
Subject: [patch 0/5][RFC - ipv4/udp checkpoint/restart] dumping/restoring the IPV4/UDP sockets

Posted by [Daniel Lezcano](#) on Wed, 06 Jun 2007 12:18:06 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi,

I would like to resurect the discussion we had concerning the socket checkpoint/restart.

I began to look how to checkpoint them. I thought the following:

The socket can be checkpointed one by one from userspace.
That will allow to provide a mechanism to application which wants to checkpoint himself (I think for example at the tcpcp project) and increase the scope of usability.

If we are inside a container, only the sockets related to the container are viewable and by the way, checkpointable, so the socket checkpoint/restart will bring mobility to the container.

The socket relies to the socket fs. We can use the inode number to identify the socket. Using that, we can checkpoint/restart a socket associated with a fd because inode number can be easily retrieved from a fstat call and we can identify orphan sockets. Inode number of orphan sockets are viewable in the /proc/net/tcp file.

The checkpoint/restart data can be transfered between kernel and userspace via the generic netlink. It is a clean and secure way to define a message format with ascendant compatibility, ie : move the socket to an os with a superior kernel version. The generic netlink message can be either used as raw data to be directly dumped to disk or can be modified from userspace for some specific purpose (I dont have examples)

The way the socket are checkpointed/restored is to stick as much as possible to user/kernel frontier in order to catch errors and bad values the sooner.
I think for example, all the kernel_setsockopt, kernel_connect, etc ... function family, it is more reliable and secure.

The following patchset is a RFC for C/R the UDP sockets. It applies to 2.6.20.

One question is pending. Should we dump/restore send and receive queue knowing the protocol is not reliable ?

Regards.

-- Daniel
--

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>
