

---

Subject: Re: IO scheduling

Posted by [Vasily Tarasov](#) on Thu, 21 Sep 2006 06:43:40 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi,

I'll try to explain you what is in Vserver concerning IO scheduling and why it isn't now in OpenVZ.

There is cfq scheduler in Linux kernel, that allows to assign IO priority to the process. It supports three classes:

real time (rt)

best effort (be)

idle class (i)

Within rt class and be class are 8 levels of priority. The more level is - more time for input/output particular process has. Additional information can be found for example at <http://www.mjmwired.net/kernel/Documentation/block/ioprio.txt>

So what do they do in Vserver?

When user sets certain IO priority to the context, Vserver framework just sets this IO priority to all processes in context! And that's all that they do, but this isn't right. Just look at these example:

1st context: 3 processes - priority be:4

2nd context: 1 process - priority be:6

So user expects that 2nd context has more IO bandwidth, but this isn't true, 'cause 1st context has more processes! And the more processes 1st context has more IO bandwidth it has.

Some time ago there were patches to do the same in OpenVZ, but do we need such implementation?

To create more sophisticated and true IO scheduling more investigation is necessary. Also there is also a big problem, 'cause pages can be written to the block device, when information about process isn't available any more...

HTH,

vass.