Subject: Re: [PATCH 0/3] Fix problem with static_key decrement
Posted by Glauber Costa on Fri, 20 Apr 2012 15:01:50 GMT
View Forum Message <> Reply to Message

On 04/19/2012 07:54 PM, Tejun Heo wrote:
> On Thu, Apr 19, 2012 at 07:49:15PM -0300, Glauber Costa wrote:
>> Hi,
>>
>> This is my proposed fix for the sock memcg static_key
>> problem raised by Kamezawa. It works for me, but I would
>> Kame, please confirm.
>
> Please detail the problem. I don't follow what's the purpose here.
>

Ok.

1) Kame found the following bug: we were decrementing the jump label when the socket limit was set back to unlimited. The problem is that the sockets outlive the memcg, so we can only do that when the last reference count is dropped. It is worth mentioning that kmem controller for memcg will have the exact same problem - I am actually updating my series with all the results of this discussion here.

2) If, however, there are no sockets in flight, mem_cgroup_put() during ->destroy() will be the last one, and the decrementing will happen there.

3) static_key updates cannot happen with the cgroup_mutex held. This is because cpusets hold it from within the cpu_hotplug.lock - that static_keys take through get_online_cpus() in its cpu hotplug handler.

4) Looking at the cpusets code, it really seems necessary, at least by now.

5) Deferring all this to worker threads as you suggested in the cpu thread - that has a similar problem - can solve this problem, but in general, will create tons of others, like windows of inconsistent information.

That's basically it.