

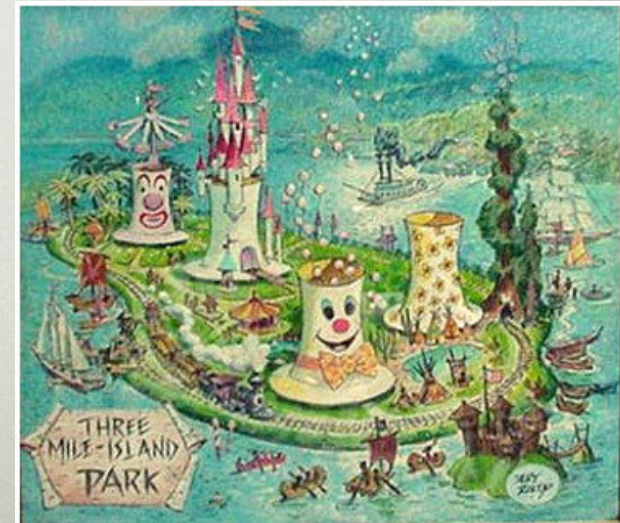
MALWARE

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

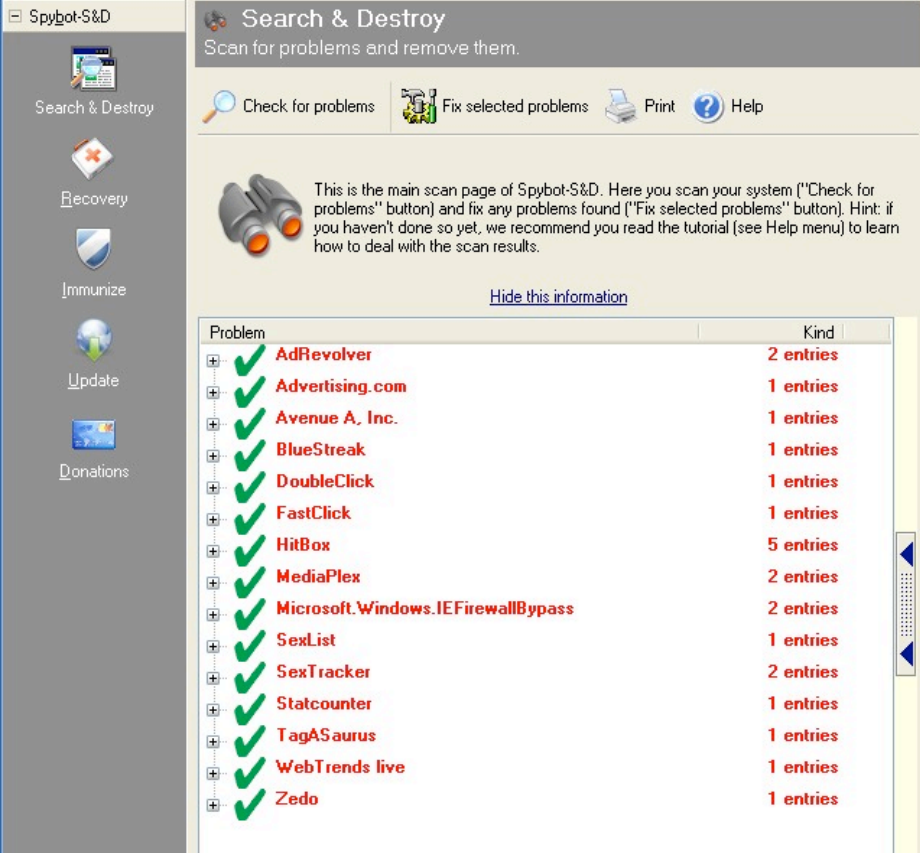
- Christians On Facebook
- Leader hacked on march 2009
- Post Islamic message
- Lost >10 000 members

The screenshot shows a Facebook group page with the following details:

- Group Name:** There is No God but ALLAH; Muhammad is the messenger of Allah
- Group Type:** This is an open group. Anyone can join and invite others to join.
- Admins:** Sarah Hs (Brazil)
- Basic Info:** Common Interest - Religion & Spirituality. Description: "La ilaha illallah". (There is no god but Allah). The phrase is the bedrock of Islam. The belief in the oneness of God is called Tawhid and it is a prerequisite for being a Muslim. Tawhid is to believe that Allah is one and there is none like Him; He has no partner; He neither begets nor is He begotten; He is indivisible in person; He is eternal; He is infinite; He has neither beginning nor end; He is All-Mighty, the All-Knowing, the All-Just, the Cherisher of all worlds, the Patron, the Guide, the Helper, the Merciful, the Compassionate, etc.
- Group Description:** "La ilaha illallahu Muhammad rasulullah". (There is no god but Allah; Muhammad is the messenger of Allah.) This phrase is the first principle of Islam. There are two parts of this declaration: (1) La ilaha illallah, (Tawhid) and (2) Muhammad rasulullah. (Risalah). The belief in Tawhid shapes and influences the entire course of our life. Risalah is the channel of communication between Allah and mankind.
- Declarations of Faith:** Several sections provide translations of Islamic declarations of faith, such as "Kalimatul Shahadah", "Kalimatul Tawhid", "Kalimatul Tawheed", "Kalimatul Tawhid", and "Kalimatul Raddul kuf".

THE MALWARE ZOO

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



The screenshot shows the Spybot-S&D Search & Destroy interface. The main window title is "Search & Destroy" with the subtitle "Scan for problems and remove them." Below the title bar, there are four buttons: "Check for problems", "Fix selected problems", "Print", and "Help". A message box with a binoculars icon states: "This is the main scan page of Spybot-S&D. Here you scan your system ('Check for problems' button) and fix any problems found ('Fix selected problems' button). Hint: if you haven't done so yet, we recommend you read the tutorial (see Help menu) to learn how to deal with the scan results." Below this message is a "Hide this information" link. The main area contains a table of detected problems.

Problem	Kind
AdRevolver	2 entries
Advertising.com	1 entries
Avenue A, Inc.	1 entries
BlueStreak	1 entries
DoubleClick	1 entries
FastClick	1 entries
HitBox	5 entries
MediaPlex	2 entries
Microsoft.Windows.IEFirewallBypass	2 entries
SexList	1 entries
SexTracker	2 entries
Statcounter	1 entries
TagASaurus	1 entries
WebTrends live	1 entries
Zedo	1 entries

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)
- Methamorpnic : 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

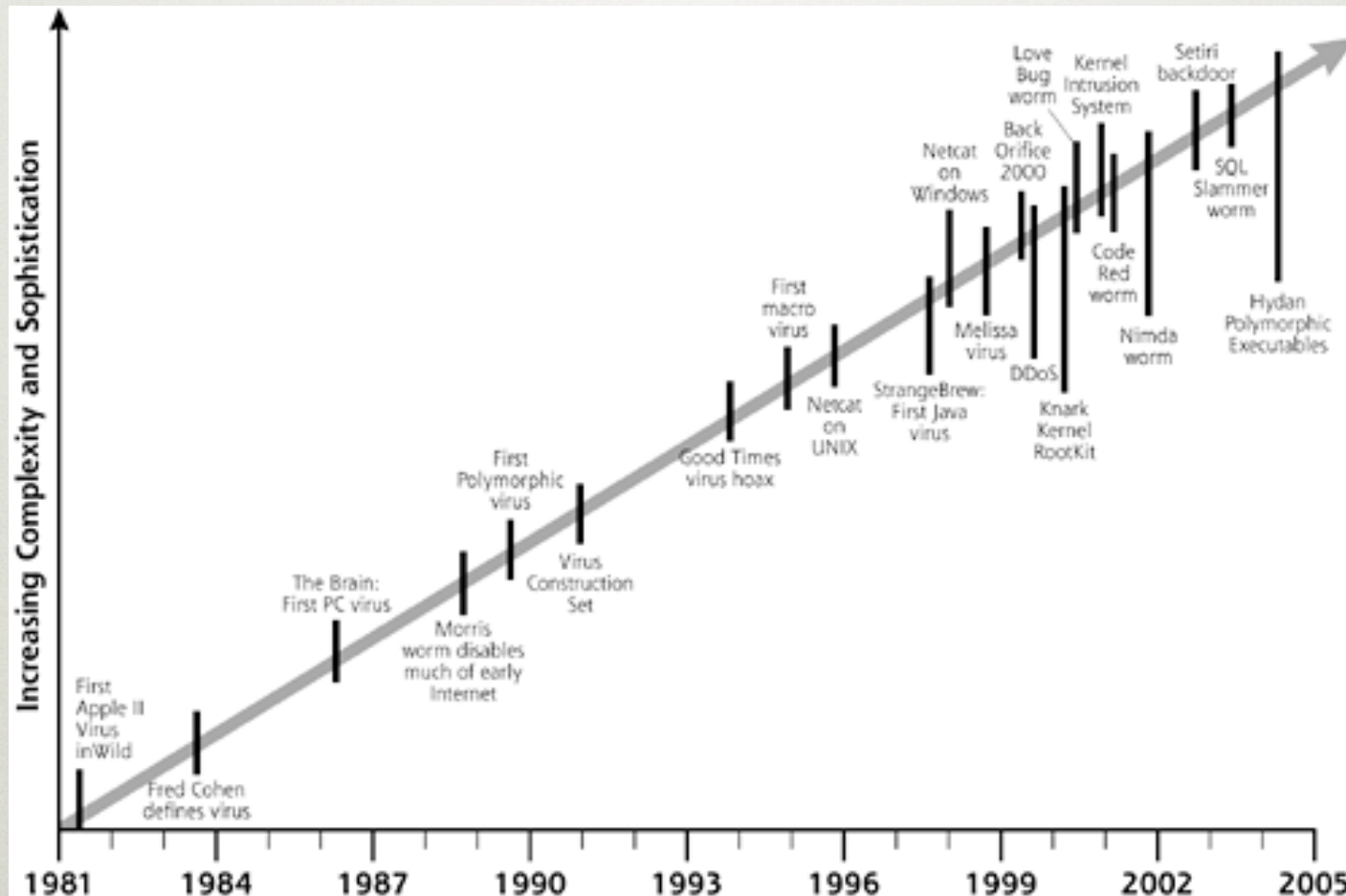
WHAT IS A WORM

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



Melissa spread by email and share

Knark rootkit made by creed demonstrate the first ideas

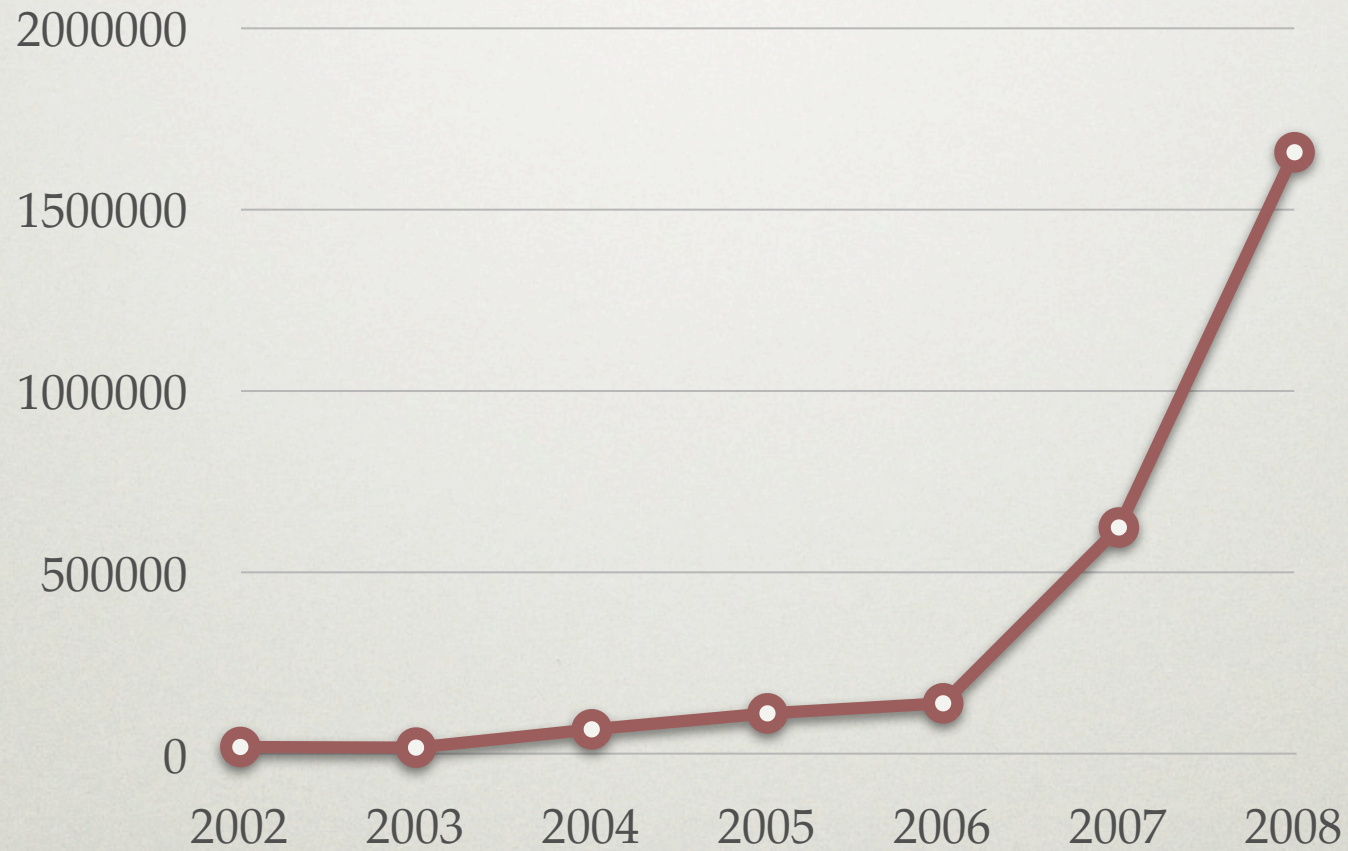
love bug vb script that abused a weakness in outlook

Kernl intrusion by optyx gui and effcient hiding

ISTORY

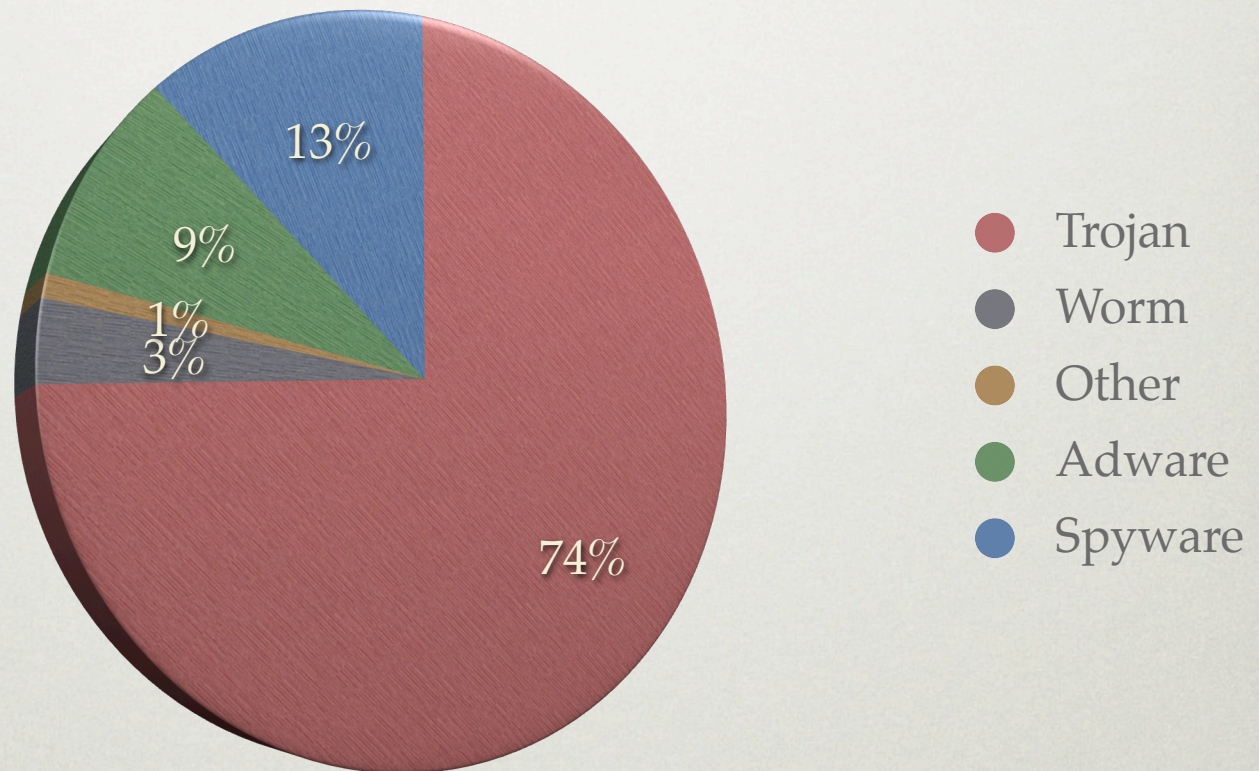
- 1981 First reported virus : Elk Cloner (Apple 2)
- 1983 Virus get defined
- 1986 First PC virus MS DOS
- 1988 First worm : Morris worm
- 1990 First polymorphic virus
- 1998 First Java virus
- 1998 Back orifice
- 1999 Melissa virus
- 1999 Zombie concept
- 1999 Knark rootkit
- 2000 love bug
- 2001 Code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

NUMBER OF MALWARE SIGNATURES



Symantec report 2009

MALWARE REPARTITION



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

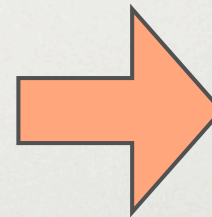
- Executable
- Interpreted file
- Kernel
- Service
- MBR
- Hypervisor



OVERWRITING MALWARE

Malware

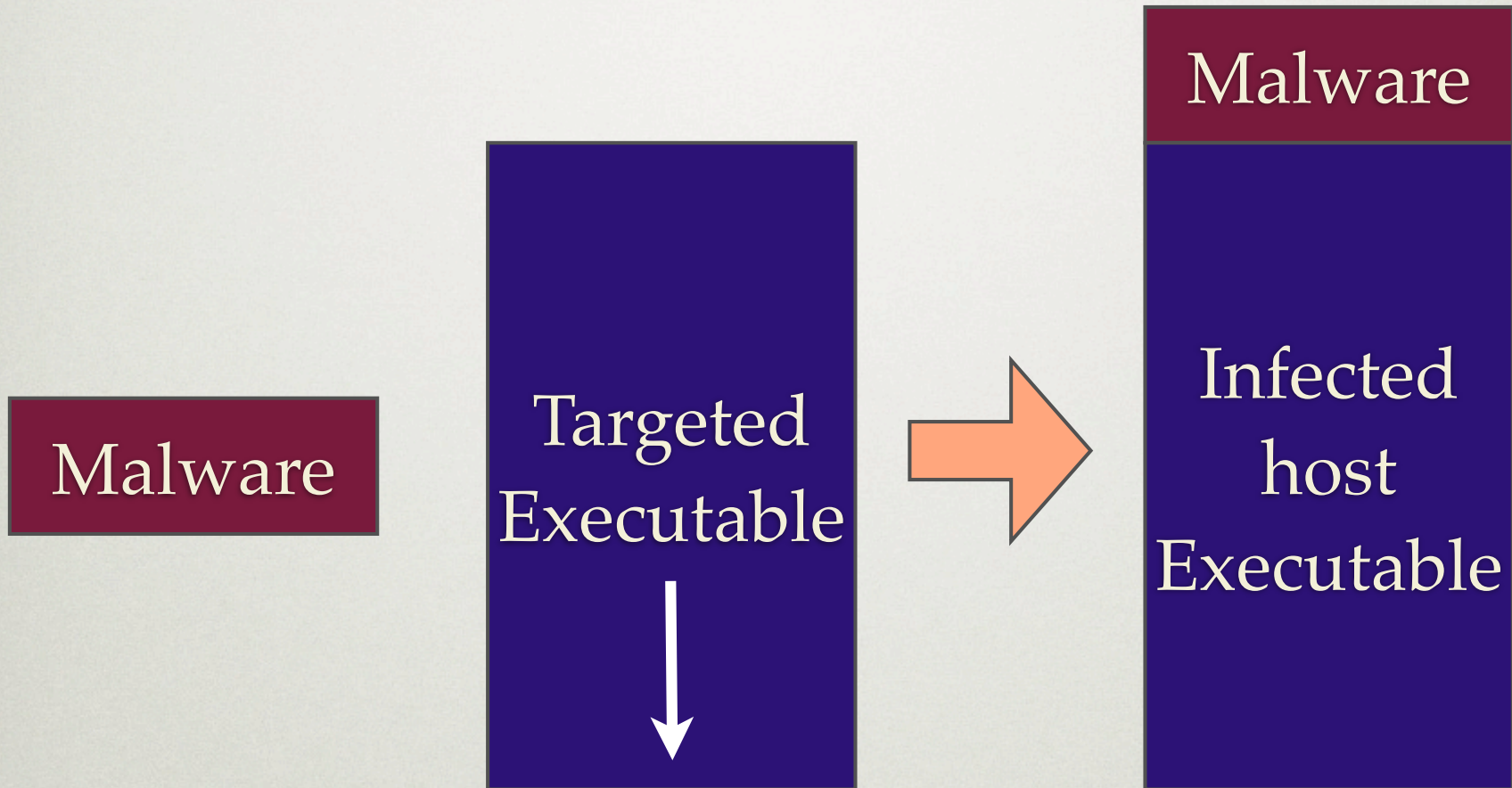
Targeted
Executable



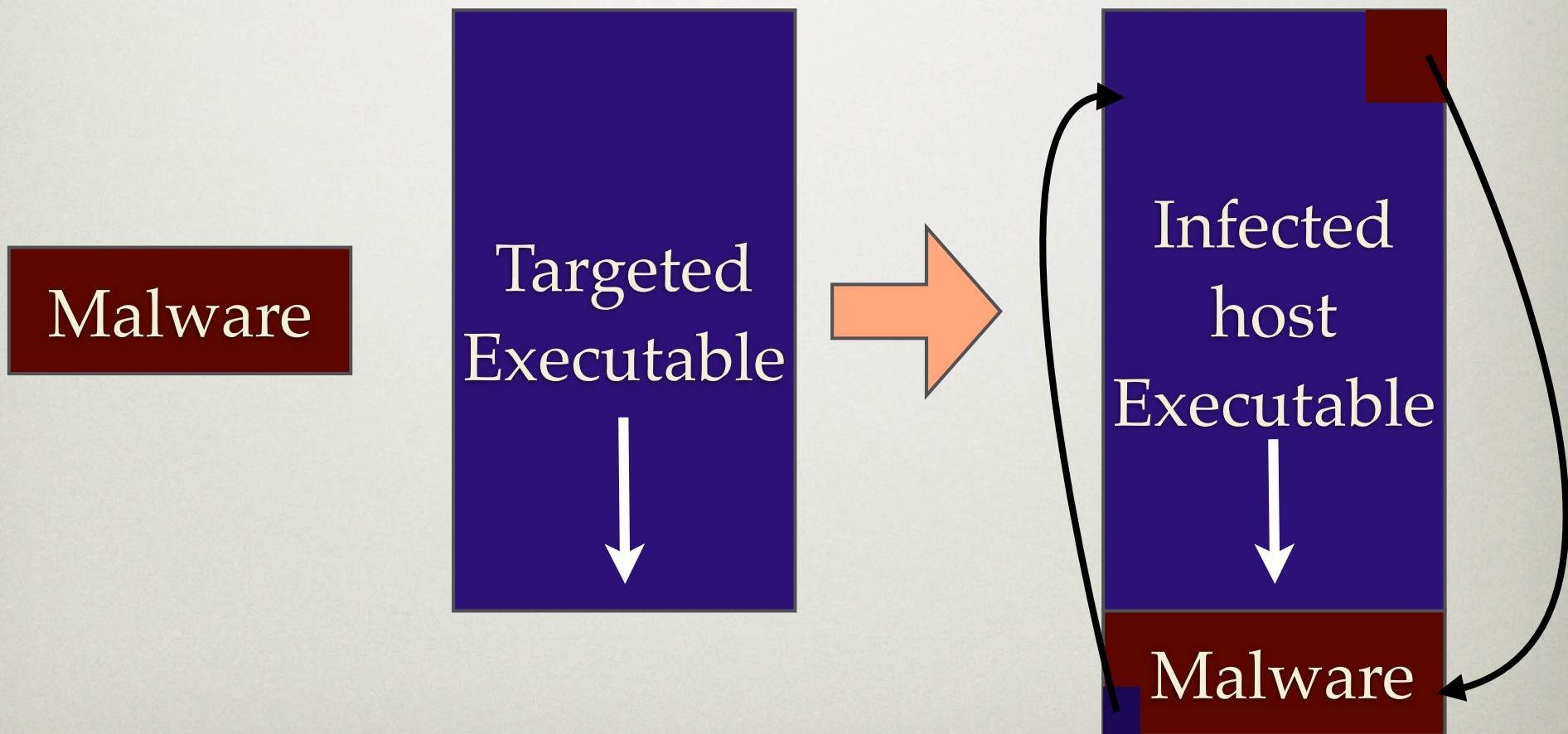
Malware



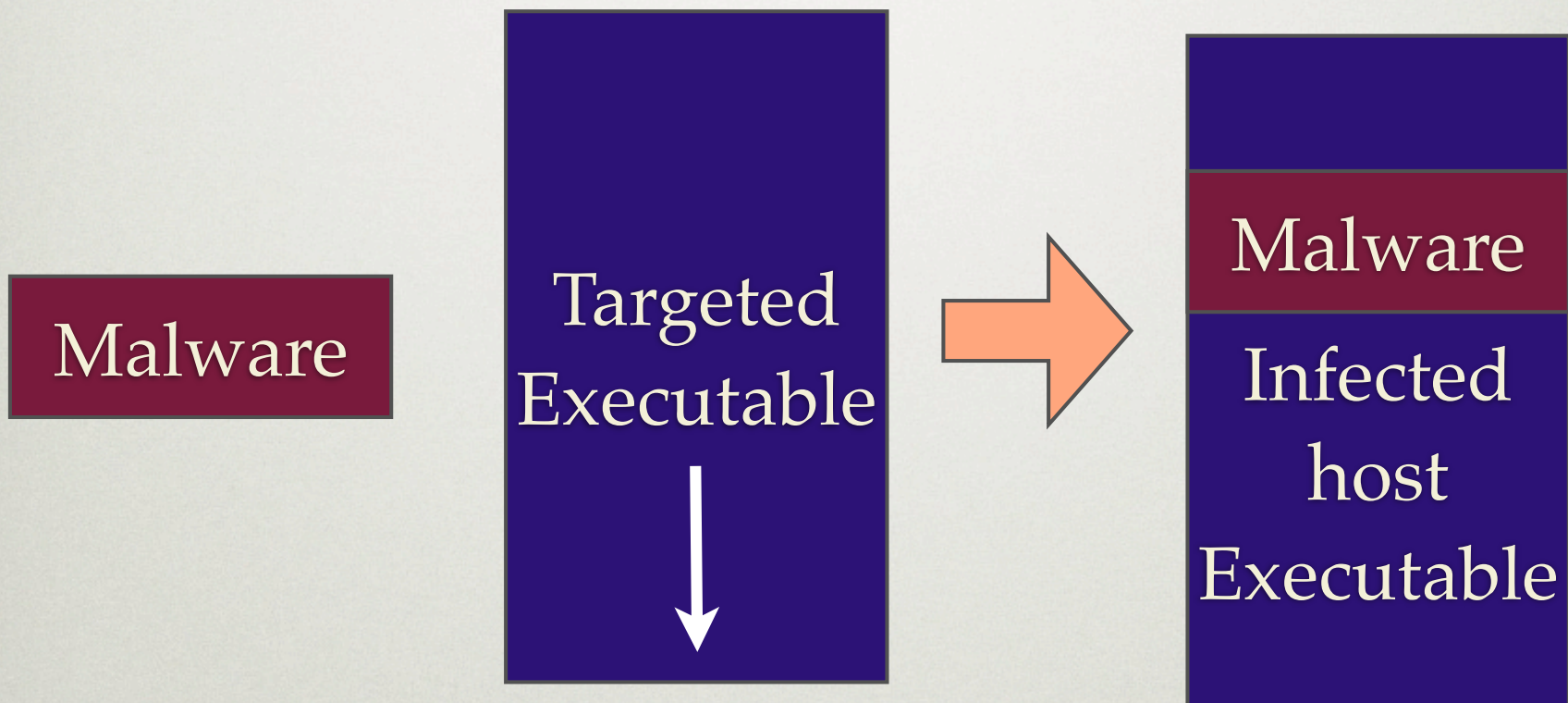
PREPENDING MALWARE



APPENDING MALWARE



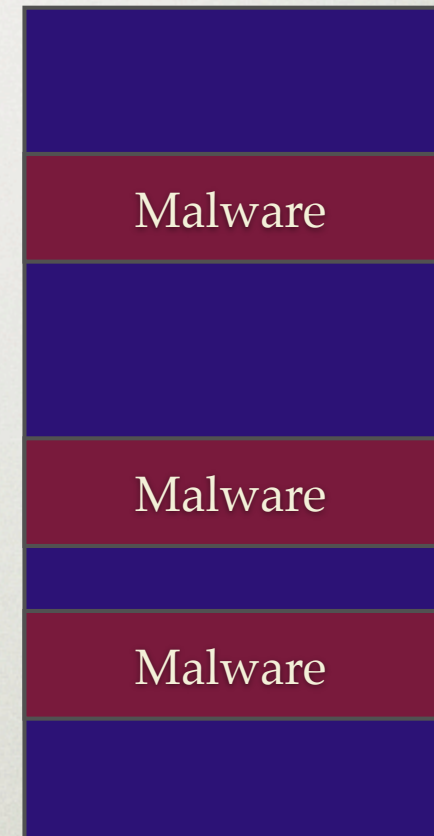
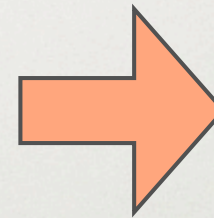
CAVITY MALWARE



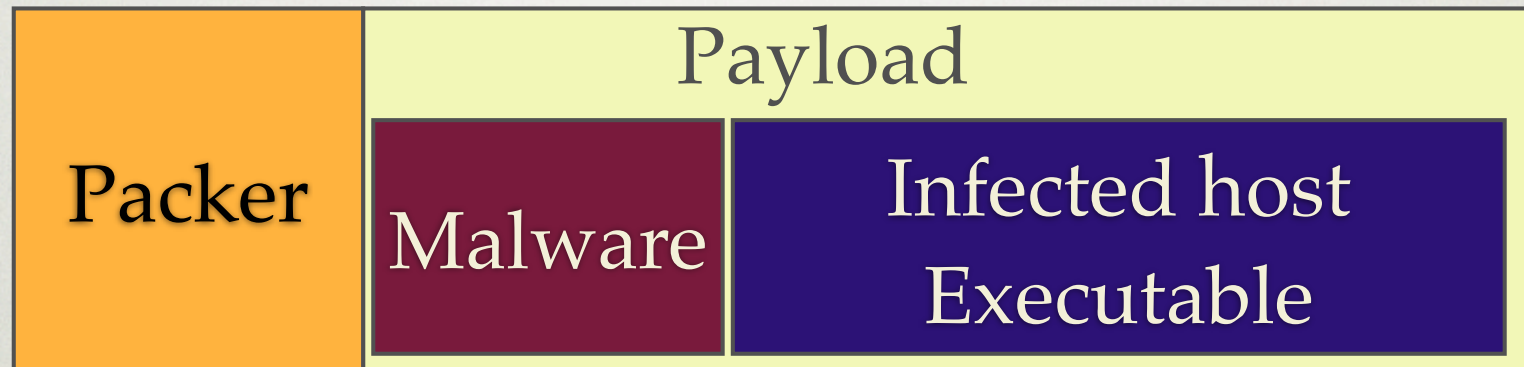
MULTI-CAVITY MALWARE

Malware

Targeted
Executable



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 known 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

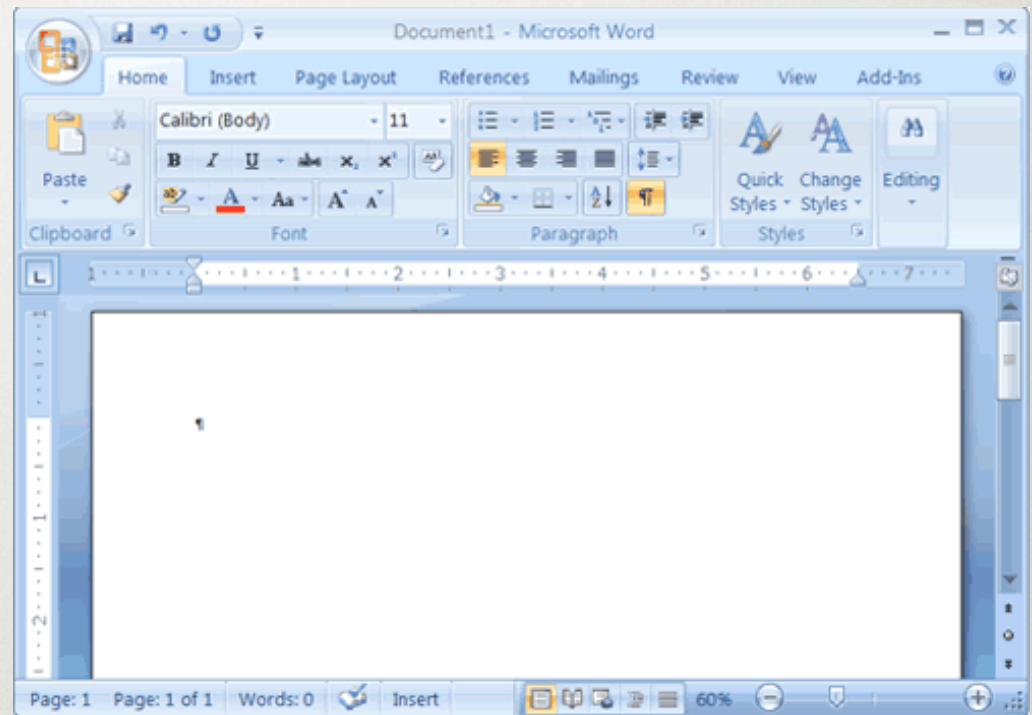
- Init.d
- /etc/rc.local
- .login .xsession
- crontab
 - crontab -e
 - /etc/crontab

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

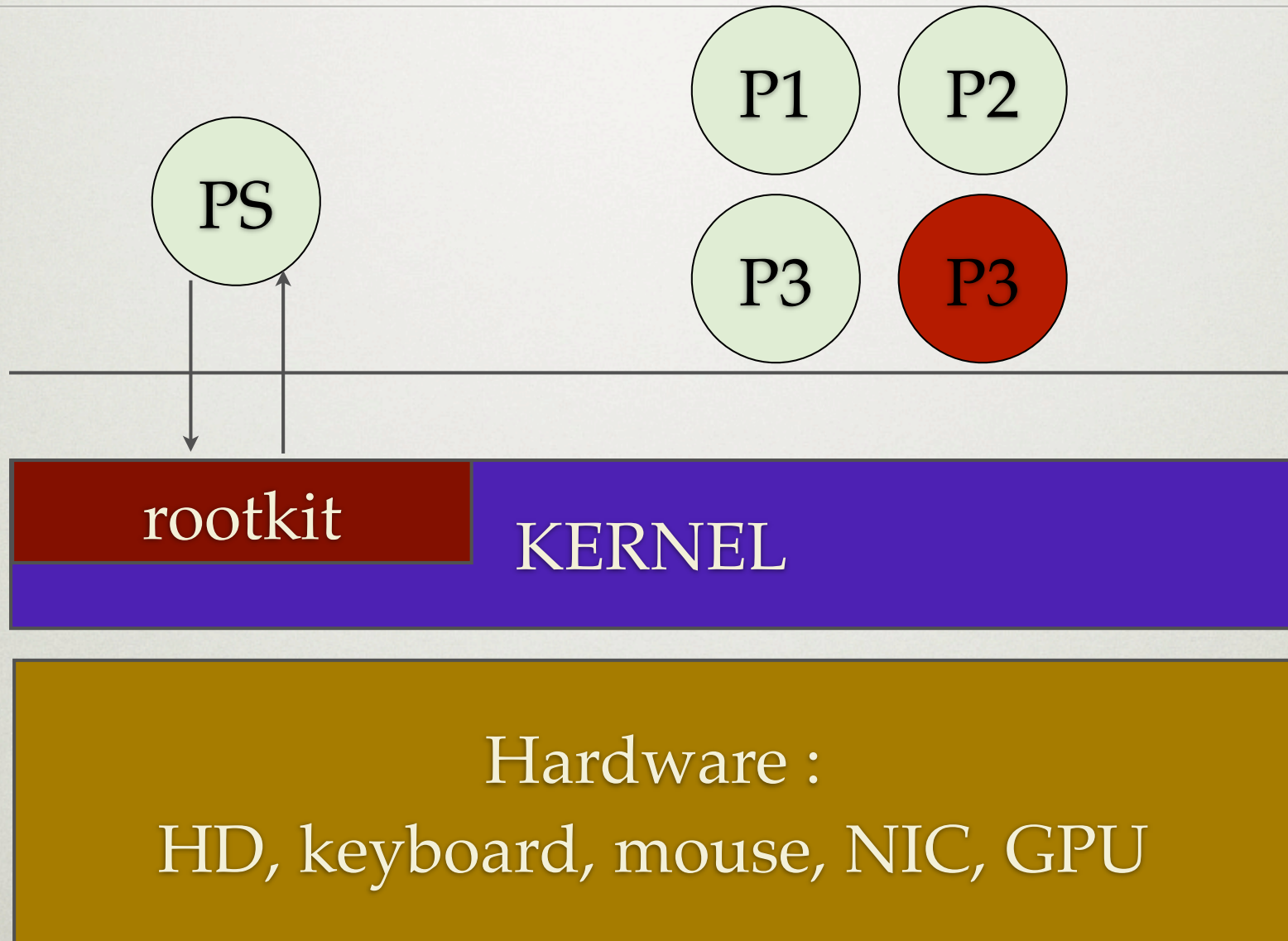
Kernel task

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

What to hide

- ➔ Process
- ➔ Files
- ➔ Network traffic

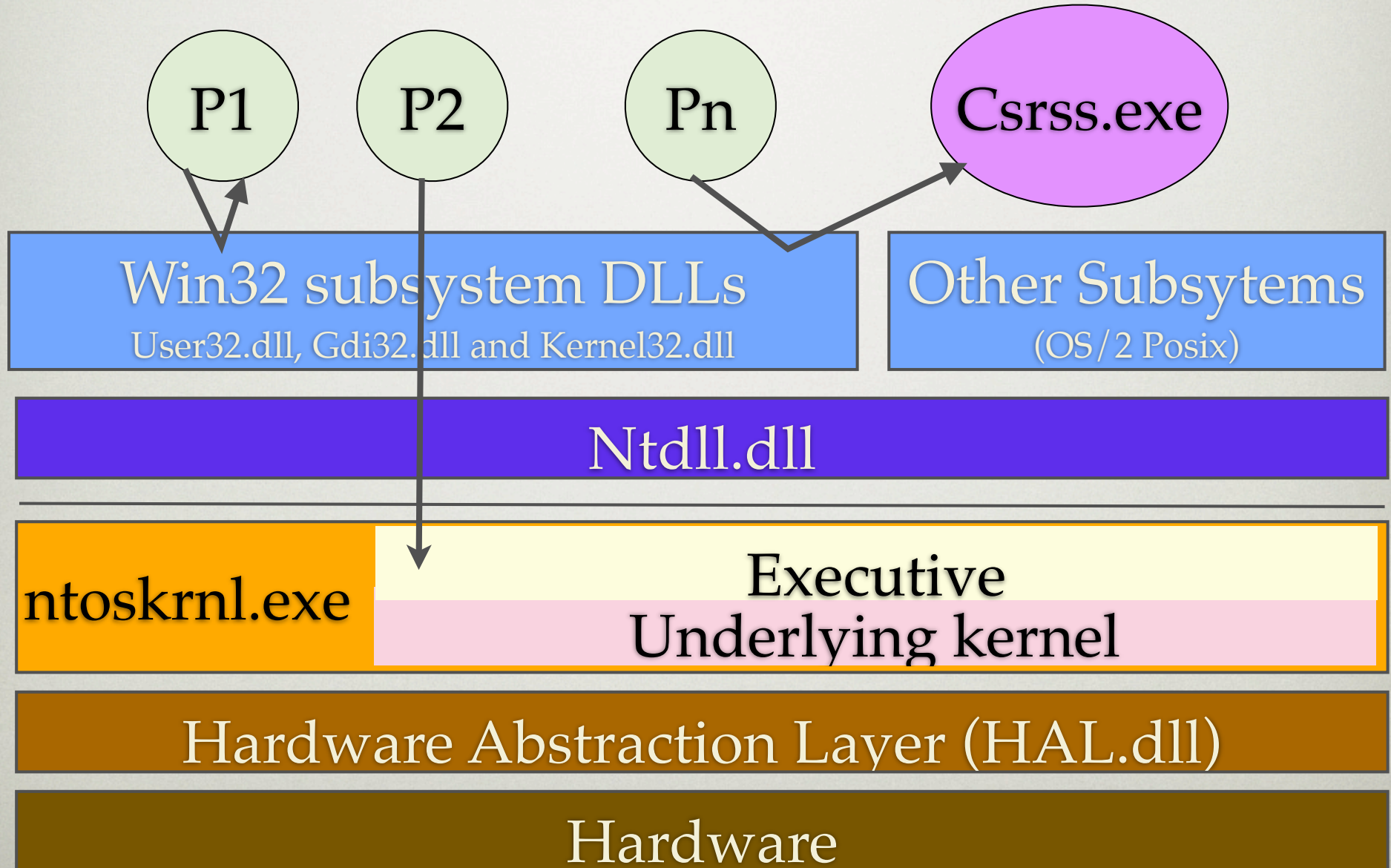
KERNEL ROOTKIT



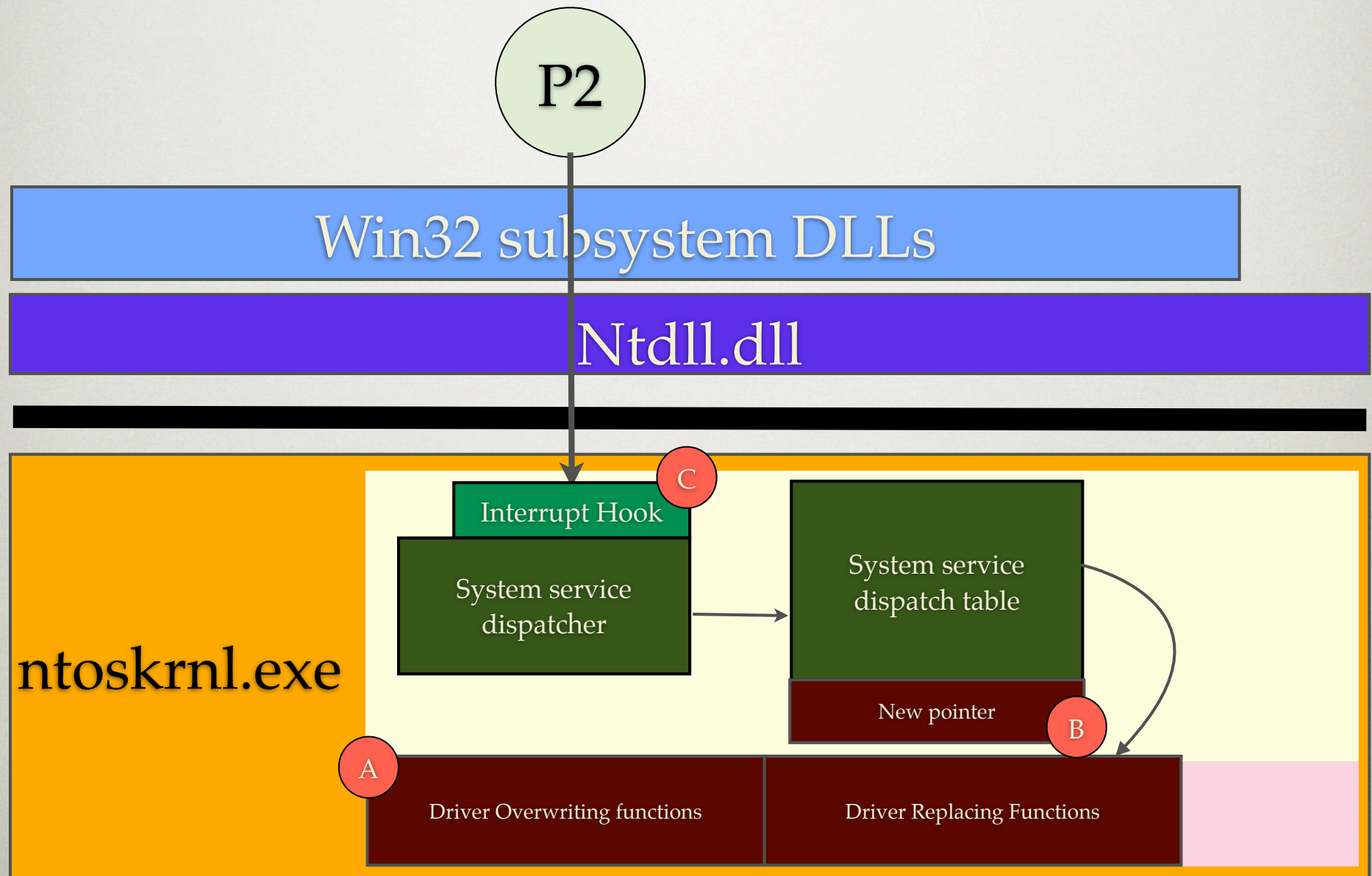
SUBVERTING TECHNIQUES

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

WINDOWS KERNEL

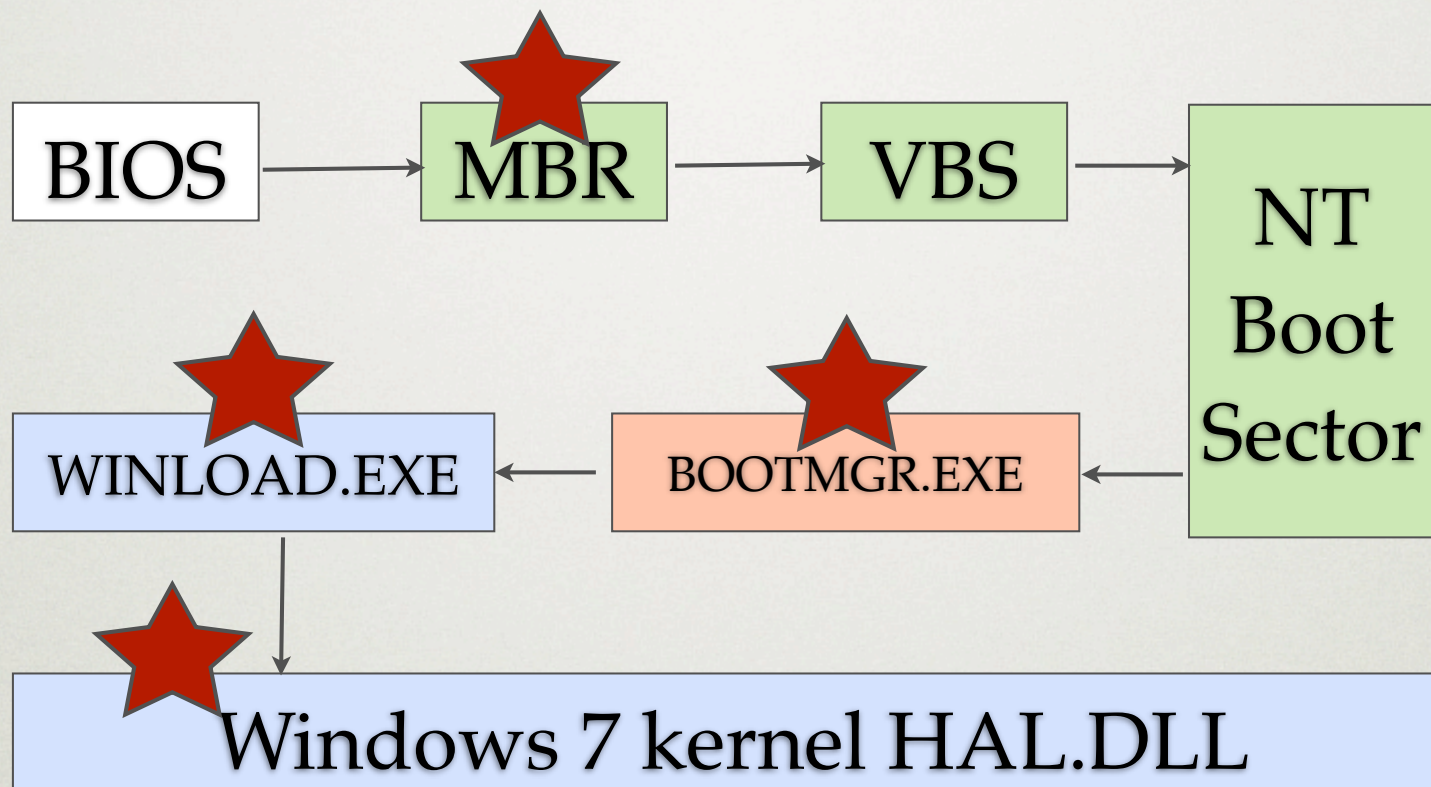


KERNEL DEVICE DRIVER



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.



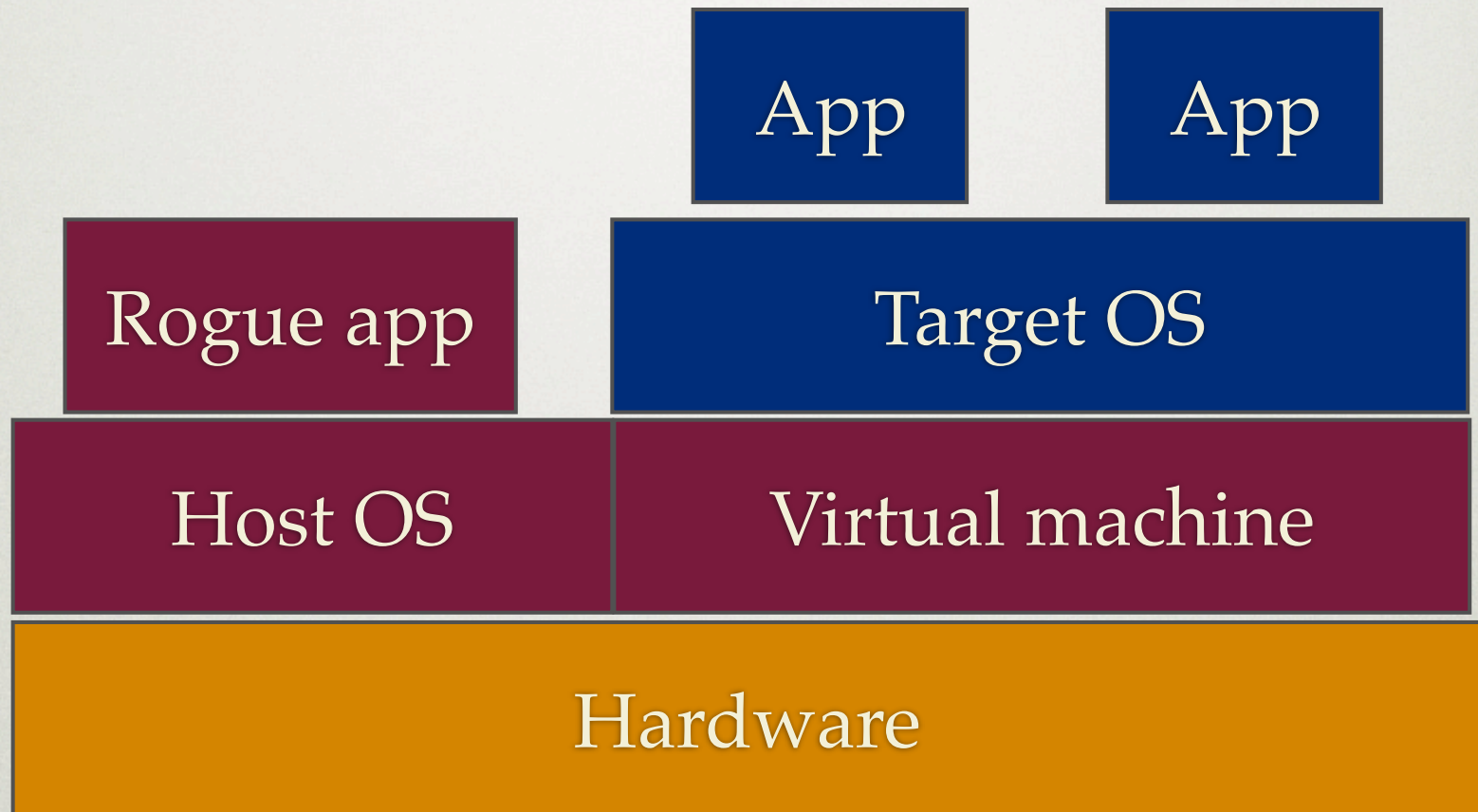
VBOOT

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

HYPERVERSITOR ROOTKIT



HYPERVERSOR ROOTKIT



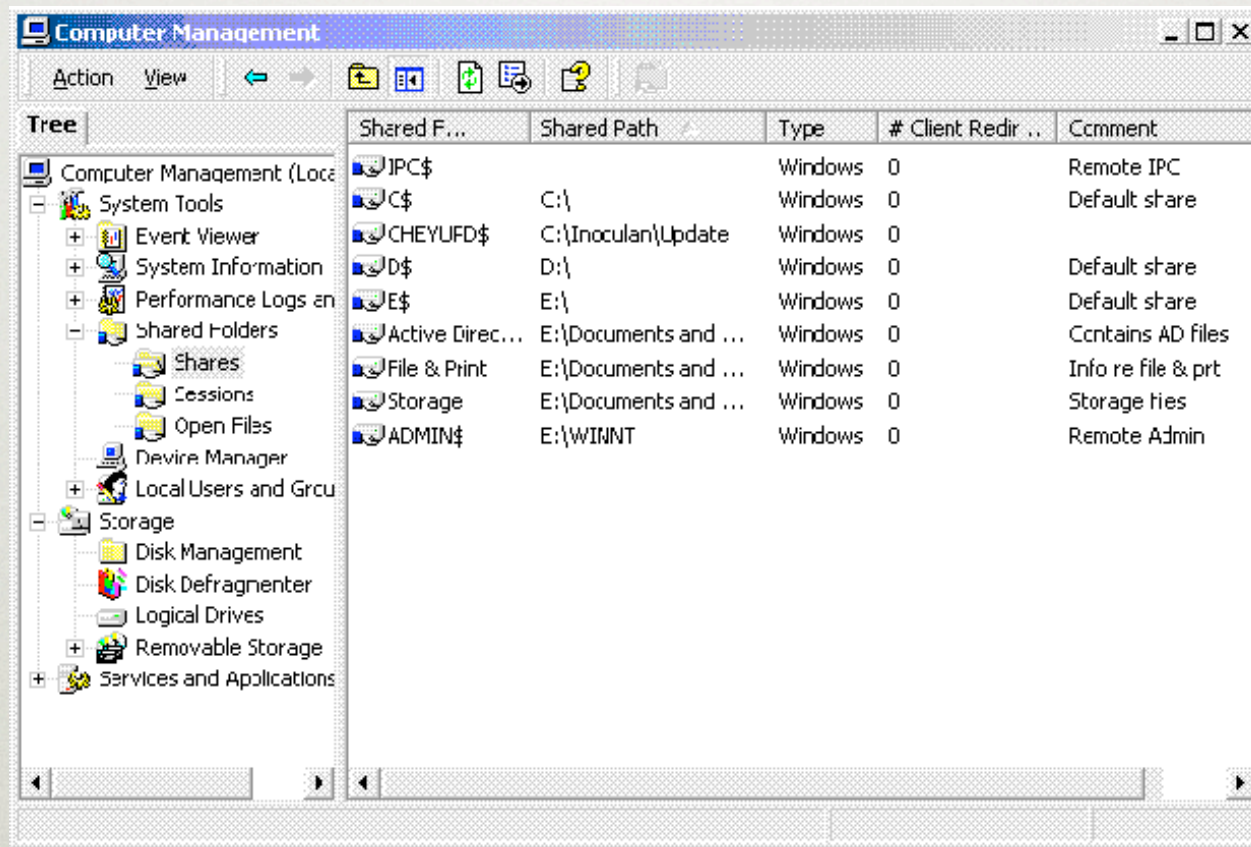
PROPAGATION VECTOR

OUTLINE

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



SHARED FOLDER

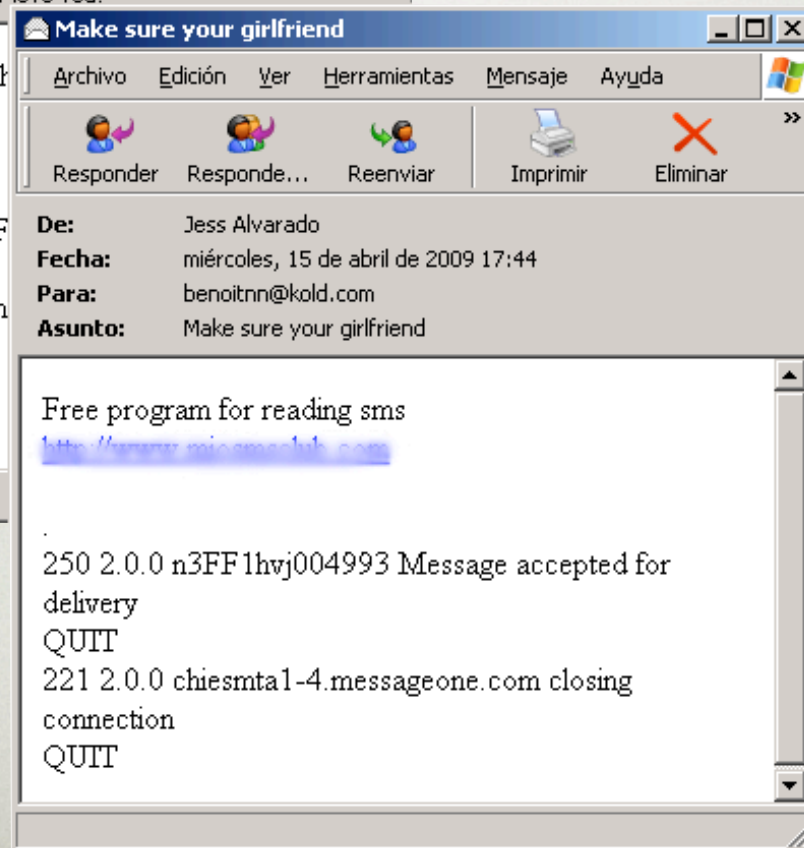


EMAIL PROPAGATION

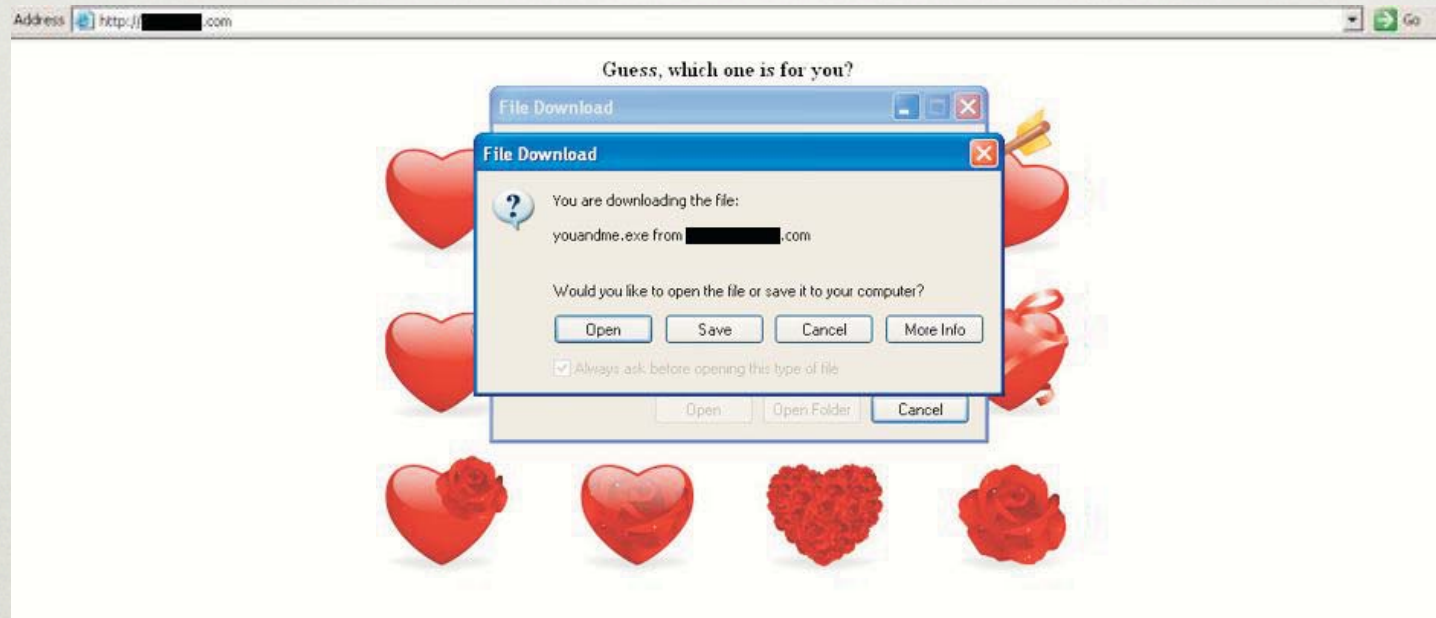


Read other people's SMS with
[http://www.micromark1.com](#)

250 OK 9C/62-06241-E7AF
QUIT
221 serin.channel4.local closing
connection
QUIT

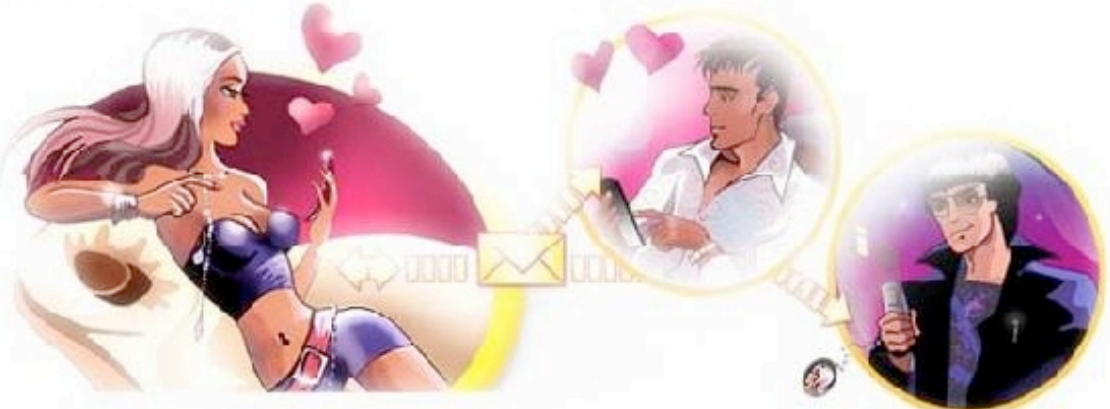


VALENTINE DAY ...



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 be able to read all SMS when you are online. Be aware of everything! This is an extremely new service!

[http://\[Removed\].com/freetrial.exe](http://[Removed].com/freetrial.exe)

Download Free Trial
© SMS Spy. All rights reserved

FAKE CODEC



FAKE ANTIVIRUS

The screenshot shows a window titled "System Security" with a blue header and a yellow sidebar. The sidebar contains icons for "System Scan", "Protection", "Privacy", "Update", and "Settings". The main area displays "System Security : System Scan" with a table of detected threats. Below the table is a "Scan progress" section with a green progress bar, a "Stop" button, and a "Save Report" button. The "Infections" count is 8.

System Security
protect your pc

Registration Update Support

System Scan

Type	Run Type	Name	Details
Spyware	C:/windows/syst...	Spyware.IEMonster.d	Steals passwords from Inter...
Adware	autorun	Zlob.PornAdvertiser.ba	Adware that displays pop-u...
Spyware	autorun	Spyware.IMMonitor	Program that can be used to...
Backdoor	C:/windows/syst...	Win32.Rbot.fm	An IRC controlled backdoor...
Trojan	autorun	Infostealer.Banker.E	Steals sensitive information fr...
Dialer	C:/windows/syst...	Dialer.Xpehbam.biz_dialer	A Dialer that loads pornogra...
Spyware	autorun	Spyware.KnownBadSites	Uses the Windows hosts file t...
Trojan	autorun	Trojan.Tooso	Trojan.Tooso is a trojan whi...

Scan progress

Scanning Stop

Path ...ings\Administrator\SendTo\Desktop (create shortcut).DeskLink

Infections **8**

Save Report Remove

Get full real-time protection with System Security

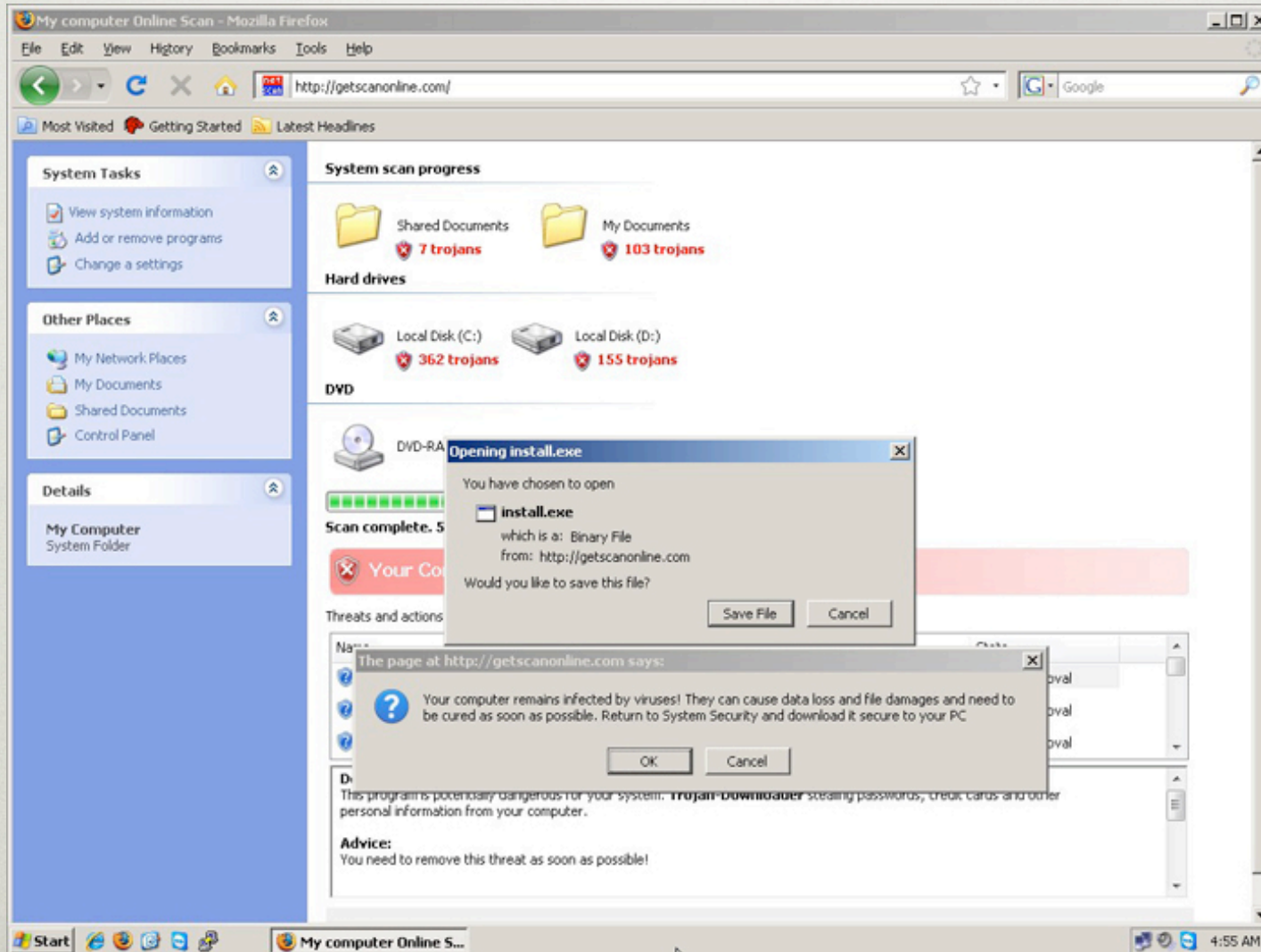
System Security 2009

HIJACK YOU BROWSER



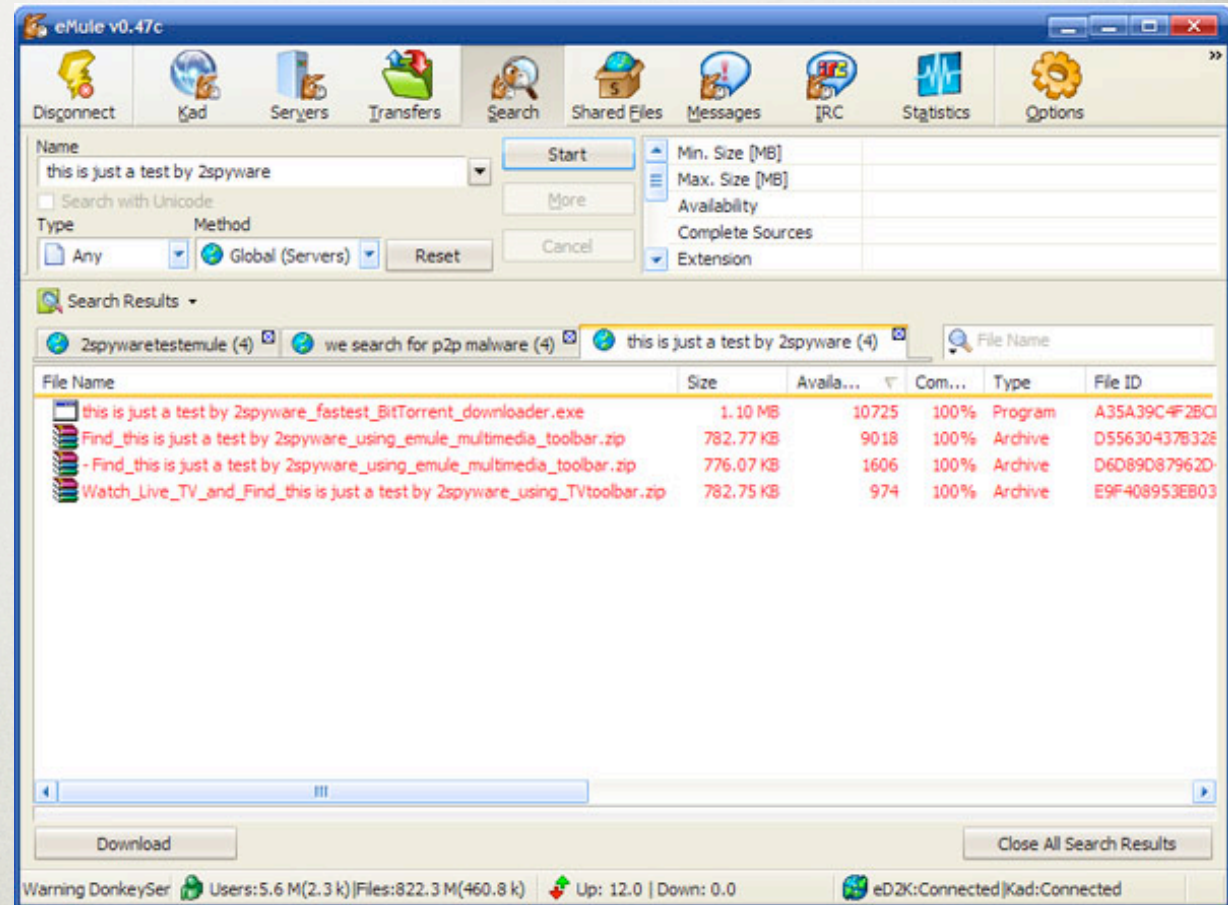
The image shows a screenshot of a Google search interface. At the top left is the Google logo. To its right is a search input field containing the text "Cinderella Full Story In Script". Further right is a "Search" button and two links: "Advanced Search" and "Preferences". Below the search bar is a horizontal line. Underneath this line, on the left, is the word "Web" and on the right, "Results 1 - 10 of about 124,000 for". Below this is a search result. The title is "[Cinderella Full Story In Script](#)". The snippet below the title reads: "Cinderella full story in script But we enjoy fairy tales not because we revel in cinderella s slums are really just less well-kept neighborhoods. full the ...". At the bottom of the snippet, there is a red rectangular box containing the text: "get-new.mee.fgu.name/liouclsuser.html - 8 hours ago - [Similar pages](#)".

FAKE PAGE !



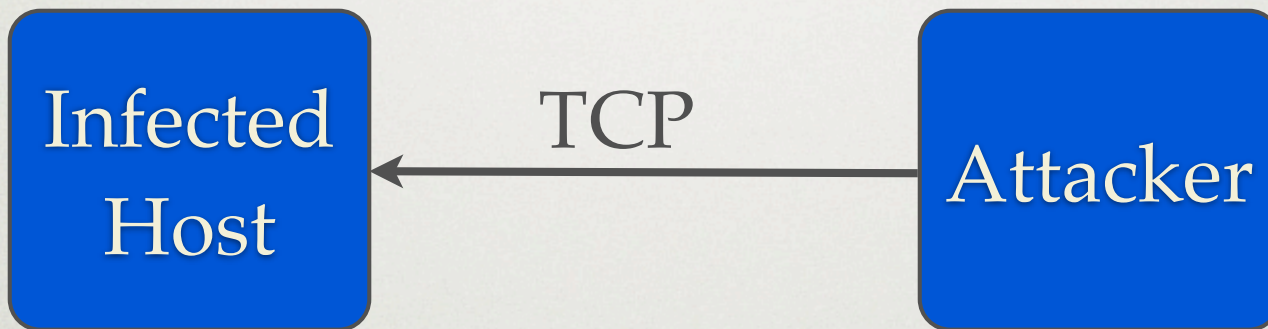
P2P FILES

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

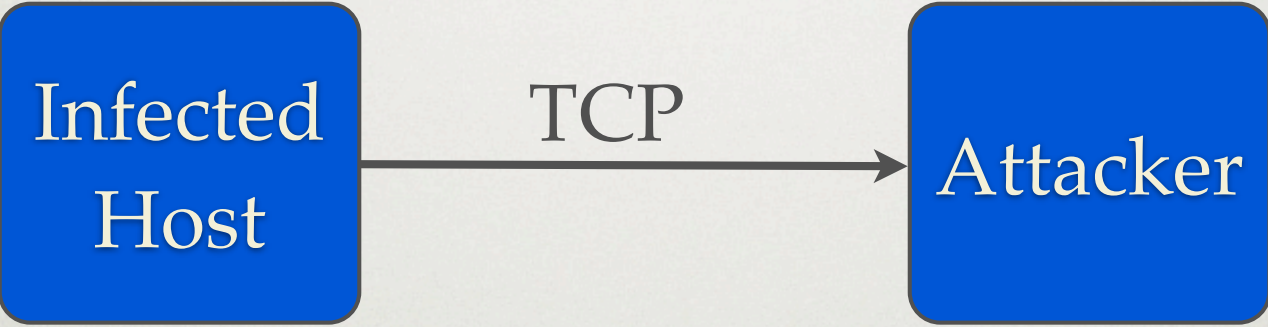


BACKDOOR

BASIC



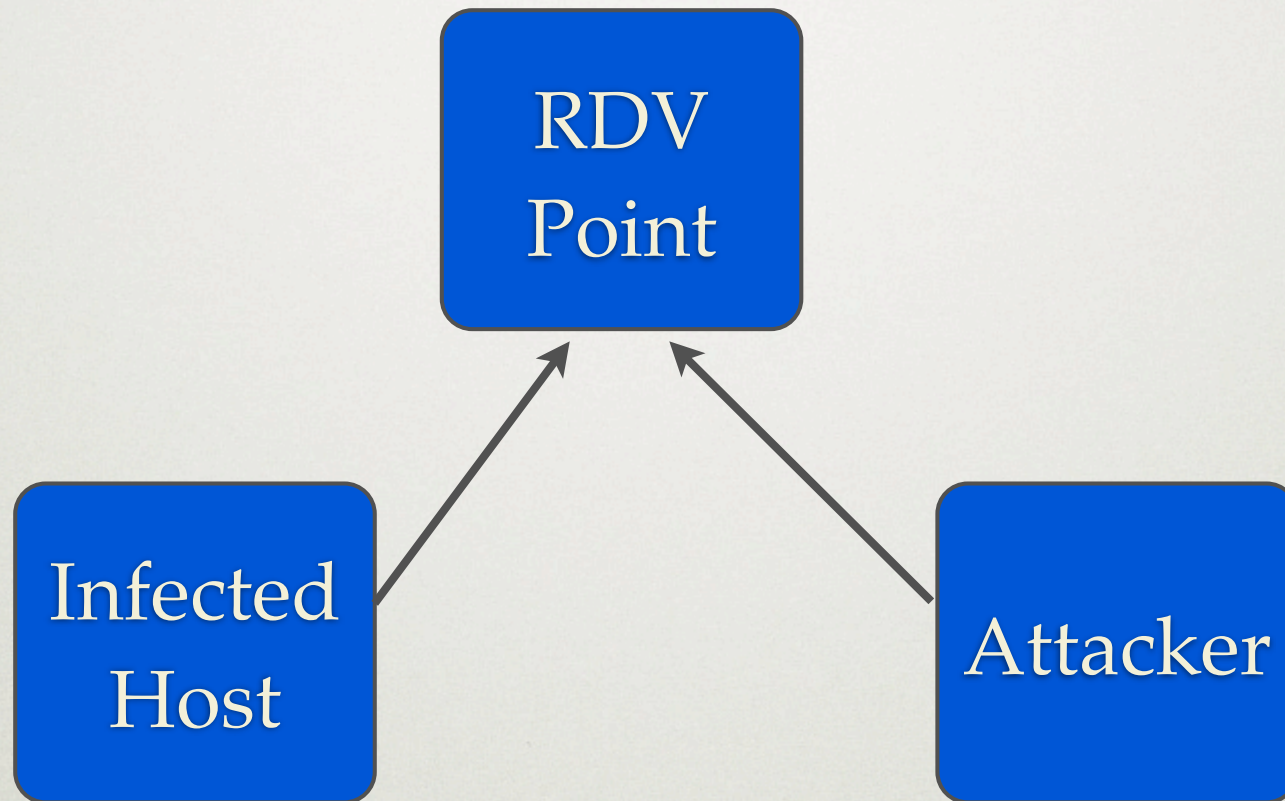
REVERSE



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

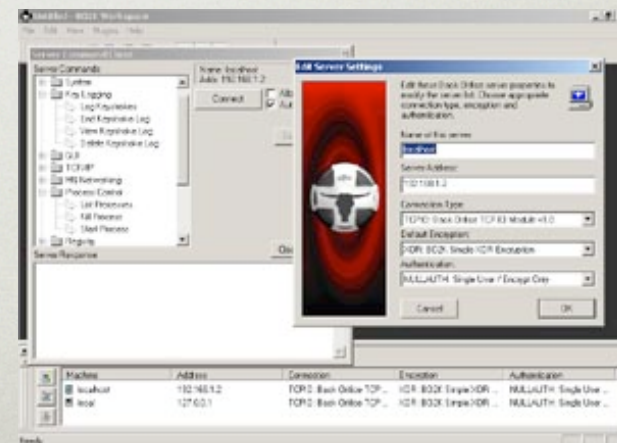
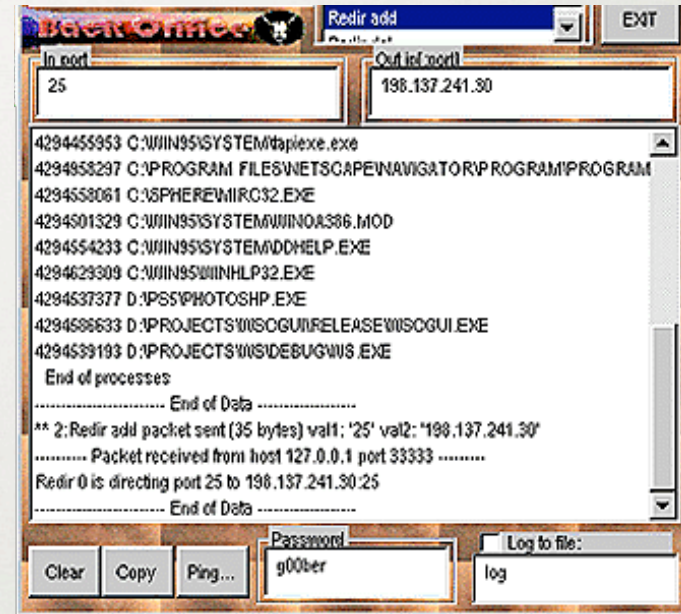


ADWARE



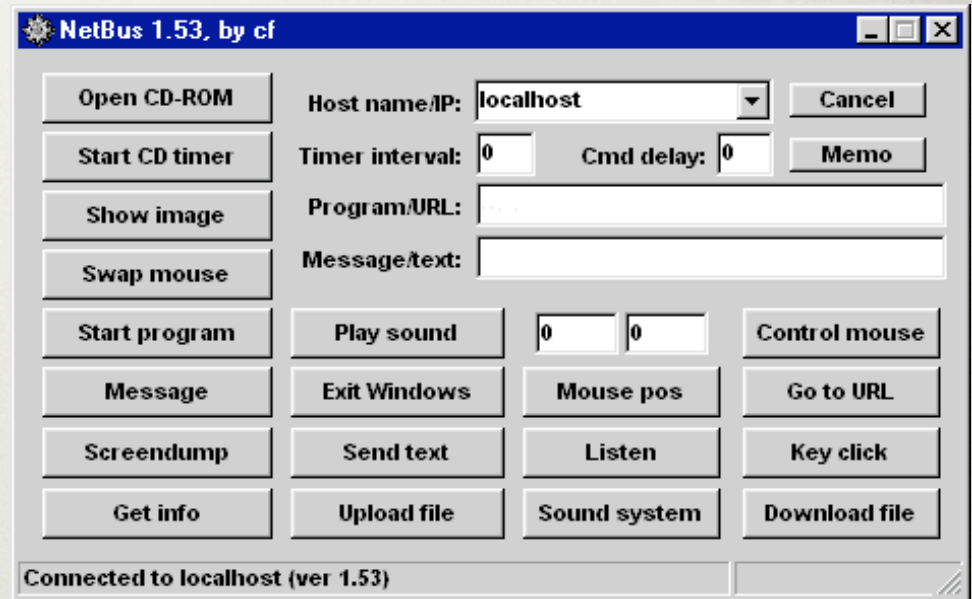
BACKORIFICE

- Defcon 1998
- new version in 2000



NETBUS

- 1998
- Used for “prank”



SYMANTEC PCANYWHERE

The screenshot shows the Symantec pcAnywhere Remote Task Manager interface. The window title is "DFULLER3-VMWARE - Symantec pcAnywhere Remote". The menu bar includes "File", "Edit", "Task", "Actions", and "Help".

The left sidebar contains two main sections:

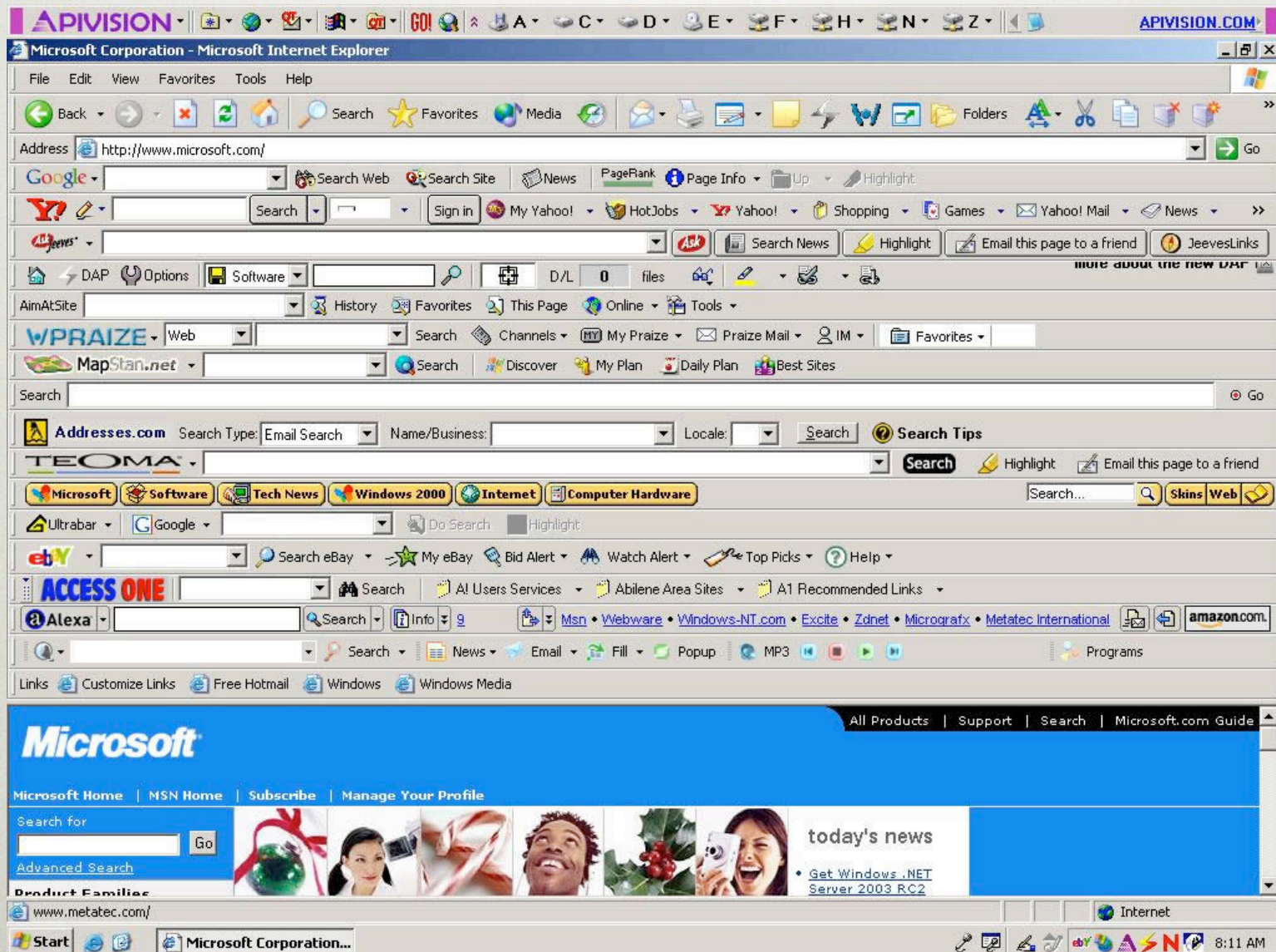
- Session Manager:** Remote Control, Remote Management, File Transfer, Command Queue, Show Chat, End Session.
- Remote Management:** Task Manager (selected), Command Prompt, Services, Edit System Files, Edit Registry, Event Log, Installed Programs, Change Computer State.
- Details:** DFULLER3-VMWARE, 192.168.142.131, TCP/IP, Encryption Level: None, Windows 2000, Connected, Time of Session: 00:00:41, Send: 2 KB, Receive: 8 KB.

The main area displays the **Task Manager** window with the **Processes** tab selected. It shows a list of running processes with the following columns: Image Name, PID, Session ID, User Name, Threads, Memory Usage, and Priority.

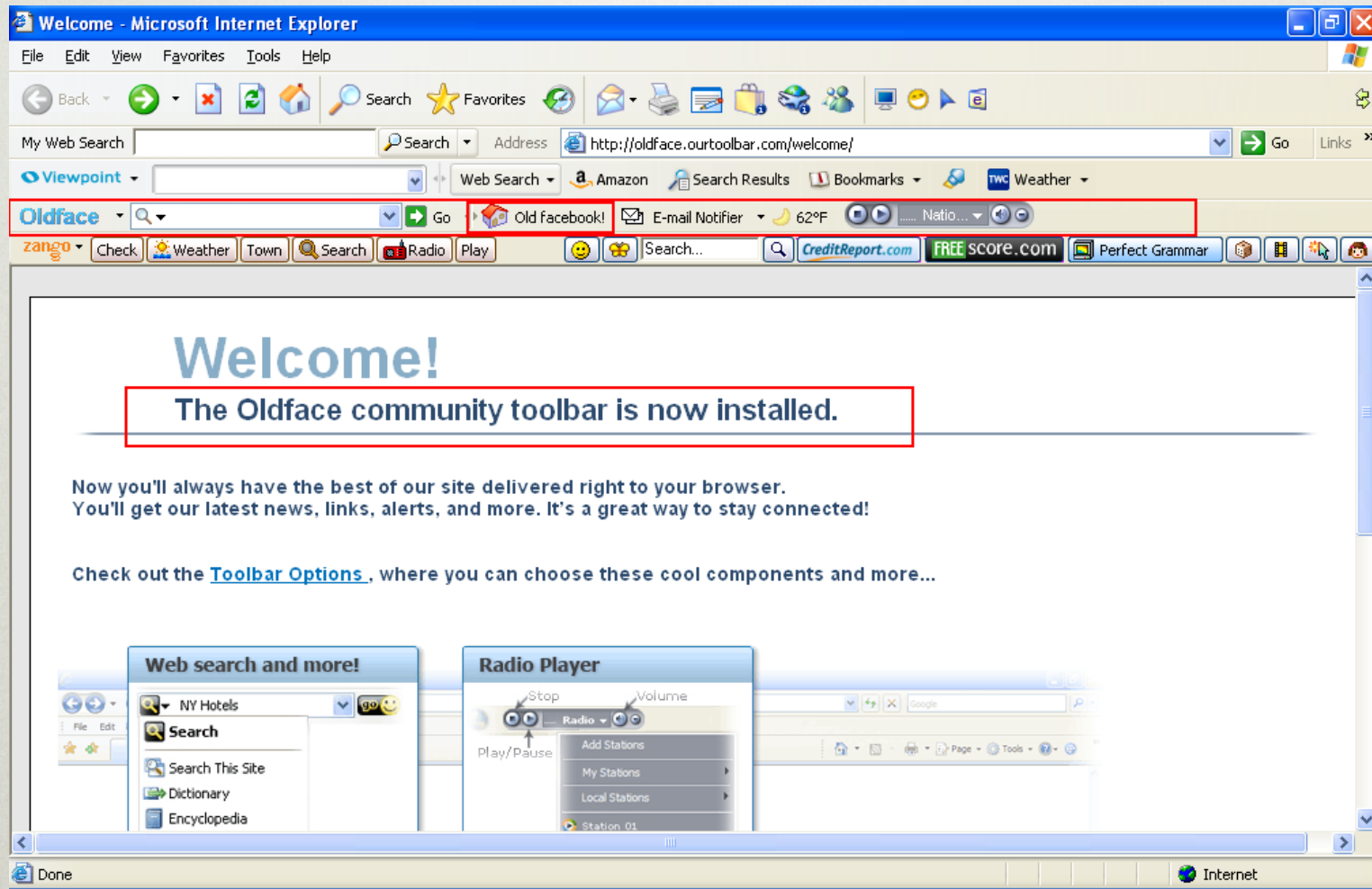
Image Name	PID	Session ID	User Name	Threads	Memory Usage	Priority
[System Process]	0	0	<access denied>	1	0 KB	Low
System	8	0	<access denied>	32	212 KB	Normal
smss.exe	136	0	NT AUTHORITY\SYSTEM	6	396 KB	AboveNormal
winlogon.exe	160	0	NT AUTHORITY\SYSTEM	18	1,956 KB	High
csrss.exe	164	0	<access denied>	8	0 KB	High
services.exe	212	0	NT AUTHORITY\SYSTEM	41	6,148 KB	Normal
lsass.exe	224	0	NT AUTHORITY\SYSTEM	19	5,616 KB	Normal
svchost.exe	404	0	NT AUTHORITY\SYSTEM	9	4,152 KB	Normal
spoolsv.exe	448	0	NT AUTHORITY\SYSTEM	11	4,508 KB	Normal
awhost32.exe	476	0	NT AUTHORITY\SYSTEM	19	8,012 KB	Normal
svchost.exe	508	0	NT AUTHORITY\SYSTEM	7	7,512 KB	Normal
regsvc.exe	548	0	NT AUTHORITY\SYSTEM	2	1,152 KB	Normal
MSTask.exe	564	0	NT AUTHORITY\SYSTEM	7	3,580 KB	Normal
VMwareService.e	596	0	NT AUTHORITY\SYSTEM	2	1,412 KB	High
WinMgmt.exe	636	0	NT AUTHORITY\SYSTEM	4	216 KB	Normal
svchost.exe	688	0	NT AUTHORITY\SYSTEM	6	5,552 KB	Normal
Explorer.EXE	816	0	DFULLER3-VMWARE\Ad...	12	7,368 KB	Normal
VMwareTray.exe	896	0	DFULLER3-VMWARE\Ad...	1	1,720 KB	Normal

At the bottom right of the Task Manager window, there are two buttons: "New Task" and "End Process".

BROWSER TOOLBAR ...

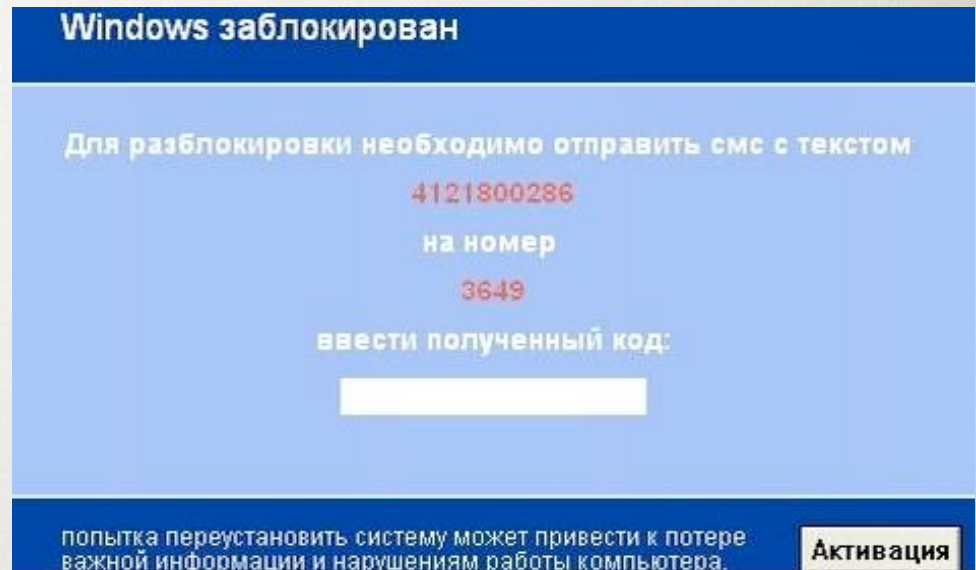


TOOLBAR AGAIN



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*

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 $P(V) = \text{true} \rightarrow \text{halt}$
 - if $P(V) = \text{false} \rightarrow \text{spread}$

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 be 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

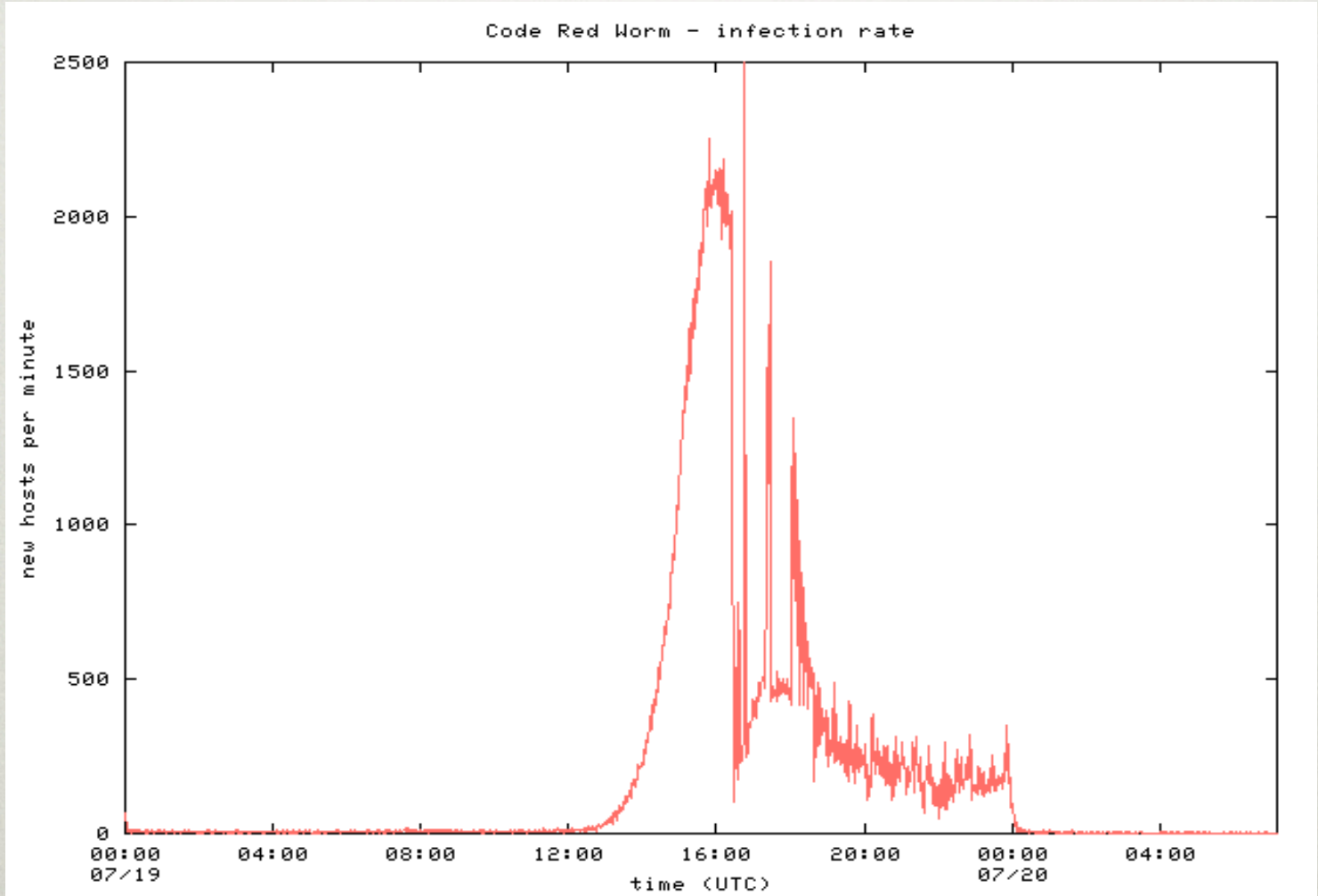
◆ Initial release of July 13

- 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 \Rightarrow *linear growth*

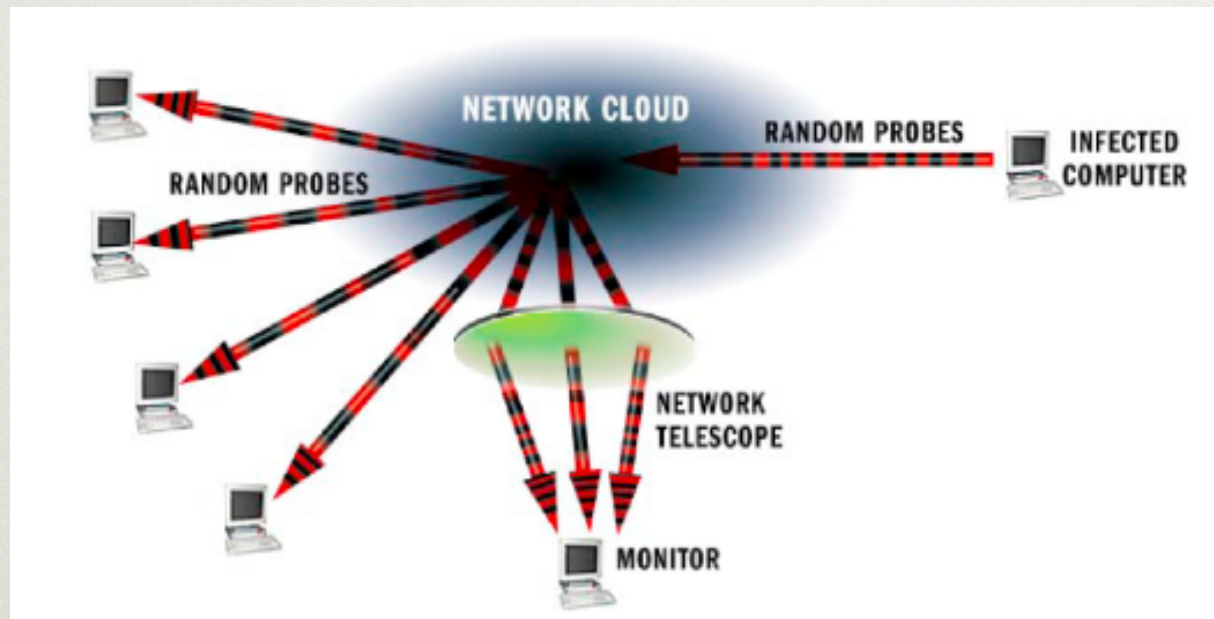
◆ Revision released July 19, 2001.

- White House responds to threat of flooding attack by changing the address of *www.whitehouse.gov*
- Causes Code Red to die for date \geq 20th of the month.
- But: this time random number generator correctly seeded

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:⁸¹1/256.

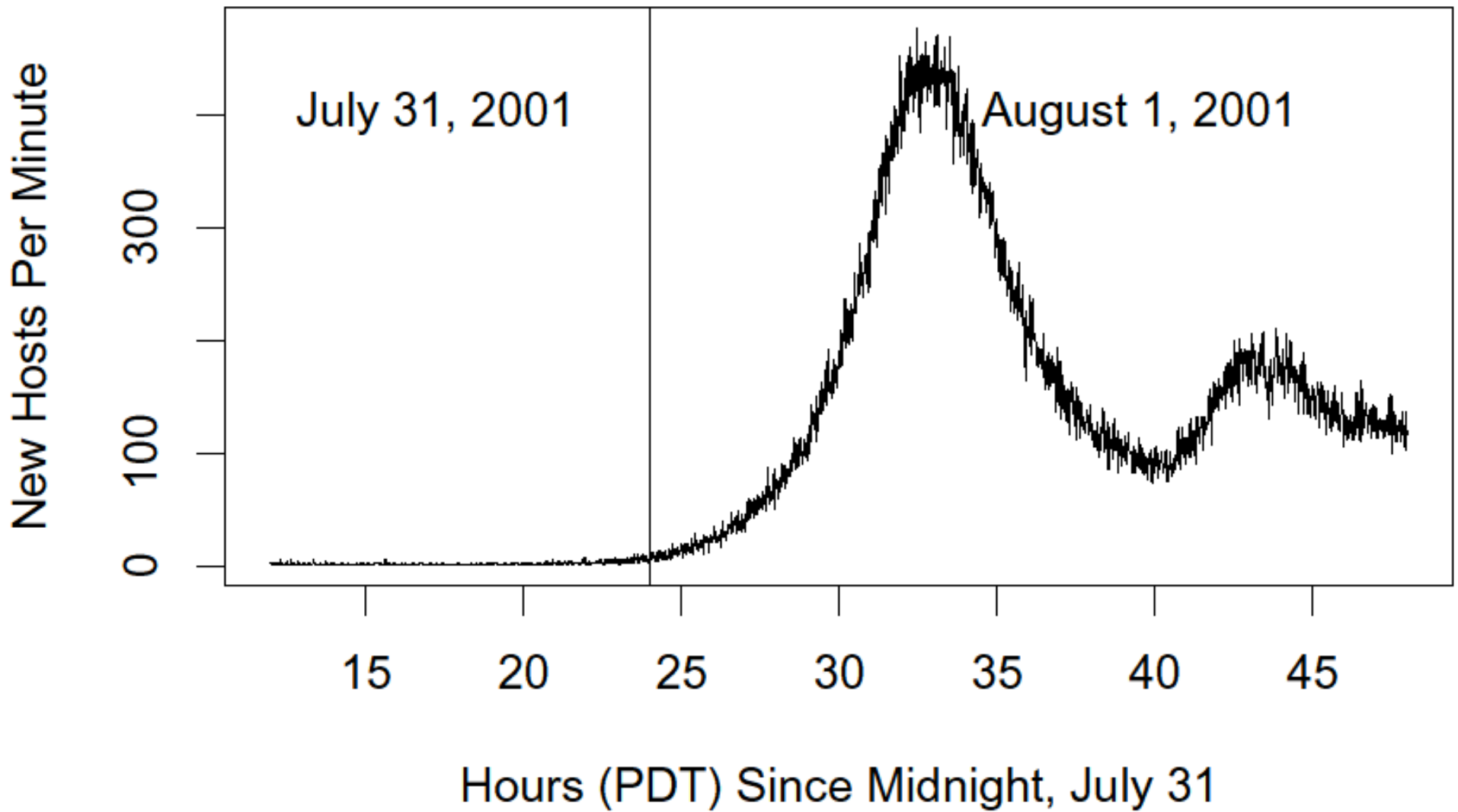
Spread of Code Red

- ◆ 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 20th), worm dies ...
... except for hosts with inaccurate clocks!
- ◆ It just takes one of these to restart the worm on August 1st ...

Slides: Vern
Paxson

Return of Code Red Worm



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

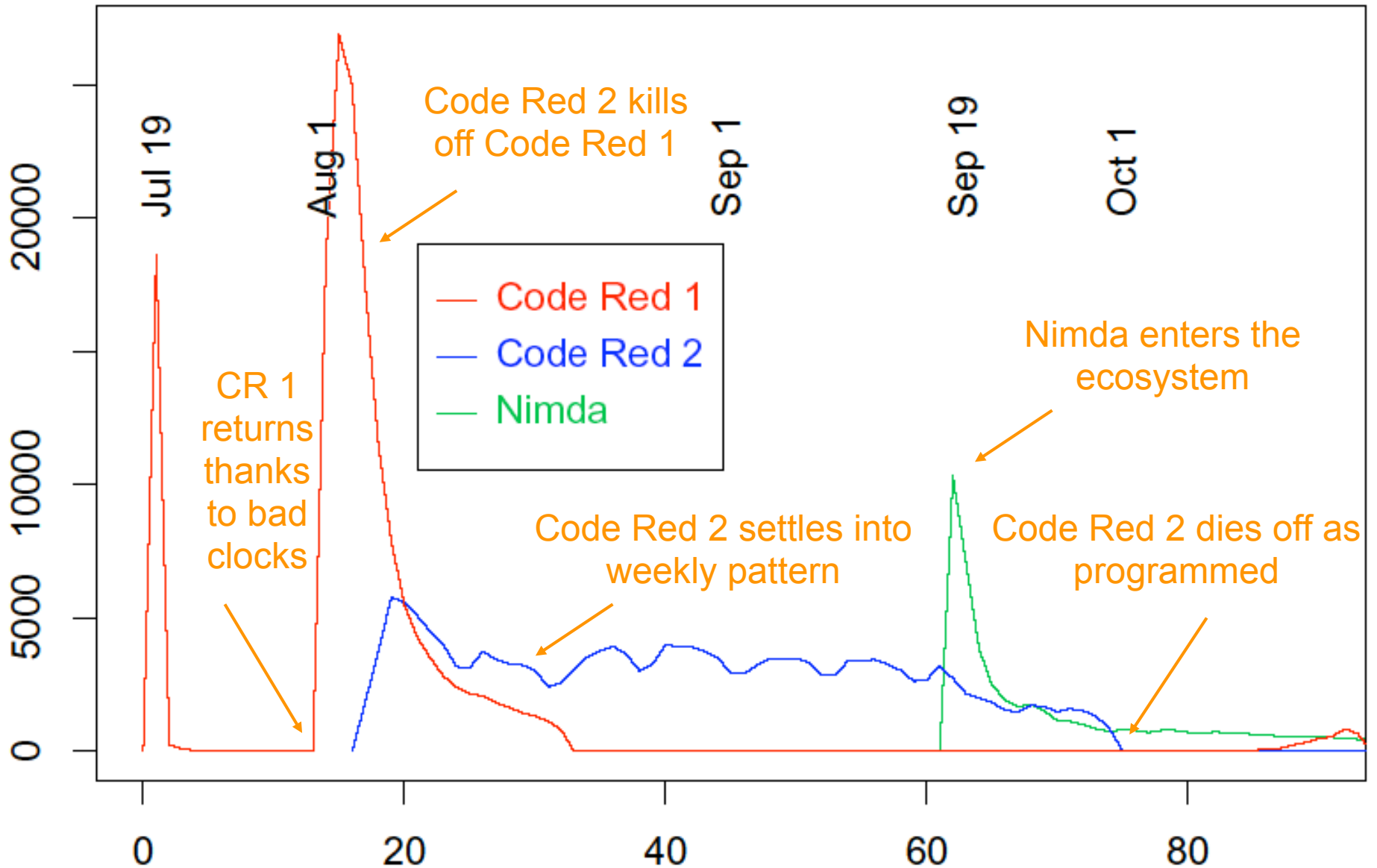
Paxson

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

Distinct Remote Hosts Attacking LBNL



Days Since July 18, 2001

Slides: Vern Paxson

HOW DO WORMS PROPAGATE?

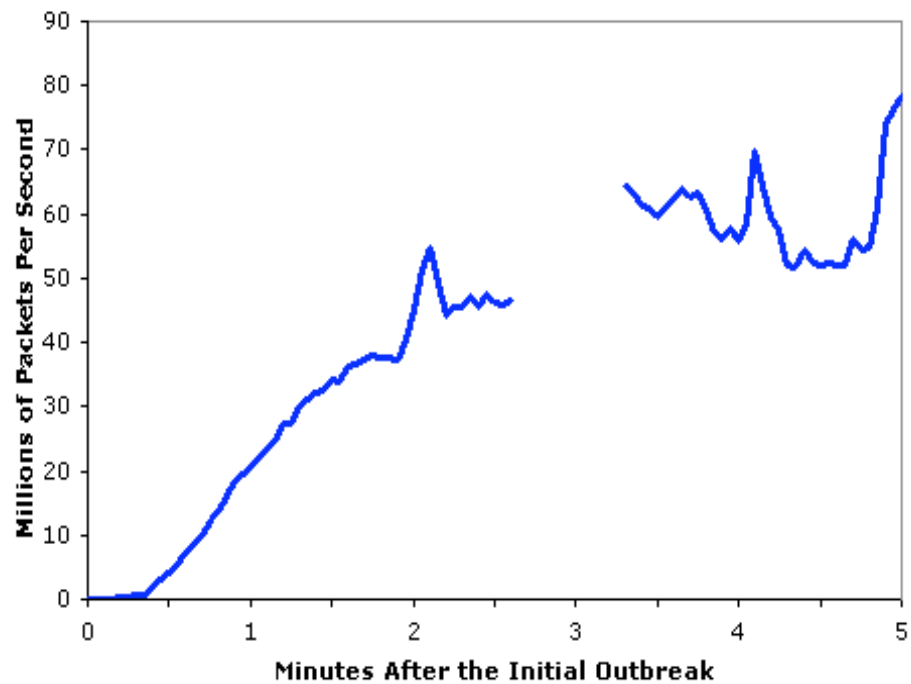
- ◆ **Scanning** worms : Worm chooses “random” address
- ◆ **Coordinated** scanning : Different worm instances scan different addresses
- ◆ **Flash** worms
 - Assemble tree of vulnerable hosts in advance, propagate along tree
 - ◆ Not observed in the wild, yet
 - ◆ Potential for 10⁶ hosts in < 2 sec ! [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

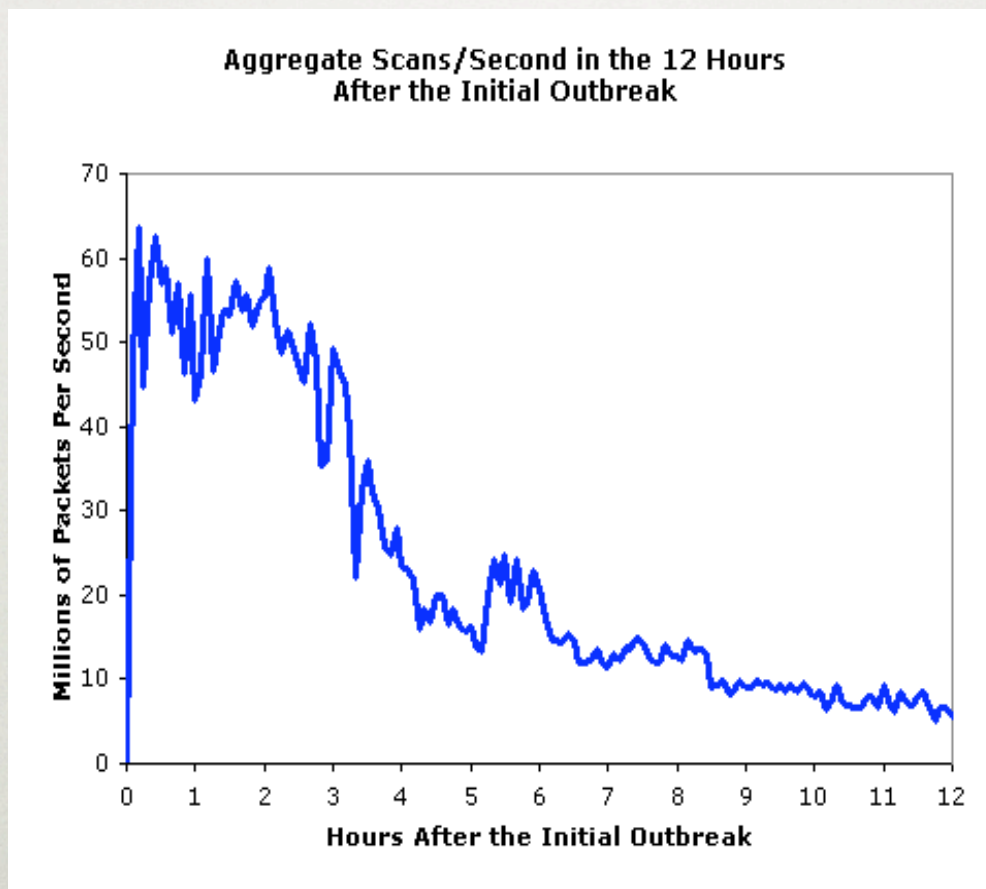
- 01 / 25 / 2003
- Vulnerability disclosed : 25 june 2002
- Better scanning algorithm
- UDP Single packet : 380bytes

SLAMMER PROPAGATION

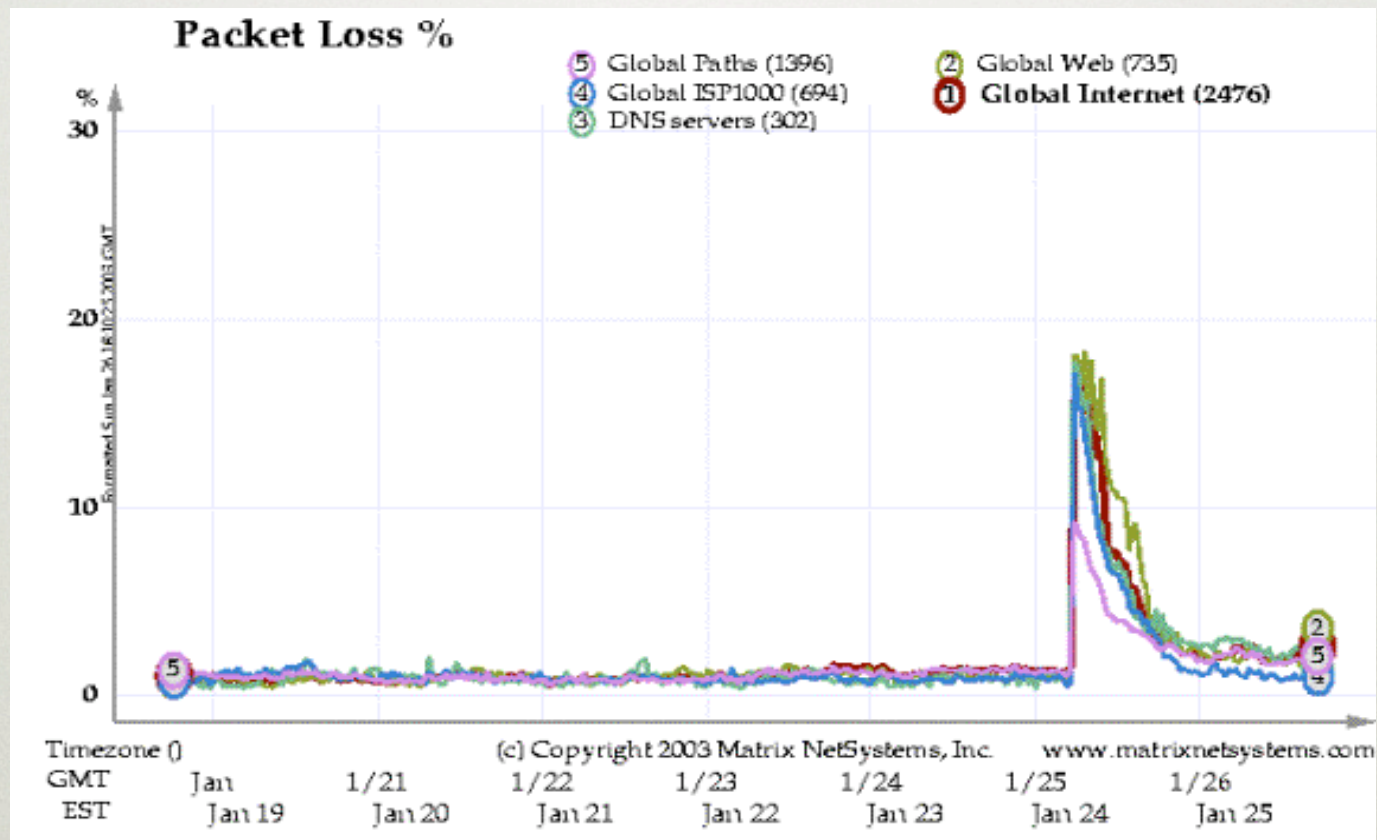
Aggregate Scans/Second in the first 5 minutes based on Incoming Connections To the WAIL Tarpit



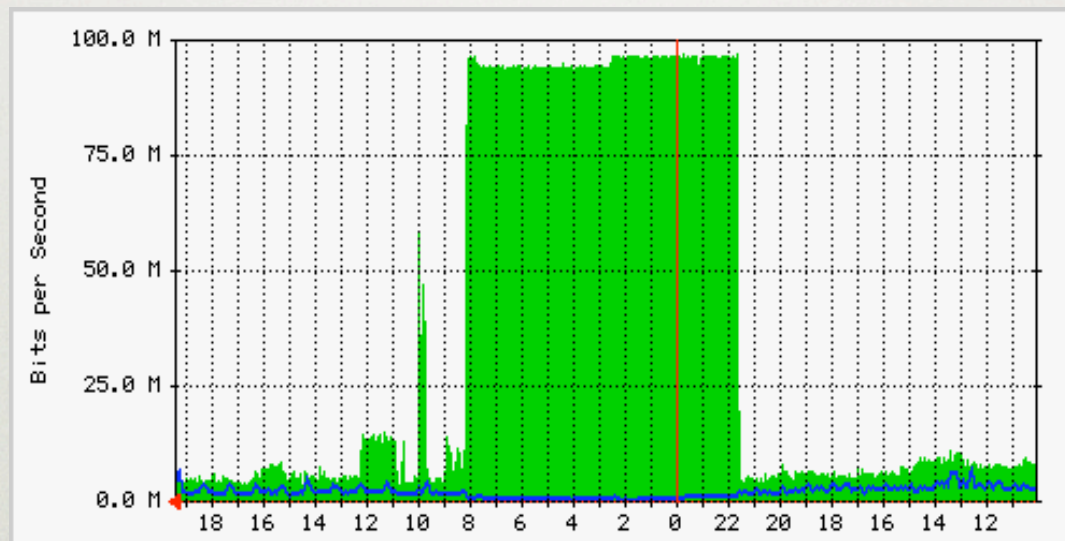
NUMBER OF SCAN/SEC



PACKET LOSS



A SERVER VIEW



CONSEQUENCES

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

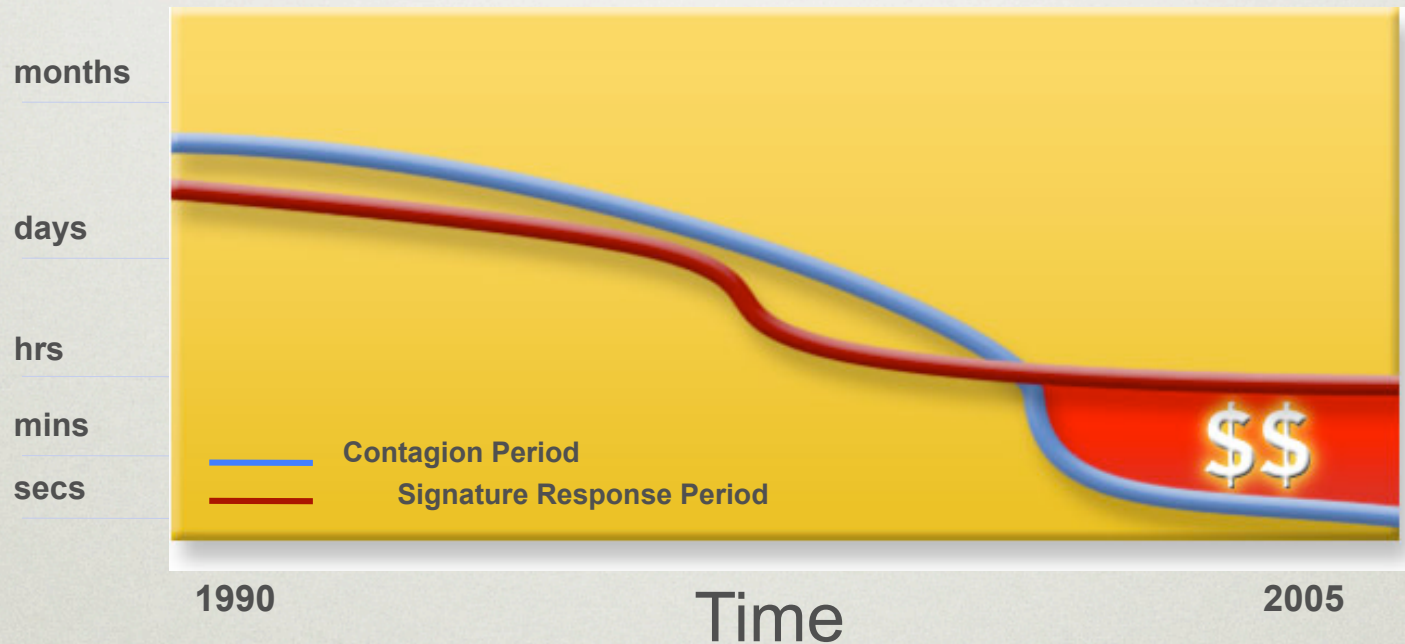
Worm Detection and Defense

- ◆ Detect via *honeypots*: collections of “honeypots” fed by a network telescope.
 - Any outbound connection from honeypot = 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

Contagion Period



Signature Response Period

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

```
PACKET HEADER
SRC: 11.12.13.14.3920 DST: 132.239.13.24.5000 PROT: TCP
PACKET PAYLOAD (CONTENT)
00F0 90 90 90 .....
0100 90 90 90 .....M?.w
0110 90 90 90 .....cd.....
0120 90 90 90 90 90 90 90 90 90 90 90 90 90 90 90 .....
0130 90 90 90 90 90 90 90 90 EB 10 5A 4A 33 C9 66 B9 .....ZJ3.f.
0140 66 01 80 34 0A 99 E2 FA EB 05 E8 EB FF FF FF 70 f..4.....p
...
```

**Kibvu.B signature captured by
Earlybird on May 14th, 2004**

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

**OBSERVATION:
HIGH-PREVALENCE STRINGS ARE RARE**

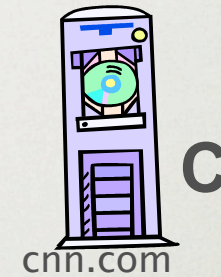
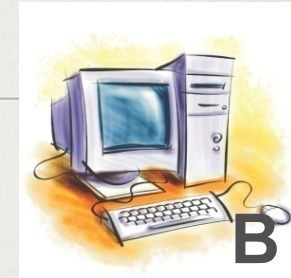
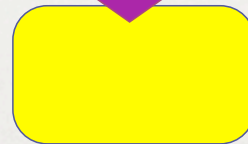
Only 0.6% of the 40 byte substrings repeat more than 3 times in a minute

(Stefan Savage, UCSD *)

THE BASIC ALGORITHM

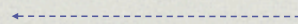


Detector in network



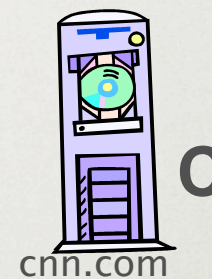
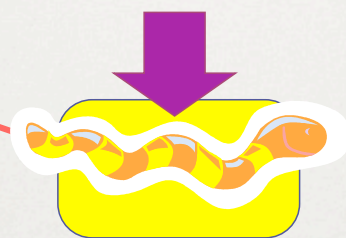
Prevalence Table

Address Dispersion Table
Sources Destinations






Detector in network

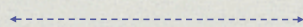


Prevalence Table

Address Dispersion Table
Sources Destinations

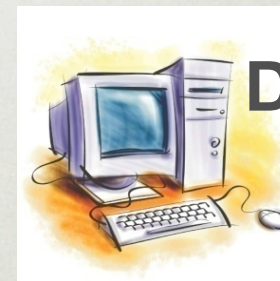
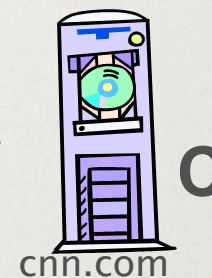
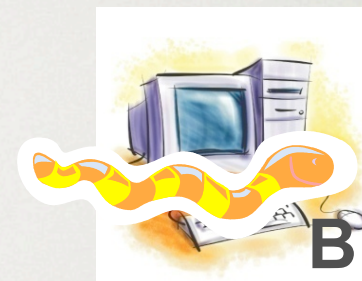
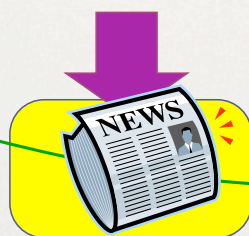
	1
---	----------

1 (A)	1 (B)
--------------	--------------







Detector in network



Prevalence Table

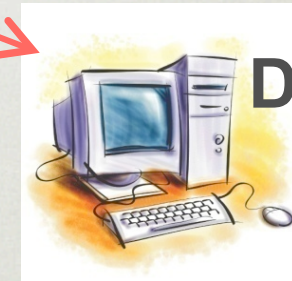
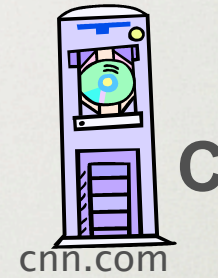
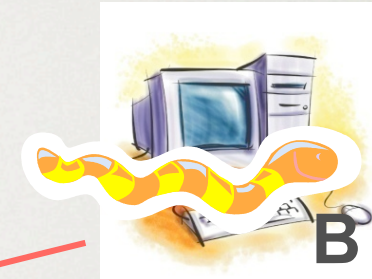
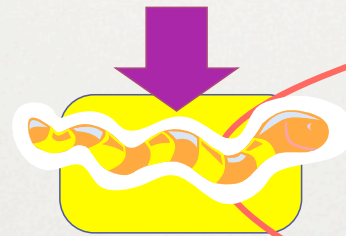
	1
	1

Address Dispersion Table
Sources Destinations



1 (A)	1 (B)
1 (C)	1 (A)



Detector in network

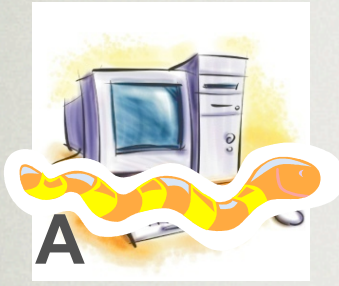


Prevalence Table

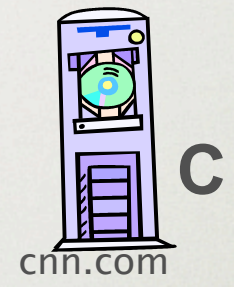
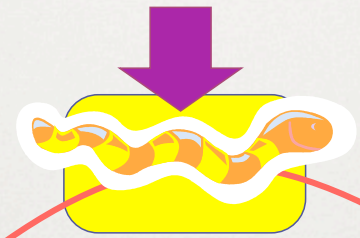
	2
	1

Address Dispersion Table
Sources Destinations

2 (A,B)	2 (B,D)
1 (C)	1 (A)





Detector in network



Prevalence Table

Address Dispersion Table
Sources Destinations

	3	\longleftrightarrow	3	3
	1		(A, B, D) 1 (C)	(B, D, E) 1 (A)

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