

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2006-08/msg10680.html>

- *From:* Mel Gorman <mel@xxxxxxxxx>
 - *Date:* Thu, 31 Aug 2006 17:17:52 +0100 (IST)
-

On Thu, 31 Aug 2006, Paul Jackson wrote:

The following patch in 2.6.18-rc4-mm3 is broken on my x86_64:

`account-for-memmap-and-optionally-the-kernel-image-as-holes.patch`

The failure is 100% reproducible.

Ok, I'm suprised that it is this patch that causes a problem. I felt the patch would either explode everywhere or just work.

The system has a pair of dual-core Intel Xeon 5100 series (Woodcrest) processors (4 logical CPUs total) and 2 GBytes of ram.

The .config is what one gets from 'make defconfig' for arch x86_64, plus the following changes:

```
===== begin =====
--- .config.def 2006-08-31 04:29:22.100311614 -0500
+++ .config 2006-08-31 04:29:03.247761750 -0500
@@ -1,7 +1,7 @@
#
# Automatically generated make config: don't edit
# Linux kernel version: 2.6.18-rc4-mm3
-# Thu Aug 31 04:29:22 2006
+# Thu Aug 31 04:07:54 2006
#
CONFIG_X86_64=y
CONFIG_64BIT=y
@@ -44,7 +44,7 @@
# CONFIG_AUDIT is not set
CONFIG_IKCONFIG=y
CONFIG_IKCONFIG_PROC=y
-# CONFIG_CPUSETS is not set
```

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

```
+CONFIG_CPUSETS=y
# CONFIG_RELAY is not set
CONFIG_INITRAMFS_SOURCE=""
CONFIG_UID16=y
@@ -205,7 +205,7 @@
# CONFIG_ACPI_ASUS is not set
# CONFIG_ACPI_IBM is not set
# CONFIG_ACPI_TOSHIBA is not set
-CONFIG_ACPI_SONY=m
+# CONFIG_ACPI_SONY is not set
CONFIG_ACPI_BLACKLIST_YEAR=0
# CONFIG_ACPI_DEBUG is not set
CONFIG_ACPI_EC=y
@@ -1270,7 +1270,11 @@
# CONFIG_REISERFS_FS_SECURITY is not set
# CONFIG_JFS_FS is not set
CONFIG_FS_POSIX_ACL=y
-# CONFIG_XFS_FS is not set
+CONFIG_XFS_FS=y
+# CONFIG_XFS_QUOTA is not set
+# CONFIG_XFS_SECURITY is not set
+# CONFIG_XFS_POSIX_ACL is not set
+# CONFIG_XFS_RT is not set
# CONFIG_GFS2_FS is not set
# CONFIG_OCFS2_FS is not set
# CONFIG_MINIX_FS is not set
===== end =====
```

Nothing very suprising there.

The boot fails with the following console output:

ok, this is interesting. It appears that the log is truncated or somehow corrupt.

```
===== begin =====
root (hd0,0)
Filesystem type is ext2fs, partition type 0x83
kernel /vmlinuz.pj2 root=/dev/sda3 console=ttyS1,115200 showopts pj2
[Linux-bzImage, setup=0x1c00, size=0x2b66e5]

Linux version 2.6.18-rc4-mm3 (pj@spandau) (gcc version 4.1.0 (SUSE Linux)) #48 SMP
Thu Aug 31 04:22:41 CDT 2006
Command line: root=/dev/sda3 console=ttyS1,115200 showopts pj2
BIOS-provided physical RAM map:
```

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

BIOS-e820: 0000000000000000 – 000000000009f000 (usable)
BIOS-e820: 000000000009f000 – 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 – 000000007f932000 (usable)
BIOS-e820: 000000007f932000 – 000000007f9d0(ACPI NVS)

Little bit missing here. I don't expect 000000007f9d0 to be truncated like that.

BIOS-e820: 000000007f9d0000 – 000000007fa42000 (usable)
BIOS-e820: 000000007fa42000 – 000000007fb2b000 (usable)

or 000000007fa42000 to have an additional 0 at the end.

BIOS-e820: 000000007fb2b000 – 000000007fb3a000 (ACPI data)
B0000000000–000000007fc00000
Bootmem setup node 0 0000000000000000–000000007fc00000

and this seems to interleave even though the bootmem setup node range would match your physical memory.

Zone PFN raProcessor #0 (Bootup-CPU)

There is information missing here. That should be Zone PFN Ranges followed by a list of active PFN ranges from your system. After that, I expect to see a message like

X pages DMA reserved
Y pages used for memmap

Do you think this is a problem with your serial console or something else? Do you see the Zone PFN ranges information when the patch is backed out? Those messages, as well as booting with loglevel=8 would really help me figure out what went pear shaped.

ACPI: LAPIC (acpi_id[0x01] lapic_id[0x06] enabled)
Processor #6
ACPapic_id[0x85] disabled)
ACPI: LAPIC (acpi_id[0x06] lapic_id[0x86] disabled)
ACPI: LAPIC (acpi_x02] high level lint[0x1])
ACPI: LAPIC_NMI (acpi_id[0x03] high level lint[0x1])
ACPI: LAPIC_NM0x08] address[0xfec00000] gsi_base[0])
IOAPIC[0]: apic_id 8, address 0xfec00000, GSI 0–23
ACPI0x0b] address[0xfec84400] gsi_base[72])
IOAPIC[3]: apic_id 11, address 0xfec84400, GSI 72–95
AUsing ACPI (MADT) for SMP configuration information
Allocating PCI resources starting at 800000ot=/dev/sda3 console=ttyS1,115200 showopts pj2
Initializing CPU#0

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

PID hash table entries: 40962 (order: 8, 1048576 bytes)
Checking aperture...
Memory: 2052128k/2093056k available (3519k kerved, 2323k data, 280k init)
Calibrating delay using timer specific routine.. 5324.66 BogoMIPS (lpj=10649332)
Mount-cache hash table entries: 256
CPU: L1 I cache: 32K, L1 D cache: 32K
CPU: L2 cache: 4096K
CPU 0/0 -> Node 0
using mwait in idle threads.
CPU: Physical Processor ID: 0
CPU: Processor Core ID: 0
CPU0: Thermal monitoring enabled (TM2)
SMP alternatives: switching to UP code
ACPI: Core revision 20060707
Using local APIC timer interrupts.
result 20781304
Detected 20.781 MHz APIC timer.
SMP alternatives: switching to SMP code
Booting processor 1/4 APIC 0x6
Initializing CPU#1
Calibrating delay using timer specific routine.. 5320.16 BogoMIPS (lpj=10640330)
CPU: L1 I cache: 32K, L1 D cache: 32K
CPU: L2 cache: 4096K
CPU 1/6 -> Node 0
CPU: Physical Processor ID: 3
CPU: Processor Core ID: 0
CPU1: Thermal monitoring enabled (TM2)
Genuine Intel(R) CPU @ 2.66GHz stepping 04
SMP alternatives: switching to SMP code
Booting processor 2/4 APIC 0x1
Initializing CPU#2
Calibrating delay using timer specific routine.. 5320.16 BogoMIPS (lpj=10640332)
CPU: L1 I cache: 32K, L1 D cache: 32K
CPU: L2 cache: 4096K
CPU 2/1 -> Node 0
CPU: Physical Processor ID: 0
CPU: Processor Core ID: 1
CPU2: Thermal monitoring enabled (TM2)
Genuine Intel(R) CPU @ 2.66GHz stepping 04
SMP alternatives: switching to SMP code
Booting processor 3/4 APIC 0x7
Initializing CPU#3
Calibrating delay using timer specific routine.. 5320.04 BogoMIPS (lpj=10640092)
CPU: L1 I cache: 32K, L1 D cache: 32K
CPU: L2 cache: 4096K
CPU 3/7 -> Node 0
CPU: Physical Processor ID: 3
CPU: Processor Core ID: 1
CPU3: Thermal monitoring enabled (TM2)
Genuine Intel(R) CPU @ 2.66GHz stepping 04
Brought up 4 CPUs

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

testing NMI watchdog ... OK.
time.c: Using 14.318180 MHz WALL HPET GTOD HPET/TSC timer.
time.c: Detected 2660.007 MHz processor.
migration_cost=30,7937
NET: Registered protocol family 16
ACPI: bus type pci registered
PCI: Using MMCONFIG at a0000000
ACPI: Interpreter enabled
ACPI: Using IOAPIC for interrupt routing
ACPI: PCI Root Bridge [PCI0] (0000:00)
PCI: Ignoring BAR0-3 of IDE controller 0000:00:1f.1
PCI: PXH quirk detected, disabling MSI for SHPC device
PCI: PXH quirk detected, disabling MSI for SHPC device
PCI: Transparent bridge - 0000:00:1e.0
ACPI: PCI Interrupt Link [LNKA] (IRQs 5 7 *10 11)
ACPI: PCI Interrupt Link [LNKB] (IRQs 5 7 10 *11)
ACPI: PCI Interrupt Link [LNKC] (IRQs 5 7 *10 11)
ACPI: PCI Interrupt Link [LNKD] (IRQs *5 7 10 11)
ACPI: PCI Interrupt Link [LNKE] (IRQs *5 7 10 11)
ACPI: PCI Interrupt Link Intel 82802 RNG detected
SCSI subsystem initialized
usbcore: registered new interface driver uirq". If it helps, post a report
hpet0: at MMIO 0xfed00000, IRQs 2, 8, 0
hpet0: 3 64-bit timeow: b8b00000-b8bffff
PCI: Bridge: 0000:03:00.2
IO window: disabled.
MEM window: disabled.
MEM window: disabled.
PREFETCH window: disabled.
PCI: Bridge: 0000:02:02.0
IO window disabled.
MEM window: disabled.
PREFETCH window: disabled.
PCI: Bridge: 0000:00:02.0
IO window: disabled.
MEM window: disabled.
PREFETCH window: disabled.
PCI: Bridge: 0000:00:05.0c:00.2
IO window: disabled.
MEM window: disabled.
PREFETCH window: disabled.
PCI: Bridge: 0000:00:1e.0
IO window: 1000-1fff
MEM window: b8c00000-0000:01:00.0[A] -> GSI 16 (level, low) -> IRQ 169
ACPI: PCI Interrupt 0000:02:00.0[A] - IRQ 169
ACPI: PCI Interrupt 0000:00:03.0[A] -> GSI 16 (level, low) -> IRQ 169
ACPI: PCI Interrupt 0000:00:07.0[A] -> GSI 16 (level, low) -> IRQ 169
IP route cache hash table entries: 65536 (order: 7, 524288 bytes)
TCP established hash table entries: 262144 (order: 10, 4194304 bytes)
TCP bind hash table entries: 65536 (order: 8, 1048576 bytes) Hash tables configured

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

```
(established 262144 bind 65536)
TCP reno registered
Total HugeTLB io scheduler noop registered
io scheduler deadline registered
io scheduler cfq registered (def0000:00:1d.7 EHCI: BIOS handoff failed (BIOS bug ?)
01010001
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
assign_interrupt_mode Found MSI capability
aer: probe of 0000:00:02.0:pcie01 failed with error 2
aer: probe of 0000:00:03.0:pcie01 failed with error 1
aer: probe of 0000:00:04.0:pcie01 failed failed with error 2
aer: probe of 0000:00:07.0:pcie01 failed with error 2
ACPI: Power Button r Device is not present [20060707]
ACPI: Getting cpuindex for acpiid 0x4
ACPI Exception (acpi_060707)
ACPI: Getting cpuindex for acpiid 0x6
ACPI Exception (acpi_processor-0681): AE_NOT_FOUReal Time Clock Driver v1.12ac
Linux apgart interface v0.101 (c) Dave Jones
Serial: 8250/1655/O 0x2f8 (irq = 3) is a 16550A
floppy0: no floppy controllers found
RAMDISK driver initialized: 16 RAM disks of 4096K size 1024 blocksize
loop: loaded (max 8 devices)
IntI 17 sharing vector 0x42 and IRQ 17
ACPI: PCI Interrupt 0000:07:00.0[A] -> GSI 18 (level, low) -> IRQ 66
e1000: 0000:07:00.0: e1000_probe: (PCI Express:2.5Gb/s:Width x4) 00:04:23:cf:2d:d2
e1000: eth0: e1000_probe: Intel(R) PRO/1000 Network Connection
GSI 18 sharing vector 0x4A and IRQ 18
ACPI: PCI Interrupt 0000:07:00.1[B] -> GSI 19 (level, low) -> IRQ 74
e1000: 0000:07:00.1:ide0 at 0x1f0-0x1f7,0x3f6 on irq 14
hda: ATAPI 24X DVD-ROM drive, 256kB Cache, UDMA(33)
Uniform CD-ROM driver Revision: 3.20
megaraid cmm: 2.20.2.7 (Release Date: Sun Jul 16 00:01:03 EST 2006)
megaraid: 2.20.4.9 (Release Date: Sun Jul 16 12:27:22 EST 2006)
megasas: 00.00.03.01 Sun May 14 22:49:52 PDT 2006
megasas: 0x1000:0x0411:0x8086:0x3501: bus 4:slot 14:func 0
ACPI: PCI Interrupt 0000:04:0e.0[A] -> GSI 18 (level, low) -> IRQ 66
scsi0 : LSI Logic SAS based MegaRAID driver
scsi 0:0:0:0: Direct-Access ATA HDT722525DLA380 A80A PQ: 0 ANSI: 5
scsi 0:0:1:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:0:2:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:0:3:0: Direct-Access Scsi 0:2:0:0: Direct-Access INTEL SROMBASAS18E 1.00 PQ:
0 ANSI: 5
scsi 0:2:1:0: Direswapper invoked oom-killer: gfp_mask=0xd1, order=0, oomkilladj=0
```

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Call Trace:

```
[<ffffffff802025bc67>] __alloc_pages+0x229/0x2b2
[<ffffffff80274e46>] cache_grow+0x134/0x333
[<ffffffff802really_probe+0x47/0xc9
[<ffffffff803eea20>] __driver_attach+0x6f/0xaf
[<ffffffff803ee214>] bfffffff803abf12>] acpi_ds_init_one_object+0x0/0x82
[<ffffffff80207046>] init+0x0/0x306
[<ffu 0 hot: high 186, batch 31 used:24
cpu 0 cold: high 62, batch 15 used:0
cpu 1 hot: high 186, 15 used:0
Node 0 Normal per-cpu: empty
Active:0 inactive:0 dirty:0 writeback:0 unstable:0 freeB high:0kB active:0kB inactive:0kB
present:0kB pages_scanned:0 all_unreclaimable? no
lowmem_res 0*2048kB 496*4096kB = 2035560kB
```

This is also garbled up. This is in show_free_areas() though and it looks like it is saying there are 496*4096kB pages currently free. Not clear at all how it managed to go OOM due to this patch.

Node 0 Normal: empty

Swap cache: add 0, delete 0, find 0/0, r swap cached

Kernel panic – not syncing: Out of memory and no killable processes...

===== end =====

Without this bad patch, the system boot continues with the following messages, slightly overlapping my presentation with the above output:

===== begin =====

...

```
ACPI: PCI Interrupt 0000:04:0e.0[A] -> GSI 18 (level, low) -> IRQ 66
scsi0 : LSI Logic SAS based MegaRAID driver
scsi 0:0:0:0: Direct-Access ATA HDT722525DLA380 A80A PQ: 0 ANSI: 5
scsi 0:0:1:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:0:2:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:0:3:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:0:4:0: Direct-Access ATA HDS725050KLA360 AB0A PQ: 0 ANSI: 5
scsi 0:2:0:0: Direct-Access INTEL SROMBSAS18E 1.00 PQ: 0 ANSI: 5
scsi 0:2:1:0: Direct-Access INTEL SROMBSAS18E 1.00 PQ: 0 ANSI: 5
SCSI devi: write through
SCSI device sda: 486326272 512-byte hdwr sectors (248999 MB)
sda: test WP fail sda1 sda2 sda3
sd 0:2:0:0: Attached scsi disk sda
SCSI device sdb: 2923825152 512-byte hdwr s assuming drive cache: write through
SCSI device sdb: 2923825152 512-byte hdwr sectors (1496998 sdb1
sd 0:2:1:0: Attached scsi disk sdb
sd 0:2:0:0: Attached scsi generic sg0 type 0
sd 0:2:aw1394: /dev/raw1394 device initialized
GSI 20 sharing vector 0x5A and IRQ 20
```

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

Re: x86_64 account-for-memmap patch in 2.6.18-rc4-mm3 doesn't boot.

ACPI: PCI Interrupt 7: debug port 1

...

===== end =====

I won't rest till it's the best ...
Programmer, Linux Scalability
Paul Jackson <pj@xxxxxxx> 1.925.600.0401

Can I see a full bootlog with the patch backed out to see if that console garbling is still there please? Have you any idea why the console garbling is happening?

Thanks

Mel Gorman
Part-time Phd Student Linux Technology Center
University of Limerick IBM Dublin Software Lab

—
To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxxx
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>