Subject: Re: cryo and mm->arg_start
Posted by Sukadev Bhattiprolu on Tue, 15 Jul 2008 21:40:50 GMT
View Forum Message <> Reply to Message

Serge E. Hallyn [serue@us.ibm.com] wrote:
| Quoting Matt Helsley (matthltc@us.ibm.com):
| >
| > On Fri, 2008-07-11 at 09:38 -0700, Dave Hansen wrote:
| > > On Fri, 2008-07-11 at 08:13 -0500, Serge E. Hallyn wrote:
| > > >
| > > > One thing we could do here is to start extending the cryo approach
| > > > with Eric's checkpoint-as-a-coredump (caac?). We generate the
| > > > tiniest of coredumps which, at first, contains nothing but
| > > > mm->arg_start and maybe a process id. It would be simplest if
| > > > it also contained a filename for the real executable,
| > >
| > > The exec model sounds reasonable to me.
| > >
| > > But, I think the filename of the exe is going to have to be in the
| > > checkpoint *already*. It is mapped by at least one of the VMAs, and
| > > will probably be dumped as a normal file-backed area.
| >
| > Yes, the file that backed the exec will be there. Note that thanks to
| > "stacking" filesystems the path to the file backing the exe is not
| > _always_ going to be the same as the path to the file which userspace
| > exec'd in the first place. You can see this by comparing
| > the /proc/<pid>/exe symlink with the file backing the VMA.
| >
| > This is important to any program which checks the /proc/self/exe
| > symlink to find out where it's installed (Java does this, for example).
| > I think it's possible to do this with a binfmt -- it's just one more
| > detail to remember.
| >
| > Cheers,
| > -Matt
|
| Let's say that before starting my checkpointable job, I did
|
| mount -t ecryptfs /home/hallyn /home/hallyn
|
| Now if the checkpointable job is /home/hallyn/somelongjob, then I think
| it's fair to say that restart can fail if /home/hallyn at the restart
| machine isn't ecryptfs-mounted.
|
| In that case, would you still think there is a problem?
|
| On the other hand, if the checkpointable job did the ecryptfs mount
| itself, then it would be expected that at restart the ecryptfs mount

| would be remounted. How that would be done I have no idea offhand.

Hmm, wonder if the new /proc/pid/mountinfo with its mount-ids would
enable us to identify the filesystems that a given process expects.

Which brings up another question. If two processes in the same container
have different mount namespaces and mount points, we would need to
reestablish the mounts during restart right ?

Suka

_____

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