Software Security 2: Malicious Software

Kehuan Zhang

IERG4130 2022

Overview of Malicious Software (Malware)

What is malicious software

- Also called Malware

  "a program that is inserted into a system, usually covertly, with the intent of compromising the confidentiality, integrity, or availability of the victim's data, applications, or operating system or otherwise annoying or disrupting the victim"

- Classification of Malware
  - How it spreads or propagates
  - What it (or payload) will do

Malware Propagation Mechanisms

- Exploit software vulnerabilities of existing system (remotely)
  - E.g., worms that exploit vulnerability of networking code
  - E.g., drive-by download attack leverages vulnerability in browsers
- Exploit weakeness of Human beings
  - Play tricks to convince victim users open untrusted files or install suspicious software
  - E.g., openning a Word document which contains malicious VBscript code
- Via file systems (once have got inside a vulnerable system)
  - E.g., Virus hidden in an executable file can affect other files

Malicious Payloads

- Corruption of system or data files
  - E.g., damage disk boot sector so disk partition table would be lost
  - Ransomware: encrypt useful document and ask for money
- Theft of service and resources
  - E.g., zombie of a botnet, code that mining bitcoin at background
- Theft of information
  - Compromise confidentiality
  - E.g., key logger, sending files back to servers controlled by attacker
- Stealthy presence for remote control in future
  - APT (Advance Persistent Threats)

Classical Malware Categories

- Virus: infect executables to spread (often through file system)
  - Just like the flu virus which will host cells to reproduce themselves
- Worms: infect machines via networks to spread
- Trojan: infect machines when executed by careless users
  - Emphasize the fact of malicious payload hidden inside
- Rootkit: infect kernels to avoid detection and removal
  - Normally require root privilege

What is Virus?

- Malicious code that propagate through infected programs / software
  - Will infect other programs that it can access and modify
  - Just like biological viruses: a small piece of DNA or RNA that will replicate itself many times by destroying host cells
- Virus can acquire the same privilege as the infected program
  - Often hidden inside the host application
  - Perform malicious operations stealthily
- Normally viruses are specific to certain OS and platform
  - Virus for Windows will not infect programs for Linux or Mac OS
  - But there are cross-platform virus
    - ★ Some are specific to document types, like PDF, Word, etc.
    - ★ Why? Because the malicious payload would be written in platform-independent languages (like JavaScript or VBScript)
  - Virus inside open source code?
- Components of a Virus
  - Infection Mechanism: the means by which a virus will replicate itself
  - Trigger: the condition that will activate the malicious payload
  - Payload: the malicious behaviour of a virus besides replicating itself

Common Infection Means or Targets

- Boot sector
  - Disk Boot Sector normally will be the first piece of code to be executed
  - It can infect objects inside the OS to be booted
  - But also goes beyond the scope of the OS, i.e., hard to detect and removal
- Executable files
  - The malicious code will have a chance to run whenever the infected host program was double-clicked and executed
- Macro
  - Some document format may support embedding of dynamic code (i.e., Macros)
  - E.g., Word can support VBScript, PDF can support JavaScript
  - Just like infection of executable files, such documents with malicious macros will leverage the power of host software (like Office or PDF Reader)
- Combination of different means
  - Make such viruses even more sophisticated and difficult to be removed completely

Life Cycle of a Virus

- Dormant phase
  - Virus may have already infected certain files and system, but does not do anything malicious
- Propagation phase
  - Virus will replicate itself by infecting other new programs and systems
  - Note that: each replicated copy may be different from parent copy, in order to evade detection.
  - The replicated copy in the new programs or systems will follow the same life cycle as its parent virus
- Triggering phase
  - Being activated by certain conditions
  - E.g., at certain time, or a event had occurred certain times
  - Just like a "logic bomb" - a digital bomb triggered by certain conditions
- Execution phase
  - Execute the malicious payload and perform malicious activities
  - E.g., steal information, send network message (to launch Denial of Service attack), etc.

Virus Strategies to Evade Detection

- Anti-static and anti-dynamic detection methods
- Encryption
  - The malicious payload will be encrypted before execution phase
  - Such encryption tools are often called *packers*
- Code obfuscation (anti static analysis)
  - To transform the malicious payload into a form that is hard to analyze
  - E.g., remove all meaningful information from the code (code stripping)
  - Or injecting some misleading or abnormal code (e.g., to cause misalignment)
  - Increase analysis complexity $\rightarrow$ values undecidable via static analysis
- Code mutations
  - Change its code to a different appearance or fingerprint
  - It is normal to see that anti-virus software will often fail to catch such kind of mutated ones
  - How to be more accurate? Try to use features that are invariant.
  - Any thought? –> behaviour? API Call sequence?
- Anti dynamic analysis
  - E.g. Execution environment detection (whether it has been debugged or run in a VM) or anti-debug

What is Worm?

- Malicious program that tries to exploit vulnerable network software to infect victim machines as many as possible
  - Mainly propagate through network
    - ★ E.g., exploiting vulnerabilities of programs on a server
  - It will try to use the infected machines as launching pad to infect all other machines that it can reach
  - Most are very aggressive, thus may used up significant of resources, especially network bandwidth
  - It may carry some malicious payload besides the code for propagations

Worm Propagation Mechanism

- E-mail or Web pages
  - E-mail: Worm.ExploreZip will send itself out via emails for propagation for E-mail
  - Web pages: Koobface will send fake Facebook messages to trick victims to watch a malicious video
- File sharing and Remote File Access
  - Similar to virus, but transmit itself via network instead of disks or USB storage
- Remote Login and Code Execution
  - Victim host may access to remote server

Target Discovery Strategy

- Random
  - Infected host will choose and attack target hosts with random IP addresses
  - Could generate huge traffic due to unsuccessful rate
- Hit-List
  - The work come with a pre-built list of target machines
  - Multiple worms may collaborate to avoid overlapping of targeting host
  - Highly targeted, very efficient
- Topological
  - Use the network information acquired from the hosts to decide the target
  - Email or phone contact list should fall into type
- Local subnet
  - Target machines are those reside in the same subnet
  - So that traffics will not go through the routers
  - And also may not trigger network firewalls to raise alarms

Example Worms in History

Name	Year	Details
Morris Worm	1988	The first worm that had significant impacts. Exploit a buffer overflow vulnerability. Author now is a professor at MIT
Melissa	1998	e-mail worm, first to include virus, worm and Trojan in one package
Code Red	July 2001	exploited Microsoft IIS bug, probes random IP addresses, consumes lots of Internet resource
SQL Slammer	Early 2003	exploited a buffer overflow vulnerability in SQL server, compact and spread rapidly
Sobig.F	Late 2003	exploited open proxy servers to turn infected machines into spam engines
Conficker	Nov 2008	exploits a Windows buffer overflow vulnerability most widespread infection since SQL Slammer
Stuxnet	2010	restricted rate of spread to reduce chance of detection, targeted industrial control systems

Other Infection methods 1 - Drive-By-Download

- What is "drive-by"?
  - Dictionary: 'The act of shooting someone from a moving car'
  - In Cyber Security: hacking users even though they are just browsing Web pages
  - More formal: exploit browser (or plug-in) vulnerabilities to download and install malware on the system when the user view a Web page controlled by the attack
- Features of drive-by download attack?
  - Unlike worms or virus, it does not try to propagate aggressively
  - It spreads and infects users who visit certain malicious Web page
  - Mainly exploit vulnerabilities inside browsers
  - Stealthy: victims were not aware of being infected (and no user actions were required)

More about Drive-by-Download Attack

- Details before victims were hacked
  - Attacks need to scan and searching for vulnerable Web site
  - Then implant (embedding) some malicious code (e.g., JavaScript) into Web pages of that vulnerable Web site
  - The malicious code would break browser's sandbox protections
- How to attract victim visits?
  - Hacking popular Web site (with large clicking volumes)
  - Malicious Advertisement
  - May even setup a malicious Web site (e.g., pirate content, pornography)
- It was very popular sometime ago
  - Because of vulnerabilities in Browsers, especially for Microsoft IE
- Methods to Evade Detection
  - Target selection
    - ★ Randomly
    - ★ or strategically (e.g., based on IP address)
  - One of my previous work on mal-advertising - "Understanding and Detecting Malicious Web Advertising" (ACM CCS 2012)

Other Infection Methods 2 - Clickjacking

- Hijack user clicks
  - Typically for set up a transparent over a legitimate page
  - When user clicks, he/she may intent to click legitimate page at bottom
  - But the malicious page at upper layer will intercept the click
  - Be used to download malware, redirect user to other places, or change security settings
- Can also be used to steal sensitive information
  - Bottom page maybe bank login page
  - But user's typing (account and password) will be intercepted by top malicious layer

Other Infection Methods 3

- Social Engineering: to trick users to assist the infection / hacking
- Spam
  - Normally, spam means unsolicited emails
  - Some spam may just carry junk information
  - But many of them may carry malware (malicious .doc or .pdf files)
- Trojan horse
  - Programs or software that contains hidden code
  - Will be triggered if executed / opened by careless users
  - The hidden code can also be used for infection (by download and install malicious payloads)
- Mobile phone Trojan
  - Becomes very popular in past several years (since introduction of iPhone)
    - ★ Mobile phones contains lots of sensitive and valuable information
  - A major channel is free apps or repackaged apps
    - ★ What is Repackaged apps? How to avoid repackaged apps?

Types of Malicious Payload

Payloads Types

- System corruption
  - Example: CIH virus in 1998 (also called Chernobyl virus)
  - when triggered, not only infect executable files, but also BIOS
- Information theft
  - Keylogger or spyware
  - Mobile Apps are also typical examples: have you ever thought about the mysterious permissions request from Android apps during installation?
- Zombies and botnet
  - Zombies (or Bots) are computers taken over and controlled by remote attackers
  - Botnet is a collection of zombies which coordinate with each other under attacker's control
  - Botnet was very common and powerful (even for today)
    - ★ DDoS attack, e.g., the Mirai Botnet
    - ★ Spamming, Click Here for an example
    - ★ Spreading malware, e.g., download and instatll some software at background
    - ★ Manipulating network traffic, e.g., Click Fraud

Payloads Types (cont.)

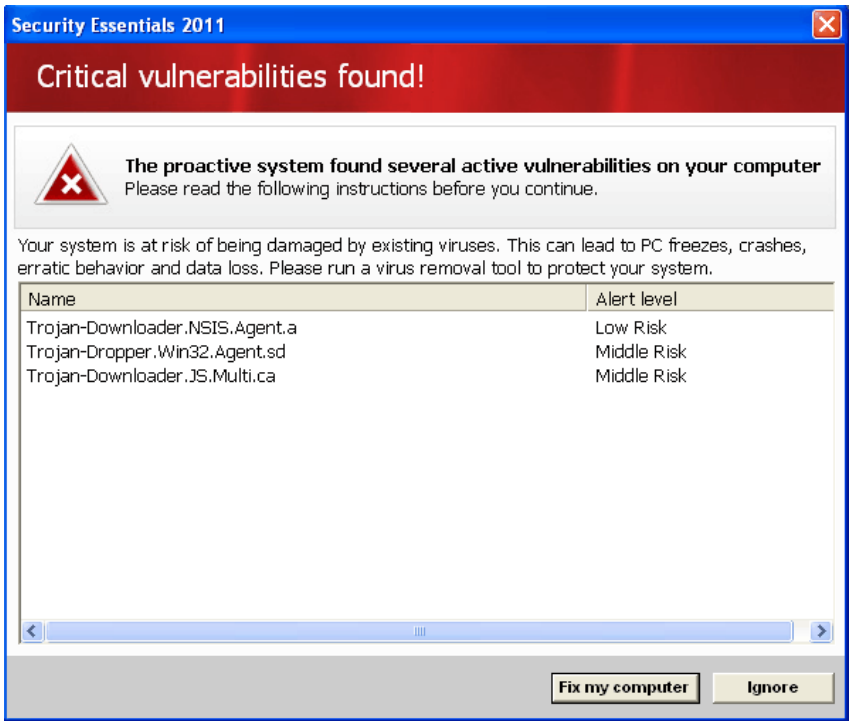
- Phishing
  - Try to leverage victim's trust on certain sources to steal sensitive information
  - E.g., email box is full and you need to upgrade (by login with your credentials)
- Difference between Spam and Phishing emails?
  - Spam emails in general are unsolicited emails. The sender may just want to send you a message, e.g., ads
  - Phishing emails are malicious ones that targeting to steal recipient's sensitive information
- Spear-Phishing
  - Highly targeted phishing attack
  - Recipients are carefully selected and studied
  - E-mails are crafted based on specific information of targeting recipient (instead of massive mailing of a generic version)

Payloads Types (cont.)

- Backdoor
  - Means a hidden entry that can bypass the authentication or access control mechanisms
  - Backdoor can result from malicious code or software vulnerability
    - ★ E.g., debugging code left behind
    - ★ E.g., Backdoor in some TP-Link routers, https://jalalsela.com/hacking-tp-link-tl-wr740n-backdoor/
  - Detecting malicious backdoor is hard, but it is possible for backdoor vulnerability
    - ★ A reserach work: "firmalice - automatic detection of authentication bypass vulnerabilities in binary firmware"
- Rootkit is a kind of backdoor to OS kernel
  - Malicious code has affected and resided in the kernel
  - To create and maintain a secret access to victim system (with root priviledge)
  - Can bypass conventional security mechanism

Payload Types (cont.)

- Rogue Security Software
  - Present fake virus scanning message
  - Will disable some useful system utilities, e.g., access to regedit, task manager, etc.



Payloads Types (cont.)

- Ransomware
  - Encrypt user data and ask for money for decryption

Malware Detection Techniques

Signature based Methods

- A signature is a set of data or code that can be used to uniquely identify a piece of malicious software
- However, malware may evolve and so the signature may change
  - Need to update the signature database frequently
  - Malware may purposely mutate itself (e.g., polymorphic or malware)
  - E.g., packer & unpacker, code-obfuscation, etc.

Behaviour based Methods

- Malware may change its appearance but needs to keep its malicious behaviour
- The key question is how to define or describe the behaviour
  - One example is using System Calls
  - Other examples: access to certain sensitive resources, request special permissions (e.g., trying to switch to privileged account)

Network based Methods

- Abnormal network traffic
  - initiate lots of network connections
  - try to explore network topology
  - large volume of outgoing traffic

New trend: Machine Learning based Approach

- Need lots of samples with labels of either malware or benign software
  - Maybe more refined labels, e.g., malware families
- The challenging part is how to convert the program into the form understandable by a ML algorithm
- Some related research works
  - Treat the input as a byte stream (one-dimension)
  - Convert into something similar to heat-map by calculating frequency of each possible byte (two-dimension)
  - Graphical models with network embedding
    - ★ extract some features from graphical representation of a program
    - ★ then form a feature vector to be fed into a machine learning model
- Limitation of Machine Learning based approach: interpretability
  - It is hard for human to understand and explain why the model will work in that way
    - ★ E.g., a piece of code is classified as malware, but why?

How to Protect Yourself - Recommended Practice

- Keep your system upgraded and patched
  - In real world, many machines has a long exposure window without being patched
  - There are researches where hackers could launch attacks based on information from patches
  - However, patching could not prevent 0-day (zero-day) attack (but zero-day vulnerability is expensive)
- Back up you data
  - Offline on a separate place, not a different folder on the same machine
- Do not click suspicious links or email attachment
  - How to recognize "suspicious" items?
- Use legal software instead of cracked ones
  - Downloading website may contain malware and launch drive-by download attack
  - Register code/serial number calculator tools may contain malware
  - For mobile platforms, there are App repackage attack
- If you thought it might be dangerous, then use virtual machine

Frequently asked questions

- Is it safer to use Mac OS or Linux than Windows?
- Shall we be worry-free if have anti-virus program installed?
- What is difference between firewall and anti-virus software?
- Any question from you?