Subject: Re: [ckrm-tech] [PATCH 4/7] UBC: syscalls (user interface)
Posted by Rohit Seth on Thu, 24 Aug 2006 01:20:35 GMT
View Forum Message <> Reply to Message

On Tue, 2006-08-22 at 12:58 +0900, Magnus Damm wrote:
> On Mon, 2006-08-21 at 18:16 -0700, Rohit Seth wrote:
> > On Mon, 2006-08-21 at 11:47 +0900, Magnus Damm wrote:
> > > On Fri, 2006-08-18 at 07:45 -0700, Dave Hansen wrote:
> > > > On Fri, 2006-08-18 at 12:08 +0400, Andrey Savochkin wrote:
> > > > >
> > > > > A) Have separate memory management for each container,
> > > > >   with separate buddy allocator, lru lists, page replacement mechanism.
> > > > >   That implies a considerable overhead, and the main challenge there
> > > > >   is sharing of pages between these separate memory managers.
> > > >
> > > > Hold on here for just a sec...
> > > >
> > > > It is quite possible to do memory management aimed at one container
> > > > while that container's memory still participates in the main VM.
> > > >
> > > > There is overhead here, as the LRU scanning mechanisms get less
> > > > efficient, but I'd rather pay a penalty at LRU scanning time than divide
> > > > up the VM, or coarsely start failing allocations.
> > >
> > > This could of course be solved with one LRU per container, which is how
> > > the CKRM memory controller implemented things about a year ago.
> >
> > Effectively Andrew's idea of faking up nodes is also giving per
> > container LRUs.
>
> Yes, but the NUMA emulation approach is using fixed size containers
> where the size is selectable at the kernel command line,
[Apologies for late reply..]

Yup, if we run with fake NUMA support for providing container
functionality then dynamic resizing will be important (and that is why I
made the initial comment of possibly using memory hot-plug)

> while the CKRM
> (and pzone) approach provides a more dynamic (and complex) solution.

...this complexity is not always a positive thing ;-) (I do like core
of CKRM stuff FWIW).

-rohit