
Subject: Re: Question : memrlimit cgroup's task_move (2.6.26-rc5-mm3)

Posted by [Balbir Singh](#) on Fri, 20 Jun 2008 13:33:55 GMT

[View Forum Message](#) <> [Reply to Message](#)

KAMEZAWA Hiroyuki wrote:

> On Thu, 19 Jun 2008 23:55:56 +0530

> Balbir Singh <balbir@linux.vnet.ibm.com> wrote:

>
>> * KAMEZAWA Hiroyuki <kamezawa.hiroyu@jp.fujitsu.com> [2008-06-19 12:14:35]:

>>

>>> I used memrlimit cgroup at the first time.

>>>

>>> May I ask a question about memrlimit cgroup ?

>>>

>> Hi, Kamezawa-San,

>>

>> Could you please review/test the patch below to see if it solves your
>> problem? If it does, I'll push it up to Andrew

>>

>

> At quick glance,

>> + /*

>> + * NOTE: Even though we do the necessary checks in can_attach(),

>> + * by the time we come here, there is a chance that we still

>> + * fail (the memrlimit cgroup has grown its usage, and the

>> + * addition of total_vm will no longer fit into its limit)

>> + */

> I don't like this kind of holes. Considering tests which are usually done

> by developpers, the problem seems not to be mentioned as "rare"..

> It seems we can easily cause Warning. right ?

>

> Even if you don't want to handle this case now, please mention as "TBD"

> rather than as "NOTE".

>

Honestly to fix this problem completely, we need transactional management in cgroups. Both can_attach() and attach() are called with cgroup_mutex held, but total_vm is changed with mmap_sem held.

What we can do is

1. Implement a routine attach_failed() in cgroups, that is called for each task for which can_attach() succeeded, if any of the can_attach() routine returns an error
2. Do the migration in can_attach() and unroll in attach_failed()

--

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>
