Subject: Re: Checkpoint/restart (was Re: [PATCH 0/4] - v2 - Object creation with a specified id)
Posted by Alexey Dobriyan on Tue, 22 Apr 2008 20:09:21 GMT
View Forum Message <> Reply to Message

On Tue, Apr 22, 2008 at 11:56:20AM -0700, Dave Hansen wrote:
> On Tue, 2008-04-22 at 23:36 +0400, Alexey Dobriyan wrote:
> > On Fri, Apr 18, 2008 at 07:44:59AM +0200, Nadia.Derbey@bull.net wrote:
> > >   . echo "LONG XX" > /proc/self/task/<my_tid>/next_id
> > >   next object to be created will have an id set to XX
> > >   . echo "LONG<n> X0 ... X<n-1>" > /proc/self/task/<my_tid>/next_id
> > >   next object to be created will have its ids set to XX0, ... X<n-1>
> > >   This is particularly useful for processes that may have several ids if
> > >   they belong to nested namespaces.
> >
> > Can we answer the following questions before merging this patch:
> >
> > a) should mainline kernel have checkpoint/restart feature at all
> > b) if yes, should it be done from kernel- or userspace?
> >
> > Until agreement will be "yes/from userspace" such patches don't make
> > sense in mainline.
>
> What do you mean by "from kernel" or "from userspace"?

By "from userspace" I mean proposed interfaces and similar: usespace by
special system calls puts some state of future object and then does
normal system call which creates aforementioned object.

This can be used for PIDs, OK.

This can be used for SystemV shmem ids. But SystemV shmem also has
(let's choose) uid/gid, atimes and actual content. How would restoration
look like?

How to restore struct task_struct::did_exec ? Do execve(2)?

A was ptracing B at checkpoint moment...

Netdevices: stats, name, all sorts of flags, hw addresses, MTU

iptables rules.

These next ids are suitable, well, only for ids which is very, very small
part of kernel state needed to restore group of processes.

> Userspace has to be involved somewhere, right? It at least need to open()
> the checkpoint file and pass it into the kernel.

Sure, but opening dumpfile is not an interesting part.