Subject: Re: containers development plans (July 10 version)
Posted by dev on Fri, 13 Jul 2007 13:10:46 GMT
View Forum Message <> Reply to Message

Paul Menage wrote:
> On 7/10/07, Serge E. Hallyn <serge@hallyn.com> wrote:
>
>>A (still under construction) list of features we expect to be worked on
>>next year looks like this:
>>   4. task containers functionality
>>   specific containers
>
>
> A couple of more container subsystem requests that have come out of
> the Linux Foundation Japan symposium, although I think they've also
> been mentioned before more than once - per-container swap and disk I/O
> scheduling.
>
> I'm not familiar enough with the current Linux disk scheduler code to
> know how easy/hard it is to add rate guarantees on a per-container
> basis, but the swap one should be easier.

Paul, OpenVZ implements 2-level disk I/O scheduler based on CFQ.
It is not that hard to implement and we will propose the patches
a bit later based on your containers stuff.

However, there is one big problem which I'm not sure mainstream
is ready to go with. Writes. Writes are usually asynchronous
and it is impossible to say which container caused the write
when it goes to the scheduler.
OpenVZ solves this by tracking dirty pages and context which made them dirty.
But it is a next step after std CFQ modifying.

> One potential issue with the swap container is how integrated should
> it be with the memory controller? I can certainly see people wanting
> to be able to use a swap controller without requiring a page-based
> memory controller (e.g. you might want to combine it with node-based
> control via cpusets instead) but adding two pointers to the mm_struct,
> one for swap controller subsystem and one for memory controller
> subsystem, seems a little bit ugly.

Thanks,
Kirill