Subject: Re: Re: BC: resource beancounters (v2)
Posted by kir on Mon, 28 Aug 2006 17:40:18 GMT
View Forum Message <> Reply to Message

Rohit Seth wrote:
> On Sat, 2006-08-26 at 17:37 +0100, Alan Cox wrote:
>
>> Ar Gwe, 2006-08-25 am 19:15 -0700, ysgrifennodd Rohit Seth:
>>
>>> Yes, sharing of pages across different containers/managers will be a
>>> problem. Why not just disallow that scenario (that is what fake nodes
>>> proposal would also end up doing).
>>>
>> Because it destroys the entire point of using containers instead of
>> something like Xen - which is sharing. Also at the point I am using
>> beancounters per user I don't want glibc per use, libX11 per use glib
>> per use gtk per user etc..
>>
>>
>>
>
> 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.