

Re: pitiful performance of an SATA150 drive

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/stable/2007-03/msg00736.html>

- *From:* Jeremy Chadwick <koitsu@xxxxxxxxxxxx>
 - *Date:* Mon, 26 Mar 2007 12:22:51 -0700
-

On Mon, Mar 26, 2007 at 02:36:27PM -0400, Mikhail Teterin wrote:

Over a year later this remains a problem -- exactly as described below...

No other SATA devices are present -- the only other IDE device is the DVD drive. My main disks are SCSI.

What's MUCH worse is that the (slowly) written data is also often corrupted... I use the drive to store our vast collection of photos and the backups. Every once in a while I encounter a corrupt JPEG file, and the backups are always corrupt somewhere. Doing something like:

```
dump 0auChf 16 0 - /home | bzip2 -9 > /store/home.0.bz2
```

always produces a corrupt file (as per ``bzip2 -t"). I used to blame the drive's temperature, but it now sits in its own enclosure and stays under 40 Celsius.

I can't reproduce the corruption you report. I run massive backups (7 levels; level 0 on Sunday, 1-6 on Mon-Sat) in our co-location facility and have always had success with restore(8). We use gzip -2 not bzip2, for what it's worth. The dumps are done over SSH.

When the drive is accessed, there are (according to `systat -vm') many thousands of interrupts 17 -- on my system these are shared between pcm0 and ehci0. Why are these triggered by accessing SATA is unclear, but the Intr's share of the CPU time is often above 80% of one processor's total (I have 4 processors).

See below for some of my stats for comparison.

As I mentioned a year ago, Knoppix was accessing the same drive at much higher speeds, so I don't believe, the problem is with the hardware...

Please, advise. Thanks!

Re: pitiful performance of an SATA150 drive

Let's compare numbers and devices, since I use SATA devices exclusively on my own home network, as well as in both of my production co-los. I'll use my home network for the below tests.

Here's the SATA controller I'm using (on-board nVidia):

```
atapci2: <nVidia nForce CK804 SATA300 controller> port
0x9e0-0x9e7,0xbe0-0xbe3,0x960-0x967,0xb60-0xb63,0xc400-0xc40f mem 0xd3001000-0xd3001fff irq
21 at device 8.0 on pci0
ata4: <ATA channel 0> on atapci2
ata5: <ATA channel 1> on atapci2
ad8: 190782MB <WDC WD2000JD-00HBB0 08.02D08> at ata4-master SATA150
ad10: 476940MB <Seagate ST3500630AS 3.AAD> at ata5-master SATA300
```

The motherboard itself is an Asus A8N-E with an AMD Athlon 64 X2 3800+ in it. Kernel is built with SMP. No overclocking.

Data taken from smartctl (ports/sysutils/smartmontools); I'm including this because it shows general information about the drives.

```
=== START OF INFORMATION SECTION ===
Model Family: Western Digital Caviar SE (Serial ATA) family
Device Model: WDC WD2000JD-00HBB0
Serial Number: WD-WCAL73023909
Firmware Version: 08.02D08
User Capacity: 200,049,647,616 bytes
Device is: In smartctl database [for details use: -P show]
ATA Version is: 6
ATA Standard is: Exact ATA specification draft version not indicated
Local Time is: Mon Mar 26 11:47:50 2007 PDT
SMART support is: Available - device has SMART capability.
SMART support is: Enabled
```

```
=== START OF INFORMATION SECTION ===
Model Family: Seagate Barracuda 7200.10 family
Device Model: ST3500630AS
Serial Number: 3QG00GQ7
Firmware Version: 3.AAD
User Capacity: 500,107,862,016 bytes
Device is: In smartctl database [for details use: -P show]
ATA Version is: 7
ATA Standard is: Exact ATA specification draft version not indicated
Local Time is: Mon Mar 26 11:48:09 2007 PDT
SMART support is: Available - device has SMART capability.
SMART support is: Enabled
```

Filesystems:

```
icarus# df -k
Filesystem 1024-blocks Used Avail Capacity Mounted on
```

Re: pitiful performance of an SATA150 drive

Re: pitiful performance of an SATA150 drive

```
/dev/ad8s1a 507630 60150 406870 13% /  
devfs 1 1 0 100% /dev  
/dev/ad8s1d 16244334 45706 14899082 0% /var  
/dev/ad8s1e 16244334 920 14943868 0% /tmp  
/dev/ad8s1f 32494668 1307402 28587694 4% /usr  
/dev/ad8s1g 115577350 1793544 104537618 2% /home  
procfs 4 4 0 100% /proc  
/dev/ad10s1d 473015558 121726480 313447834 28% /storage  
devfs 1 1 0 100% /var/named/dev
```

Pseudo-benchmarks:

```
icarus# time dd if=/dev/ad8s1a of=/dev/null bs=1m  
512+0 records in  
512+0 records out  
536870912 bytes transferred in 11.292704 secs (47541396 bytes/sec)
```

```
icarus# time dd if=/dev/ad10s1d of=/dev/null bs=1m  
^C6798+0 records in  
6798+0 records out  
7128219648 bytes transferred in 92.150703 secs (77353937 bytes/sec)  
0.007u 1.319s 1:32.15 1.4% 27+2956k 0+0io 0pf+0w
```

I consider these numbers pretty decent. The WD drive isn't performing as nice as I'd expect, but the Seagate drive definitely does.

Adjusting the block size in dd doesn't make any difference; I've tried 16k, 32k, 64k, and 1m.

There have been discussions in the past about using dd as a disk I/O tester on FreeBSD (vs. Linux), compared to a utility like bonnie++. Those may apply here, but I think your problem is elsewhere and not with dd on Linux vs. FreeBSD. :)

Regarding interrupt usage:

The above SATA controller is on irq 21, which is also shared with ohci0 on the system. I fired off:

```
icarus# time dd if=/dev/ad10s1d of=/dev/null bs=1m  
^C9988+0 records in  
9988+0 records out  
10473177088 bytes transferred in 135.268101 secs (77425328 bytes/sec)  
0.000u 1.948s 2:15.26 1.4% 24+2695k 0+0io 0pf+0w
```

In one window, and did this in the other:

```
icarus# bash -c "while true; do vmstat -i | grep irq21 && sleep 1; done"  
irq21: ohci0+ 3838384 1  
irq21: ohci0+ 3839576 1
```

Re: pitiful performance of an SATA150 drive

Re: pitiful performance of an SATA150 drive

```
irq21: ohci0+ 3840763 1
irq21: ohci0+ 3841948 1
irq21: ohci0+ 3843131 1
irq21: ohci0+ 3844318 1
irq21: ohci0+ 3845513 1
irq21: ohci0+ 3846703 1
irq21: ohci0+ 3847879 1
irq21: ohci0+ 3849080 1
irq21: ohci0+ 3850258 1
irq21: ohci0+ 3851445 1
irq21: ohci0+ 3852643 1
irq21: ohci0+ 3853607 1
=== Hit ^C to stop the dd here ===
irq21: ohci0+ 3853607 1
irq21: ohci0+ 3853607 1
irq21: ohci0+ 3853609 1
irq21: ohci0+ 3853609 1
irq21: ohci0+ 3853617 1
irq21: ohci0+ 3853617 1
```

Interrupt usage looks about what I'd expect; nothing spiralling out of control, that's for sure.

Are you sure this isn't some weird motherboard problem? Your system obviously uses an APIC; can you toggle usage of it in the BIOS and see if your problem goes away?

--

| Jeremy Chadwick jdc at parodius.com |
| Parodius Networking <http://www.parodius.com/> |
| UNIX Systems Administrator Mountain View, CA, USA |
| Making life hard for others since 1977. PGP: 4BD6C0CB |

freebsd-stable@xxxxxxxxxxx mailing list
<http://lists.freebsd.org/mailman/listinfo/freebsd-stable>
To unsubscribe, send any mail to "freebsd-stable-unsubscribe@xxxxxxxxxxx"