

```
version 2.6.9-1.667 (bhcompile@twtey.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 0000000000e40000 - 0000000000100000 (reserved)
BIOS-e820: 0000000001000000 - 00000000010ef000 (usable)
BIOS-e820: 00000000010ef000 - 00000000010efc00 (ACPI data)
BIOS-e820: 00000000010efc00 - 00000000010ff000 (ACPI NVS)
BIOS-e820: 00000000010ff000 - 0000000001000000 (usable)
BIOS-e820: 00000000010fec0000 - 00000000010fec1000 (reserved)
BIOS-e820: 00000000010fee0000 - 00000000010fee01000 (reserved)
BIOS-e820: 00000000010fee0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd
mapped 4G/4G trampoline to f
Initializing CPU#0
CPU 0 irqstacks, hard-020
PID hash table entries:
.....
```

# Debugging & Time v Linuxovém kernelu

David Majda ([david@majda.cz](mailto:david@majda.cz))

prezentace na předmět Linux Kernel (SW075) na MFF UK

12. 12. 2005

všechny informace v prezentaci se týkají jádra 2.6.14.3

```
kernel version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 0000000000e40000 - 0000000000100000 (reserved)
BIOS-e820: 0000000001000000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fec000 (ACPI data)
BIOS-e820: 0000000000fec000 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 000000000fec0000 - 000000000fec10000 (reserved)
BIOS-e820: 000000000fee0000 - 000000000fee01000 (reserved)
BIOS-e820: 000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTBL$ 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=LABEL=
mapped 4G/4G trampoline to 4f000000
Initializing CPU#0
CPU 0 irqstacks, hard-0220
PID hash table entries: 128
.....
```

# Debugging



```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fec000 (ACPI data)
BIOS-e820: 0000000000fec000 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000001000000 (usable)
BIOS-e820: 000000000fec0000 - 000000000fec1000 (reserved)
BIOS-e820: 000000000fee0000 - 000000000fee1000 (reserved)
BIOS-e820: 000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard-0x00000000
PID hash table entries: 128
```

# Kernel Debugging

- Kdy je třeba ladit kernel?
  - jsem kernelový vývojář
  - píšu driver
  - kernel mi padá a nevím proč
  - chci vědět, jak to funguje uvnitř
- Kdy není třeba ladit kernel?
  - jsem-li normální programátor a kernel považuju za fungující černou skříňku implementující ±UNIXové API

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec1000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x3108
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G trampoline to
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:

```

# printk()

- výpis hlášení a chyb do konzole, dmesg a syslogu
- analogie `printf()` – stejné použití, obdobná syntaxe formátovacího řetězce

„You will know when you are a real kernel hacker when you start typing `printf` as `printk` in your user programs :)“

`/usr/src/linux/Documentation/kernel-hacking.tmpl`

- lze používat i v interrupt kontextu
- 1kB buffer – přetečení je problém uživatele
- zápis do `/dev/kmsg` způsobí `printk()`
- parametry jádra
  - `earlyprintk=vga|serial[,ttySn[,baudrate]]`
  - `time` – prefixuje výpisy časem

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec00000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee00000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G trampoline to f
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:
.....
```

# printk()

- nastavení úrovně výpisu:

```
/* system is unusable */
#define KERN_EMERG          "<0>"
/* action must be taken immediately */
#define KERN_ALERT         "<1>"
/* critical conditions */
#define KERN_CRIT          "<2>"
/* error conditions */
#define KERN_ERR           "<3>"
/* warning conditions */
#define KERN_WARNING       "<4>"
/* normal but significant condition */
#define KERN_NOTICE        "<5>"
/* informational */
#define KERN_INFO          "<6>"
/* debug-level messages */
#define KERN_DEBUG         "<7>"
```

- použití:

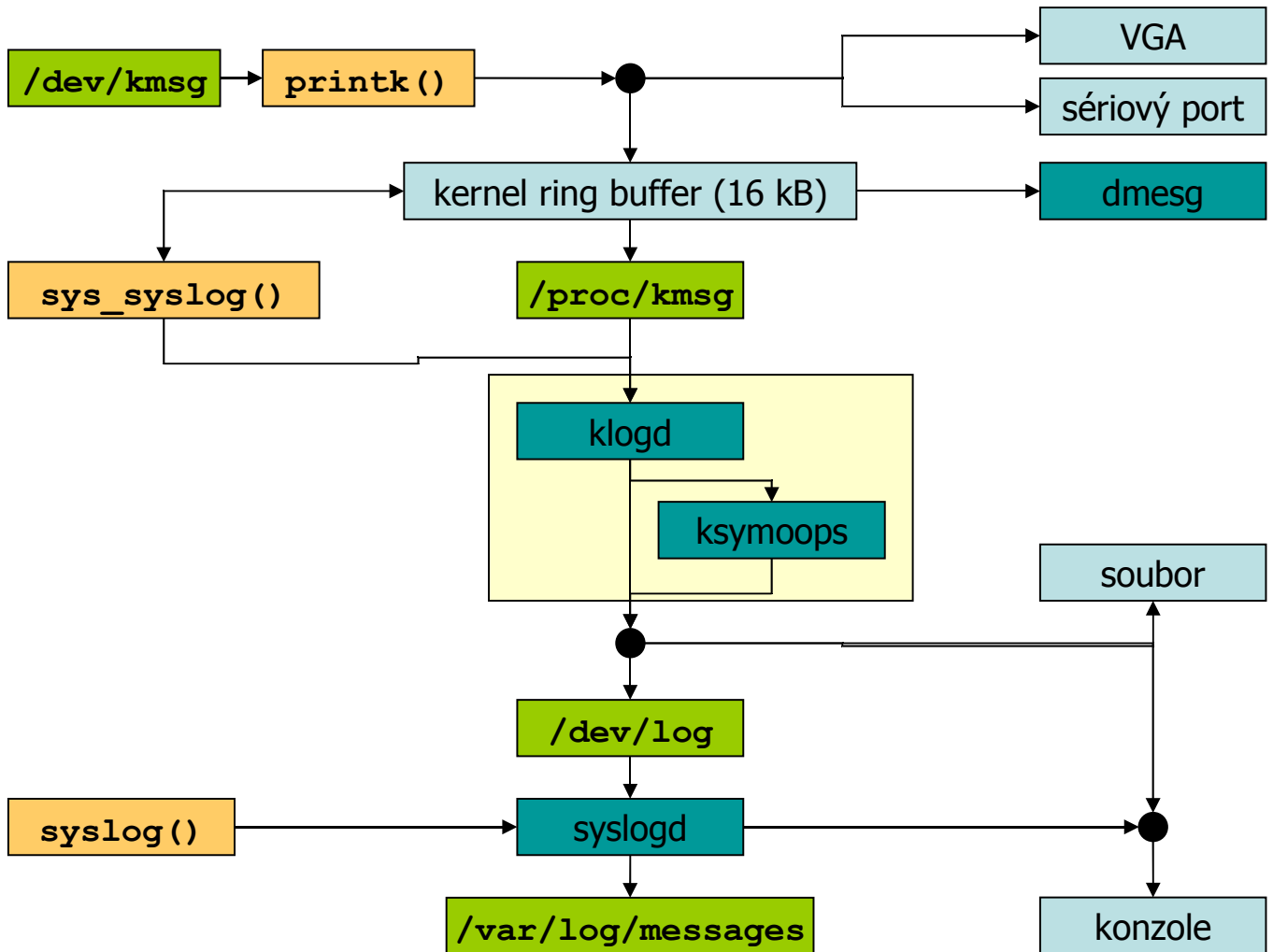
```
printk(KERN_ERR "%s%d: adapter kernel panic.\n",
        dev->name, instance);
```

```

version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000001000000 (usable)
BIOS-e820: 000000000fec0000 - 000000000fec1000 (reserved)
BIOS-e820: 000000000fee0000 - 000000000fee0100 (reserved)
BIOS-e820: 000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=LABEL=
mapped 4G/4G trampoline to ffff
Initializing CPU#0
CPU 0 irqstacks, hard-020
PID hash table entries:

```

# Velké schéma logování



```
kernel version 2.6.9-1.667 (bhcompile@twetty.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef000 (usable)
BIOS-e820: 00000000001ef000 - 00000000001efc00 (ACPI data)
BIOS-e820: 00000000001efc00 - 00000000001ff000 (ACPI NVS)
BIOS-e820: 00000000001ff000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
  DMA zone: 4096 pages, LIFO batch:1
  Normal zone: 61440 pages, LIFO batch:15
  HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTBL 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G trampoline to 0x
Initializing CPU#0
CPU 0 irqstacks, hard-0x
PID hash table entries:
#
```

# Oops!

- „segfault v kernelu“ – přístup k paměti, kam se přistupovat nesmí
- segfault je pro program fatální, pro kernel nikoliv
  - např. Oops v ovladači může nechat kernel ve stabilním stavu
  - ale obecně oopsnutý kernel nelze považovat za nadále stabilní
- způsobí výpis ladicích informací (registry procesoru, umístění tabulek deskriptorů stránek, call trace...)
- `klogd` odchyťává Oops, mění na logy, případně automaticky (pomocí `ksymoops`) nahrazuje adresy za názvy funkcí dle `System.map`

# Oops – příklad

- příklad Oops:

```
Unable to handle kernel paging request at virtual address 211e2018
c0129577
*pde = 00000000
Oops: 0000
CPU: 0
EIP: 0010:[<c0129577>] Not tainted
Using defaults from ksymoops -t elf32-i386 -a i386
EFLAGS: 00010083
eax: d7ee5000 ebx: b420e080 ecx: c164e000 edx: c1615d04
esi: c16073d0 edi: 00000246 ebp: 000001f0 esp: d7c5de84
ds: 0018 es: 0018 ss: 0018
Process mount (pid: 25, stackpage=d7c5d000)
Stack: 00000000 c0309c00 000001f0 00000000 c01fadb7 c16073d0 000001f0 c1615a40
c1615700 c1615a40 c01fa126 00000001 000001f0 00000000 c022f793 c1615a40
00000001 00000000 000001f0 d7b6fde0 d7c5df14 0000006e bfffec0c 00000018
Call Trace: [<c01fadb7>] [<c01fa126>] [<c022f793>] [<c01f8acb>] [<c01f8720>]
[<c01f9450>] [<c0106d40>] [<c0106c4f>]
Code: 8b 44 81 18 89 41 14 83 f8 ff 75 1d 8b 41 04 8b 11 89 42 04
```

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000ef0000 (usable)
BIOS-e820: 0000000000ef0000 - 0000000000ef0000 (ACPI data)
BIOS-e820: 0000000000ef0000 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee10000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTBL 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x3008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard-0x00000000
PID hash table entries: 128
.....
```

# System.map

- `/boot/System.map-x.y.z`
- tabulka symbolů (zejm. globální proměnné a funkce)
- použití:
  - při kernel Oops
  - debugger
- vytvořen při kompilaci kernelu (skript `scripts/mksysmap`)
- příklad
  - formát: `adresa typ jméno`

```
c0344204 b pirq_table
c0344208 b pirq_router
c034420c b pirq_router_dev
c0344220 b ascii_buffer
c0344224 b ascii_buf_bytes
```

- podrobnosti: `man nm`

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000fff000 (ACPI NVS)
BIOS-e820: 0000000000fff000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec1000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee1000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd1
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard-0/0
PID hash table entries: 1024
```

# klogd

- čte zprávy od kernelu
  - z `/proc/kmsg`, je-li namountován proc FS
  - jinak voláním `sys_syslog()`
  - lze ovlivnit parametrem
- předá zprávy dál
  - syslogd
  - do souboru (lze nastavit)
  - na konzoli (lze filtrovat prioritu)
- odchytává Oopsy
- Oopsy lze prohnat přes `ksymoops` – náhrada adres za symboly

```
version 2.6.9-1.667 (bhcompile@twetty.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee01000 (reserved)
BIOS-e820: 00000000fffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd
mapped 4G/4G trampoline to f
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:

```

# syslogd

- standardní UNIXový logovací démon
- zápis přes UNIX socket (`/dev/log`) nebo internet sockety (remote debugging)
- konfigurace v `/etc/syslog.conf`
- zápis knihovními funkcemi:

```
void openlog(const char *ident, int option,
             int facility);
void syslog(int priority, const char *format, ...);
void vsyslog(int priority, const char *format,
             va_list ap);
void closelog(void);
int setlogmask(int mask);
```

version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers  
14:41:25 EST 2004  
BIOS-provided physical RAM map:  
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)  
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)  
BIOS-e820: 00000000000a0000 - 00000000000e0000 (reserved)  
BIOS-e820: 00000000000e0000 - 0000000000100000 (reserved)  
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)  
BIOS-e820: 0000000000fef000 - 0000000000fff000 (ACPI data)  
BIOS-e820: 0000000000fff000 - 0000000000fff000 (ACPI NVS)  
BIOS-e820: 0000000000fff000 - 0000000010000000 (usable)  
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)  
BIOS-e820: 00000000fee00000 - 00000000fee01000 (reserved)  
BIOS-e820: 00000000ffe00000 - 0000000100000000 (reserved)  
OMB HIGHMEM available.  
256MB LOWMEM available.  
zapping low mappings.  
On node 0 totalpages: 65536  
DMA zone: 4096 pages, LIFO batch:1  
Normal zone: 61440 pages, LIFO batch:15  
HighMem zone: 0 pages, LIFO batch:1  
DMI present.  
ACPI: RSDP (v000 PTLTD  
ACPI: RSDT (v001 PTLTD RSDT 0x06040000  
ACPI: FADT (v001 INTEL 440BX 0x06040000  
ACPI: MADT (v001 PTLTD APIC 0x06040000  
ACPI: BOOT (v001 PTLTD \$SBFTRM 0x06040000  
ACPI: DSDT (v001 PTLTD Custom 0x06040000  
ACPI: PM-Timer IO Port: 0x1008  
Built 1 zonelists  
Kernel command line: ro root=hd0  
mapped 4G/4G trampoline to f0000000  
Initializing CPU#0  
CPU 0 irqstacks, hard-0/0  
PID hash table entries: 128

# Když chci ohlásit chybu...

- **BUG()**

– obecně (v `include/asm-generic/bugs.h`):

```
#define BUG() do { \  
    printk("kernel BUG at %s:%d!\n", \  
           __FILE__, __LINE__); \  
    panic("BUG!"); \  
} while (0)
```

– na různých architekturách předefinováno

- **BUG\_ON()**

– přibližně:

```
#define BUG_ON(condition) \  
    if ((condition) != 0) BUG();
```



```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee10000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x00004000
ACPI: FADT (v001 INTEL 440BX 0x00004000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTBLX 0x00000000
ACPI: DSDT (v001 PTLTD Custom 0x00000000
ACPI: PM-Timer IO Port: 0x3008
Built 1 zonelists
Kernel command line: ro root=hd=0
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard=0x00000000
PID hash table entries: 1024
hash...

```

# kgdb

- kgdb – běžné ladění (breakpointy, watches...) kernelu pomocí gdb
- potřeba 2 stroje (jeden testovaný, druhý monitorovací), propojené sériovým kabelem
- architektury i386, x86\_64, ppc, s390
- nutný patch do kernelu (gdb\_stub, fault hndlery, komunikace po kabelu), a modifikovaná verze GDB
- spustit testovaný stroj s modifikovným kernelem, pak spustit monitorovací stroj:

```
(gdb) target remote /dev/ttyS1
Remote debugging using /dev/ttyS1
breakpoint () at gdbstub.c:1153
1153      }
(gdb) c

Continuing.

...
```

- přenáší se konzolové výpisy
- ladí se jako v obyčejném gdb

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000a0000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e0000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000ef0000 (usable)
BIOS-e820: 0000000000ef0000 - 0000000000efc000 (ACPI data)
BIOS-e820: 0000000000efc000 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000001000000 (usable)
BIOS-e820: 0000000001000000 - 0000000001fec0000 (reserved)
BIOS-e820: 0000000001fec0000 - 0000000001fee01000 (reserved)
BIOS-e820: 0000000001fee01000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd1
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard-0x00000000
PID hash table entries: 128
```

# Problémy s atomicitou a

- `in_interrupt()`
  - jsme v interrupt kontextu?
- `in_atomic()`
  - jsme v atomickém kontextu?
  - atomický kontext = když určitě nebudeme přepřánováni (spinlock, interrupt handler,...)
- `might_sleep()`
  - „anotace“ funkcí, které se mohou uspat
  - pokud je zavoláno v atomickém kontextu, vypíše stack trace, jinak nedělá nic
- `might_sleep_if(cond)`
  - přibližně:

```
#define might_sleep_if(cond) \
    if (cond) might_sleep();
```

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef000 (usable)
BIOS-e820: 00000000001ef000 - 00000000001efc00 (ACPI data)
BIOS-e820: 00000000001efc00 - 00000000001ff000 (ACPI NVS)
BIOS-e820: 00000000001ff000 - 0000000010000000 (usable)
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee10000 (reserved)
BIOS-e820: 00000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x00000000
ACPI: DSMT (v001 PTLTD Custom 0x00000000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G trampoline to
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:

```

# Magic SysRq

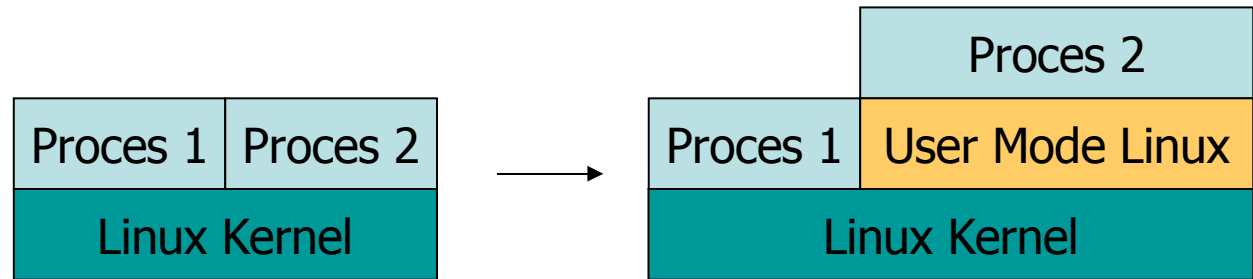
- „magická“ kombinace kláves, pomocí níž lze poslat několik základních příkazů kernelu
- hodí se, když se kernel nějak „poškodí“, zasekne apod.
- x86: Alt+SysRq+*cmd*, kde *cmd* je:
  - o = shutdown
  - b = reboot (bez sync/umount)
  - e = všem kromě init pošle SIGTERM
  - i = všem kromě init pošle SIGKILL
  - h = nápověda
- ...
- zneužitelné, na produkčních strojích nepoužívat!
  - zkompilovat kernel bez podpory Magic SysRq (`CONFIG_MAGIC_SYSRQ`)
  - vypnout v konfiguraci (`sysctl -w kernel.sysrq=0`, nebo `/etc/sysctl.conf`)

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000001000000 (usable)
BIOS-e820: 0000000000fec00000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee00000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTBL $
ACPI: DSDT (v001 PTLTD Custom
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to f
Initializing CPU#0
CPU 0 irqstacks, hard-0
PID hash table entries:

```

# User Mode Linux

- spustí Linux jako uživatelský proces v Linuxu



- jen na x86
- na co mi to je?
  - non-root může ke kernelu
  - snadné ladění, gprof, gcov,...
  - lze zkoušet cokoliv bez obav ze shození systému – „sandbox“
  - rychlejší než „klasická“ simulace/virtualizace (max. 20% zpomalení)
  - hezká hračka :-)



```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 0000000000e40000 - 0000000000100000 (reserved)
BIOS-e820: 0000000001000000 - 000000000fef0000 (usable)
BIOS-e820: 000000000fef0000 - 000000000fefc000 (ACPI data)
BIOS-e820: 000000000fefc000 - 000000000ff00000 (ACPI NVS)
BIOS-e820: 000000000ff00000 - 0000000010000000 (usable)
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee10000 (reserved)
BIOS-e820: 00000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTBL$ 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=LABEL=
mapped 4G/4G trampoline to 4f000000
Initializing CPU#0
CPU 0 irqstacks, hard-0200
PID hash table entries: 1024
##### 14:41:25 EST 2004
```

# Time



```
kernel version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers... 20041017 08:0...
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fec000 (ACPI data)
BIOS-e820: 0000000000fec000 - 0000000000f00000 (ACPI NVS)
BIOS-e820: 0000000000f00000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec00000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee00000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTBL 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G topoline to 4G
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:
#
```

# HZ

- makro, určující frekvenci generování přerušení od časovače (IRQ 0)
- hodnota ovlivňuje:
  - přesnost timerů
  - granularitu timeslice při plánování (a tedy latenci procesů)
  - výkonost `poll()` a `select()` (interně využívají timer)
- granularita × overhead
- nastavitelné v konfiguraci při kompilaci kernelu (`CONFIG_HZ`), default = 250
- dříve HZ=100
- mimo kernel (`/proc, times(),...`) HZ=100 (kvůli kompatibilitě)

```
version 2.6.9-1.667 (bhcompile@twetty.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef000 (usable)
BIOS-e820: 00000000001ef000 - 00000000001fec00 (ACPI data)
BIOS-e820: 00000000001fec00 - 00000000001ff000 (ACPI NVS)
BIOS-e820: 00000000001ff000 - 0000000000100000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec1000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee1000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000000ffe0000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
Highmem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC 0x06040000
ACPI: BOOT (v001 PTLTD $SBFTBL 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x3008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G transpeline to 4G
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:

```

# jiffies

- jiffy = moment, okamžik, minutka
- globální proměnná kernelu
- při každém tiku časovače inkrementována = počet tiků hodin
- nyní 64bitová verze (**jiffies\_64**), na většině architektur je 32bitová verze spodních 32 bitů z 64bitové
  - přístup k 64 bitům není atomický => používat zámek `xtime_lock` resp. `get_jiffies_64()`
- inicializována na **INITIAL\_JIFFIES** ~ -5 minut (aby wrapnula a projevily se buggy)

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee0000 - 0000000000fee10000 (reserved)
BIOS-e820: 0000000000ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x00000000
ACPI: DSDT (v001 PTLTD Custom 0x00000000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd=
mapped 4G/4G trampoline to f
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:

```

# RTC

- Real Time Clock
- hodiny, vestavěné v počítači (typicky v chipsetu)
- běží, i když je počítač vypnutý
- generují přerušení (IRQ 8) s nastavitelnou frekvencí
  - 2 Hz–8192 Hz, jen mocniny 2
- lze též generovat signál při každém updatu hodin (každou sekundu) a alarmy
- přerušení se předávají do `/dev/rtc`
  - znakové zařízení, jen pro čtení
  - unsigned longs, vždy informace o typu přerušení a o počtu přerušení daného typu od posledního čtení
- stavové informace v `/proc/driver/rtc`

```
version 2.6.9-1.667 (bhcompile@twetty.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef000 (usable)
BIOS-e820: 00000000001ef000 - 00000000001efc00 (ACPI data)
BIOS-e820: 00000000001efc00 - 00000000001ff000 (ACPI NVS)
BIOS-e820: 00000000001ff000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec00000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee00000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTBL $acpi
ACPI: DSDT (v001 PTLTD Custom 0x00000000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=
mapped 4G/4G transpale to 4G
Initializing CPU#0
CPU 0 irqstacks, hard-0
PID hash table entries:

```

# RTC – použití z user-space

- `#include \`  
`<linux/include/linux/rtc.h>`
- nastavení frekvence, času, alarmů a další funkcí přes `ioctl()`
  - frekvenci nad 64 Hz může nastavit jen root
  - konzervativní omezení – 486/33 MHz začne nestíhat až při 1024 Hz
- proces pak typicky zavolá `read()/select()` na `/dev/rtc`, spí dokud nenastane přerušení a přečte informace o něm
- při externí synchronizaci času stroje (např. přes NTP) se každých 11 minut přerušení na chvíli zablokuje => nutno s tím počítat

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fefc00 (ACPI data)
BIOS-e820: 0000000000fefc00 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee10000 (reserved)
BIOS-e820: 00000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:
hash=0x00000000
```

# Obsluha přerušení časovače

- `timer_interrupt()`
  - voláno se zakázanými přerušenými
  - skutečná práce se odvede v dalších volaných funkcích
    - broadcast přerušení ostatním procesorům v případě SMP
    - inkrementace jiffies
    - update systémového loadu
    - update časů u procesů
    - spuštění timerů

...

```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef000 (usable)
BIOS-e820: 00000000001ef000 - 00000000001efc00 (ACPI data)
BIOS-e820: 00000000001efc00 - 00000000001ff000 (ACPI NVS)
BIOS-e820: 00000000001ff000 - 0000000001000000 (usable)
BIOS-e820: 00000000001fec0000 - 00000000001fec1000 (reserved)
BIOS-e820: 00000000001fee0000 - 00000000001fee0100 (reserved)
BIOS-e820: 00000000001ffe0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTRM 0x00000000
ACPI: DSDT (v001 PTLTD Custom 0x00000000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard=0
PID hash table entries:
#####
```

# Kernel timers

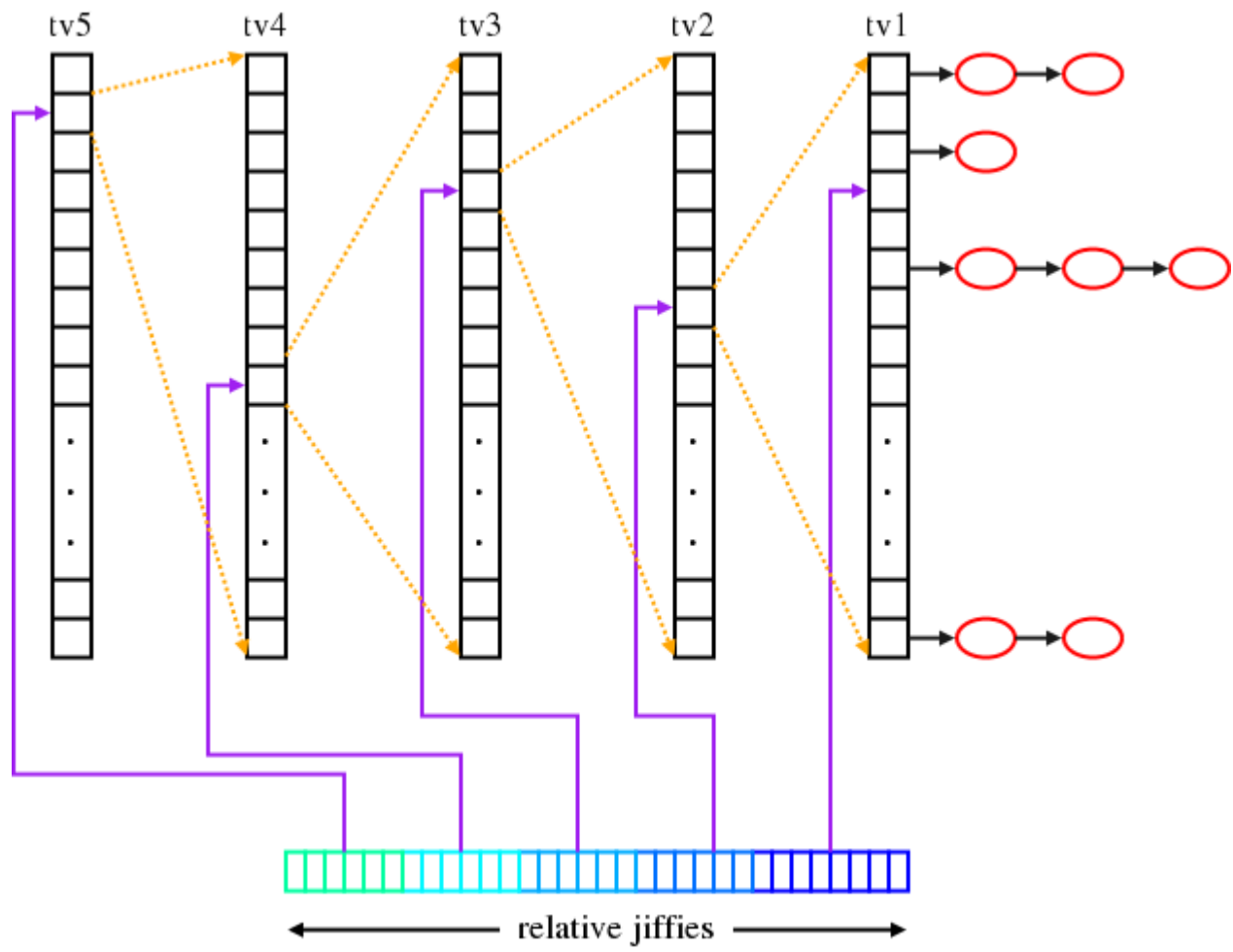
- timeout
  - jednou, netřeba přesně, hodně expiruje
- timer
  - víckrát, potřeba přesně, zřídka expiruje
- přesnost: podle HZ
- API:

```
extern void add_timer_on(struct timer_list *timer,
    int cpu);
static inline void add_timer(struct timer_list * timer)
extern int del_timer(struct timer_list * timer);
extern int mod_timer(struct timer_list *timer,
    unsigned long expires);

struct timer_list {
    ...
    unsigned long expires;
    void (*function)(unsigned long);
    unsigned long data;
    ...
};
```

```
version 2.6.9-1.667 (bhcompile@twetty.build.redhat.com) (gcc vers... 20041017 08:05:14
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 0000000000e40000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef0000 (usable)
BIOS-e820: 00000000001ef0000 - 00000000001fec000 (ACPI data)
BIOS-e820: 00000000001fec000 - 00000000001ff0000 (ACPI NVS)
BIOS-e820: 00000000001ff0000 - 0000000010000000 (usable)
BIOS-e820: 00000000fec00000 - 00000000fec10000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee01000 (reserved)
BIOS-e820: 00000000ffe00000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
HighMem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC MADT
ACPI: BOOT (v001 PTLTD $SBFTRM $acpi
ACPI: DSDT (v001 PTLTD Custom $acpi
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=RAW
mapped 4G/4G trampoline to ffff
Initializing CPU#0
CPU 0 irqstacks, hard-020
PID hash table entries:
.....
```

# Jak si kernel eviduje timery?



```
version 2.6.9-1.667 (bhcompile@tweety.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 0000000000fef000 (usable)
BIOS-e820: 0000000000fef000 - 0000000000fec000 (ACPI data)
BIOS-e820: 0000000000fec000 - 0000000000ff0000 (ACPI NVS)
BIOS-e820: 0000000000ff0000 - 0000000010000000 (usable)
BIOS-e820: 0000000000fec0000 - 0000000000fec10000 (reserved)
BIOS-e820: 0000000000fee00000 - 0000000000fee01000 (reserved)
BIOS-e820: 0000000000ffe00000 - 0000000010000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
Highmem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD
ACPI: BOOT (v001 PTLTD $SBFTRM 0x06040000
ACPI: DSDT (v001 PTLTD Custom 0x06040000
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd0
mapped 4G/4G trampoline to 0x00000000
Initializing CPU#0
CPU 0 irqstacks, hard=00000000
PID hash table entries: 1024
```

# BogoMIPS

- Bogus Millions of Instructions per Second
- „the number of million times per second a processor can do absolutely nothing“ :-)
- velmi hrubý odhad rychlosti procesoru
- používá se pro kalibraci délky interních smyček (busy loops)
- hodnota se měří a vypisuje při startu (fce `calibrate_delay`)
  - příklad na mém PC (Pentium M, 1.8 MHz, Linux ve VMWare):

```
[dmajda@localhost /]$ dmesg | grep BogoMIPS
Calibrating delay loop... 3637.24 BogoMIPS (lpj=1818624)
```

- lpj = loops per jiffy
- BogoMips mini-howto: <http://www.clifton.nl/index.html?bogomips.html>

```
version 2.6.9-1.667 (bhcompile@twetey.build.redhat.com) (gcc vers
14:41:25 EST 2004
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009f800 (usable)
BIOS-e820: 000000000009f800 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000dc000 - 00000000000e0000 (reserved)
BIOS-e820: 00000000000e4000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 00000000001ef0000 (usable)
BIOS-e820: 00000000001ef0000 - 00000000001efc000 (ACPI data)
BIOS-e820: 00000000001efc000 - 00000000001ff0000 (ACPI NVS)
BIOS-e820: 00000000001ff0000 - 00000000001000000 (usable)
BIOS-e820: 00000000001fec0000 - 00000000001fec10000 (reserved)
BIOS-e820: 00000000001fee0000 - 00000000001fee01000 (reserved)
BIOS-e820: 00000000001fee0000 - 0000000100000000 (reserved)
OMB HIGHMEM available.
256MB LOWMEM available.
zapping low mappings.
On node 0 totalpages: 65536
DMA zone: 4096 pages, LIFO batch:1
Normal zone: 61440 pages, LIFO batch:15
Highmem zone: 0 pages, LIFO batch:1
DMI present.
ACPI: RSDP (v000 PTLTD
ACPI: RSDT (v001 PTLTD RSDT 0x06040000
ACPI: FADT (v001 INTEL 440BX 0x06040000
ACPI: MADT (v001 PTLTD APIC
ACPI: BOOT (v001 PTLTD $SBFTBL $acpi
ACPI: DSDT (v001 PTLTD $dswtbl $acpi
ACPI: PM-Timer IO Port: 0x1008
Built 1 zonelists
Kernel command line: ro root=hd
mapped 4G/4G trampoline to 0
Initializing CPU#0
CPU 0 irqstacks, hard-0
PID hash table entries:
hashtable size: 0
```

# Odkud jsem čerpal

- `/usr/src/linux & grep :-)`
- Google, Google, Google
  - příliš mnoho zdrojů, abych je zde vyjmenoval
  - spousta stránek zastaralých, neoznačují popisovanou verzi jádra, často protiřečící si informace
  - na druhou stranu se toho dá najít překvapivě hodně
  - dobrý zdroj: archivy různých mailing-listů, komentáře v mailech k patchům