Subject: Re: [PATCH v2 02/13] memcg: Kernel memory accounting infrastructure.
Posted by Greg Thelen on Tue, 13 Mar 2012 17:00:58 GMT
View Forum Message <> Reply to Message

Glauber Costa <glommer@parallels.com> writes:
> 2) For the kernel itself, we are mostly concerned that a malicious container may
> pin into memory big amounts of kernel memory which is, ultimately,
> unreclaimable. In particular, with overcommit allowed scenarios, you can fill
> the whole physical memory (or at least a significant part) with those objects,
> well beyond your softlimit allowance, making the creation of further containers
> impossible.
> With user memory, you can reclaim the cgroup back to its place. With kernel
> memory, you can't.

In overcommit situations the page allocator starts failing even though memcg page can charge pages. When page allocations fail the oom killer plays a role. Page allocations can fail even without malicious usage of kernel memory (e.g. lots of mlock or anon without swap can fill a machine). I assume that the kernel memory pinned the malicious containers will be freed or at least become reclaimable once the processes in malicious containers are killed (oom or otherwise). We have been making use of the oom killer to save a system from irreconcilable overcommit situations.