
Subject: Re: Need help to debug freeze on kernel side (somehow related to lxc)
Posted by [pva](#) on Fri, 05 Nov 2010 12:41:04 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Matt, for your help!

I've changed a bit subject to make it more clear that lxc freezer itself have no relation (I've did checks you provided, just to be sure). Now server froze again and I had some time to gather a bit of information. Yet still I'm unsure what to do with this freeze.

By freeze I mean that ``ps aux`` output freeze at some point and I'm unable to kill it with `ctrl+C`. `strace` pointed that it hangs on reading `/proc/3780/cmdline` file (environ file is unreadable too). exe symlink pointed on `/usr/sbin/sshd` and this time I was unable to ssh, while previously it was possible (so different processes occur in the same situation from time to time). This process does not belongs to cgroup (it's in `/` cgroup). `kill/kill -9 3780` did nothing. I've tried to gather more proc information from `/proc/3780` (in attachment), also there is `kern.log` with some `sysrq` information (memory info, kernel dump and similar). Could you help me to see what other information could be of interest here? How to find out where sshd hanged and why? I thought `/proc/3780/syscall` could help here, but I failed to find what this file has inside and numbers there are not addresses of functions in `System.map` (or `grep` was unable to find them). Any suggestions, please?

With best regards,

--

Peter.

> > Hi. We are using lxc to separate different services into containers: for
> > this discussions we have apache+php, mysql, nginx containers to serve
> > our web application. After upgrade (I think from kernel 2.6.32 into
> > something newer, now we are using 2.6.35, but tried 34 too) we've
> > experience following issue: at some point nginx starts to show us "504
> > Gateway Time-out" error and while it is possible to ssh on server ``ps
> > aux`` hangs (with no ability to stop it), it is impossible to restart
> > apache container (hangs on stop) and the only way to fix this is to
> > restart server using `sysrq` or power button. At the same time there is
> > nothing in the logs. I suspect apache starts to eat lots of memory but
> > oom killer somehow freezes container but I don't have any proves. What
>
> The OOM killer does not freeze tasks. Now if the tasks were already
> frozen and if the OOM killer selected them then I can see how that

> would be a problem. However, again I doubt that's what's happening here
> for several reasons.

>

> 1. lxc doesn't arbitrarily freeze tasks -- unless you were checkpointing
> or freezing the task yourself (or using a custom script to do
> so), the tasks in the container's cgroup should not be frozen.

>

> 2. If the task(s) are frozen then by definition they are not allocating
> memory. At best they're pinning the memory they've already
> allocated before being frozen. [The tasks will respond to
> kill signals when thawed.]

>

> > could you suggest to debug this issue? What sysrq information could be
> > useful here?

>

> [Cc'ing lxc-users@lists.sf.net for lxc-specific debugging ideas/advice.]

>

> Here's some info on collecting and diagnosing the state of the freezer
> so that hopefully we can eliminate your concerns about it being involved
> and confirm what I've said above:

>

> If you want to figure out if the cgroup freezer is involved at all
> debugging it requires that you be in the "host". Find out which
> process ids are your apache/nginx/etc processes. Then look at their
> cgroups in /proc/<pid>/cgroup. Keep in mind that the "/" in those
> paths isn't the same as "/" -- it's the directory the cgroup
> subsystems are mounted at (see /proc/mounts to figure out where).
> You want the line that says "freezer".

>

> Look at the cgroups mount point with the freezer subsystem in the
> cgroup(s) of these processes (it'll say "freezer" in the mount options).
> Confirm that your pids are listed in the cgroup by looking at the tasks
> file.

>

> If the freezer.state file of those cgroups contains the word "THAWED"
> then the problem lies elsewhere. If the freezer.state says "FREEZING"
> or "FROZEN" however then you'll want to look at the state of the
> processes. Some or all should be in the "D" state while "FREEZING".
> All should be in "D" state while "FROZEN".

>

> "FREEZING" is an intermediate state however so it's not possible to
> determine if there's a bug based purely on the info collected so far.
> The best you can do with "FREEZING" is try and write "FROZEN" into
> freezer.state one or more times and see if it 'eventually' succeeds
> -- say within 10 seconds or 20 attempts, whichever takes longer.
> If it doesn't then you need to strace the processes and see if any
> are stuck in a syscall -- vfork perhaps. You can also try writing
> "THAWED". If it doesn't thaw on the first try then there's a bug.

>
> Whenever you write a new state to freezer.state you should read the
> file again to find out whether the state change took place. Some
> transitions are handled lazily and only take place when you ask for
> the state by reading it.
>
> That's the way to figure out if the freezer is involved and, if so,
> where it's stuck.
>
> Cheers,
> -Matt Helsley

1:blkio,freezer,devices,memory,cpuacct,cpu,ns,debug,cpuset:/

7d0bc56000-7d0bcc000 r-xp 00000000 fe:00 221	/usr/sbin/sshd
7d0bec9000-7d0becb000 r--p 00073000 fe:00 221	/usr/sbin/sshd
7d0becb000-7d0becc000 rw-p 00075000 fe:00 221	/usr/sbin/sshd
7d0becc000-7d0befd000 rw-p 00000000 00:00 0	[heap]
312a1beb000-312a1bf6000 r-xp 00000000 08:03 7686	/lib64/libnss_files-2.12.1.so
(deleted)	
312a1bf6000-312a1df6000 ---p 0000b000 08:03 7686	/lib64/libnss_files-2.12.1.so
(deleted)	
312a1df6000-312a1df7000 r--p 0000b000 08:03 7686	/lib64/libnss_files-2.12.1.so
(deleted)	
312a1df7000-312a1df8000 rw-p 0000c000 08:03 7686	/lib64/libnss_files-2.12.1.so
(deleted)	
312a1df8000-312a1e02000 r-xp 00000000 08:03 7684	/lib64/libnss_nis-2.12.1.so
(deleted)	
312a1e02000-312a2001000 ---p 0000a000 08:03 7684	/lib64/libnss_nis-2.12.1.so
(deleted)	
312a2001000-312a2002000 r--p 00009000 08:03 7684	/lib64/libnss_nis-2.12.1.so
(deleted)	
312a2002000-312a2003000 rw-p 0000a000 08:03 7684	/lib64/libnss_nis-2.12.1.so
(deleted)	
312a2003000-312a2018000 r-xp 00000000 08:03 7687	/lib64/libnsl-2.12.1.so
(deleted)	
312a2018000-312a2217000 ---p 00015000 08:03 7687	/lib64/libnsl-2.12.1.so
(deleted)	
312a2217000-312a2218000 r--p 00014000 08:03 7687	/lib64/libnsl-2.12.1.so
(deleted)	
312a2218000-312a2219000 rw-p 00015000 08:03 7687	/lib64/libnsl-2.12.1.so
(deleted)	
312a2219000-312a221b000 rw-p 00000000 00:00 0	
312a221b000-312a2222000 r-xp 00000000 08:03 7591	/lib64/libnss_compat-2.12.1.so (deleted)
312a2222000-312a2421000 ---p 00007000 08:03 7591	/lib64/libnss_compat-2.12.1.so (deleted)

312a2421000-312a2422000 r--p 00006000 08:03 7591 /lib64/libnss_compat-2.12.1.so (deleted)	
312a2422000-312a2423000 rw-p 00007000 08:03 7591 /lib64/libnss_compat-2.12.1.so (deleted)	
312a2423000-312a2425000 r-xp 00000000 08:03 7682 (deleted)	/lib64/libdl-2.12.1.so
312a2425000-312a2625000 ---p 00002000 08:03 7682 (deleted)	/lib64/libdl-2.12.1.so
312a2625000-312a2626000 r--p 00002000 08:03 7682 (deleted)	/lib64/libdl-2.12.1.so
312a2626000-312a2627000 rw-p 00003000 08:03 7682 (deleted)	/lib64/libdl-2.12.1.so
312a2627000-312a2784000 r-xp 00000000 08:03 7689 (deleted)	/lib64/libc-2.12.1.so
312a2784000-312a2983000 ---p 0015d000 08:03 7689 (deleted)	/lib64/libc-2.12.1.so
312a2983000-312a2987000 r--p 0015c000 08:03 7689 (deleted)	/lib64/libc-2.12.1.so
312a2987000-312a2988000 rw-p 00160000 08:03 7689 (deleted)	/lib64/libc-2.12.1.so
312a2988000-312a298d000 rw-p 00000000 00:00 0	
312a298d000-312a2995000 r-xp 00000000 08:03 7473 (deleted)	/lib64/libcrypt-2.12.1.so
312a2995000-312a2b94000 ---p 00008000 08:03 7473 (deleted)	/lib64/libcrypt-2.12.1.so
312a2b94000-312a2b95000 r--p 00007000 08:03 7473 (deleted)	/lib64/libcrypt-2.12.1.so
312a2b95000-312a2b96000 rw-p 00008000 08:03 7473 (deleted)	/lib64/libcrypt-2.12.1.so
312a2b96000-312a2bc4000 rw-p 00000000 00:00 0	
312a2bc4000-312a2bc6000 r-xp 00000000 08:03 7483 (deleted)	/lib64/libutil-2.12.1.so
312a2bc6000-312a2dc5000 ---p 00002000 08:03 7483 (deleted)	/lib64/libutil-2.12.1.so
312a2dc5000-312a2dc6000 r--p 00001000 08:03 7483 (deleted)	/lib64/libutil-2.12.1.so
312a2dc6000-312a2dc7000 rw-p 00002000 08:03 7483 (deleted)	/lib64/libutil-2.12.1.so
312a2dc7000-312a2ddf000 r-xp 00000000 08:03 86	/lib64/libz.so.1.2.5
312a2ddf000-312a2fde000 ---p 00018000 08:03 86	/lib64/libz.so.1.2.5
312a2fde000-312a2fdf000 r--p 00017000 08:03 86	/lib64/libz.so.1.2.5
312a2fdf000-312a2fe0000 rw-p 00018000 08:03 86	/lib64/libz.so.1.2.5
312a2fe0000-312a3194000 r-xp 00000000 fe:00 11819	/usr/lib64/libcrypto.so.1.0.0
312a3194000-312a3393000 ---p 001b4000 fe:00 11819	/usr/lib64/libcrypto.so.1.0.0
312a3393000-312a33ad000 r--p 001b3000 fe:00 11819	/usr/lib64/libcrypto.so.1.0.0
312a33ad000-312a33b7000 rw-p 001cd000 fe:00 11819 /usr/lib64/libcrypto.so.1.0.0	
312a33b7000-312a33ba000 rw-p 00000000 00:00 0	

312a33ba000-312a33c7000 r-xp 00000000 08:03 7172	/lib64/libpam.so.0.82.3
312a33c7000-312a35c6000 ---p 0000d000 08:03 7172	/lib64/libpam.so.0.82.3
312a35c6000-312a35c7000 r--p 0000c000 08:03 7172	/lib64/libpam.so.0.82.3
312a35c7000-312a35c8000 rw-p 0000d000 08:03 7172	/lib64/libpam.so.0.82.3
312a35c8000-312a35d0000 r-xp 00000000 08:03 5808	/lib64/libwrap.so.0.7.6
312a35d0000-312a37d0000 ---p 00008000 08:03 5808	/lib64/libwrap.so.0.7.6
312a37d0000-312a37d1000 r--p 00008000 08:03 5808	/lib64/libwrap.so.0.7.6
312a37d1000-312a37d2000 rw-p 00009000 08:03 5808	/lib64/libwrap.so.0.7.6
312a37d2000-312a37f0000 r-xp 00000000 08:03 7485	/lib64/ld-2.12.1.so (deleted)
312a39dd000-312a39e2000 rw-p 00000000 00:00 0	
312a39ed000-312a39ee000 rw-p 00000000 00:00 0	
312a39ee000-312a39ef000 r-xp 00000000 00:00 0	[vdso]
312a39ef000-312a39f0000 r--p 0001d000 08:03 7485	/lib64/ld-2.12.1.so (deleted)
312a39f0000-312a39f1000 rw-p 0001e000 08:03 7485	/lib64/ld-2.12.1.so (deleted)
312a39f1000-312a39f2000 rw-p 00000000 00:00 0	
38c1aa4b000-38c1aa6c000 rw-p 00000000 00:00 0	[stack]
ffffffff600000-ffffffff601000 r--p 00000000 00:00 0	[vsyscall]

sshd (3780, #threads: 1)

```

-----
se.exec_start          :    28127759.101429
se.vruntime            :    44278157.160161
se.sum_exec_runtime    :           0.025137
se.statistics.wait_start :           0.000000
se.statistics.sleep_start :           0.000000
se.statistics.block_start :    28127759.101429
se.statistics.sleep_max :           0.000000
se.statistics.block_max :           0.000000
se.statistics.exec_max  :           0.025137
se.statistics.slice_max :           0.000000
se.statistics.wait_max  :           0.000000
se.statistics.wait_sum  :           0.000000
se.statistics.wait_count :              1
se.statistics.iowait_sum :           0.000000
se.statistics.iowait_count :              0
sched_info.bkl_count    :              0
se.nr_migrations        :              1
se.statistics.nr_migrations_cold :              0
se.statistics.nr_failed_migrations_affine:              0
se.statistics.nr_failed_migrations_running:              0
se.statistics.nr_failed_migrations_hot:              0
se.statistics.nr_forced_migrations :              0
se.statistics.nr_wakeups :              0
se.statistics.nr_wakeups_sync :              0
se.statistics.nr_wakeups_migrate :              0
se.statistics.nr_wakeups_local :              0
se.statistics.nr_wakeups_remote :              0
se.statistics.nr_wakeups_affine :              0

```

```
se.statistics.nr_wakeups_affine_attempts: 0
se.statistics.nr_wakeups_passive : 0
se.statistics.nr_wakeups_idle : 0
avg_atom : 0.025137
avg_per_cpu : 0.025137
nr_switches : 1
nr_voluntary_switches : 1
nr_involuntary_switches : 0
se.load.weight : 1024
policy : 0
prio : 120
clock-delta : 81
```

25137 0 1

3780 (sshd) D 6788 6788 6788 0 -1 4202560 2 0 0 0 0 0 0 20 0 1 0 2812775 30244864 22
18446744073709551615 537068396544 537068867916 3900277437424 3900277434256
3378569316729 256 0 4096 81925 0 0 0 17 7 0 0 0 0 0

7384 22 0 116 0 146 0

Name: sshd
State: D (disk sleep)
Tgid: 3780
Pid: 3780
PPid: 6788
TracerPid: 0
Uid: 0 0 0 0
Gid: 0 0 0 0
FDSize: 64
Groups:
VmPeak: 29536 kB
VmSize: 29536 kB
VmLck: 0 kB
VmHWM: 88 kB
VmRSS: 88 kB
VmData: 452 kB
VmStk: 132 kB
VmExe: 464 kB
VmLib: 3684 kB
VmPTE: 60 kB
VmSwap: 328 kB
Threads: 1
SigQ: 10/31080
SigPnd: 0000000000000100
ShdPnd: 0000000000004100
SigBlk: 0000000000000000
SigIgn: 0000000000001000

SigCgt: 0000000000014005
CapInh: 0000000000000000
CapPrm: ffffffff
CapEff: ffffffff
CapBnd: ffffffff
Cpus_allowed: ff
Cpus_allowed_list: 0-7
Mems_allowed: 1
Mems_allowed_list: 0
voluntary_ctxt_switches: 1
nonvoluntary_ctxt_switches: 0
PaX: PeMRs

-1 0x38c1aa6a790 0x312a26c8979

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>

File Attachments

- 1) [3780-cgroup](#), downloaded 407 times
 - 2) [3780-maps](#), downloaded 407 times
 - 3) [3780-sched](#), downloaded 398 times
 - 4) [3780-schedstat](#), downloaded 434 times
 - 5) [3780-stat](#), downloaded 432 times
 - 6) [3780-statm](#), downloaded 416 times
 - 7) [3780-status](#), downloaded 411 times
 - 8) [3780-syscall](#), downloaded 437 times
 - 9) [kern.log](#), downloaded 440 times
-