IOP Conference Series: Materials Science and Engineering

Review of Hybrid Analysis Technique for Malware Detection

To cite this article: Yus Kamalrul Bin Mohamed Yunus and Syahrulanuar Bin Ngah 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **769** 012075

Review of Hybrid Analysis Technique for Malware Detection

Yus Kamalrul Bin Mohamed Yunus[1], Syahrulanuar Bin Ngah[2]

[1,2]Faculty of Computing, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia

E-mail: yus.kamalrul96@gmail.com, syahrulanuar@ump.edu.my.

Abstract. Malware is a problem spread out worldwide. Current techniques to analyze these malware are static analysis technique and dynamic analysis technique. Later, the two analysis technique is combined into a technique called hybrid analysis technique. This paper discusses on the current analysis technique and introduces a new approach towards the hybrid analysis technique by introducing memory analysis technique into it. The expected outcome of producing memory analysis technique in hybrid analysis technique will be discussed in the latter part of this paper.

1. Introduction

Malware is the term used by security professionals and is commonly used in the Information Technology (IT) field in regards to any malicious software that enters a system without the authorization of the computer administrator and/or user of the said system. This term emerges after combining the word “Malicious” and “Software” thus creating the word malware. This type of attack is growing in volume and evolving in complexity. With the current explosion of Internet of Things and Internet of Everything, threats from malware is increasing and require efficient monitoring and detection of the affected device. Effective analysis method is needed to match with these data-intensive systems and its scale. Currently, there are many types of malware detector had been created by security companies and researchers; these detectors heavily depend on the techniques being implemented in them.

Several types of threads that is referred as malware is trojan, ransomware and botnet [1]. Trojan is a type of malware that mislead user of its true intent and is often disguised as a legitimate computer program. The term is derived from the famous story of a wooden horse used by ancient Greek to penetrate the city of Troy. Once activated, the Trojan malware can delete, block, modify and copy data, also it can disrupt computer and network performance. To date, Trojans can perform many actions, therefore, it can be further categorized as backdoor Trojan, exploit Trojan, rootkit Trojan, Distributed Denial of Service (DDOS) Trojan, and but not limited to proxy Trojan. Ransomware on the other hand is a type of malware that threatens and at the same time block users from accessing their data on their system by encrypting it [2]. Most of the time, the attacker will demand ransom in cryptocurrency in order to release the files back to the victim. It is divided into several types such as scareware ransomware, locker ransomware and crypto ransomware [2]. In recent years, the use of ransomware as an attack vector has grown internationally. Attacker generates a key pair and place the public key in the ransomware. The ransomware is then released to the victim. To carry out the attack, the malware generates a random asymmetric key while it proceeds to encrypt victim's data and in certain cases render the system unusable. It puts up a message to victim that includes the cyphertext and steps on how to pay the ransom. Victim will then send the cyphertext back to the attacker with the amount of cryptocurrency the attacker demand. The attacker then deciphers the cyphertext with a private key, and sends the asymmetric key to the victim. The data is decrypted using the key provided. At times, upon receiving their ransom, the attacker did not send any key to victim resulting victim with monetary loss alongside data loss. The last example is botnet, botnet is a collection of internet connected device such as PC, IOT device or mobile phones where its security has been compromised and its control is handled by malicious third parties. The usual communication channel used by botnet is either Internet Relay Chat (IRC) or Hyper Text Transfer Protocol (HTTP). Once infected, victims machine will be used to carry out massive DDOS attacks, email spam to millions of users and generate fake internet traffic. Botnet is a commodity for cyber criminal because it is rented out for a variety of purposes.

Malware analysis technique came in three classifications, static analysis technique, dynamic analysis technique and hybrid analysis technique. These techniques have been around for many years and are being used by security analyst and professional to detect malware in a device. This three analysis technique will be discussed in more depth in the next sections. Figure 1.1 shows types of malware analysis technique and the methods it uses.

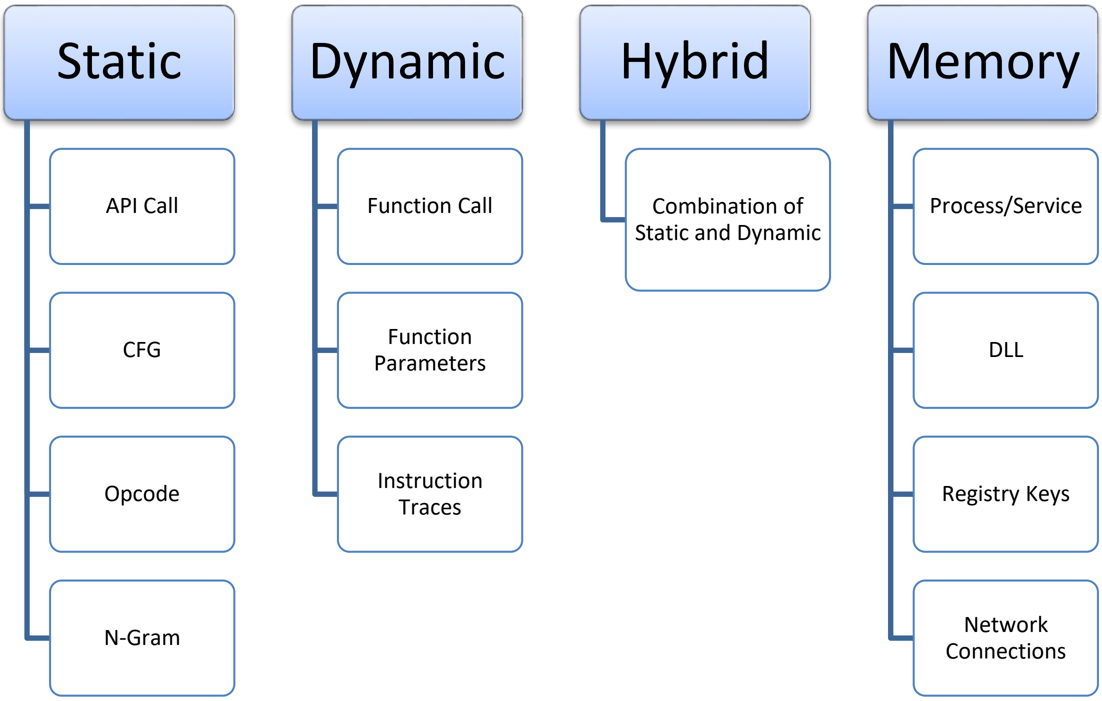

Figure 1.1 Malware Analysis Techniques

The aim of this paper is to discuss the current malware analysis technique also the approach of combining hybrid analysis technique with memory analysis technique. The proposed method in this paper will be adopted as the future work by the authors.

2. Static Analysis Technique

This analysis technique works by analyzing the code sections of the malware in order to gain knowledge on how the malware is operating [3]. This analysis procedure is done without executing the malware. Using reverse engineering tools, the malware is dissected in order to analyze its code and later, it can be re-build. Due to not executing the malware, static analysis technique consumes fewer resource and time [4]. The different tools used to perform this analysis are debugger, dissembler, decompiler and source code analyzer. The purpose of analyzing the code is to identify the inner working of the malware; the unique algorithm in the malware will help identify the malware and can be analyzed further to eliminate the malware.

This analysis technique can be further broken down into three approaches, signature based approach, permission based approach and heuristic approach. Signature based approach are used mostly on commercial anti-malware product or on-the shelf product. It extracts the malware pattern and creates a unique signature that represents the extracted malware. A code, executable or program is flagged as a malware if the signature matches with the existing malware signature. The major drawback of this technique is that it only detects existing or known malware type, if a new malware is introduced in the system where the signature is not yet available, this approach will not detect the malware. Also due to limited library of signature, most of malware is undetected.

The second approach is permission based approach. This approach is used to control by approving or denying access right of any application. When a program is installed, as default settings, it will not have any access towards data stored in the device and does not have any impact towards system security.

The last approach searches for the instruction those are not present in the infected program or code rather than searching for a certain signature. This makes finding new variant of malware easier. Some of the technique in heuristic approach is file based heuristic, rule based heuristic and generic signature analysis.

The drawback of using this analysis technique is code obfuscation. Malware that uses code obfuscation will not be detected by this technique as it can’t find the malware signature and instructions. Aside from that, malware that uses anti-debug and anti-disassembly will not be detected using this method.

3. Dynamic Analysis Technique

The analysis process is done during the execution of the malware code or the infected file. The malware code or infected file is analyzed in a sandbox environment or virtualized environment such as VMware player, Oracle Virtual Machine (VM) VirtualBox or Hyper-V environment. The environment can be a simulator, emulator, sandbox and more. The analysis can be done by monitoring certain components which is function calls, information flow, function parameter and tracing instructions. Malware or infected file need to be analyzed in an invisible environment due to some malware came with anti-virtual machine and anti-emulator technique [5]. When the malware code detects the environments, it will behave normally and will not execute any malicious activity.

A little different from static analysis technique, this analysis technique can be broken down into two approaches, anomaly-based approach and taint analysis. In order to detect malicious behavior of applications, this approach uses machine learning. Existing malware features are used to train a model for unknown malware. The tools provide deep analysis, as a result, it require more resource to operate. The drawback is application must be installed in a system in order to detect malicious software. Also, if a legitimate application invokes more system call than the usual call set in the model, the application can be falsely detected as malware.

User input can be very dangerous if it is not properly checked. Taint approach checks for variable that user input can modify. A scientific technique called “dynamic taint analysis” used by one of the application that uses this approach [4]. Information of interest will be marked as “taint”. It stays with information when used. The tracking system then tracks the movement of that taint.

The drawback of using this method is it consumes resources while being time consuming. Other than that, malware that incorporated anti-virtual machine and anti-emulator will slip pass through this method.

4. Hybrid Analysis Technique

This analysis technique is introduced to overcome the limitation available in both static and dynamic analysis technique. It starts by analyzing the signature of any malware code and continue by combining it with other behavioral pattern parameters to enhance malware analysis [3]. Due to this reason, it overcomes both the shortcoming of static and dynamic analysis technique. This increases the ability in detecting malicious software correctly. In the same time, this analysis technique has almost all of the strength of static and hybrid technique.

In detecting malware in android application, mobile sandbox is an example of hybrid analysis technique. Static analysis will analyze the APK file, user permission and identifying suspicious code. While the dynamic analysis using emulator will be used to run the suspicious APK file to check the application behavior. Figure 4.1 shows the mobile sandbox.

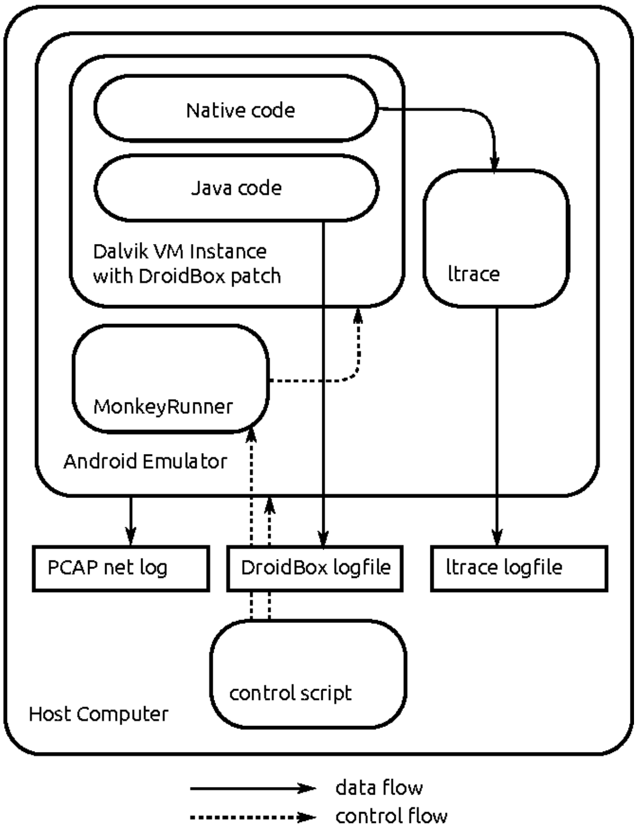

Figure 4.1 Mobile Sandbox

The limitation of using this method is this technique is resource hungry. This method is time consuming as it is running both static and dynamic technique as hybrid technique.

5. Proposed Method

Proposed method is combination of hybrid analysis technique with memory analysis technique. Although hybrid analysis technique analyze the signature of malware and then combine it with behavioral pattern parameter but it only analyzes the malware that is stored on the disk and the same malware runs in the memory. Malware that sits and run in the memory is either not checked or high probability that their signature and behavioral pattern is not the same as the classic malware that sits in the hard disk and run in the memory. Previously, computer hijacker will send malware and open a backdoor into unsuspecting user computer and start mining bitcoin, a type of cryptocurrency, without user knowledge.

Earlier 2019, ESET security team had found out that attacker injected malicious JavaScript code to website ads. Now, the website contains malicious ads. User who visited the website will have their system used for Monero or Dash cryptocurrency mining as bitcoin mining will alert user because it will skyrocket processor usage. This is all happening in the memory as the browser is running in the memory and all tabs and page is in the memory.

Memory analysis technique is able to examine malware code outside of function normal scope [3]. It uses the image of memory to analyze operating system, running programs and the general state the computer. In digital forensics on regards of memory, there are two steps which is memory acquisition and memory analysis. In acquisition, memory image is obtained by acquisition of volatile memory to non-volatile memory [8]. In analysis, the memory image is analyzed thoroughly while looking for malicious activity in the image. Memory analysis especially offline memory analysis approach is a process that can be repeated and can be followed to acquire precise digital evidence for research and court use [9]. Researchers in regards of memory forensics had proposed several approaches such as trigger-based memory analysis approach and Application Programming Interface (API) trigger-based memory analysis approach. Memory forensics technique able to monitor behavior of Malware in regards to API hooking, Dynamic-link Library (DLL) injection and hidden processes.

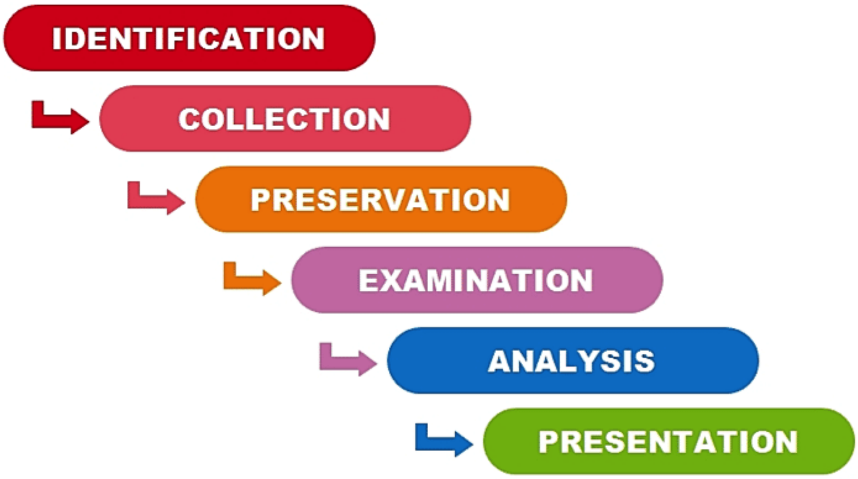

Figure 5.1 Memory Acquisitions for Digital Forensics

Malware uses API hooking to interrupt function calls. Most common type of API hooking are Import Address Table (IAT) hooks, inline API hooks, Interrupt Descriptor Table (IDT) hook, System Service Dispatch Table (SSDT) hook and Input/Output Request Packet (IRP) hook. DLL injection inserts malicious code into a legitimate process. DLL injection categories are classic DLL, registry modification injection, and Windows hooking function injection.

The process of the proposed technique starts by capturing the contents of memory by using forensics tool. After the capturing process, the memory contents will be further analyzed by using hybrid analysis technique. The process if memory capture is done manually as the automation of this process is still being explored.

Malware programs nowadays frequently contain checking criteria. This will determine the existence of certain directories or files on a machine. When the checking criteria are met, the malware runs part of its code matching with the criteria. Other type of malware may need internet connection or that a specific mutex object does not exist in order for it to execute its code [6].

There are several problems that need to be addressed during the research process. One of the problems is steps to capture the memory contents before some of the malware starts to wipe its tracks and disappear from the system. The steps to address the problem are still being explored. Other than that, anti-forensics tools are being executed within the malware to prevent memory acquisition and analysis, some went into extra lengths by planting fake artifacts in-memory [7].

The steps mentioned earlier were just a discussion on how the process will be done theoretically as the research on implementing the proposed technique will be done in future work. Therefore, during the execution phase of the future work, majority of the steps discussed will be revised to improve the execution and implementation of the proposed technique while at the same time conducting analysis on the effectiveness and performance of the proposed technique.

Expected outcome from this proposed technique is a more robust malware analysis method. The combination of hybrid analysis technique and memory analysis technique will certainly increase resource usage but it will detect more malware than only using hybrid analysis technique. This new method will be expected to detect more malware than hybrid analysis technique alone.

6. Conclusion

Malware is critical threat to a system either it is a computer mobile phone or IOT device. This paper introduces briefly on three types of malware, Trojan, ransomware and botnet. Other than that, this paper also briefly discuss on three analysis technique available for malware analysis which is static analysis technique, dynamic analysis technique and hybrid analysis technique. The drawback of each analysis technique is also provided briefly.

New proposed method for malware analysis technique is introduced. The method proposed to work by combining hybrid analysis technique and memory analysis technique. While proposing the new method, a brief introduction to memory analysis technique is also provided. Malware change their signature frequently has present many open research on how to detect them and what will be the best way to detect them. These researchers help in providing information to security professionals and researchers to tweak current analysis technique and helps in a new analysis technique creation process.

Acknowledgments

The work is funded by Universiti Malaysia Pahang under Grant ID RDU1703294. The authors would like to thank all reviewers for their constructive reviews, and Universiti Malaysia Pahang for the support.

References

[1] Jyoti Landage, Prof. M. P. Wankhade,Malware and Malware Detection Techniques : A Survey, International Journal Of Engineering Research & Technology (IJERT) Volume 02, Issue 12 (2013)

[2] Sultan, Hirra. A Survey On Ransomware: Evolution, Growth, And Impact. International Journal of Advanced Research in Computer Science vol 9, No. 2 (2018).

[3] Rao, V.R., & Hande, K. A comparative study of static , dynamic and hybrid analysis techniques for android malware detection (2017).

[4] Rami Sihwail,Khairuddin Omar and Khairul Akram Zainol Ariffin,"A Survey on Malware Analysis Techniques: Static, Dynamic, Hybrid and Memory Analysis," International Journal on Advanced Science, Engineering and Information Technology, vol. 8, no. 4-2, pp. 1662-1671, (2018).

[5] Tahir, R. A study on malware and malware detection techniques. International Journal Education and Management Engineering (IJEME) 8(2), 20–30 (2018)

[6] A. Moser, C. Kruegel and E. Kirda. Exploring Multiple Execution Paths for Malware Analysis. 2007 IEEE Symposium on Security and Privacy *(SP '07)*, Berkeley, CA, 2007, pp. 231-245 (2007).

[7] Teller, Tomer and Adi Hayon. Enhancing Automated Malware Analysis Machines with Memory Analysis. (2014).

[8] Arends, Jan, and Ing Kerstin. "Malware Analysis." (2018).

[9] Sali VR, Khanuja HK. RAM Forensics: The Analysis and Extraction of Malicious Processes from Memory Image Using GUI Based Memory Forensic Toolkit. In2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) 2018 Aug 16 (pp. 1-6). IEEE.