**ABSTRACT**

Malware is worldwide outbreak which tends to mitigate attacks and collects private data.Studies said that the impact were surpass.There are a variety of malware by which it attacks on honest users.Malware detectors are basic tools to protect against malware.It is therefore important to study malware detection techniques and identify the type of malware attacked on your system.In this paper we will study what actually malware is and how is attacks the users.The survey provides information about different types of malware attacked on the system.In this paper we will study and analyze different malware detection techniques.There are many malware detection techniques given in this survey with an idea of comparison and decision making to distinguish among the malware.The survey also provides a reference to the user for more information.

**INTRODUCTION**

1. **What is Malware**?

The term malware is derived from combination of two words:**Mal**icious soft**ware**.It is a software that creates mischief or hurts the user.Programs are also called as malware if they secretly act against the interests of the user.For example: at a point Sony music Compact discs silently installed a roolkit on purchasers’ computers with the intention of copying ans they faced a lot of problems in the future. Malware are different from normal programs in a way that most of them have the ability to spread itself in the network,remain undetectable,cause changes/damage to the infected system or network,persistence.

1. **What does malware do?**

These malicious programs can perform a variety of different functions such as stealing,encrypting or deleting sensitive data,alerting or hijacking core computing functions and monitoring users’ computers activity without their knowledge.They have the ability to bring down computer performance to knees and can cause a destruction to the network.The most common way of attacking the system is Internet as users visits a lot of websites and there may a chance that some websites are malicious.

1. **How malware spread?**

Malware authors use variety of techniques to spread malware that infect devices and networks.The most common way is Internet.If a user is visiting some malicious website then by downloading pirated contents like emails,videos,advertisements, movies, pirated games or software, or any files then there is high probability that malware enters your system.If its an advanced malware then it directly access the Operating systems.then it starts to encrypt files and record personal information Malware can also transmit by offline like if a user plugged in a pen drive which contains malicious code or via a CD/DVD.

1. **How is Malware delivered?**

Malware is typically delivered in the form of a link or file over email and requires the user to click on the link or open the file to execute the malware.Malware has actually been a threat to individuals and organizations since the early 1970s when the Creeper virus first appeared. Some strains are sent over email via a link or executable file. Others are delivered via instant messaging or social media. Even mobile phones are vulnerable to attack. It is mandatory that organizations are aware of all vulnerabilities so they can lay down an effective line of defense.

1. **Malware types?**

Malware is a collective term of all kinds of threats.Many believe that malware is same but each malware is distinct and have different attacking techniques.The following are the types of malware:

1. Adware
2. Bots
3. Rootkits
4. Spyware
5. Trojan horse
6. Viruses

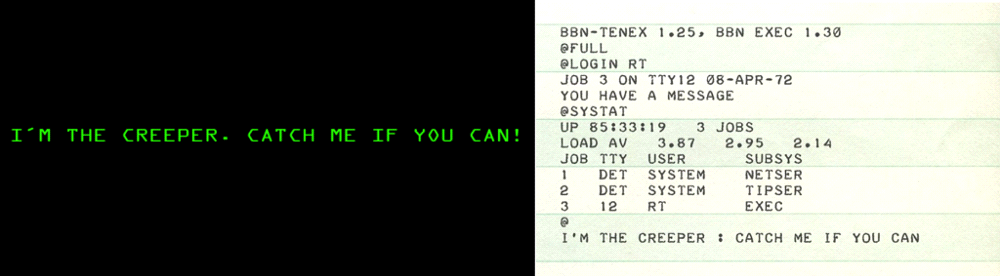
Vii. Worms

**HISTORY**

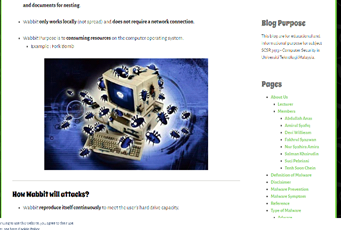
Early malware was primitive,and it was spread by offline via floppy disks(which is carried from computer to computer by human needs).

Here is a list of some early versions of malware :

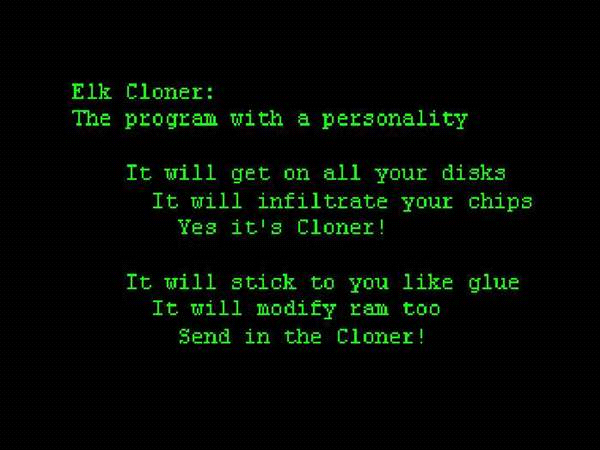
* **1971 Creeper:** An experiment designed to test how a program might move between one computer to other computer.It will run on the tenex operating system and spread through the ARPANET.The systems which are infected by the creeper they display the message “IM THE CREEPER :CATCH ME IF YOU CAN”.when the creeper program will run it will try to print the file but it doesn’t print then it will find another tenex system to open a connection and to pickup it self and transfer to the other machine then it will start running on new machine.



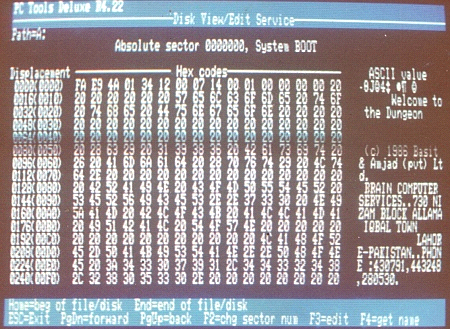
* **1974 – Wabbit**: It is a self-replicating program ,it will make a multiple copies of itself on a computer until it slow down the system, and it also makes this replication process until the system performance was reduced and it might eventually crash the system. Researchers named this virus “wabbit” (rabbit) because of the speed at which it was able to replicate.It does not require any program and documents for nesting.It works Locally and does not require a network connection.The main purpose of wabbit is to consume resources on the computer operating system and it will reproduce itself continuously to meet the user hard drive capacity. Ex: fork bomb



* **1982 –  Elk Cloner:** It was the earliest wide spreaded self replicating microcomputer virus which stores on floppy disks to affect the PC’s .when the computer booted from the floppy disks which was infected, then the copy of the virus was placed into the computer memory. when the uninfected disk was inserted, the entire disk operating system(DOS) along with the Elk cloner virus would be copied to the uninfected disk. In this way the virus transfer from one disk to another disk.



* **1986 –  Brain Boot Sector Virus**: It is the first computer virus to infect the MS-DOS. It was the standard name to the computer virus. Brain effects the IBM PC by replacing the boot sector(boot sector is a sector of a data storage device) of a floppy disk with the copy of boot sector virus and the origin boot sector is moved to another disks and it was marked as bad. Normally Infected disks will have 5 kilobytes of bad boot sectors and the disk label was changes to @Brain.



* 1988 — Morris Worm: It was self replicating worm written by a student at Cornell University who named Robert Tappan Morris. According to him the purpose of this worm is to measure the size of the precursor of “internet” of time – ARPANET(The Advanced Research Projects Agency Network).Although it caused denial of service for about 10% out of 60,000 computers which €are connected to ARPANET. The worm spread by making use of vulnerabilities in UNIX send mail, finger, and rsh/rexec as well as by guessing weak passwords.

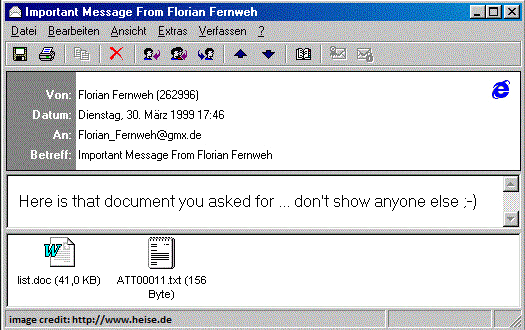
Before it spreads to other system. This worm checks that “the computer is infected or not” and was running a Morris Worm process. If it was infected then Morris worm will re-infect it 1 in 7 times so that the user will not avoid Morris worm infection by creating fake Morris worm process.

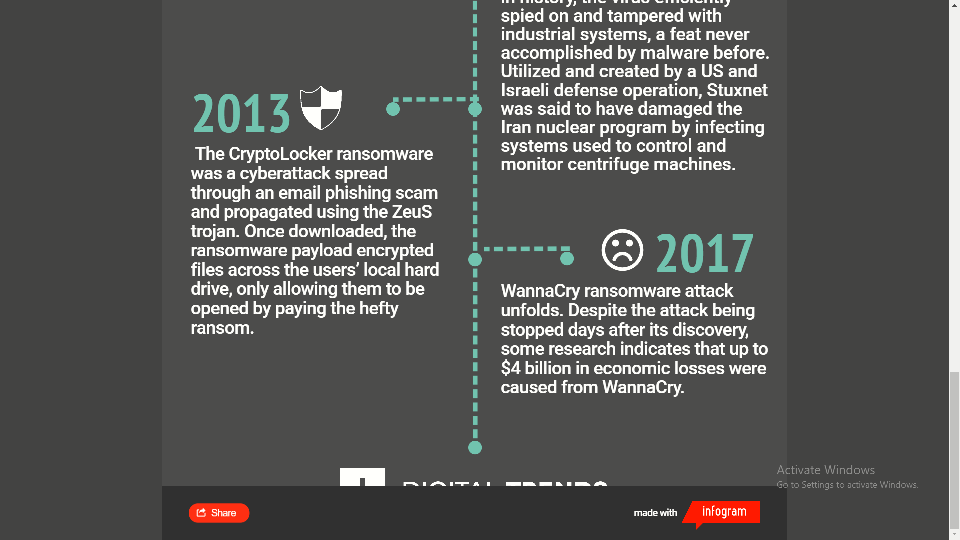
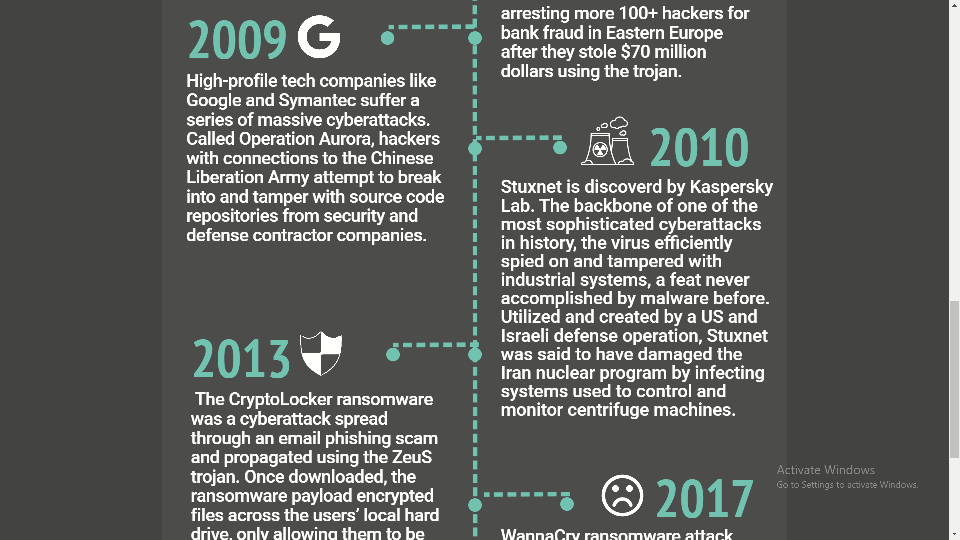
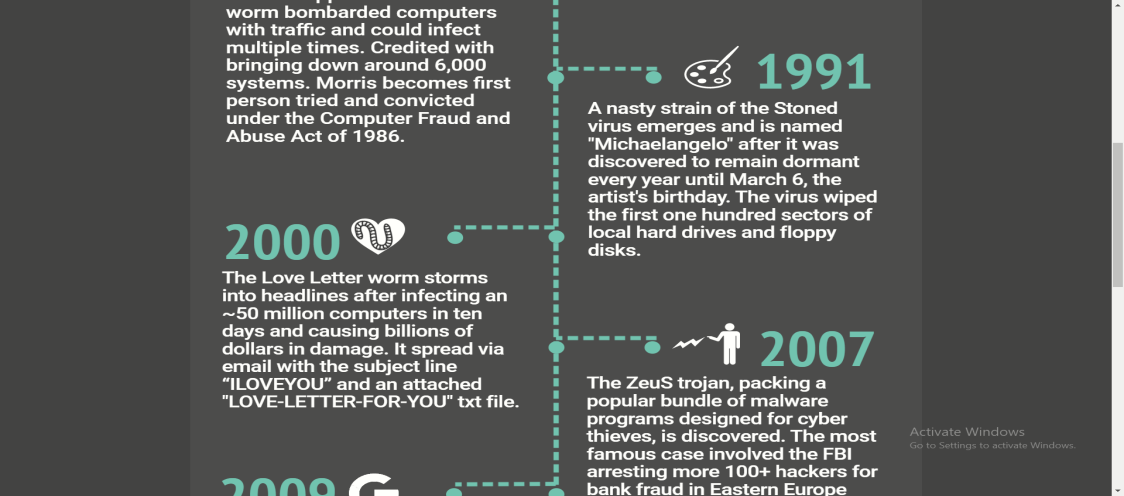
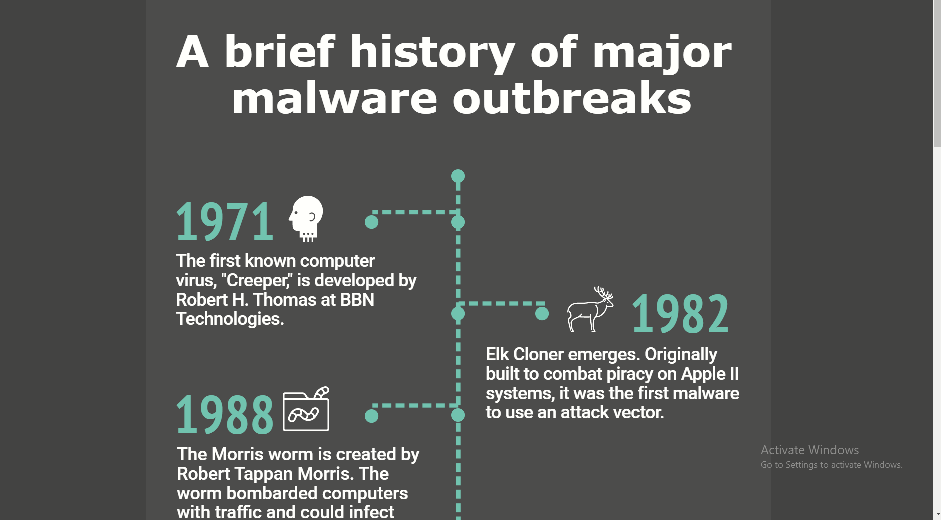


* **1991 — Michelangelo Virus**: It will re-write the first 100 sectors of the hard disk so by this the file allocation table will be destroyed and computer cannot boot.it is having destructive payload where all the data will be replaced with the random characters so that the data will not be recovered and it damages the disks And this will be done only when computer is booted on his(the artist Michelangelo) birthday dated March 6 and the virus will not check Master Boot Record(MBR) is infected previously or not. If MBR is infected by same virus then it will move the previous virus to the original MBR where it was stored so that recovery of MBR is not possible.



* **1999 — Melissa Virus:** it was the fast spreading macro virus that is distributed as an email attachment. So when email is opened it disables protection. It mostly targeted on Microsoft Word and Outlook-based systems. The virus infects the computer through the emails with the title "Important Message From", followed by current username. By clicking the message the body contains "Here's that document you asked for. Don't show anyone else ;” and below there will be a document title with the name list.doc which contains a list of pornographic sites and go with logins for each. Then it will mail itself to the first 50 contacts of users contact list and then disables the multiple protection features on on Microsoft Word and Outlook-based.





**TYPES OF MALWARE**

**Computer Virus:**



Computer Virus is a type of computer program , so when program is executed it replicates itself by modifying other computer programs by inserting its own code , when this replication is succeed then the affected areas are said to be infected with computer virus. It is designed to infect objects on the disk .It can corrupt particular file or software often triggered by individuals action.

*Computer Virus Example:*

If there is a word document and the virus has corrupted the documents then it becomes junk and we cant retrieve the information or even it can corrupt the word processor s/w completely .At this time if we try to copy the document from one computer to another then it damages the other computer.Once a virus is attached to a program, file, or document then virus will be in inactive state until circumstances cause the computer to execute its code. In order for a virus to infect your computer, you have to run the infected program, which in turn causes the virus code to be executed.  It can spread through emails , text messages , Internet file downloads, links etc.

**How to recognize your computer infected with virus?**

Here are the list of points to check whether your computer is infected with virus or not:

* Frequent pop-up windows.
* Changes to your homepage
* Mass emails being sent from your email account.
* Frequent crashes
* Unusually slow computer performance
* Unknown programs that start up when you turn on your computer.
* Unusual activities like password changes.

**Worms:**



A worm is malicious program that spreads the copies of itself from computer to computer, but instead of writing code in multiple times ,if it is installed once then it looks for the other system to infect This worms can enter into our systems as attachments in spam emails. Once it is opened it provides some link to a malicious website or it is automatically gets downloaded and starts infecting the systems. It may modify, delete the files or it may also inject malicious software. Sometimes It may keep more and more copies of itself and overloads the hard drive space. It harm the system.It can also steal data, install a backdoor, and allow a hacker to gain control over a computer and its system settings.

*How to recognize if computer has a worm?:*

Here are the list of points to check whether your computer has a worm or not:

1. **check your hard drive space**. When worms repeatedly replicate themselves, they start to use up the free space on your computer.

2. **Monitor speed and performance**:If some programs are crashing or not running properly then that could be a red flag that a worm is eating up your processing power.

3. **Keep on checking new files and old files:** The computer worm function is to delete and replace files on a computer.

*Worm example:*

Morris worm – it was the self replicating worm written by a student at Cornell University who named Robert Tappan Morris. He wanted to discover “how big the internet was”. But the code contained bugs and caused variety of problems and damaged around 60,000 unix computers, financial damage between 10$ million and 100$million.

**Trojan horses:**



Trojan horse : it is a malicious program that does not replicate but remains unnoticed and collects the username and passwords. It gets downloaded when we install any software or any mp3 songs or any games from a strange website. So it enters into the system without the knowledge of user and it helps criminals to get access over the user data and causes serious damage to the system. It has same privilege as the user has. As it has same privilege it can delete or modify the files or it might install some malware software unnoticeable. Most of the criminals uses Trojan because it is unnoticeable. It performs the activities such as creating backdoor to get access, disabling the firewalls, antivirus etc.

*How do Trojans work?*

This is an example of how the Trojans works:

We might think you’ve received an email from someone you know and click on what looks like a legitimate attachment. But you’ve been fooled. The email is from a cyber criminal, and the file you clicked on — and downloaded and opened — has gone on to install malware on your device

E g : Backdoor Trojan permits system access by an uninvited party potentially allowing remote administration of system often include a key logger which records every key pressed to find the users confidential data.

*Here are some common types of Trojan malware*:

*1. Backdoor Trojan:* The trojan will create the backdoor so that the attacker will gain access to the systems and the attacker downloads your data and he might add malicious software into the systems.

*2. Distributed Denial of Service attack trojan:* these trojans will carry the attacks against the computer networks which gets installed or downloaded via spam emails and turns those systems into a parts of botnets. These trojans will have a backdoor through which hackers gets access to the devices and it activates their botnet army to attack coordinately. Once it is activated then it starts generating traffic to the server, website by sending request continuously. By sending continuous request the server or the website will shutdown and visitors or the users cannot access them.

*3. Downloader Trojans :* it targets already infected computers. It does not have any backdoor but it performs the actions which benefits to the hacker. These trojans downloads the variety of files and programs on to the hard drive. That includes unwanted apps, or upgrades for the installed malware.

*4. Fake Antivirus Trojans:* These trojans are download as same way the other trojans via email attachments, visiting to malicious links. Once installed it acts as antivirus software and continuously informs the user about the non-existing security threats found in their device. It is same like ransomware. It prompts the user to pay money to download the full version anti virus software. And pop ups will continuously appreas with false alerts even the user is working anything else.

*5.Infostealer Trojan:* It steals the data from the infected device

*Trojan horse example:*

In 2007: Strom worm a backdoor Trojan created by Russian hakers, that was distributed to the millions of pc’s worldwide via emails about a fictional strom wave that was killing many people in Europe. These trojans are used to create botnet and attacks against the websites and servers and cyber security companies

***Ransomware***

Ransomware is a type of malware  that threatens to publish the victim's data or it block access to it unless demanded amount of money is paid. While some simple ransomware may lock the system in a way which is not difficult for a knowledgeable person to reverse, more advanced malware techniques are used they are cryptoviral extortion .The victim organization is stopped totally until it pays, but there is no guarantee that payment will result in the necessary decryption key or that the decryption key provided will function properly.

Common types of ransomware:

*1.Crypto Locker :* It encrypts the user data of each file with the randomly generated symmetric key and this symmetric key is encrypted with the public asymmetric key.

Key and adds this to the file. Once all 70 common types encrypted then it displays

the message demanding for money to decrypt the encrypted file, if they wont return money then symmetric key will be deleted and data will be lost.

*2. Scareware:* It acts as the anti virus software. It finds the issues on your computer and it demands money to recover it. Some types of scareware may lock your computer.

*3. Raas(Ransom as a service):* it is the malware hosted by the hacker to steal the files, information and system hostage and then requests for a ransom to return the access to the victim.

*4. Mac ransomware:* the ransomware called KeRanger attacked Mac operating systems. This malicious software infected apple user systems through the app named Transmission which encrypts the victim files after launched.

*5. Ransomware on mobile devices:* Ransomware attacked mobile largely in the year 2014. It enters through the malicious app that says it has been locked due to illegal activity.

**Who are the targets of ransomware attacks?**

Here are the four target groups:

a) Groups that are recognise as having lesser security teams: universities can come into this category because they have high file-sharing with less security.

b) Orgations that pays quickly: Government agencies, medical facilities, banks, etc this kind of groups can come into this category because they need quick access to their files. So they may willing to pay to get the access quickly.

c) Organisations that having sensititive data: firms that are having sensitive data are targeted because cybercriminals on the legal argunent could result if data is hacked and inform them about data is held for ransom.

d) Bussinesses in the Western market: Cybercriminals will target those who pay much i.e, corporate entities. This involves focusing on the United states and the united kingdoms, Canada due to greater wealth and pc use.

Example : Crypto Locker : This is a ransomware that was first seen in 2007. It enters via infected email attachments. Once it is entered it searches for a files to encrypt and hold it for ransom

It caused 5,00,000 computers. Security companies managed to seize the worldwide network of home computers that are used to spread the Crypto Locker. So by doing this the controlled the part of criminal network and snatch the data

Without knowing to the criminal. Later this actions led to development of online where the victim can get the unlock key and release the data without any payment

**Spyware:**



Spyware is the malicious or unwanted software that enters your device, stealing your internet data usage, and the sensitive information. It is mainly designed to get access over your computer without the user’s knowledge. It gathers your personnel information and relays it to the external users or to the other data firms. Its aim is to track your every data and sell your internet usage data. It also captures your credit card details and bank account information, or it steals your personnel identity.Some Spywares install other malicious software like trojan and keep backdoor as well.Even they can install a keylogger to keep track of credit card details and other personal details.

* According to Norton Cyber Security Insights Report Global Results. In 2017, there are 978 million people in 20 countries were affected by the cyber crime.
* Spyware is a threat to organisations and to the every individual users because it tracks the private data.

For example: key logger it is a malicious program that keeps the track of everything we type and sends to criminal who has created the key logger.

**How do I get spyware?**

Spyware can effect Pc’s, Android or IOS, Mac’s, devices and even windows are more suspectable to attacks.

These are the common ways to become to recognize whether spyware is present or not:

* Accepting pop-up or prompt window without knowing it.
* Downloading the pirated software.
* Accepting unknown sender emails attachments.
* Downloading movies, songs or games from untrustworthy source.

How to prevent spyware?

* These are the 4 ways to prevent from spyware.
* Don’t click on pop-up advertisements.
* Don’t download the software from unreadable source.
* Don’t open email attachments from unknown senders.
* Don’t download files from the untrustworthy source.

*Spyware example:*

Dark Hotel it targeted businesses leaders and government leaders using hotel WI-FI. They used different types of malware to gain the access to those powerful people’s systems. Once he got the access then attacker installs the keylogger to capture everything including the Usernames and passwords and some other sensitive data.

**Adware:** One of the most reason for adware is to collect the information about the user to make the advertisements based on that. It is called adware when it is on the computer and it is called as madware when it is on the mobile or tablet. This adware or madware slows down your systems and even it might prone to crashing.

1) Slow computer: it takes longer time to launch and run the programs or to load the documents or files, anything else you do, then computer have an adware. It slows down device processor and takes lots of memory space. In this way it decreases the overall performance.

2) Attack with ads: normally it is usual to get the advertisements when browsing. It’s even normal to those ads that related to something you searched before. If the ads are getting continuously and unable to close those or they are redirected to full page ads, then your system is infected with the spyware.

3) Constant crashing: If the programs are continuously crashing then your system is infected with the adware.

4) Browser home page changes: It is also known for changing the browsers homepage. It may redirect to a new page to install the more adware and other forms of malware on to the system.

5) Slow internet connection: It slows down the internet connection because it downloads lots of advertisements from the internet.

Adware example:

Fireball: it is a adware that infected 250 million computers and devices in the year of 2017. It hijacks the computer browser to change the search engines and starts tracking the web activity.

**Rogue software:**

It  is a form of malicious software and it can be considered as a Internet fraud because it misleads users into believing that there is a virus on their computer and aims to convince them to pay for a fake malware removal tool that actually installs malware on their computer .it manipulates users through fear and it is form of ransomware.

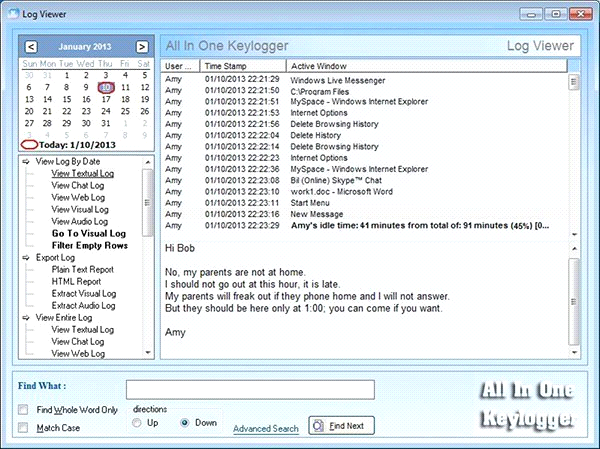
**Key loggers:**

A key logger (also known as keystroke logger) is software that tracks the keys stroke on your keyboard in a secret way so that user don’t know that actions are being monitored. This is usually done with malicious intent to collect account information, credit card numbers, user names, passwords, and other private data. It can cause serious threat to the users by intercepting passwords and some other confidential data entered via Keyboard. Most workstation keyboards plug into the back of the computer, so that user cannot see the connections. A key logger also enter as module which already installed inside the keyboard itself. Whenever the user types, everything will be collected and save as a text in its own small hard drive consists of several gigabytes of memory.



Fig: KEY-LOGGER

A keylogger software installed two files in same directory. One file was the DDL file which does all the recordings and the other one was an executable file which installs the DDL file and activate it to function. The keylogger program will record everything that user types and informs to one who installed that keylogger. Besides keystrokes, Some key logging programs also have the functionality to record the user’s data, such as noticing things which are copied to clipboard and taking screen-shots of the user's screen.



*Types of keyloggers:*

The keylogger is divided as the software based and the hardware based.

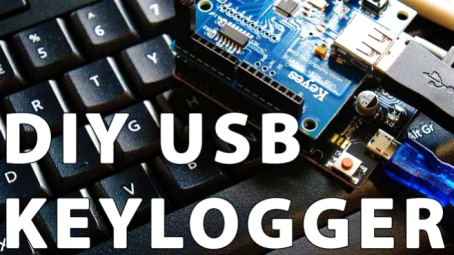
These are the common keyloggers with both of these categories

**1) API based keyloggers:** this keylogger software will use the API(application programming interface) keyboard to record all the keystrokes. Whenever you type a key then a notification will be sent to that application saying that you are typing so that every character will be seen on the screen. The API based keyloggers intercept notifications and captures everything as separate one. And the logs are kept on the small hard drive for easy retrieval to hacker.

**2) Form Grabbing based keyloggers**: This Grabbing based keyloggers logs the data from the web forms whenever user submits his/her form to the website. It is similar to the API based keyloggers as they breaks(interrupt) the submission notification and logs all the information which you have entered in the form. This includes like Full name, login credentials, email address, phone number of a user. This whole process takes places when the user clicks on submit button or Enter button and it is completed before the form submitted to the website.

**3) Kernel based keyloggers:** This kernel based keyloggers slows down the core of the computers kernel operating system which is very difficult to detect and remove. So this keylogger software hides inside the operating system and records every keystrokes as they pass through the kernel operating system. They are distributed through Rootkits, or through the other malicious software that bypass kernel operating systems and targets the hardware.

**4) Hardware keyloggers:** These are the devices that use the electric circuit inside the keyboard to log the keystrokes. So these are built in keyboard and also available as the USB connector for PC’s and the Mini-PCI card for laptops. This keylogger will have the internal memory to store the keystrokes i.e, the hackers must have the physical access to the keyboard to get the data.



**5) Acoustic keyloggers:** Acoustic keyloggers are complex and rarely used. They utilize the principles of a acoustic crypt analysis to record the keystrokes on hardware. Each key whatever the user types has unique acoustic signature. Their might be a little bit of difference but individual signatures are analyzed through the statistical methods

*How key Loggers enter into a system:*

Key loggers will spread as the same as other malicious software. This keyloggers are purchased knowingly to threat someone. Keyloggers enter through the emails, instant messages or by visiting the malicious websites.

*How to detect key logger:*

Key loggers are difficult to detect. But it can be detected if the users has following issues with the systems.

Slow performance when browsing the web.

Mouse or keystroke pause or showing nothing whatever the user types or it shows the error screen when loading the graphics or web pages.

**Botnet:**

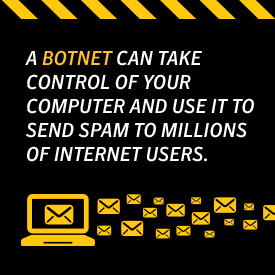
It is a good technology for bad intentions. Botnet is nothing but the string of connected devices to perform the tasks. It maintains the chat rooms or takes the overall control of computer. This botnets are dangerous in internet. They are the workhorses of the Internet. They are connected computers performing a number of repetitive tasks to keep websites going.

***What happens if malicious botnets attacks the system?***

In such situations system is directly hacked but sometimes a program crawls the internet to exploit the security and starts hacking automatically.

***What botnets do in the system?***

The botnets adds user to their website. It happens through downloading the drive-by software or by tricking you to download the trojan horse on the computer. Once it is installed it gets contact with its master computer and let it know that everything is ready. Therefore the computer is now under control of the one who created the botnet into the system.



Once the owner of botnet gets the control of the user’s computer, they use the user systems to perform some criminal tasks.

These are the common tasks that are executed by botnets:

* Launching distributed denial-of-service (DDoS) attacks to shut down websites.
* Sending spam emails to millions of internet users.
* Generating the fake traffic over the internet over a third party websites for financial gain.
* Replacing banner ads in user web browser targeting the user.
* Pop-ups ads that are designed to get user to pay for the removal of the botnet through a phony anti-spyware package.

**MALWARE DETECTOR**

A malware detector is used to detect the malware present in the system by using some detection techniques. Malware detector protects the system from malicious activity.Its not necessary that malware detector sits on the same system on which it tries to detect the malware .Sometimes it resides on another system also and try to detect the abnormal behaviour of the system.By using some detection techniques it checks for malware.Malware detectors take two inputs. First one is the knowledge to recognize the maliciousness of the program and the other is program which has to be tested.Once the malware detector knows what is malicious behavior(abnormal behavior) and the program under inspection, it can use its techniques to identify whether the program is malicious or not. Techniques used for detecting malware can be categorized into two categories anomaly-based detection and signature-based detection.

**MALWARE DETECTION TECHNIQUES**

Malware detection Techniques can be divided into anomaly based detection and signature based detection.Anomaly detection is again divided into specification-based detection. Specification-based systems uses some rule set to determine a valid behavior.If a behavior is valid then it is termed as un-malicious and normal.If not,then there is some anomaly and it turned as malicious.Signature-based systems used some characterization to determine the malicity of a program. Figure I shows the relationship between the various types of malware detection . Each of the detection techniques can have one of three different approaches: static, dynamic, or hybrid (see Fig 1).As anomaly and specification based malware detection checks for malware by gathering information and then analyzing.The static approach uses the syntax and structural based programs to determine the malware. Like in signature-based detection,the static approach looks for the amount of bytes the program is taking,the syntax of the codes to determine the maliciousness in the program whereas the dynamic based approach looks for the run time environment.It means if checks for the malicity before execution of the code or may be under execution of the code and the hybrid is the combination of the both static-based and dynamic-based.

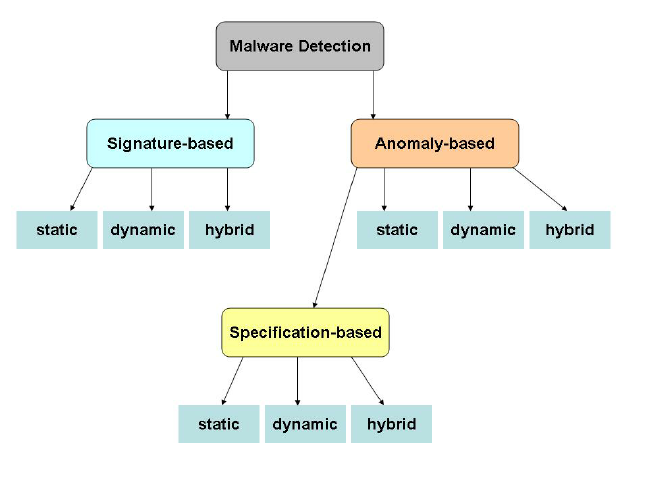


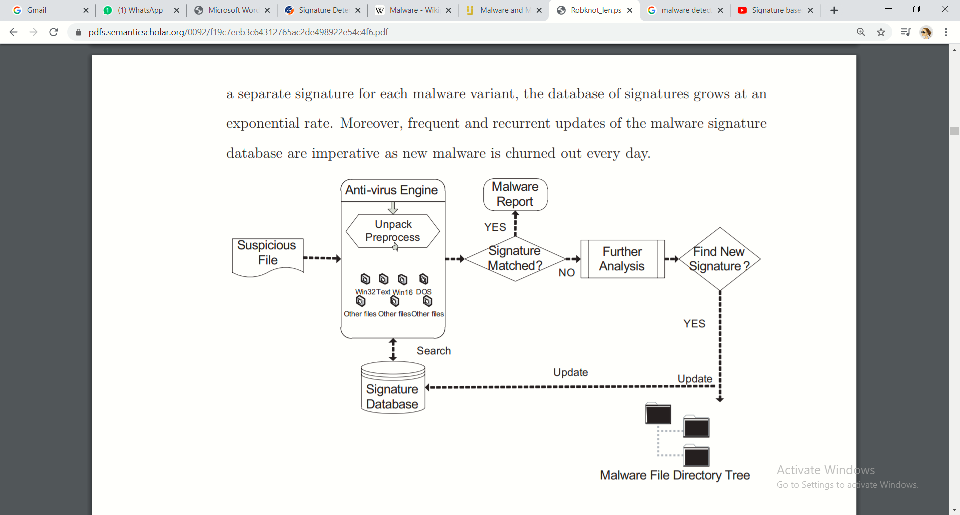
Fig I: A classification of malware detection techniques

The remaining part will describe about the anomaly-based detections with some examples and followed by specification-based and signature based.

**SIGNATURE BASED DETECTION**

It is also called as Misuse detection. It maintains the database of signature(signature is nothing but a series of bits that make a unique number which is used to detect fingerprints and identify specific viruses) detects malware by comparing pattern with the database. Mostly Antivirus software is designed by using this technique . All files are translated to high level language code to find the signatures and this signatures are stored in signature database. Then this is code is examined and some of the characteristics are taken out. Then these characteristics are used to construct the signature of a malware . when the antivirus software is installed in a computer then signature of a familiar code is updated and refreshed continuously so that this it can detect the known occurrences of malware correctly. The advantage of this technique is that it can identify the known occurrences of malware correctly and it takes the less number of resources to identify the malware.The disadvantage of this technique is that it cannot detect the new and unknown instances of malware because we cannot find the signatures for that types of malware.

*How does the flow of signature based detection works?*



The suspicious file is sent to the Anti-virus Engine, then it checks whether the file is packed or unpacked ,if the file is packed then it unpacks the file and send that file to scanner otherwise if the file is unpacked then it directly sends that file to scanner without making any changes. this unpacking is a built in function in Antivirus engine. now Antivirus engine asses the code which is in assembler language(human readable form)of a file with the malware signature database.if the file signature or a pattern is matches with the database then that file is considered as harmful. then a report is generated .if that file signature or a pattern doesn’t match with the database then that file is send to further analysis to find a new signature ,if they find a new signature then that new signature is updated in the signature database so when another file arrives it also undergo this process it also checks with new signature ,similarly this process continues when a new file is encountered . so this is the process of information flow of signature based detection.

*Types of Signature based detection:*

* Static Analysis of Signature based detection
* Dynamic Analysis of Signature based detection
* Hybrid Analysis of Signature based detection

*Static Analysis of Signature based malware detection*

In this the program is examined for the some sequence of code so that it would give information about the main aim of malicious program.It is used to give the approximate run time behavior of a program.it uses the its knowledge and compares the program with the known signatures to find whether the program is malicious or not. Mostly the malware is spread in the form of binary code which is used to provide the information about a program .

**Advantages and disadvantages**:

**Advantages:**

**1)More Time and Resource Consumption:** In static analysis we won’t run the malware because it takes more time and more resources are required for it. In static analysis the time taken to break up the code is associated to size of the code.

**2)Global View:** Global view of a code is nothing but overall view of code.The use of global view in static analysis is that it can identify the logic of executable code without running it.

**3)Easily Accessible Form:** In static analysis the disassembled code is inspected and it is converted to assembler instructions like in the form of functions which is easily accessible

**4)Stable and Repeatable:** The code which is created during static analysis is stable which means on that code it is easy to apply new techniques on it. The repeating procedure of disassembling provides the flexibility to static analysis.

**5)Safety and Data Independent:** If Once we take the disassembled code information of the file then we do not need to run the original files again. So in this way it provides safety to our system.

**Disadvantages:**

**1)Limitation of Software Reverse-engineering(copying)Techniques:** In the modern malware program which is written in high level programming language if we change a single bit of a code in the malware program then it causes greater change in binary code.

**2)Susceptible to Inaccuracies due to obfuscation and polymorphic Techniques:** The automated obfuscation(It is programming technique in which intentionally code is made unclear to prevent copying) tools implement techniques like changing the order of instructions. By this malware designers will have an advantage and they easily create new malware .

**3)Conservative Approximation:** In this a technique is implemented with some estimations which are not changeable and this estimations may lead to a miscalculation.

*Dynamic Analysis of Signature based detection*

In Dynamic signature-based malware detection the malware is detected by using specialized monitoring mechanisms.The malware is identified by API calls which are made by executable. Application Programming Interface is an communicating layer between Windows environment and executable. Because of some unclear techniques which are used in malware we cannot separate the part of a program but there is a chance that we can separate a part of program by using API calls in following ways they are:

1. *Data preprocessing:*

The programs which are executing uses the API calls. Like to search for a file which is infected by the virus they call FindFirstFileA, FindNextFileA and FindClose. So by these sequence of API calls we can generate the signature for the malicious program. To trace the API calls the software was created which monitors and displays API calls and their services was used. The output of this method is given as file which contains the list of API calls..

2.*Signature generation by API call tracing:*

The Signature for a malware is created by using critical API call The signature of a program based on API call tracing can be represented as two components they are

i)call frequency

ii) interaction of the critical API calls.

Call frequency is used to show how a critical calls are distributed,and interaction of the critical API calls is used to show the mapping nature of interaction of malware. To show mapping nature we use directed graph.All critical API calls are divided into some classes to create the signature of malware.If we have API calls like CallNextHookEx , isDebuggerPresent and CreateProcess then we will get the AFH as the part of the signature. By combining API calls into classes it will allow to represent the functionality as set of functions.

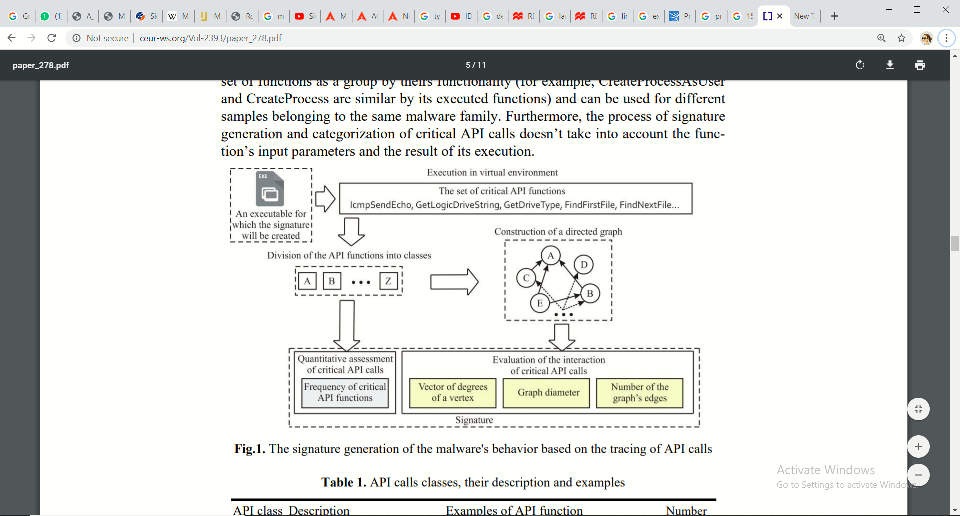
. 

Fig: Creating a signature for malware by tracing the API calls.

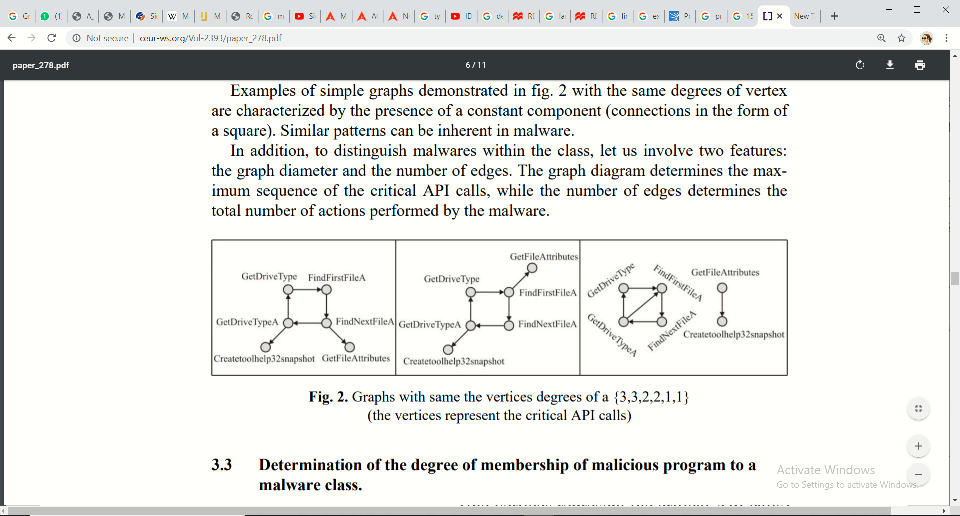
3*.Determination of the degree of membership of malicious program to a malware class:*

It shows the relationship between the malware program and one of the malware classes in form of the number of critical API calls. it will allows separating the malware programs within the class.To know the difference between malware within the class and without class we create two features for a graph they are

a)the graph with diameter

b) number of edges.

The graph diagram shows the maximum sequence of the critical API calls and the number of edges shows the total number of actions performed by the malware.



4.*Determination of the degree of membership of malicious program to a malware class:*

In this signature needs the frequencies of the critical API calls and those calls should contain the degree of membership of malicious program to a malware class. The process for the degree of membership of malicious program to a malware class is depends upon the difference between the number of API calls performed by malicious program and applications in the process of their own execution.

5.*Detection of a malicious program represented by signature of program behavior based on API call tracing*:

To detect the malicious program we need to find the difference between the frequencies of the critical API calls of a program and the frequency of critical API calls for each class. Then we get membership degree of the program to each of the classes as output. Then the signature which is matching with class can be identified by finding the minimum value of difference between membership degree to each of the malware classes.

Rule based IDS Approach

In this we use the state transition diagrams to know the attacks that are done by malicious code.The data is send to the preprocessor which changes the data format so that it can be analyzed by transition diagrams.and this data is compared with state diagrams.

Behavioral Approach to Worm Detection

In this we use different kind of behavioral signatures. one is base signature. base signature can be identified by observing the data flow from inside and outside of a node. when a server changes into client it is considered as base signature. so worm must be act like a client to other host to that it can infect the more systems with same functionality.

Alpha-in and Alpha out is other type of signature. in this worms will send the same kind of data to various nodes .if that data is not same then it will enter and exit the data flow links.

Another type of signature is fan out. It keeps a minimum value for number of connectors a host can have at a given time.To create a signature for this type of a behavior we need to keep minimum value on the following

(a) trees depth

(b) Number of successors in the tree

(c) Average branching factor

(d) Time taken to reach a particular tree depth

**Advantages and disadvantages:**

**Advantages:**

**Effectiveness and precision:** It uses the real values during run time so it gives the effective results.and it gives exact runtime behavior of code. **Simplicity**: Only single path execution is considered in this dynamic analysis so it is more simpler than static analysis.

**Runtime Behavioral Information:** The malware is assessed by using the runtime information given by code during dynamic analysis.

**No Unpacking:** In this unpacked code is executed automatically and runs the code .so no unpacking is done in this method it directly executes the code.

**Robust to Code Obfuscation:** In this the obfuscated(unclear) code does not make any changes in the behavioral information which is collected during the execution of a code.

**Disadvantages:**

**1)Limited View of Malware:**It is not possible to check all execution paths and variable values during dynamic analysis so because of this reason it might not allow to view of the total malware

**2)Trace Dependence and Execution Time Period:** The main disadvantage of dynamic analysis is that it is trace dependent . The malicious behaviour ‘s can not be collected in dynamic analysis with in the given period of time.

**3)Lack of Interactive Behaviour Information:**malware runs automatically without any human acknowledgement or interaction.so it is lack of interaction with human beings.

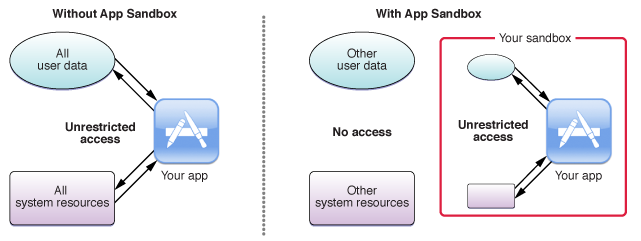
**4)Time and Resource Consumption**: In Dynamic analysis when we want to analyse a huge amount malware files it takes a lot of time and many resources are required for it.

*Hybrid Analysis of Signature based detection:*

Hybrid Analysis is a combination of static and dynamic analysis. It will combine run time data which is taken from dynamic analysis and static analysis to identify the malicious functionality in the applications. It combines the static features which are obtained while analyzing the application. So because of this the detection more accurately. it makes the system slow and the analysis process is time consuming. Tools used in hybrid analysis approach:

**Mobile Sandbox:**

It is a combination of static and dynamic analysis. In this we use analysis of APK file for static analysis. In this the anti viruses takes the user permissions and send the manifest.xml file for identifying the suspicious code that is scanned. In dynamic analysis an emulator is used for running the suspicious code and check the behavior of the code. The sandbox is a security mechanism for separating running programs. It is used to execute untrusted programs or code from untrusted without causing harm to the operating system.



1. **Andrubis :** In Andrubis framework to perform the dynamic analysis the result of static analysis is used .so first we need to perform static analysis and then dynamic analysis. In static analysis this framework is concentrated on android manifest.xml file and byte code. The information which comes from static analysis is used for dynamic analysis. In dynamic analysis they do analysis like stimulation, taint tracing etc.

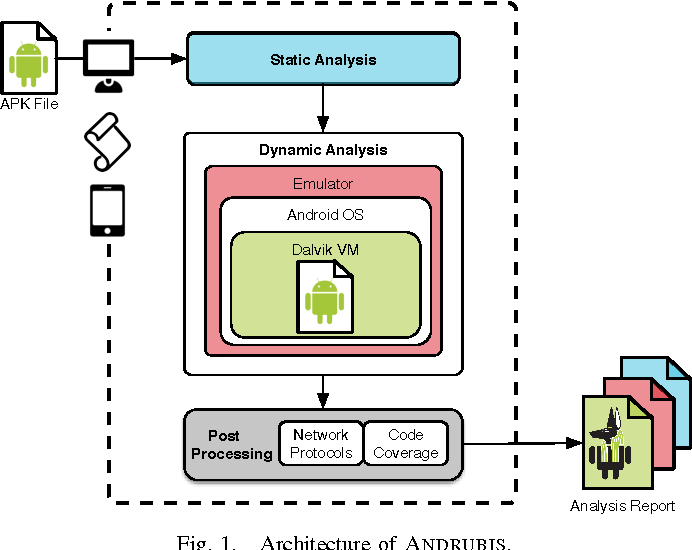


Fig: Andrubis framework

**ANOMALY - BASED DETECTION**

Anomaly detection process mainly consists of two phases - a learning(training) phase and a monitor(detection) phase.In the learning phase the detector tries to learn the normal behavior of the system when encounters by a operation. The detector can be trained by the behavior of the host or the PUI(Potentially Unwanted Installations (**spyware**)) or both.An advantage of anomaly based detection is the possession to detect the zero - day attacks.

***Zero-Days attack*** : The term zero day refer to the frangibility itself, or an attack that has zero days between the time the frangibility is happened and the first attack.

The two fundamentals drawback of this is *high false rate* and the *complexion involved in the training phase* to distinguish normal behavior and abnormal.

Fig-2 will tell us that why anomalies system is alone not good for malware detection.

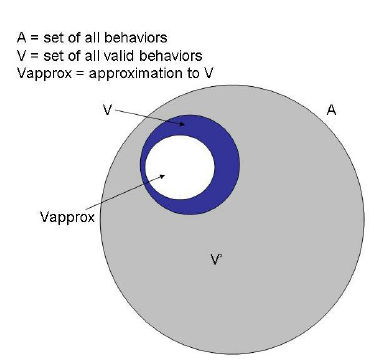


Fig-2: behaviour characterization of anomaly based detection

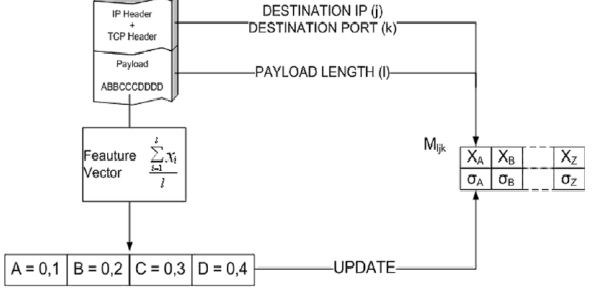
As shown, V is the set of valid behaviors of a system, and V’ is the set of invalid behaviors. Anomaly-based detection attempts to approximate the execution.The execution to valid behaviors is set as Vapprox which is shown in the above figure. valid behavior can be changed as malicious because Vapprox is an conjecture.

For example, if an estimation is never encountered in learning phase, an estimation seen during the monitoring phase gives an false alarm. It increases the high false rate with this detection technique.The count of unseen behavior in anomaly is not zero.Therefore, the probability of giving a *false positive* is not zero.

*Dynamic Anomaly -Based Detection Process*

In this detection process,data collected from the execution of program is used to perceive malicious code. Firstly the Inspection phase scans the program under inspection during its implementation, checking for maliciousness with what the detector has understood in the training phase.

*PAYL*

**

*Fig1 :PAYL Architecture*

Wang and Stolfo present PAYL which is the most prominent payload based anomaly based NIDS(N*etwork intrusion detection system,****NIDS****is a system that attempts to detect hacking activities, denial of service attacks or port scans on a computer network or a computer itself)* which calculates the expected payload on each service port of a system.

PAYL works in 2- step.First packets are classified according to the payload length then an n-gram analysis is applied to payload.The working of PAYL is byte frequency distribution which allows for a centroid model to be made for each. centroid model is calculated in learning phase. Then the detector contrast incoming payloads with the centroid model, and calculates the *Mahalanobis* distance between them. The Mahalanobis distance recognizes mean values of a feature vector, variance and co-variance resulting a strong statistical measure of similarity. So, If we have large Mahalanobis distance value i.e the incoming payload is having a large distance from centroid model then the incoming payload is malicious.

*Data Mining Approaches for Intrusion Detectors*

*Intrusion detection*: An **intrusion detection** system (IDS) is software application that keep track on a network for malicious activity or policy violations.

Lee and Stolfo suggested association rules and frequent event, for the use of intrusion detection systems. Frequent event as well as association rules are combine known as rule set. Rule set defines which behavior is normal and which doesn’t.Base detection emissary used some part of target. Meta detection used base detection and output of intrusions.The output forms audit data .On this audit data detection system execute the behavior.All authors use tcpdump data to learn the normal behavior.Through manual assessment of normal and abnormal data, they can identify that it can detect the possibility of abnormal activity.

*Short Sequences of System Calls*

Hofmeyr et al. advance a technique that inspect system call sequences to recognize maliciousness in the system. We need to make profiles to show the normal behavior . Normal is defined as short sequences of system calls. We need to calculate Hamming distance to tell how a system call sequence chose another. We need to set the threshold to determine whether its anomalous or not. If we get a large hamming value then its malicious.Authors can properly check encroachment that tries to stop many UNIX programs.

*Using computer Forensic methods for Privacy- invasive software*

Boldt and Carlson gives privacy-invasive software (PIS). Spyware and Adware are basic types of PIS. Boldt and team used the Forensic Tool Kit (FTK) to recognize PIS.Initially authors create a system free of PIS, a clean system. And this is considered the baseline of the system. And it is main target. If the baseline is taped, task is performed to PIS on the target.Like, surfing on the browser. And snapshot recorded at definite intervals and Ad-Aware is most popular PIS removal tool, and so we assess Ad-Aware by forensic and static analysis methods.In this technique, Boldt and his team noticed that Ad-Aware generates false positives as well as false negatives.

*FSA For detecting Anomalous programs*

Sekar et al. made a Finite State Automata (FSA) method to classify anomaly inspection. Every node in FSA constitute a state (program counter) in the PUI which the algorithm takes as input to learn valid behavior faster and good detection. Transitions in the FSA are done by *system calls(* *a****system call****is the programmatic way in which a computer program requests a service from the kernel of the operating system it is executed on)*The working of FSA based approach to detect anomaly is to construct the FSA ,for this the program is carry out multiple times. Whenever a system call is plead, a new transition is added to the automation.And system calls are obstructed, and the program state is recorded. If any error occurs,an anomaly occurred. After that, the algorithm sees for a valid transition from FSA’s current state to the newly cite system call. If no transition found then there is an anomaly. Else if above two steps executed properly then the FSA is moved to the next state.

Author compares it to n-gram method.And found that it have lower false-positive rate than n-gram.

*NATE*

Taylor and Alves-Foss present cost efficient approach for detecting anomaly. NATE stands for Network Analysis of Anomalous Traffic Events.Their method focus on attack which ruined network protocol destructibility.With an supposition that anomalous packets tend to have a large number of *syn*, *fin*, and *reset* packets, while having a low number of *ack* packets. The information flows from source IP to port is called as session.To check their technique they used the MIT Lincoln Labs data, namely the FTP, HTTP, and SMTP data. This technique can be different in real.Authors used Mahalanobis distance to estimate the distance between the known attacks and the normal.

*Static Anomaly based detection*

In Static anomaly based detection we can possibly get the maliciousness of the program without even executing it.

*Fileprint Analysis*

This process is proposed by Li et al. This analysis is also called as n-gram analysis for detecting the malware in files.As we know that a detection method undergoes in two phases - so in learning phase model or set of models are combined to characterize the various file types based on the structural and memory aspects.The author test this by considering a benign files that have regular byte composition for their respective type.For instance if benign file is having unique byte distribution then from .doc and .exe files.Then this can be marked as suspicious and it has to undergo several other methods to detect the malware presence.Li et al. experienced that requesting 1-gram analysis to PDF files inserted with malware productively to COTS AV scanner . There technique detection rates are between 72.1% and 94.5% percent for PDF file that had embedded malware, whereas the COTS AV scanner had a detection rate of zero effectively.If the malware is present then we get alarms before downloading it.

Since it is possible to keep the malware in between the .pdf .At that time the user must open the pdf to view content at that point also this method works.

*Hybrid Anomaly-Based Detection*

*Strider Ghost Buster*

Wang at el developed a malware detector which is also called as ghostware. It is a malware which secrete itself in the Operating systems .It hides himself by ambush the queries of them and alter so that it cant be traced.For example, if user ask to list the files in the current directory, then there is chance that ghostware delete any of its assets from results given by dir command.Author tells about*“cross-view diff-based*” method to inspecting these malware. They gave two ways of scanning for malware, firstly inside-the-box method and the other one is outside-the-box approach.There are many functions which return values which creates many opportunities for the ghostware to interrupt. The authors’ decided to compare the results from a high-level system call like “dir” to a low-level access(*Master File Table* (MFT) directly) of the same data without using a system call. This process is known the “cross-view diff-based” approach. In the inside-the-box method the high-level is compared with the low-level results running in the same machine.But what if the ghostware has taken the control in the operating system.At that case we can’t trust the machine .So we chose to do outside the box approach. In that another pure host do the low-level access without anyone’s knowledge and the comparison between clean host is done with high level host.If there is any difference between the high-level and low - level scans or vice-versa then ghostware is present.Mostly the results of inside the box approach is zero but there are chances that outside the box method produces some false positives.

**SPECIFICATION BASED DETECTION**

This process is derived from anomaly based detection. specification based detection approximates the requirements of application or system where as other detection techniques approaches the implementation of files or system. There is a training phase which aims to learn the all valid behaviour of a program or system which needs to be inspected. There is a tool called panorama which captures the system wide information flow of the program under inspection over a system and checks a behaviour against a valid set of rule to detect malicious activity. . Specification based Detection is based on the analysis of the behaviour that are described in the system specification. Known, unknown and new malware can be detected by the specification detection technique but it is a tedious process.

Specification based Detection is further divided into four types they are:

* Dynamic Specification based Detection
* Static Specification based Detection
* Hybrid Specification based Detection

*Dynamic specification Based detection process*

This method which comes under dynamic specification based their behaviour is observed during runtime to find the harmful programs.

Minority security-critical programs:

There is a monitor called as Distributed program Execution monitor. In this monitor, trace policies are generated for 15 unix programs and one trace policy was constructed for the remote file distribution client program.

ACT: Attachment chain tracing:

After that system admin will clear the number of layers and that should be recognized from any host i before manufacturing the decision of how many hosts are contaminated. Lets take an example in case system admin selects three layers to be the malicious at the moment of decision then hosts regarding to some host R at every layer would have to be proclaimed doubtful previous to all hosts would be cleared to be contaminated.

Automated detection of vulnerabilities in privilege programs:

Derived from a program specification its runtime behaviour concluded to whether it is hostile or not. Programs specification is interpreted into System calls that is differentiated to the system calls of the PUI. System call of the PUI are seized by the operating system.

Process Behaviour Monitoring:

In this program is physically interpreted into an Auditing Specification Language(ASL) and it is assembled into a C++ class and it links with an framework that represents system calls and it generates the system call detection engine. Every system call implored at run time is represented and it will sent to the system call detection engine. Detection engine compares with the system calls that are being made at run time against the ASL specifications.

Enlisting Hardware in Malicious code injection:

In this method the processor has its own stack known as Secure return address stack(SRAS). With the help of SRAS the processor identifies attacks based upon whether the stack in the memory is correspondent with SRAS. And SRAS must be placed securely so that it is not being adjusted by malware. The call and return address happens in a LIFO order but its not possible every time it means it does not follow the LIFO order. One way to handle this exception is not permit non LIFO behaviour and another one is to force the user to reassemble the application and only to allow specific types of non LIFO behaviour.

Mitigating XSS Attacks from the Client-side:

The easiest and the most effective **client**-**side** solution to the **XSS** problem for users is to deactivate JavaScript in their browsers. Unfortunately, this solution is often not feasible because a large number of web sites uses JavaScript for navigation and enhanced presentation of information.

Dynamic information Flow Tracking:

In this data is be specified as either safe or unsafe or genuine or fake. The privacy policy is marked by the operating system it recognizes the fake data. Operating system have done this by keeping a bit of already recognized data. The processor verifies that the control in the program is not unexpected.If it identifies the control is based upon fake data by verifying the bit given by the os with the recognized data,then processor will throws an exception that is seized by the operating system.And it destroys the process which made exception.

Using Instruction Block signatures:

In this some of the resources are allocated to processors to check that the most secure instructions are only being executed. In this method instruction block signatures are confirmed at runtime. And these signatures are coded by a secret processor key.

Fast detection of Scanning worms:

In this, entries are used to represent whether a host is clean or contagious or un contagious. A host is considered closer to an contagious entries for unsuccessful connection and moved further from the clean entry. Observation of scanning worms. Each host starts with 10 credits in credit based approach. A host is credited with the 2 credits for every successful connection and 1 credit is subtracted for every unsuccessful connection. And in case host connection credits placed to 0 then connection requests from that particular host request will be blocked.

Protecting Against unexpected system calls:

This method holds Interrupt Address table(IAT). The IAT carries the system call number with respect to the system call that are seem in the executable and also it carries the address after the system call.

Preventing SQL injection Attacks:

The first step in preventing a SQL injection attack is to establish which (if any) of your applications are vulnerable. The best way to do this is to launch your own attacks to see whether they are successful. But SQL is a complex language, so it is not a trivial task to construct code snippets that can be injected into a query to attempt to compromise a database.

*Static specification based detection*

The structural properties of the PUI are used by the static specification based detection to determine the maliciousness.

Static Detection of Malicious code in Executables:

Static analysis is carried out directly on the binary code. Static analysis offers techniques for predicting properties of the behaviour of programs without running them. The static analysis of a given binary executable is achieved in three major steps: construction of an intermediate representation, flow-based analysis that catches security oriented program behaviour, and static verification of critical behaviours against security policies (model checking)

Static Analysis of Binaries:

**Binary analysis** (**code** review) is a form of **static analysis** that deals only with the **binary** executable of an application without visibility into the source **code**.

While the reconstruction of the control-flow graph of a binary has received wide attention, thechallenge of categorizing code into defect-free

Compiler Approach to Mal code Detection:

The assembler takes input as source code and generates binary code and also annotations that contains types of assembly code. The annotations also contains information about the behaviour of assembly code. A component which is referred as verifiable certifies that the types agree with those which are found in the executable.The verifiable concludes whether the executable is secure, based on security policies. It is not allowed to execute if the executable is not secure.

*Hybrid Specification based detection*

Detection of malicious executables(DOME):

This method was introduced by the Robet et al. It was planned to expose infected and unclear code. DOME is classified into 2 types. In one step, detection of malicious executable uniformly pre processes the PUI. Pre processing includes (1)saving system call address, (2) names, (3)each system call followed by the address of the instruction. The third point preserved by the DOME are the system calls return address that are terminatable. In 2nd step, DOME perfects the terminatable at runtime,and also it certifies that all system calls that are made at runtime match with those recorded in a first step of static analysis accomplished.

Intrusion detection via static analysis:

This model examines source code. It finds a CFG that indicates its system call trace. Digraph model was introduced by Wagner and Dean where system calls are verified to check whether system call was malicious or not. It is always better to verify the system call arguments as it makes the system call more perfect and number of correct possible paths for a given model will be decreases.

*StackGuard:*

StackGuard is an extension of the assembler that verifies the changes that occurs to the active return addresses. It places a canary word.Before returning address canary word is verified if any changes are been observed then control will not transfered to return the address.

Spike:

Spike is a substructure introduced to users monitors the behavior of the files,with the aim of finding malicious behaviour. observation is processed by instrumenting the operating system services.The instrumentation is performed with the help of the drifters.drifters will have 2 components those are the code break point and the other one is the instrument.

**CONCLUSION**

In this paper we discussed about what actually malware is followed by the categories of malware.We find that though malware is troublesome there is a way to transpire and make system benevolent.There is vast history of malware as firstly it created its impact in 1970s and continues till now.But now we have methods to recognize malware.We saw several ways to identify the malware present in the system and define some techniques to remove them.We even disclose the target of various malware and wrote about the prevention from them. There are several techniques to detect malware and abandon it from causing harm to honest user’s system(can be mobile).Among the categories the anomaly based detection is having too less techniques to identify malware.Among each malware detection technique there is some pros and cons.The cons are giving false alarms of malware and considering a malicious behavior as normal.But these techniques won’t stop here.With growing world the researchers encounters more new ways to detect malware with cost-efficiently.