Subject: Re: [RFC][PATCH 2/7] RSS controller core
Posted by ebiederm on Sun, 11 Mar 2007 19:34:42 GMT
View Forum Message <> Reply to Message

Andrew Morton <akpm@linux-foundation.org> writes:

> Yep. Straightforward machine partitioning. An attractive thing is that it
> 100% reuses existing page reclaim, unaltered.

And misses every resource sharing opportunity in sight. Except for
filtering the which pages are eligible for reclaim an RSS limit should
not need to change the existing reclaim logic, and with things like the
memory zones we have had that kind of restriction in the reclaim logic
for a long time. So filtering out ineligible pages isn't anything new.

>> imho memzone approach is inconvinient for pages sharing and shares accounting.
>> it also makes memory management more strict, forbids overcommiting
>> per-container etc.
>
> umm, who said they were requirements?
>
>> Maybe you have some ideas how we can decide on this?
>
> We need to work out what the requirements are before we can settle on an
> implementation.

If you are talking about RSS limits the term is well defined. The
number of pages you can have mapped into your set of address space at
any given time.

Unless I'm totally blind that isn't what the patchset implements. A
true RSS limit over multiple processes has a lot of potential to be
generally useful, is very understandable, doesn't affect kernel cache
decisions so largely performance should not be affected. There is a
little more overhead in the fault logic but that is a moderately
expensive path anyway.

> Sigh. Who is running this show? Anyone?

Someone is supposed to run the show? :)

> You can actually do a form of overcommittment by allowing multiple
> containers to share one or more of the zones. Whether that is sufficient
> or suitable I don't know. That depends on the requirements, and we haven't
> even discussed those, let alone agreed to them.

Another really nasty issue is the container term as the resource guys
are using the term in a subtlety different way then it has been used

with namespaces leading to several threads where the participants talked past each other. We need a different term to designate the group of tasks a resource controller is dealing with.

The whole filesystem interface also is over general and makes it too easy to express the hard things (like move an existing task from one group of tasks to another) leading to code complications.

On the up side I think the code the focus is likely in the right place to start delivering usable code.

Eric

_____

Containers mailing list
Containers@lists.osdl.org
https://lists.osdl.org/mailman/listinfo/containers