
Subject: Re: [RFC/PATCH] cgroup swap subsystem
Posted by [KAMEZAWA Hiroyuki](#) on Thu, 06 Mar 2008 08:33:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Thu, 06 Mar 2008 11:20:17 +0300
Pavel Emelyanov <xemul@openvz.org> wrote:

> KAMEZAWA Hiroyuki wrote:
> > On Wed, 05 Mar 2008 17:14:12 +0300
> > Pavel Emelyanov <xemul@openvz.org> wrote:
> >>> Strongly agree. Nobody's interested in swap as such: it's just
> >>> secondary memory, where RAM is primary memory. People want to
> >>> control memory as the sum of the two; and I expect they may also
> >>> want to control primary memory (all that the current memcg does)
> >>> within that. I wonder if such nesting of limits fits easily
> >>> into cgroups or will be problematic.
> >> This nesting would affect the res_counter abstraction, not the
> >> cgroup infrastructure. Current design of resource counters doesn't
> >> allow for such thing, but the extension is a couple-of-lines patch :)
> >>
> > IMHO, keeping res_counter simple is better.
> >
> > Is this kind of new entry in mem_cgroup not good ?
> > ==
> > struct mem_cgroup {
> > ...
> > struct res_counter memory_limit.
> > struct res_counter swap_limit.
> > ..
> > }
> >
> > I meant the same thing actually. By "nesting would affect" I
> > meant, that we might want to make res_counters hierarchical.
> >
> > That would kill two birds with one stone - we will make a true
> > hierarchical memory accounting and let charging of two counters
> > with one call.

Hierarchical res_counter makes sense.
Making it in simple/reasonable style will be our challenge.

Thanks,
-Kame

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>

Subject: Re: [RFC/PATCH] cgroup swap subsystem
Posted by [Pavel Emelianov](#) on Thu, 06 Mar 2008 08:38:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

KAMEZAWA Hiroyuki wrote:

> On Thu, 06 Mar 2008 11:20:17 +0300

> Pavel Emelyanov <xemul@openvz.org> wrote:

>

>> KAMEZAWA Hiroyuki wrote:

>>> On Wed, 05 Mar 2008 17:14:12 +0300

>>> Pavel Emelyanov <xemul@openvz.org> wrote:

>>>> Strongly agree. Nobody's interested in swap as such: it's just

>>>> secondary memory, where RAM is primary memory. People want to

>>>> control memory as the sum of the two; and I expect they may also

>>>> want to control primary memory (all that the current memcg does)

>>>> within that. I wonder if such nesting of limits fits easily

>>>> into cgroups or will be problematic.

>>>> This nesting would affect the res_counter abstraction, not the

>>>> cgroup infrastructure. Current design of resource counters doesn't

>>>> allow for such thing, but the extension is a couple-of-lines patch :)

>>>>

>>> IMHO, keeping res_counter simple is better.

>>>

>>> Is this kind of new entry in mem_cgroup not good ?

>>> ==

>>> struct mem_cgroup {

>>> ...

>>> struct res_counter memory_limit.

>>> struct res_counter swap_limit.

>>> ..

>>> }

>> I meant the same thing actually. By "nesting would affect" I

>> meant, that we might want to make res_counters hierarchical.

>>

>> That would kill two birds with one stone - we will make a true

>> hierarchical memory accounting and let charging of two counters

>> with one call.

>

> Hierarchical res_counter makes sense.

> Making it in simple/reasonable style will be our challenge.

I have this in my TODO list. Since this is not so urgent, then if you don't mind I can prepare the patches next week - after I set the git tree up. This change doesn't seem that big.

> Thanks,

> -Kame

>

>

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>

Subject: Re: Re: [RFC/PATCH] cgroup swap subsystem
Posted by [Paul Menage](#) on Thu, 06 Mar 2008 08:48:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Thu, Mar 6, 2008 at 12:38 AM, Pavel Emelyanov <xemul@openvz.org> wrote:

> > Hierarchical res_counter makes sense.
> > Making it in simple/reasonable style will be our challenge.
>
> I have this in my TODO list. Since this is not so urgent, then if you
> don't mind I can prepare the patches next week - after I set the git
> tree up. This change doesn't seem that big.
>

The change that you're referring to is allowing a cgroup to have a total memory limit for itself and all its children, and then giving that cgroup's children separate memory limits within that overall limit?

Paul

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>

Subject: Re: Re: [RFC/PATCH] cgroup swap subsystem
Posted by [Pavel Emelianov](#) on Thu, 06 Mar 2008 08:50:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

Paul Menage wrote:

> On Thu, Mar 6, 2008 at 12:38 AM, Pavel Emelyanov <xemul@openvz.org> wrote:
>> > Hierarchical res_counter makes sense.
>> > Making it in simple/reasonable style will be our challenge.
>>
>> I have this in my TODO list. Since this is not so urgent, then if you
>> don't mind I can prepare the patches next week - after I set the git
>> tree up. This change doesn't seem that big.
>>
>
> The change that you're referring to is allowing a cgroup to have a

> total memory limit for itself and all its children, and then giving
> that cgroup's children separate memory limits within that overall
> limit?

Yup. Isn't this reasonable?

Without this, if I'm a task in a 1GB limited cgroup, I can create a new one, set 2GB limit and spawn a kid into it (or move there myself) and be happy with 2GB of memory... With the proposed change, even if I set a 2GB for a subgroup it will not pass `_my_` (1GB) limit.

> Paul
>

Containers mailing list
Containers@lists.linux-foundation.org
<https://lists.linux-foundation.org/mailman/listinfo/containers>
