Subject: Kernel panel with TC Traffic shaping module
Posted by HaroldB on Sun, 24 Sep 2006 02:48:42 GMT
View Forum Message <> Reply to Message

Hi, I am getting this error with Commercial VZ and openVZ kernels:

such as: 2.6.8-022stab078.14-enterprise

Here is the kernel panic:
Quote:
Unable to handle kernel paging request at virtual address 00100100
printing eip:
023bdfcd
*pde = 00003001
Oops: 0000 [#1]
SMP
Modules linked in: vzrst vzcpt ip_vzredir vzredir vzcompat vzfs vzevent vzlist vznet vzstat ip_vznetstat vznetstat vzmon af_packet ip_nat_ftp ip_conntrack_ftp iptable_nat sch_sfq cls_u32 ipt_TOS ipt_LOG sch_cbq vzdquota ipt_state ip_conntrack ipt_length ipt_ttl ipt_tcpmss ipt_TCPMSS iptable_mangle ipt_multiport ipt_limit ipt_tos ipt_REJECT vzdev iptable_filter ip_tables thermal processor fan button battery asus_acpi ac ohci_hcd usbcore shpchp tg3 floppy ide_cd cdrom
CPU: 0, VCPU: 0:1
EIP: 0060:[<023bdfcd>]   Tainted: P
EFLAGS: 00010286   (2.6.8-022stab067.1-enterprise)
EIP is at qdisc_lookup+0x3d/0x60
eax: 001000d0   ebx: 001000d0   ecx: c0bb8118   edx: 00100100
esi: 00010000   edi: c0bb8000   ebp: 91529800   esp: 9445ebf8
ds: 007b   es: 007b   ss: 0068
Process tc (pid: 2418, veid=0, threadinfo=9445e000 task=4a42c770)
Stack: 00000000 91529810 023beb67 c0bb8000 00010000 bf9b3800 00000006 00000008
       024e8348 91529810 00000cec 1afa1220 000005dc bf9b3800 00000000 ffffffff
       c0bb8000 fff8f000 00000001 bf72ff00 0000045c 024e7bc0 91529800 023b862a
Call Trace:
[<023beb67>] tc_modify_qdisc+0x107/0x4d0
[<023b862a>] rtnetlink_rcv+0x31a/0x3e0
[<0216cf56>] rw_vm+0x116/0x330
[<023c3190>] netlink_data_ready+0x60/0x70
[<023c2e42>] netlink_sendmsg+0x512/0x5d0
[<0211f6a0>] default_wake_function+0x0/0x20
[<023c30b9>] netlink_recvmsg+0x1b9/0x230
[<0211f6a0>] default_wake_function+0x0/0x20
[<023a612d>] sock_sendmsg+0x9d/0xc0
[<0216cf56>] rw_vm+0x116/0x330
[<023ac3fc>] verify_iovec+0x3c/0xa0
[<023a7c3e>] sys_sendmsg+0x18e/0x1f0
[<0215f102>] handle_mm_fault+0x132/0x1e0

[<0216cf56>] rw_vm+0x116/0x330
[<0211c28b>] vcpu_put+0x8b/0x110
[<0216d4ba>] get_user_size+0x3a/0x80
[<023a8132>] sys_socketcall+0x242/0x260
[<021083a3>] do_IRQ+0x133/0x1d0
Code: 8b 40 30 0f 18 00 90 39 ca 75 e8 31 c0 5b 5e c3 89 d8 eb f9

When this happens, I need to reboot the server for it to come back to life. Can anyone point me in the right direction to avoid this? I am using rules such as those in the wiki. For each IP on my server I am running:

Quote:

```
$ID=1;  # gets incremented on each IP
$LINERATE="100mbit";
$IP="x.x.x.x" #gets changed on each IP
$THROTTLERATE="10000kbit" # 10 megabits per second limit per IP
 tc qdisc add dev eth0 root handle 1: cbq avpkt 1000 bandwidth $LINERATE  #max rate for interface
 tc qdisc add dev eth1 root handle 1: cbq avpkt 1000 bandwidth $LINERATE  #max rate for interface
 tc class add dev eth0 parent 1: classid 1:$ID cbq rate $THROTTLERATE allot 1500 prio 5 bounded isolated
 tc class add dev eth1 parent 1: classid 1:$ID cbq rate $THROTTLERATE allot 1500 prio 5 bounded isolated
 tc filter add dev eth0 parent 1: protocol ip prio 16 u32 match ip src $IP flowid 1:$ID
 tc filter add dev eth1 parent 1: protocol ip prio 16 u32 match ip src $IP flowid 1:$ID
 tc qdisc add dev eth0 parent 1:$ID sfq perturb 10
 tc qdisc add dev eth1 parent 1:$ID sfq perturb 10
iptables -I FORWARD 1 -o eth0 -s $IP -m limit --limit 200/sec -j ACCEPT
iptables -I FORWARD 2 -o eth0 -s $IP -j DROP
iptables -I FORWARD 1 -o eth1 -s $IP -m limit --limit 200/sec -j ACCEPT
iptables -I FORWARD 2 -o eth1 -s $IP -j DROP
```

Each 30 minutes, I am flushing the rules with:

Quote:

```
 tc qdisc del dev eth0 root 2>/dev/null
 tc qdisc del dev eth1 root 2>/dev/null
```

and then adding them again with the above rules.

Could flushing and adding the rules over and over cause any problems? Is there a limit to the number of rules, say I have 500 IPs on this server, is that too many rules?

Thanks!
Harold

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by dlzinc on Sun, 24 Sep 2006 19:09:28 GMT
View Forum Message <> Reply to Message

Hmm.. out of curiosity... does it work properly if you don't use an enterprise kernel? (eg -smp) If you use a test kernel?

Flushing and re-adding rules *shouldn't* cause a problem.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by HaroldB on Mon, 25 Sep 2006 00:37:14 GMT
View Forum Message <> Reply to Message

Do you mean use v2.6.16? Isn't this kernel version exploitable?

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by dlzinc on Mon, 25 Sep 2006 00:57:04 GMT
View Forum Message <> Reply to Message

Yes. 2.6.16.27 is what the current -test is based off of.

I don't *believe* 2.6.16.27 has any exploitable security problems except for SCTP (which, someone correct me if I'm wrong, is not included in the OVZ kernel)

My "shot in the dark" is that it might be a bug between tc and >4GB memory (enterprise kernel). Alternatively, your box might have bad memory (or a bad memory controller...)

From the crashdump it looks like it's dying when you're running the delete-rules.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by HaroldB on Mon, 25 Sep 2006 01:02:30 GMT
View Forum Message <> Reply to Message

Hi, I am receiving this panic on more than one server, all of which have more than 4GB. Due to this, its unlikely to be a bad hardware component.

Are you saying that I should be able to patch linux-2.6.16.27 with patch-026test017-combined? I think this patch can only work against 2.6.16, which is exploitable. Please tell me your thoughts.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by Vasily Tarasov on Tue, 26 Sep 2006 11:00:42 GMT
View Forum Message <> Reply to Message

Hello,

This is definitely kernel BUG. Thank you for reporting it.
I filled the bug: http://bugzilla.openvz.org/show_bug.cgi?id=278
You can add yourself in CC to follow the progress on BUG.

Thanks,
vass.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by Vasily Tarasov on Tue, 26 Sep 2006 11:23:13 GMT
View Forum Message <> Reply to Message

Harold!

You sad that you've used 2.6.8-022stab078.14-enterprise kernel, but in calltrace, you posted, we find out that the version of running kernel is 2.6.8-022stab067.1-enterprise!

Please, try newer kernel and then inform us about results.

Thanks!

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by HaroldB on Tue, 26 Sep 2006 13:56:45 GMT
View Forum Message <> Reply to Message

Hi Vass, that panic was from a commercial VZ kernel. I get the same panic on openvz kernels as well. I will try and catch one of the openvz kernel panics. Currently, I am researching how TC rules are established using the commercial VZ scripts, the "shaperon" switch to init.d/vz. I feel the tc rules may differ and be more stable.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by Vasily Tarasov on Tue, 26 Sep 2006 14:19:54 GMT
View Forum Message <> Reply to Message

However it worths to carry out your researches on latest kernel, 'cause the problem could be fixed in later releases. And according to oops you're using old kernel: 067.1.

Subject: Re: Kernel panel with TC Traffic shaping module
Posted by HaroldB on Tue, 26 Sep 2006 14:21:55 GMT
View Forum Message <> Reply to Message

Yes I understand this. Like I said, I have gotten the same panic on kernel:
2.6.8-022stab078.14-enterprise

I'll try to catch it on an openvz node running that kernel and paste it here. However, the panic has most of the same data in it.

Subject: Panic Capture of latest stable openvz kernel
Posted by HaroldB on Tue, 26 Sep 2006 15:31:15 GMT
View Forum Message <> Reply to Message

Here we go:

2.6.8-022stab078.14-enterprise

Unable to handle kernel paging request at virtual address 00100100
printing eip:
023d782d
*pde = 00003001
Oops: 0000 [#1]
SMP
Modules linked in: sch_sfq cls_u32 sch_cbq simfs vzdquota af_packet parport_pc lp parport i2c_dev i2c_core sunrpc vznetdev vzmon vzdev ipt_REDIRECT ip_nat_ftp ip_conntrack_ftp ipt_state ipt_length ipt_ttl ipt_tcpmss ipt_TCPMSS ipt_limit ipt_LOG ipt_TOS ipt_tos ipt_REJECT ipt_multiport iptable_filter iptable_mangle iptable_nat ip_conntrack ip_tables thermal processor fan button battery asus_acpi ac ohci_hcd usbcore shpchp e100 mii tg3 floppy ide_cd cdrom
CPU: 2, VCPU: 0:3
EIP: 0060:[<023d782d>] Not tainted
EFLAGS: 00010286 (2.6.8-022stab078.14-enterprise)
EIP is at qdisc_lookup+0x3d/0x60
eax: 001000d0 ebx: 001000d0 ecx: 07f73118 edx: 00100100
esi: 00010000 edi: 07f73000 ebp: 7373d000 esp: 96bdfbf8
ds: 007b es: 007b ss: 0068
Process tc (pid: 2631, veid=0, threadinfo=96bde000 task=d9dbe240)
Stack: 00000000 7373d010 023d83c7 07f73000 00010000 eb5fa400 00000006 00000008
02508048 7373d010 00000c28 32b502e4 000005dc eb5fa400 00000000 ffffffff
07f73000 fff4b000 00000001 e969dc80 0000045c 025078c0 7373d000 023d1d06
Call Trace:
[<023d83c7>] tc_modify_qdisc+0x107/0x4d0
[<023d1d06>] rtnetlink_rcv+0x336/0x400
[<0216c6d6>] rw_vm+0x116/0x330
[<023dc9f0>] netlink_data_ready+0x60/0x70
[<023dc6a2>] netlink_sendmsg+0x512/0x5d0

[<0211ed70>] default_wake_function+0x0/0x20
[<023dc919>] netlink_recvmsg+0x1b9/0x230
[<0211ed70>] default_wake_function+0x0/0x20
[<023bf99d>] sock_sendmsg+0x9d/0xc0
[<0216c6d6>] rw_vm+0x116/0x330
[<023c5cdc>] verify_iovec+0x3c/0xa0
[<023c151e>] sys_sendmsg+0x18e/0x1f0
[<0215e802>] handle_mm_fault+0x142/0x220
[<0216c6d6>] rw_vm+0x116/0x330
[<0216cc3a>] get_user_size+0x3a/0x80
[<023c1a12>] sys_socketcall+0x242/0x260
[<02119f10>] do_page_fault+0x0/0x62a
Code: 8b 40 30 0f 18 00 90 39 ca 75 e8 31 c0 5b 5e c3 89 d8 eb f9