Subject: Re: [RFC PATCH 0/5] memcg: VM overcommit accounting and handling
Posted by KAMEZAWA Hiroyuki on Tue, 10 Jun 2008 00:14:27 GMT
View Forum Message <> Reply to Message

On Tue, 10 Jun 2008 01:32:58 +0200
Andrea Righi <righi.andrea@gmail.com> wrote:

>
> Provide distinct cgroup VM overcommit accounting and handling using the memory
> resource controller.
>

Could you explain the benefits of this even when we have memrlimit controller ?
(If unsure, see 2.6.26-rc5-mm1 and search memrlimit controller.)

And this kind of virtual-address-handling things should be implemented on
memrlimit controller (means not on memory-resource-controller.).
It seems this patch doesn't need to handle page_group.

Considering hierarchy, putting several kinds of features on one controller is
not good, I think. Balbir, how do you think ?

Thanks,
-Kame

> Patchset against latest Linus git tree.
>
> This patchset allows to set different per-cgroup overcommit rules and,
> according to them, it's possible to return a memory allocation failure (ENOMEM)
> to the applications, instead of always triggering the OOM killer via
> mem_cgroup_out_of_memory() when cgroup memory limits are exceeded.
>
> Default overcommit settings are taken from vm.overcommit_memory and
> vm.overcommit_ratio sysctl values. Child cgroups initially inherits the VM
> overcommit parent's settings.
>
> Cgroup overcommit settings can be overridden using memory.overcommit_memory and
> memory.overcommit_ratio files under the cgroup filesystem.
>
> For example:
>
> 1. Initialize a cgroup with 50MB memory limit:
> # mount -t cgroup none /cgroups -o memory
> # mkdir /cgroups/0
> # /bin/echo $$ > /cgroups/0/tasks
> # /bin/echo 50M > /cgroups/0/memory.limit_in_bytes
>

> 2. Use the "never overcommit" policy with 50% ratio:
> # /bin/echo 2 > /cgroups/0/memory.overcommit_memory
> # /bin/echo 50 > /cgroups/0/memory.overcommit_ratio
>
> Assuming we have no swap space, cgroup 0 can allocate up to 25MB of virtual
> memory. If that limit is exceeded all the further allocation attempts made by
> userspace applications will receive a -ENOMEM.
>
> 4. Show committed VM statistics:
> # cat /cgroups/0/memory.overcommit_as
> CommitLimit: 25600 kB
> Committed_AS: 9844 kB
>
> 5. Use "always overcommmit":
> # /bin/echo 1 > /cgroups/0/memory.overcommit_memory
>
> This is very similar to the default memory controller configuration: overcommit
> is allowed, but when there's no more available memory oom-killer is invoked.
>
> TODO:
> - shared memory is not taken in account (i.e. files in tmpfs)
>
> -Andrea
> --
> To unsubscribe from this list: send the line "unsubscribe linux-kernel" in
> the body of a message to majordomo@vger.kernel.org
> More majordomo info at http://vger.kernel.org/majordomo-info.html
> Please read the FAQ at http://www.tux.org/lkml/
>

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

---

Subject: Re: [RFC PATCH 0/5] memcg: VM overcommit accounting and handling
Posted by Balbir Singh on Tue, 10 Jun 2008 05:13:23 GMT
View Forum Message <> Reply to Message

KAMEZAWA Hiroyuki wrote:
> On Tue, 10 Jun 2008 01:32:58 +0200
> Andrea Righi <righi.andrea@gmail.com> wrote:
>
>> Provide distinct cgroup VM overcommit accounting and handling using the memory
>> resource controller.
>>
>

> Could you explain the benefits of this even when we have memrlimit controller ?
> (If unsure, see 2.6.26-rc5-mm1 and search memrlimit controller.)
>
> And this kind of virtual-address-handling things should be implemented on
> memrlimit controller (means not on memory-resource-controller.).
> It seems this patch doesn't need to handle page_group.
>
> Considering hierarchy, putting several kinds of features on one controller is
> not good, I think. Balbir, how do you think ?
>

I would tend to agree. With the memrlimit controller, can't we do this in user
space now? Figure out the overcommit value and based on that setup the memrlimit?

--
Warm Regards,
Balbir Singh
Linux Technology Center
IBM, ISTL

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

---

Subject: Re: [RFC PATCH 0/5] memcg: VM overcommit accounting and handling
Posted by Pavel Emelianov on Tue, 10 Jun 2008 07:52:25 GMT
View Forum Message <> Reply to Message

Balbir Singh wrote:
> KAMEZAWA Hiroyuki wrote:
>> On Tue, 10 Jun 2008 01:32:58 +0200
>> Andrea Righi <righi.andrea@gmail.com> wrote:
>>
>>> Provide distinct cgroup VM overcommit accounting and handling using the memory
>>> resource controller.
>>>
>> Could you explain the benefits of this even when we have memrlimit controller ?
>> (If unsure, see 2.6.26-rc5-mm1 and search memrlimit controller.)
>>
>> And this kind of virtual-address-handling things should be implemented on
>> memrlimit controller (means not on memory-resource-controller.).
>> It seems this patch doesn't need to handle page_group.
>>
>> Considering hierarchy, putting several kinds of features on one controller is
>> not good, I think. Balbir, how do you think ?
>>

>
> I would tend to agree. With the memrlimit controller, can't we do this in user
> space now? Figure out the overcommit value and based on that setup the memrlimit?

I also agree with Balbir and Kamezawa. Separate controller for VM (i.e. vma-s
lengths) is more preferable, rather than yet another fancy feature on top of
the existing rss one.

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

---

Subject: Re: [RFC PATCH 0/5] memcg: VM overcommit accounting and handling
Posted by Andrea Righi on Tue, 10 Jun 2008 08:30:29 GMT
View Forum Message <> Reply to Message

Pavel Emelyanov wrote:
> Balbir Singh wrote:
>> KAMEZAWA Hiroyuki wrote:
>>> On Tue, 10 Jun 2008 01:32:58 +0200
>>> Andrea Righi <righi.andrea@gmail.com> wrote:
>>>
>>>> Provide distinct cgroup VM overcommit accounting and handling using the memory
>>>> resource controller.
>>>>
>>> Could you explain the benefits of this even when we have memrlimit controller ?
>>> (If unsure, see 2.6.26-rc5-mm1 and search memrlimit controller.)
>>>
>>> And this kind of virtual-address-handling things should be implemented on
>>> memrlimit controller (means not on memory-resource-controller.).
>>> It seems this patch doesn't need to handle page_group.
>>>
>>> Considering hierarchy, putting several kinds of features on one controller is
>>> not good, I think. Balbir, how do you think ?
>>>
>> I would tend to agree. With the memrlimit controller, can't we do this in user
>> space now? Figure out the overcommit value and based on that setup the memrlimit?
>
> I also agree with Balbir and Kamezawa. Separate controller for VM (i.e. vma-s
> lengths) is more preferable, rather than yet another fancy feature on top of
> the existing rss one.
>

Yep! it seems I totally miss the memrlimit controller. I was trying to
implement pretty the same functionalities, using a different approach.
However, I agree that a separate controller seems to be a better

solution.

Thank you all for pointing in the right direction. I'll test memrlimit controller and give a feedback.

-Andrea

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