

Network priority cgroup

The Network priority cgroup provides an interface to allow an administrator to dynamically set the priority of network traffic generated by various applications

Nominally, an application would set the priority of its traffic via the `SO_PRIORITY` socket option. This however, is not always possible because:

1. The application may not have been coded to set this value
2. The priority of application traffic is often a site-specific administrative decision rather than an application defined one.

This cgroup allows an administrator to assign a process to a group which defines the priority of egress traffic on a given interface. Network priority groups can be created by first mounting the cgroup filesystem:

```
# mount -t cgroup -onet_prio none /sys/fs/cgroup/net_prio
```

With the above step, the initial group acting as the parent accounting group becomes visible at `/sys/fs/cgroup/net_prio`. This group includes all tasks in the system. `/sys/fs/cgroup/net_prio/tasks` lists the tasks in this cgroup.

Each `net_prio` cgroup contains two files that are subsystem specific

`net_prio.prioidx`

This file is read-only, and is simply informative. It contains a unique integer value that the kernel uses as an internal representation of this cgroup.

`net_prio.ifpriomap`

This file contains a map of the priorities assigned to traffic originating from processes in this group and egressing the system on various interfaces. It contains a list of tuples in the form `<iface priority>`. Contents of this file can be modified by echoing a string into the file using the same tuple format. For example:

```
echo "eth0 5" > /sys/fs/cgroups/net_prio/iscsi/net_prio.ifpriomap
```

This command would force any traffic originating from processes belonging to the `iscsi net_prio` cgroup and egressing on interface `eth0` to have the priority of said traffic set to the value 5. The parent accounting group also has a writeable `'net_prio.ifpriomap'` file that can be used to set a system default priority.

Priorities are set immediately prior to queueing a frame to the device queueing discipline (qdisc) so priorities will be assigned prior to the hardware queue selection being made.

One usage for the `net_prio` cgroup is with `mqprio` qdisc allowing application traffic to be steered to hardware/driver based traffic classes. These mappings can then be managed by administrators or other networking protocols such as DCBX.

A new `net_prio` cgroup inherits the parent's configuration.