Subject: Re: [PATCH v2 02/13] memcg: Kernel memory accounting infrastructure.
Posted by Glauber Costa on Tue, 13 Mar 2012 17:31:40 GMT

On 03/13/2012 09:00 PM, Greg Thelen wrote:
> 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.
If you overcommit mem+swap, yes. If you overcommit mem, no: reclaim
happens first. And we don't have that option with pinned kernel memory.

Of course you *can* run your system without swap, but the whole thing
exists exactly because there is a large enough # of ppl who wants to be
able to overcommit their physical memory, without failing allocations.