

---

Subject: CONFIG\_4KSTACKS

Posted by [christoph](#) on Wed, 09 Aug 2006 14:28:34 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I was just wondering why the CONFIG\_4KSTACKS option was removed from OpenVZ Kernel as of kernel-022stab078.10. I found no bug related to that in Bugzilla.

<http://openvz.org/news/updates/kernel-022stab078.10>

I've seen some stack overflows with kernel-022stab070 in /var/log/messages.

See here for example:

```
Jul 31 05:40:31 node1 kernel: Stack overflow 304 task=svn (e7eb5350) [<c0107766>]
check_stack_overflow+0x66/0x80
Jul 31 05:40:31 node1 kernel: =====
Jul 31 05:40:31 node1 kernel: [<c0107e51>] do_IRQ+0x41/0x1d0
Jul 31 05:40:31 node1 kernel: [<c01070ed>] do_nmi+0x4d/0x70
Jul 31 05:40:31 node1 kernel: [<c0426f5c>] common_interrupt+0x18/0x20
Jul 31 05:40:31 node1 kernel: [<f4905312>] e1000_xmit_frame+0x642/0xc10 [e1000]
Jul 31 05:40:31 node1 kernel: [<c03d4b93>] qdisc_restart+0x83/0x1e0
Jul 31 05:40:31 node1 kernel: [<c03c70ff>] dev_queue_xmit+0xdf/0x360
Jul 31 05:40:31 node1 kernel: [<c03e6cc8>] ip_finish_output2+0xa8/0x1a0
Jul 31 05:40:31 node1 kernel: [<c03e6c20>] ip_finish_output2+0x0/0x1a0
...
```

Thank you,  
Christoph

---

---

Subject: Re: CONFIG\_4KSTACKS

Posted by [dev](#) on Thu, 10 Aug 2006 15:19:27 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Can you check your messages in the log and post full calltraces (not cut down as this one) here please?

Well, we had some strange issues looking like stack overflows and there are some messages which can be found by google, but had no real confirmation of stack overflows.

The main reason why we switched it back to 8k stacks is that it provides slightly better performance under high load when lots of interrupts are generated and stack switching introduces noticable overhead (2-4%).

---

---

Subject: Re: CONFIG\_4KSTACKS

Posted by [christoph](#) on Thu, 10 Aug 2006 16:25:03 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

ok, here is the full trace.

```
Jul 31 05:40:31 node1 kernel: Stack overflow 304 task=svn (e7eb5350) [<c0107766>]
check_stack_overflow+0x66/0x80
Jul 31 05:40:31 node1 kernel: =====
Jul 31 05:40:31 node1 kernel: [<c0107e51>] do_IRQ+0x41/0x1d0
Jul 31 05:40:31 node1 kernel: [<c01070ed>] do_nmi+0x4d/0x70
Jul 31 05:40:31 node1 kernel: [<c0426f5c>] common_interrupt+0x18/0x20
Jul 31 05:40:31 node1 kernel: [<f4905312>] e1000_xmit_frame+0x642/0xc10 [e1000]
Jul 31 05:40:31 node1 kernel: [<c03d4b93>] qdisc_restart+0x83/0x1e0
Jul 31 05:40:31 node1 kernel: [<c03c70ff>] dev_queue_xmit+0xdf/0x360
Jul 31 05:40:31 node1 kernel: [<c03e6cc8>] ip_finish_output2+0xa8/0x1a0
Jul 31 05:40:31 node1 kernel: [<c03e6c20>] ip_finish_output2+0x0/0x1a0
Jul 31 05:40:31 node1 kernel: [<c03e6c20>] ip_finish_output2+0x0/0x1a0
Jul 31 05:40:31 node1 kernel: [<c03d240e>] nf_hook_slow+0xee/0x130
Jul 31 05:40:31 node1 kernel: [<c03e6c20>] ip_finish_output2+0x0/0x1a0
Jul 31 05:40:32 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:32 node1 kernel: [<c03e4695>] ip_finish_output+0x1f5/0x200
Jul 31 05:40:32 node1 kernel: [<c03e6c20>] ip_finish_output2+0x0/0x1a0
Jul 31 05:40:32 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:32 node1 kernel: [<c03e6c04>] dst_output+0x14/0x30
Jul 31 05:40:32 node1 kernel: [<c03d240e>] nf_hook_slow+0xee/0x130
Jul 31 05:40:32 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:32 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:32 node1 kernel: [<c03e4d5c>] ip_queue_xmit+0x44c/0x560
Jul 31 05:40:33 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:33 node1 kernel: [<c03e6bf0>] dst_output+0x0/0x30
Jul 31 05:40:33 node1 kernel: [<c03e4d5c>] ip_queue_xmit+0x44c/0x560
Jul 31 05:40:33 node1 kernel: [<c03d4b27>] qdisc_restart+0x17/0x1e0
Jul 31 05:40:33 node1 kernel: [<c03f6b78>] tcp_transmit_skb+0x598/0x950
Jul 31 05:40:33 node1 kernel: [<c03c090f>] sk_reset_timer+0x1f/0x30
Jul 31 05:40:33 node1 kernel: [<c03f7882>] tcp_write_xmit+0x182/0x320
Jul 31 05:40:33 node1 kernel: [<c03f44fb>] __tcp_data_snd_check+0xeb/0x100
Jul 31 05:40:33 node1 kernel: [<c03f4e6d>] tcp_rcv_established+0x4fd/0xa10
Jul 31 05:40:34 node1 kernel: [<c03fee72>] tcp_v4_do_rcv+0x162/0x170
Jul 31 05:40:34 node1 kernel: [<c03c03c5>] __release_sock+0x45/0x70
Jul 31 05:40:34 node1 kernel: [<c03c0bb8>] release_sock+0x78/0x80
Jul 31 05:40:34 node1 kernel: [<c03ea75c>] tcp_sendmsg+0x4bc/0x12f0
Jul 31 05:40:34 node1 kernel: [<c040f85d>] inet_sendmsg+0x4d/0x60
Jul 31 05:40:34 node1 kernel: [<c03bce8d>] sock_sendmsg+0x9d/0xc0
Jul 31 05:40:34 node1 kernel: [<c03f7882>] tcp_write_xmit+0x182/0x320
Jul 31 05:40:34 node1 kernel: [<c03c103f>] alloc_skb+0x5f/0x130
Jul 31 05:40:34 node1 kernel: [<c026a08b>] as_merged_request+0x4b/0x210
Jul 31 05:40:34 node1 kernel: [<c03bcef6>] kernel_sendmsg+0x46/0x60
```

Jul 31 05:40:34 node1 kernel: [<f4d10595>] drbd\_send+0xb5/0x260 [drbd]  
Jul 31 05:40:34 node1 kernel: [<f4d0ff8f>] drbd\_send\_dblock+0x21f/0x4d0 [drbd]  
Jul 31 05:40:34 node1 kernel: [<c017265b>] bio\_clone+0x1b/0x90  
Jul 31 05:40:34 node1 kernel: [<f4d09d73>] drbd\_make\_request\_common+0x4a3/0x920 [drbd]  
Jul 31 05:40:34 node1 kernel: [<f4d0a2d6>] drbd\_make\_request\_26+0xe6/0x29d [drbd]  
Jul 31 05:40:34 node1 kernel: [<c02640bb>] generic\_make\_request+0x16b/0x1f0  
Jul 31 05:40:34 node1 kernel: [<c014dcd1>] mempool\_alloc+0x81/0x140  
Jul 31 05:40:35 node1 kernel: [<c0121f40>] autoremove\_wake\_function+0x0/0x60  
Jul 31 05:40:35 node1 kernel: [<c02641b0>] submit\_bio+0x70/0x130  
Jul 31 05:40:35 node1 kernel: [<c0172507>] bio\_alloc+0xe7/0x1e0  
Jul 31 05:40:35 node1 kernel: [<c0171cda>] submit\_bh+0x13a/0x1a0  
Jul 31 05:40:35 node1 kernel: [<c0171daf>] ll\_rw\_block+0x6f/0x90  
Jul 31 05:40:35 node1 kernel: [<c01f9000>] \_\_flush\_batch+0x30/0x70  
Jul 31 05:40:35 node1 kernel: [<c01f9103>] \_\_flush\_buffer+0xc3/0x200  
Jul 31 05:40:35 node1 kernel: [<c026007b>] fw\_register\_class\_device+0x7b/0x190  
Jul 31 05:40:35 node1 kernel: [<c01f935a>] log\_do\_checkpoint+0x11a/0x260  
Jul 31 05:40:35 node1 kernel: [<c01f8d63>] \_\_log\_wait\_for\_space+0x113/0x130  
Jul 31 05:40:35 node1 kernel: [<c01f226e>] start\_this\_handle+0x10e/0x4b0  
Jul 31 05:40:35 node1 kernel: [<c0121f40>] autoremove\_wake\_function+0x0/0x60  
Jul 31 05:40:35 node1 kernel: [<c0121f40>] autoremove\_wake\_function+0x0/0x60  
Jul 31 05:40:35 node1 kernel: [<c01f274d>] journal\_start+0xed/0x120  
Jul 31 05:40:35 node1 kernel: [<c01ea114>] \_\_ext3\_journal\_stop+0x24/0x50  
Jul 31 05:40:36 node1 kernel: [<c01e2eba>] ext3\_ordered\_writpage+0x6a/0x200  
Jul 31 05:40:36 node1 kernel: [<c01e2e20>] bput\_one+0x0/0x10  
Jul 31 05:40:36 node1 kernel: [<c0156d02>] pageout+0xc2/0x110  
Jul 31 05:40:36 node1 kernel: [<c014a897>] wake\_up\_page+0x17/0x50  
Jul 31 05:40:36 node1 kernel: [<c0156fc1>] shrink\_list+0x271/0x520  
Jul 31 05:40:36 node1 kernel: [<c0155d78>] \_\_pagevec\_release+0x28/0x40  
Jul 31 05:40:36 node1 kernel: [<c0157402>] shrink\_cache+0x192/0x400  
Jul 31 05:40:36 node1 kernel: [<c0157e0a>] shrink\_zone+0xba/0xf0  
Jul 31 05:40:36 node1 kernel: [<c0157eac>] shrink\_caches+0x6c/0x70  
Jul 31 05:40:36 node1 kernel: [<c0157fb0>] try\_to\_free\_pages+0x100/0x2b0  
Jul 31 05:40:36 node1 kernel: [<c014f5f9>] \_\_alloc\_pages+0x349/0x410  
Jul 31 05:40:36 node1 kernel: [<c014c9b5>] generic\_file\_aio\_write\_nolock+0x355/0xbb0  
Jul 31 05:40:36 node1 kernel: [<c01159d6>] smp\_apic\_timer\_interrupt+0xb6/0xd0  
Jul 31 05:40:36 node1 kernel: [<c0426fde>] apic\_timer\_interrupt+0x1a/0x20  
Jul 31 05:40:36 node1 kernel: [<f4ceb928>] vefs\_init\_inode\_native+0x38/0x120 [vzfs]  
Jul 31 05:40:37 node1 kernel: [<c014d319>] generic\_file\_aio\_write+0x79/0xb0  
Jul 31 05:40:37 node1 kernel: [<c01df844>] ext3\_file\_write+0x44/0xd0  
Jul 31 05:40:37 node1 kernel: [<c016cc10>] do\_sync\_write+0x80/0xb0  
Jul 31 05:40:37 node1 kernel: [<c011a252>] schedule\_vcpu+0x72/0x330  
Jul 31 05:40:37 node1 kernel: [<c016ccf8>] vfs\_write+0xb8/0x130  
Jul 31 05:40:37 node1 kernel: [<c016ccf8>] vfs\_write+0xb8/0x130  
Jul 31 05:40:37 node1 kernel: [<c016ce41>] sys\_write+0x51/0x80  
Jul 31 05:40:37 node1 kernel: [<c04265fb>] syscall\_call+0x7/0xb

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [dev](#) on Fri, 11 Aug 2006 08:49:42 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

and exact kernel version please

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [christoph](#) on Fri, 11 Aug 2006 08:54:04 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

here it is:

2.6.8-022stab070.9-smp

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [dev](#) on Fri, 11 Aug 2006 08:58:51 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I send this call traces to DRBD developers, as don't see much VZ specific here. Thanks for info.

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [christoph](#) on Fri, 11 Aug 2006 09:01:23 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I already talked to DRBD development.  
They don't think it is related to DRBD and they suggested increasing kernel stack size.

Christoph

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [Vasily Tarasov](#) on Fri, 11 Aug 2006 10:04:52 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Sorry, but where have you got this kernel from?  
I mean there is no such kernel version here on site?  
Have you compiled it yourself?

---

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [christoph](#) on Fri, 11 Aug 2006 10:07:56 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

it's a Virtuozzo kernel...

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [dev](#) on Fri, 11 Aug 2006 10:14:00 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

))) yeah, it is easier to think the problem is somewhere else...

for me it doesn't look as the bug introduced specifically by DRBD, but the mainstream problem caused by DRBD usage. DRBD makes networking calls from deep inside of block layer. So this call trace is DRBD-specific. And I don't think any distro would increase kernel stack due to DRBD only as it is always considered as a bug.

Vasiliy will take a look at it.

---

Subject: Re: CONFIG\_4KSTACKS  
Posted by [dev](#) on Fri, 11 Aug 2006 13:11:25 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Here what Vasiliy found. Right column is stack usage in functions from call trace. It really looks like it is unsafe to use DRBD with 4k stacks :/

```
Jul 31 05:40:31 node1 kernel: [<f4905312>] e1000_xmit_frame+0x642/0xc10 [e1000] 60 A84
Jul 31 05:40:31 node1 kernel: [<c03d4b93>] qdisc_restart+0x83/0x1e0 c
Jul 31 05:40:31 node1 kernel: [<c03c70ff>] dev_queue_xmit+0xdf/0x360 14
Jul 31 05:40:31 node1 kernel: [<c03e6cc8>] ip_finish_output2+0xa8/0x1a0 c A04
Jul 31 05:40:31 node1 kernel: [<c03d240e>] nf_hook_slow+0xee/0x130 28
Jul 31 05:40:32 node1 kernel: [<c03e4695>] ip_finish_output+0x1f5/0x200 1c
Jul 31 05:40:32 node1 kernel: [<c03e6c04>] dst_output+0x14/0x30 4
Jul 31 05:40:32 node1 kernel: [<c03d240e>] nf_hook_slow+0xee/0x130 28
Jul 31 05:40:33 node1 kernel: [<c03e4d5c>] ip_queue_xmit+0x44c/0x560 cc
Jul 31 05:40:33 node1 kernel: [<c03f6b78>] tcp_transmit_skb+0x598/0x950 34
Jul 31 05:40:33 node1 kernel: [<c03f7882>] tcp_write_xmit+0x182/0x320 1c
Jul 31 05:40:33 node1 kernel: [<c03f44fb>] __tcp_data_snd_check+0xeb/0x100 1c
Jul 31 05:40:33 node1 kernel: [<c03f4e6d>] tcp_rcv_established+0x4fd/0xa10 20
Jul 31 05:40:34 node1 kernel: [<c03fee72>] tcp_v4_do_rcv+0x162/0x170 10
Jul 31 05:40:34 node1 kernel: [<c03c03c5>] __release_sock+0x45/0x70 8
Jul 31 05:40:34 node1 kernel: [<c03c0bb8>] release_sock+0x78/0x80 c
Jul 31 05:40:34 node1 kernel: [<c03ea75c>] tcp_sendmsg+0x4bc/0x12f0 68
Jul 31 05:40:34 node1 kernel: [<c040f85d>] inet_sendmsg+0x4d/0x60 14
Jul 31 05:40:34 node1 kernel: [<c03bce8d>] sock_sendmsg+0x9d/0xc0 b0 790
```

Jul 31 05:40:34 node1 kernel: [<c03bcef6>] kernel\_sendmsg+0x46/0x60 14  
Jul 31 05:40:34 node1 kernel: [<f4d10595>] drbd\_send+0xb5/0x260 [drbd] 40  
Jul 31 05:40:34 node1 kernel: [<f4d0ff8f>] drbd\_send\_dblock+0x21f/0x4d0 [drbd] 48  
Jul 31 05:40:34 node1 kernel: [<f4d09d73>] drbd\_make\_request\_common+0x4a3/0x920  
[drbd] 64  
Jul 31 05:40:34 node1 kernel: [<f4d0a2d6>] drbd\_make\_request\_26+0xe6/0x29d [drbd] 18  
Jul 31 05:40:34 node1 kernel: [<c02640bb>] generic\_make\_request+0x16b/0x1f0 4c  
Jul 31 05:40:35 node1 kernel: [<c02641b0>] submit\_bio+0x70/0x130 48  
Jul 31 05:40:35 node1 kernel: [<c0171cda>] submit\_bh+0x13a/0x1a0 1c  
Jul 31 05:40:35 node1 kernel: [<c0171daf>] ll\_rw\_block+0x6f/0x90 8  
Jul 31 05:40:35 node1 kernel: [<c01f9000>] \_\_flush\_batch+0x30/0x70 c  
Jul 31 05:40:35 node1 kernel: [<c01f9103>] \_\_flush\_buffer+0xc3/0x200 30  
Jul 31 05:40:35 node1 kernel: [<c01f935a>] log\_do\_checkpoint+0x11a/0x260 12c  
Jul 31 05:40:35 node1 kernel: [<c01f8d63>] \_\_log\_wait\_for\_space+0x113/0x130 1c  
Jul 31 05:40:35 node1 kernel: [<c01f226e>] start\_this\_handle+0x10e/0x4b0  
Jul 31 05:40:35 node1 kernel: [<c01f274d>] journal\_start+0xed/0x120 28 38C  
Jul 31 05:40:36 node1 kernel: [<c01e2eba>] ext3\_ordered\_writepage+0x6a/0x200 1C  
Jul 31 05:40:36 node1 kernel: [<c0156d02>] pageout+0xc2/0x110 48  
Jul 31 05:40:36 node1 kernel: [<c0156fc1>] shrink\_list+0x271/0x520 60  
Jul 31 05:40:36 node1 kernel: [<c0157402>] shrink\_cache+0x192/0x400 60  
Jul 31 05:40:36 node1 kernel: [<c0157e0a>] shrink\_zone+0xba/0xf0 8  
Jul 31 05:40:36 node1 kernel: [<c0157eac>] shrink\_caches+0x6c/0x70 8  
Jul 31 05:40:36 node1 kernel: [<c0157fb0>] try\_to\_free\_pages+0x100/0x2b0 64  
Jul 31 05:40:36 node1 kernel: [<c014f5f9>] \_\_alloc\_pages+0x349/0x410 40  
Jul 31 05:40:36 node1 kernel: [<c014c9b5>] generic\_file\_aio\_write\_nolock+0x355/0xbb0 bc  
Jul 31 05:40:37 node1 kernel: [<c014d319>] generic\_file\_aio\_write+0x79/0xb0 64  
Jul 31 05:40:37 node1 kernel: [<c01df844>] ext3\_file\_write+0x44/0xd0 20  
Jul 31 05:40:37 node1 kernel: [<c016ccf8>] vfs\_write+0xb8/0x130 28  
Jul 31 05:40:37 node1 kernel: [<c016ce41>] sys\_write+0x51/0x80 24  
Jul 31 05:40:37 node1 kernel: [<c04265fb>] syscall\_call+0x7/0xb

---