Malware: Malicious Software

Viruses, Worms, Trojans, Rootkits

- Malware can be classified into several categories, depending on propagation and concealment
- Propagation
  - Virus: human-assisted propagation (e.g., open email attachment)
  - Worm: automatic propagation without human assistance
- Concealment
  - Rootkit: modifies operating system to hide its existence
  - Trojan: provides desirable functionality but hides malicious operation
- Various types of payloads, ranging from annoyance to crime

Insider Attacks

- An **insider attack** is a security breach that is caused or facilitated by someone who is a part of the very organization that controls or builds the asset that should be protected.
- In the case of malware, an insider attack refers to a security hole that is created in a software system by one of its programmers.

Backdoors

- A **backdoor,** which is also sometimes called a **trapdoor,** is a hidden feature or command in a program that allows a user to perform actions he or she would not normally be allowed to do.
- When used in a normal way, this program performs completely as expected and advertised.
- But if the hidden feature is activated, the program does something unexpected, often in violation of security policies, such as performing a privilege escalation.
- Benign example: **Easter Eggs** in DVDs and software

Logic Bombs

- A **logic bomb** is a program that performs a malicious action as a result of a certain logic condition.
- The classic example of a logic bomb is a programmer coding up the software for the payroll system who puts in code that makes the program crash should it ever process two consecutive payrolls without paying him.
- Another classic example combines a logic bomb with a backdoor, where a programmer puts in a logic bomb that will crash the program on a certain date.

The Omega Engineering Logic Bomb

- An example of a logic bomb that was actually triggered and caused damage is one that programmer Tim Lloyd was convicted of using on his former employer, Omega Engineering Corporation. On July 31, 1996, a logic bomb was triggered on the server for Omega Engineering's manufacturing operations, which ultimately cost the company millions of dollars in damages and led to it laying off many of its employees.

The Omega Bomb Code

- The Logic Behind the Omega Engineering Time Bomb included the following strings:
- 7/30/96
  - Event that triggered the bomb
- F:
  - Focused attention to volume F, which had critical files
- F:\LOGIN\LOGIN 12345
  - Login a fictitious user, 12345 (the back door)
- CD \PUBLIC
  - Moves to the public folder of programs
- FIX.EXE /Y F:*.*
  - Run a program, called FIX, which actually deletes everything
- PURGE F:\/ALL
  - Prevent recovery of the deleted files

Defenses against Insider Attacks

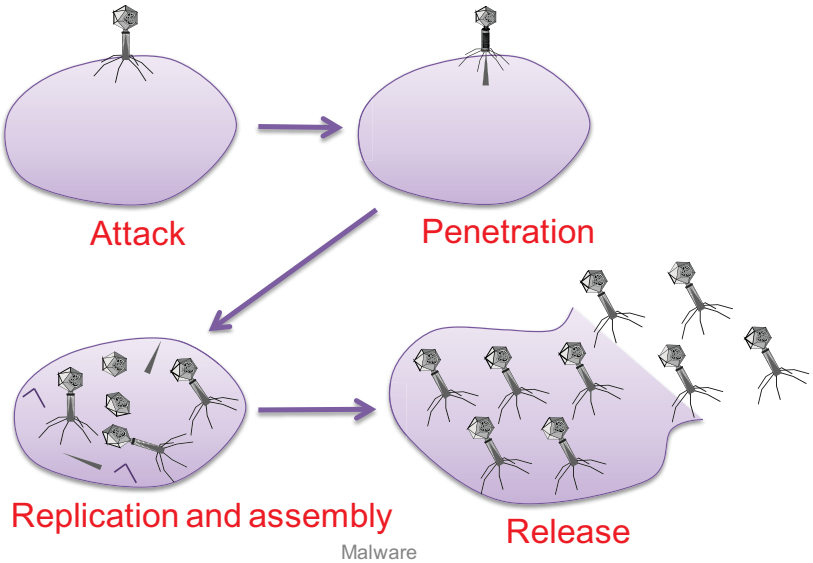
- Avoid single points of failure.
- Use code walk-throughs.
- Use archiving and reporting tools.
- Limit authority and permissions.
- Physically secure critical systems.
- Monitor employee behavior.
- Control software installations.

Computer Viruses

- A **computer virus** is computer code that can replicate itself by modifying other files or programs to insert code that is capable of further replication.
- This self-replication property is what distinguishes computer viruses from other kinds of malware, such as logic bombs.
- Another distinguishing property of a virus is that replication requires some type of **user assistance,** such as clicking on an email attachment or sharing a USB drive.

Biological Analogy

- Computer viruses share some properties with Biological viruses



Early History

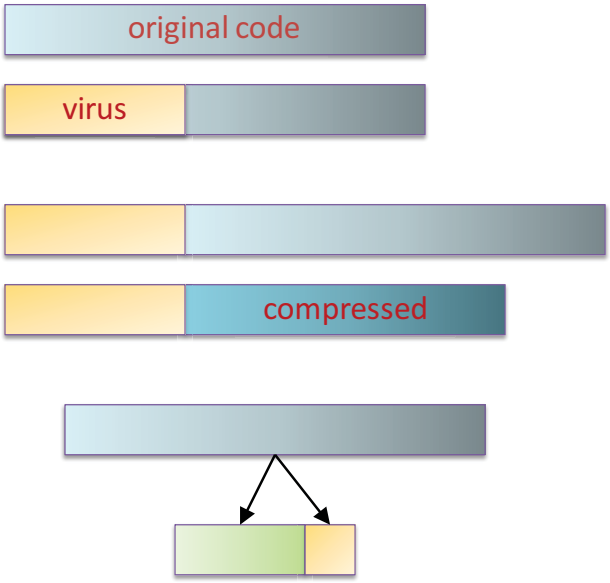
- 1972 sci-fi novel "When HARLIE Was One" features a program called VIRUS that reproduces itself
- First academic use of term virus by PhD student Fred Cohen in 1984, who credits advisor Len Adleman with coining it
- In 1982, high-school student Rich Skrenta wrote first virus released in the wild: Elk Cloner, a boot sector virus
- (c)Brain, by Basit and Amjood Farooq Alvi in 1986, credited with being the first virus to infect PCs

Virus Phases

- **Dormant phase.** During this phase, the virus just exists—the virus is laying low and avoiding detection.
- **Propagation phase.** During this phase, the virus is replicating itself, infecting new files on new systems.
- **Triggering phase.** In this phase, some logical condition causes the virus to move from a dormant or propagation phase to perform its intended action.
- **Action phase.** In this phase, the virus performs the malicious action that it was designed to perform, called **payload.**
  - This action could include something seemingly innocent, like displaying a silly picture on a computer's screen, or something quite malicious, such as deleting all essential files on the hard drive.

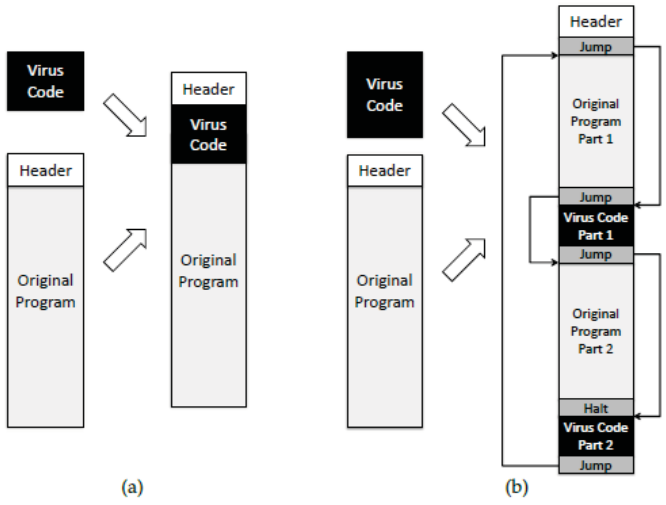
Infection Types

- Overwriting
  - Destroys original code
- Pre-pending
  - Keeps original code, possibly compressed
- Infection of libraries
  - Allows virus to be memory resident
  - E.g., kernel32.dll
- Macro viruses
  - Infects MS Office documents
  - Often installs in main document template



Degrees of Complication

- Viruses have various degrees of complication in how they can insert themselves in computer code.



Concealment

- Encrypted virus
  - Decryption engine + encrypted body
  - Randomly generate encryption key
  - Detection looks for decryption engine
- Polymorphic virus
  - Encrypted virus with random variations of the decryption engine (e.g., padding code)
  - Detection using CPU emulator
- Metamorphic virus
  - Different virus bodies
  - Approaches include code permutation and instruction replacement
  - Challenging to detect

Computer Worms

- A **computer worm** is a malware program that spreads copies of itself without the need to inject itself in other programs, and usually without human interaction.
- Thus, computer worms are technically not computer viruses (since they don't infect other programs), but some people nevertheless confuse the terms, since both spread by self-replication.
- In most cases, a computer worm will carry a malicious payload, such as deleting files or installing a backdoor.

Early History

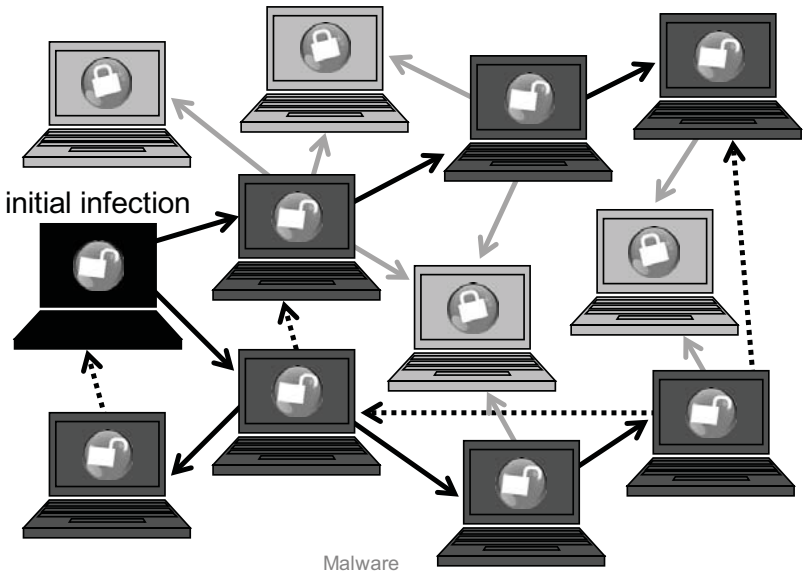
- First worms built in the labs of John Shock and Jon Hepps at Xerox PARC in the early 80s
- CHRISTMA EXEC written in REXX, released in December 1987, and targeting IBM VM/CMS systems was the first worm to use e-mail service
- The first internet worm was the Morris Worm, written by Cornell student Robert Tappan Morris and released on November 2, 1988

Worm Development

- Identify vulnerability still unpatched
- Write code for
  - Exploit of vulnerability
  - Generation of target list
    - Random hosts on the internet
    - Hosts on LAN
    - Divide-and-conquer
  - Installation and execution of payload
  - Querying/reporting if a host is infected
- Initial deployment on a private network
- Worm template
  - Generate target list
  - For each host on target list
    - Check if infected
    - Check if vulnerable
    - Infect
    - Recur

Worm Propagation

- Worms propagate by finding and infecting vulnerable hosts.
  - They need a way to tell if a host is vulnerable
  - They need a way to tell if a host is already infected.



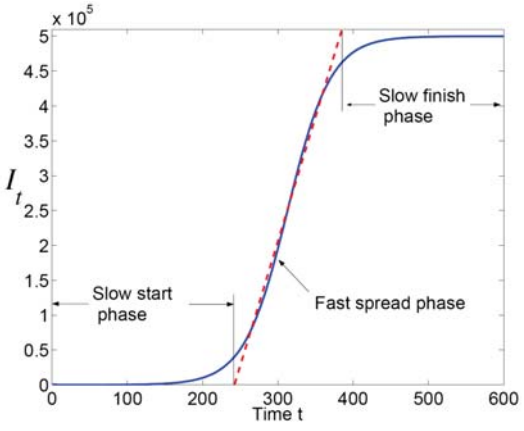
Propagation: Theory

- Classic epidemic model
  - N: total number of vulnerable hosts
  - $I(t)$: number of infected hosts at time t
  - $S(t)$: number of susceptible hosts at time t
  - $I(t) + S(t) = N$
  - β: infection rate
- Differential equation for $I(t)$:

$$dI/dt = \beta I(t)\, S(t)$$

- More accurate models adjust propagation rate over time

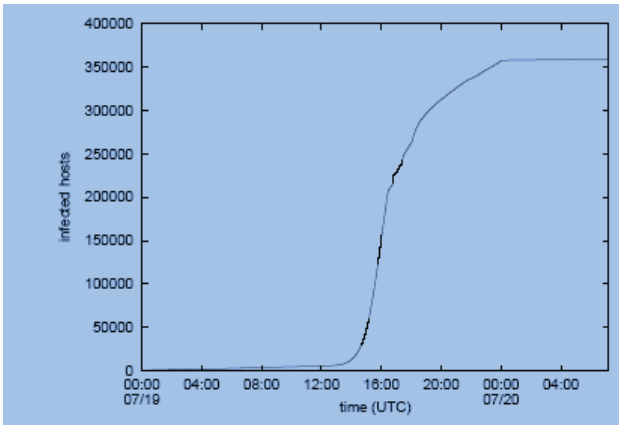
Source:
Cliff C. Zou, Weibo Gong, Don Towsley, and Lixin Gao. The Monitoring and Early Detection of Internet Worms, IEEE/ACM Transactions on Networking, 2005.



Propagation: Practice

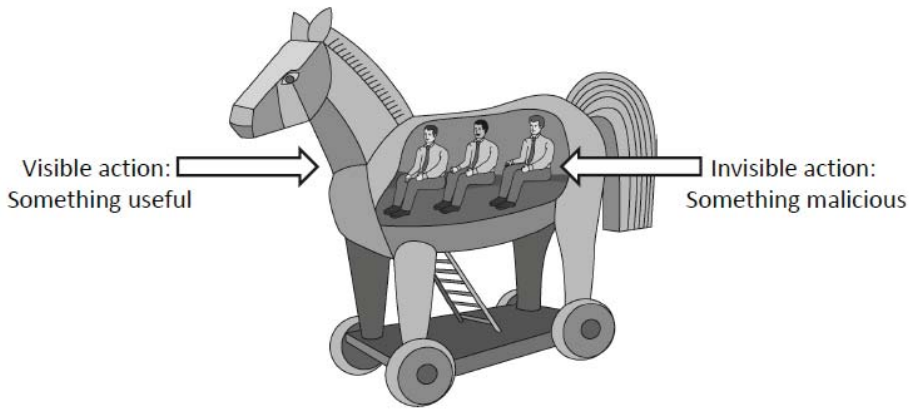
- Cumulative total of unique IP addresses infected by the first outbreak of Code-RedI v2 on July 19-20, 2001

Source:
David Moore, Colleen Shannon, and Jeffery Brown. Code-Red: a case study on the spread and victims of an Internet worm, CAIDA, 2002



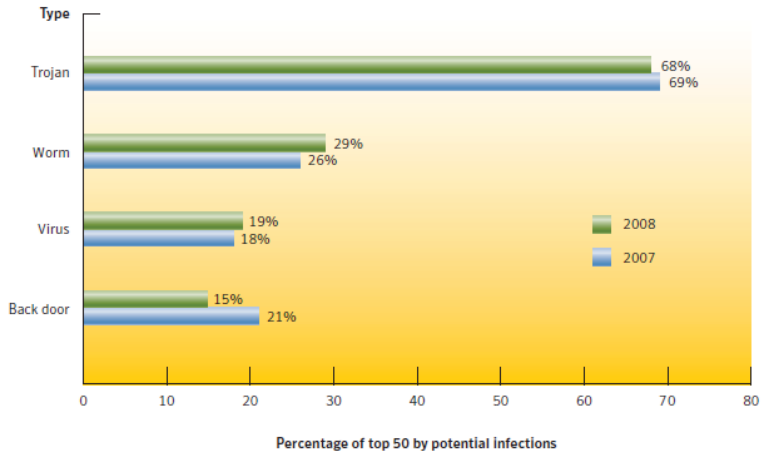
Trojan Horses

- A **Trojan horse (or Trojan)** is a malware program that appears to perform some useful task, but which also does something with negative consequences (e.g., launches a keylogger).
- Trojan horses can be installed as part of the payload of other malware but are often installed by a user or administrator, either deliberately or accidentally.



Current Trends

- Trojans currently have largest infection potential
  - Often exploit browser vulnerabilities
  - Typically used to download other malware in multi-stage attacks



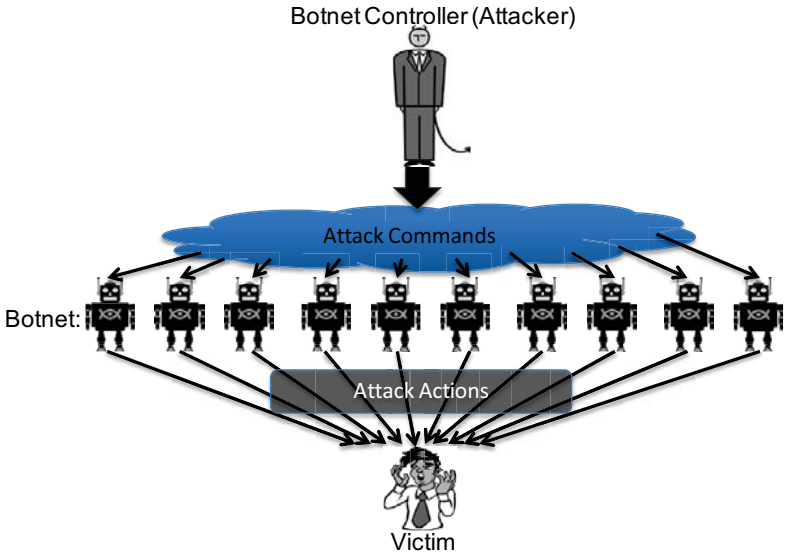
Source: Symantec Internet Security Threat Report, April 2009

Rootkits

- A rootkit modifies the operating system to hide its existence
  - E.g., modifies file system exploration utilities
  - Hard to detect using software that relies on the OS itself
- RootkitRevealer
  - By Bryce Cogswell and Mark Russinovich (Sysinternals)
  - Two scans of file system
  - High-level scan using the Windows API
  - Raw scan using disk access methods
  - Discrepancy reveals presence of rootkit
  - Could be defeated by rootkit that intercepts and modifies results of raw scan operations

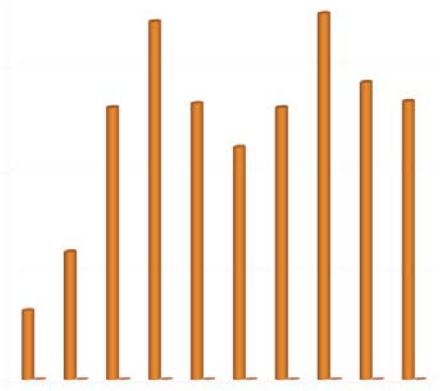
Malware Zombies

- Malware can turn a computer in to a **zombie**, which is a machine that is controlled externally to perform malicious attacks, usually as a part of a **botnet**.



Financial Impact

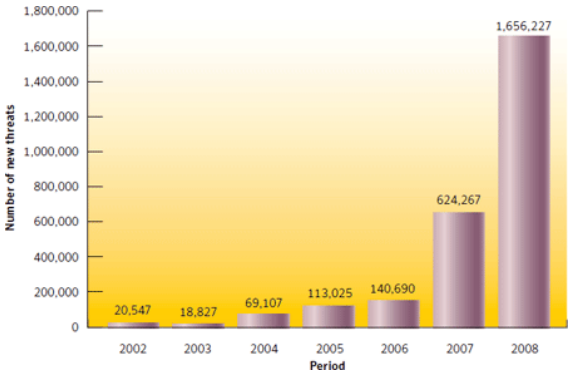
- Malware often affects a large user population
- Significant financial impact, though estimates vary widely, up to $100B per year (mi2g)
- Examples
  - LoveBug (2000) caused $8.75B in damages and shut down the British parliament
  - In 2004, 8% of emails infected by W32/MyDoom.A at its peak
  - In February 2006, the Russian Stock Exchange was taken down by a virus.



Economics of Malware

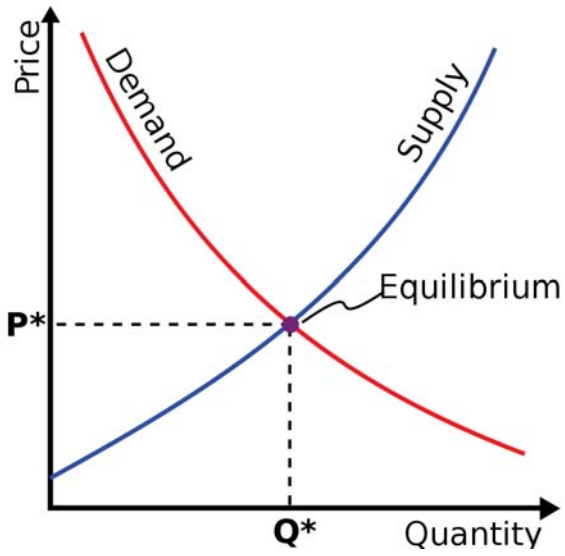
- New malware threats have grown from 20K to 1.7M in the period 2002-2008
- Most of the growth has been from 2006 to 2008
- Number of new threats per year appears to be growing an exponential rate.

Source: Symantec Internet Security Threat Report, April 2009

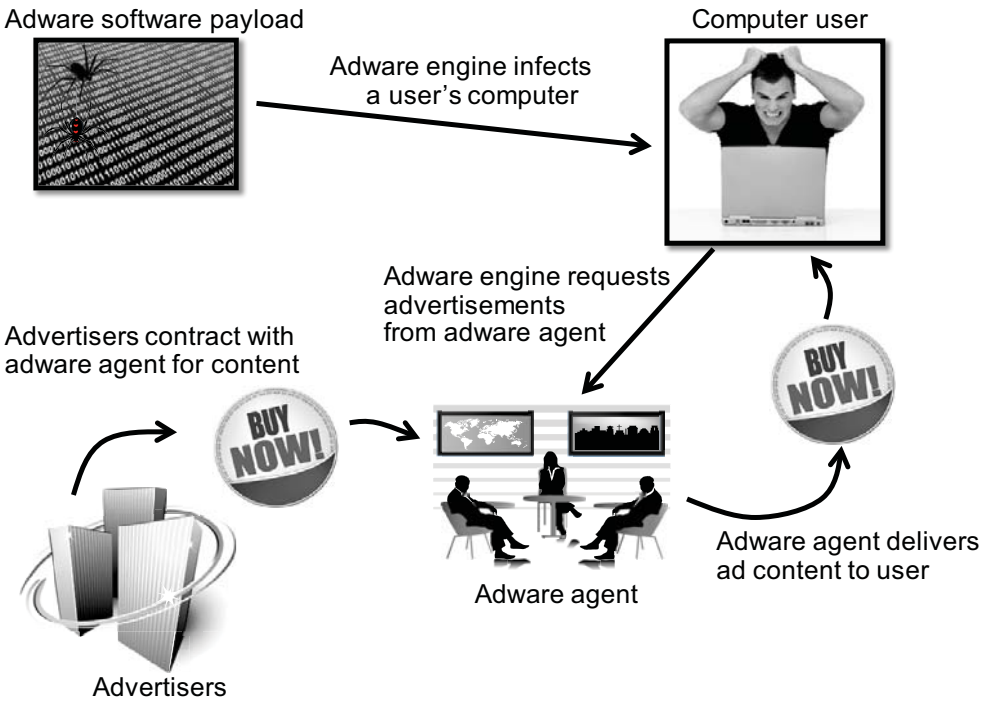


Professional Malware

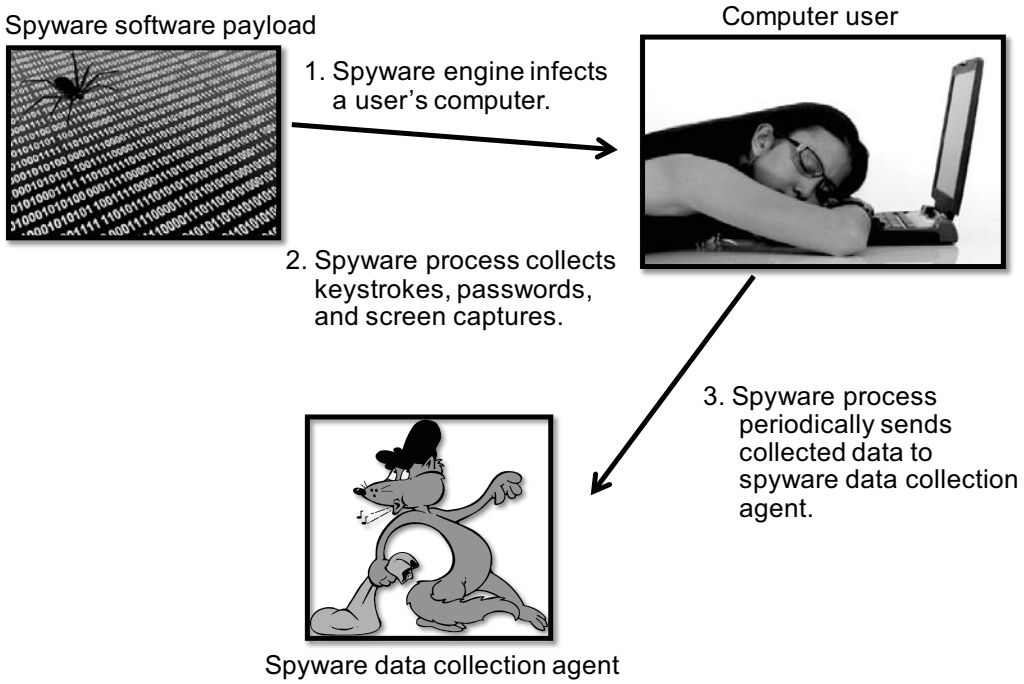
- Growth in professional cybercrime and online fraud has led to demand for professionally developed malware
- New malware is often a custom-designed variations of known exploits, so the malware designer can sell different "products" to his/her customers.
- Like every product, professional malware is subject to the laws of supply and demand.
  - Recent studies put the price of a software keystroke logger at $23 and a botnet use at $225.



Adware



Spyware

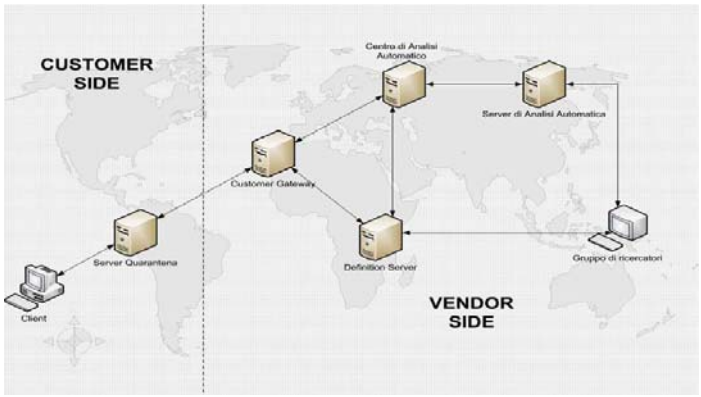


Signatures: A Malware Countermeasure

- Scan compare the analyzed object with a database of signatures
- A signature is a virus fingerprint
  - E.g.,a string with a sequence of instructions specific for each virus
  - Different from a digital signature
- A file is infected if there is a signature inside its code
  - Fast pattern matching techniques to search for signatures
- All the signatures together create the malware database that usually is proprietary

Signatures Database

- Common Malware Enumeration (CME)
  - aims to provide unique, common identifiers to new virus threats
  - Hosted by MITRE
  - http://cme.mitre.org/data/list.html
- Digital Immune System (DIS)
  - Create automatically new signatures



White/Black Listing

- Maintain database of cryptographic hashes for
  - Operating system files
  - Popular applications
  - Known infected files
- Compute hash of each file
- Look up into database
- Needs to protect the integrity of the database

Heuristic Analysis

- Useful to identify new and "zero day" malware
- Code analysis
  - Based on the instructions, the antivirus can determine whether or not the program is malicious, i.e., program contains instruction to delete system files,
- Execution emulation
  - Run code in isolated emulation environment
  - Monitor actions that target file takes
  - If the actions are harmful, mark as virus
- Heuristic methods can trigger false alarms

Shield vs. On-demand

- Shield
  - Background process (service/daemon)
  - Scans each time a file is touched (open, copy, execute, etc.)
- On-demand
  - Scan on explicit user request or according to regular schedule
  - On a suspicious file, directory, drive, etc.

Performance test of scan techniques
- Comparative: check the number of already known viruses that are found and the time to perform the scan
- Retrospective: test the proactive detection of the scanner for unknown viruses, to verify which vendor uses better heuristics

Anti-viruses are ranked using both parameters:
http://www.av-comparatives.org/

Online vs Offline Anti Virus Software

Online
- Free browser plug-in
- Authentication through third party certificate (i.e. VeriSign)
- No shielding
- Software and signatures update at each scan
- Poorly configurable
- Scan needs internet connection
- Report collected by the company that offers the service

Offline
- Paid annual subscription
- Installed on the OS
- Software distributed securely by the vendor online or a retailer
- System shielding
- Scheduled software and signatures updates
- Easily configurable
- Scan without internet connection
- Report collected locally and may be sent to vendor

Quarantine

- A suspicious file can be isolated in a folder called quarantine:
  - E.g,. if the result of the heuristic analysis is positive and you are waiting for db signatures update
- The suspicious file is not deleted but made harmless: the user can decide when to remove it or eventually restore for a false positive
  - Interacting with a file in quarantine it is possible only through the antivirus program
- The file in quarantine is harmless because it is encrypted
- Usually the quarantine technique is proprietary and the details are kept secret

Static vs. Dynamic Analysis

Static Analysis

- Checks the code without trying to execute it
- Quick scan in white list
- Filtering: scan with different antivirus and check if they return same result with different name
- Weeding: remove the correct part of files as junk to better identify the virus
- Code analysis: check binary code to understand if it is an executable, e.g., PE
- Disassembling: check if the byte code shows something unusual

Dynamic Analysis

- Check the execution of codes inside a virtual sandbox
- Monitor
  - File changes
  - Registry changes
  - Processes and threads
  - Networks ports

Virus Detection is Undecidable

- Theoretical result by Fred Cohen (1987)
- Virus abstractly modeled as program that eventually executes infect
- Code for infect may be generated at runtime
- Proof by contradiction similar to that of the halting problem

- Suppose program isVirus(P) determines whether program P is a virus
- Define new program Q as follows:

```
if (not isVirus(Q))
  infect
stop
```

- Running isVirus on Q achieves a contradiction

Other Undecidable Detection Problems

- Detection of a virus
  - by its appearance
  - by its behavior
- Detection of an evolution of a known virus
- Detection of a triggering mechanism
  - by its appearance
  - by its behavior
- Detection of a virus detector
  - by its appearance
  - by its behavior
- Detection of an evolution of
  - a known virus
  - a known triggering mechanism
  - a virus detector

Resources

- Computer Emergency Response Team
  - Research center funded by the US federal government
  - Vulnerabilities database
- Symantec
  - Reports on malware trends
  - Database of malware
- Art of Computer Virus Research and Defense by Peter Szor