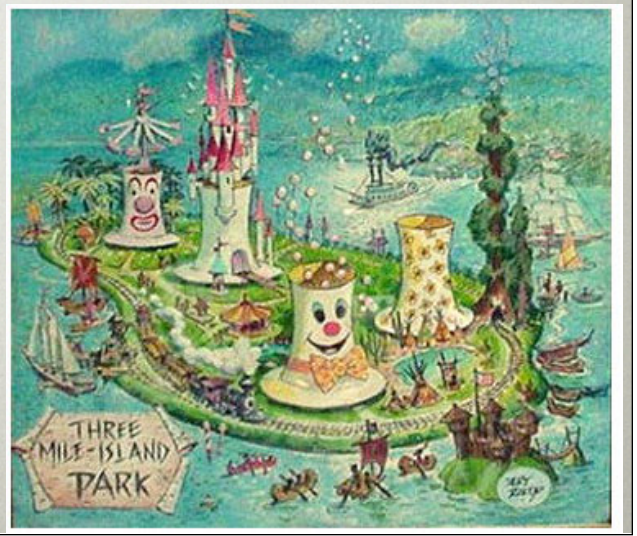
Malware

CS155 Spring 2009

Elie Bursztein

Welcome to the zoo

- What malware are
- How do they infect hosts
- How do they hide
- How do they propagate
- Zoo visit !
- How to detect them
- Worms



What is a malware ?

A Malware is a set of instructions that run on your computer and make your system do something that an attacker wants it to do.

What it is good for ?

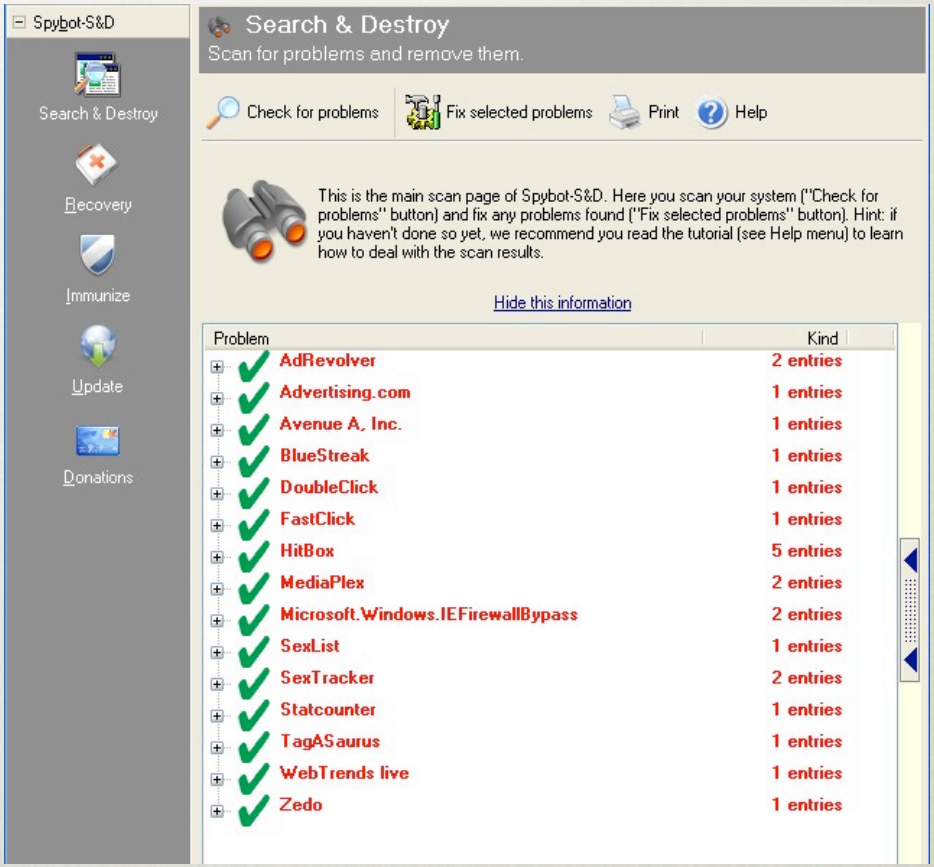
- Steal personal information
- Delete files
- Click fraud
- Steal software serial numbers
- Use your computer as relay

A recent illustration

- Christians On Facebook
- Leader hacked on march 2009
  - Post Islamic message
  - Lost >10 000 members

The Malware Zoo

- Virus
- Backdoor
- Trojan horse
- Rootkit
- Scareware
- Adware
- Worm



What is a Virus ?

a program that can infect other programs by modifying them to include a, possibly evolved, version of itself

Fred Cohen 1983

Some Virus Type

- Polymorphic : uses a polymorphic engine to mutate while keeping the original algorithm intact (packer)
- Methamorpic : Change after each infection

What is a trojan

A trojan describes the class of malware that appears to perform a desirable function but in fact performs undisclosed malicious functions that allow unauthorized access to the victim computer

Wikipedia

What is rootkit

A root kit is a component that uses stealth to maintain a persistent and undetectable presence on the machine

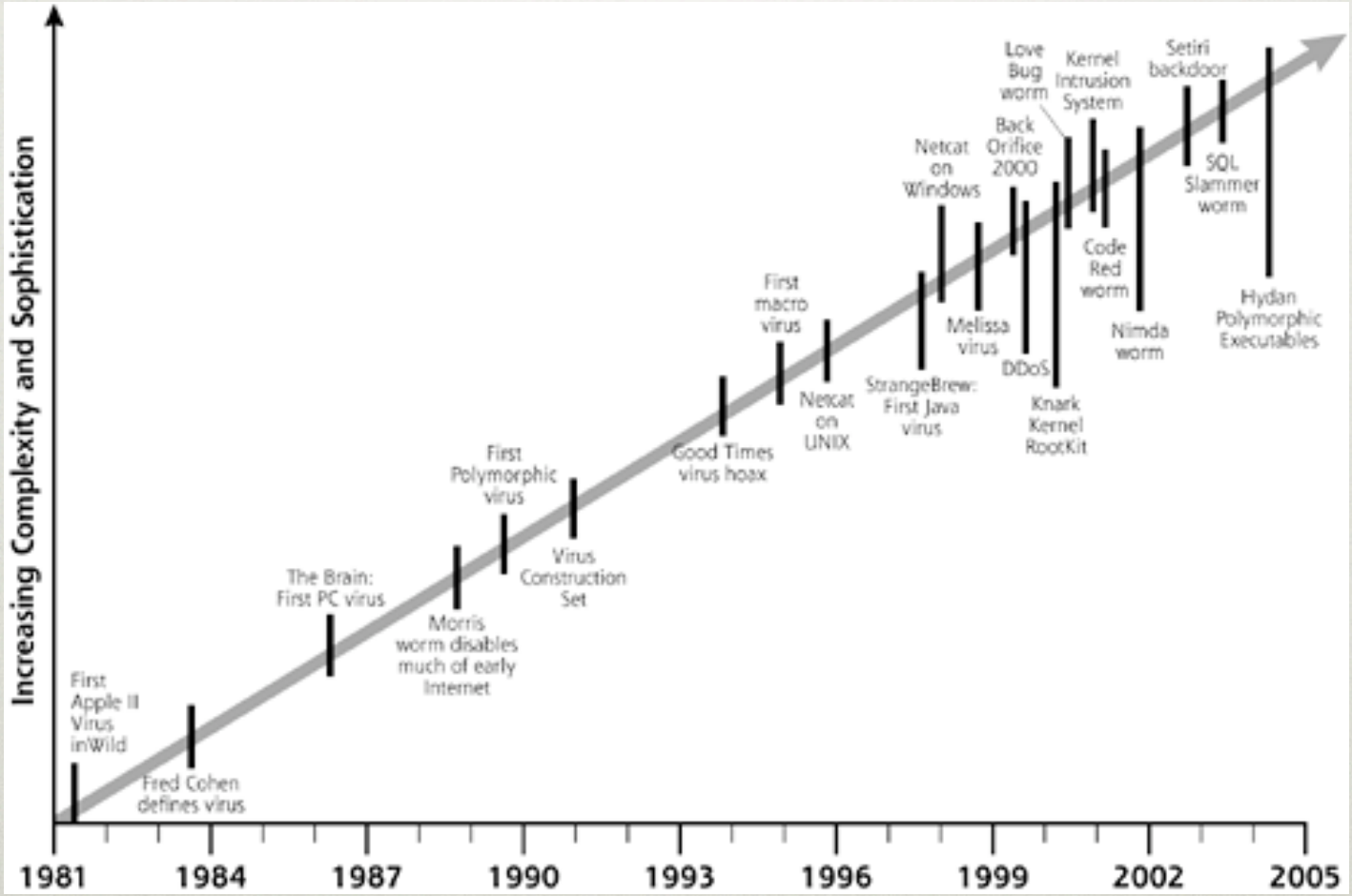
Symantec

What is a worm

A computer worm is a self-replicating computer program. It uses a network to send copies of itself to other nodes and do so without any user intervention.

Image courtesy of: Tech Tips.com

Almost 30 years of Malware



From Malware fighting malicious code

Melissa spread by email and share

Knark rootkit made by creed demonstrate the first ideas

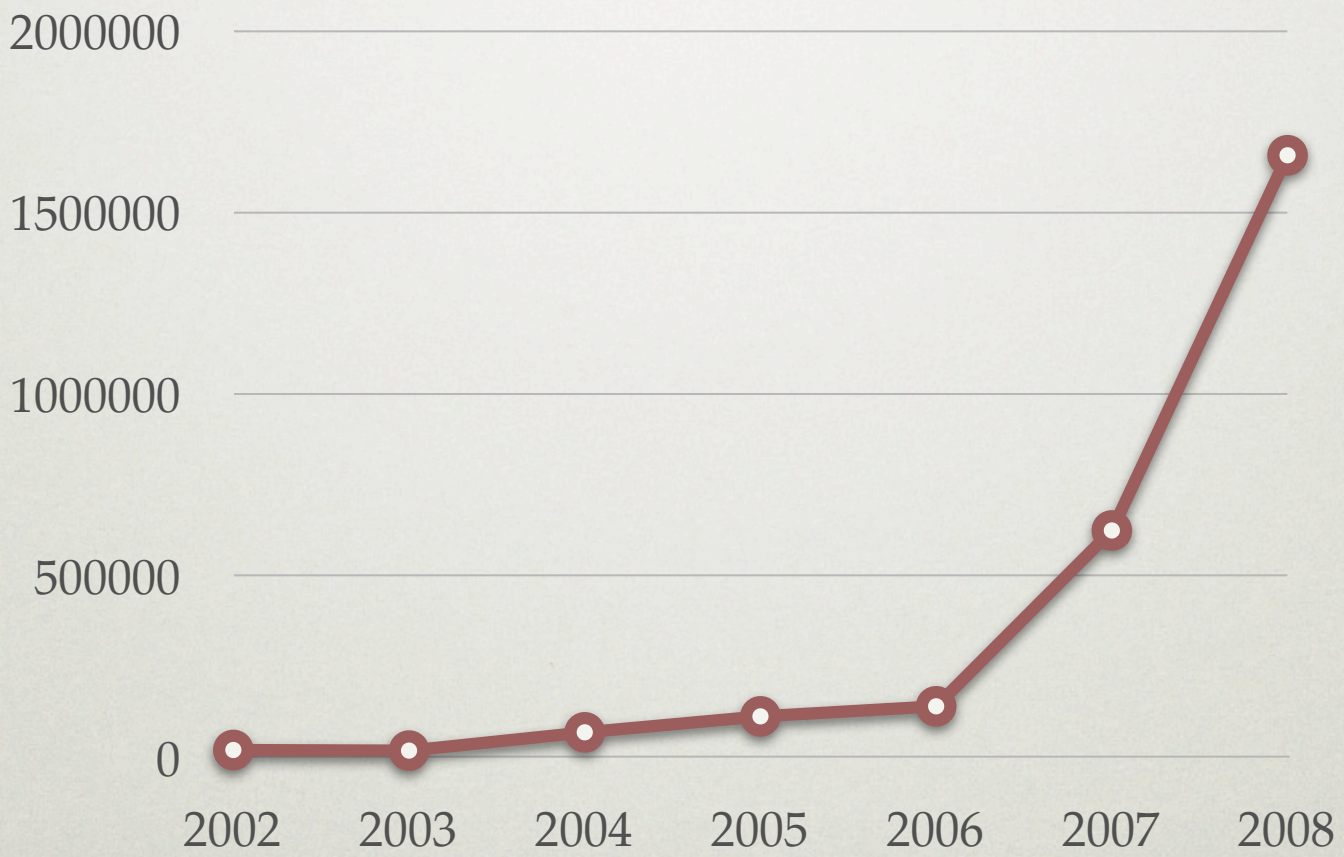
love bug vb script that abused a weakness in outlook

Kernl intrusion by optyx gui and efficent hidding

ISTORY

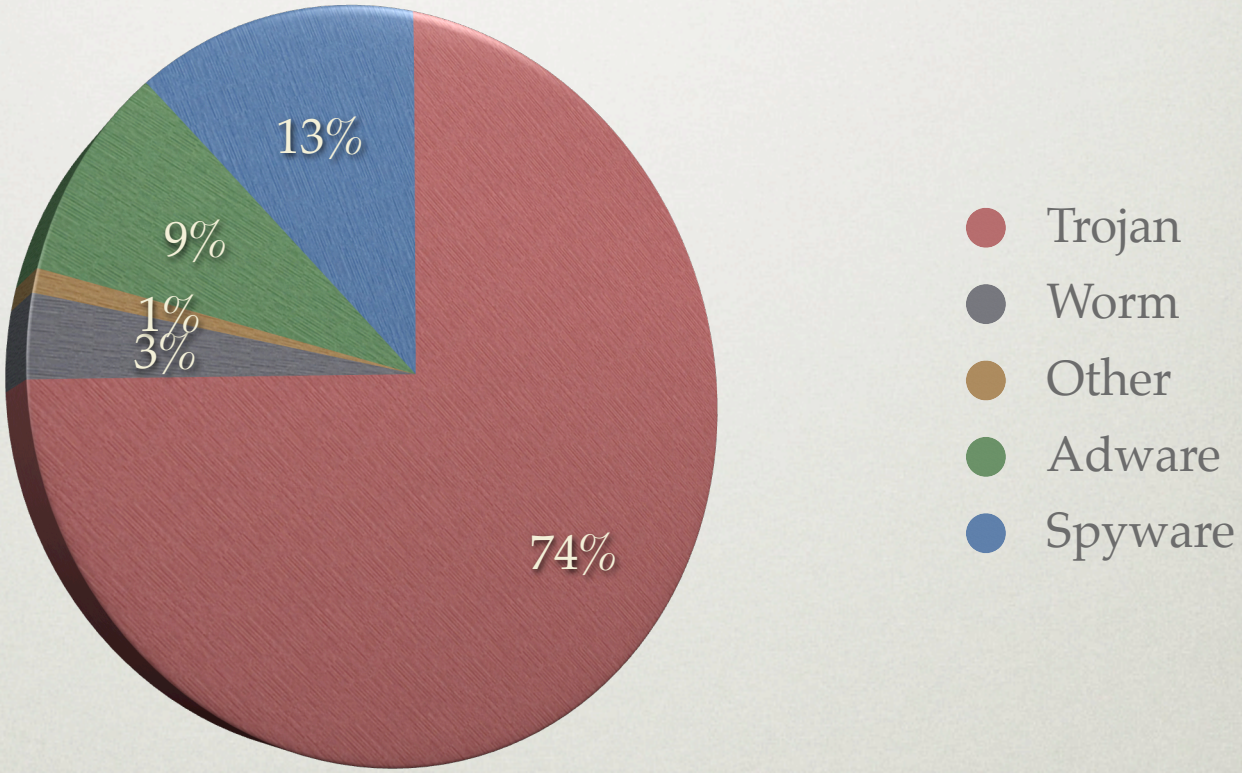
- 1981 First reported virus : Elk Cloner (Apple 2)
- 1983 Virus get defined
- 1986 First PC virus MS DOS
- 1988 First worm : Morris worm
- 1990 First polymorphic virus
- 1998 First Java virus
- 1998 Back orifice
- 1999 Melissa virus
- 1999 Zombie concept
- 1999 Knark rootkit
- 2000 love bug
- 2001 Code Red Worm
- 2001 Kernel Intrusion System
- 2001 Nimda worm
- 2003 SQL Slammer worm

Number of malware signatures



Symantec report 2009

Malware Repartition



Panda Q1 report 2009

Infection methods

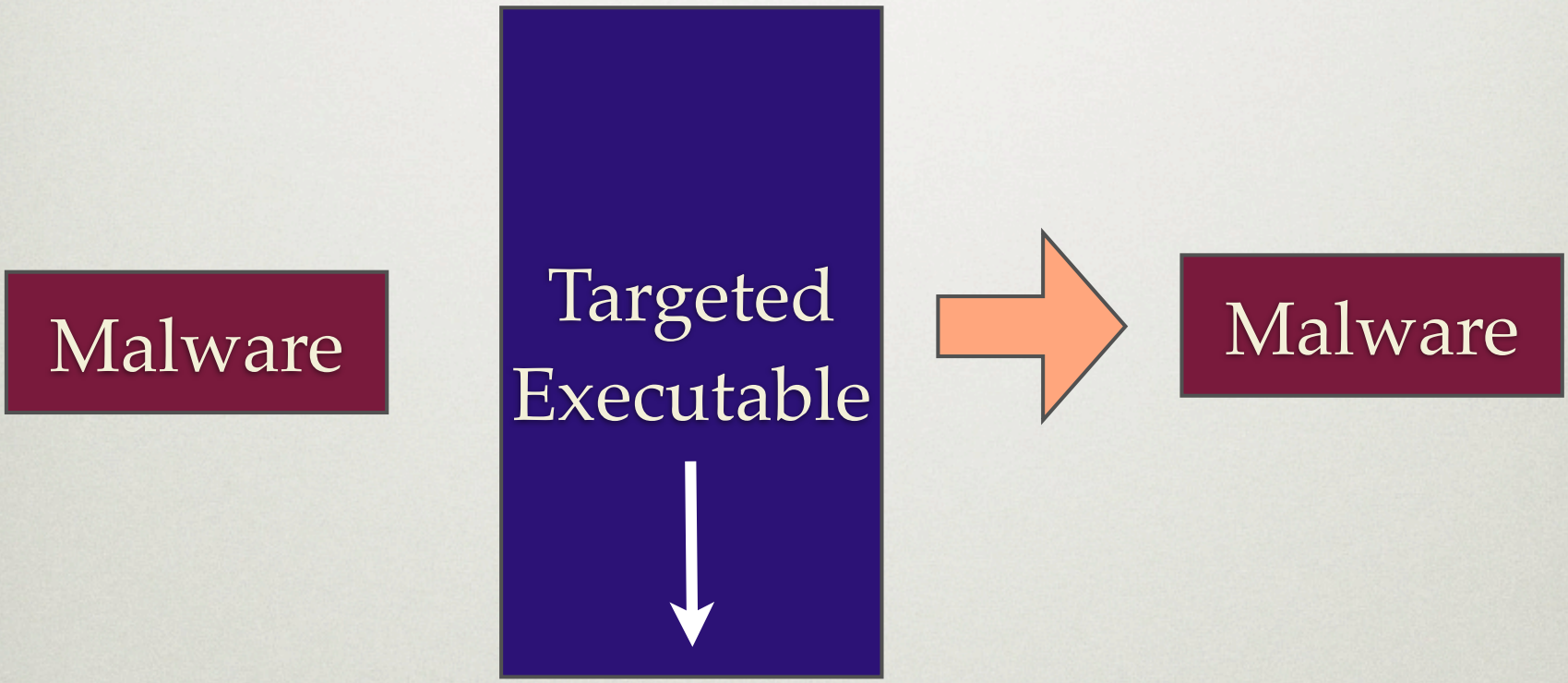
Outline

- What malware are
- How do they infect hosts
- How do they propagate
- Zoo visit !
- How to detect them
- Worms

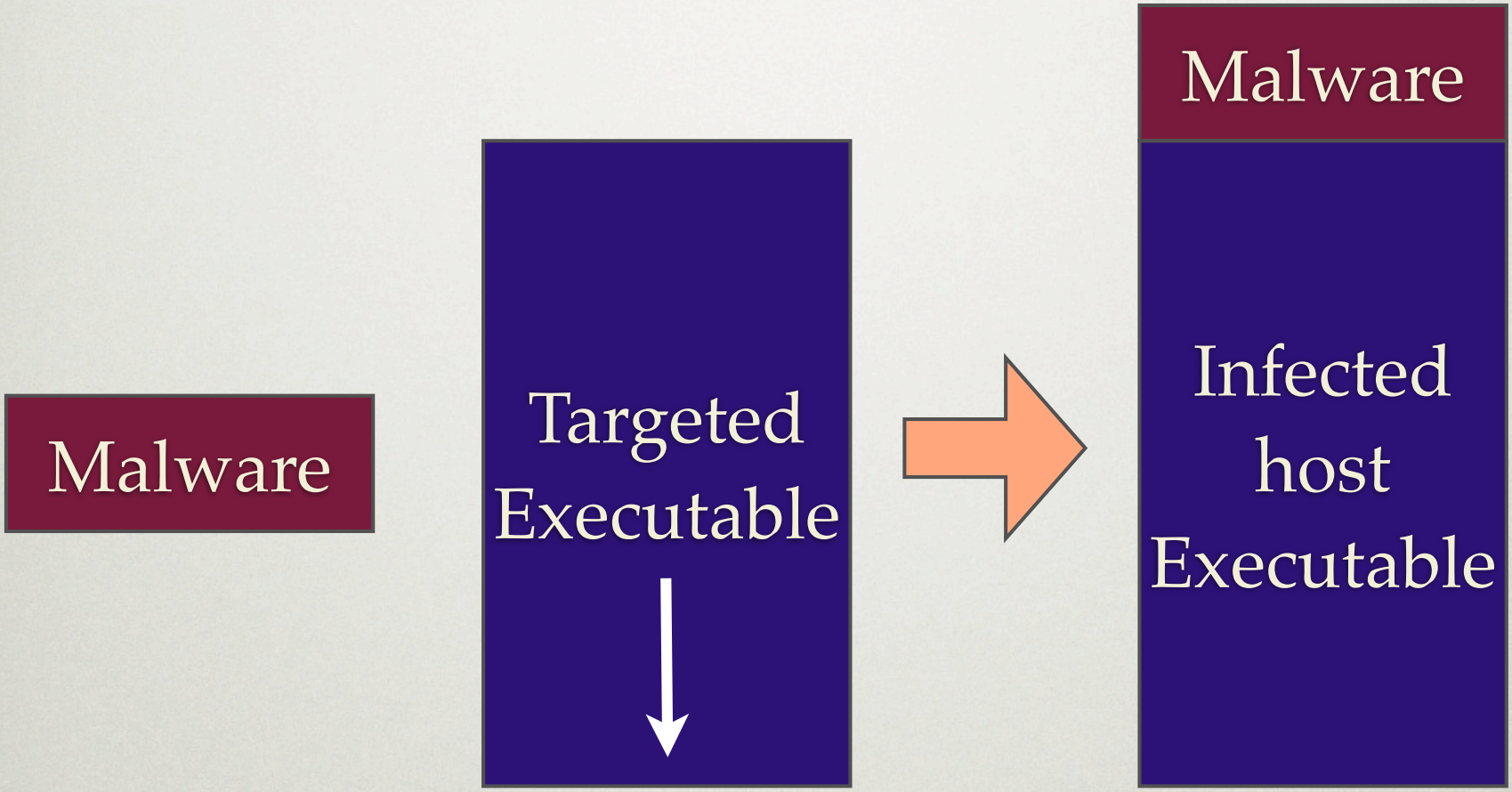
What to Infect

- Executable
- Interpreted file
- Kernel
- Service
- MBR
- Hypervisor

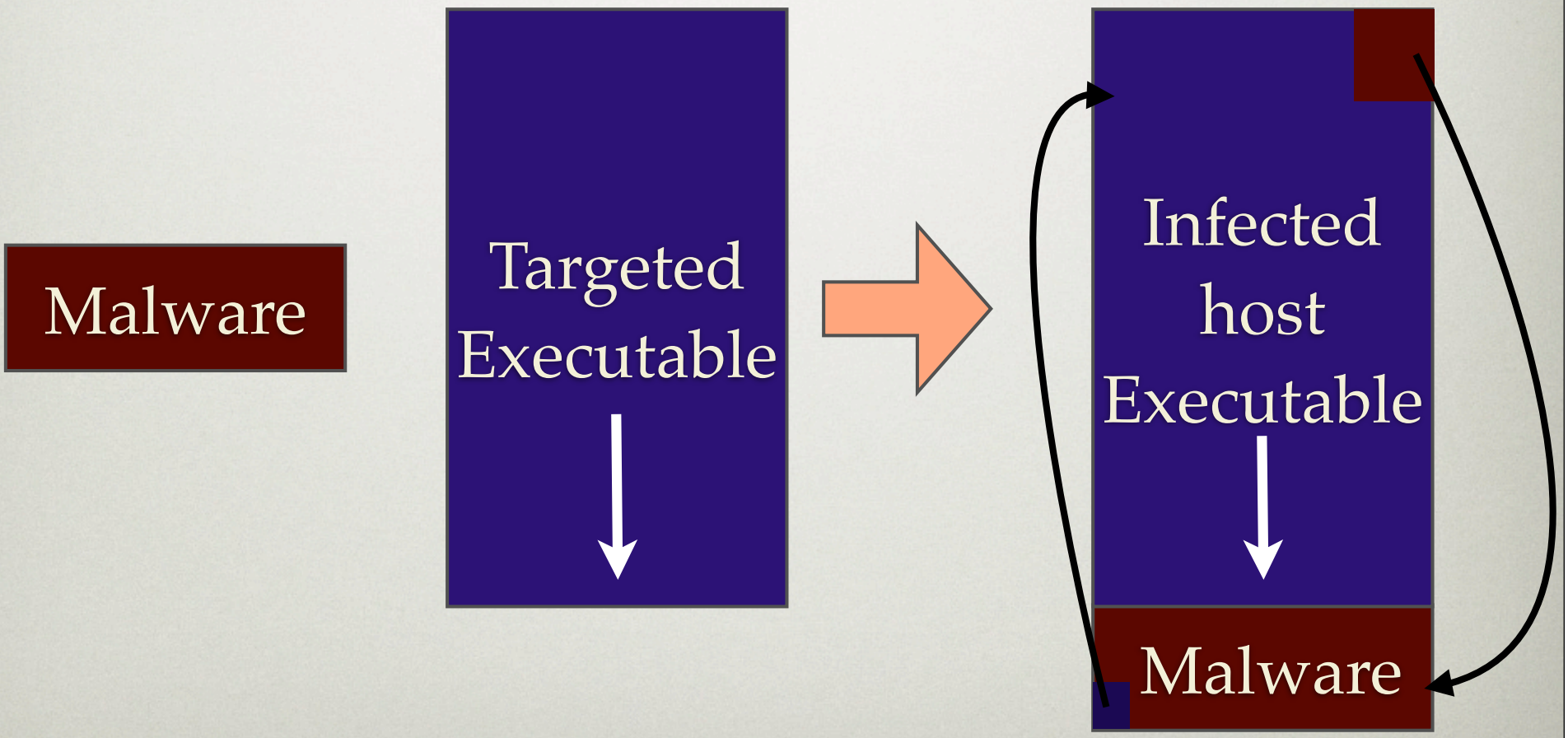
Overwriting malware



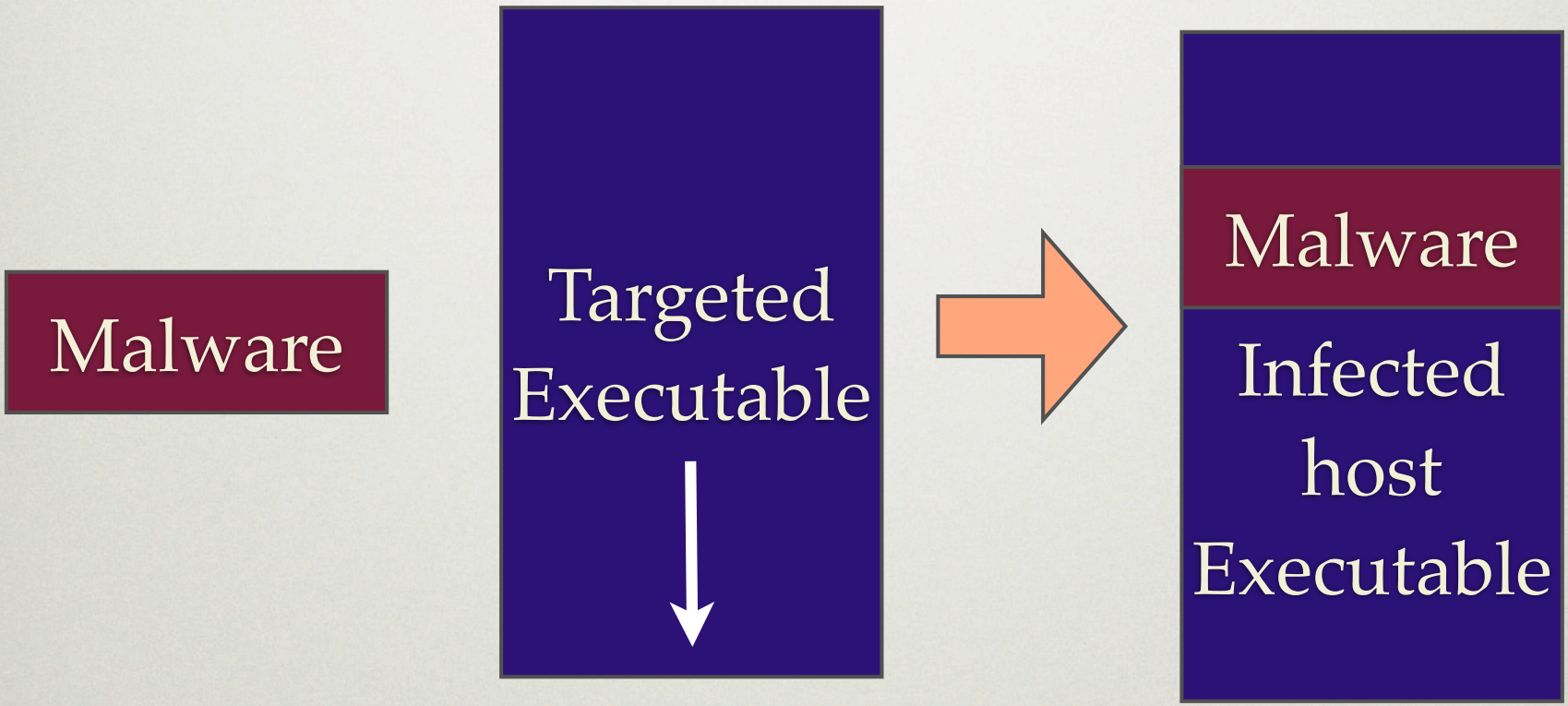
PREPENDING MALWARE



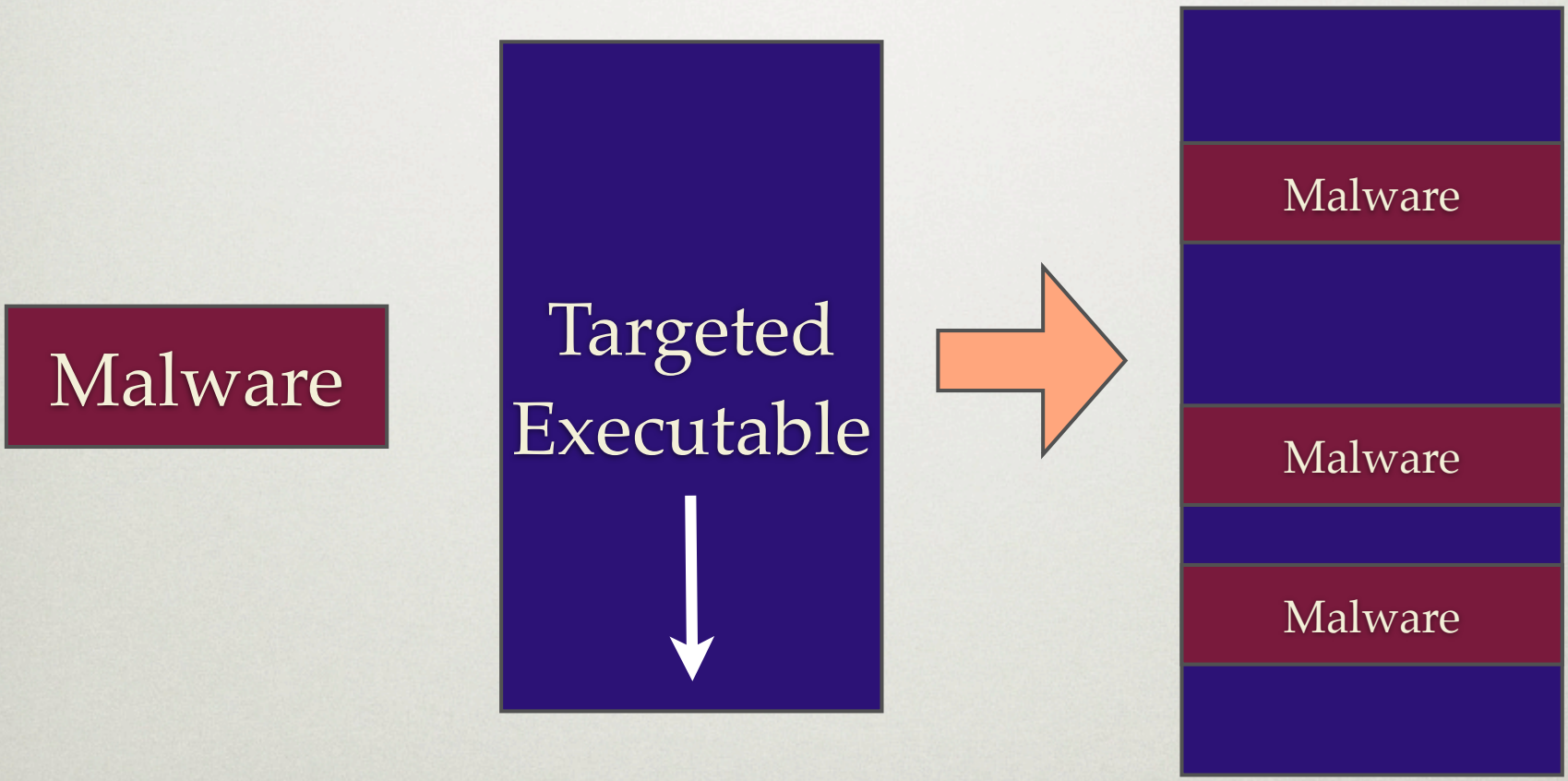
APPENDING MALWARE



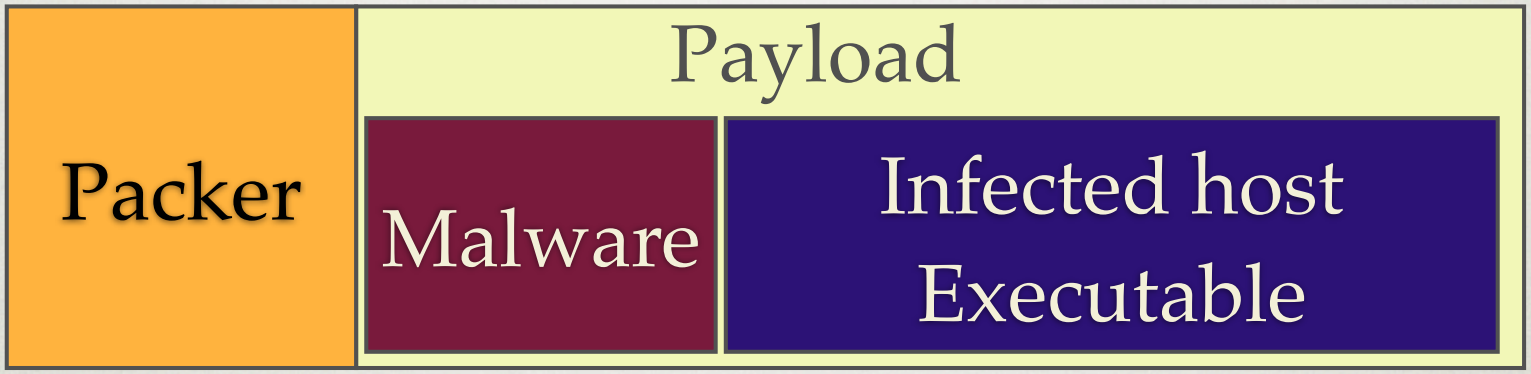
Cavity malware



Multi-Cavity malware



Packers



Packer functionalities

- Compress
- Encrypt
- Randomize (polymorphism)
- Anti-debug technique (int / fake jmp)
- Add-junk
- Anti-VM
- Virtualization

Auto start

- Folder auto-start : C:\Documents and Settings\[user_name]\Start Menu \Programs\Startup
- Win.ini : run=[backdoor]" or "load=[backdoor]".
- System.ini : shell=”myexplorer.exe”
- Wininit
- Config.sys

Auto start cont.

- Assign know extension (.doc) to the malware
- Add a Registry key such as *HKCU\SOFTWARE\Microsoft\Windows \CurrentVersion\Run*
- Add a task in the task scheduler
- Run as service

Unix autostart

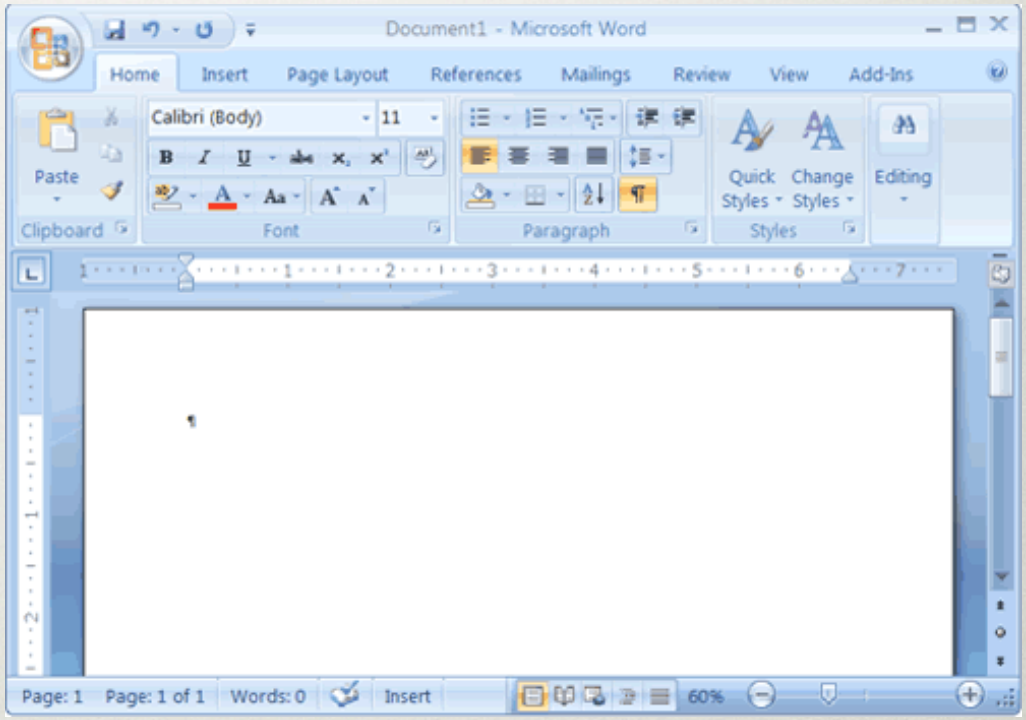
- Init.d
- /etc/rc.local
- .login .xsession
- crontab
  - crontab -e
  - /etc/crontab

Macro virus

- Use the builtin script engine
- Example of call back used (word)
  - AutoExec()
  - AutoClose()
  - AutoOpen()
  - AutoNew()

Document based malware

- MS Office
- Open Office
- Acrobat



Userland root kit

- Perform
  - login
  - sshd
  - passwd
- Hide activity
  - ps
  - netstat
  - ls
  - find
  - du

Subverting the Kernel

Kernel task

- Process management
- File access
- Memory management
- Network management

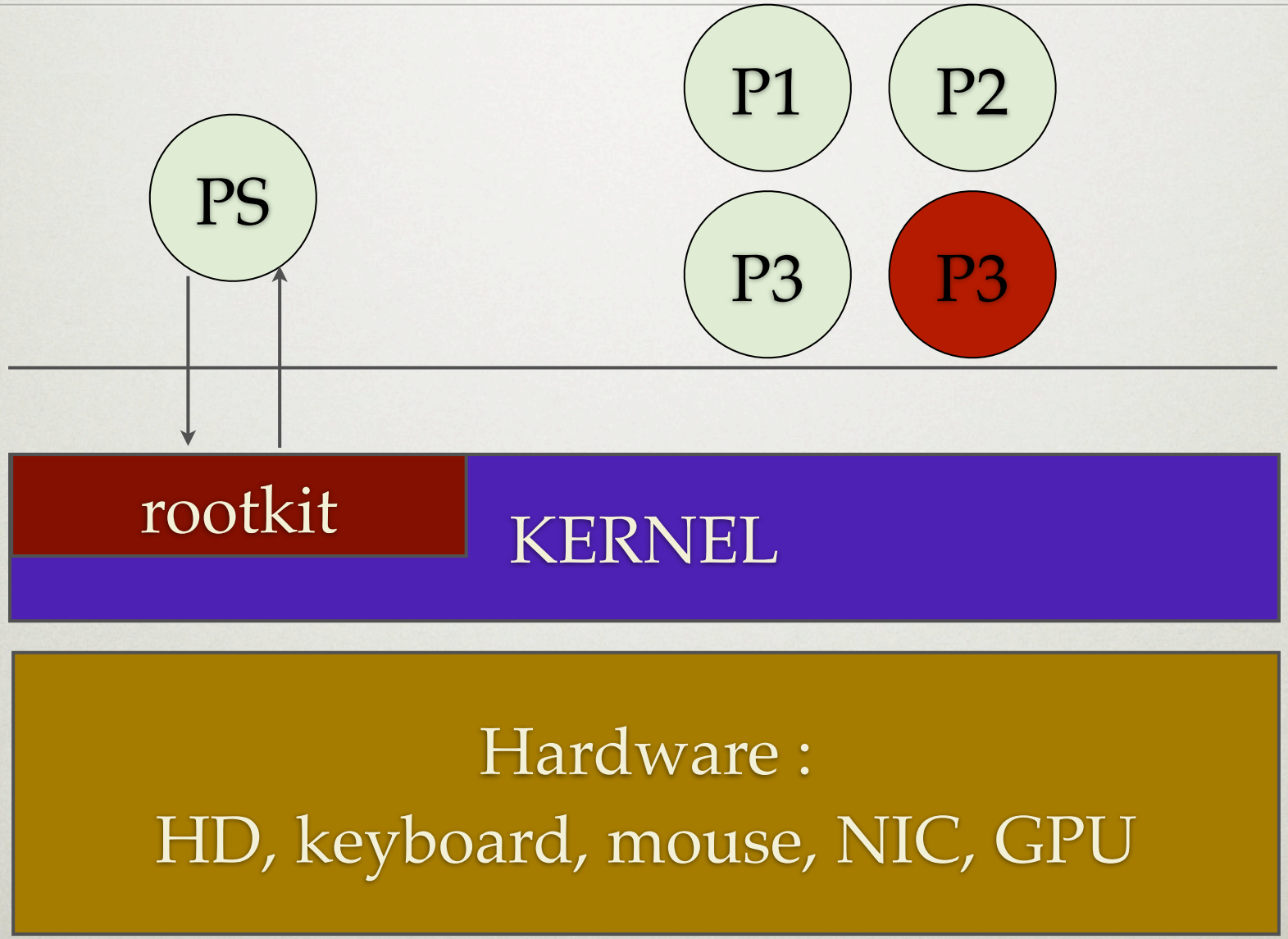
What to hide

➡ Process

➡ Files

➡ Network traffic

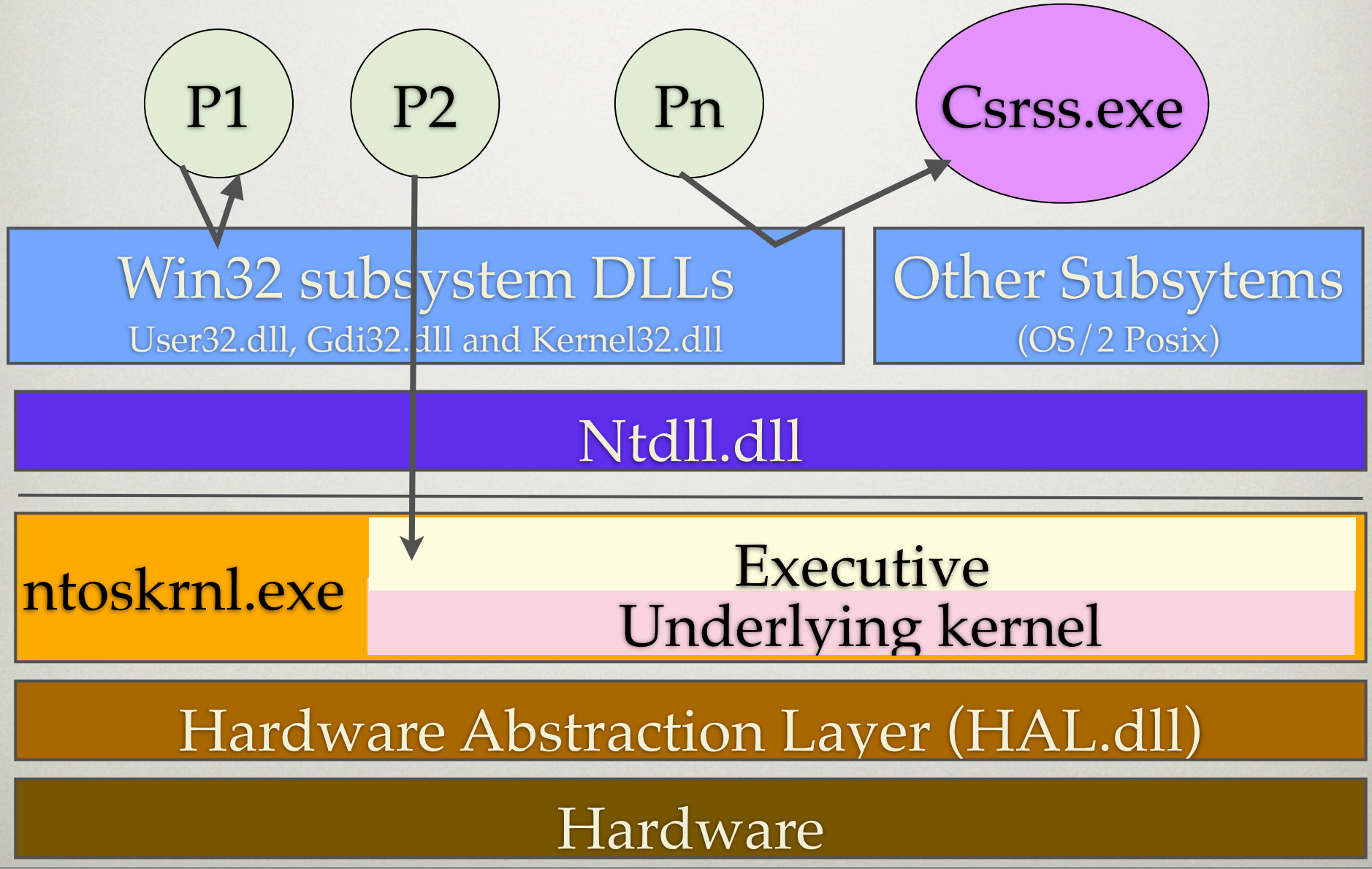
Kernel rootkit



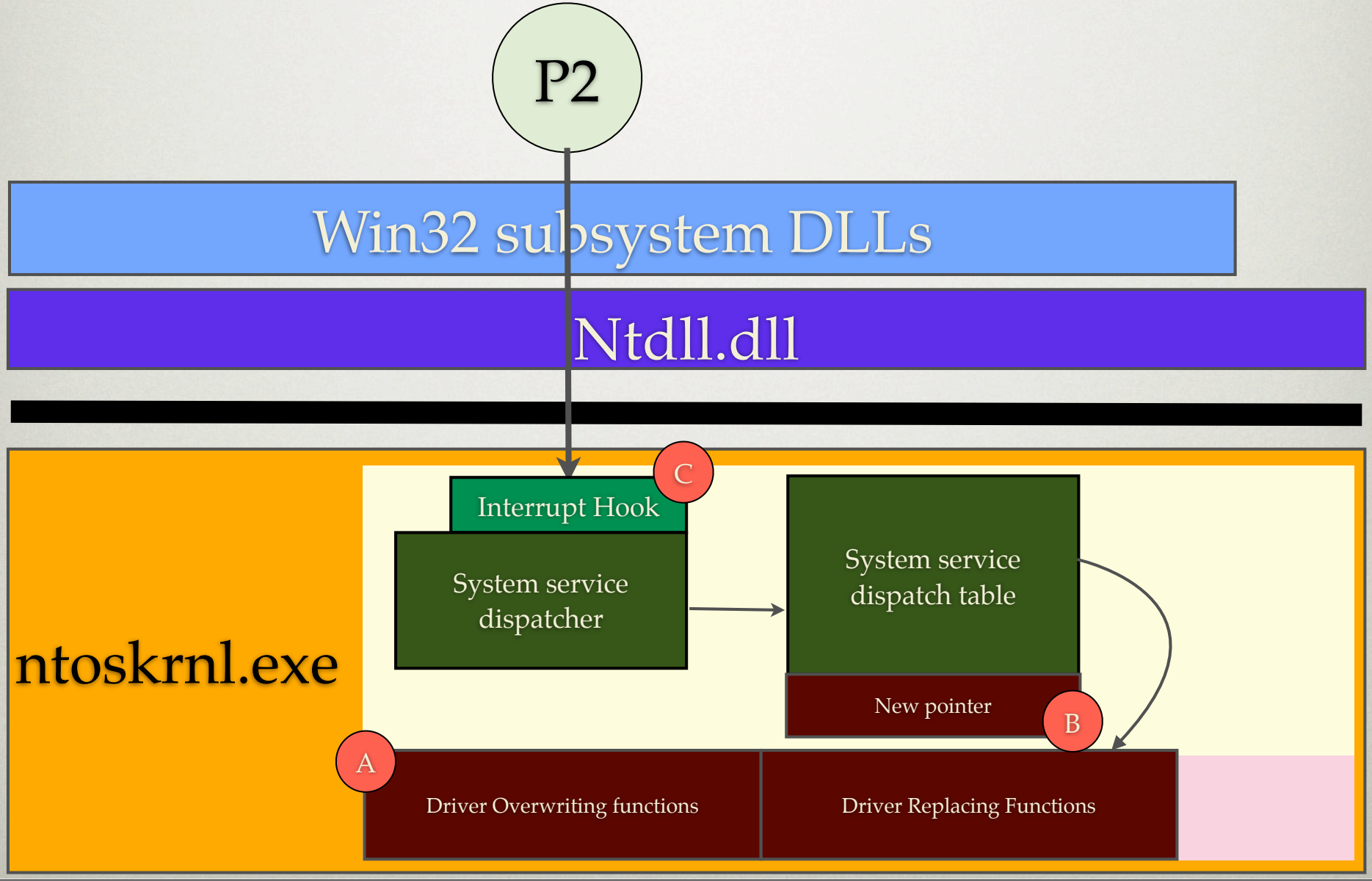
Subverting techniques

- Kernel patch
- Loadable Kernel Module
- Kernel memory patching (/dev/kmem)

Windows Kernel

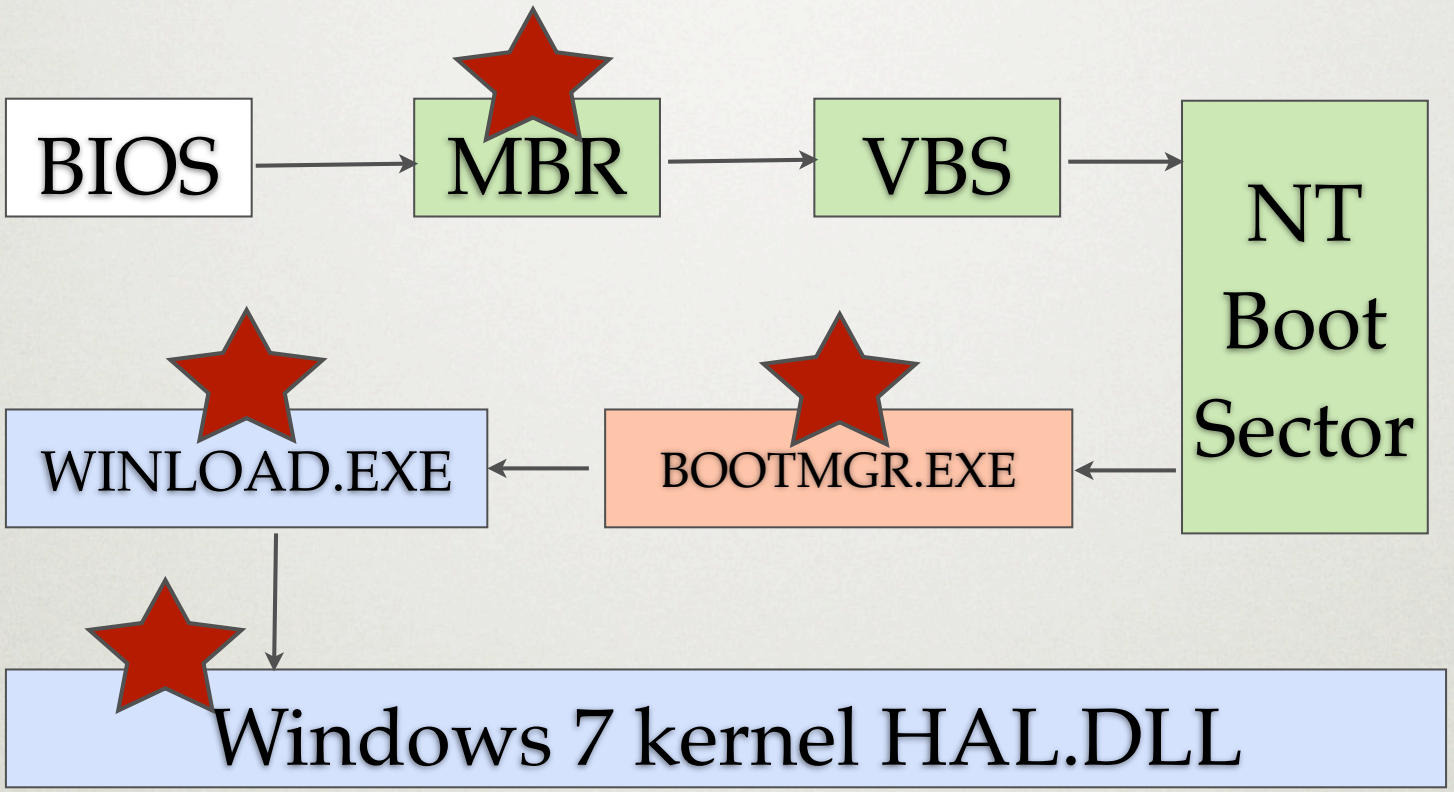


Kernel Device driver



MBR/Bootkit

Bootkits can be used to avoid all protections of an OS, because OS consider that the system was in trusted stated at the moment the OS boot loader took control.

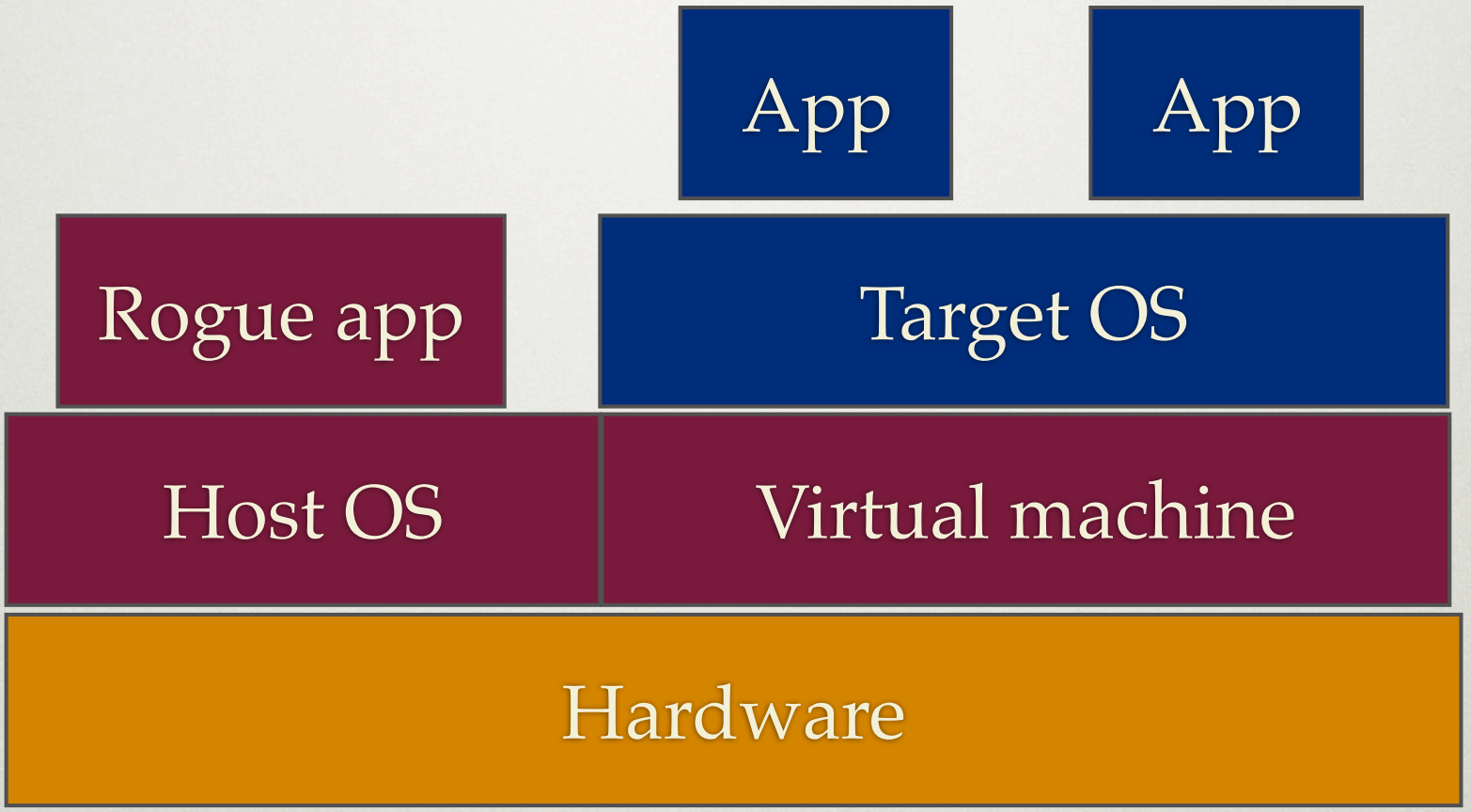

BIOS
MBR
VBS
NT Boot Sector
WINLOAD.EXE
BOOTMGR.EXE
Windows 7 kernel HAL.DLL

Vboot

- Work on every Windows (vista,7)
- 3ko
- Bypass checks by letting them run and then do inflight patching
- Communicate via ping

Hypervisor rootkit

Hypervisor rootkit

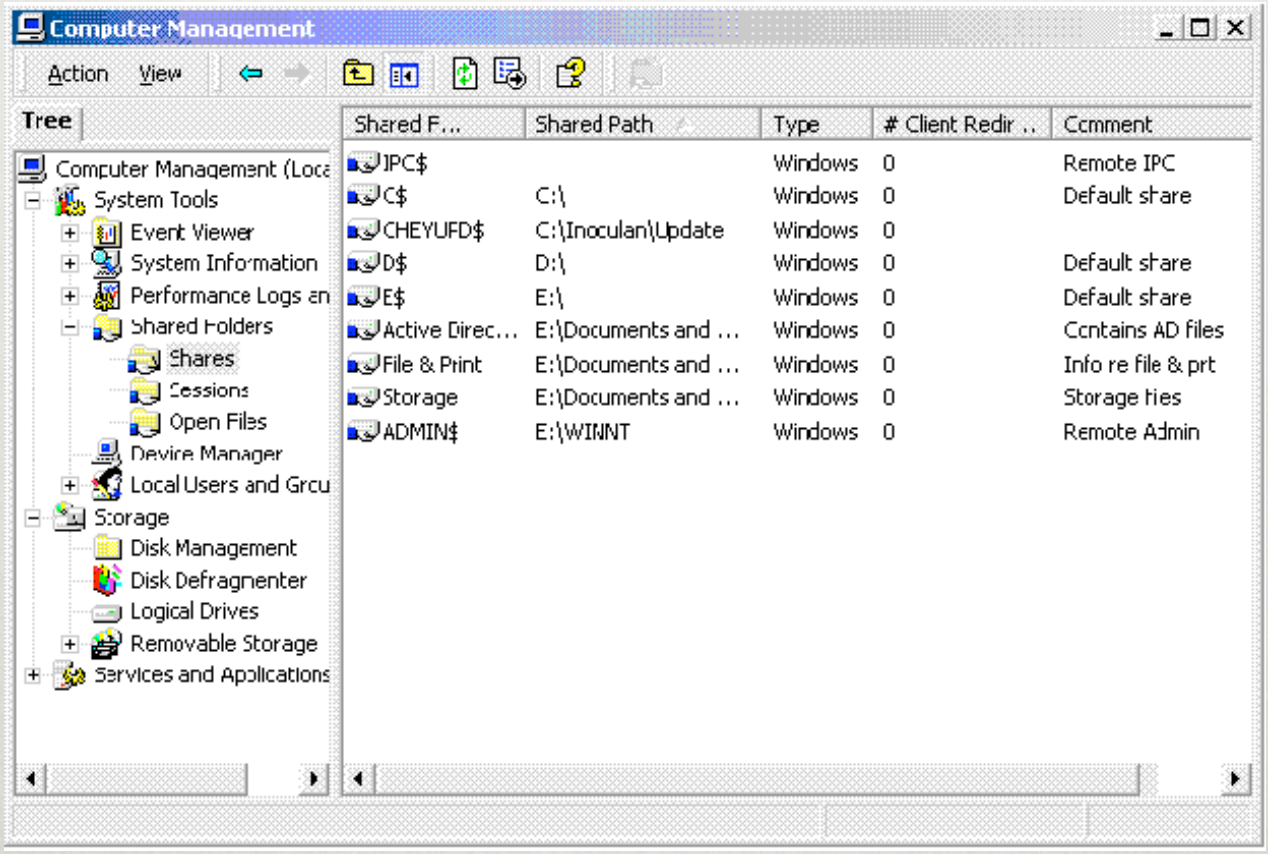


Propagation Vector

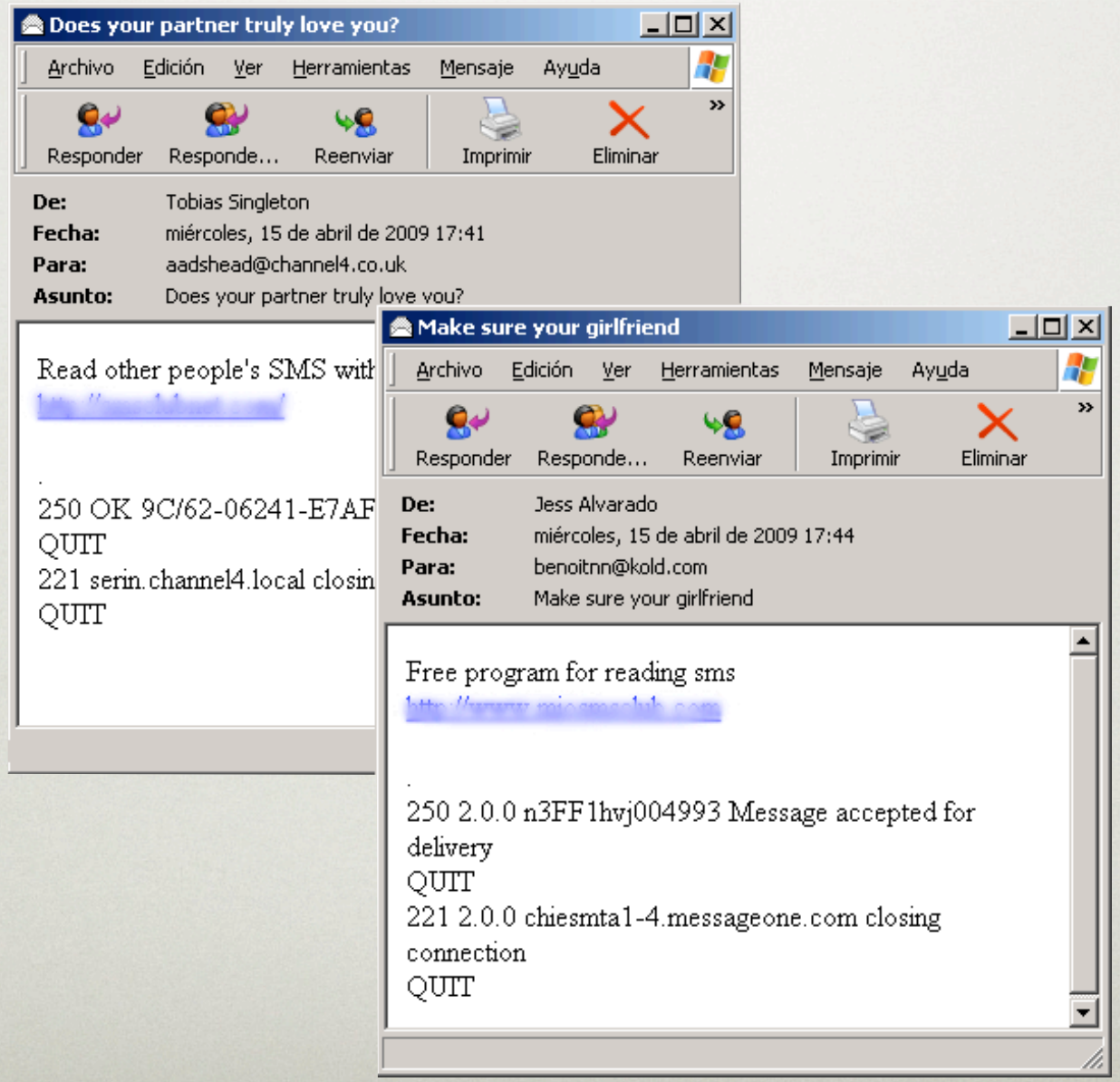
Outline

- What malware are
- How do they infect hosts
- How do they propagate
- Zoo visit !
- How to detect them
- Worms

Shared folder

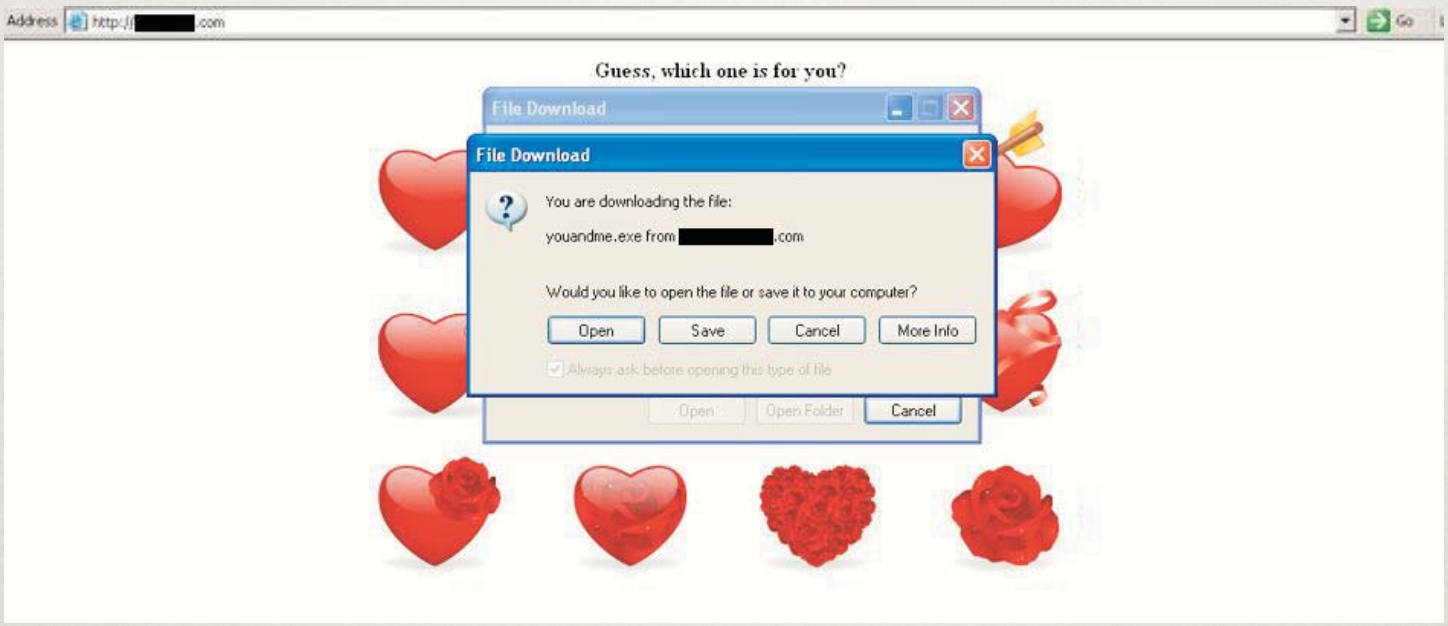


Email propagation



from pandalab blog

Valentine day ...



Waledac malicious domain from pandalab blog

Email again

Are you interested in reading other people's sms?

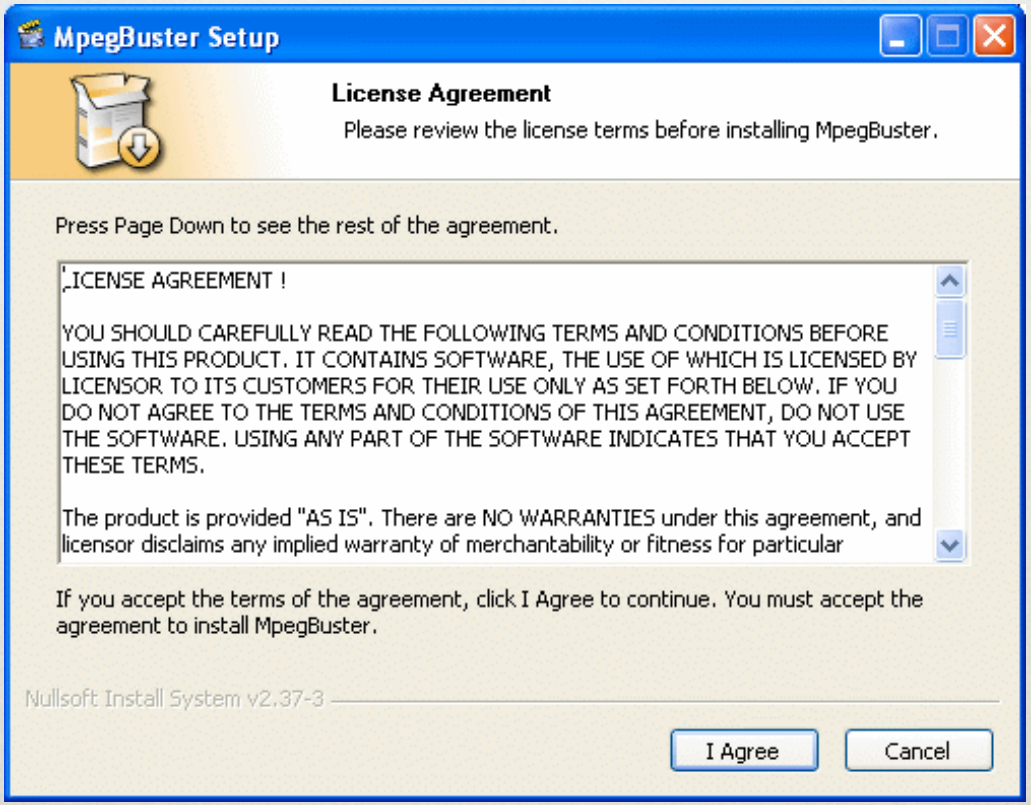
Get Your Free 30-Day Trial!

Do you want to test your partner or just to read somebody's SMS? This program is exactly what you need then! It's so easy! You don't need to install it at the mobile phone of your partner.
Just download the program and you will able to read all SMS when you are online. Be aware of everything! This is an extremely new service!

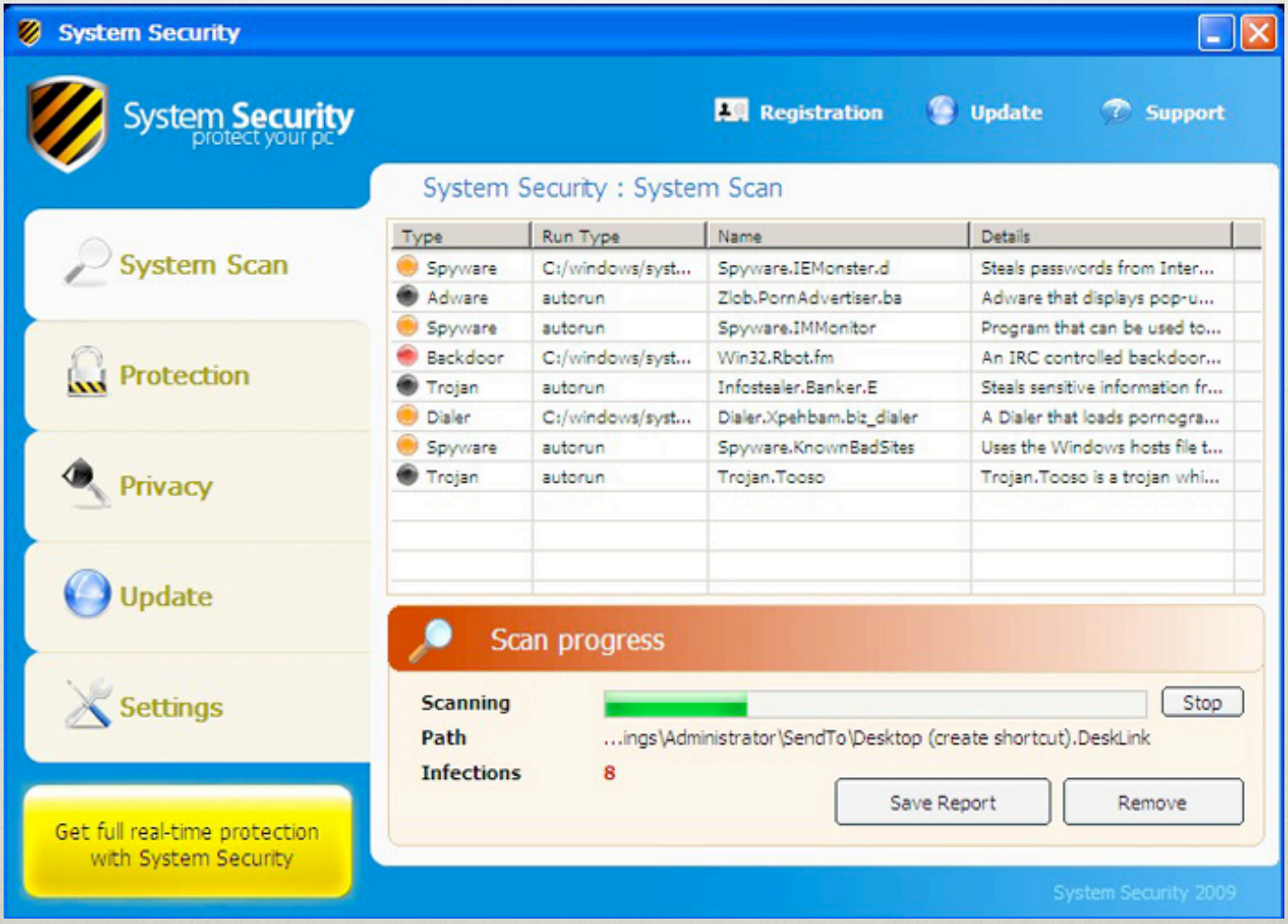
http://[Removed].com/freetrial.exe

Download Free Trial

Fake codec


MpegBuster Setup
License Agreement
Please review the license terms before installing MpegBuster.
Press Page Down to see the rest of the agreement.
LICENSE AGREEMENT !
YOU SHOULD CAREFULLY READ THE FOLLOWING TERMS AND CONDITIONS BEFORE USING THIS PRODUCT. IT CONTAINS SOFTWARE, THE USE OF WHICH IS LICENSED BY LICENSOR TO ITS CUSTOMERS FOR THEIR USE ONLY AS SET FORTH BELOW. IF YOU DO NOT AGREE TO THE TERMS AND CONDITIONS OF THIS AGREEMENT, DO NOT USE THE SOFTWARE. USING ANY PART OF THE SOFTWARE INDICATES THAT YOU ACCEPT THESE TERMS.
The product is provided "AS IS". There are NO WARRANTIES under this agreement, and licensor disclaims any implied warranty of merchantability or fitness for particular
If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install MpegBuster.
Nullsoft Install System v2.37-3
I Agree
Cancel

Fake antivirus

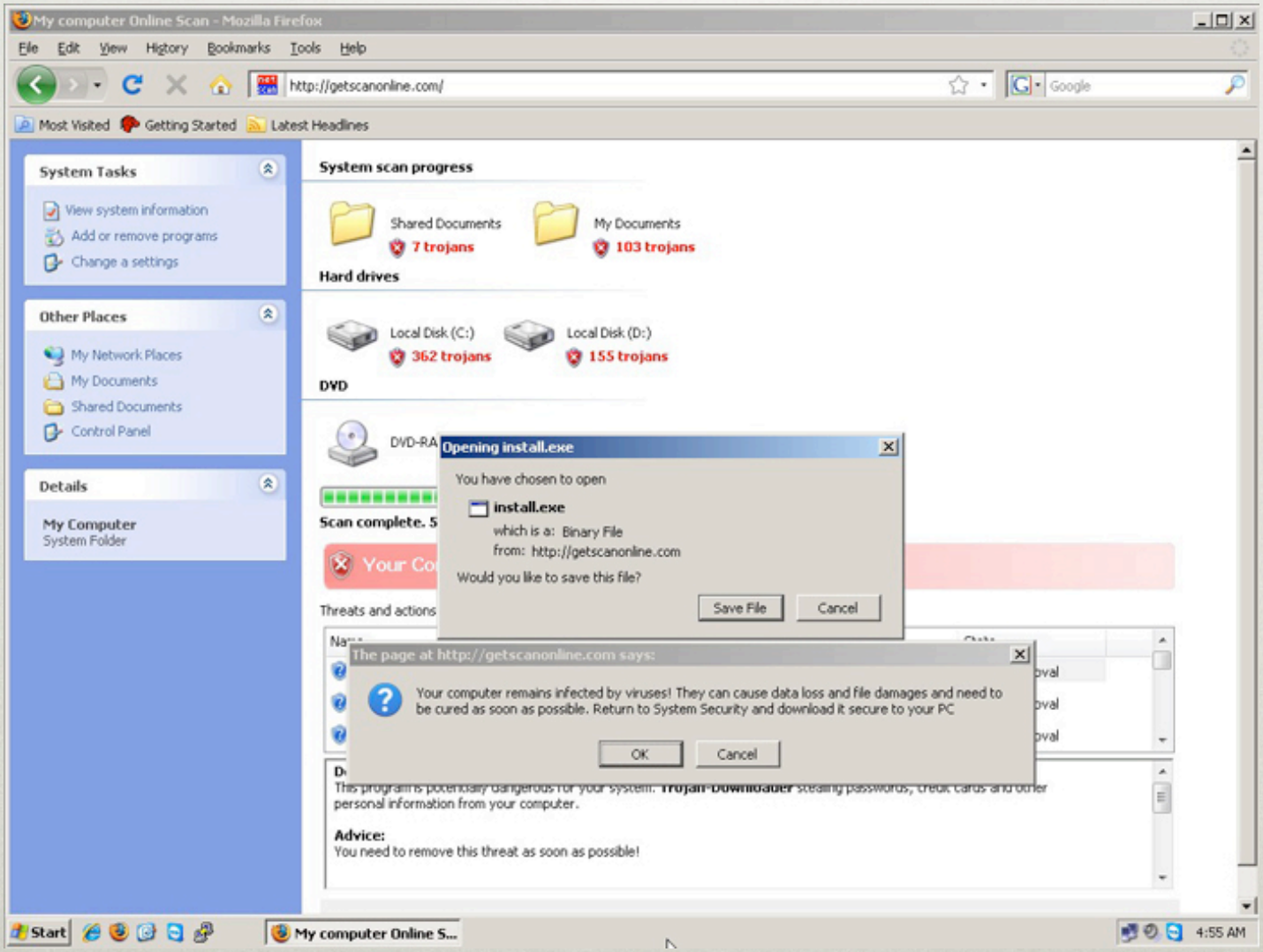


from pandalab blog

Hijack you browser

from pandalab blog

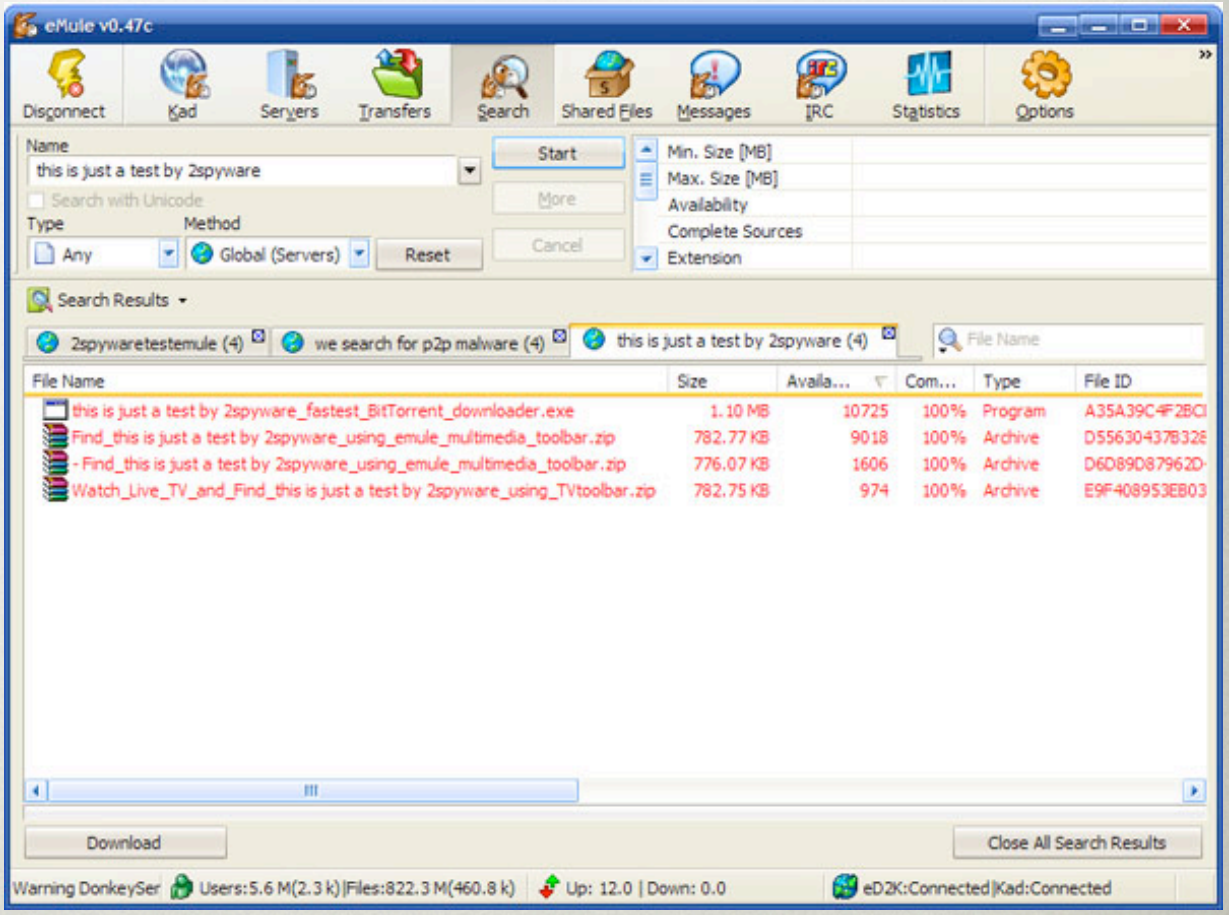
Fake page !



from pandalab blog

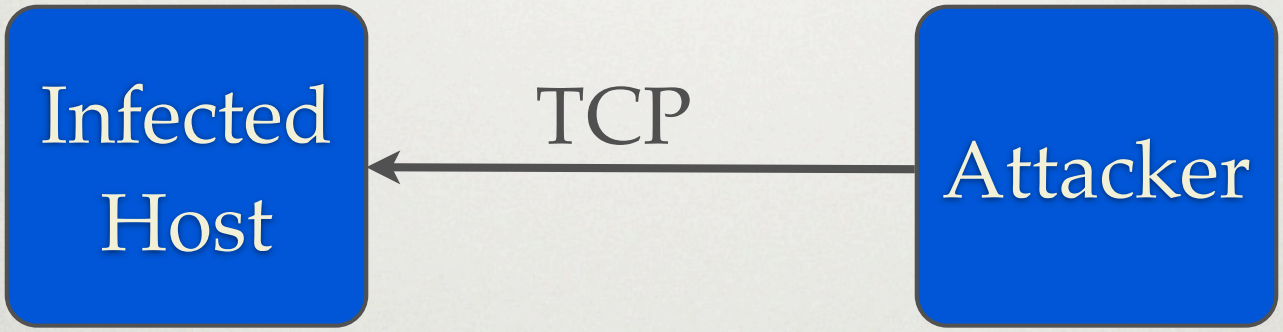
P2P Files

- Popular query
- 35.5% are malwares
  (Kalafut 2006)

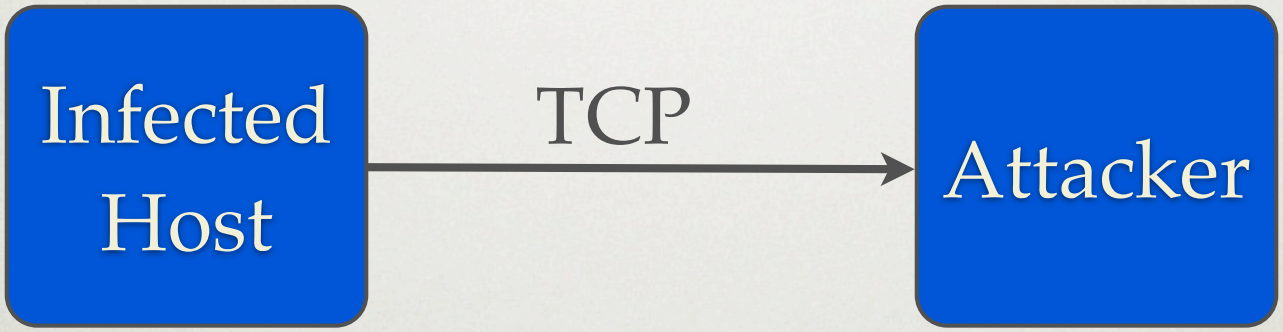


Backdoor

Basic

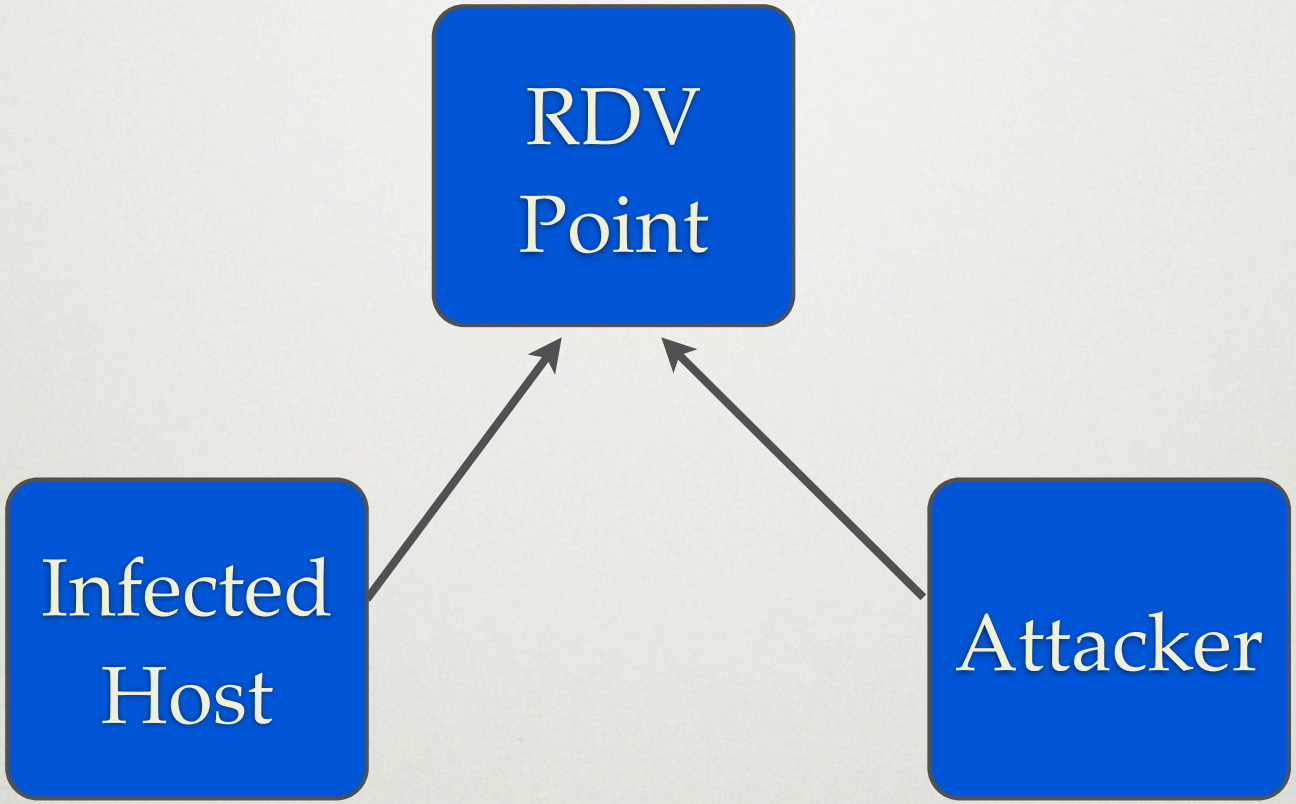


Reverse



COVERT

Rendez vous backdoor



Bestiary

Outline

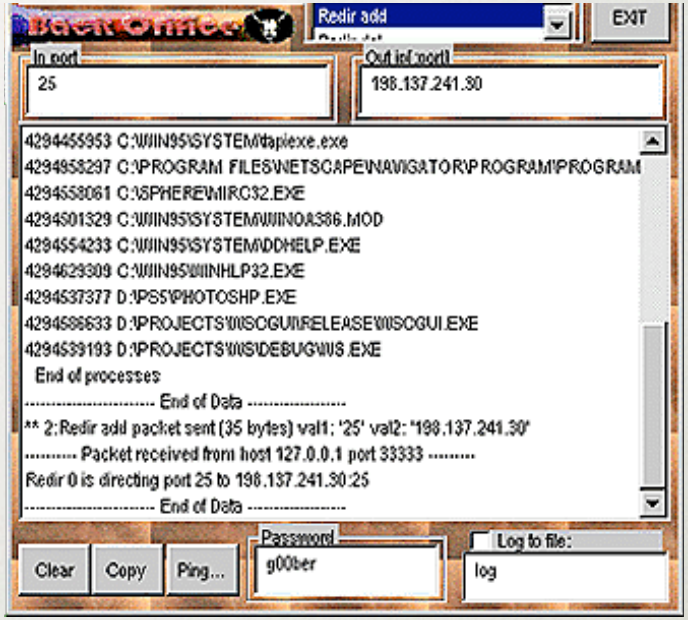
- What malware are
- How do they infect hosts
- How do they propagate
- Zoo visit !
- How to detect them
- Worms

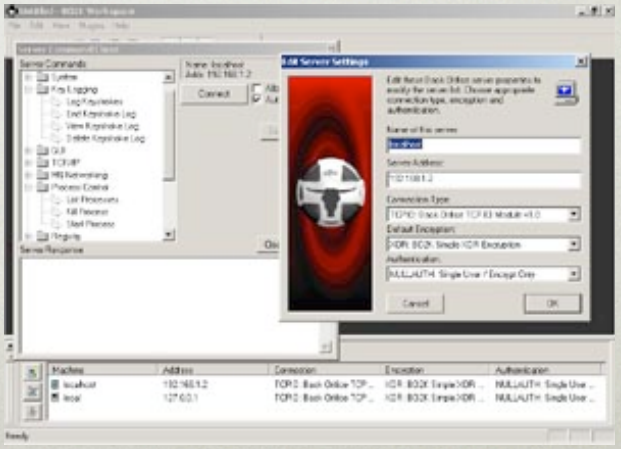
Adware


Free connection
Free modem
Click here to find out more
BT
INSWEB
Lower your insurance costs.
Enter to WIN!
XEROX
WIN an $11,000 USD
Business Upgrade Package
or a Xerox Phaser® 8400
Color Printer
World's fastest color printer
for under $1,000 USD.
Phaser 8400
Color Printer
$999
Quotes
of coffee,
TRY 2 RISK-FREE ISSUES!
15 FREE
POWER
GUIDES
Netscape What's New
Science Proves It: Drinking Causes...
Teen Girls Beware: Popular Boys Do THIS
Look What Was Found in Stone Age Cave
Strange Space Object Mystery Deepens
Did You See What Bush Dared to Wear?
If You Get This E-Mail Scam, Delete It
Before you go, did you know....
You can add AOL® for Broadband to any high-speed cable or DSL connection!
Works with and enhances any basic high-speed connection
Built-in protection for you and your family
The best on-demand and exclusive programming online
GET A FREE TRIAL!
Upto 45 days FREE
Click here for details
Home-Owners Click Here
AMERIQUEST MORTGAGE COMPANY®
Are You Paying Too
Rates are at Hist
Get cas
your h
HomeLoanCenter.com
TRAVELZOO
This Week's
on the Interne
Released APRIL 14,
Song
CheapCaribbean.c
ReliaQuot
A better way to buy life insura
Take this free test!

BackOrifice

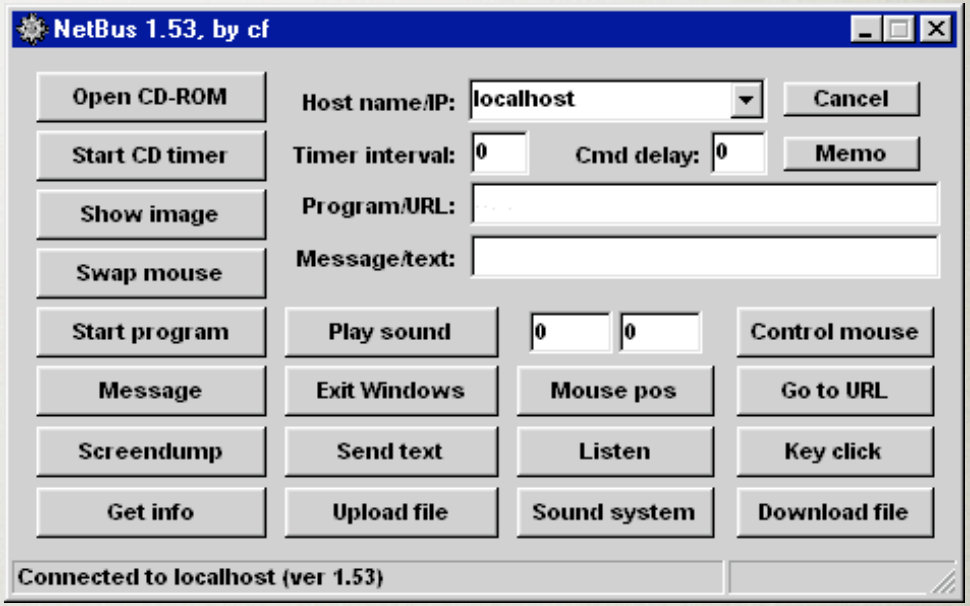
- Defcon 1998
- new version in 2000



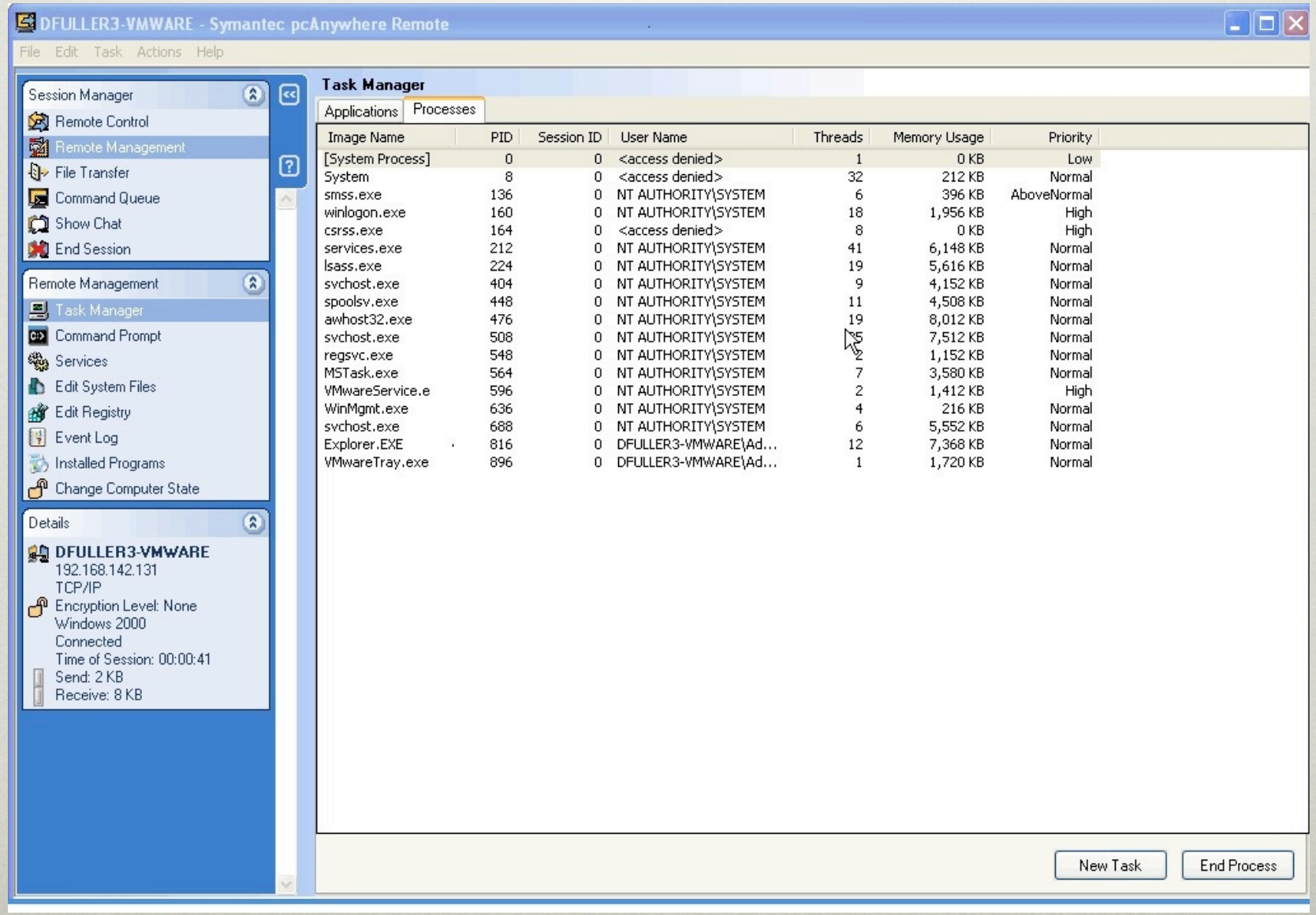


Netbus

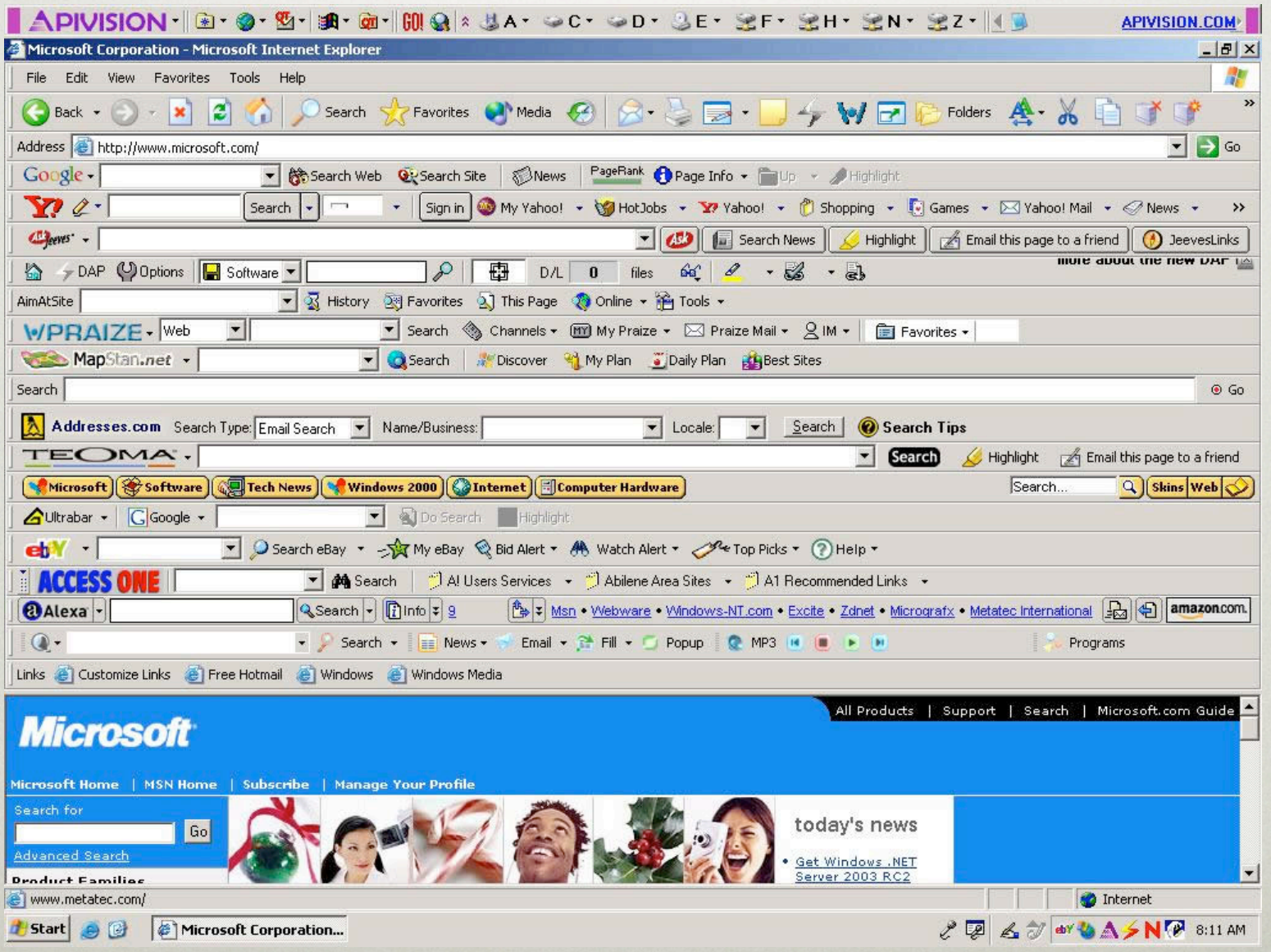
- 1998
- Used for “prank”



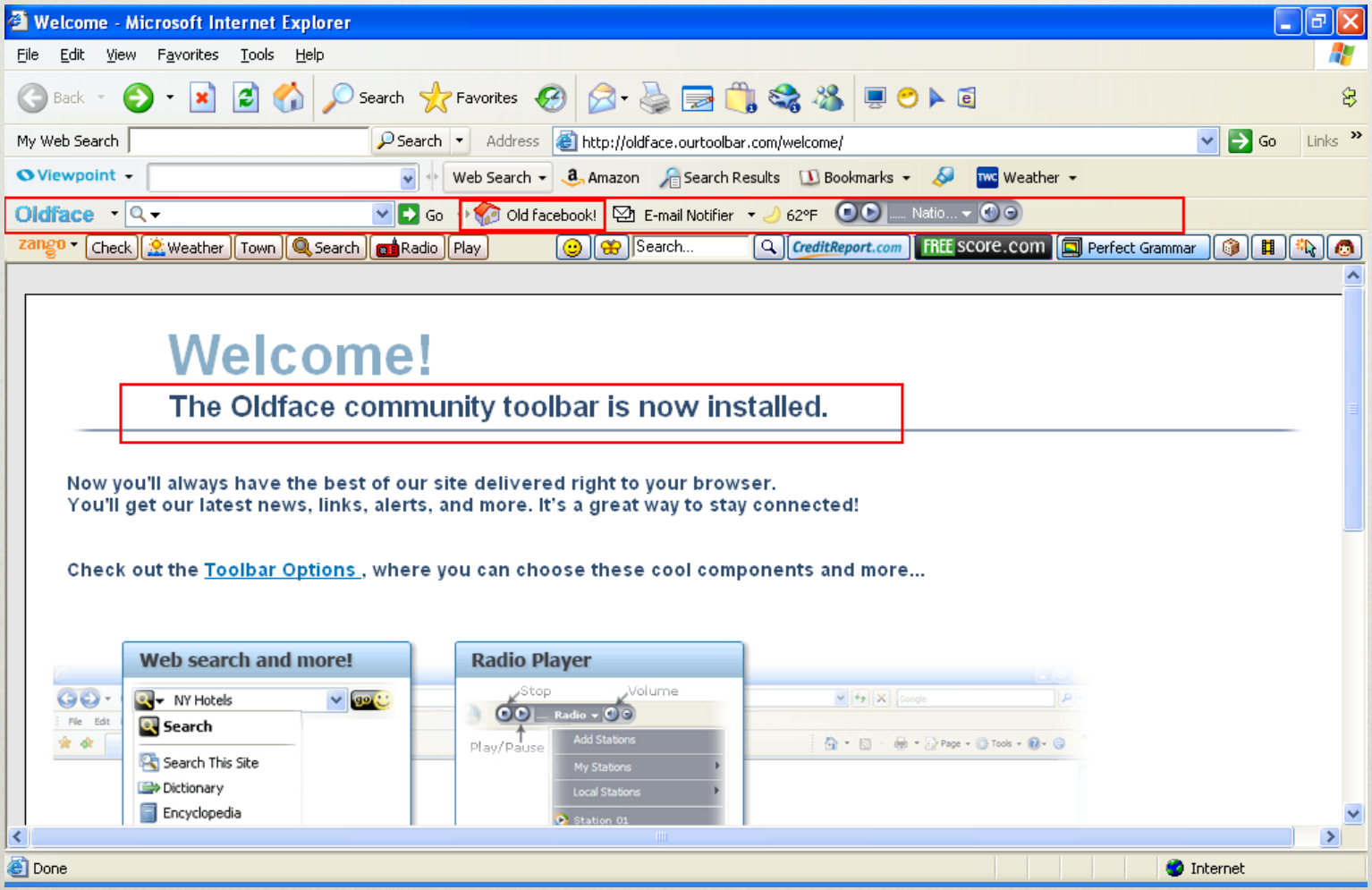
Symantec pcAnywhere



Browser Toolbar ...

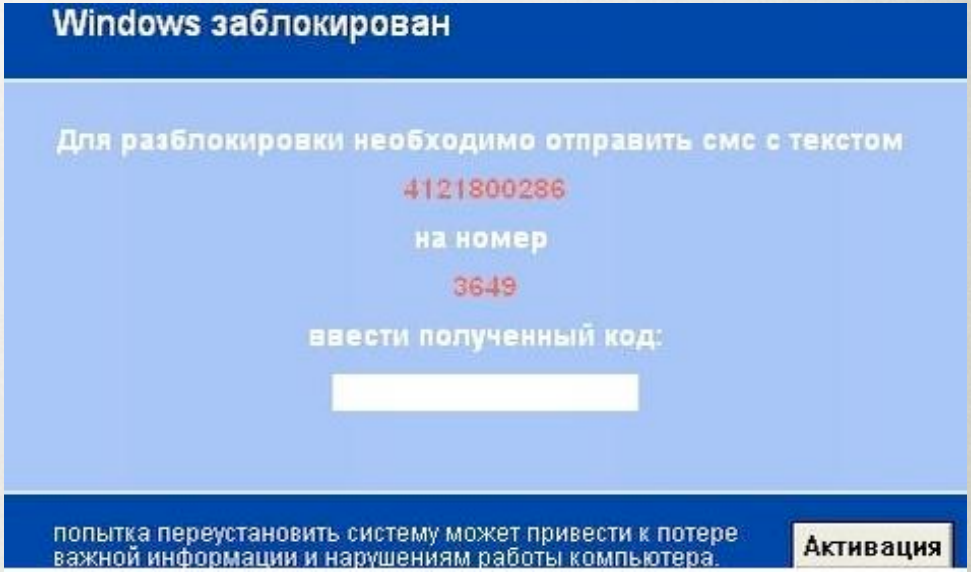


Toolbar again



Ransomware

- Trj/SMSlock.A
- Russian ransomware
- April 2009



To unlock you need to send an SMS with the text
4121800286
to the number
3649
Enter the resulting code:

Any attempt to reinstall the system may lead to loss of important information and computer damage

from pandalab blog

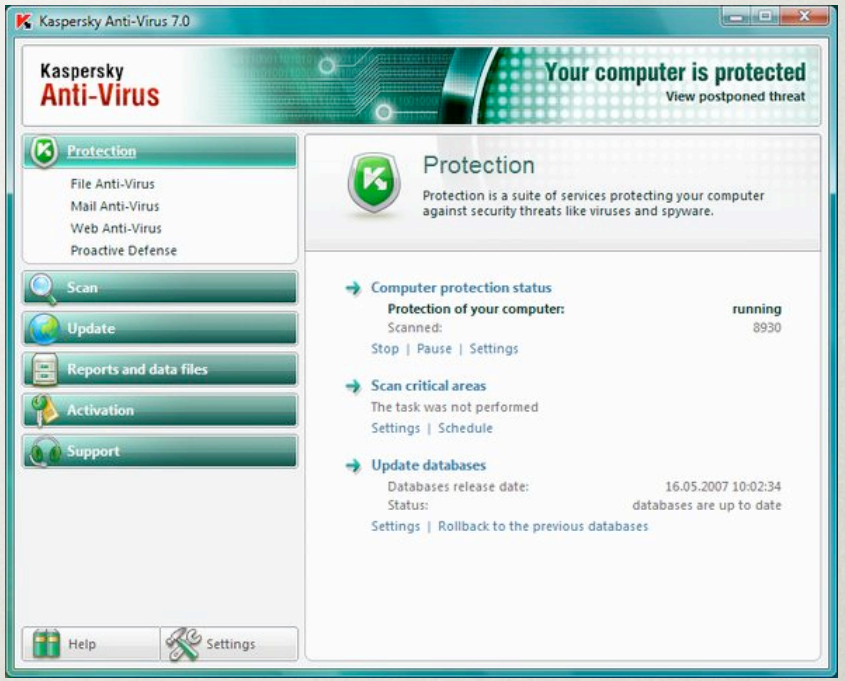
Detection

Outline

- What malware are
- How do they infect hosts
- How do they propagate
- Zoo visit !
- How to detect them
- Worms

Anti-virus

- Analyze system behavior
- Analyze binary to decide if it a virus
- Type :
  - Scanner
  - Real time monitor



Impossibility result

- It is not possible to build a perfect virus/malware detector (Cohen)

Impossibility result

- Diagonal argument
- P is a perfect detection program
- V is a virus
- V can call P
  - if P(V) = true -> halt
  - if P(V) = false -> spread

Virus signature

- Find a string that can identify the virus
- Fingerprint like

Heuristics

- Analyze program behavior
  - Network access
  - File open
  - Attempt to delete file
  - Attempt to modify the boot sector

Checksum

- Compute a checksum for
  - Good binary
  - Configuration file
- Detect change by comparing checksum
- At some point there will more malware than “goodware” ...

Sandbox analysis

- Running the executable in a VM
- Observe it
  - File activity
  - Network
  - Memory

Dealing with Packer

- Launch the exe
- Wait until it is unpack
- Dump the memory

Worms

Outline

- What malware are
- How do they infect hosts
- How do they propagate
- Zoo visit !
- How to detect them
- Worms

Worm

- A worm is self-replicating software designed to spread through the network
  - Typically, exploit security flaws in widely used services
  - Can cause enormous damage
    - Launch DDOS attacks, install bot networks
    - Access sensitive information
    - Cause confusion by corrupting the sensitive information
- Worm vs Virus vs Trojan horse

Cost of worm attacks

- Morris worm, 1988
  - Infected approximately 6,000 machines
    - 10% of computers connected to the Internet
  - cost ~ $10 million in downtime and cleanup
- Code Red worm, July 16 2001

Internet Worm (First major attack)

- Released November 1988
  - Program spread through Digital, Sun workstations
  - Exploited Unix security vulnerabilities
    - VAX computers and SUN-3 workstations running versions 4.2 and 4.3 Berkeley UNIX code

Some historical worms of note

Worm	Date	Distinction
Morris	11/88	Used multiple vulnerabilities, propagate to "nearby" sys
ADM	5/98	Random scanning of IP address space
Ramen	1/01	Exploited three vulnerabilities
Lion	3/01	Stealthy, rootkit worm
Cheese	6/01	Vigilante worm that secured vulnerable systems
Code Red	7/01	First sig Windows worm; Completely memory resident
Walk	8/01	Recompiled source code locally
Nimda	9/01	Windows worm: client-to-server, c-to-c, s-to-s, …
Scalper	6/02	11 days after announcement of vulnerability; peer-to-peer network of compromised systems
Slammer	1/03	Used a single UDP packet for explosive growth

Kienzle and Elder

Increasing propagation speed

- Code Red, July 2001
  - Affects Microsoft Index Server 2.0,
    - Windows 2000 Indexing service on Windows NT 4.0.
    - Windows 2000 that run IIS 4.0 and 5.0 Web servers
  - Exploits known buffer overflow in Idq.dll
  - Vulnerable population (360,000 servers) infected in 14 hours
- SQL Slammer, January 2003
  - Affects in Microsoft SQL 2000
  - Exploits known buffer overflow vulnerability

Code Red

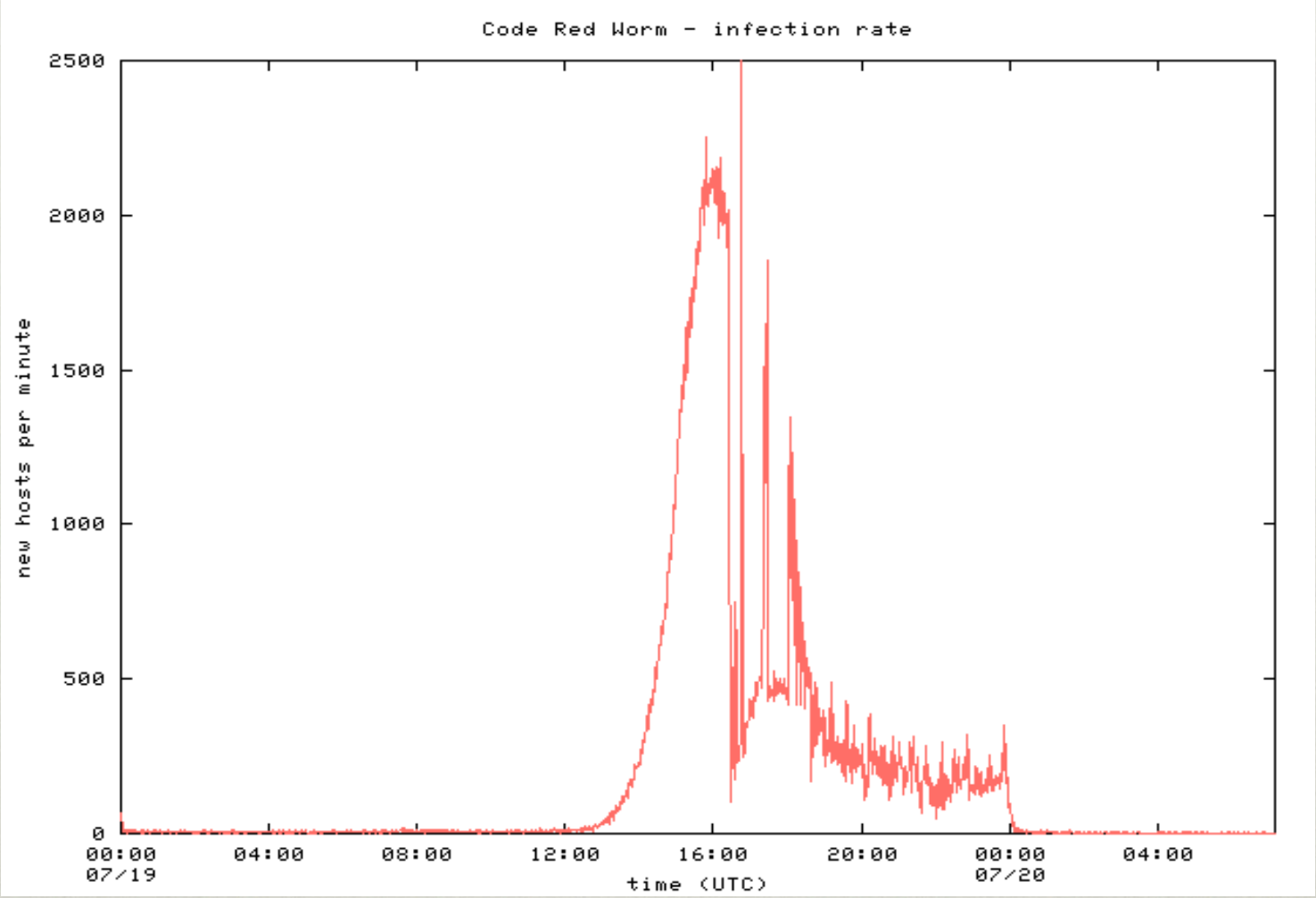
- Initial version released July 13, 2001
  - Sends its code as an HTTP request
  - HTTP request exploits buffer overflow
  - Malicious code is not stored in a file
    - Placed in memory and then run
- When executed,
  - Worm checks for the file C:\Notworm

Code Red of July 13 and July 19

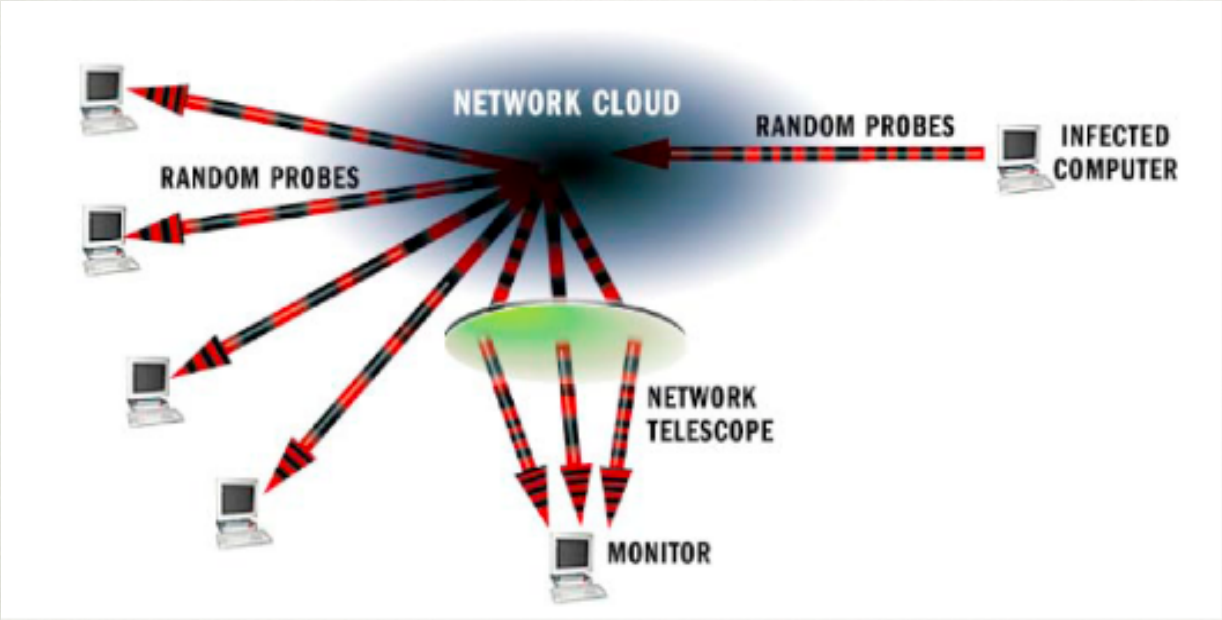
- Initial release of July 13
  - 1st through 20th month: Spread
    - via random scan of 32-bit IP addr space
  - 20th through end of each month: attack.
    - Flooding attack against 198.137.240.91 (*www.whitehouse.gov)*
  - Failure to seed random number generator ⇒ *linear growth*
- Revision released July 19, 2001.
  - White House responds to threat of flooding attack by changing the address of *www.whitehouse.gov*
  - Causes Code Red to die for date ≥ 20th of the month.
  - But: this time random number generator correctly seeded

Slides: Vern Paxson

Infection rate



Measuring activity: network telescope



- Monitor cross-section of Internet address space, measure traffic
  - "Backscatter" from DOS floods
  - Attackers probing blindly
  - Random scanning from worms
- LBNL's cross-section: 1/32,768 of Internet
- UCSD, UWisc's cross-section: 1/256.

Spread of Code Red

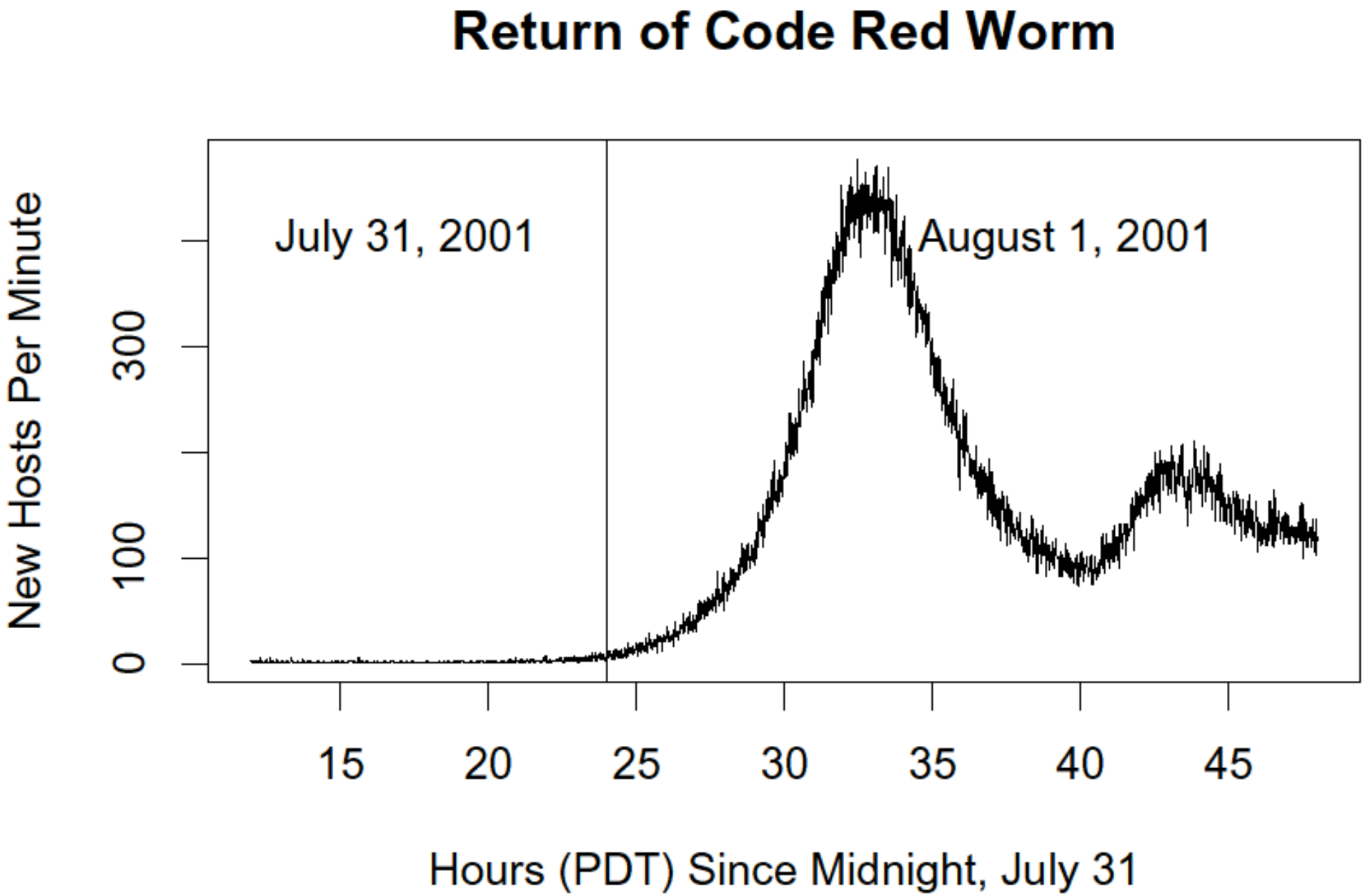
- Network telescopes estimate of # infected hosts: 360K. (Beware DHCP & NAT)
- Course of infection fits classic *logistic*.
- Note: larger the vulnerable population, *faster* the worm spreads.

- That night ($\Rightarrow 20^{th}$), worm dies …

  … except for hosts with inaccurate clocks!

- It just takes one of these to restart the worm on August 1^{st} …

Slides: Vern Paxson



Slides: Vern Paxson

Code Red 2

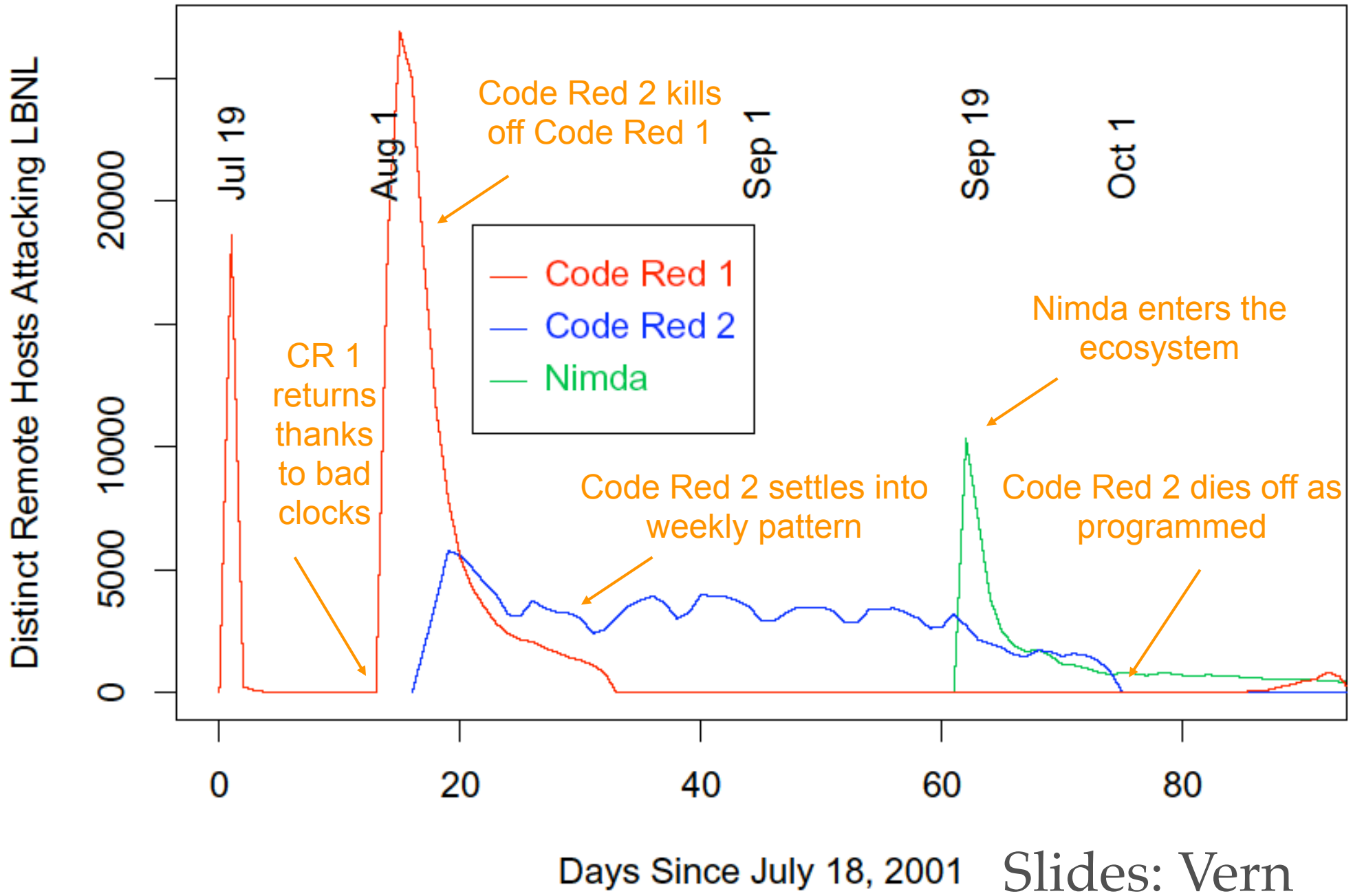
- Released August 4, 2001.
- Comment in code: “Code Red 2.”
  - But in fact completely different code base.
- Payload: a root backdoor, resilient to reboots.
- Bug: crashes NT, only works on Windows 2000.
- *Localized scanning*: prefers nearby addresses.
- Kills Code Red 1.
- Safety valve: programmed to die Oct 1, 2001.

Slides: Vern Paxson

Striving for Greater Virulence: Nimda

- Released September 18, 2001.
- Multi-mode spreading:
  - attack IIS servers via infected clients
  - email itself to address book as a virus
  - copy itself across open network shares
  - modifying Web pages on infected servers w/ client exploit
  - scanning for Code Red II backdoors (!)
- worms form an ecosystem!
- Leaped across firewalls.

Slides: Vern Paxson



Slides: Vern Paxson

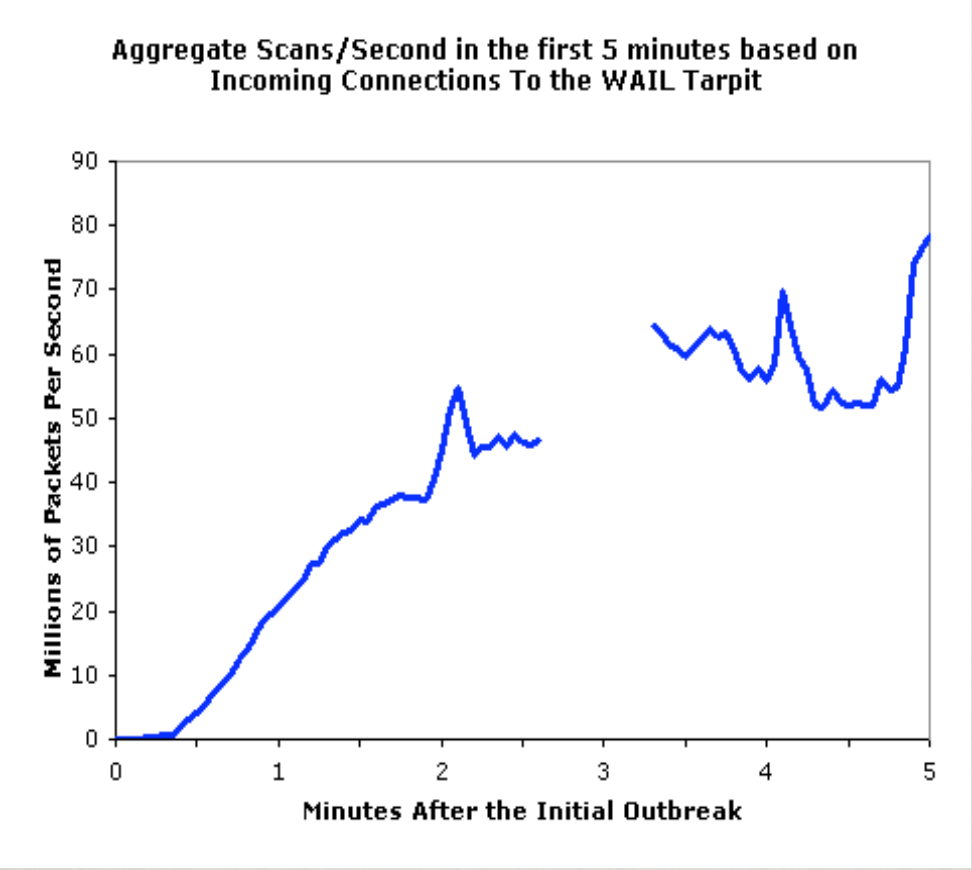
How do worms propagate?

- **Scanning** worms : Worm chooses "random" address
- **Coordinated** scanning : Different worm instances scan different addresses
- **Flash** worms
  - Assemble tree of vulnerable hosts in advance, propagate along tree
    - Not observed in the wild, yet
    - Potential for 106 hosts in < 2 sec ! [Staniford]
- **Meta-server** worm :Ask server for hosts to infect (e.g., Google for "powered by phpbb")
- **Topological** worm: Use information from infected hosts (web server logs, email address books, config files, SSH "known hosts")
- **Contagion** worm : Propagate parasitically along with normally initiated communication

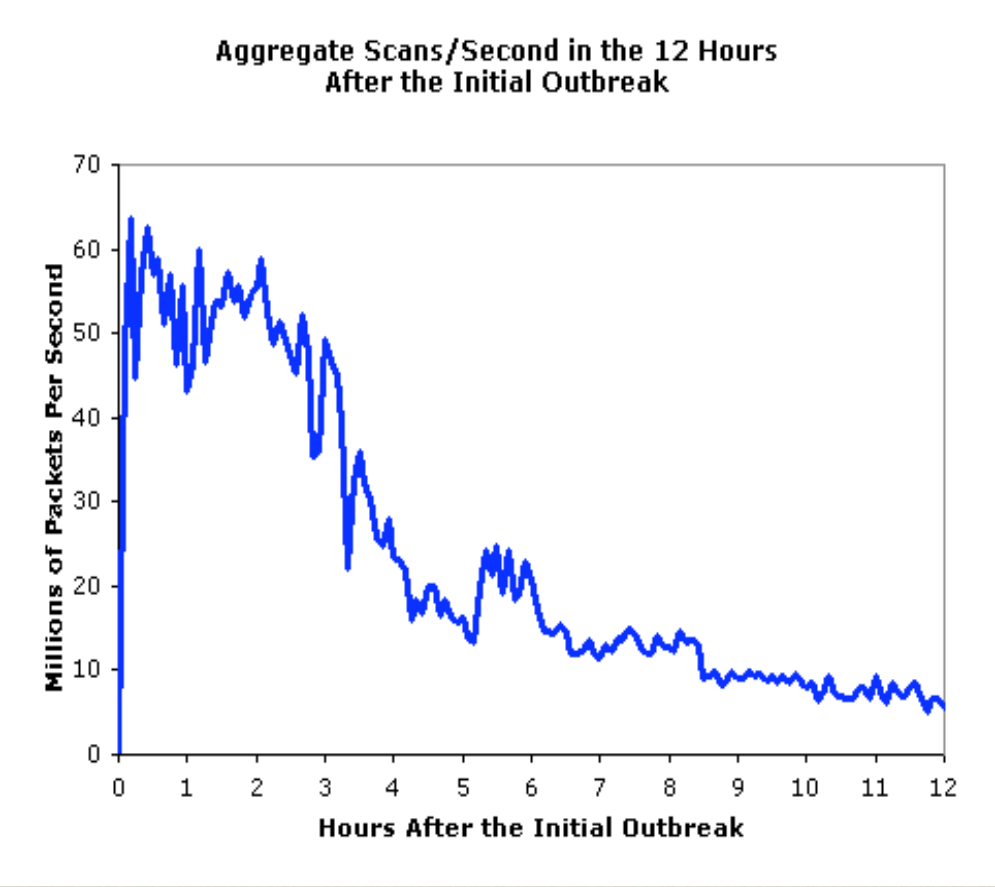
SLAMMER

- 01/25/2003
- Vulnerability disclosed : 25 june 2002
- Better scanning algorithm
- UDP Single packet : 380bytes

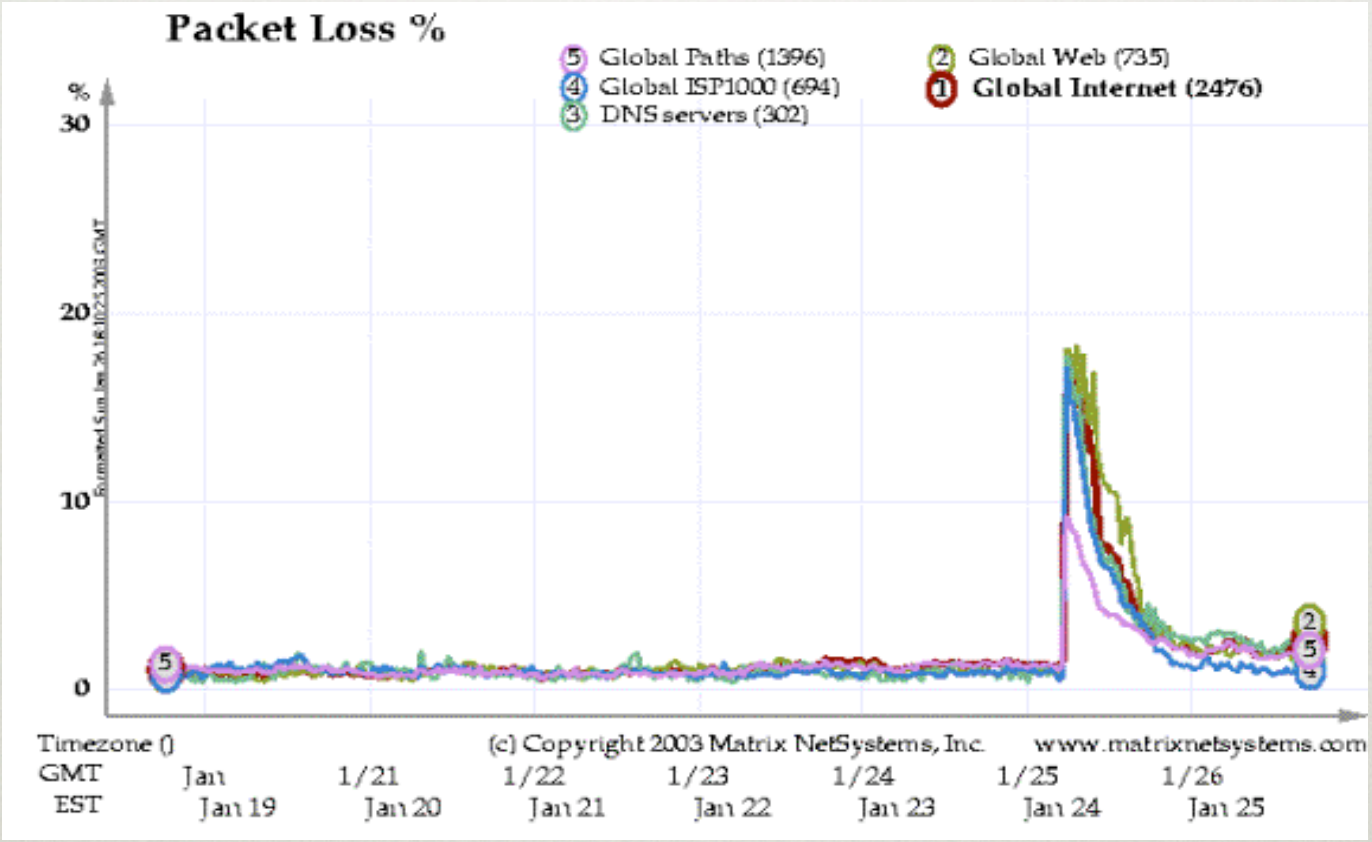
Slammer propagation



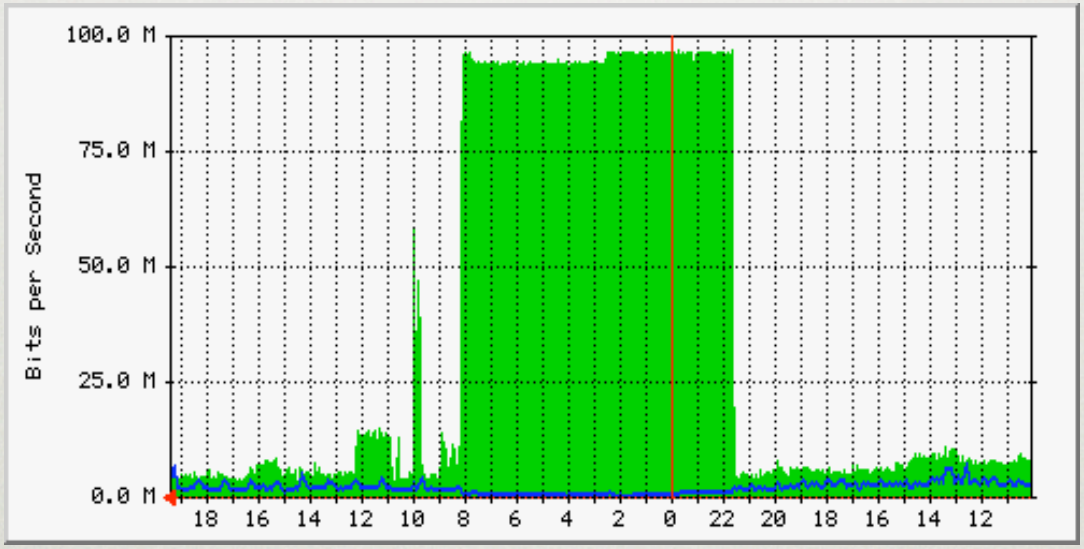
Number of scan/sec



Packet loss


Packet Loss %
5 Global Paths (1396)
4 Global ISP1000 (694)
3 DNS servers (302)
2 Global Web (735)
1 Global Internet (2476)
%
30
20
10
0
Timezone ()
GMT
EST
(c) Copyright 2003 Matrix NetSystems, Inc.
www.matrixnetsystems.com
Jan
1/21
1/22
1/23
1/24
1/25
1/26
Jan 19
Jan 20
Jan 21
Jan 22
Jan 23
Jan 24
Jan 25

A server view



Consequences

- ATM systems not available
- Phone network overloaded (no 911!)
- 5 DNS root down
- Planes delayed

Worm Detection and Defense

- Detect via *honeyfarms*: collections of "honeypots" fed by a network telescope.
  - Any outbound connection from honeyfarm = worm.

    (at least, that's the theory)
  - Distill *signature* from inbound/outbound traffic.
  - If telescope covers N addresses, expect detection when worm has infected 1/N of population.
- Thwart via *scan suppressors*: network elements that block traffic from hosts that make failed connection attempts to too many other hosts
  - 5 minutes to several weeks to write a signature
  - Several hours or more for testing

Need for automation

- Current threats can spread faster than defenses can reaction
- Manual capture/analyze/signature/rollout model too slow

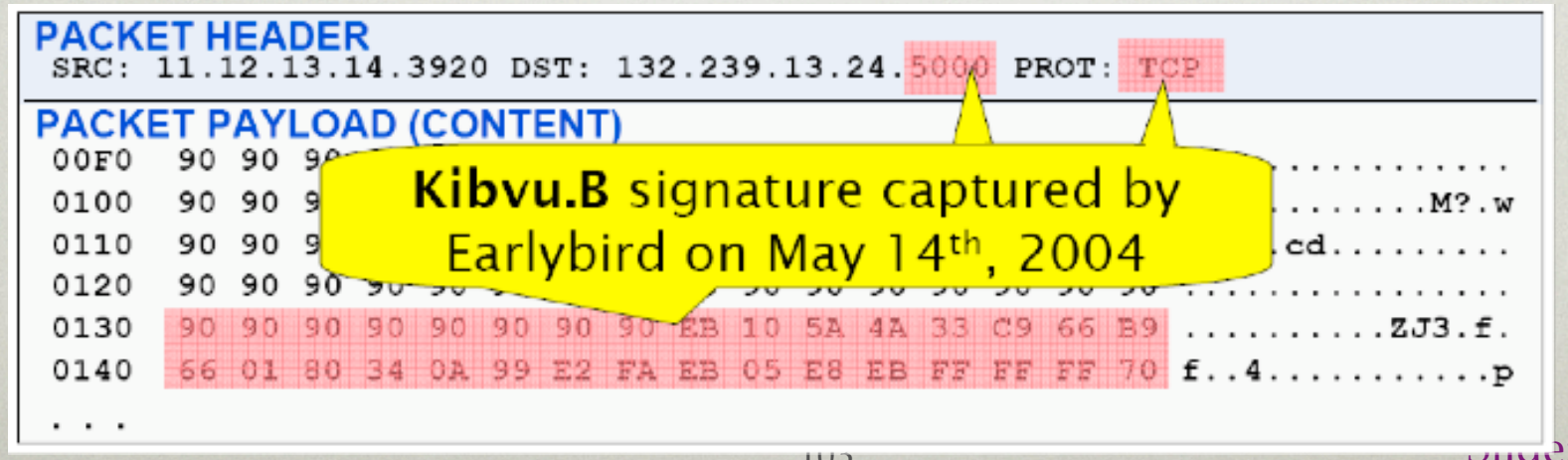
Slide: Carey Nachenberg, Symantec

Signature inference

- Challenge
  - need to automatically learn a content "signature" for each new worm – potentially in less than a second!
- Some proposed solutions
  - Singh et al, Automated Worm Fingerprinting, OSDI '04
  - Kim et al, Autograph: Toward Automated, Distributed Worm Signature Detection, USENIX Sec '04

Signature inference

- Monitor network and look for strings common to traffic with worm-like behavior
  - Signatures can then be used for content filtering



Slide: S Savage

Content sifting

- Assume there exists some (relatively) unique invariant bitstring W across all instances of a particular worm (true today, not tomorrow...)
- Two consequences
  - **Content Prevalence**: W will be more common in traffic than other bitstrings of the same length
  - **Address Dispersion**: the set of packets containing W will address a disproportionate number of distinct sources and destinations
- Content sifting: find W's with high content prevalence and high address dispersion and drop that traffic

Slide: S Savage

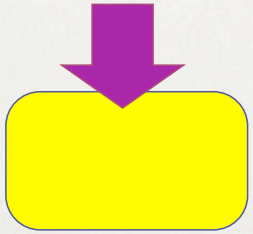
Observation:
High-prevalence strings are rare

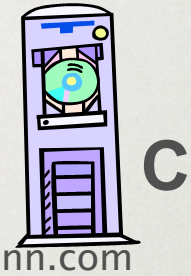
Only 0.6% of the 40 byte **substrings repeat more than 3 times** in a minute

(Stefan Savage, UCSD *)

The basic algorithm

Detector in network



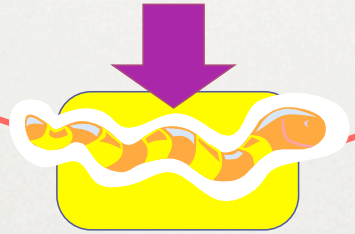


Prevalence Table

Address **Dispersion** Table

Sources Destinations

Detector in network



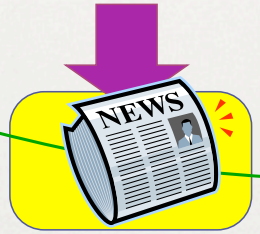
Prevalence Table

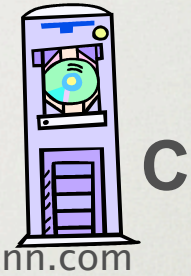
	1

Address **Dispersion** Table

Sources	Destinations
1 (A)	**1 (B)**

Detector in network





Prevalence Table

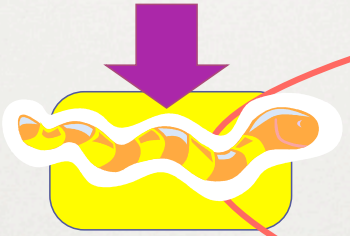
	**1**
	**1**

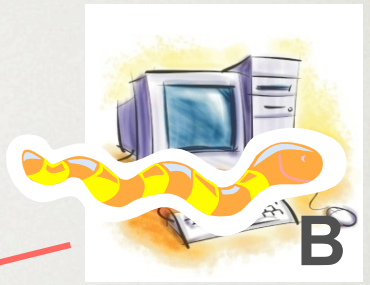
Address **Dispersion** Table

Sources	Destinations
1 (A)	**1 (B)**
1 (C)	**1 (A)**

(Stefan Savage, UCSD *)

Detector in network





Prevalence Table

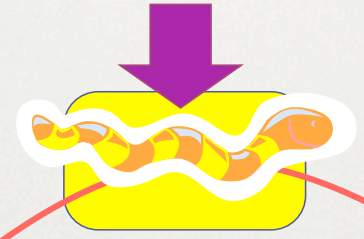
	**2**
	**1**

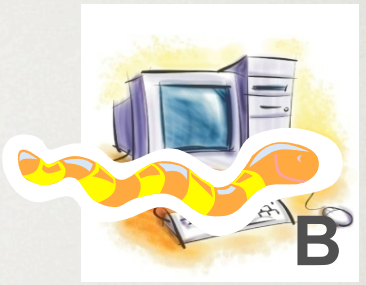
Address **Dispersion** Table

Sources	Destinations
2 (A,B)	**2 (B,D)**
1 (C)	**1 (A)**

(Stefan Savage, UCSD *)

Detector in network





Prevalence Table

3	(A,B,D)
1	(C)

Address **Dispersion** Table

Sources	Destinations
3	3
(B,D,E)	(A,B,D)
1 (A)	1 (C)

Challenges

- Computation
  - To support a 1Gbps line rate we have 12us to process each packet, at 10Gbps 1.2us, at 40Gbps…
    - Dominated by memory references; state expensive
  - Content sifting requires looking at every byte in a packet
- State
  - On a fully-loaded 1Gbps link a naïve implementation can easily consume 100MB/sec for table
  - Computation/memory duality: on high-speed (ASIC) implementation, latency requirements may limit state to on-chip SRAM