Subject: Re: [PATCH (resubmit)][BRIDGE] Properly dereference the br_should_route_hook
Posted by paulmck on Thu, 29 Nov 2007 14:36:50 GMT
View Forum Message <> Reply to Message

On Fri, Nov 30, 2007 at 12:04:20AM +1100, Herbert Xu wrote:
> On Tue, Nov 27, 2007 at 07:21:08PM +0300, Pavel Emelyanov wrote:
> > This hook is protected with the RCU, so simple
> >
> > if (br_should_route_hook)
> > br_should_route_hook(...)
> >
> > is not enough on some architectures.
> >
> > Use the rcu_dereference/rcu_assign_pointer in this case.
> >
> > Fixed Stephen's comment concerning using the typeof().
> >
> > Signed-off-by: Pavel Emelyanov <xemul@openvz.org>
>
> Applied to net-2.6. Thanks Pavel!
>
> > static void __exit ebtable_broute_fini(void)
> > {
> > - br_should_route_hook = NULL;
> > + rcu_assign_pointer(br_should_route_hook, NULL);
>
> Just for the record, rcu_assign_pointer is never necessary when
> you're assigning NULL. The reason is that rcu_assign_pointer serves
> as a barrier between the initialisation of the content of what you're
> assigning and the actual assignment. Since NULL does not need to be
> initialised you don't need the barrier :)

Of course, if the rcu_assign_pointer() of NULL is not on a hot code
path, the extra memory barrier might not be hurting enough to care.

> Hmm, perhaps we could even build this logic into rcu_assign_pointer.

That certainly is an interesting tradeoff... Save a memory barrier
when assigning NULL, but pay an extra test and branch in all cases.
Though it does make for a simpler rule -- just use rcu_assign_pointer()
in all cases. Of course, if almost all rcu_assign_pointer() executions
assign non-NULL pointers, the optimal strategy would be to leave the
implementation of rcu_assign_pointer() alone, and simply enforce use
of rcu_assign_pointer(), even if the pointer being assigned is NULL.

For a rough guess, if fewer than a few percent of rcu_assign_pointer()
executions assign NULL, then it is best to simply change the rule.

If more than about ten percent of rcu_assign_pointer() executions
assign NULL, then it would make sense to put the check into the
rcu_assign_pointer() primitive. The percentages would be of dynamic
executions, rather than static counts of lines of code.

So, any intuitions on what fraction of the time rcu_assign_pointer()
is assigning NULL? Failing that, what workload should be used to take
the measurements? ;-)

> Then again, who still uses an Alpha? Mine died years ago :)

Although rcu_dereference() does a memory barrier only on Alpha, that of
rcu_assign_pointer() is needed on any machine that does not preserve store
order (Itanium, POWER, ARM, some MIPS boxes according to rumor, ...).

	Thanx, Paul

> Cheers,
> --
> Visit Openswan at http://www.openswan.org/
> Email: Herbert Xu ~{PmV>HI~} <herbert@gondor.apana.org.au>
> Home Page: http://gondor.apana.org.au/~herbert/
> PGP Key: http://gondor.apana.org.au/~herbert/pubkey.txt
> -
> To unsubscribe from this list: send the line "unsubscribe netdev" in
> the body of a message to majordomo@vger.kernel.org
> More majordomo info at http://vger.kernel.org/majordomo-info.html