Subject: Re: [PATCH v9 0/9] Request for inclusion: per-cgroup tcp memory pressure controls
Posted by KAMEZAWA Hiroyuki on Thu, 15 Dec 2011 05:40:19 GMT
View Forum Message <> Reply to Message

On Mon, 12 Dec 2011 11:47:00 +0400
Glauber Costa <glommer@parallels.com> wrote:

> Hi,
>
> This series fixes all the few comments raised in the last round,
> and seem to have acquired consensus from the memcg side.
>
> Dave, do you think it is acceptable now from the networking PoV?
> In case positive, would you prefer merging this trough your tree,
> or acking this so a cgroup maintainer can do it?
>

I met this bug at _1st_ run. Please enable _all_ debug options!.

>From this log, your mem_cgroup_sockets_init() is totally buggy because
tcp init routine will call percpu_alloc() under read_lock.

please fix. I'll turn off the config for today's linux-next.

Regards,
-Kame

==
Dec 15 14:47:15 bluextal kernel: [228.386054] BUG: sleeping function called from invalid context at k
ernel/mutex.c:271
Dec 15 14:47:15 bluextal kernel: [228.386286] in_atomic(): 1, irqs_disabled(): 0, pid: 2692, name: cg
create
Dec 15 14:47:15 bluextal kernel: [228.386438] 4 locks held by cgcreate/2692:
Dec 15 14:47:15 bluextal kernel: [228.386580] #0: (&sb->s_type->i_mutex_key#15/1){+.+.+.}, at: [<ff
ffffff8118717e>] kern_path_create+0x8e/0x150
Dec 15 14:47:15 bluextal kernel: [228.387238] #1: (cgroup_mutex){+.+.+.}, at: [<ffffffff810c2907>]
cgroup_mkdir+0x77/0x3f0
Dec 15 14:47:15 bluextal kernel: [228.387755] #2: (&sb->s_type->i_mutex_key#15/2){+.+.+.}, at: [<ff
ffffff810be168>] cgroup_create_file+0xc8/0xf0
Dec 15 14:47:15 bluextal kernel: [228.388401] #3: (proto_list_lock){++++.+}, at: [<ffffffff814e7fe4

>] mem_cgroup_sockets_init+0x24/0xf0
Dec 15 14:47:15 bluextal kernel: [228.388922] Pid: 2692, comm: cgcreate Not tainted 3.2.0-rc5-next-20
111214+ #34
Dec 15 14:47:15 bluextal kernel: [228.389154] Call Trace:
Dec 15 14:47:15 bluextal kernel: [228.389296] [<ffffffff81080f07>] __might_sleep+0x107/0x140
Dec 15 14:47:15 bluextal kernel: [228.389446] [<ffffffff815ce15f>] mutex_lock_nested+0x2f/0x50
Dec 15 14:47:15 bluextal kernel: [228.389597] [<ffffffff811354fd>] pcpu_alloc+0x4d/0xa20
Dec 15 14:47:15 bluextal kernel: [228.389745] [<ffffffff810a5759>] ? debug_check_no_locks_freed+0xa9/0x180
Dec 15 14:47:15 bluextal kernel: [228.389897] [<ffffffff810a5615>] ? trace_hardirqs_on_caller+0x105/0x190
Dec 15 14:47:15 bluextal kernel: [228.390057] [<ffffffff810a56ad>] ? trace_hardirqs_on+0xd/0x10
Dec 15 14:47:15 bluextal kernel: [228.390206] [<ffffffff810a3600>] ? lockdep_init_map+0x50/0x140
Dec 15 14:47:15 bluextal kernel: [228.390356] [<ffffffff81135f00>] __alloc_percpu+0x10/0x20
Dec 15 14:47:15 bluextal kernel: [228.390506] [<ffffffff812fda18>] __percpu_counter_init+0x58/0xc0
Dec 15 14:47:15 bluextal kernel: [228.390658] [<ffffffff8158fb26>] tcp_init_cgroup+0xd6/0x140
Dec 15 14:47:15 bluextal kernel: [228.390808] [<ffffffff814e8014>] mem_cgroup_sockets_init+0x54/0xf0
Dec 15 14:47:15 bluextal kernel: [228.390961] [<ffffffff8116b47b>] mem_cgroup_populate+0xab/0xc0
Dec 15 14:47:15 bluextal kernel: [228.391120] [<ffffffff810c053a>] cgroup_populate_dir+0x7a/0x110
Dec 15 14:47:15 bluextal kernel: [228.391271] [<ffffffff810c2b16>] cgroup_mkdir+0x286/0x3f0
Dec 15 14:47:15 bluextal kernel: [228.391420] [<ffffffff811836b4>] vfs_mkdir+0xa4/0xe0
Dec 15 14:47:15 bluextal kernel: [228.391566] [<ffffffff8118747d>] sys_mkdirat+0xcd/0xe0
Dec 15 14:47:15 bluextal kernel: [228.391714] [<ffffffff811874a8>] sys_mkdir+0x18/0x20
Dec 15 14:47:15 bluextal kernel: [228.391862] [<ffffffff815d86cf>] tracesys+0xdd/0xe2
==