

Have I broken my RAID5 by simulating a disk failure?

Source: <http://linux.derkeiler.com/Newsgroups/comp.os.linux.questions/2003-10/0109.html>

From: Brian Marriage (brian.marriage_at_snellwilcox.com)

Date: 10/14/03

Date: 14 Oct 2003 06:36:30 -0700

By merely **simulating** disk failures, I seem to have screwed up my new RAID...

I've recently put together my first software RAID5 on RH9 using 4 disks, following the simple instructions in the Samba HOWTO on www.tldp.org. I got it working, put my data on it. Then I wanted to make sure it could withstand any (single) disk failing. So I (again) followed the instructions: Powered down, unplugged a disk, restarted. Some messages about a disk missing, but everything still worked (as expected). Powered down, reconnected disk (/dev/hde), restarted, and then used `raidhotadd /dev/hde`. Seemed happy. Now to try the next disk - /dev/hdc. I now think this is maybe where I went wrong - I think I should have let it spend 3 hours reconstruct everything, but I just assumed that it had nothing to do: The data would still be there - after all, I only **simulated** a failure...

Powered up with /dev/hdc disconnected. RAID not at all happy. Heart sinks. Now what? Lots of fiddling around (non-destructive fiddling, I'm fairly confident!), and no progress. It now seems to think that /dev/hde is a spare disk! I now turn to the ever-helpful community (am I grovelling enough?) in the hope someone can help (please?)

I now have 2 disks that the RAID thinks are dead, but I would have thought were both still fine if only I could persuade it to use them. Is there something I can do? Here is a dump of `dmesg` after a boot-up (all disks connected):

```
Linux version 2.4.20-8 (bhcompile@porky.devel.redhat.com) (gcc version 3.2.2
20030222 (Red Hat Linux 3.2.2-5)) #1 Thu Mar 13 17:54:28 EST 2003
BIOS-provided physical RAM map:
BIOS-e820: 0000000000000000 - 000000000009fc00 (usable)
BIOS-e820: 000000000009fc00 - 00000000000a0000 (reserved)
BIOS-e820: 00000000000f0000 - 0000000000100000 (reserved)
BIOS-e820: 0000000000100000 - 000000001fff0000 (usable)
BIOS-e820: 000000001fff0000 - 000000001fff3000 (ACPI NVS)
```

comp.os.linux.questions: Have I broken my RAID5 by simulating a disk failure?

```
BIOS-e820: 000000001fff3000 - 0000000020000000 (ACPI data)
BIOS-e820: 00000000fec00000 - 00000000fec01000 (reserved)
BIOS-e820: 00000000fee00000 - 00000000fee01000 (reserved)
BIOS-e820: 00000000ffff0000 - 0000000100000000 (reserved)
0MB HIGHMEM available.
511MB LOWMEM available.
On node 0 totalpages: 131056
zone(0): 4096 pages.
zone(1): 126960 pages.
zone(2): 0 pages.
Kernel command line: ro root=LABEL=/
Initializing CPU#0
Detected 551.255 MHz processor.
Console: colour VGA+ 80x25
Calibrating delay loop... 1101.00 BogoMIPS
Memory: 511232k/524224k available (1347k kernel code, 10428k reserved, 999k
data, 132k init, 0k highmem)
Dentry cache hash table entries: 65536 (order: 7, 524288 bytes)
Inode cache hash table entries: 32768 (order: 6, 262144 bytes)
Mount cache hash table entries: 512 (order: 0, 4096 bytes)
Buffer-cache hash table entries: 32768 (order: 5, 131072 bytes)
Page-cache hash table entries: 131072 (order: 7, 524288 bytes)
CPU: L1 I cache: 16K, L1 D cache: 16K
CPU: L2 cache: 128K
Intel machine check architecture supported.
Intel machine check reporting enabled on CPU#0.
CPU: After generic, caps: 0183fbff 00000000 00000000 00000000
CPU: Common caps: 0183fbff 00000000 00000000 00000000
CPU: Intel Celeron (Mendocino) stepping 05
Enabling fast FPU save and restore... done.
Checking 'hlt' instruction... OK.
POSIX conformance testing by UNIFIX
mtrr: v1.40 (20010327) Richard Gooch (rgooch@atnf.csiro.au)
mtrr: detected mtrr type: Intel
PCI: PCI BIOS revision 2.10 entry at 0xfb5c0, last bus=1
PCI: Using configuration type 1
PCI: Probing PCI hardware
PCI: Using IRQ router PIIX [8086/7110] at 00:07.0
Limiting direct PCI/PCI transfers.
isapnp: Scanning for PnP cards...
isapnp: No Plug & Play device found
Linux NET4.0 for Linux 2.4
Based upon Swansea University Computer Society NET3.039
Initializing RT netlink socket
apm: BIOS version 1.2 Flags 0x07 (Driver version 1.16)
Starting kswapd
VFS: Disk quotas vquot_6.5.1
pty: 2048 Unix98 ptys configured
Serial driver version 5.05c (2001-07-08) with MANY_PORTS MULTIPORT SHARE_IRQ
SERIAL_PCI ISAPNP enabled
ttyS0 at 0x03f8 (irq = 4) is a 16550A
```

Have I broken my RAID5 by simulating a disk failure?

comp.os.linux.questions: Have I broken my RAID5 by simulating a disk failure?

```
ttyS1 at 0x02f8 (irq = 3) is a 16550A
Real Time Clock Driver v1.10e
Floppy drive(s): fd0 is 1.44M
FDC 0 is a post-1991 82077
NET4: Frame Diverter 0.46
RAMDISK driver initialized: 16 RAM disks of 4096K size 1024 blocksize
Uniform Multi-Platform E-IDE driver Revision: 7.00beta-2.4
ide: Assuming 33MHz system bus speed for PIO modes; override with idebus=xx
PIIX4: IDE controller at PCI slot 00:07.1
PIIX4: chipset revision 1
PIIX4: not 100% native mode: will probe irqs later
  ide0: BM-DMA at 0xf000-0xf007, BIOS settings: hda:DMA, hdb:pio
SiI3112 Serial ATA: IDE controller at PCI slot 00:09.0
PCI: Found IRQ 10 for device 00:09.0
PCI: Sharing IRQ 10 with 00:07.2
SiI3112 Serial ATA: chipset revision 2
SiI3112 Serial ATA: not 100% native mode: will probe irqs later
  ide1: MMIO-DMA at 0xe080d000-0xe080d007, BIOS settings: hdc:pio, hdd:pio
  ide2: MMIO-DMA at 0xe080d008-0xe080d00f, BIOS settings: hde:pio, hdf:pio
HPT366: onboard version of chipset, pin1=1 pin2=2
HPT366: IDE controller at PCI slot 00:13.0
PCI: Found IRQ 15 for device 00:13.0
PCI: Sharing IRQ 15 with 00:0b.0
PCI: Sharing IRQ 15 with 00:13.1
HPT366: chipset revision 1
HPT366: not 100% native mode: will probe irqs later
  ide3: BM-DMA at 0xd800-0xd807, BIOS settings: hdg:DMA, hdh:DMA
PCI: Found IRQ 15 for device 00:13.1
PCI: Sharing IRQ 15 with 00:0b.0
PCI: Sharing IRQ 15 with 00:13.0
  ide4: BM-DMA at 0xe400-0xe407, BIOS settings: hdi:pio, hdj:pio
hda: QUANTUM FIREBALLct15 15, ATA DISK drive
blk: queue c03c9f40, I/O limit 4095Mb (mask 0xffffffff)
hdc: ST3120026AS, ATA DISK drive
hde: ST3120026AS, ATA DISK drive
hdg: IC35L120AVV207-0, ATA DISK drive
hdh: IC35L120AVV207-0, ATA DISK drive
blk: queue c03cac60, I/O limit 4095Mb (mask 0xffffffff)
blk: queue c03cada4, I/O limit 4095Mb (mask 0xffffffff)
ide0 at 0x1f0-0x1f7,0x3f6 on irq 14
ide1 at 0xe080d080-0xe080d087,0xe080d08a on irq 10
ide2 at 0xe080d0c0-0xe080d0c7,0xe080d0ca on irq 10
ide3 at 0xd000-0xd007,0xd402 on irq 15
hda: host protected area => 1
hda: 29336832 sectors (15020 MB) w/418KiB Cache, CHS=1826/255/63, UDMA(33)
hdc: host protected area => 1
hdc: 234441648 sectors (120034 MB) w/8192KiB Cache, CHS=14593/255/63
hde: host protected area => 1
hde: 234441648 sectors (120034 MB) w/8192KiB Cache, CHS=14593/255/63
hdg: host protected area => 1
hdg: 241254720 sectors (123522 MB) w/1821KiB Cache, CHS=15017/255/63,
```

Have I broken my RAID5 by simulating a disk failure?

comp.os.linux.questions: Have I broken my RAID5 by simulating a disk failure?

```
UDMA(66)
hdh: host protected area => 1
hdh: 241254720 sectors (123522 MB) w/1821KiB Cache, CHS=15017/255/63,
UDMA(66)
ide-floppy driver 0.99.newide
Partition check:
hda: hda1 hda2 hda3
hdc: hdc1
hde: hde1
hdg: hdg1
hdh: hdh1
ide-floppy driver 0.99.newide
md: md driver 0.90.0 MAX_MD_DEVS=256, MD_SB_DISKS=27
md: Autodetecting RAID arrays.
[events: 00000017]
[events: 00000019]
[events: 00000019]
[events: 00000019]
md: autorun ...
md: considering hdh1 ...
md: adding hdh1 ...
md: adding hdg1 ...
md: adding hde1 ...
md: adding hdc1 ...
md: created md0
md: bind<hdc1,1>
md: bind<hde1,2>
md: bind<hdg1,3>
md: bind<hdh1,4>
md: running: <hdh1><hdg1><hde1><hdc1>
md: hdh1's event counter: 00000019
md: hdg1's event counter: 00000019
md: hde1's event counter: 00000019
md: hdc1's event counter: 00000017
md: superblock update time inconsistency -- using the most recent one
md: freshest: hdh1
md: kicking non-fresh hdc1 from array!
md: unbind<hdc1,3>
md: export_rdev(hdc1)
md0: removing former faulty hdc1!
kmod: failed to exec /sbin/modprobe -s -k md-personality-4, errno = 2
md: personality 4 is not loaded!
md :do_md_run() returned -22
md: md0 stopped.
md: unbind<hdh1,2>
md: export_rdev(hdh1)
md: unbind<hdg1,1>
md: export_rdev(hdg1)
md: unbind<hde1,0>
md: export_rdev(hde1)
md: ... autorun DONE.
```

Have I broken my RAID5 by simulating a disk failure?

comp.os.linux.questions: Have I broken my RAID5 by simulating a disk failure?

NET4: Linux TCP/IP 1.0 for NET4.0
IP Protocols: ICMP, UDP, TCP, IGMP
IP: routing cache hash table of 4096 buckets, 32Kbytes
TCP: Hash tables configured (established 32768 bind 65536)
Linux IP multicast router 0.06 plus PIM-SM
NET4: Unix domain sockets 1.0/SMP for Linux NET4.0.
RAMDISK: Compressed image found at block 0
Freeing initrd memory: 261k freed
VFS: Mounted root (ext2 filesystem).
SCSI subsystem driver Revision: 1.00
PCI: Found IRQ 15 for device 00:0b.0
PCI: Sharing IRQ 15 with 00:13.0
PCI: Sharing IRQ 15 with 00:13.1
scsi0 : AdvanSys SCSI 3.3G: PCI Ultra-Wide: PCIMEM 0xE0848000-0xE084803F, IRQ 0xF
Journalled Block Device driver loaded
kjournald starting. Commit interval 5 seconds
EXT3-fs: mounted filesystem with ordered data mode.
Freeing unused kernel memory: 132k freed
usb.c: registered new driver usbdevfs
usb.c: registered new driver hub
usb-uhci.c: \$Revision: 1.275 \$ time 17:59:01 Mar 13 2003
usb-uhci.c: High bandwidth mode enabled
PCI: Found IRQ 10 for device 00:07.2
PCI: Sharing IRQ 10 with 00:09.0
usb-uhci.c: USB UHCI at I/O 0xb000, IRQ 10
usb-uhci.c: Detected 2 ports
usb.c: new USB bus registered, assigned bus number 1
hub.c: USB hub found
hub.c: 2 ports detected
usb-uhci.c: v1.275:USB Universal Host Controller Interface driver
usb.c: registered new driver hiddev
usb.c: registered new driver hid
hid-core.c: v1.8.1 Andreas Gal, Vojtech Pavlik <vojtech@suse.cz>
hid-core.c: USB HID support drivers
mice: PS/2 mouse device common for all mice
EXT3 FS 2.4-0.9.19, 19 August 2002 on ide0(3,2), internal journal
Adding Swap: 1044216k swap-space (priority -1)
kjournald starting. Commit interval 5 seconds
EXT3 FS 2.4-0.9.19, 19 August 2002 on ide0(3,1), internal journal
EXT3-fs: mounted filesystem with ordered data mode.
ip_tables: (C) 2000-2002 Netfilter core team
natsemi dp8381x driver, version 1.07+LK1.0.17, Sep 27, 2002
originally by Donald Becker <becker@scyld.com>
<http://www.scyld.com/network/natsemi.html>
2.4.x kernel port by Jeff Garzik, Tjeerd Mulder
PCI: Found IRQ 11 for device 00:0f.0
divert: allocating divert_blk for eth0
eth0: NatSemi DP8381[56] at 0xe091e000, 00:09:5b:20:19:0a, IRQ 11.
ip_tables: (C) 2000-2002 Netfilter core team
eth0: link up.

Have I broken my RAID5 by simulating a disk failure?

And this is what dmesg adds after a raidstart /dev/md0 (which, incidentally, returns silently):

```
[events: 00000017]
[events: 00000019]
[events: 00000019]
[events: 00000019]
md: autorun ...
md: considering hde1 ...
md: adding hde1 ...
md: adding hdh1 ...
md: adding hdg1 ...
md: adding hdc1 ...
md: created md0
md: bind<hdc1,1>
md: bind<hdg1,2>
md: bind<hdh1,3>
md: bind<hde1,4>
md: running: <hde1><hdh1><hdg1><hdc1>
md: hde1's event counter: 00000019
md: hdh1's event counter: 00000019
md: hdg1's event counter: 00000019
md: hdc1's event counter: 00000017
md: superblock update time inconsistency -- using the most recent one
md: freshest: hde1
md: kicking non-fresh hdc1 from array!
md: unbind<hdc1,3>
md: export_rdev(hdc1)
md0: removing former faulty hdc1!
raid5: measuring checksumming speed
 8regs : 1018.000 MB/sec
32regs : 519.200 MB/sec
8regs_prefetch: 1018.000 MB/sec
32regs_prefetch: 519.600 MB/sec
pII_mmx : 1254.000 MB/sec
p5_mmx : 1327.200 MB/sec
raid5: using function: p5_mmx (1327.200 MB/sec)
md: raid5 personality registered as nr 4
md0: max total readahead window set to 744k
md0: 3 data-disks, max readahead per data-disk: 248k
raid5: spare disk hde1
raid5: device hdh1 operational as raid disk 3
raid5: device hdg1 operational as raid disk 2
raid5: not enough operational devices for md0 (2/4 failed)
RAID5 conf printout:
--- rd:4 wd:2 fd:2
disk 0, s:0, o:0, n:0 rd:0 us:1 dev:[dev 00:00]
disk 1, s:0, o:0, n:1 rd:1 us:1 dev:[dev 00:00]
```

comp.os.linux.questions: Have I broken my RAID5 by simulating a disk failure?

```
disk 2, s:0, o:1, n:2 rd:2 us:1 dev:hdg1
disk 3, s:0, o:1, n:3 rd:3 us:1 dev:hdh1
raid5: failed to run raid set md0
md: pers->run() failed ...
md :do_md_run() returned -22
md: md0 stopped.
md: unbind<hde1,2>
md: export_rdev(hde1)
md: unbind<hdh1,1>
md: export_rdev(hdh1)
md: unbind<hdg1,0>
md: export_rdev(hdg1)
md: ... autorun DONE.
```

Lastly, my /etc/raidtab:

```
raiddev /dev/md0
    raid-level 5
    nr-raid-disks 4
    nr-spare-disks 0
    persistent-superblock 1
    parity-algorithm left-symmetric
    chunk-size 32
    device /dev/hdc 1
    raid-disk 0
    device /dev/hde 1
    raid-disk 1
    device /dev/hdg 1
    raid-disk 2
    device /dev/hdh 1
    raid-disk 3
```

I really hope that someone can help me. I have a copy of 3/4 of the data on this raid. It'd be a real bumner to lose the rest. Please help!

Thanks,
Brian