Subject: Re: [PATCH] various dst_ifdown routines to catch refcounting bugs
Posted by ebiederm on Thu, 27 Sep 2007 16:27:43 GMT
View Forum Message <> Reply to Message

"Denis V. Lunev" <den@openvz.org> writes:

> Moving dst entries into init_net.loopback_dev is not a good thing.
> This hides obvious and non-obvious ref-counting bugs.

Acked-by: "Eric W. Biederman" <ebiederm@xmission.com>

To be clear using init_net.loopback is currently safe because we don't
have any destination cache entries for anything except the initial
network namespace.

I have not yet made this change simply because I haven't gotten around
to this part in my patches.

I do have a question I would like to bring up, because I like avoiding
explicit references to loopback_dev when I can.

/* Dirty hack. We did it in 2.2 (in __dst_free),
 * we have _very_ good reasons not to repeat
 * this mistake in 2.3, but we have no choice
 * now. _It_ _is_ _explicit_ _deliberate_
 * _race_ _condition_.
 *
 * Commented and originally written by Alexey.
 */

What is the race that is talked about in that comment. Can we just
assign NULL instead of the loopback device when we bring a route down.
My gut feeling is that something like:
 dst->input = dst->output = dst_discard;
may be enough.   But I don't know where the deliberate race is.

I haven't traced this all of the way through but from the obvious
parts I just get this nagging feeling that something isn't quite
right.

Eric