Subject: Re: Re: BC: resource beancounters (v2)
Posted by Rohit Seth on Mon, 28 Aug 2006 22:28:09 GMT
View Forum Message <> Reply to Message

On Mon, 2006-08-28 at 21:41 +0400, Kir Kolyshkin wrote:
> Rohit Seth wrote:
> >
> > I'm not saying per use glibc etc. That will indeed be useless and bring
> > it to virtualization world. Just like fake node, one should be allowed
> > to use pages that are already in (for example) page cache- so that you
> > don't end up duplicating all shared stuff. But as far as charging is
> > concerned, charge it to container who either got the page in page cache
> > OR if FS based semantics exist then charge it to the container where the
> > file belongs. What I was suggesting is to not charge a page to
> > different counters.
> >
>
> Consider the following simple scenario: there are 50 containers
> (numbered, say, 1 to 50) all sharing a single installation of Fedora
> Core 5. They all run sshd, apache, syslogd, crond and some other stuff
> like that. This is actually quite a real scenario.
>
> In the world that you propose the container which was unlucky to start
> first (probably the one with ID of either 1 or 50) will be charged for
> all the memory, and all the
> others will have most of their memory for free. And in such a world
> per-container memory accounting or limiting is just not possible.

If you are only having task based accounting then yes the first
container using a page will be charged. And when it hit its limit then
it will inactivate some of the pages. If some other container now uses
the same page (that got inactivated) again then this next container will
be charged for that page.

Though if we have file/directory based accounting then shared pages
belonging to /usr/lib or /usr/bin can go to a common container.

-rohit