be seen in Fig. 7. Finally, the client will reboot and goes back into windows environment for the next analysis. At this point, TWMAN will automate generate the analysis reports. Fig. 4 is the default boot screenshot of TWMAN, which shows five option: boot local hard-disk, dd to save infect client image, dd to restore clean client image, clonezilla save clean client image, and clonezilla restore client image. Table II shows the uploaded analysis report of the TWMAN and Fig. 8 shows CWSandBox analysis report.

Figure 6. Start dd save infect client image to Server

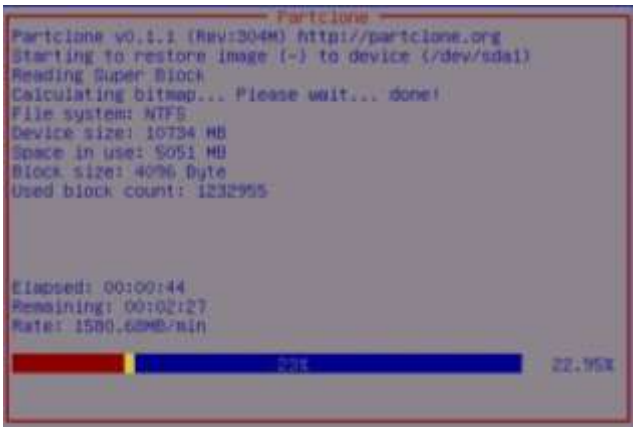

Figure 7. Use Clonezilla restore clear OS

Figure 8. CWSandBox analysis report

IV. EXPERIMENTAL RESULTS

In order to test and verify that the real environment of our approach can get correct malware behavioral information, the malware are analyzed on two sandbox system, which are VM-based and Real-OS analysis technique with CWSandbox. We choose 4840 malwares (please see Fig. 9) to analyze the malware behavior in TWMAN and CWSandBox. All the reports are presented on the website, as shown in Fig. 10.

TABLE II. TAIWAN MALWARE ANALYSIS REPORT.

```
Taiwan Malware Analysis Net, TWMAN - Analysis Report
>> Summary report for 7d99b0e9108065ad5700a899a1fe3441 created at Jan
12 14:06:37 CST 2011 <<

>> Host file changes <<
>> Registry Run Key changes<<
+++ /forensics/7d99b0e9108065ad5700a899a1fe3441/run-rip.txt    2010-07-
12 05:05:06.000000000 +0800
+Cryptographic Service|C:\WINDOWS\system32\enjqtk.exe
>> Registry Service Key changes <<
+{33F3B709-064F-4FF7-95BD-434D50D67CCC}
+{DAACA8DA-DD27-4F5F-8163-1AE2BF76188D}
>> ssdeep info (  Fuzzy Hashing )<<
192:5E0HVFnXlKyFMOJd1BPSXPe2T7GCkE80XuPH:KwznVKyqCvPSryi
Zu/,"7d99b0e9108065ad5700a899a1fe3441"
7d99b0e9108065ad5700a899a1fe3441 matches
/forensics/ssdeep.db:883658b543256aa701707ac85a8e3306 (58)
7d99b0e9108065ad5700a899a1fe3441 matches
/forensics/ssdeep.db:547aac09b48d57cc8b1542ca1c6c6402 (75)
7d99b0e9108065ad5700a899a1fe3441 matches
/forensics/ssdeep.db:96924b5583a0fa34a66d81257939a5d6 (65)
7d99b0e9108065ad5700a899a1fe3441 matches
 matches /forensics/ssdeep.db:33d8d205001b83f96a9480ce51d54b60 (60)
7d99b0e9108065ad5700a899a1fe3441 matches
/forensics/ssdeep.db:1bebbe1e398adff9baab1962ec5827e0 (100)

>> Connect IP <<
IP 192.168.0.110 > 201.196.11.244: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.171.67: ICMP echo request, id 512, length 41
IP 192.168.0.110.123 > 207.46.197.32.123: UDP, length 48
IP 192.168.0.110 > 201.196.46.244: ICMP echo request, id 512
IP 192.168.0.110 > 201.196.28.60: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.38.41: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.207.12: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.202.32: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.13.108: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.168.206: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.85.157: ICMP echo request, id 512, length 41
IP 192.168.0.110 > 201.196.108.43: ICMP echo request, id 512, length 41

>> NPASCAN v1.7  <<
Current User : TWMAN-SINGLE-01\Administrator
Current IP   : 127.0.0.1
Start Time   : 12 Jan 2011 14:13:01
------------------Start Scan-----------------------
掃瞄完成!!未偵測到相關惡意程式!
------------------End Scan-----------------------

>> CWSandBox VirusScan Report  <<
VSCAN Version:3.2.1861.2 (Feb 22 2009  19:30:04);run at:: Jul 11 12:54:40
2010
defs version: 5444 (2009-10-12T17:47:12)
command line: c:\SBScanV3\vscan /l c:\virus.txt /def c:\SBDefsV3
C:\WINDOWS\system32\sandnet.exe
 [      53], Malicious      ,CRC8_BL  ,            Trojan.Win32.Generic!BT,
4150696,C:\WINDOWS\system32\sandnet.exe
1 objects processed in 0 secs, 0 fps
1 threats detected, 0 suspicious files

>> Advanced Intrusion Detection Environment<<
Summary:
  Total number of files:          28292
  Added files:                    333
  Removed files:                  664
  Changed files:                  51
---------------------------------------------------------------------------------
Taiwan Malware Analysis Net, TWMAN - Analysis Report
Last Update 2011-01-01"
E-Mail : TWMAN@nchc.narl.org.tw
Developed By TonTon | http://TWMAN.nchc.org.tw "
```

A. *Analysis Result 1*

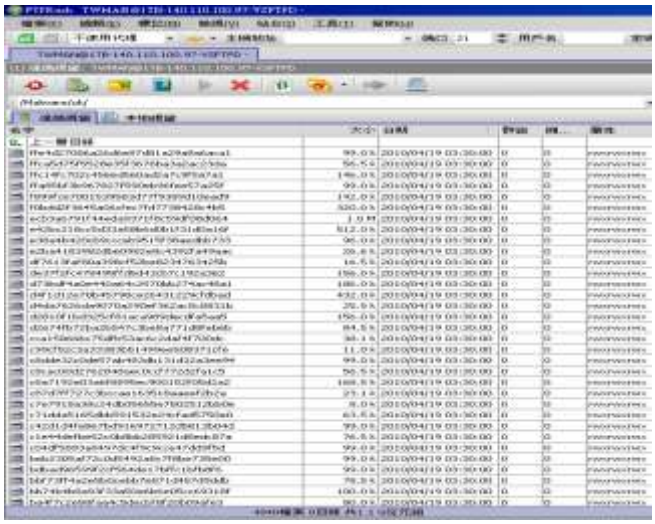

Figure 9. 4840 malwares in our analysis experimental

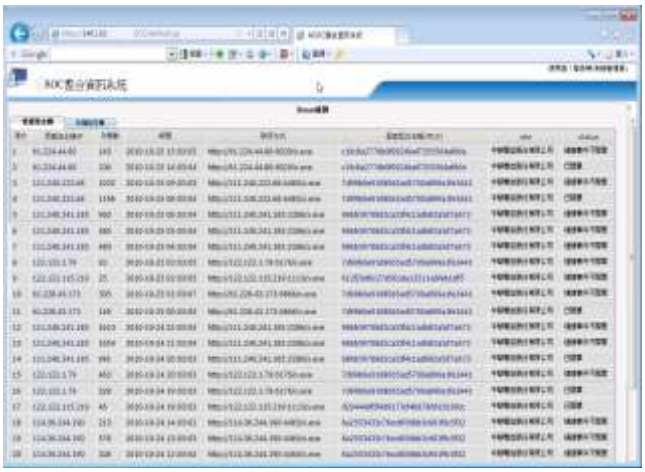

Figure 10. All reports presented on the website: twman.nchc.org.tw.

From the observation of the report, we found an interested result. TWMAN can detect many suspicious behavior that is not detected by the VM-base, sandbox environment, and upload to Virus Total from the same malware (MD5 is 98eeb5e2e889ddc8307a4f49fe277771). For example, TWMAN found that the malware changes many files, registers, and connect to internet. Those ip are 203.69.113.26, 199.7.48.190 and 199.7.51.190. We also found the ip 199.7.48.0-199.7.63.255 are VeriSign Global Registry Services, which is the trusted provider of Internet infrastructure services for the networked world, and another ip 203.69.113.26 is HINET ISP ADSL (please see Fig. 11-14).

B. *Analysis Result 2*

From the observation of the another report, we have also found an interested result. TWMAN can detect many suspicious behavior that is not detected by the VM-base, sandbox environment (CWSandBox), and upload to Virus Total from the same malware (MD5 is 08f95d2993506a939c374d47e31fad3c). For example, TWMAN found that the malware changes many files, registers, and connect to internet. Those ip are 203.69.113.43,

199.7.51.190 and 199.7.52.190 (please see Fig. 17-20). We also found the ip 199.7.48.0-199.7.63.255 are VeriSign Global Registry Services, which is the trusted provider of Internet infrastructure services for the networked world, and another ip 203.69.113.43 is HINET ISP ADSL (please see Fig. 15-18).

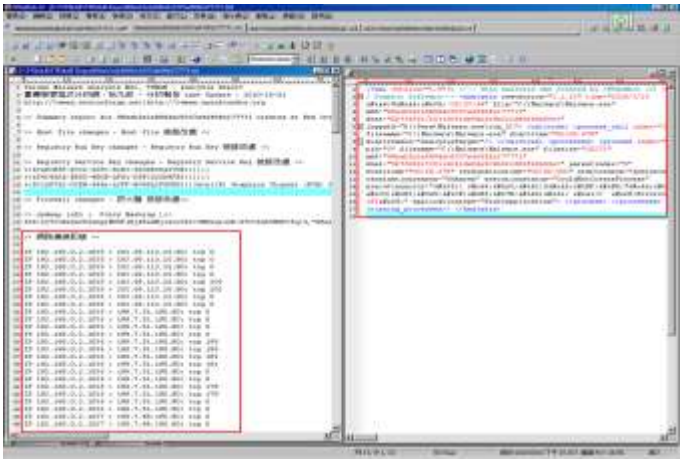

Figure 11. TWMAN and CWSandBox analysis report

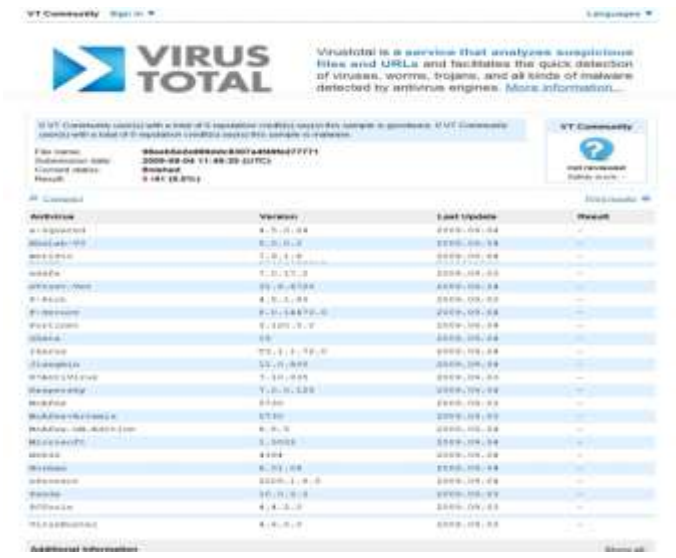

Figure 12. Virustotal analysis result

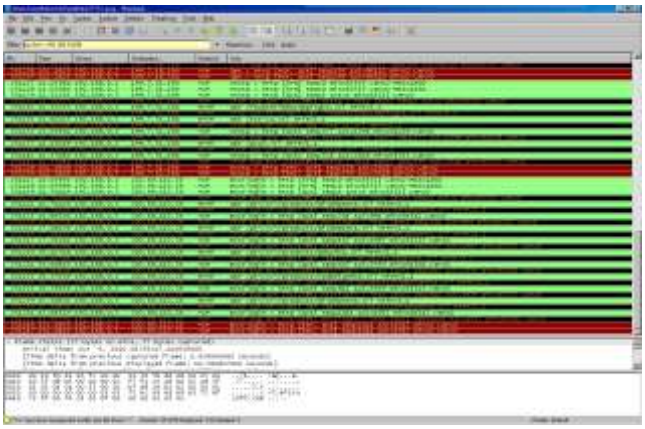

Figure 13. Tcpdump result

Figure 14. Whois result

Figure 15. Virustotal analysis results

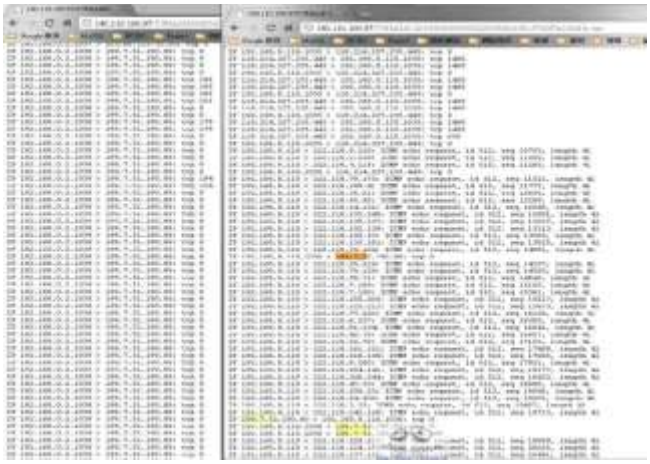

Figure 16. Analysis report of TWMAN and CWSandBox

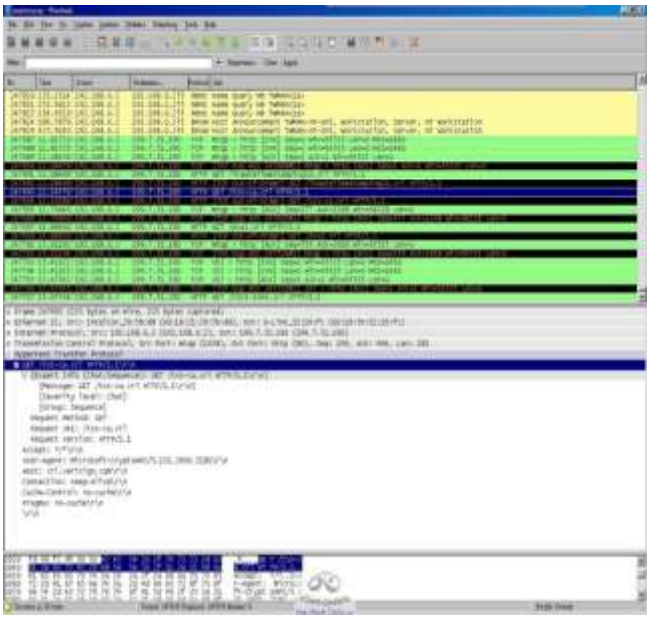

Figure 17. Tcpdump result

IP 位址：203.69.113.43
PTR：--

註冊局 WHOIS 主機：whois.apnic.net:43

% [whois.apnic.net node-4]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html

person: HINET Network-Adm
address: CHTD, Chunghwa Telecom Co., Ltd.
address: Data-Bldg. 6F, No. 21, Sec. 21, Hsin-Yi Rd.,
address: Taipei Taiwan 100
fax-no: +886 2 2344 2513
fax-no: +886 2 2395 5671
e-mail: network-adm (at) hinet.net
nic-hdl: HN27-AP
remarks: same as TWNIC nic-handle HN184-TW
mnt-by: MAINT-TW-TWNIC
changed: hostmaster (at) twnic.net 20000721
source: APNIC

person: HINET Network-Center
address: CHTD, Chunghwa Telecom Co., Ltd.
address: Data-Bldg. 6F, No. 21, Sec. 21, Hsin-Yi Rd.,
address: Taipei Taiwan 100
country: TW

Figure 18. whois result

V. Conclusion and Future Works

Although, there are many research issues on malware behavioral analysis, but according to our experiment result, we believe different analysis environments have different behavior of malware. So, we release this tool on Source Forge (http://twman.sourceforge.net) and Open Foundry (http://twman.openfoundry.org), and hope to get more research partners (please see Fig. 19). Surprisingly, we get help and require from Military Police Command Forensic Science Center (http://afpc.mnd.gov.tw/english/index.aspx) and the Investigation Bureau of the Ministry of Justice in Taiwan (http://www.mjib.gov.tw/en/). Up till now, TWMAN have been build in three partners and run smooth. In additional, we are developing the multi-clients and multi-OS and integrate malware collection of TWMAN (please see Fig. 20), and hope this system will get more correct analysis result and more security solution.

Figure 19. Project state of TWMAN

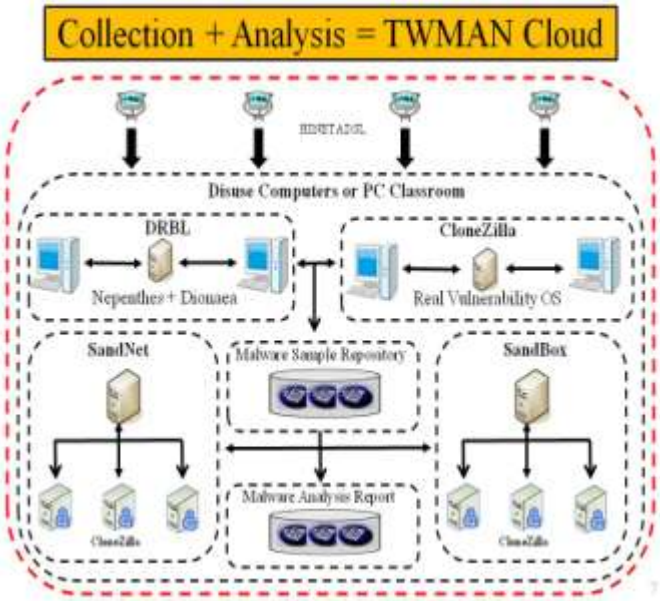

Figure 20. The strucrt of future TWMAN

By following the development processes of this study, there are several specifications of markup languages for creation of information resources as, Web Ontology Language (OWL) and Fuzzy Markup Language (FML) [15], Which can be expanded to solve more complex, practical problems and exploited to malware behavioral and ontologically depict the real world knowledge used by agents to infer additional information and provide advanced services to end users. Examples include through OWL to describe registry key value changes, network connection and file changes to build the ontology (please see Fig. 21) [6]. Combine FML to build the behavioral rules (please see Fig. 22), and then utilize the same pattern of behavior to determine the virus and malicious program to provide better protection.

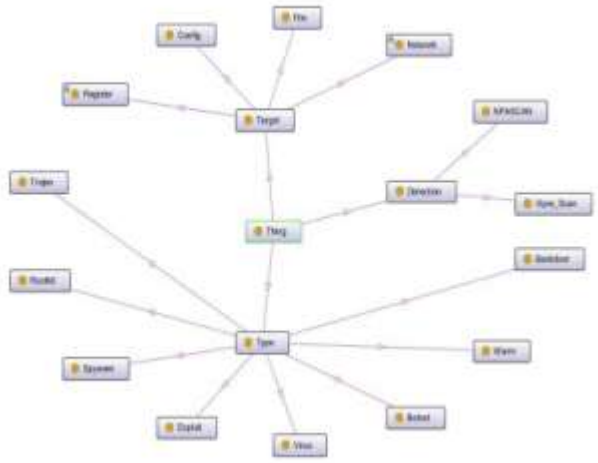

Figure 21. Use OWL to describe malware behavioural ontology

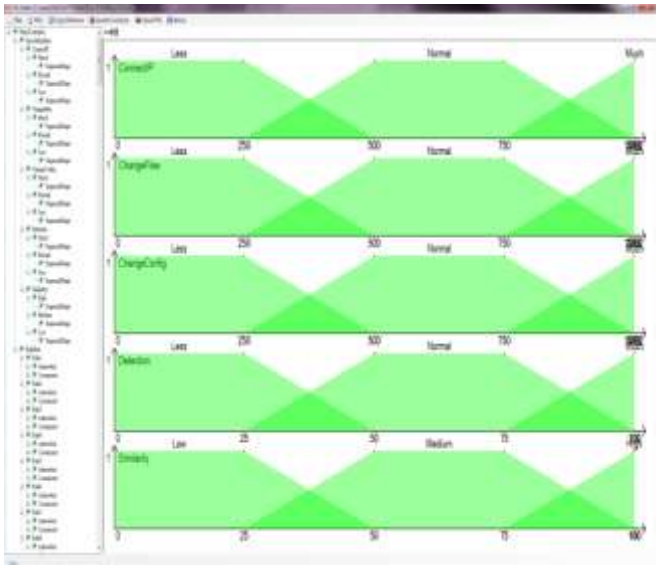

Figure 22. Use FML to build fuzzy malware behavioral rules

Figure 23. Google Custom Search API of TWMAN

Furthermore, a Web-based system can be simply developed by utilizing fuzzy ontology-based applications supported by Protégé API, Google Custom Search API, and Facebook develop API(please see Fig. 23, 24). Consequently, an inferred ontology can be considered as an intelligence knowledge base. In the future, The TWMAN can integrate with a human thinking semantic model in order to take prompt and appropriate measures. Therefore, we consider TWMAN could be a powerful tool to improve the network safety. Therefore, we consider TWMAN could be a powerful tool to improve the network safety.

Figure 24. TWMAN website (http://twman.nchc.org.tw)

ACKNOWLEDGMENT

The authors would like to thank the National Center for High Performance Center of Taiwan for financially supporting this open source research project. In addition, the authors would like to thank National Science Council of Taiwan for financially supporting this research under the Grant NSC 99-2923-E-024-003-MY3, NSC 98-2221-E-024-009-MY3, NSC 99-2911-I-024-004 and NSC 99-2218-E-492-006 .

REFERENCES

[1] Hengli Zhao, Ming Xu, Ning Zheng, Jingjing Yao, and Q. Ho, "Malicious Executables Classification Based on Behavioral Factor Analysis," presented at the 2010 International Conference on e-Education, e-Business, e-Management and e-Learning, Sanya, China, 2010.

[2] Wikipedia. *Storm botnet*. http://en.wikipedia.org/wiki/Storm_botnet

[3] F.-S. Corporation. F-Secure Reports Amount of Malware Grew by 100% during 2007. Available: http://www.f-secure.com/f-secure/pressroom/news/fs_news_20071204_1_eng.html

[4] J. Stewart, "Behavioural malware analysis using Sandnets," *Computer Fraud & Security,* vol. 2006, no. Issue, pp. 4-6, December 2006.

[5] S. Corp. Symantec Internet Security Threat Report: Trends for July-December 2007 (Executive Summary).

[6] H. D. Huang, T. Y. Chuang, Y. L. Tsai, and C. S. Lee, "Ontology-based Intelligent System for Malware Behavioral Analysis," presented at the 2010 IEEE World Congress on Computational Intelligence (WCCI2010), Barcelona, Spain, 2010.

[7] S. Software. (2007). *CWSandbox user guide v 2.1.13*.

[8] C. Willems, T. Holz, and F. Freiling, "Toward automated dynamic malware analysis using CWSandbox," *IEEE Security & Privacy,* vol. 5, no. Issue, pp. 32-39, 2007.

[9] A. Vasudevan, "MalTRAK: Tracking and Eliminating Unknown Malware," presented at the Computer Security Applications Conference, 2008. ACSAC 2008. Annual, Anaheim, CA,, 2008.

[10] G. Jacob., H. Debar., and E. Filiol., "Behavioral detection of malware: from a survey towards an established taxonomy," *Journal in Computer Virology,* vol. 4, no. Issue, pp. 251-266, 2008.

[11] K. Rieck, T. Holz, C. Willems, P. Düssel, and P. Laskov, "Learning and classification of malware behavior," presented at the Fifth Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 08), 2008.

[12] J. Clausing, "Building an automated behavioral malware analysis environment using open source software," SANS Institute Reading Room2009.

[13] J. Van Randwyk, L. Ken Chiang Lloyd, and K. Vanderveen, "Farm: An automated malware analysis environment," presented at the Security Technology, 2008. ICCST 2008. 42nd Annual IEEE International Carnahan Conference on, Prague, 2008.

[14] J. Stewart. *Truman - the reusable unknown malware analysis net*. http://www.secureworks.com/research/tools/truman.html

[15] C.S. Lee, M.H. Wang, G. Acampora, C. Y. Hsu, and H. Hagras, "Diet assessment based on type-2 fuzzy ontology and fuzzy markup language," *International Journal of Intelligent Systems,* vol. 25, no. 12, Wiley, 2010.