Subject: Re: [RFC][PATCH] Pid namespaces vs locks interaction
Posted by Brad Boyer on Sat, 08 Dec 2007 22:21:19 GMT
View Forum Message <> Reply to Message

On Thu, Dec 06, 2007 at 09:51:30AM -0600, Serge E. Hallyn wrote:
> Quoting Vitaliy Gusev (vgusev@openvz.org):
> > fl_pid is used by nfs, fuse and gfs2. For instance nfs keeps in fl_pid some
> > unique id to identify locking process between hosts - it is not a process
> > pid.
>
> Ok, but so the struct user_flock->fl_pid is being set to the task's
> virtual pid, while the struct kernel_flock->fl_pid is being set to
> task->tgid for nfsd use.
>
> Why can't nfs just generate a uniqueid from the struct pid when it
> needs it?
>
> Fuse just seems to copy the pid to report it to userspace, so it would
> just copy pid_vnr(kernel_flock->pid) into user_flock->fl_pid.
>
> Anyway I haven't looked at all the uses of struct fl_pid, but you
> can always get the pidnr back from the struct pid if needed so there
> should be no problem.

Perhaps we could add a sysid field like some unix systems have. Here is the flock structure documentation from Sun:

The flock structure contains at least the following elements:

```
short   l_type;       /* lock operation type */
short   l_whence;     /* lock base indicator */
off_t   l_start;      /* starting offset from base */
off_t   l_len;        /* lock length; l_len == 0 means
                      until end of file */
int     l_sysid;      /* system ID running process holding lock */
pid_t   l_pid;        /* process ID of process holding lock */
```

Using the sysid could show that the pid field refers to a separate namespace, and might also be useful for NFS to show that the lock is really held by a process on a different system. This would also be something we could export to user space in a way that some programs are already written to expect and handle properly.

Brad Boyer
flar@allandria.com

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