CPU Accounting Controller

The CPU accounting controller is used to group tasks using cgroups and account the CPU usage of these groups of tasks.

The CPU accounting controller supports multi-hierarchy groups. An accounting group accumulates the CPU usage of all of its child groups and the tasks directly present in its group.

Accounting groups can be created by first mounting the cgroup filesystem:

```
# mount -t cgroup -ocpuacct none /sys/fs/cgroup
```

With the above step, the initial or the parent accounting group becomes visible at /sys/fs/cgroup. At bootup, this group includes all the tasks in the system. /sys/fs/cgroup/tasks lists the tasks in this cgroup. /sys/fs/cgroup/cpuacct.usage gives the CPU time (in nanoseconds) obtained by this group which is essentially the CPU time obtained by all the tasks in the system.

New accounting groups can be created under the parent group /sys/fs/cgroup:

```
# cd /sys/fs/cgroup
# mkdir g1
# echo $$ > g1/tasks
```

The above steps create a new group g1 and move the current shell process (bash) into it. CPU time consumed by this bash and its children can be obtained from g1/cpuacct.usage and the same is accumulated in /sys/fs/cgroup/cpuacct.usage also.

cpuacct.stat file lists a few statistics which further divide the CPU time obtained by the cgroup into user and system times. Currently the following statistics are supported:

user: Time spent by tasks of the cgroup in user mode. system: Time spent by tasks of the cgroup in kernel mode.

user and system are in USER_HZ unit.

cpuacct controller uses percpu_counter interface to collect user and system times. This has two side effects:

- It is theoretically possible to see wrong values for user and system times. This is because percpu_counter_read() on 32bit systems isn't safe against concurrent writes.
- It is possible to see slightly outdated values for user and system times due to the batch processing nature of percpu_counter.