Subject: Re: Re: [PATCHSET] 2.6.20-lxc8
Posted by ebiederm on Wed, 28 Mar 2007 00:31:59 GMT
View Forum Message <> Reply to Message

Benjamin Thery <benjamin.thery@bull.net> writes:

> Hi,
>
> Yesterday, I applied a patch similar to Kirill's one that skip skb_cow() in
> ip_forward when the device is a etun, and it does help a lot.
>
> With the patch the cpu load increase is reduced by 50%. Part of the problem is
> "solved".

>
> Here are the figures for netperf:
>
> (Host A -> Host B
> Host A is running kernel 2.6.20-rc5-netns.i386)
>
>   Throughput   CPU load
>
> - without container: 719.78   10.45
> - inside a container (no patch)   719.37   21.88
> - inside a container with patch: 728.93   15.41
>
> The CPU load with the ip_forward patch is now "only" 50% higher (10% compared to
> 15%) than the reference case without container.
>
> The throughput is even better (I repeated the test a few times and I always got
> better results from inside the container).
>
> (1) Why skb_cow() performs the copy?
>
> I also added some traces to understand why skb_cow() does copy the skb: is it
> insufficient headroom or that the skb has been cloned previously?
> In our case, the condition is always that the "TCP skb" is marked as cloned.
> It is likely that these skb have been cloned in tcp_skb_transmit().

Hmm. I wonder if there is any way we could possibly detect or avoid that case.
It sounds like a general routing code issue if the copy is unnecessary.

> (2) Who consumes the other 5% percent cpu?
>
> With the patch installed oprofile reports that pskb_expand_head() (called by
> skb_cow) has disappeared from the top cpu consumers list.
>

> Now, the remaining symbol that shows unusual activity is
> csum_partial_copy_generic().
> I'd like to find who is the caller, unfortunately, this one is harder to
> track. It is written in assembler and called by "static inline" routines and
> Systemtap doesn't like that. :(
>
>
> So, that was the current status.
> I'm continuing my investigations.

Thanks. I would recommend testing a setup using the in kernel ethernet bridging.
It is a completely different path and it should not have much less of a potential
to process packets before they get to the destination network namespace.

Of course if we can improve our routing performance that would be good but
there are limits to what we can correctly do.

Eric

_____