Subject: Re: [IPC]: Logical refcount loop in ipc ns -> massive leakage
Posted by ebiederm on Sun, 04 Feb 2007 08:28:37 GMT
View Forum Message <> Reply to Message

Kirill Korotaev <dev@sw.ru> writes:

> Guys,
>
> Though I have no patch in the hands for mainstream,
> I feel a responsibility to report one majore problem
> related to IPC namespace design.
>
> The problem is about refcounting scheme which is used.
> There is a leak in IPC namespace due to refcounting loop:
> shm segment holds a file, file holds namespace,
> namespace holds shm segment. Loop.
> I suppose the problem is not only IPC-related
> and will happen with some other namespaces as well so should
> be a good lesson for us.
>
> The question is how to fix this.
>
> In OpenVZ we always used 2 different refcounters exactly for this purposes:
> process counter and reference counter.
> When the process counter becomes zero (i.e. the last process from the
> namespace dies) namespace objects are destroyed and cleanuped.
> And the reference counter on the namespace as always protects the structure
> memory only.
>
> How to fix this in mainstream?
> Sure the same approach as above can be used. However, I dislike
> the idea of adding process-counter to each namespace requiring this.
> Any ideas?

I'm still looking and refining, but here is what I have so far:

The struct file that is used appears impossible for user space
to get at directly. Therefore I believe we can instead increment
and decrement the namespace count at the same places we increment
and decrement shm_nattach. Ideally we would only increment the
namespace count when shm_nattach goes from 0 to 1 and we would
only decrement the namespace count when shm_nattach goes from 1 to 0.

Does that make sense?

Eric

_____
Containers mailing list