Subject: Re: [RFC][PATCH] Pid namespaces vs locks interaction
Posted by serue on Thu, 13 Dec 2007 16:40:55 GMT
View Forum Message <> Reply to Message

Quoting Vitaliy Gusev (vgusev@openvz.org):
> On 12 December 2007 21:42:25 Serge E. Hallyn wrote:
> > Ok sorry - by letting this thread sit a few days I lost track of where
> > we were.
> >
> > I see now, so you're saying fl_pid for nfs is not in fact a task pid.
> > It's a magically derived unique id. (And you say it is unique across
> > all the nfs clients?)
>
> It is unique for pair client,server.
>
> >
> > So does the p in fl_pid stand for something, or could we rename it to
> > fl_id or fl_uniqueid?
>
> If fl_pid will be renamed with fl_uniqueid or something, it still need
> accessing from fs/locks.c: cat /proc/locks shows pids which also are NFS
> pids (unique id).
>
> For example, let's look the /proc/locks in my system (NFS-server) when do
> flock on a NFS client:
>
> 1: POSIX ADVISORY WRITE 2 08:06:63116 0 EOF
> 2: POSIX ADVISORY WRITE 7047 08:09:1899694 0 EOF
> 3: FLOCK ADVISORY WRITE 3334 08:06:110497 0 EOF
> 4: FLOCK ADVISORY WRITE 3265 08:06:94786 0 EOF
> 5: POSIX ADVISORY WRITE 2582 08:06:110462 0 EOF
>
> It indicates that process with pid 2 has a posix lock. Really it is a NFS
> unique id. Problem can be solved by using pid of lockd.
>
> > Maybe that's too much bother, but so long as we're bothering with a pid
> > cleanup at all it seems worth it to me. On the other hand maybe
> > J. Bruce Fields was right and we should accept the fact that the
> > flock->fl_pid shouldn't be taken too seriously, and leave it be.
>
> Mix pids from some namespaces is not good. We can store process pid seen from

Agreed, and that was the basis for my earlier objection.

It sounds like it's clear to all people smarter than I that fl_pid is
not really a pid, so there is no reason for changing the name. And your
patch (contrary to my earlier read of it) only translates fl_nspid into
fl_pids in temporary flocks being passed to userspace, through fcntl and

/proc/locks.

So I completely withdraw my objection.

Except, for the sake of other cognitively challenged types like myself,
could you add a comment by fl_pid and fl_nspid in fs.h, to the effect of

```
 unsigned int fl_pid;  /* unique id and sometimes global pid */
 struct pid *fl_nspid; /* to calculate owner pid_nr for userspace */
```

(or something more accurate if I'm off)?

So after all that,

Acked-by: Serge Hallyn <serue@us.ibm.com>

(sorry)

thanks,
-serge

> init namespace to the flock->fl_pid (instead pid from the current namespace).
> Thus fcntl(F_GETLK,...) and "cat /proc/locks" will show global pids.   But
> some LTP tests can fail.

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