I.J. Education and Management Engineering, **2018, 2, 20-30**
Published Online March 2018 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijeme.2018.02.03

Available online at http://www.mecs-press.net/ijeme

A Study on Malware and Malware Detection Techniques

Rabia Tahir

Department of Computer Science, Virtual University of Pakistan

Received: 09 October 2017; Accepted: 19 December 2017; Published: 08 March 2018

Abstract

The impact of malicious software are getting worse day by day. Malicious software or malwares are programs that are created to harm, interrupt or damage computers, networks and other resources associated with it. Malwares are transferred in computers without the knowledge of its owner. Mostly the medium used to spread malwares are networks and portable devices. Malwares are always been a threat to digital world but with a rapid increase in the use of internet, the impacts of the malwares become severe and cannot be ignored. A lot of malware detectors have been created, the effectiveness of these detectors depend upon the techniques being used. Although researchers are developing latest technologies for the timely detection of malwares but still malware creators always stay one step ahead. In this paper, a detailed review of malwares types are provided, malware analysis and detection techniques are studied and compared. Furthermore, malware obfuscation techniques have also been presented.

Index Terms: Malware, malware analysis techniques, malware detection techniques.

1. Introduction

The term malware is short for of malicious software, as the name suggests malwares are intended to harm computers and computer users by stealing information, corrupting files or by just doing mischievous activities to annoy users. Malware is widely spreading, it is suggested, there is massive increase in security incidents of computer[1,2].Development of networks is hindered by malware. Malware targets the applications that are runs over the internet. As almost all fields of life uses internet to improve its quality of service increase the need to detect and deactivate malwares as early as possible so that the negative results created by these malwares can be avoided. The malware that have propagation ability are most dangerous one because there is not central control so defending them is not an easy task. Studies show that the malwares are one of the most significant

* Corresponding author.
E-mail address: ms160400256@vu.edu.pk

threat to computer security[3].Malware creators are always come with new ideas. They develop malwares in such a way that they changed themselves from time to time so that they cannot be detected easily.

Malware writers always try to write programs that cannot be easily detectable, with passage of type they made improvement in the techniques that are used to hide or morphed the malicious code successfully. These ideas starts with simple encryption and then go towards oligiomorphic, polymorphic and metamorphic viruses.

Malware detectors are used to detect these malwares and antivirus scanners are one of the way to detect some of them but with progression of malware development techniques, malware detectors use a number of techniques to avoid the disastrous effects of these software. Due to the limitation of the existing malware detection techniques, the machine learning and data mining methods are combined with existing detection methods to add the efficiency in the detection process[18]. Signature based detection methods are good in detecting the known malwares but they are unable to detect unknown malwares and polymorphic malware because they can change their signatures. signature based detection can also not detect new malware as their signature are not developed at this stage. Although heuristic based detection methods can detect new, known as well as unknown malwares but they have high rate of false positive and negative which leads us to development of more accurate detections methods. Due to the rapid increase in polymorphic malwares, the heuristic based detection techniques are combined with machine learning method to get more accurate and efficient detection of malware.

This paper presents a brief study of malwares, overview of different kinds of malware, camouflage evolution in malware, malware obfuscation techniques, malware analysis techniques and malware detection methods. The comparison among analysis and detection techniques have also been made and future trends in detection techniques are also identified.

2. Types of Malware

Malwares exist in different forms, they are broadly categorized in following classes. They are not mutually exclusive although many of them exist in more than one class.

Virus: Virus infects computers and other files by replicating itself. It cannot exist independently so it attaches with other files more precisely executable files and application and due to its replicating features , it spread across files and even computers through network.it cause system performance degradation and denial of service [4].

Worms: Worms are malicious piece of code that exist independently. they have feature to replicate itself. They propagates through storage devices and emails, also consume network and computer resources which leads to system degradation in performance. As they can create multiple copies of themselves, antivirus scanners can identify these codes because of multiple existence.[6]

Trojan Horse: Trojan Horse behaves like a useful program but it has harmful purpose. They do not replicate themselves but it transferred in a computer by internet interaction like downloading. It steals sensitive information, observe activity of users and can delete and alter or corrupt files on the system where it resides.[5]

Rootkit: Rootkit take control of the operating system such that it can hide itself or can make a safe environment for other malwares to hide in the system. Basically it a masking techniques to cheat antivirus so that they cannot find malwares in the system and consider them as normal applications.

Spyware: Spyware are used to steal someone personal information or keep the watch on user's activities. It is installed without the knowledge of system owner and secretly collect the information and send it back to the creator. Even company with big names like google also use spyware to collect the required information of their users[7].

Adware: Adware are quite annoying most of the time as it plays advertisement on user computer without its permission and interrupt its current activity. Basic purpose of the adware is to get financial benefit. It does as much harmful as other malwares.

Cookies: Cookies are in form of text file and contain information that is stored by web browser on users

computer for future use. Apparently cookies are not harmful but they become a threat when it is used by some spyware.

Sniffers: They are the software that observe and record the network traffic. They analyse different fields of packets and collect information for preparation of the malware attack.

Botnet: Bot is a software that allow attacker to control an infected computer .A network of infected computers controlled by hackers/attackers to do malicious activities without the knowledge of owner. They can make denial of service attacks, send spam messages, steal information

Keyloggers: It is kind of spyware that is used to record key strokes to steal passwords , credit card details and other important and sensitive figures. It transferred in a computer when some other malicious software is installed or any infected site is visited by user.

Spam: Junk emails are another names of spams. These identical emails are created and send to multiple recipients at the same time. It consumes a lot of bandwidth and also cause to slow down the system.

Ransomware: Now a days Ransomware are the major threat for internet industry. Ransomware are malwares that take control of your PC by encrypting your data, stop some application and don't allow you use your operating system until you fulfil their demands. The demands is mostly in terms of money. After paying money, it is still not guaranteed that you will get your control back.

3. Camouflage Evolution in Malware

To understand and develop malware analysis and detection techniques, it is strongly recommended to study Camouflage of malware. Camouflage of malware refers to concealment of the malware to hide them from malware detectors as long as possible. There are number of techniques used by malware writers which vary from simple tactic like encryption to complex and advanced one like metamorphic[19].

3.1. Encryption

Malware writers always want to make their program to not be noticed and detected by malware detectors. To easiest technique that they use to camouflage is encryption. It is the first technique that is used for concealment of malwares [8]. It consist of encryption and decryption module. Each time encryption is done with different key while decryption remain the same. As uniqueness is not provided in decryption technique so that there is possibility of their detection. In 1987 the first encrypted malware CASCADE was appeared [9]. Structure of encrypted virus is shown in Fig.1

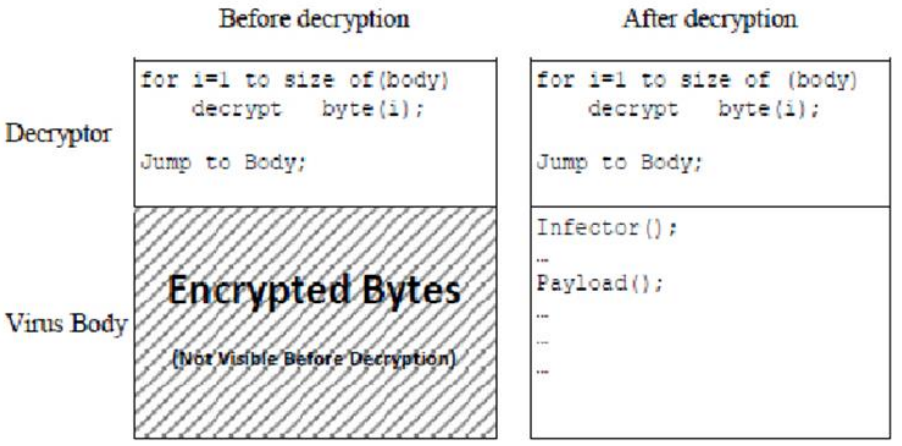

Fig.1 Structure of Encrypted Virus

The main purpose of this technique is to avoid antivirus detection and static code analysis. This method also delay the process of investigation.

3.2. Oligiomorphism

First oligiomorphic virus was appeared in 1990, it named as Whale and it was a DOS virus[10].This is considered as an advancement in camouflage in malwares. As in encryption, the decryptor remain consistent for each infection while on oligiomorphism works with unique decryptor with each infection. It is also considered as advancement in encryption technique also known as semi polymorphic. Although oligiomorphism provides different decryptor from a list of decryption for each new attack, still there are chance to caught by antivirus by checking all the decryptor.

3.3. Polymorphic

In 1990, First Polymorphic virus 1260 was developed by Mark Washburn. Polymorphic virus are combination of encryption and oligiomorphism but they are more complex than other viruses. So it is very hard to detect by antiviruses as they change their appearance with each copy. Number of decryptor that they can generate are not limited. This kind of viruses use different obfuscation techniques to change its appearance. This procedure of change is done by mutation engine.

3.4. Metamorphism

Encryption is not a part of metamorphic virus rather in this generation the content of the malware alters. This is the reason that there is no need of decryptor. It also implements a mutation engine like polymorphism but it change its whole body rather by only altering the decryptor. The basic idea is the syntax change on each new copy while semantics remains the same i.e. the apparently virus change on each infection but the meaning or working remains the same. First metamorphic virus ACG was developed in 1998 for DOS

4. Malware Obfuscation Techniques

The technique that malware programmers use to make the malware difficult to read and understand is known as Obfuscation. The basic purpose of this technique is to hide the malicious behavior of malware. Obfuscation techniques are categorized in different ways by different researchers. Most common techniques are divided in following six categories[11].

4.1. Dead code insertion

This is easiest way to change code without affecting its meaning. In this technique, garbage code or statements are inserted in the code by using NOP statements and push followed by pop. These statement used in such a sequence that it does not affect the semantics of code.

4.2. Instruction replacement

In this technique, instruction are replaced with the other instructions that generate the same meaning just like synonyms in natural languages. For example all following instructions have same effect on register eax. Each instruction set the register value to zero.

```
move eax,0
xor eax, eax
and eax,0
sub eax,eax
```

In this technique instruction are substituted with equivalent instructions which make the detection of these malwares very hard.

4.3. Register reassignment

This technique reassign register in every copy without changing the semantics of the virus. It is a simplest technique but when combine with other technique can make it very difficult to detect.

4.4. Subroutine reordering

A set of instructions in piece of code is permuted in this technique such that code changes in its appearance but the behavior remains the same .An example of subroutine follows by its reordering is described as follows:

```
//Subroutine
mov eax, 0A
push ecx
add esi, ebx

//reordering
add esi, ebx
mov eax,0A
push ecx
```

4.5. Code transposition

In this technique, the flow of the instructions in the original piece of code is rearranged such that there semantic does not change[12].There are two approaches that can be used to do code transposition. One is reorder the instructions randomly and to recover the original code the unconditional statements and jumps are used while in other method the instructions that are independent and have no impact on other instructions are chosen and reordered. Second approach is difficult to implement but it is much more effective than first one.

4.6. Code integration

As the name indicates, in this malware obfuscation technique, the malicious code integrate or embed itself into the program that need to be effected. This is very effective technique, the original program is decomposed and malicious code is added in it such that it cannot be detected easily[13].

5. Malware Detector

A program that is designed to detect malicious programs and code are known as malware detector. The general malware detection function can be defined as

$$D(p) = \begin{cases} malicous & if\ p\ contains\ malicous\ code \\ bengin & otherwise \end{cases}$$

D is the function that can check the an applications or programs (p) is either a benign Program or a malicious program.

$$D(p) = \begin{cases} malicous\ if\ p\ contains\ malcode \\ bengin\ \ if\ p\ is\ a\ normal\ program \\ undecidable\ \ if\ D\ fails\ to\ determine\ p \end{cases}$$

Sometimes D cannot decide a program is harmful or harmless application because the malware is just a new invention and cannot be detected by malware detector so we can rewrite the above defined function as follows:

6. Malware Analysis Techniques

Malware analysis is a step towards malware detection. To detect malwares , first we have to analyze , how malware perform its function and what is the purpose behind malware development so that this kind understanding about malware make it easy for the developers of malware detectors to implement the defensive functionality. Malware analysis techniques are divided into three categories on the basis of time and technique to do the analysis.

6.1. Static Analysis

When a software or piece of code is analyzed without executing, this kind of analysis is called static analysis or code analysis. Static information is extracted from the code to determine either the software contains malicious code or not. In this technique, the software is reverse engineered by using different tools and the structure of the malicious code is analyzed to understand how it works. Different tools that can be used to perform static analysis are debugger, dissembler, decompiler and source code analyzers. Methods that are used in performing static analysis include File Format Inspection, String Extraction, Fingerprinting, AV scanning and Disassembly.

6.2. Dynamic Analysis

When functionality of software is analyzed and observed by executing it is known as dynamic or behavioral analysis[14].It can be done by tracing the function calls, control flows and also by analyzing the instructions and parameters of functions. Malicious code are run in a virtual environment to observe its behavior and to design the actions against these negative behavior. The tools that are used for dynamic analysis are sandbox, simulator, emulators RegShot, Process Explorer. Dynamic analysis is more effective the static analysis as in this technique, the infected software is executed in virtual machine for monitoring purposes. This can detect many kinds of malwares easily. This type of analysis takes more time as we have to design the environment to execute and test a malicious software.

6.3. Hybrid

It combines both static and dynamic analysis techniques so can take the benefits if both approaches. Firstly a software is observed by code analysis by checking the malware signature and then it is run in virtual environment to observe its actual behavior.

The comparison between static and dynamic analysis techniques have been shown in Table I.

Table 1. Comparison of Static and Dynamic Analysis

Static Analysis	Dynamic Analysis
Fast and safe	Time consuming and vulnerable
Good in analyzing multipath malware	Difficult to analyze the multipath malware
Cannot analyze obfuscated and polymorphic malware	Cannot analyze obfuscated and polymorphic malware
Low level of false positive (Accuracy is high)	High level of false positive(Accuracy is low)

7. Malware Detection Techniques

Malware detection techniques can be divided into three broad categories, signature based, heuristic based and specification based. These techniques identify and detect malware and take countermeasures against those malwares for the safety of computer systems from a potential loss data and resources.

7.1. Signature based

When a malware is written, a sequence of bit generally known as signature is embedded in its code which can later use to identify from which family this malware belongs to. The signature based detection technique is used by most of the antivirus programs. The antivirus program disassemble the code of the infected file and search for the pattern that belong to a malware family[15].signatures of the malwares are maintained in database and then further used for comparison in detection process. This kind of detection technique is also known as string or pattern scanning or matching. It can be static, dynamic or hybrid as well.

7.2. Heuristic based

The heuristic based detection detects or differentiate between the normal and abnormal behavior of a system so that ultimately the known and unknown malware attacks can be identified and resolved. The heuristic based detection process consists of two steps. In first step, the behavior of the system is observed in the absence of attack and keep a record of the important information that can be verified and checked in case of attack. This difference is watched out in second step to detect the malware of a particular family.

Behavior detector that is used in heuristic based technique consist of the following three basic components.

Data collection: As the name suggests, this component deals with the collection of data either static or dynamic.

Interpretation: This component interpret the data collected from data collection component and convert it into intermediate form.

Matching algorithm: This component is used for matching the behavior signature with the converted information in the interpretation component.Behavior detector is shown in Fig.2 which explains the functionality how all these components work together.

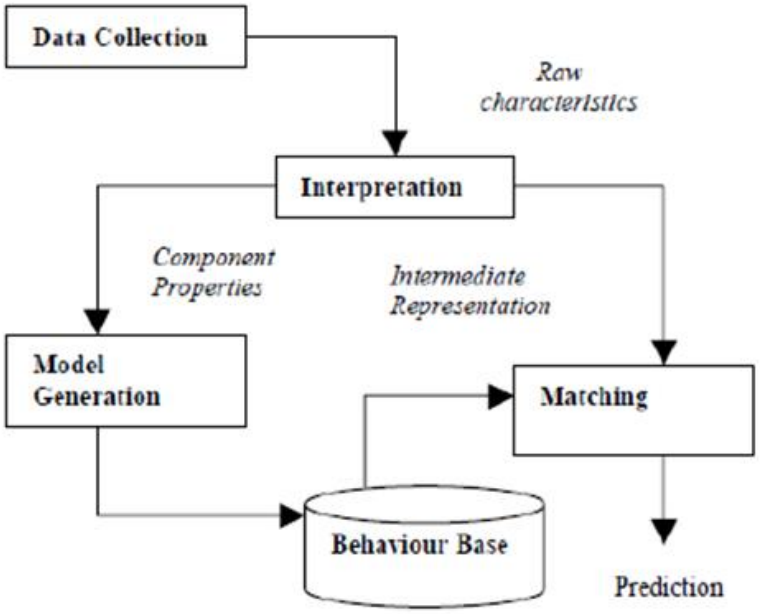

Fig.2. Behavior Detector[16]

Although heuristic based detection is an efficient method but there are limitations associated with method as well as it needs more resources and level of false positive is also high. This method is also known as proactive technique, also named as behavior or anomaly detection techniques. Different type of analysis techniques are done before performing heuristic based detection that are file based ,weight based ,rule based and generic signature analysis.

7.3. Specification based

In specification based detection technique, applications are monitored according to their specification and checks for the normal and abnormal behavior. This technique is derived from heuristic based technique but main difference is heuristic based detection techniques used machine learning and AI methods to detect valid and invalid activity of a legitimate program but in specification based detection technique is based on the analysis of the behavior that are described in the system specification[17]. This method is somehow manual comparison of the normal activities of some system. It overcomes the limitation of the heuristic based techniques by lowering the level of false positive and increasing the level of false negative.

The advantages and disadvantages of the three detection techniques are shown in Table II.

Fig.3. classification and relationship of the malware analysis and detection techniques are shown, each of malware detection technique can wither be static, dynamic or hybrid and also the specification based detection technique is derived from heuristic based detection technique.

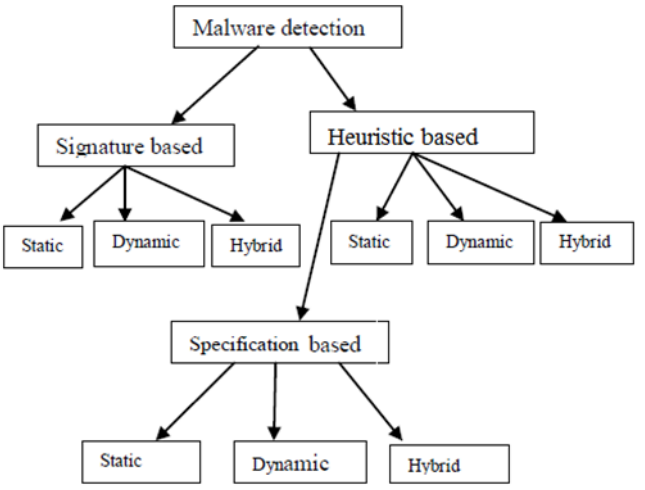

Fig.3. Categories of Malware Detection Techniques

Table 2. Advantages and Disadvantages of Malware Detection Techniques

	Advantages	**Disadvantages**
Signature based	-Known malwares can be detected easily -Used less resources as compared to other techniques	-Unknown malwares cannot be detected.
Heuristic based	-Known and unknown new malware can be detected	-Data need to be updated regarding new and unknown malwares. -Need more resources in terms of time and space -level of false positive is high.
Specification based	-Known , unknown and new malware can be detected -Level of false positive is low	-level of false negative is high -not efficient in detection of new malwares. -specification development is time consuming

8. Conclusion

Internet industry is growing rapidly, users and organization need secure and safe activities over the internet. Malware are the biggest threat for today's electronic world as they are harmful for the users by stealing their information, corrupting data and disabling the network and systems by malicious attacks. This paper presents and review the malware types which are always been a danger for computer world. Day by day the malware writers are improving and evolving camouflage techniques from simple encrypted virus to extreme complex and difficult to detect polymorphic and metamorphic viruses. Two main analysis techniques are also discussed in the paper. Static analysis perform better in multipath malwares, also their accuracy level is high as compare to dynamic analysis. While dynamic analysis can analyze obfuscated and polymorphic virus but there accuracy is not up to the level. Three major malware detection techniques are discussed along their pros and cons. While comparing these techniques it is noted that signature based detection methods are good at detecting known malwares with fewer resources while heuristic and specification based detection methods can also detect new and unknown malwares but with heuristic based approach the level of false positive is high and also need more resources. The issue with the specification based detection approach is development of the specification data of the legitimate program which is time and space consuming process. Due to the limitation of heuristic based approach, new methods are being used in detection which combine the existing techniques with the machine

learning and data mining methods.

Due to the limitation of the existing malware detection techniques, the machine learning and data mining methods are combined with existing detection methods to add the efficiency in the detection process[18]. Signature based detection methods are good in detecting the known malwares but they are unable to detect unknown malwares and polymorphic malware because they can change their signatures. signature based detection can also not detect new malware as their signature are not developed at this stage. Although heuristic based detection methods can detect new, known as well as unknown malwares but they have high rate of false positive and negative which leads us to development of more accurate detections methods. Due to the rapid increase in polymorphic malwares, the heuristic based detection techniques are combined with machine learning method to get more accurate and efficient detection of malware. Various kinds of malware detection technologies are being used in industry according to the requirements for example Host based intrusion detection system, Network based intrusion detection system

Hybrid intrusion detection system, Agent based intrusion detection system, Web based intrusion detection system, Application protocol based intrusion detection system and multi-agent P2P intrusion detection system.

References

[1] Adelstein, Frank, Matthew Stillerman, and Dexter Kozen. "Malicious code detection for open firmware." Computer Security Applications Conference, 2002. Proceedings. 18th Annual. IEEE, 2002.

[2] Bergeron, Jean, et al. "Static detection of malicious code in executable programs." Int. J. of Req. Eng 2001.184-189 (2001): 79.

[3] William, Stallings. Computer Security: Principles And Practice. Pearson Education India, 2008.

[4] Spafford, Eugene. "The internet worm incident." ESEC'89 (1989): 446-468.

[5] Idika, Nwokedi, and Aditya P. Mathur. "A survey of malware detection techniques." Purdue University 48 (2007).

[6] Li, Jun, and Shad Stafford. "Detecting smart, self-propagating Internet worms." Communications and Network Security (CNS), 2014 IEEE Conference on. IEEE, 2014.

[7] Yin, Heng, et al. "Panorama: capturing system-wide information flow for malware detection and analysis." Proceedings of the 14th ACM conference on Computer and communications security. ACM, 2007.

[8] You, Ilsun, and Kangbin Yim. "Malware obfuscation techniques: A brief survey." Broadband, Wireless Computing, Communication and Applications (BWCCA), 2010 International Conference on. IEEE, 2010.

[9] Beaucamps, Philippe. "Advanced polymorphic techniques." International Journal of Computer Science 2.3 (2007): 194-205.

[10] Szor, Peter. The art of computer virus research and defense. Pearson Education, 2005.

[11] You, Ilsun, and Kangbin Yim. "Malware obfuscation techniques: A brief survey." Broadband, Wireless Computing, Communication and Applications (BWCCA), 2010 International Conference on. IEEE, 2010.

[12] Christodorescu, Mihai, and Somesh Jha. Static analysis of executables to detect malicious patterns. Wisconsin Univ-Madison Dept of Computer Sciences, 2006.

[13] Konstantinou, E., and S. Wolthusen. Metamorphic Virus: Analysis and Detection Technical Report. RHUL-MA-2008-02 Department of Mathematics Royal Holloway, University of London, 2008.

[14] Elhadi, Ammar AE, Mohd A. Maarof, and Ahmed H. Osman. "Malware detection based on hybrid signature behaviour application programming interface call graph." American Journal of Applied Sciences 9.3 (2012): 283.

[15] Landage, Jyoti, and M. P. Wankhade. "Malware and malware detection techniques: A survey." International Journal of Engineering Research and Technology (IJERT) 2.12 (2013): 2278-0181.

[16] Jacob, Gr égoire, Herv é Debar, and Eric Filiol. "Behavioral detection of malware: from a survey towards an established taxonomy." Journal in computer Virology 4.3 (2008): 251-266.

[17] Robiah, Y., et al. "A new generic taxonomy on hybrid malware detection technique." arXiv preprint arXiv: 0909.4860 (2009).
[18] Chumachenko, Kateryna. "Machine Learning Methods for Malware Detection and Classification." (2017).
[19] Rad, Babak Bashari, Maslin Masrom, and Suhaimi Ibrahim. "Camouflage in malware: from encryption to metamorphism." International Journal of Computer Science and Network Security 12.8 (2012): 74-83.

Authors' Profiles

Rabia Tahir (born October 13,1984) is currently doing MS in Computer Science from Virtual University of Pakistan .She was an ex instructor of Virtual University of Pakistan in Department of Computer Sciences. Her area of interest in network security and Databases.

How to cite this paper: Rabia Tahir,"A Study on Malware and Malware Detection Techniques", International Journal of Education and Management Engineering(IJEME), Vol.8, No.2, pp.20-30, 2018.DOI: 10.5815/ijeme.2018.02.03