Subject: Re: [PATCH] BC: resource beancounters (v2)
Posted by Nick Piggin on Sat, 26 Aug 2006 03:55:43 GMT
View Forum Message <> Reply to Message

Alan Cox wrote:
> Ar Sad, 2006-08-26 am 01:14 +1000, ysgrifennodd Nick Piggin:
>
>>I still think doing simple accounting per-page would be a better way to
>>go than trying to pin down all "user allocatable" kernel allocations.
>>And would require all of about 2 hooks in the page allocator. And would
>>track *actual* RAM allocated by that container.
>
>
> You have a variety of kernel objects you want to worry about and they
> have very differing properties.
>
> Some are basically shared resources - page cache, dentries, inodes, etc
> and can be balanced pretty well by the kernel (ok the dentries are a bit
> of a problem right now). Others are very specific "owned" resources -
> like file handles, sockets and vmas.

That's true (OTOH I'd argue it would still be very useful for things
like pagecache, so one container can't start a couple of 'dd' loops
and turn everyone else to crap). And while the sharing may not be
exactly captured, statistically things should balance over time.

So I'm not arguing about _also_ accounting resources that are limited
in other ways (than just the RAM they consume).

But as a DoS protection measure on RAM usage, trying to account all
kernel allocations that are user triggerable just sounds hard to
maintain, holey, ugly, invsive (and not perfect either -- in fact it
still isn't clear to me that it is any better than my proposal).

>
> Tracking actual RAM use by container/user/.. isn't actually that
> interesting. It's also inconveniently sub page granularity.

If it isn't interesting, then I don't think we want it (at least, until
someone does get an interest in it).

>
> Its a whole seperate question whether you want a separate bean counter
> limit for sockets, file handles, vmas etc.

Yeah that's fair enough. We obviously want to avoid exposing limits on
things that it doesn't make sense to limit, or that is a kernel
implementation detail as much as possible.

eg. so I would be happy to limit virtual address, less happy to limit
vmas alone (unless that is in the context of accounting their RAM usage
or their implied vaddr charge).

--
SUSE Labs, Novell Inc.
Send instant messages to your online friends http://au.messenger.yahoo.com