Malware: Types, Analysis and Classifications

By **Harjeevan Singh Gill**

Abstract—Malware is becoming an everyday challenge to detect with more complex malware designs and advancements of computing infrastructures allowing malware to easily take various forms within it. More proficient attacker now use anti-analysis techniques such as packing or obfuscation making malware harder to detect. The paper explains the types of malware, compares the types of analysis and explains how malware is classified.

I. Introduction

Malware is always ever evolving; hackers are always finding new ways to be able to deploy their code on other people's devices in order to obtain their goal. Hacker's motives are different and can be as simple as stealing data to slowing done a business to the point they cannot operate for a certain time. Ultimately malware in some way affects the performance of a device, which in turn can result in a loss of money and time. It is important to understand how malware works in order to prevent it from entering and staying in a system.

II. Types of malware

When computer viruses are deployed, it is coded to self-replicate inserting programed code into computer programs on the device. There are many types of virus that each specialise in different areas of a device. Examples include: file infected can target executable files, that can slow down programs and system files when run, browser action that redirects the user to a malicious website. Network virus that uses network connections to replicate itself on shared resources. With the aid of social engineering viruses are ever evolving as they try to avoid detection of more antivirus software. In 2000 The ILOVEYOU virus infected estimated 10 million PCs, which was a love letter written in a text file. The virus would send the letter to every email on a PC. The phising attack if success full would replace media files on the computer with its own code allowing the attacker to steal passwords so they could gain access to the users finances. It is estimated this virus collected around 15 billion dollars [12].

A worm requires computer networks in order to infect other devices and exploits the security vulnerabilities on the target device to gain access. The worm will duplicate itself on the target device and then try to gain access to another one. WannaCry is a ransom ware worm that infected millions around the world because it exploited security vulnerability if the user did not update their windows operating system. Worms usually are deployed to slow down network traffic and cause network disruption.

Unlike worms and viruses Trojan horses usually don't try to inject malicious code into files. They instead trick the user to have a fake virus removed for money by pretending to be legitimate software [14]. The malware is tends to be spread through mobile apps and phishing.

Ransom ware is spread through many different ways: exploiting security vulnerabilities, phishing, malvertising and visiting infected website. The purpose of this malware is to deny access to some or all of the computer system until a ransom is paid. This is usually paid with cryptocurreny as it is virtual untraceable. An event such as the WannaCry attack on the NHS lead to systems being down for days costing £92 million [1].

Backdoor attack if done successfully allows the attacker to gain remote access to a victim's computer, which can lead to finding sensitive data that can later corrupted, deleted or transferred. This malware is usually used when an attacker does not want the victim to know that he has remote access to their computer and required the bypass of authentication. The malware can be stored in a hidden part of a program like a Trojan horse, firmware, operating systems, code or hardware such as a router. A normal use of this is network administrators needing a way to reset employee passwords.

RAM scraper gains information that is temporarily stored on the Ram of a device. Data is often not encrypted for a short time resulting in sensitive data being able to be viewed, resulting in the targets often being point-of-sale (POS) systems such as cash registers when dealing with debit/credit cards.

Fireless malware works in memory and infects legitimate programs but does not leave a footprint and or try to use files to store itself. Instead, the malware goes to the memory and aims to exploit PowerShell and other automated tools used by system administrators to avoid detection. Usually, a system reboot will clear the malware from the device as it is not stored on the hard drive. One of the more recent attacks was the Equifax breach [9].

Greyware can be subcategorised into three: spyware, adware and unwanted remote access tools [15,16]. Greyware does not have to be malicious but it's intended goal is to slow down the performance of the computer and create security risks. Spyware secretly monitors the device obtaining data, either for individual use or to sell the data on such as passwords. Adware is software that displays pop up advertisements. The attacker

GAINS revenue every time the advert is show or clicked and the links may be malicious. Unwanted remote access tools can be installed without the users knowledge. The attacker will be able to use the victim's computer and can take it over without being sat there [17,18]. This will give the attacker full access of the computer allowing for easy access to the sensitive data even when the computer is off.

Keyloggers are available on computers and smartphones. This malware records which keys are being pressed in order to obtain sensitive information. Data such as login information and debit cards is the main data targets that the hackers wants sent back to them.

Cyptojacking uses the victims computer to mine cryptocurrency. It is often downloaded onto the victims computer with the aid of malvertising. This is another type of malware that is malicious, where it is added to legitimate advertising websites and blends in as a normal advert. Adverts can be made to look attractive with offers with products that will persuade victims to view the deals.

If a computer can be controlled remotely because of malware it is considered to be a bot. The bot can then be used for illegal activity and be traced back to the bot computer rather than the hackers [19-21]. A botnet is a group of bots which are primarily used to perform denial of services attacks (DDosS) and spreading other types of malware. A bot master is the cybercriminals controlling the machines remotely. They are commonly used for email spam, fraud campaigns, DDoD attacks, mine Bitcoin, keyloggers and to rent to other hackers.

Crimewares function is to access a victims financial accounts to steal funds and make unauthorised transactions to simulate cybercrime without the victim knowing. Social engineering can be used to obtain sensitive information using this malware.

Rootkits usually gain access through phishing, password cracking or exploiting security to gain access to the computer. Once access is acquired the rootkit and finds another software to hide its code in and can be hard for anti-virus software's to find. The purpose of this is to gain unauthorised access to the computer and can be installed with administrator privileges for the hacker.

Hijackware hinders a user using the internet. It will install unwanted toolbars as well as display pop up ads constantly. The malware can also decide what webpage the user goes to by redirecting them to a specific webpage [5].

There are many more malware types but many of today's malware attacks are hybrid malware where it is a combination of malware's being used to attack. For example a Trojan attack may also have adware within it. Bugs can also open up vulnerabilities for malware exploitation, with EternalBule vulnerability leading to the WannaCry ransomware cryptoworm. Miral malware targets the internet of things and was very successful in 2016 to target DNS provider Dyn with a DDos attack as most devices have default username and passwords. This is a prime example of negligence when setting up devices.

III. Malware analysis

A. What is malware analysis?

It is important to analyse a suspicious file or URL to help detect or mitigate. This is vitally important to security analysts and incident responders to use malware analysis to understand the attack. The key benefits of malware analysis are identify the source of the attack, identify the security threat level of the malware, identify the exploits the malware uses to gain access allowing for patch the network, assess the damage done, improve efficiency of indicators of compromise and reveal hidden compromises in the network.

B. Types of malware analysis

Static analysis main disadvantage is static analysis struggle against sophisticated malware as it does not analyse code when it is running. Instead, it looks for files with malicious intent and identifies packed, files, libraries, infrastructure, file name, type and size. MD5 hashes (used for fingerprinting) can be used to compare with a database for known malware by using filename, and hash number as this is based on the contents of the file [5]. This analysis is used for efficient, transparency and quality of analysis. Efficiency allows many files to be checked quickly and extract malware in a reasonable time; some malware are designed to purposefully slow themselves in order to increase detection time. Sophisticated malware have the ability to detect a sandbox thus transparency is important. The information extracted and tool uses to analyse the information will result in the quality of analysis [6].
Dynamic analysis (behavioural) occurs in a secure environment (sandbox) to execute suspicious files that are flag for malicious code. Using a sandbox allows the network defender to run the malicious code without their network being infected, giving them the ability to see how the malware works; the malware could change IP addresses, file path locations, new registry keys, domain names ect. The network defender can also see if the malware is communicating with the hacker and try to find the location of the hackers sever. Debugging can also allow break down the steps of the malware whist it is being run. A database can be created to store attempted attacks and new discovered attackers can be compared with them to best find a solution to mitigate the malware [6].

C. Code analysis

Code analysis is a technique that requires understanding of the operating system and programming language of the malware. The focus is looking at the code and inner working of the binary. This technique shows analysis that both static and dynamic cannot.

D. Memory analysis

Memory analysis looks at the computer RAM for forensic artifacts and is useful for detecting stealth and evasive capabilities of malware. It gives a good understanding of malware behaviour after infection and is a forensic technique.

E. Hybrid analysis

Hybrid analysis has the ability to detect sophisticated malware as it uses both static and dynamic analysis. Malware now has the ability to go undetected by both static and dynamic with newer versions preventing the code to run in virtual software or if debugging is attempted. Malware may require user input or be delayed trying to avoid detection. With both of these techniques being used more indicators of compromise can be extracted and is has the most effectiveness out of the three [4][7].

Attacker's often use use Obfuscation to modify malware to evade malware detectors. The common malware technique used by hackers can encrypt itself and decrypt itself after it passes the malware detector. A polymorphic virus is an example of this that uses nop-insertion and code transposition to decrypt code in a loop using place jumping to maintain original semantics [3][13].

IV. Stages involved with malware analysis

There are four stages of when analysing a malware: fully automated analysis, static property analysis, interactive behaviour analysis, manual code analysis.

- Stage one: Fully automated malware analysis is one of the best ways to process files on a large scale, because of the speed and has the ability to provide quick answers in a report. The report usually describes network traffic, file activity, the malicious code registry keys, mutex values and more. The suspicious files are accessed to determine what would happen to the network if infected. The key benefit of the fully automated analysis allows for rapid incident response time to handle malware. This is a much cheaper option than using a cyber-analyst to respond to each incident, instead more complex incidents are handled by them.
- Stage two: Static properties analysis determines if further stages of analysis need to be taken and is quick due to no running of code [23]. It is useful for analysts as static properties can sometimes be sufficiently for looking at indicators of compromise and is usually the first stage in when analysing.
- Stage three: Interactive behaviour analysis is usually the third stage where the malicious code is found and is run to observe the behaviour in an isolated environment. This will show how the file system, registry, process and network activities work. One of the important values to watch is memory usage as it can show when the malicious code is executed and how often. Interaction with the program can yield better understanding of the malicious program; tools such as Wireshark, Process Hacker/Monitor and ProDot.

  It is important to do multiple testing as the malicious program may be programed to attack certain servers in the network or change its attack every time it's executed. Having an analysis interacting with the program in creative ways is more time consuming that automated methods, however the analyst will spend more time and look for more than what the automation is coded to find. This stage requires more skill compared to stages one and two [8].
- Stage four: Manual reverse coding is a time-consuming task that requires a specialist to debug and is done to understand exactly how the malware program works. As it is time consuming and a rare skill most malware is not decoded. However, it is possible to reverse the code to provide insight that the other stages cannot such as: decoding encrypted data stored or transferred other capabilities the program has that didn't show when interactive behaviour is used and understanding the logic of the program.

V. Malware analysis use case

- Malware detection: Identifying code and using deep behavioural analysis will allow functionality threats to be detected. This information can help with future tools relating to threat alerts.
- Incident response: The main benefit is effectiveness and efficiency of the response, as the malware attack is likely to be known.
- Threats and alerts: Security teams benefit greatly as newer malware attacks can be detected earlier in the attack cycle. This improves the response time and less chance of network breach.
- Malware hunting: The process of malware hunting is to find had to find malware and find ways to mitigate it. The analysis will help to expose the programs behaviour and find network connections, port access, file access and domains.

Analysis of malware is vitally important to keep on top of the ever-evolving threats, allowing business to put appropriate security measures and incident response plans in place. Reports from malware analysis allow business to mitigate vulnerabilities quicker helping to reduce costs and recovery time. It is important to follow the stages listed above as it is best practice as each stage has more depth of analysis than the last, therefore more time consuming. If the stages were not in this order time would be wasted by not filtering out relevant files [3].

VI. Classifications

Malware classification use deep learning algorithms to categorise the malware. There are many types of ways the malware can be looked at to train the deep learning algorithms. The final aim of the classification it to determine the threat level of the malware and find what family it belongs to [25]. The result of this allows security engineers find the best way to deal with specific malware attacks.

Classifications can be grouped into two categories dynamic and static to find features of the malware. Features are extracted to be used for machine learning or data mining tools

are used to create a classification for the malware types. The attackers motive can also be used for classification. Crimeware classification could be malware designed to steal personal, business or proprietary information in order to steal something of financial value. [7].

By disassembling malware low-level static features such as strings, byte-sequences and opcode sequences can be extracted. High-level static features can also be extracted by using control flow graphs, functions used in the malware and API call sequences. Opcode sequences change the malware to the point it is undetectable or hard to detect with static features with obfuscation technique [24]. API call sequences are one of the more common techniques that are used to extract data from and can be done from either static or dynamic analysis. There may be multiple call paths the malware uses therefore is important static analysis is done thoroughly before information is extracted. Dynamic analysis provides more important and reliable results as the features can be seen as code is running. Usually a single path is discovered, however this can be hard to obtain when using a virtual environment; code could become inactive or pre-code execution is deploy at a different time. Dynamic analysis overall is better when using large datasets [10].

Malware samples firstly have to be identified to what family they belong to. As malware contains binary indicators and strings these can be used to help determine the malware classification. Malware sometimes can be based on general characteristics, however, is mostly classified on the unique strings and binary indicators in the binary. YARA tool can be used to identify and classify malware. Malware researchers use this to create YARA rules bases on textual or binary information from the malware samples and use this to scan files to look for the malware [2][11].

VII. Conclusion

With the vast amount of malware types, it is important to use malware analysis to identify and counter new and old malware. Malware analysis is a vital when looking at malware as it will show exactly how malware works, reacts to interaction and the types of functions it has not only to deploy malware but also to evade anti-malware software. With the aid of malware classification security engineers can help to determine the best way to counter and detect newer malware variants. This will improve the time it takes to counter the malware as well as reduce costs and overall damage to infected devices.

References

[1] Acronis International GmbH (2021) The top 5 ransomware attacks in the UK and their hidden costs on business. Acronis International GmbH. https://www.acronis.com/en-gb/articles/ransomware-attacks/ Date Ac- cessed: 11/11/2021

[2] I. Ghafir and V. Prenosil, “Blacklist-based Malicious IP Traffic Detection,” Global Conference on Communication Technologies (GCCT). Thuckalay, India: pp. 229-233, 2015.

[3] Black, P. et al (2021) Malware Variant Identification Using Incremental Clustering. Electronics 10(14). https://www.mdpi.com/2079-9292/10/14/1628/htm Date Accessed:12/11/2021

[4] I. Ghafir, M. Husak and V. Prenosil, “A Survey on Intrusion Detection and Prevention Systems,” IEEE/UREL conference, Zvule, Czech Republic, pp. 10-14, 2014.

[5] Christodorescu, M. et al (2005) Semantics-aware malware detection. Institute of Electrical and Electronics Engineers 1(1) 1-23.https://dl-acm-org.brad.idm.oclc.org/doi/10.1145/1387673.1387674 Date Accessed:21/11/2021

[6] Damodaran, A. et al (2017) A comparison of static, dynamic, and hybrid analysis for malware detection. Journal of Computer Virology and Hacking Techniques 13(1) 1-14.

[7] I. Ghafir and V. Prenosil, “Advanced Persistent Threat Attack Detection: An Overview,” International Journal of Advances in Computer Networks and Its Security (IJCNS), vol. 4(4), pp. 50-54, 2014.

[8] K. A, M., 2018. Learning malware analysis. Birmingham, UK: Packt Publishing, pp.6-75.

[9] Leon, R. S. (2021) Hypervisor-assisted dynamic malware analysis. Cybersecurity 4(1).https://link.springer.com/article/10.1186/s42400-021- 00083-9 Date Accessed:11/11/2021.

[10] I. Ghafir and V. Prenosil, “DNS query failure and algorithmically generated domain-flux detection,” International Conference on Frontiers of Communications, Networks and Applications. Kuala Lumpur, Malaysia, pp. 1-5, 2014.

[11] Monnappa, K. A. (2018) Learning Malware Analysis: Explore the Concepts, Tools, and Techniques to Analyze and Investigate Windows Malware. First Edition. Packt Publishing Limited.

[12] Mukherjee, L. (2020) Intro to Malware Analysis: What It Is How It Works. https://sectigostore.com/blog/malware-analysis-what-it- is-how-it-works/ Date Accessed: 26/11/2021.

[13] I. Ghafir, J. Svoboda and V. Prenosil, “Tor-based malware and Tor connection detection,” International Conference on Frontiers of Communications, Networks and Applications. Kuala Lumpur, Malaysia, pp. 1-6, 2014.

[14] Musarubra US LLC (2021) What Is Fileless Malware? https://www.mcafee.com/enterprise/en-gb/security-awareness/ransomware/what-is-fileless-malware.html Date Accessed: 26/11/2021

[15] Pektas, A. and Acarman, T. (2018) Malware classification based on API calls and behaviour analysis. IET Information Security 12(2) 107-117.

[16] I. Ghafir, J. Svoboda, V. Prenosil, “A Survey on Botnet Command and Control Traffic Detection,” International Journal of Advances in Computer Networks and Its Security (IJCNS), vol. 5(2), pp. 75-80, 2015.

[17] Sudhakar and Kumar, S. (2021) MCFT-CNN: Malware classi- fication with fine-tune convolution neural networks using tradi- tional and transfer learning in Internet of Things. Future Genera- tion Computer Systems 125(1) 334-351. https://www-sciencedirect-com.brad.idm.oclc.org/science/article/pii/S0167739X21002247?via Date Accessed: 14/11/2021.

[18] VILLAS-BOAS, A., 2019, May 28. A laptop infected with 6 of the most dangerous computer viruses in history was sold at auction to an anonymous buyer for $1.345 million — here’s what each virus can do. Business Insider.

[19] S. Eltanani and I. Ghafir, "Aerial Wireless Networks: Proposed Solution for Coverage Optimisation," IEEE Conference on Computer Communications Workshops”, IEEE, 2021.

[20] I. Ghafir and V. Prenosil, “Advanced Persistent Threat and Spear Phishing Emails.” International Conference Distance Learning, Simulation and Communication. Brno, Czech Republic, pp. 34-41, 2015.

[21] M. Hammoudeh, I. Ghafir, A.Bounceur and T. Rawlinson, “Continuous Monitoring in Mission-Critical Applications Using the Internet of Things and Blockchain,” International Conference on Future Networks and Distributed Systems. Paris, France, 2019.

[22] I. Ghafir and V. Prenosil, “DNS traffic analysis for malicious domains detection,” International Conference on Signal Processing and Integrated networks. Noida, India: pp. 613 - 618, 2015.

[23] U. Raza, J. Lomax, I. Ghafir, R. Kharel and B. Whiteside, “An IoT and Business Processes Based Approach for the Monitoring and Control of High Value-Added Manufacturing Processes,” International Conference on Future Networks and Distributed Systems. Cambridge, United Kingdom, 2017.

[24] I. Ghafir, V. Prenosil, M. Hammoudeh, F. J. Aparicio-Navarro, K. Rabie and A. Jabban, “Disguised Executable Files in Spear-Phishing Emails: Detecting the Point of Entry in Advanced Persistent Threat.” International Conference on Future Networks and Distributed Systems. Amman, Jordan, 2018.

[25] You, I. and Yim, K. (2010) Malware Obfuscation Techniques: A Brief Survey. Institute of Electrical and Electronics Engineers 1(1) 4-192.