Subject: Re: megaraid_mbox: garbage in file
Posted by vaverin on Fri, 05 May 2006 18:14:25 GMT

View Forum Message <> Reply to Message

James Bottomley wrote:
> On Fri, 2006-05-05 at 09:37 +0400, Vasily Averin wrote:
>>The issue is that the correctly finished scsi read command return me garbage
>>(repeated 0 ...127 -- see hexdump in my first letter) instead correct file content.
>>"attempt to access beyond end of device" messages occurs due the same garbage
>>readed from the Indirect block. I found this garbage present in data buffers
>>beginning at megaraid driver functions.
>>
>>I would note that if I read the same file by using dd with bs=1024 or bs=512 --
>>I get correct file content.
>>
>>When I use kernel with 4Gb memory limit -- the same cat command return me
>>correct file content too, without any garbage.
>>
>>Question is what it is the strange garbage? Have you seen it earlier?
>>Is it possible that it is some driver-related issue or it is broken hardware?
>>And why I can workaround this issue by using only 4Gb memory?
>
> This is really odd ... if the controller can't reach *any* memory above
> 32 bits, then, on an 8GB machine you'd expect corruption all over the
> place since most user pages come from the top of highmem.
>
> The first thing to try, since you have an opteron system, is to get rid
> of highmem entirely and use a 64 bit kernel (just to make sure we're not
> running into some annoying dma_addr_t conversion problem).

Unfortunately it is customers node, and I'm not able to re-install 64-bit
distribution to load 64-bit kernel. Of course I'll ask customer about this, but
it will be done later.

> Then, I
> suppose if that doesn't work, try printing out the actual contents of
> the sg list to see what the physical memory location of the page
> containing the corrupt block is.

I've already done such experiment:
On 2.6.8-based virtuozzo kernel I've added following code to
megaraid_mbox_display_scb function:

```
  virt = page_address(sg[i].page) + sg[i].offset;
  printk("mbox sg%d: page %p off %d addr %llx len %d "
         "virt %p first %08x page->flags %08x\n",
   i, sg[i].page, sg[i].offset, sg[i].dma_address, sg[i].length,
   virt, virt == NULL ? 0: *(int *)virt, sg[i].page->flags);
```

and get the following results

```
May  4 02:51:38 vpsn002 kernel:
 megaraid mailbox: status:0x0 cmd:0xa7 id:0x25 sec:0x1a
   lba:0x33f624ac addr:0xffffffff ld:128 sg:4
 scsi cmnd: 0x28 0x00 0x33 0xf6 0x24 0xac 0x00 0x00 0x1a 0x00
 mbox request_buffer eafde340 use_sg 4
 mbox sg0: page 077a0474 off 0 addr 1fd575000 len 4096 virt ff15a000
   first 03020100 page->flags 40020101
 mbox sg1: page 077b5738 off 0 addr 1fdede000 len 4096 virt ff141000
   first 03020100 page->flags 40020101
 mbox sg2: page 077ad500 off 0 addr 1fdb40000 len 4096 virt ff056000
   first 03020100 page->flags 40020101
 mbox sg3: page 030d46e8 off 1024 addr 5e6a400 len 1024 virt 07e6a400
   first 03020100 page->flags 20001004
```

"first 03020100" shows that data in the all sg buffers is already corrupted.
Also I would note that page for last 1Kb buffer is not Highmem.

If you want I can reproduce this experiment on 2.6.16 kernel too.

> This could also be a firmware problem, I suppose, but I haven't seen any
> similar reports.

Thank you,
 Vasily Averin

SWsoft Virtuozzo/OpenVZ Linux kernel team