Subject: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Tue, 22 Nov 2011 08:52:18 GMT
View Forum Message <> Reply to Message

Hello,
 since 2.6.32 branch is no longer maintained:
"
Also, from now (30 August 2011) we no longer maintain the following kernel branches:

* 2.6.27
* 2.6.32
"
 we have switched to RHEL6 branch, which seems to run fine, and solves some
long-running problems with 2.6.32 (vSwap, problem with accounting of
mmaped files usage).
 All was nice until some heavier loaded servers came online with RHEL6, and
- they started crashing. And then came the upgrade train:
stab036.1 => stab037.1 => stab039.10 => stab040.1 => stab042.1 etc

 With one of the problems we caught, we were told to switch from stable to
testing kernels (now I see that that testing kernel later became stable,
so while confusing, it makes some sense).
 All those kernels (and stab039.11, which from description should be
latest stable) exhibit the same problem/class of problems - when put under
stress, they crash.
 It's quite easy to recreate, now that we've spent some time tracking it down,
just start the machine with for example:

 stress --cpu 12 --io 16 --vm 32 -d 24 --hdd-bytes 10G
and maybe bonnie++ running in loop, and in few minutes/few hours you've got
dead machines spewing something like:

[1515.249585] BUG: scheduling while atomic: stress/2054/0xffff8800
[1515.250189] BUG: unable to handle kernel paging request at ffffffc047118e0
[1515.250189] IP: [<ffffffff8105620e>] account_system_time+0x9e/0x1f0
[1515.250189] PGD 1a27067 PUD 0
[1515.250189] Thread overran stack, or stack corrupted
[1515.250189] Oops: 0000 [#1] SMP

or maybe:
[1876.747809] BUG: unable to handle kernel paging request at 00000006000000bd
[1876.747815] IP: [<ffffffff8105a4fe>] select_task_rq_fair+0x32e/0xa20
[1876.747823] PGD 12d089067 PUD 0
[1876.747826] Oops: 0000 [#1] SMP

or
[38764.623677] BUG: unable to handle kernel paging request at 000000000001e440
[38764.623677] IP: [<ffffffff814c8efe>] _spin_lock+0xe/0x30

[38764.623677] PGD 12c7b4067 PUD 12c7b5067 PMD 0
[38764.623677] Oops: 0002 [#2] SMP
[38764.623677] last sysfs file: /sys/devices/virtual/block/ram9/stat
[38764.623677] CPU 1
or sometimes strangely affecting HP smart array, and causing it to
disconnect it's raids (I don't understand how that's possible, but it
doesn't happen with old openvz)

Under the same load, classic 2.6.32-openvz kernels do just fine (although
my personal feeling is that rhel6 is way more snappy under such a load).

It usually takes less then few hours for rhel6 kernel to crash, although
with lighter load it might take weeks or months.

Should we continue testing 'stable' branch, or maybe fixes are more likely
to be expected in testing 042.x?

best regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by kir on Tue, 22 Nov 2011 10:34:59 GMT
View Forum Message <> Reply to Message

On 11/22/2011 12:52 PM, Dariush Pietrzak wrote:
> Hello,
> since 2.6.32 branch is no longer maintained:
> "
> Also, from now (30 August 2011) we no longer maintain the following kernel branches:
>
> * 2.6.27
> * 2.6.32
> "
> we have switched to RHEL6 branch, which seems to run fine, and solves some
> long-running problems with 2.6.32 (vSwap, problem with accounting of
> mmaped files usage).
> All was nice until some heavier loaded servers came online with RHEL6, and
> - they started crashing. And then came the upgrade train:
> stab036.1 => stab037.1 => stab039.10 => stab040.1 => stab042.1 etc
>
> With one of the problems we caught, we were told to switch from stable to
> testing kernels (now I see that that testing kernel later became stable,
> so while confusing, it makes some sense).

> All those kernels (and stab039.11, which from description should be
> latest stable) exhibit the same problem/class of problems - when put under
> stress, they crash.
> It's quite easy to recreate, now that we've spent some time tracking it down,
> just start the machine with for example:
>
> stress --cpu 12 --io 16 --vm 32 -d 24 --hdd-bytes 10G
> and maybe bonnie++ running in loop, and in few minutes/few hours you've got
> dead machines spewing something like:
>
> [1515.249585] BUG: scheduling while atomic: stress/2054/0xffff8800
> [1515.250189] BUG: unable to handle kernel paging request at ffffffffc047118e0
> [1515.250189] IP: [<ffffffff8105620e>] account_system_time+0x9e/0x1f0
> [1515.250189] PGD 1a27067 PUD 0
> [1515.250189] Thread overran stack, or stack corrupted
> [1515.250189] Oops: 0000 [#1] SMP
>
> or maybe:
> [1876.747809] BUG: unable to handle kernel paging request at 00000006000000bd
> [1876.747815] IP: [<ffffffff8105a4fe>] select_task_rq_fair+0x32e/0xa20
> [1876.747823] PGD 12d089067 PUD 0
> [1876.747826] Oops: 0000 [#1] SMP
>
> or
> [38764.623677] BUG: unable to handle kernel paging request at 000000000001e440
> [38764.623677] IP: [<ffffffff814c8efe>] _spin_lock+0xe/0x30
> [38764.623677] PGD 12c7b4067 PUD 12c7b5067 PMD 0
> [38764.623677] Oops: 0002 [#2] SMP
> [38764.623677] last sysfs file: /sys/devices/virtual/block/ram9/stat
> [38764.623677] CPU 1
> or sometimes strangely affecting HP smart array, and causing it to
> disconnect it's raids (I don't understand how that's possible, but it
> doesn't happen with old openvz)
>
> Under the same load, classic 2.6.32-openvz kernels do just fine (although
> my personal feeling is that rhel6 is way more snappy under such a load).
>
> It usually takes less then few hours for rhel6 kernel to crash, although
> with lighter load it might take weeks or months.

I am very sad to hear this. Could you please file a bug to
bugzilla.openvz.org so our kernel guys will start working on that?

>
>
> Should we continue testing 'stable' branch, or maybe fixes are more likely
> to be expected in testing 042.x?
Well it depends. What we have in -testing branch is indeed testing, so

there can be more fixes but more bugs. Generally, if you have multiple machines, I recommend to have a few (perhaps less important ones) running rhel6-testing kernels, while having all the other ones at rhel6 (stable) branch.

The thing is, those -testing kernels are actually candidates for stable repo, and as you can see some of them are then moved to stable (after we do some internal testing to make sure there are no regressions etc).

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Wed, 23 Nov 2011 10:52:11 GMT
View Forum Message <> Reply to Message

Zitat von Kir Kolyshkin <kir@openvz.org>:

> On 11/22/2011 12:52 PM, Dariush Pietrzak wrote:
>> Hello,
>> since 2.6.32 branch is no longer maintained:
>> "
>> Also, from now (30 August 2011) we no longer maintain the following
>> kernel branches:
>>
>> * 2.6.27
>> * 2.6.32
>> "
>> we have switched to RHEL6 branch, which seems to run fine, and solves some
>> long-running problems with 2.6.32 (vSwap, problem with accounting of
>> mmaped files usage).
>> All was nice until some heavier loaded servers came online with RHEL6, and
>> - they started crashing. And then came the upgrade train:
>> stab036.1 => stab037.1 => stab039.10 => stab040.1 => stab042.1 etc
>>
>> With one of the problems we caught, we were told to switch from stable to
>> testing kernels (now I see that that testing kernel later became stable,
>> so while confusing, it makes some sense).
>> All those kernels (and stab039.11, which from description should be
>> latest stable) exhibit the same problem/class of problems - when put under
>> stress, they crash.
>> It's quite easy to recreate, now that we've spent some time
>> tracking it down,
>> just start the machine with for example:
>>
>> stress --cpu 12 --io 16 --vm 32 -d 24 --hdd-bytes 10G
>> and maybe bonnie++ running in loop, and in few minutes/few hours you've got
>> dead machines spewing something like:
>>

>> [1515.249585] BUG: scheduling while atomic: stress/2054/0xffff8800
>> [1515.250189] BUG: unable to handle kernel paging request at
>> ffffffc047118e0
>> [1515.250189] IP: [<ffffffff8105620e>] account_system_time+0x9e/0x1f0
>> [1515.250189] PGD 1a27067 PUD 0
>> [1515.250189] Thread overran stack, or stack corrupted
>> [1515.250189] Oops: 0000 [#1] SMP
>>
>> or maybe:
>> [1876.747809] BUG: unable to handle kernel paging request at
>> 00000006000000bd
>> [1876.747815] IP: [<ffffffff8105a4fe>] select_task_rq_fair+0x32e/0xa20
>> [1876.747823] PGD 12d089067 PUD 0
>> [1876.747826] Oops: 0000 [#1] SMP
>>
>> or
>> [38764.623677] BUG: unable to handle kernel paging request at
>> 000000000001e440
>> [38764.623677] IP: [<ffffffff814c8efe>] _spin_lock+0xe/0x30
>> [38764.623677] PGD 12c7b4067 PUD 12c7b5067 PMD 0
>> [38764.623677] Oops: 0002 [#2] SMP
>> [38764.623677] last sysfs file: /sys/devices/virtual/block/ram9/stat
>> [38764.623677] CPU 1
>> or sometimes strangely affecting HP smart array, and causing it to
>> disconnect it's raids (I don't understand how that's possible, but it
>> doesn't happen with old openvz)
>>
>> Under the same load, classic 2.6.32-openvz kernels do just fine (although
>> my personal feeling is that rhel6 is way more snappy under such a load).
>>
>> It usually takes less then few hours for rhel6 kernel to crash, although
>> with lighter load it might take weeks or months.
>
> I am very sad to hear this. Could you please file a bug to
> bugzilla.openvz.org so our kernel guys will start working on that?
>

Sad but true it looks like the RHEL6 based kernels have many rough
edges. We tried to move from some stable Ubuntu 8.04 based OpenVZ
server to RHEL6 based ones, primarly to get better IPv6 support. After
some test we got different kernel panics like this one
http://bugzilla.openvz.org/show_bug.cgi?id=2095 and another one when
using ipt_recent iptable rules inside the VE.
So basically ip(6)tables is not usable inside VE with RHEL6 based kernels :-(
We have also tried the Debian 6 included openvz kernel which works
fine regarding iptables, but got unkillable processes (vsftpd, apache)
spinning at 100% CPU from time to time.

So we have to stick with the Ubuntu 8.04 (2.6.24) OpenVZ until RHEL6
based line really reaches "stable".

Regards

Andreas

File Attachments
1) smime.p7s, downloaded 583 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Wed, 23 Nov 2011 12:31:52 GMT
View Forum Message <> Reply to Message

> I am very sad to hear this. Could you please file a bug to
> bugzilla.openvz.org so our kernel guys will start working on that?

Looking at bugzilla there are many other similiar reports, one of mine has
been closed as fixed, but then returned in exactly the same function after
just 6 minutes of stress-testing new kernel.
It's easy to reproduce, just put enough load on the system.

It looks really troubling, both vSwap and 042.x branches look very nice
feature-wise, even vzmigrate seems to work fine, which is no small feat,
but it kinda feels like stability has been sacrificed to get there.

best regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by kir on Wed, 23 Nov 2011 14:42:25 GMT
View Forum Message <> Reply to Message

On 11/23/2011 04:31 PM, Dariush Pietrzak wrote:
>> I am very sad to hear this. Could you please file a bug to
>> bugzilla.openvz.org so our kernel guys will start working on that?
> Looking at bugzilla there are many other similiar reports, one of mine has
> been closed as fixed, but then returned in exactly the same function after
> just 6 minutes of stress-testing new kernel.
> It's easy to reproduce, just put enough load on the system.

Have you reopened it already? Can you provide bug number?

>
> It looks really troubling, both vSwap and 042.x branches look very nice
> feature-wise, even vzmigrate seems to work fine, which is no small feat,
> but it kinda feels like stability has been sacrificed to get there.
>
> best regards, Eyck

Guys,

I do understand reasons for your frustration, but so far I have only seen one specific bug mentioned in this thread, namely http://bugzilla.openvz.org/2095 it was filed yesterday and there is a patch already available for testing. Any other statements like "there are many bugs", "this kernel is unstable" are just not specific enough for me to deal with.

If there are bugs, they need to be reported and fixed, and we, OpenVZ team, partly rely on you, our users. We do have internal QA but can't possibly test all the use cases and scenarios.

Specifically, we rely on having bug reports from you, with full kernel logs (see http://wiki.openvz.org/Remote_console_setup), test cases (as specific and reproducible as possible), and ideally your ability to test patches that developers provide and report your results back to bugzilla.

We treat bug reports very seriously, and we do our best to reproduce your bugs locally and fix them. Again, please be specific and refer to exact bugs in bugzilla when you are having problems with kernel stability, otherwise it's not helpful and I can't do much about it.

Kir.

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Wed, 23 Nov 2011 15:21:57 GMT
View Forum Message <> Reply to Message

Zitat von Kir Kolyshkin <kir@openvz.org>:

> On 11/23/2011 04:31 PM, Dariush Pietrzak wrote:
>>> I am very sad to hear this. Could you please file a bug to
>>> bugzilla.openvz.org so our kernel guys will start working on that?
>> Looking at bugzilla there are many other similiar reports, one of mine has
>> been closed as fixed, but then returned in exactly the same function after
>> just 6 minutes of stress-testing new kernel.

>> It's easy to reproduce, just put enough load on the system.
>
> Have you reopened it already? Can you provide bug number?
>
>>
>> It looks really troubling, both vSwap and 042.x branches look very nice
>> feature-wise, even vzmigrate seems to work fine, which is no small feat,
>> but it kinda feels like stability has been sacrificed to get there.
>>
>> best regards, Eyck
>
> Guys,
>
> I do understand reasons for your frustration, but so far I have only
> seen one specific bug mentioned in this thread, namely
> http://bugzilla.openvz.org/2095 it was filed yesterday and there is
> a patch already available for testing. Any other statements like
> "there are many bugs", "this kernel is unstable" are just not
> specific enough for me to deal with.
>
> If there are bugs, they need to be reported and fixed, and we,
> OpenVZ team, partly rely on you, our users. We do have internal QA
> but can't possibly test all the use cases and scenarios.
>
> Specifically, we rely on having bug reports from you, with full
> kernel logs (see http://wiki.openvz.org/Remote_console_setup), test
> cases (as specific and reproducible as possible), and ideally your
> ability to test patches that developers provide and report your
> results back to bugzilla.
>
> We treat bug reports very seriously, and we do our best to reproduce
> your bugs locally and fix them. Again, please be specific and refer
> to exact bugs in bugzilla when you are having problems with kernel
> stability, otherwise it's not helpful and I can't do much about it.
>
> Kir.

It was no offense intended from my side. I'm totaly aware what we got
for free from you and your team. I was only wondering way my short
poking around revealed two kernel panics without any esoteric
configuration/load involved. So i was wondering if i try to upgrade
too early or if no one is using the RHEL6 OpenVZ kernels seriously
until now. The first panic is already reported the second will follow
as soon as my test server has finished fschk.

Regards

Andreas

File Attachments
1) smime.p7s, downloaded 589 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Wed, 23 Nov 2011 16:59:29 GMT
View Forum Message <> Reply to Message

Zitat von Kir Kolyshkin <kir@openvz.org>:

> On 11/23/2011 04:31 PM, Dariush Pietrzak wrote:
>>> I am very sad to hear this. Could you please file a bug to
>>> bugzilla.openvz.org so our kernel guys will start working on that?
>> Looking at bugzilla there are many other similiar reports, one of mine has
>> been closed as fixed, but then returned in exactly the same function after
>> just 6 minutes of stress-testing new kernel.
>> It's easy to reproduce, just put enough load on the system.
>
> Have you reopened it already? Can you provide bug number?
>
>>
>> It looks really troubling, both vSwap and 042.x branches look very nice
>> feature-wise, even vzmigrate seems to work fine, which is no small feat,
>> but it kinda feels like stability has been sacrificed to get there.
>>
>> best regards, Eyck
>
> Guys,
>
> I do understand reasons for your frustration, but so far I have only
> seen one specific bug mentioned in this thread, namely
> http://bugzilla.openvz.org/2095 it was filed yesterday and there is
> a patch already available for testing. Any other statements like
> "there are many bugs", "this kernel is unstable" are just not
> specific enough for me to deal with.
>
> If there are bugs, they need to be reported and fixed, and we,
> OpenVZ team, partly rely on you, our users. We do have internal QA
> but can't possibly test all the use cases and scenarios.
>
> Specifically, we rely on having bug reports from you, with full
> kernel logs (see http://wiki.openvz.org/Remote_console_setup), test
> cases (as specific and reproducible as possible), and ideally your
> ability to test patches that developers provide and report your
> results back to bugzilla.

Okay, can someone with a bugzilla account please confirm and create a

bug with this one:

Kernel (uname -a): 2.6.32-042stab039.11 as x86_64 installed on CentOS 6 HN

steps to reproduce:
- use a Ubuntu 10.04 i386 (32 bit) template and create a VE
- load ipt_recent in vz.conf
- start the VE and use something like this "iptables -A INPUT -p tcp
--dport 25 -m conntrack --ctstate NEW -m recent --name SMTP --set"
inside the VE
- exit the VE and execute vzctl stop <VE-ID> at the HN

The result is the following kernel panic

[202.003789] libfcoe_device_notification: NETDEV_UNREGISTER venet0
[202.050580] libfcoe_device_notification: NETDEV_UNREGISTER lo
[202.089227] BUG: unable to handle kernel NULL pointer dereference
at 0000000000000038
[202.089250] IP: [<ffffffffa05690ff>] fini_ipt_recent+0x1f/0x50 [xt_recent]
[202.089271] PGD 6bcb2067 PUD 64d37067 PMD 0
[202.089290] Oops: 0000 [#1] SMP
[202.089309] last sysfs file:
/sys/devices/pci0000:00/0000:00:07.0/net/eth0/type
[202.089322] CPU 0
[202.089329] Modules linked in: netconsole configfs vzethdev simfs
vzrst nf_nat vzcpt nfs lockd fscache nfs_acl auth_rpcgss vzdquota
xt_conntrack ip6t_REJECT ip6table_mangle ip6table_filter ip6_tables
nf_conntrack_ftp nf_conntrack_ipv4 nf_conntrack nf_defrag_ipv4
xt_recent xt_length xt_hl xt_tcpmss xt_TCPMSS iptable_mangle
iptable_filter xt_multiport xt_limit xt_dscp ipt_REJECT ip_tables
vzevent fcoe libfcoe libfc scsi_transport_fc scsi_tgt 8021q garp stp
llc sunrpc vznetdev vzmon vzdev ipv6 ppdev parport_pc parport k10temp
hwmon edac_core edac_mce_amd shpchp sg snd_hda_codec_via snd_hda_intel
snd_hda_codec snd_hwdep snd_seq snd_seq_device snd_pcm snd_timer snd
soundcore snd_page_alloc i2c_nforce2 ext4 mbcache jbd2 sr_mod cdrom
sd_mod crc_t10dif pata_amd ata_generic pata_acpi sata_nv forcedeth
nouveau ttm drm_kms_helper drm i2c_algo_bit i2c_core video output
dm_mod [last unloaded: scsi_wait_scan]
[202.089691]
[202.089696] Modules linked in: netconsole configfs vzethdev simfs
vzrst nf_nat vzcpt nfs lockd fscache nfs_acl auth_rpcgss vzdquota
xt_conntrack ip6t_REJECT ip6table_mangle ip6table_filter ip6_tables
nf_conntrack_ftp nf_conntrack_ipv4 nf_conntrack nf_defrag_ipv4
xt_recent xt_length xt_hl xt_tcpmss xt_TCPMSS iptable_mangle
iptable_filter xt_multiport xt_limit xt_dscp ipt_REJECT ip_tables
vzevent fcoe libfcoe libfc scsi_transport_fc scsi_tgt 8021q garp stp
llc sunrpc vznetdev vzmon vzdev ipv6 ppdev parport_pc parport k10temp
hwmon edac_core edac_mce_amd shpchp sg snd_hda_codec_via snd_hda_intel

snd_hda_codec snd_hwdep snd_seq snd_seq_device snd_pcm snd_timer snd
soundcore snd_page_alloc i2c_nforce2 ext4 mbcache jbd2 sr_mod cdrom
sd_mod crc_t10dif pata_amd ata_generic pata_acpi sata_nv forcedeth
nouveau ttm drm_kms_helper drm i2c_algo_bit i2c_core video output
dm_mod [last unloaded: scsi_wait_scan]
[202.090036] Pid: 25, comm: netns Not tainted 2.6.32-042stab039.11
#1 042stab039_11 To Be Filled By O.E.M.
[202.090047] RIP: 0010:[<ffffffffa05690ff>] [<ffffffffa05690ff>]
fini_ipt_recent+0x1f/0x50 [xt_recent]
[202.090065] RSP: 0018:ffff88006f115ce0 EFLAGS: 00010282
[202.090073] RAX: 0000000000000000 RBX: ffff8800378b3800 RCX:
0000000000000000
[202.090083] RDX: 0000000000000003 RSI: 0000000000000000 RDI:
ffffffffa056a3ac
[202.090094] RBP: ffff88006f115cf0 R08: 0000000000000000 R09:
00000000000001f8
[202.090103] R10: ffff8800378a6000 R11: 0000000000000000 R12:
ffff88006bb058e0
[202.090112] R13: ffff88006bb05000 R14: ffff88006bb058e0 R15:
0000000000000080
[202.090123] FS: 00007f22b35fe700(0000) GS:ffff880002600000(0000)
knlGS:00000000b772b8d0
[202.090193] CS: 0010 DS: 0018 ES: 0018 CR0: 000000008005003b
[202.090193] CR2: 0000000000000038 CR3: 00000000372e1000 CR4:
00000000000006f0
[202.090193] DR0: 0000000000000000 DR1: 0000000000000000 DR2:
0000000000000000
[202.090193] DR3: 0000000000000000 DR6: 00000000ffff0ff0 DR7:
0000000000000400
[202.090193] Process netns (pid: 25, veid=0, threadinfo
ffff88006f114000, task ffff88006f112ec0)
[202.090193] Stack:
[202.090193] ffff88006bb058e0 ffff8800378b3800 ffff88006f115d30
ffffffffa05692e9
[202.090193] <0> ffff88006b099800 0000000000000158 ffff88006b099958
ffff88006b099800
[202.090193] <0> ffff88006b099800 ffffffffa0545660 ffff88006f115d60
ffffffffa05222c5
[202.090193] Call Trace:
[202.090193] [<ffffffffa05692e9>] recent_mt_destroy+0x149/0x150 [xt_recent]
[202.090193] [<ffffffffa05222c5>] cleanup_match+0x45/0x60 [ip_tables]
[202.090193] [<ffffffff810a3f77>] ? uncharge_beancounter+0x57/0x70
[202.090193] [<ffffffffa05223d5>] cleanup_entry+0x65/0xc0 [ip_tables]
[202.090193] [<ffffffffa0524def>] ipt_unregister_table+0x5f/0x90
[ip_tables]
[202.090193] [<ffffffffa054502c>] iptable_filter_net_exit+0x2c/0x30
[iptable_filter]
[202.090193] [<ffffffff8140a94e>] cleanup_net+0x8e/0xe0

[202.090193] [<ffffffff8140a8c0>] ? cleanup_net+0x0/0xe0
[202.090193] [<ffffffff8108c220>] worker_thread+0x190/0x2d0
[202.090193] [<ffffffff81092680>] ? autoremove_wake_function+0x0/0x40
[202.090193] [<ffffffff8108c090>] ? worker_thread+0x0/0x2d0
[202.090193] [<ffffffff810920a6>] kthread+0x96/0xa0
[202.090193] [<ffffffff8100c2ca>] child_rip+0xa/0x20
[202.090193] [<ffffffff81092010>] ? kthread+0x0/0xa0
[202.090193] [<ffffffff8100c2c0>] ? child_rip+0x0/0x20
[202.090193] Code: 1c 24 c9 c3 0f 1f 84 00 00 00 00 00 55 48 89 e5
53 48 83 ec 08 0f 1f 44 00 00 48 8b 87 70 03 00 00 48 89 fb 48 c7 c7
ac a3 56 a0 <48> 8b 70 38 e8 68 68 c8 e0 48 8b bb 20 02 00 00 e8 ec 9f
c0 e0
[202.090193] RIP [<ffffffffa05690ff>] fini_ipt_recent+0x1f/0x50 [xt_recent]
[202.090193] RSP <ffff88006f115ce0>
[202.090193] CR2: 0000000000000038
[202.120364] ---[end trace 6b08bce91c3d45d5]---
[202.121379] Kernel panic - not syncing: Fatal exception
[202.122386] Pid: 25, comm: netns Tainted: G   D
---------------- 2.6.32-042stab039.11 #1
[202.122389] Call Trace:
[202.122398] [<ffffffff814c3a31>] ? panic+0x78/0x143
[202.122402] [<ffffffff814c7d14>] ? oops_end+0xe4/0x100
[202.122407] [<ffffffff81040c5b>] ? no_context+0xfb/0x260
[202.122410] [<ffffffff81040ed5>] ? __bad_area_nosemaphore+0x115/0x1e0
[202.122413] [<ffffffff81040fb3>] ? bad_area_nosemaphore+0x13/0x20
[202.122417] [<ffffffff8104168d>] ? __do_page_fault+0x31d/0x480
[202.122420] [<ffffffff814c427a>] ? thread_return+0x4e/0x854

Thanks

Andreas

File Attachments

1) smime.p7s, downloaded 543 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by kir on Wed, 23 Nov 2011 17:13:27 GMT
View Forum Message <> Reply to Message

On 11/23/2011 08:59 PM, lst_hoe02@kwsoft.de wrote:
> Zitat von Kir Kolyshkin<kir@openvz.org>:
>
>> On 11/23/2011 04:31 PM, Dariush Pietrzak wrote:
>>>> I am very sad to hear this. Could you please file a bug to
>>>> bugzilla.openvz.org so our kernel guys will start working on that?

>>> Looking at bugzilla there are many other similiar reports, one of mine has
>>> been closed as fixed, but then returned in exactly the same function after
>>> just 6 minutes of stress-testing new kernel.
>>> It's easy to reproduce, just put enough load on the system.
>> Have you reopened it already? Can you provide bug number?
>>
>>> It looks really troubling, both vSwap and 042.x branches look very nice
>>> feature-wise, even vzmigrate seems to work fine, which is no small feat,
>>> but it kinda feels like stability has been sacrificed to get there.
>>>
>>> best regards, Eyck
>> Guys,
>>
>> I do understand reasons for your frustration, but so far I have only
>> seen one specific bug mentioned in this thread, namely
>> http://bugzilla.openvz.org/2095 it was filed yesterday and there is
>> a patch already available for testing. Any other statements like
>> "there are many bugs", "this kernel is unstable" are just not
>> specific enough for me to deal with.
>>
>> If there are bugs, they need to be reported and fixed, and we,
>> OpenVZ team, partly rely on you, our users. We do have internal QA
>> but can't possibly test all the use cases and scenarios.
>>
>> Specifically, we rely on having bug reports from you, with full
>> kernel logs (see http://wiki.openvz.org/Remote_console_setup), test
>> cases (as specific and reproducible as possible), and ideally your
>> ability to test patches that developers provide and report your
>> results back to bugzilla.
> Okay, can someone with a bugzilla account please confirm and create a
> bug with this one:

Pardon my curiosity, but why you need someone to act as your proxy filing bugs into bugzilla? I mean, I could create a bug, then a developer will ask you for some additional info, and I will have to ask you and then copy/paste your reply to the bug report, and so on and so forth. Why make things more complicated?

Bugzilla accounts are free and instant, just go to http://bugzilla.openvz.org/createaccount.cgi and enter your email.

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Sharp on Wed, 23 Nov 2011 17:20:06 GMT
View Forum Message <> Reply to Message

On Wed, Nov 23, 2011 at 8:59 PM, <lst_hoe02@kwsoft.de> wrote:

> Okay, can someone with a bugzilla account please confirm and create a bug
> with this one:

You should sign up there and create the bug yourself.

That's the whole point of bug reporting -- interaction with devs. They
can ask you to provide some more details or they can give you a patch
to test. They will need *you* directly to make things fast, not the
middleman.

--
SY, Ilya A. Otyutskiy aka Sharp

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Wed, 23 Nov 2011 17:25:13 GMT
View Forum Message <> Reply to Message

Zitat von Kir Kolyshkin <kir@openvz.org>:

> On 11/23/2011 08:59 PM, lst_hoe02@kwsoft.de wrote:
>> Zitat von Kir Kolyshkin<kir@openvz.org>:
>>
>>> On 11/23/2011 04:31 PM, Dariush Pietrzak wrote:
>>>> I am very sad to hear this. Could you please file a bug to
>>>> bugzilla.openvz.org so our kernel guys will start working on that?
>>>> Looking at bugzilla there are many other similiar reports, one
>>>> of mine has
>>>> been closed as fixed, but then returned in exactly the same function after
>>>> just 6 minutes of stress-testing new kernel.
>>>> It's easy to reproduce, just put enough load on the system.
>>> Have you reopened it already? Can you provide bug number?
>>>
>>>> It looks really troubling, both vSwap and 042.x branches look very nice
>>>> feature-wise, even vzmigrate seems to work fine, which is no small feat,
>>>> but it kinda feels like stability has been sacrificed to get there.
>>>>
>>>> best regards, Eyck
>>> Guys,
>>>
>>> I do understand reasons for your frustration, but so far I have only
>>> seen one specific bug mentioned in this thread, namely
>>> http://bugzilla.openvz.org/2095 it was filed yesterday and there is
>>> a patch already available for testing. Any other statements like
>>> "there are many bugs", "this kernel is unstable" are just not
>>> specific enough for me to deal with.
>>>

>>> If there are bugs, they need to be reported and fixed, and we,
>>> OpenVZ team, partly rely on you, our users. We do have internal QA
>>> but can't possibly test all the use cases and scenarios.
>>>
>>> Specifically, we rely on having bug reports from you, with full
>>> kernel logs (see http://wiki.openvz.org/Remote_console_setup), test
>>> cases (as specific and reproducible as possible), and ideally your
>>> ability to test patches that developers provide and report your
>>> results back to bugzilla.
>> Okay, can someone with a bugzilla account please confirm and create a
>> bug with this one:
>
> Pardon my curiosity, but why you need someone to act as your proxy
> filing bugs into bugzilla? I mean, I could create a bug, then a
> developer will ask you for some additional info, and I will have to
> ask you and then copy/paste your reply to the bug report, and so on
> and so forth. Why make things more complicated?
>
> Bugzilla accounts are free and instant, just go to
> http://bugzilla.openvz.org/createaccount.cgi and enter your email.

I already have countless accounts at numerous
bugzilla/forums/whatever, so i try to avoid creating throw away
accounts (Karteileichen) as much as possible. As some of the
developers must confirm/reproduce the bug anyway it was may impression
it would be smart to avoid just-another-account-somewhere.

But if it helps, so it be

Regards

Andreas

File Attachments

1) smime.p7s, downloaded 531 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by kir on Wed, 23 Nov 2011 20:29:12 GMT
View Forum Message <> Reply to Message

On 11/23/2011 09:25 PM, lst_hoe02@kwsoft.de wrote:
> I already have countless accounts at numerous
> bugzilla/forums/whatever, so i try to avoid creating throw away
> accounts (Karteileichen) as much as possible. As some of the
> developers must confirm/reproduce the bug anyway it was may impression
> it would be smart to avoid just-another-account-somewhere.

I am with you here, and I always wanted to add openid support to our
bugzilla instance. Unfortunately it looks like openid support for
bugzilla is taking forever to be finished. You can see my name in votes
for bugzilla.mozilla.org bug #294608 "Support OpenID as a an account
source and login verification method" at
https://bugzilla.mozilla.org/page.cgi?id=voting/bug.html&amp ;bug_id=294608

While this is still a work in progress, I'm afraid registering at our
bugzilla is the only way to have the OpenVZ bugs you came across fixed.

> But if it helps, so it be

Thank you!

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Thu, 24 Nov 2011 09:44:59 GMT
View Forum Message <> Reply to Message

> Have you reopened it already? Can you provide bug number?

BUG 2080.
I reopened, got told that that is completely different issue, then
encountered exactly the same issue on supposedly fixed kernel,
so reopened again.

> Any other statements like
> "there are many bugs", "this kernel is unstable" are just not
> specific enough for me to deal with.

That's why I wanted to provide a way to reproduce the problem, I would
imagine that overnight stresstest would already be a part of your internal
QA.
This got through our own QA probably because we were running only
'stress' app, only when we added parallel bonnie+ (which we did, because
production machines that were crashing all had significant IO on them as
common thing).

> If there are bugs, they need to be reported and fixed, and we,
> OpenVZ team, partly rely on you, our users. We do have internal QA
> but can't possibly test all the use cases and scenarios.
>
> Specifically, we rely on having bug reports from you, with full
> kernel logs (see http://wiki.openvz.org/Remote_console_setup), test
> cases (as specific and reproducible as possible), and ideally your
> ability to test patches that developers provide and report your

> results back to bugzilla.
>
> We treat bug reports very seriously, and we do our best to reproduce
> your bugs locally and fix them. Again, please be specific and refer

Thats very nice and correct, can you please tell me if the issue I can see here has been reproduced and is too hard to fix, or maybe it's not reproducible, and then, how can I help in reproducing them.

As I said, from my point of view, the issue is trivially reproducible, results in crashes manifesting themselves in few different ways, so I assume that means multiple different bugs, or something more fundamental.

best regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Thu, 24 Nov 2011 11:27:23 GMT
View Forum Message <> Reply to Message

Zitat von Dariush Pietrzak <ml-openvz-eyck@kuszelas.eu>:

>> Have you reopened it already? Can you provide bug number?
>
> BUG 2080.
> I reopened, got told that that is completely different issue, then
> encountered exactly the same issue on supposedly fixed kernel,
> so reopened again.
>
>> Any other statements like
>> "there are many bugs", "this kernel is unstable" are just not
>> specific enough for me to deal with.
>
> That's why I wanted to provide a way to reproduce the problem, I would
> imagine that overnight stresstest would already be a part of your internal
> QA.
> This got through our own QA probably because we were running only
> 'stress' app, only when we added parallel bonnie+ (which we did, because
> production machines that were crashing all had significant IO on them as
> common thing).

Just out of curiousity i use my kernel crash-test setup to test with "stress" and "bonnie". I simply use the OpenVZ-Kernel with two container (ubuntu-10.04) and let one run stress and the other bonnie.

The load is at 15 but the machine is humming along since around 4
hours...
Is it possible that your problem arise from the io devices used?

Regards

Andreas

File Attachments

1) smime.p7s, downloaded 552 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Thu, 24 Nov 2011 12:15:39 GMT
View Forum Message <> Reply to Message

> Just out of curiousity i use my kernel crash-test setup to test with
> "stress" and "bonnie". I simply use the OpenVZ-Kernel with two
> container (ubuntu-10.04) and let one run stress and the other
> bonnie. The load is at 15 but the machine is humming along since
> around 4 hours...
With such low load we also couldn't crash it in timely matter.
With lightly loaded machines we endured months without crash.

I use this:
stress -c 22 -i 24 -m 8 -d 20 --hdd-bytes 10G
and this:
while (true)
do
bonnie++ -d /fs/v/bonnie/ -c 8 -b -f -u root
echo next
done
in parallel, I don't even have to run it inside containers.
(test machine is single 4-core Xeon E5320, with 4G ram and two 146G raid 1s joined by lvm. With loadavg 50-80 we get crashes after few hours).

> Is it possible that your problem arise from the io devices used?

Possible, but unlikely, we first noticed crashed using FC devices, and then moved to testing on small P400i with 256M ram. One of the most affected machines used P410i controller, which is very similiar and the same generation as P400i.
I can re-test on FC again.

And while IO load seems to be neccessary to cause crash, resulting oops-es are similiar, very often account_system_time appears:

[38766.228063] panic occurred, switching back to text console
[38766.228063] BUG: scheduling while atomic: stress/1962/0x10000100
(this is identical to what we saw in production, only with 'java' instead
of 'stress')

[38766.227505] BUG: unable to handle kernel paging request at 0000000000021300
[38766.227509] IP: [<ffffffff81050ec4>] update_curr+0x154/0x200
[38766.227514] PGD 12c7b4067 PUD 12c7b5067 PMD 0

[38764.623677] BUG: unable to handle kernel paging request at 000000000001e440
[38764.623677] IP: [<ffffffff814c8efe>] _spin_lock+0xe/0x30

[38764.599189] BUG: unable to handle kernel paging request at 0000000000019550
[38764.599189] IP: [<ffffffff8105674f>] account_system_time+0xaf/0x1f0

[1876.747809] BUG: unable to handle kernel paging request at 00000006000000bd
[1876.747815] IP: [<ffffffff8105a4fe>] select_task_rq_fair+0x32e/0xa20

[1515.270063] BUG: unable to handle kernel paging request at 00000004047118e0
[1515.270063] IP: [<ffffffff81050aad>] task_rq_lock+0x4d/0xa0

best regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by MailingListe on Fri, 25 Nov 2011 09:39:39 GMT
View Forum Message <> Reply to Message

Zitat von Dariush Pietrzak <ml-openvz-eyck@kuszelas.eu>:

>> Just out of curiousity i use my kernel crash-test setup to test with
>> "stress" and "bonnie". I simply use the OpenVZ-Kernel with two
>> container (ubuntu-10.04) and let one run stress and the other
>> bonnie. The load is at 15 but the machine is humming along since
>> around 4 hours...
> With such low load we also couldn't crash it in timely matter.
> With lightly loaded machines we endured months without crash.
>
> I use this:
> stress -c 22 -i 24 -m 8 -d 20 --hdd-bytes 10G
> and this:
> while (true)
> do
> bonnie++ -d /fs/v/bonnie/ -c 8 -b -f -u root

> echo next
> done
> in parallel, I don't even have to run it inside containers.
> (test machine is single 4-core Xeon E5320, with 4G ram and two 146G raid 1s
> joined by lvm. With loadavg 50-80 we get crashes after few hours).

So i tried loading the kernel harder. With a load of about 48 it was
still stable, if i raise the number for c,i,m,d even more, the OOM
killer jumps in and from that point the whole machine freeze, but does
not panic???

Regards

Andreas

File Attachments
1) smime.p7s, downloaded 552 times

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Fri, 25 Nov 2011 13:26:02 GMT
View Forum Message <> Reply to Message

> >(test machine is single 4-core Xeon E5320, with 4G ram and two 146G raid 1s
> >joined by lvm. With loadavg 50-80 we get crashes after few hours).
>
> So i tried loading the kernel harder. With a load of about 48 it was
> still stable, if i raise the number for c,i,m,d even more, the OOM
> killer jumps in and from that point the whole machine freeze, but
> does not panic???

 That would be good first sign, maybe the parameters I provided are
specific for that DL360 machine, will test again on something larger.

regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
 Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Dariush Pietrzak on Wed, 30 Nov 2011 09:34:27 GMT
View Forum Message <> Reply to Message

> still stable, if i raise the number for c,i,m,d even more, the OOM

Just an update - with 042stab039.11 + bdi patch I was also unable to re-create the original problem, we also tried much higher loads (memtester 52G 100 + stress -c 240 -i 24 -m 48 --vm-bytes 1024MB -d 20 --hdd-bytes 12G + bonnie++) and the problem does not return.
We did encounter another 'hp smartarray disconnecting raids' problems on another test machine, but no kernel oops, and the problem seems hardware-related, thanks.

best regards, Eyck
--
Key fingerprint = 40D0 9FFB 9939 7320 8294 05E0 BCC7 02C4 75CC 50D9
Total Existance Failure

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Stephen Balukoff on Thu, 01 Dec 2011 22:03:55 GMT
View Forum Message <> Reply to Message

We're also seeing a big increase in instability since moving to the RHEL 6 kernels. Specifically, our typical platform consists of a Supermicro motherboard with dual 12-core AMD procs (ie. 24 in one system); The most frustrating part is that the symptom we're seeing is highly intermittent (sometimes it takes 10 minutes to trigger, sometimes several days), and doesn't result in a kernel panic or dump per se. Instead what we're seeing is an unresponsive system (still responding to ping, but all services on the box are unresponsive), with this scrolling by on the console:

BUG: soft lockup - CPU#22 stuck for 67s! [node:585441]
BUG: soft lockup - CPU#23 stuck for 68s! [node:585419]

(multiple times per second, repeating all the different process numbers and many different processes running within containers).

We're going to file a bug report on this, of course, but wondered if there was anything else we can do here to get any other information which can help the devs to come up with the cause and hopefully fix for the above? (Again, we're not getting a panic, and we're not able to do anything on the console.)

Thanks,
Stephen

--
Stephen Balukoff
Blue Box Group, LLC

(800)613-4305 x807

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Stephen Balukoff on Thu, 01 Dec 2011 22:06:48 GMT
View Forum Message <> Reply to Message

Oh! And for what it's worth, we're seeing this on both the latest stable RHEL6 kernel, as well as the latest testing RHEL6 kernel available in the repositories for download. (That is, 042stab39.11 and 042stab044.1 respectively).

Stephen

On Thu, Dec 1, 2011 at 2:03 PM, Stephen Balukoff <sbalukoff@bluebox.net> wrote:
> We're also seeing a big increase in instability since moving to the
> RHEL 6 kernels. Specifically, our typical platform consists of a
> Supermicro motherboard with dual 12-core AMD procs (ie. 24 in one
> system); The most frustrating part is that the symptom we're seeing
> is highly intermittent (sometimes it takes 10 minutes to trigger,
> sometimes several days), and doesn't result in a kernel panic or dump
> per se. Instead what we're seeing is an unresponsive system (still
> responding to ping, but all services on the box are unresponsive),
> with this scrolling by on the console:
>
> BUG: soft lockup - CPU#22 stuck for 67s! [node:585441]
> BUG: soft lockup - CPU#23 stuck for 68s! [node:585419]
>
> (multiple times per second, repeating all the different process
> numbers and many different processes running within containers).
>
> We're going to file a bug report on this, of course, but wondered if
> there was anything else we can do here to get any other information
> which can help the devs to come up with the cause and hopefully fix
> for the above? (Again, we're not getting a panic, and we're not able
> to do anything on the console.)
>
> Thanks,
> Stephen
>
>
> --
> Stephen Balukoff
> Blue Box Group, LLC
> (800)613-4305 x807

--
Stephen Balukoff
Blue Box Group, LLC
(800)613-4305 x807

Subject: Re: Is there a stable OpenVZ kernel, and which should be fit for production
Posted by Stephen Balukoff on Fri, 02 Dec 2011 00:39:57 GMT
View Forum Message <> Reply to Message

Ok, y'all:

We managed to get a call trace. I've opened the following bug on this issue: http://bugzilla.openvz.org/show_bug.cgi?id=2110

Anything else I can do or provide to get traction on getting a developer to look at this? (This is a complete show-stopper for our Scientific Linux 6.1 OpenVZ roll-out.)

Stephen

On Thu, Dec 1, 2011 at 2:06 PM, Stephen Balukoff <sbalukoff@bluebox.net> wrote:
> Oh! And for what it's worth, we're seeing this on both the latest
> stable RHEL6 kernel, as well as the latest testing RHEL6 kernel
> available in the repositories for download. (That is, 042stab39.11
> and 042stab044.1 respectively).
>
> Stephen
>
> On Thu, Dec 1, 2011 at 2:03 PM, Stephen Balukoff <sbalukoff@bluebox.net> wrote:
>> We're also seeing a big increase in instability since moving to the
>> RHEL 6 kernels. Specifically, our typical platform consists of a
>> Supermicro motherboard with dual 12-core AMD procs (ie. 24 in one
>> system); The most frustrating part is that the symptom we're seeing
>> is highly intermittent (sometimes it takes 10 minutes to trigger,
>> sometimes several days), and doesn't result in a kernel panic or dump
>> per se. Instead what we're seeing is an unresponsive system (still
>> responding to ping, but all services on the box are unresponsive),
>> with this scrolling by on the console:
>>
>> BUG: soft lockup - CPU#22 stuck for 67s! [node:585441]
>> BUG: soft lockup - CPU#23 stuck for 68s! [node:585419]
>>
>> (multiple times per second, repeating all the different process
>> numbers and many different processes running within containers).
>>

>> We're going to file a bug report on this, of course, but wondered if
>> there was anything else we can do here to get any other information
>> which can help the devs to come up with the cause and hopefully fix
>> for the above? (Again, we're not getting a panic, and we're not able
>> to do anything on the console.)
>>
>> Thanks,
>> Stephen
>>
>>
>> --
>> Stephen Balukoff
>> Blue Box Group, LLC
>> (800)613-4305 x807
>
>
>
> --
> Stephen Balukoff
> Blue Box Group, LLC
> (800)613-4305 x807

--
Stephen Balukoff
Blue Box Group, LLC
(800)613-4305 x807