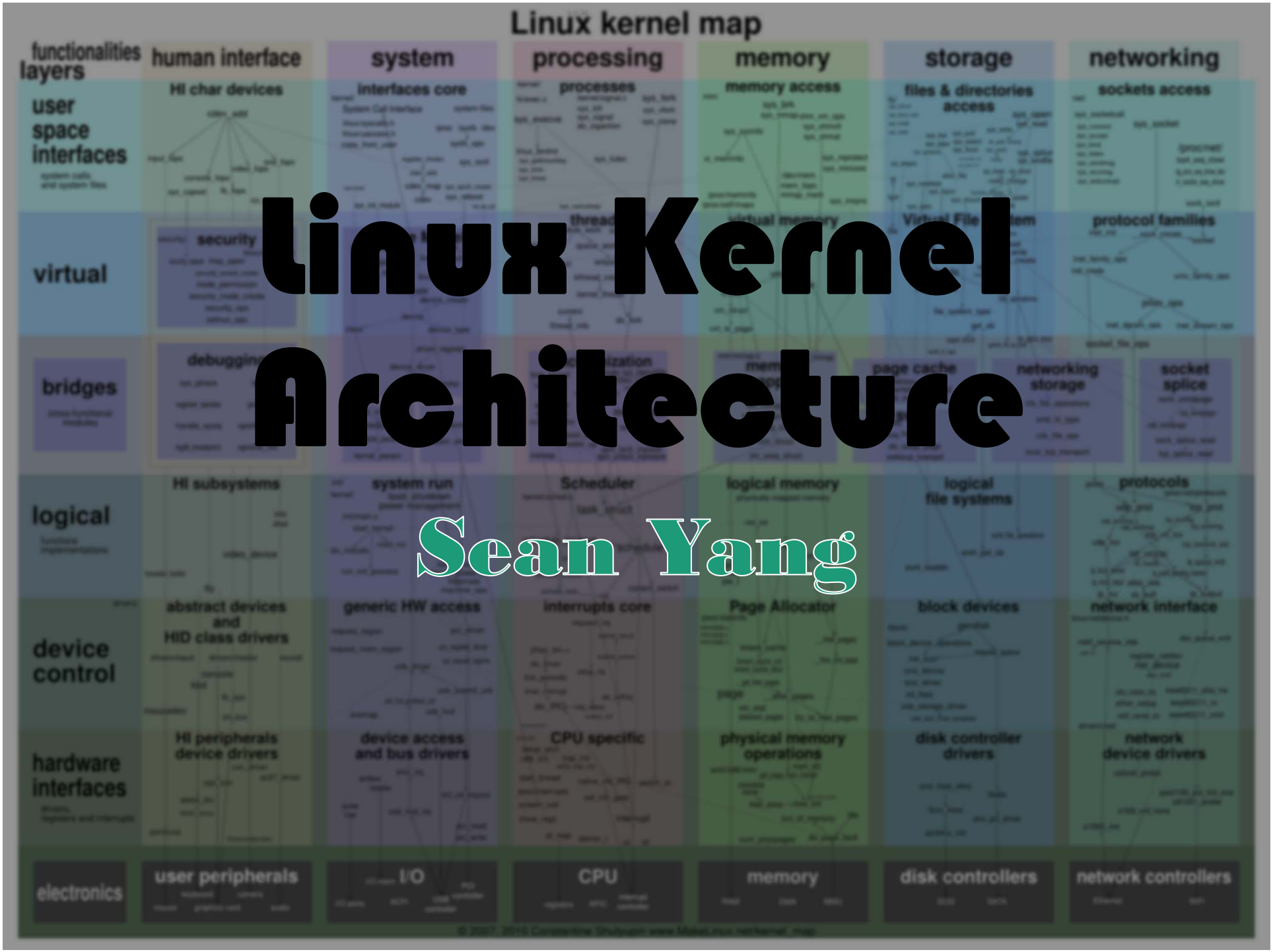
Linux Kernel Architecture

Sean Yang



What are benefits of knowing Linux kernel?

CVE Details

The ultimate security vulnerability datasource

Log In Register Reset Password Activate Account

Home

Browse :

Vendors

Products

Vulnerabilities By Date

Vulnerabilities By Type

Reports :

CVSS Score Report

CVSS Score Distribution

Search :

Vendor Search

Product Search

Version Search

Vulnerability Search

By Microsoft References

Linux » Linux Kernel : Security Vulnerabilities (CVSS Score between 7 and 7

CVSS Scores Greater Than: 0 1 2 3 4 5 6 7 8 9

Sort Results By : CVE Number Descending CVE Number Ascending CVSS Score Number

Total number of vulnerabilities : **268** Page : 1 (This Page) 2 3 4 5

Copy Results Download Results Select Table

#	CVE ID	CWE ID	# of Exploits	Vulnerability Type(s)	Publish Date	Update Date	Score
1	CVE-2015-5364	399		DoS	2015-08-31	2015-08-31	7.8
	The (1) udp_recvmsg and (2) udpv6_recvmsg functions in the Linux kernel before 4.0.6 do not prop service (system hang) via incorrect checksums within a UDP packet flood.						
2	CVE-2015-5157	264		+Priv	2015-08-31	2015-08-31	7.2
	arch/x86/entry/entry_64.S in the Linux kernel before 4.1.6 on the x86_64 platform mishandles IRET						
					2015-08-31	2015-09-01	7.2
	drivers/vhost/scsi.c in the Linux kernel bef HOST_SCSI_SET_ENDPOINT ioctl call. NOTI						
					2015-06-07	2015-06-08	7.8
	zwpan/ozusbsvc1.c in the OZWPAN driver i ket.						
					2015-08-31	2015-08-31	7.2
	6 on the x86_64 platform improperly relies						
					2015-08-08	2015-08-25	7.2
	s/pipe.c in the Linux kernel before 3.16 do s to cause a denial of service (system cras						

What can you do?

- A User

What can you do?

- A technician

What can you do?

- An administrator

What can you do?

- A system programmer

Also you can make the system

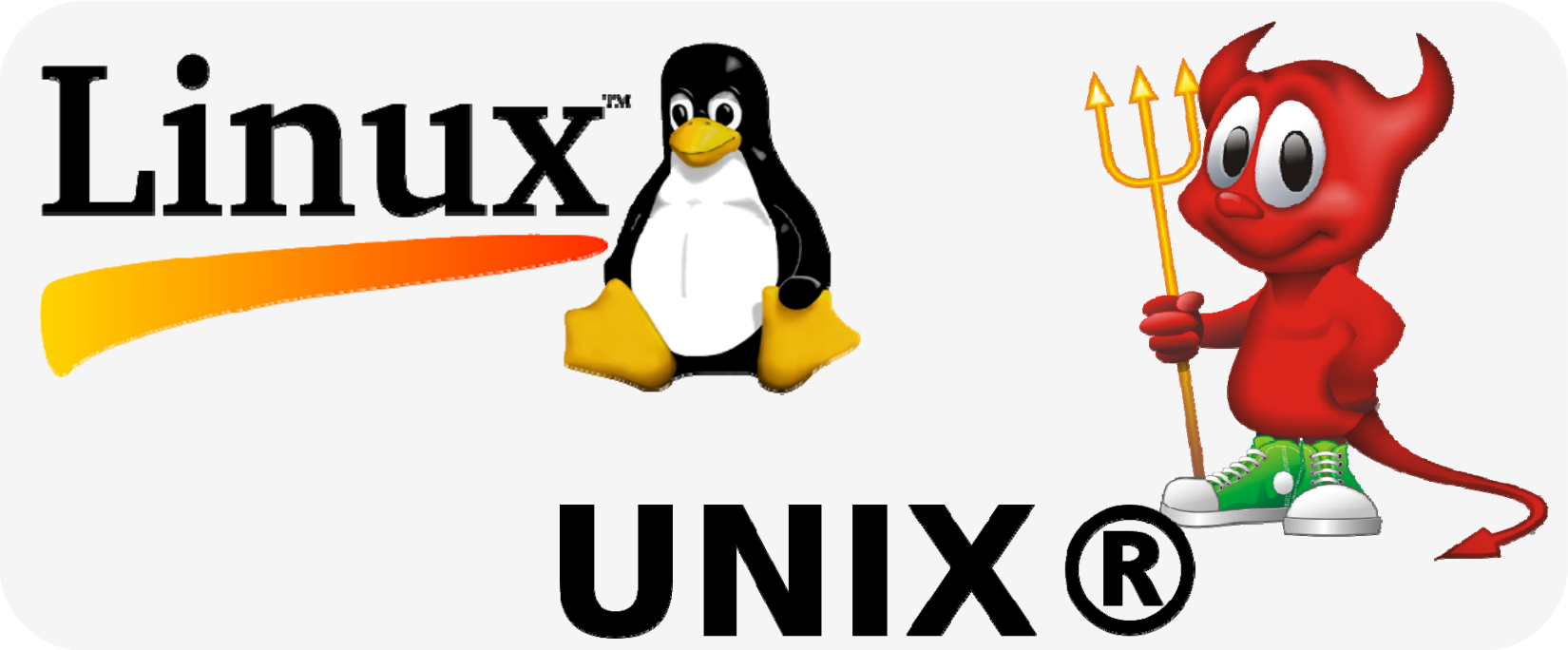
Start your engine!

UNI

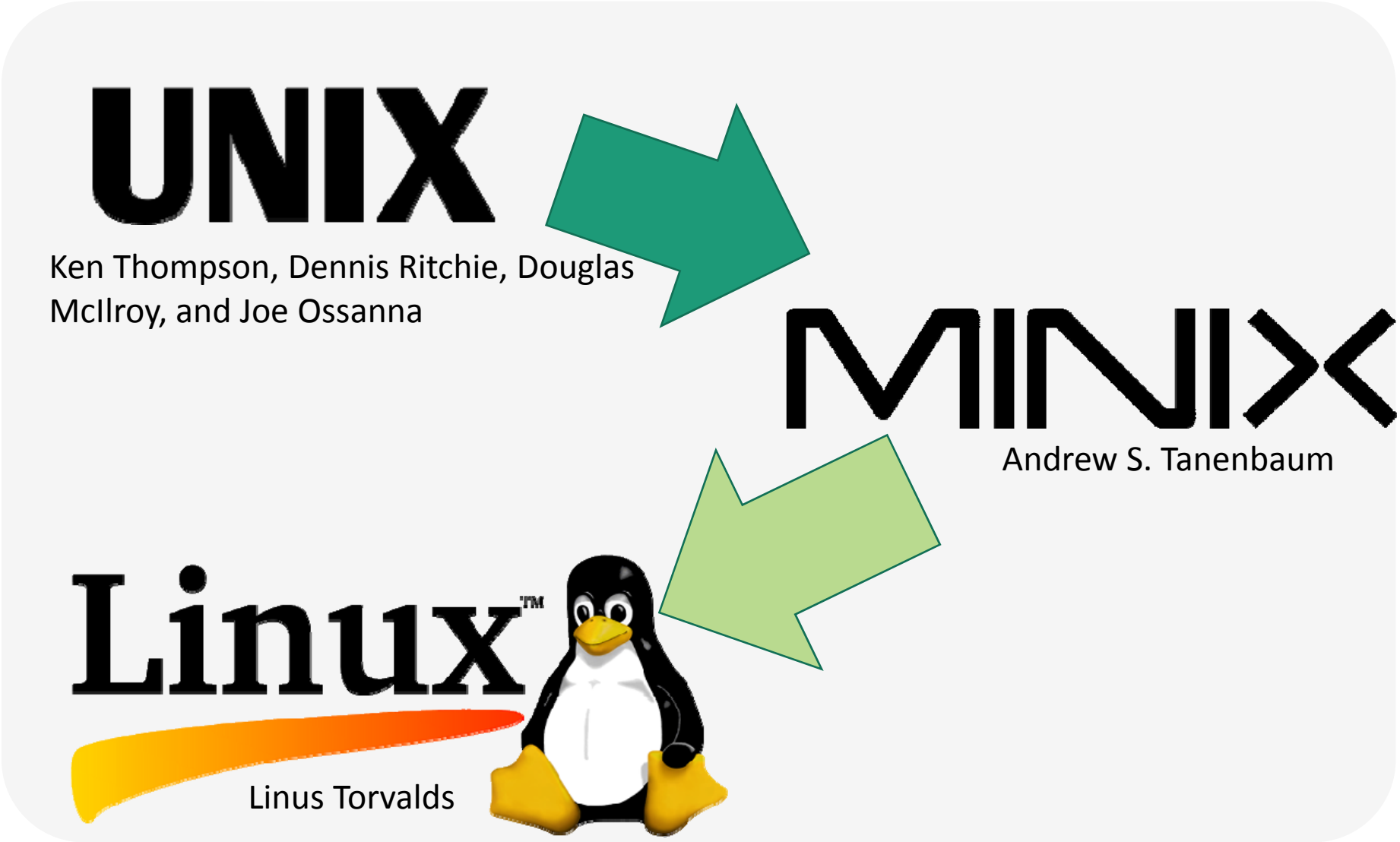
00011110 00011110 00011110 00011110 000111

Why talking about UNIX?

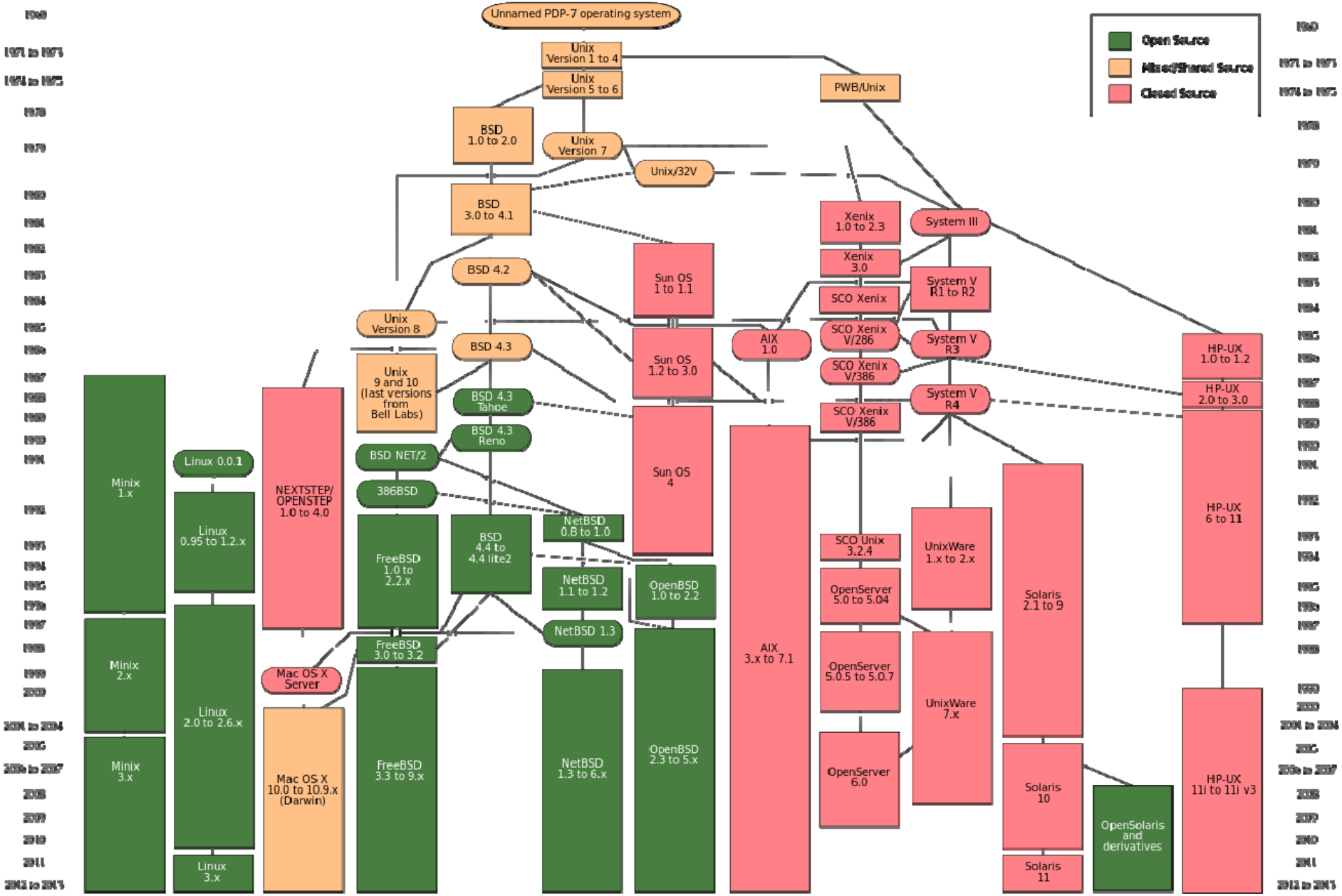
- Linux* is a member of the large family of Unix-like operating systems

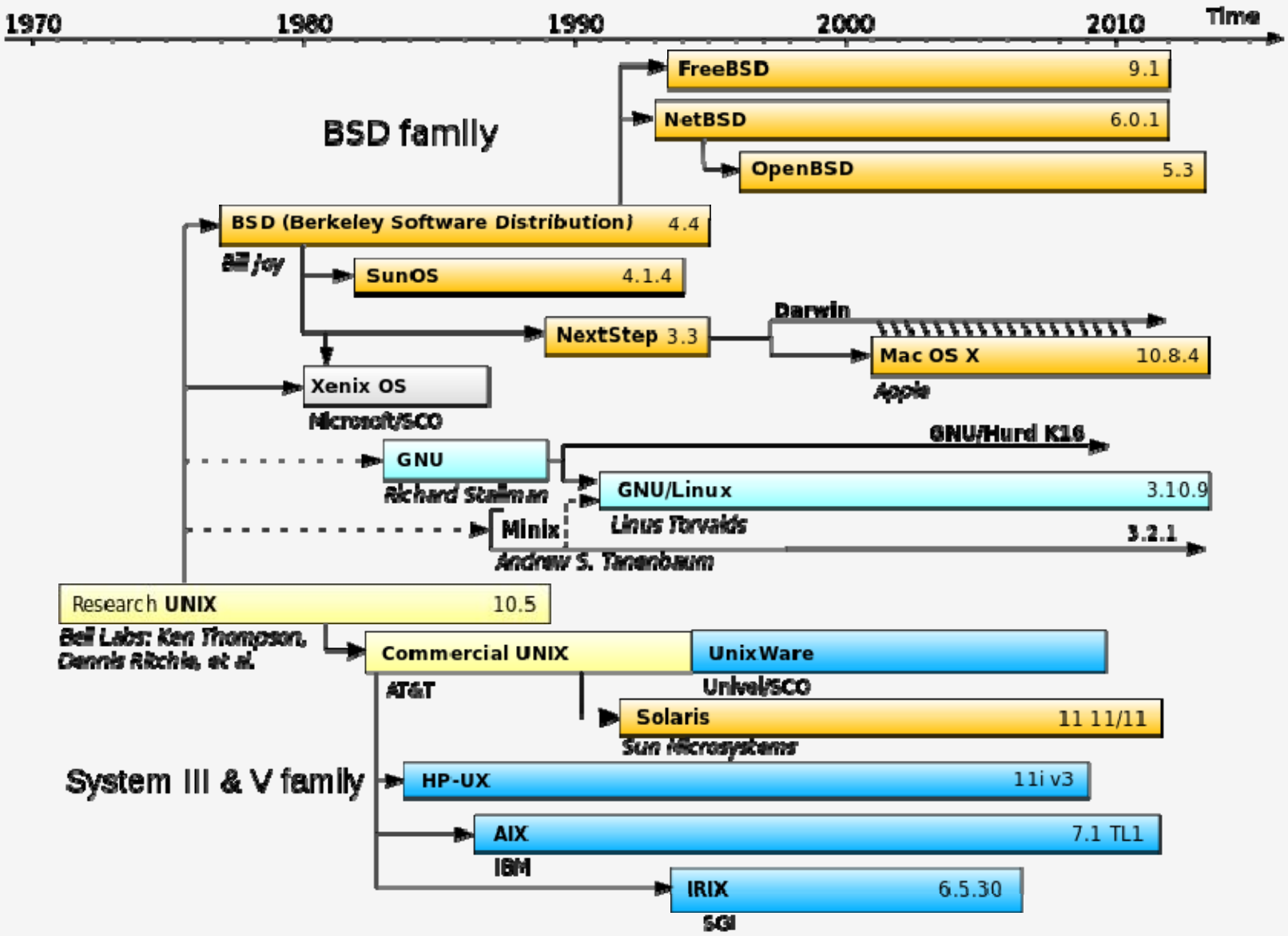


Relationship between UNIX and Linux



Unix Like OSes




1970
1980
1990
2000
2010
Time
BSD family
FreeBSD 9.1
NetBSD 6.0.1
OpenBSD 5.3
BSD (Berkeley Software Distribution) 4.4
Bill Joy
SunOS 4.1.4
NextStep 3.3
Darwin
Mac OS X 10.8.4
Apple
Xenix OS
Microsoft/SCO
GNU
Richard Stallman
GNU/Hurd K16
GNU/Linux 3.10.9
Linus Torvalds
Minix 3.2.1
Andrew S. Tanenbaum
Research UNIX 10.5
Bell Labs: Ken Thompson, Dennis Ritchie, et al.
Commercial UNIX
AT&T
UnixWare
Univel/SCO
Solaris 11 11/11
Sun Microsystems
System III & V family
HP-UX 11i v3
AIX 7.1 TL1
IBM
IRIX 6.5.30
SGI

What's common on the UNIX like OSes? – Similar philosophy

"Unix is simple and coherent, but it takes a genius (or at any rate a programmer) to understand and appreciate the simplicity."

— Dennis Ritchie

"Note from the authors: Yes, we have lost our minds. Be forewarned: You will lose yours too."

— Benny Goodheart & James Cox

What's common on the UNIX like OSes? – Similar philosophy

- Characterized by a modular design that is sometimes called the "Unix philosophy":
  - the operating system provides a set of simple tools that each perform a limited, well-defined function
  - with a unified filesystem as the main means of communication
  - and a shell scripting and command language to combine the tools to perform complex workflows.

What are differences?

UNIX

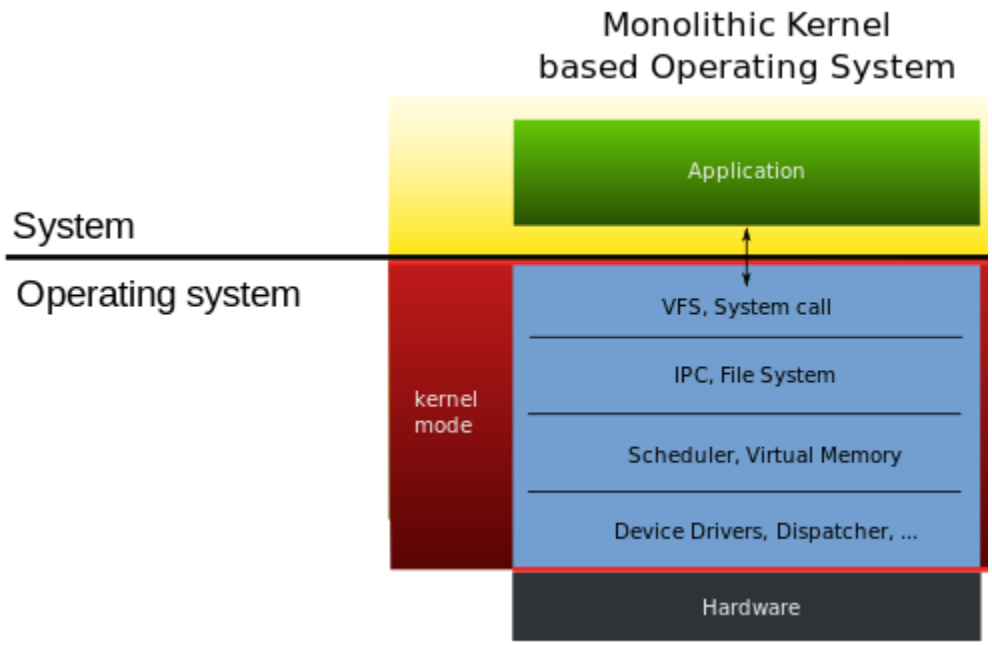
Monolithic Kernel

MINIX

Microkernel

Linux™

Monolithic Kernel (modular)



Examples: UNIX/Linux

Pro: Better performance

Con: Large and complex kernel

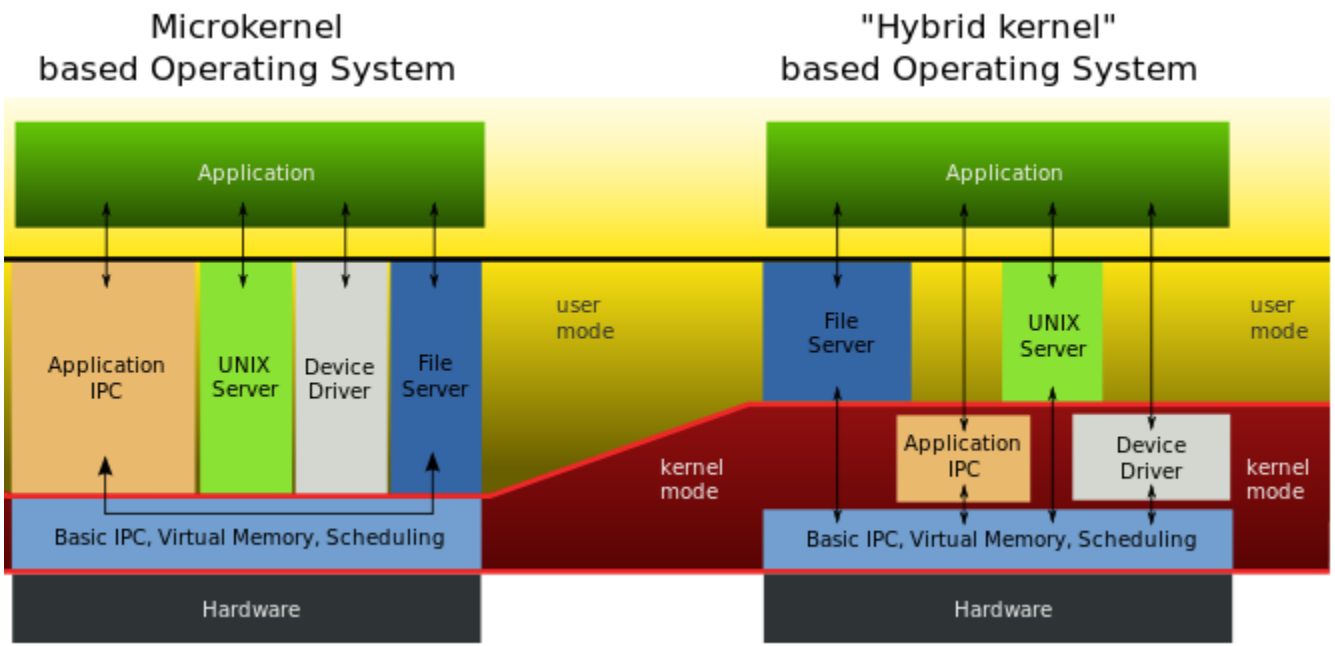
Example: Windows

Example: MINIX

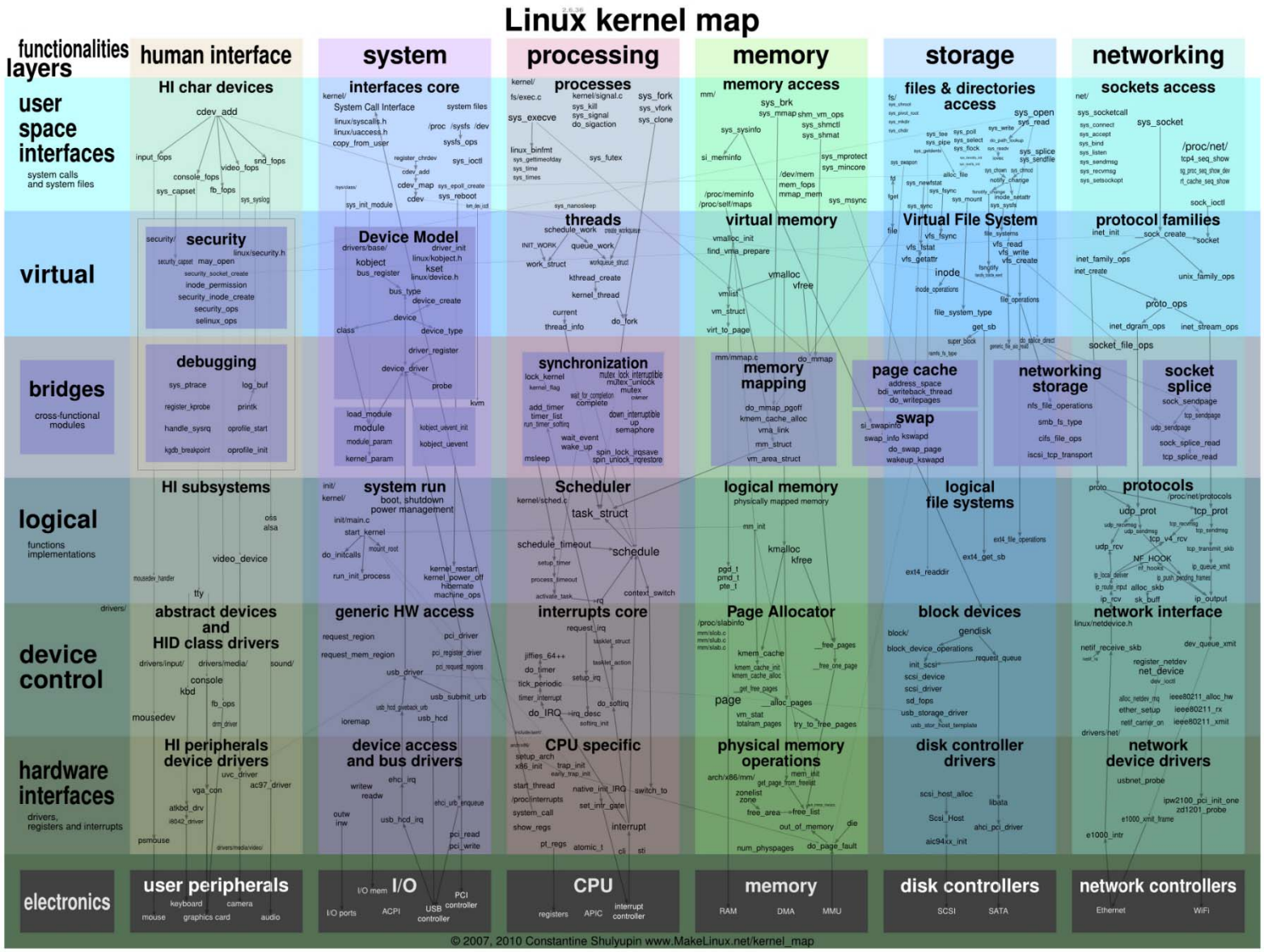
Pro: Demand very small set of functions

Easily ported to another system

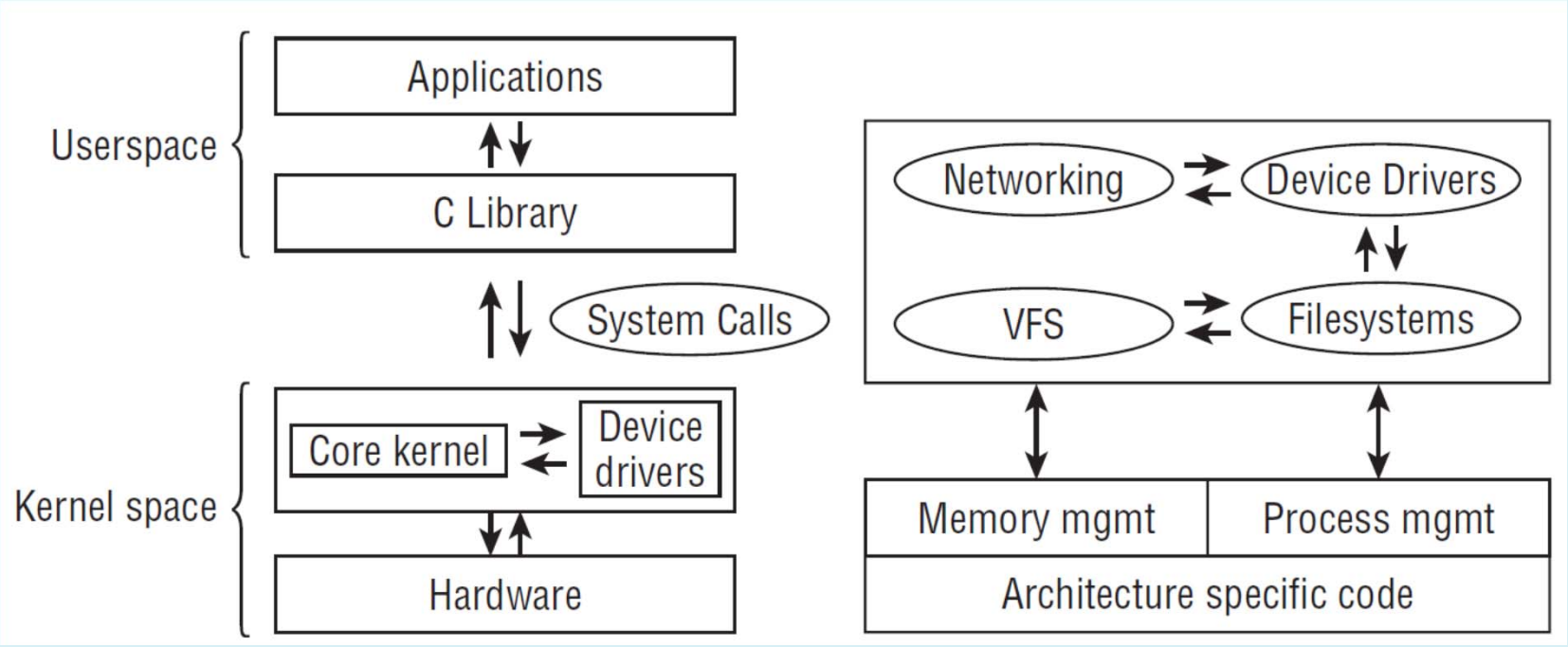
Con: Slower



Linux Kernel Architecture

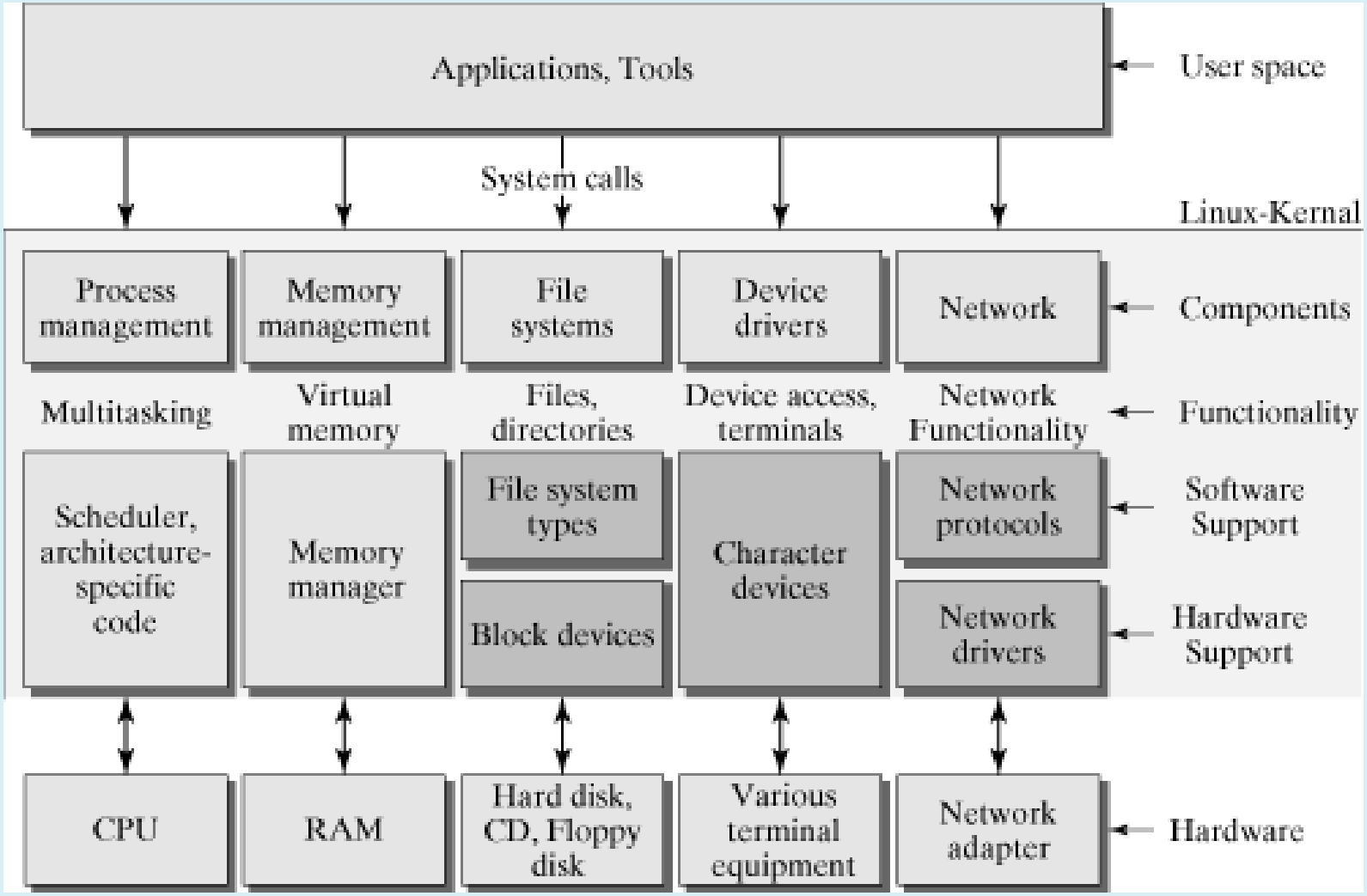


Linux Kernel Architecture



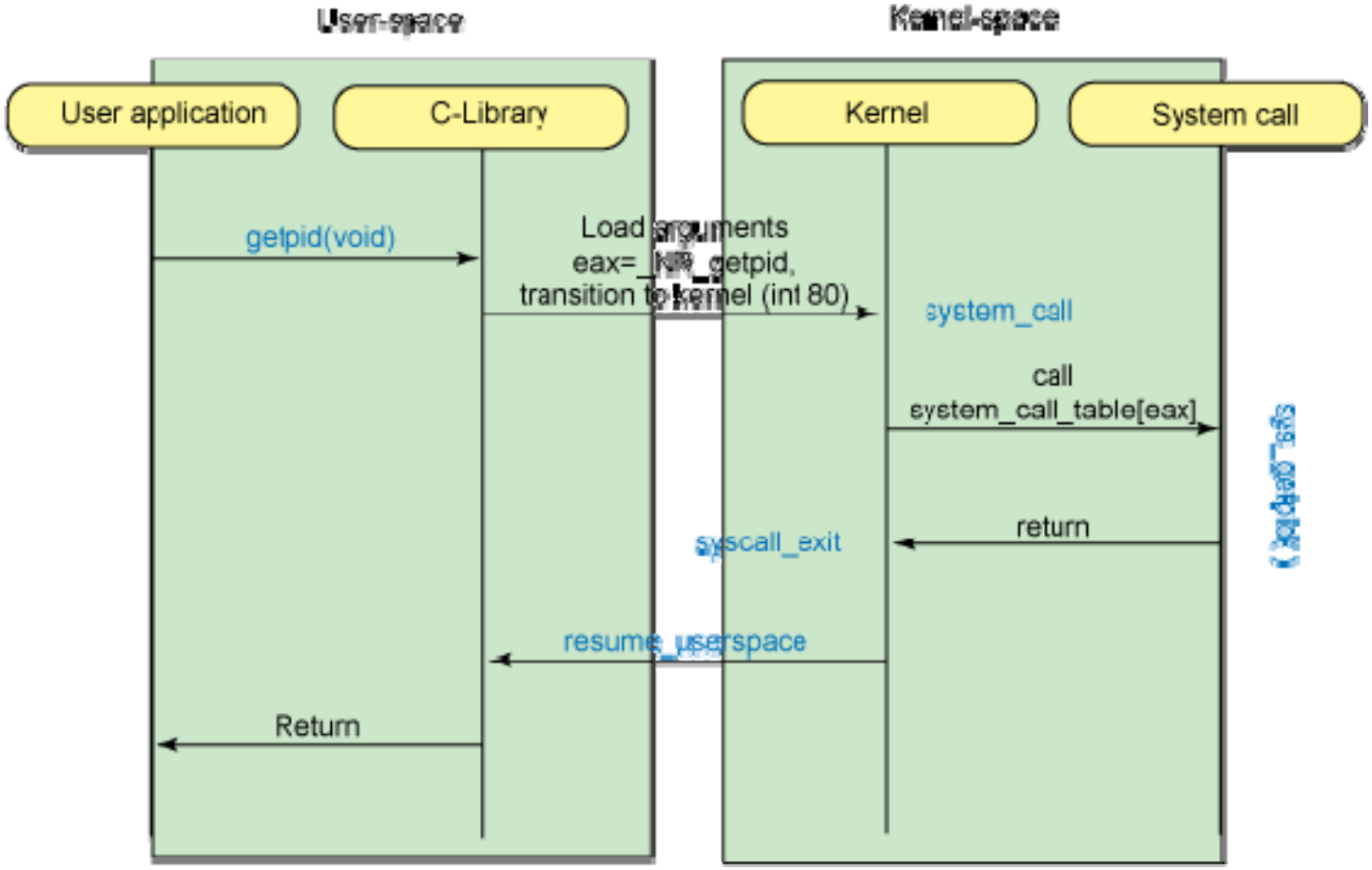
High-level view of Linux Kernel Architecture

Linux Kernel Architecture



What's going on?

- When application try to get its process id



What's going on?

- **Let's see in much lower level.**

hello.asm

```
section .data
   msg db      "hello, world!"

section .text
   global _start
_start:
   mov     rax, 1
   mov     rdi, 1
   mov     rsi, msg
   mov     rdx, 13
   syscall
   mov    rax, 60
   mov    rdi, 0
   syscall
```

To assemble and run

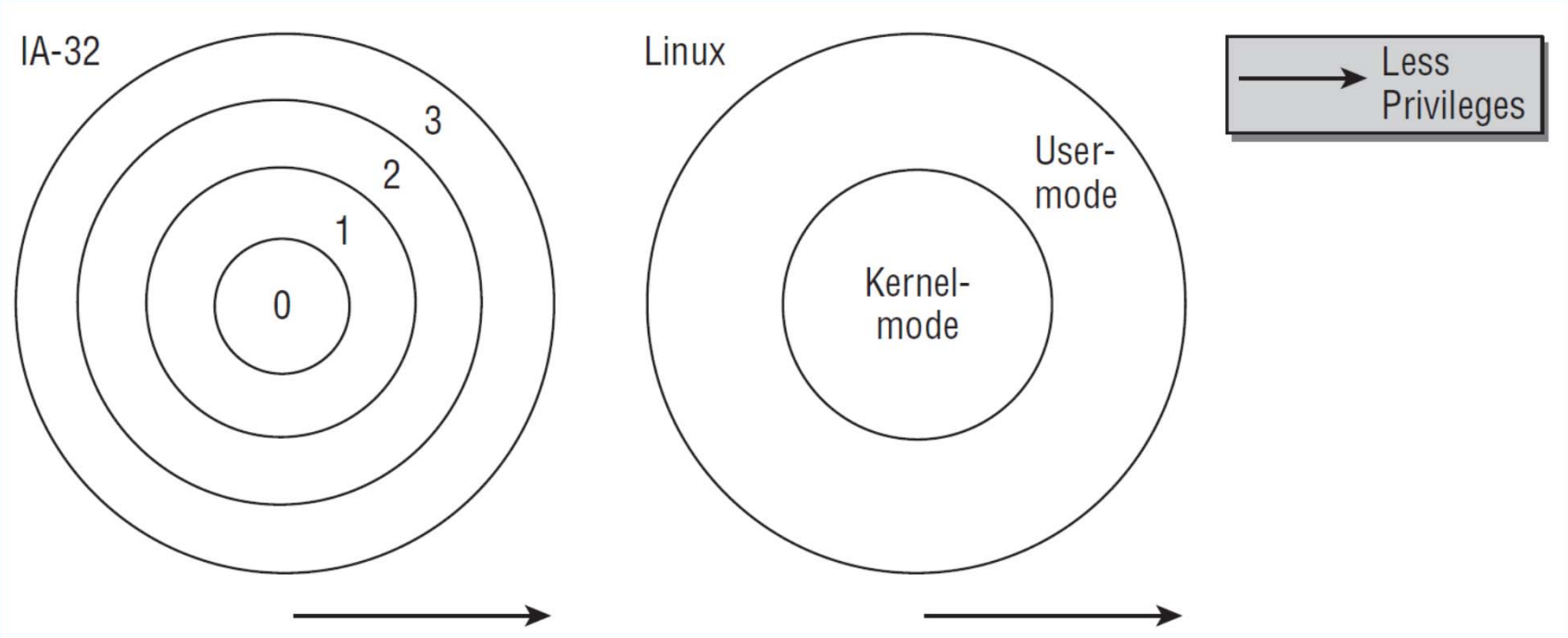
```
$ nasm -f elf64 -o hello.o hello.asm
$ ld -o hello hello.o
```

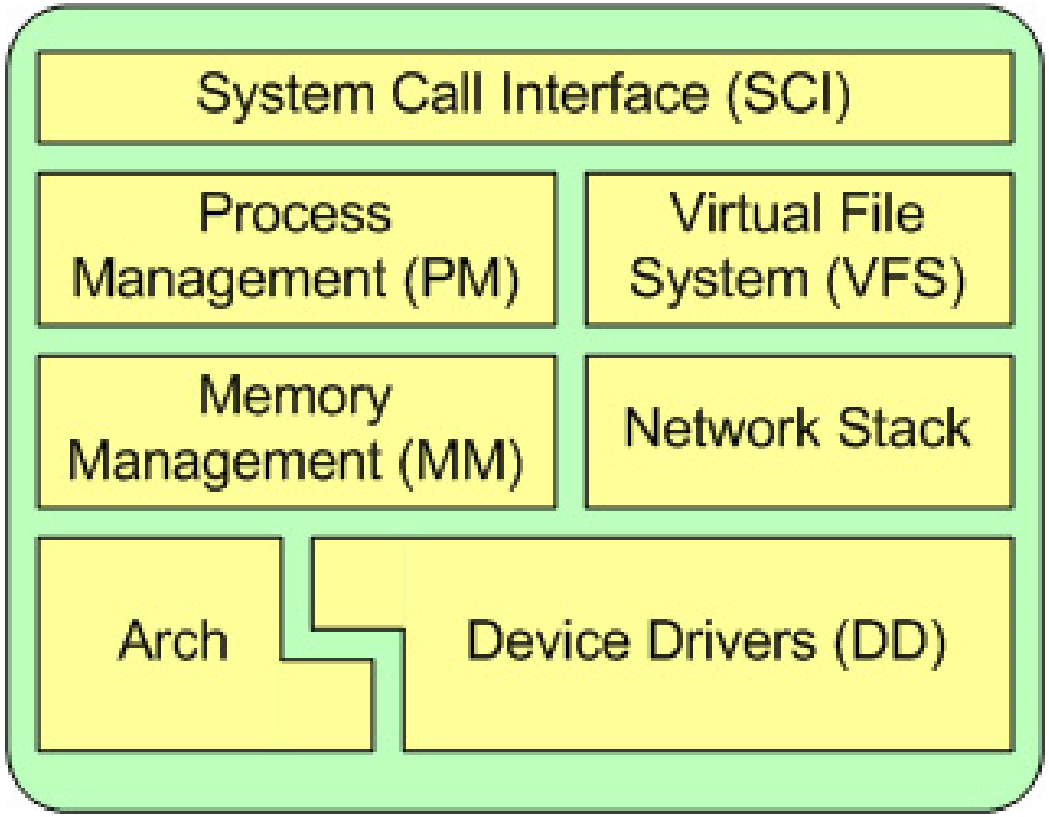
%rax	System call	%rdi	%rsi	%rdx
0	sys_read	unsigned int fd	char *buf	size_t count
1	sys_write	unsigned int fd	const char *buf	size_t count

60	sys_exit	int error_code

Linux system call table can be found:
http://blog.rchapman.org/post/36801038863/linux-system-call-table-for-x86-64

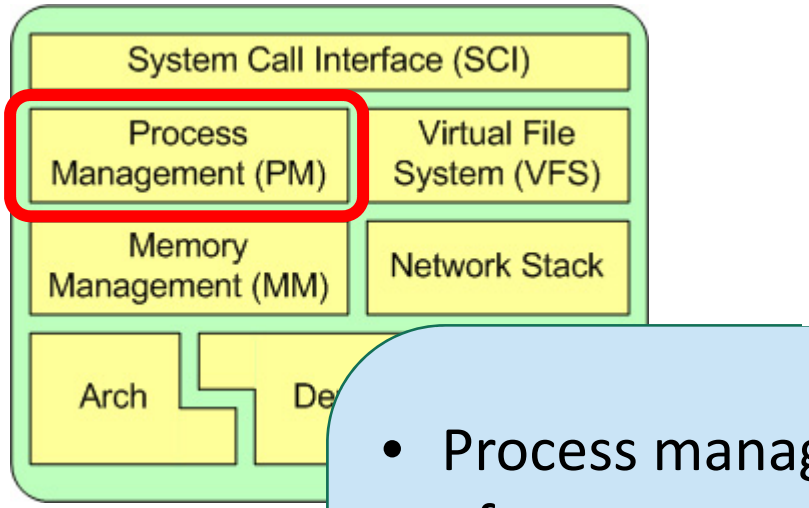
OS Privilege Mode





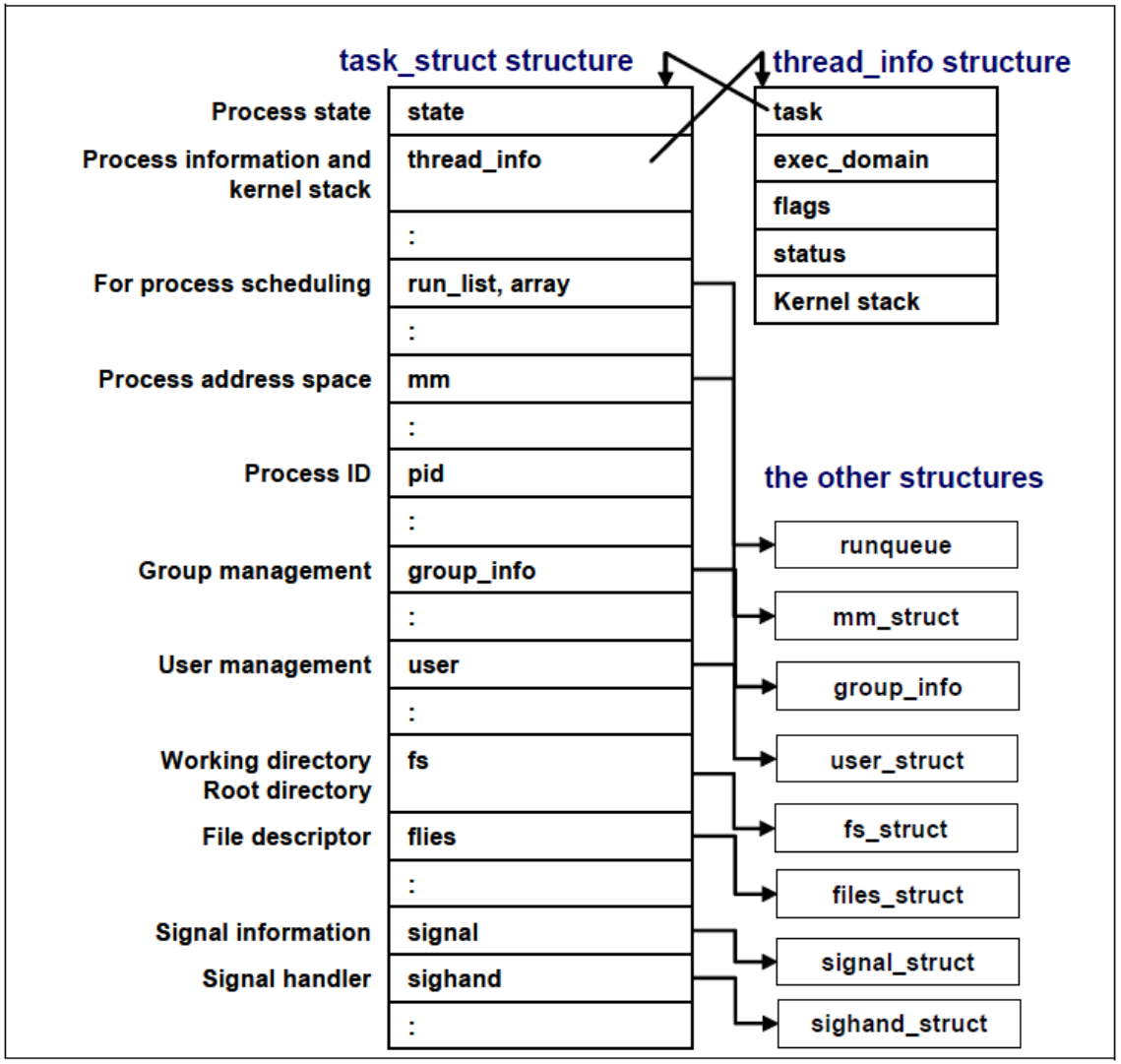
Let’s take a look at the Linux kernel

Process Management



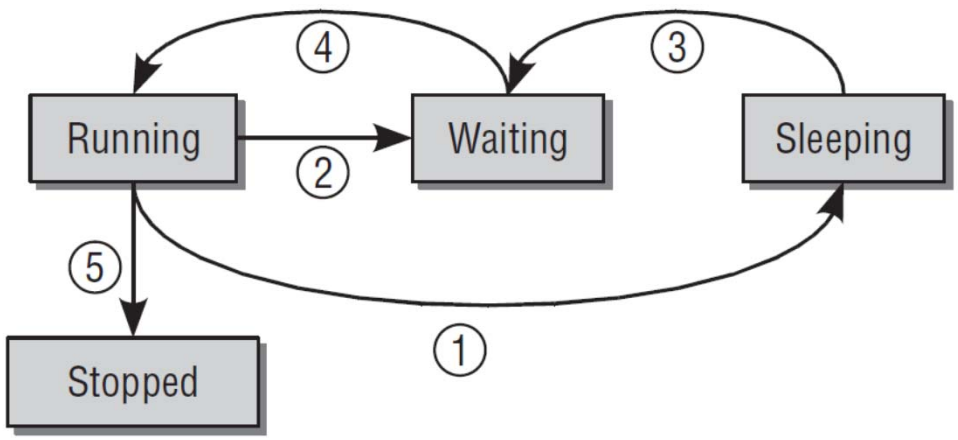
- Process management is focused on the execution of processes.
- In the kernel, these are called threads and represent an individual virtualization of the processor (thread code, data, stack, and CPU registers).

Process Management



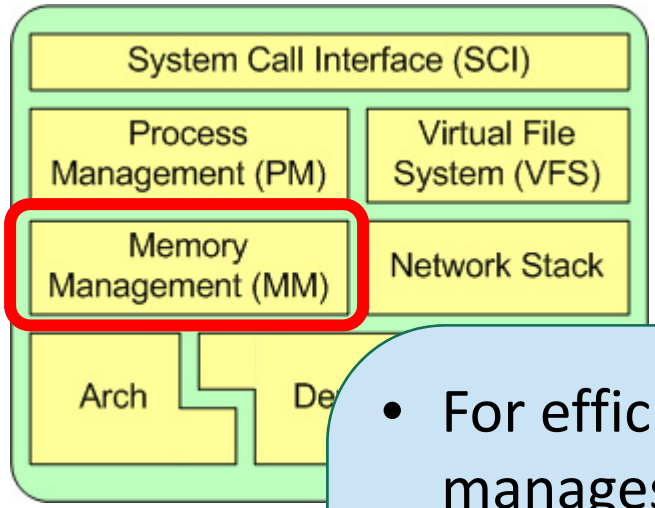
- Process information structure for Linux

Process Management



- Process states transitions

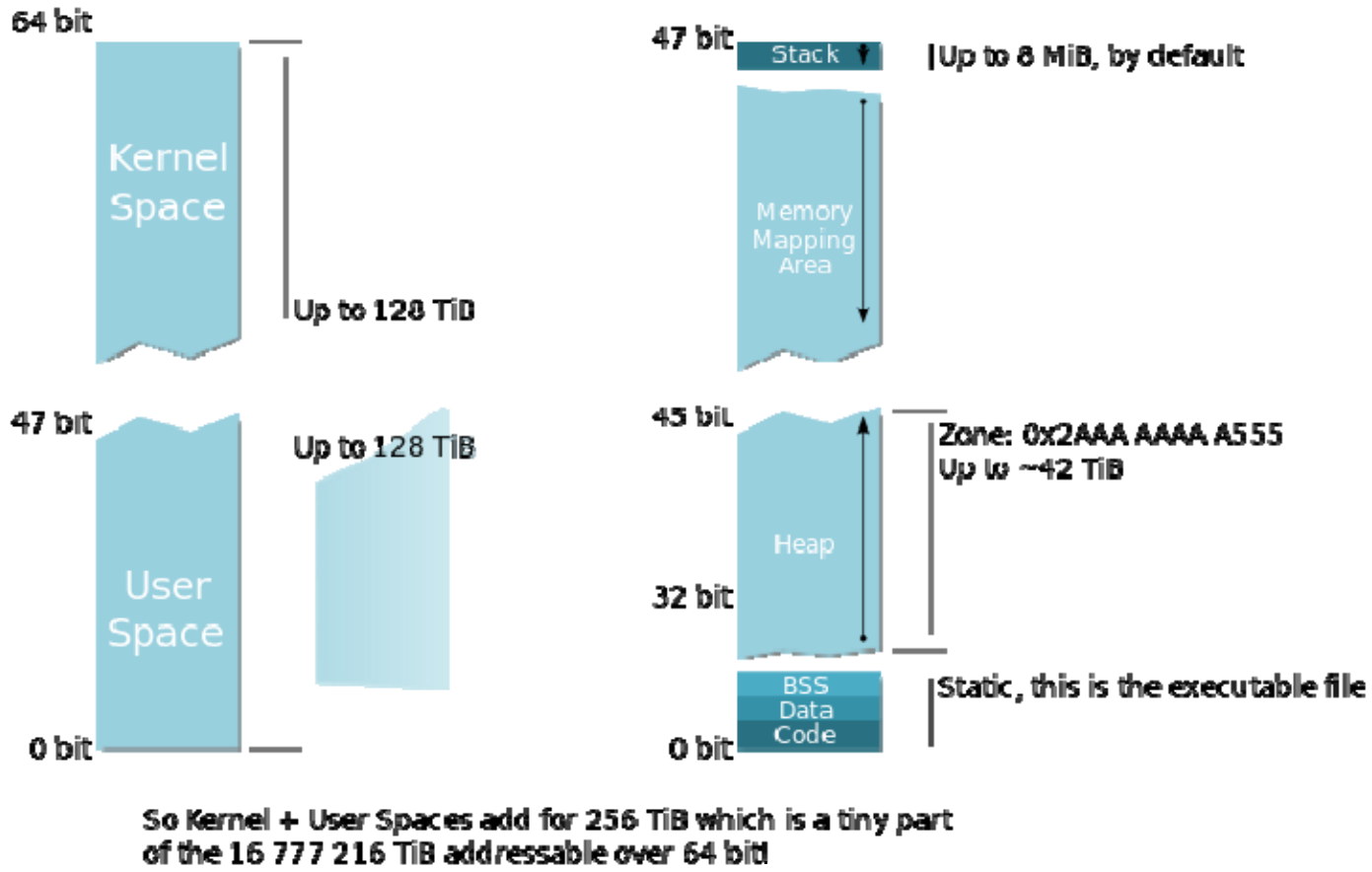
Memory Management



- For efficiency, given the way that the hardware manages virtual memory, memory is managed in what are called pages (4KB in size for most architectures).

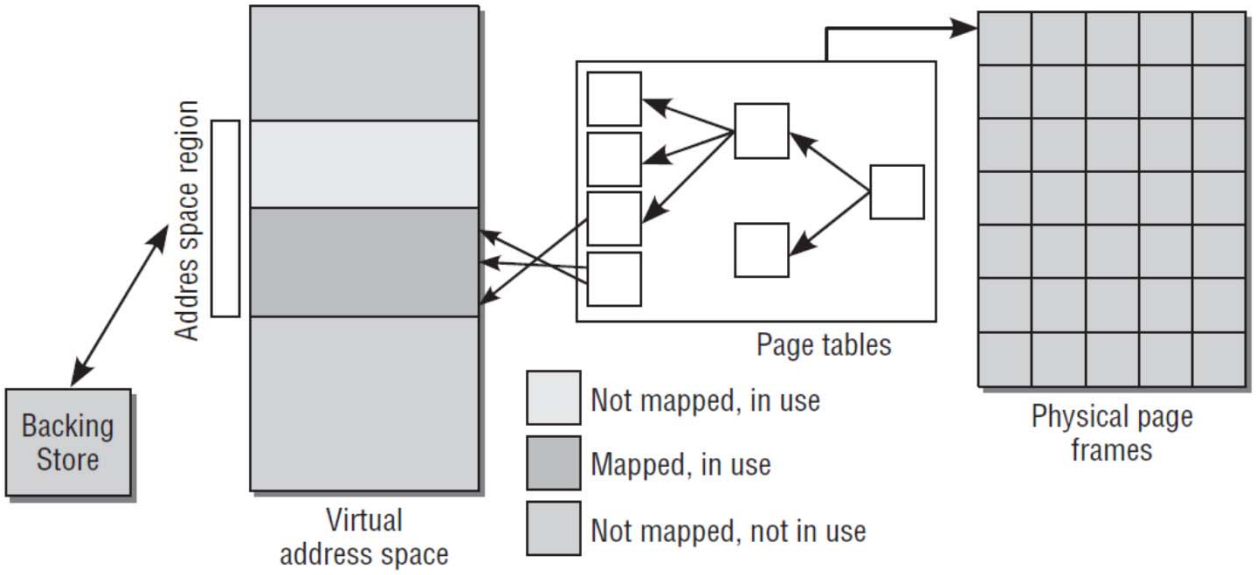
- Linux includes the means to manage the available memory, as well as the hardware mechanisms for physical and virtual mappings.

Memory Management



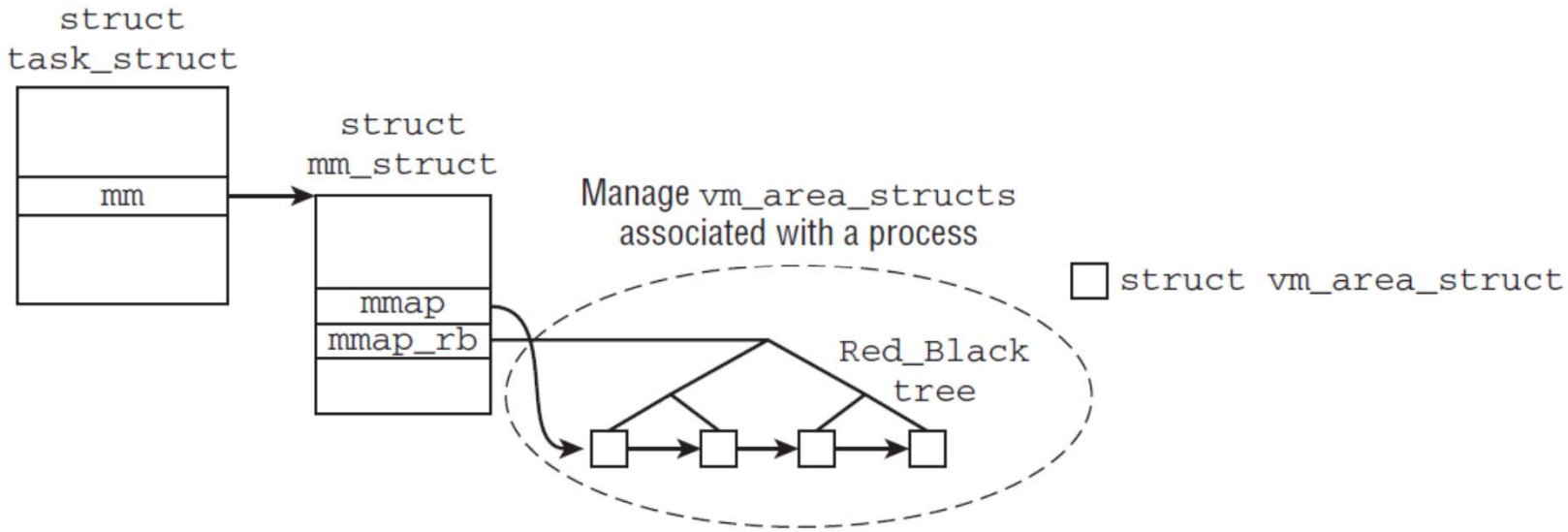
64 bit virtual memory address layout

Memory Management



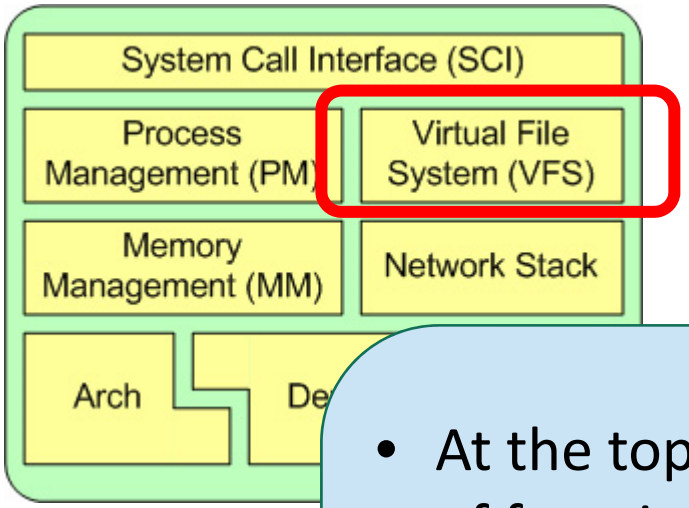
Virtual memory paging

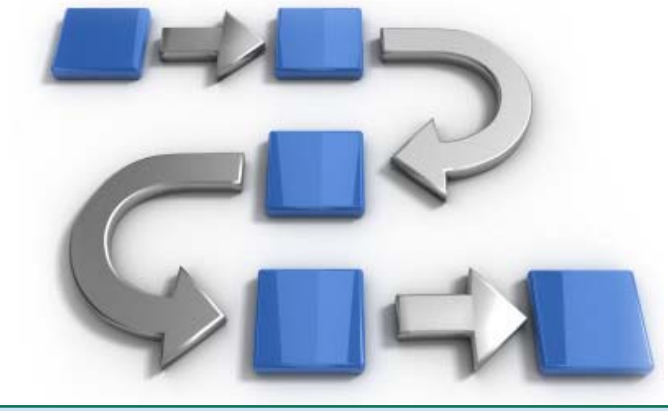
Memory Management



- Virtual memory structures

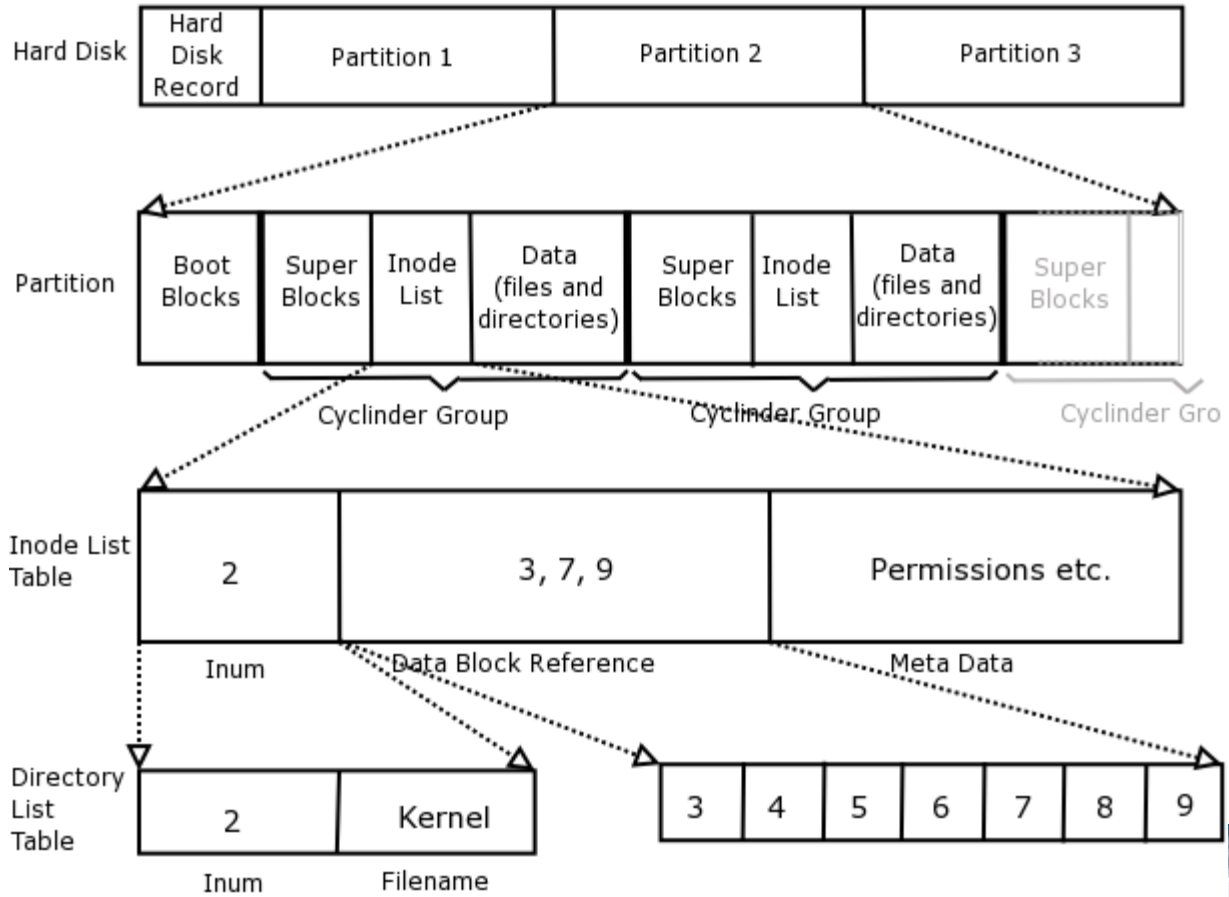
File System



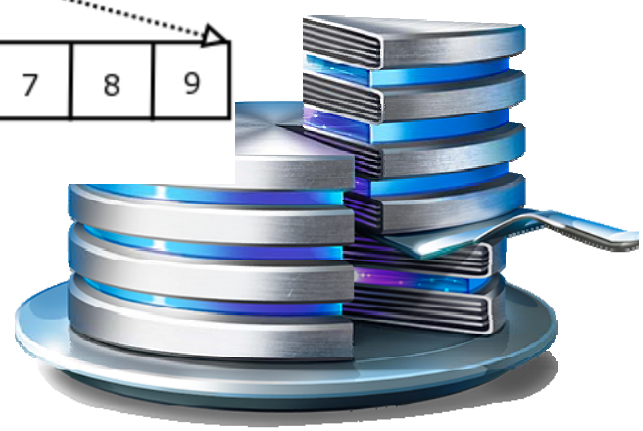


- At the top of the VFS is a common API abstraction of functions such as open, close, read, and write.
- At the bottom of the VFS are the file system abstractions that define how the upper-layer functions are implemented. These are plug-ins for the given file system (of which over 50 exist).

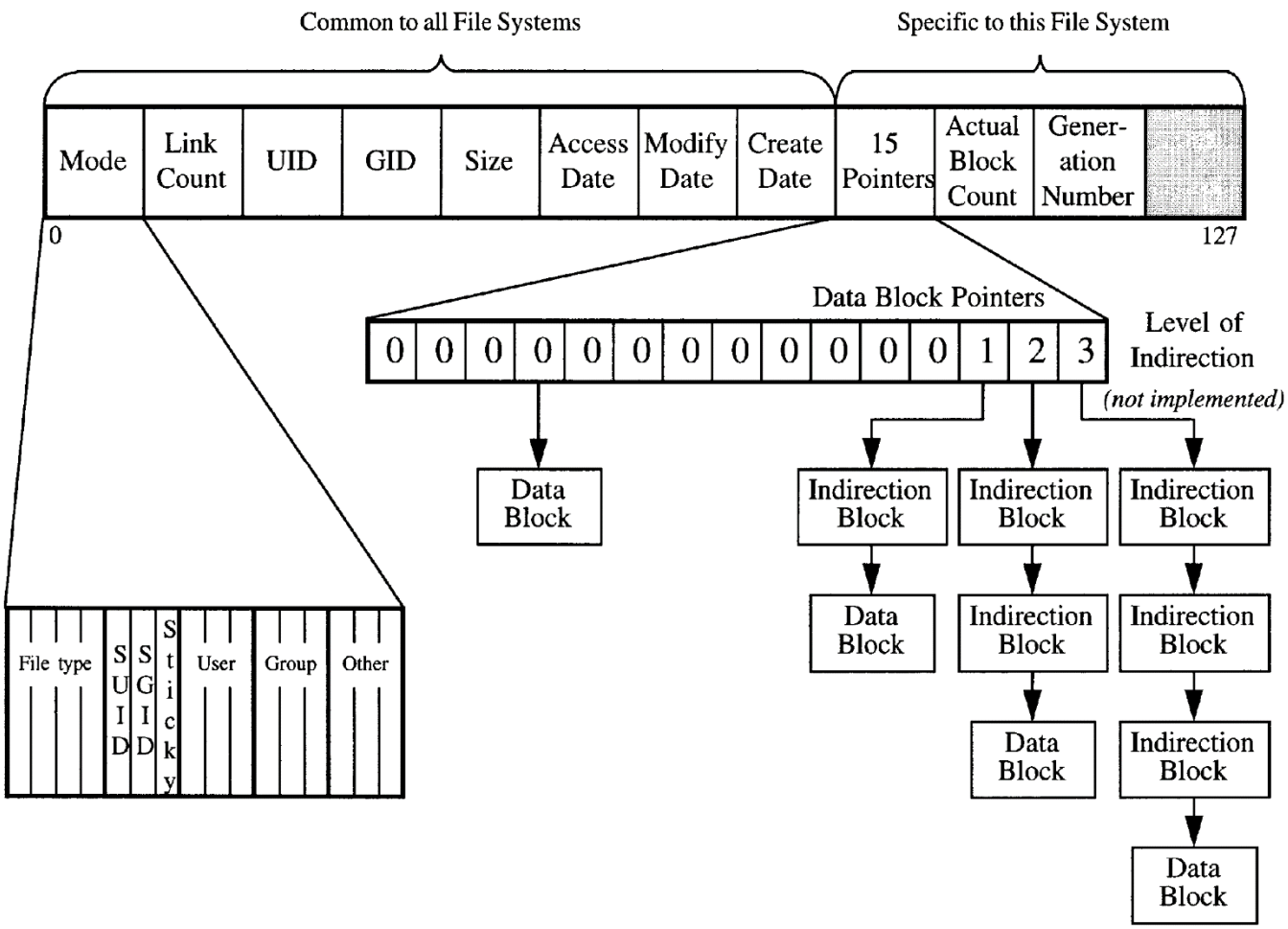
File System Layout

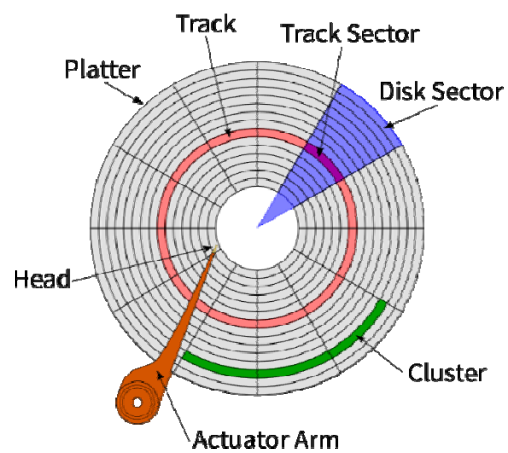


File System Layout



Disk Inode

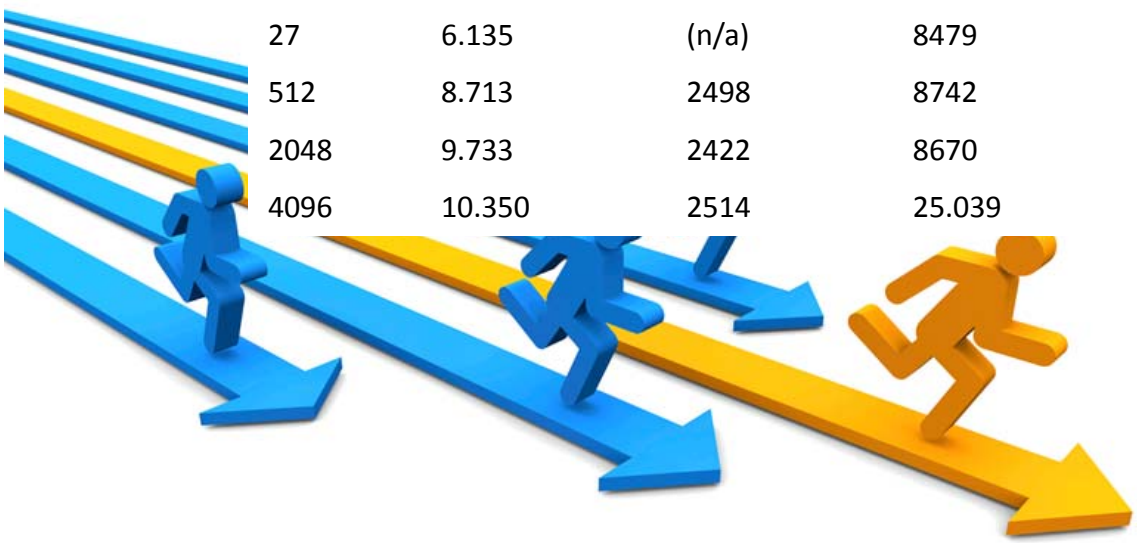

Common to all File Systems
Specific to this File System
Mode
Link Count
UID
GID
Size
Access Date
Modify Date
Create Date
15 Pointers
Actual Block Count
Gener-ation Number
0
127
Data Block Pointers
0 0 0 0 0 0 0 0 0 0 0 0 1 2 3
Level of Indirection
(not implemented)
Data Block
Indirection Block
Indirection Block
Indirection Block
Data Block
Indirection Block
Indirection Block
Data Block
Indirection Block
Data Block
File type
SUID
SGID
Sticky
User
Group
Other


Track
Track Sector
Platter
Disk Sector
Head
Cluster
Actuator Arm

How this knowledge helps?

- Performance tuning
  - Change memory paging size
    - Recompile the kernel with a proper page size options
  - Change disk block size
    - Create a new file system with different options

Block Size	Cached Normal	Cached Direct	Uncached Normal	Uncached Direct
1	5.887	(n/a)	8306	(n/a)
27	6.135	(n/a)	8479	(n/a)
512	8.713	2498	8742	2579
2048	9.733	2422	8670	2721
4096	10.350	2514	25.039	2560



Results:
no bs= 78s 144584+0 records
bs=512 78s 144584+0 records
bs=1k 38s 72292+0 records
bs=2k 38s 36146+0 records
bs=4k 38s 18073+0 records
bs=5k 39s 14458+1 records
bs=50k 38s 1445+1 records
bs=500k 39s 144+1 records
bs=512k 39s 144+1 records
bs=1M 39s 72+1 records
bs=5M 39s 14+1 records
bs=10M 39s 7+1 records

How this knowledge helps?

- Performance tuning
  - Use customizable Linux distro

How this knowledge helps?

- **Performance tuning**
  - **Unload unnecessary modules**

Module Commands

depmod - handle dependency descriptions for loadable kernel modules.
insmod - install loadable kernel module.
lsmod - list loaded modules.
modinfo - display information about a kernel module.
modprobe - high level handling of loadable modules.
rmmod - unload loadable modules.

How this knowledge helps?

- **Performance tuning**
  - **Unload unnecessary modules**

```
$ lsmod
Module                  Size  Used by
pci_stub               12622  1
joydev                 17381  0
hid_generic            12548  0
usbhid                 52659  0
hid                   106148  2 hid_generic,usbhid
eeepc_wmi              13151  0
parport_pc             32701  0
```

How this knowledge helps?

- **Secure Systems**

1	CVE-2015-5364	399	DoS	2015-08-31	2015-08-31	7.8

The (1) udp_recvmsg and (2) udpv6_recvmsg functions in the Linux kernel before 4.
attackers to caus a denial of ervice (system ha g) via incorrect checksums within

	CVE-2015-5157	264	+Priv	2015-08-31	2015-08-31	7.2

arch/x86/entry/entry_64.S in the Linux kernel before 4.1.6 on the x86_64 platform r
execution, which might allow local users to gain privileges by triggering an NMI.

3	CVE-2015-4036	119	DoS Overflow Mem. Corr.	2015-08-31	2015-09-01	7.2

Array index error in the tcm_vhost_make_tpg function in drivers/vhost/scsi.c in the l
service (memory corruption) or possibly have unspecified other impact via a crafted
renamed to vhost_scsi_make_tpg before the vulnerability was announced.

4	CVE-2015-4003	189	DoS	2015-06-07	2015-06-08	7.8

The oz_usb_handle_ep_data function in drivers/staging/ozwpan/ozusbsvc1.c in the C
cause a denial of service (divide-by-zero error and system crash) via a crafted pack

5	CVE-2015-3290	264	+Priv	2015-08-31	2015-08-31	7.2

arch/x86/entry/entry_64.S in the Linux kernel before 4.1.6 on the x86_64 platform i
users to gain privileges by triggering an NMI within a certain instruction window.

6	CVE-2015-1805	17	DoS +Priv	2015-08-08	2015-08-25	7.2

The (1) pipe_read and (2) pipe_write implementations in fs/pipe.c in the Linux kerne
__copy_to_user_inatomic and __copy_from_user_inatomic calls, which allows local u
via a crafted application, aka an "I/O vector array overrun."

7	CVE-2015-1465	17	DoS	2015-04-05	2015-04-09	7.8

The IPv4 implementation in the Linux kernel before 3.18.8 does not properly conside
lookups in the absence of caching, which allows remote attackers to cause a denial c

8	CVE-2015-1212		DoS	2015-02-06	2015-03-11	7.5

Multiple unspecified vulnerabilities in Google Chrome before 40.0.2214.111 on Windc
cause a denial of service or possibly have other impact via unknown vectors.

9	CVE-2015-1211	264	+Priv	2015-02-06	2015-03-11	7.5

The OriginCanAccessServiceWorkers function in content/browser/service_worker/se
Windows, OS X, and Linux and before 40.0.2214.109 on Android does not properly r
remote attackers to gain privileges via a filesystem: URI.

10	CVE-2015-1209		DoS	2015-02-06	2015-03-11	7.5

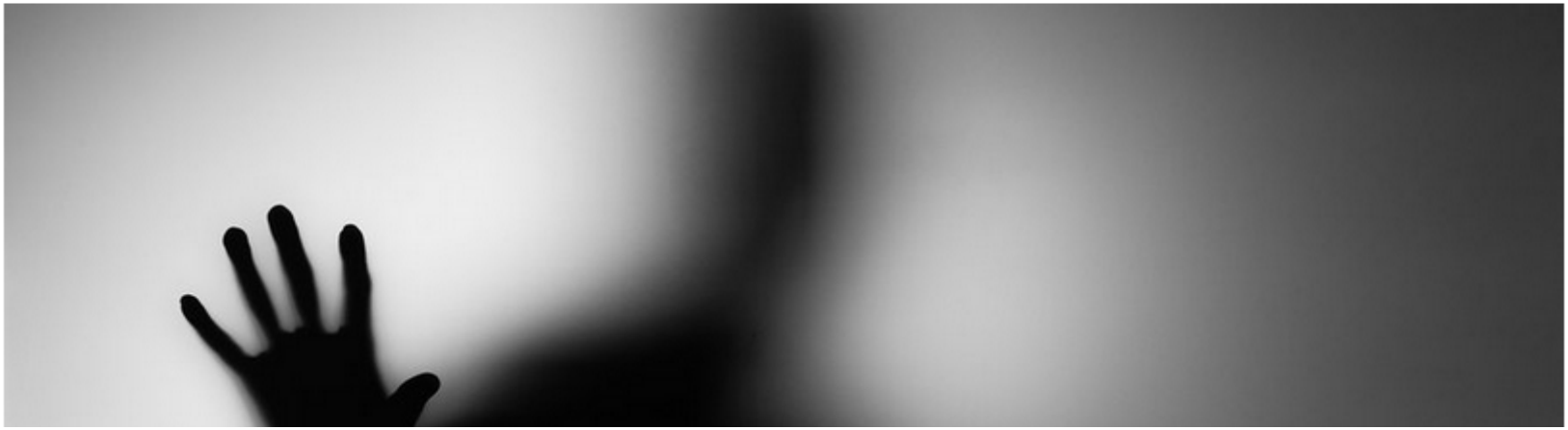
Use-after-free vulnerability in the VisibleSelection::nonBoundaryShadowTreeRootNo
Blink, as used in Google Chrome before 40.0.2214.111 on Windows, OS X, and Linu
denial of service or possibly have unspecified other impact via crafted JavaScript co

11	CVE-2015-0274	19	DoS +Priv	2015-03-16	2015-03-26	7.2

The XFS implementation in the Linux kernel before 3.15 improperly uses an old size
a denial of service (transaction overrun and data corruption) or possibly gain privile

How this knowledge helps?

- Kernel hardening
  - Disable and blacklist Linux modules



Ghost in the Machine: Linux Zero-Day Vulnerability Opens Door for Attack

BY PAMELA COBB • JANUARY 29, 2015

Questions?

Thank you