Subject: Re: [DRAFT] Container mini-summit notes v0.01
Posted by Oren Laadan on Tue, 30 Oct 2007 04:35:15 GMT

sukadev@us.ibm.com wrote:
> Oren Laadan [orenl@cs.columbia.edu] wrote:
> |
> | (sorry from the delay, been away :)
> |
> | Eric W. Biederman wrote:
> | > "Serge E. Hallyn" <serue@us.ibm.com> writes:
> | >
> | >> Sorry, I was focusing on the virtual server needs.
> | >>
> | >> devpts is it's own fs so I was fully expecting to make it mountable
> | >> multiple times so a container can have it's own /dev/pts/0. So what
> | >> other virtual devices would we want to be able to rec-reate for a
> | >> migrated application? (I wonder (a) what gregkh will say about having
> | >> a device namespace, and (b) what the sysfs implications will be)
> | >
> | > Depends. There are things like the loop device that could be interesting.
> | > There may be some others. I haven't looked at it enough detail to get
> | > beyond the fact that in some sense it isn't just limited to pts devices.
> | >
> | > A multimount devpts is interesting though.
> |
> | Devices I had to deal with (in zap) so far - to be able to ckpt/restart
> | (and migrate) a desktop session:
> |
> | * /dev/rtc (e.g. for mplayer)
> |
> | * /dev/dsp
> |
> | * /dev/random ? (to isolate entropy pools ?)
> |
> | * virtual consoles - e.g. in zap, an X server that uses a virtual device
> | runs inside a pod/container/VE (and X per-se requires a virtual console)
> |
> | * virtual terminals - e.g. in zap we allow access to a pod from the host
> | without a need to run 'sshd' inside and setup a network in the pod. (Then
> | with a suitable utility and network access to the host, this also allows
> | sort of remote (a-la serial) console access).
> | >From inside the pod it looks like /dev/tty{1,2,..}, so one can run 'getty'
> | processes inside the pod. From the outside (for the admin, e.g.) it is an
> | extended /dev/tty that has an extra ioctl to multiplex access, so the
> | admin (program) can ask to be connected to tty X of pod Y, and it will
> | connect to that console (like connecting via serial line).
>

> This sounds really interesting. Were these devices part of a complete
> device namespace ? IOW, does say /dev/tty2 in each pods have the same
> major/minor number (4,2) ? Does each '/dev/tty2' have a separate entry
> in sysfs ?

yes, they are virtualization-aware (keep in mind that this was done
before the recent work on namespaces), by having the open() method check
in which pod (namespace) it is called and act accordingly. So /dev/zty2
(zty stands for zap-tty) has the same maj/min in all pods. while at this
moment it is not integrated with sysfs, I see no reason not to do so.

>
>
> | The main advantage is that as a virtual device it can be migrated (with
> | its buffers, if not empty, as they reside inside the pod) so upon restart
> | they go with the 'getty' processes that use them. The (old) admin will
> | see the line dropped, and the (new) admin after the migration can connect
> | at the new machine.
> |
> | Oren.
> |
> |
> | _____
> | Containers mailing list
> | Containers@lists.linux-foundation.org
> | https://lists.linux-foundation.org/mailman/listinfo/containers
_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers