
Subject: Re: [PATCH v2 13/13] memcg: Document kernel memory accounting.
Posted by [Glauber Costa](#) on Sun, 11 Mar 2012 10:42:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 03/10/2012 12:39 AM, Suleiman Souhlal wrote:

```
> Signed-off-by: Suleiman Souhlal<suleiman@google.com>
> ---
> Documentation/cgroups/memory.txt | 44 ++++++-----
> 1 files changed, 40 insertions(+), 4 deletions(-)
>
> diff --git a/Documentation/cgroups/memory.txt b/Documentation/cgroups/memory.txt
> index 4c95c00..73f2e38 100644
> --- a/Documentation/cgroups/memory.txt
> +++ b/Documentation/cgroups/memory.txt
> @@ -74,6 +74,11 @@ Brief summary of control files.
>
>  memory.kmem.tcp.limit_in_bytes # set/show hard limit for tcp buf memory
>  memory.kmem.tcp.usage_in_bytes # show current tcp buf memory allocation
> + memory.kmem.usage_in_bytes # show current kernel memory usage
> + memory.kmem.limit_in_bytes # show/set limit of kernel memory usage
> + memory.kmem.independent_kmem_limit # show/set control of kernel memory limit
> +      (See 2.7 for details)
> + memory.kmem.slabinfo # show cgroup's slabinfo
>
> 1. History
>
> @@ -265,11 +270,19 @@ the amount of kernel memory used by the system. Kernel memory is
fundamentally
> different than user memory, since it can't be swapped out, which makes it
> possible to DoS the system by consuming too much of this precious resource.
>
> -Kernel memory limits are not imposed for the root cgroup. Usage for the root
> -cgroup may or may not be accounted.
> +Kernel memory limits are not imposed for the root cgroup.
>
> -Currently no soft limit is implemented for kernel memory. It is future work
> -to trigger slab reclaim when those limits are reached.
> +A cgroup's kernel memory is counted into its memory.kmem.usage_in_bytes.
> +
> +memory.kmem.independent_kmem_limit controls whether or not kernel memory
> +should also be counted into the cgroup's memory.usage_in_bytes.
> +If it is set, it is possible to specify a limit for kernel memory with
> +memory.kmem.limit_in_bytes.
> +
> +Upon cgroup deletion, all the remaining kernel memory becomes unaccounted.
> +
> +An accounted kernel memory allocation may trigger reclaim in that cgroup,
> +and may also OOM.
```

Why delete the softlimit bit? Since we're not shrinking, at least for the independent kmem case, we effectively don't do softlimits here. The file for it does not even exist...

```
>
> 2.7.1 Current Kernel Memory resources accounted
>
> @@ -279,6 +292,29 @@ per cgroup, instead of globally.
>
> * tcp memory pressure: sockets memory pressure for the tcp protocol.
>
> +* slab memory.
> +
> +2.7.1.1 Slab memory accounting
> +
> +Any slab type created with the SLAB_MEMCG_ACCT kmem_cache_create() flag
> +is accounted.
> +
> +Slab gets accounted on a per-page basis, which is done by using per-cgroup
> +kmem_caches. These per-cgroup kmem_caches get created on-demand, the first
> +time a specific kmem_cache gets used by a cgroup.
> +
> +Only slab memory that can be attributed to a cgroup gets accounted in this
> +fashion.
> +
> +A per-cgroup kmem_cache is named like the original, with the cgroup's name
> +in parentheses.
> +
> +When a cgroup is destroyed, all its kmem_caches get migrated to the root
> +cgroup, and "dead" is appended to their name, to indicate that they are not
> +going to be used for new allocations.
> +These dead caches automatically get removed once there are no more active
> +slab objects in them.
> +
> 3. User Interface
>
> 0. Configuration
```
