Subject: Re: [ckrm-tech] [PATCH 4/7] UBC: syscalls (user interface)
Posted by Magnus Damm on Mon, 21 Aug 2006 02:38:40 GMT
View Forum Message <> Reply to Message

On Fri, 2006-08-18 at 09:42 -0700, Andrew Morton wrote:
> On Fri, 18 Aug 2006 07:45:48 -0700
> Dave Hansen <haveblue@us.ibm.com> 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.
> >
>
> I have this mad idea that you can divide a 128GB machine up into 256 fake
> NUMA nodes, then you use each "node" as a 512MB unit of memory allocation.
> So that 4.5GB job would be placed within an exclusive cpuset which has nine
> "mems" (what are these called?) and voila: the job has a hard 4.5GB limit,
> no kernel changes needed.
>
> Unfortunately this is not testable because numa=fake=256 doesn't come even
> vaguely close to working. Am trying to get that fixed.

You may be looking for the NUMA emulation patches posted here:

http://marc.theaimsgroup.com/?l=linux-mm&m=1128065875018 84&w=2

There is a slightly updated x86_64 version here too:

http://marc.theaimsgroup.com/?l=linux-mm&m=1131613865203 42&w=2

/ magnus