Subject: Re: [NETNS] Oops in register_pernet_operations() with CONFIG NET NS=n

Posted by den on Thu, 25 Oct 2007 14:00:33 GMT

View Forum Message <> Reply to Message

The patch attached should help. The idea is simple. The "init" should be called only once without NETNS. Period. No need for any lists.

I'll resend it to Dave after the ACK.

```
Regards,
Den
Benjamin Thery wrote:
> Hello Pavel,
>
> I've found a problem with one of your patch related to netns:
 * [NETNS] Move some code into __init section when CONFIG_NET_NS=n (v2)
   http://www.spinics.net/lists/netdev/msg43310.html
> This patch introduces the net init/ net exit/ net initdata
> defines to save some memory when CONFIG_NET_NS is not set.
> Cedric Le Goater reported he had a *non-fatal* oops when booting
> a 2.6.23-mm1-lxc1 kernel with CONFIG NET NS=n. (2.6.23-mm1-lxc1
> contains the NETNS49 patchset). The oops occured when modules
> related to iptables were loaded after the boot completes.
>
> The problem is the following:
> - Your patch adds the __net_initdata attribute to pernet_operations
  structures.
> - pernet_operations are registered via register_pernet_subsys() and
  linked in the pernet list during boot.
> - At the end of boot, pernet operations are freed (because of the
     _net_initdata attribute), and the pernet_list (or first_device list)
  points to freed memory.
>
> - After boot, network modules which are netns-aware try to register
  themselves with register_pernet_subsys() and ...KABOOM... page
  fault when accessing pernet_list (or first_device list).
   (I reproduce Cedric's oops with the command: iptables --list)
>
> This is not a problem right now in 2.6.23-mm1 or net-2.6 because
> there are very few netns-aware network subsystems merged and they
```

```
> are all initialized during boot. But it will be problematic when
> we will merge netns code for subsystems which can be built as
> modules (eg. iptables, ...). I'm not sure we can use
> __net_init_data for pernet_operations then.
> Maybe we can add some checks in register_pernet_operations
> when CONFIG_NET_NS=n.
> I haven't found a fix yet.
> Regards,
> Benjamin
diff --git a/net/core/net_namespace.c b/net/core/net_namespace.c
index 6f71db8..9ba4809 100644
--- a/net/core/net namespace.c
+++ b/net/core/net_namespace.c
@ @ -186,7 +186,11 @ @ static int register_pernet_operations(struct list_head *list,
 int error;
 error = 0;
+#ifdef CONFIG_NET_NS
 list_add_tail(&ops->list, list);
+#endif
+ INIT_LIST_HEAD(&ops->list);
+#endif
for_each_net(net) {
 if (ops->init) {
  error = ops->init(net);
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers
```