Subject: Re: [PATCH v5 0/8] per-cgroup tcp buffer pressure settings
Posted by KAMEZAWA Hiroyuki on Wed, 05 Oct 2011 00:29:54 GMT
View Forum Message <> Reply to Message

On Tue, 4 Oct 2011 16:17:52 +0400
Glauber Costa <glommer@parallels.com> wrote:

> [[v3: merge Kirill's suggestions, + a destroy-related bugfix]]
> [[v4: Fix a bug with non-mounted cgroups + disallow task movement]]
> [[v5: Compile bug with modular ipv6 + tcp files in bytes]]
>
> Kame, Kirill,
>
> I am submitting this again merging most of your comments. I've decided to
> leave some of them out:
> * I am not using res_counters for allocated_memory. Besides being more
>   expensive than what we need, to make it work in a nice way, we'd have
>   to change the !cgroup code, including other protocols than tcp. Also,
>
> * I am not using failcnt and max_usage_in_bytes for it. I believe the value
>   of those lies more in the allocation than in the pressure control. Besides,
>   fail conditions lie mostly outside of the memory cgroup's control. (Actually,
>   a soft_limit makes a lot of sense, and I do plan to introduce it in a follow
>   up series)
>
> If you agree with the above, and there are any other pressing issues, let me
> know and I will address them ASAP. Otherwise, let's discuss it. I'm always open.
>

I'm not familar with reuqirements of users. So, I appreciate your choices.
What I adivse you here is taking a deep breath. Making new version every day
is not good for reviewing process ;)
(It's now -rc8 and merge will not be so quick, anyway.)

At this stage, my concern is view of interfaces and documenation, and future plans.

Let me give a try explanation by myself. (Correct me ;)
I added some questions but I'm sorry you've already answered.

New interfaces are 5 files. All files exists only for non-root memory cgroup.

1. memory.independent_kmem_limit
2. memory.kmem.usage_in_bytes
3. memory.kmem.limit_in_bytes
4. memory.kmem.tcp.limit_in_bytes
5. memory.kmem.tcp.usage_in_bytes

* memory.independent_kmem_limit

If 1, kmem_limit_in_bytes/kmem_usage_in_bytes works.
If 0, kmem_limit_in_bytes/kmem_usage_in_bytes doesn't work and all kmem
  usages are controlled under memory.limit_in_bytes.

Question:
- What happens when parent/chidlren cgroup has different indepedent_kmem_limit ?
- What happens at creating a new cgroup with use_hierarchy==1.

* memory.kmem_limit_in_bytes/memory.kmem.tcp.limit_in_bytes

Both files works independently for _Now_. And memory.kmem_usage_in_bytes and memory.kmem_tcp.usage_in_bytes has no relationships.

In future plan, kmem.usage_in_bytes should includes tcp.kmem_usage_in_bytes.
And kmem.limit_in_bytes should be the limiation of sum of all kmem.xxxx.limit_in_bytes.

Question:
- Why this integration is difficult ?
  Can't tcp-limit-code borrows some amount of charges in batch from kmem_limit
  and use it ?

- Don't you need a stat file to indicate "tcp memory pressure works!" ?
  It can be obtained already ?

Thanks,
-Kame