Subject: Re: "strong" Disk I/O prioritisation
Posted by Vasily Tarasov on Thu, 10 Apr 2008 06:34:08 GMT
View Forum Message <> Reply to Message

Hello Tony,

you wrote:
Quote:I want to be able to give absolute I/O priority to one (or a few) VE, and have it so that all I/O for other VEs gets serviced only when idle.

It is impossible to assign an absolute I/O priority to VE at the moment. When you use --ioprio option of vzctl, you assign a relative share of time, during which the VE in question can work with a block device.

Also current implementation doesn't support "idle" class of VEs, which are serviced only if nobody else uses the block device. This is in future plans.

you wrote
Quote:As a workaround, I tried using ionice within the VE. Even after adding the sys_admin and sys_nice capabilities, I still got
ionice -c3 id
ioprio_set: Operation not permitted

That's a good point! Thank you for noticing. The thing is that before prioritization was introduced in OpenVZ, sys_ioprio_set() system call (which is used by ionice utility) was prohibited inside VE for understandable reasons. But now, we can allow that!
You can comment
if (!ve_is_super(get_exec_env()))
return -EPERM; check in ./fs/ioprio.c file to check if it will help. I personally think, that setting a priority of the processes in VE will not help in your situation a lot.

Now several words about your tests and their results:
1) When you do cat for the first time, some parts or even the whole file is in cache. So, 2nd time you do cat, it is not read from the disk, but from the main memory. It can introduce significant distortions.
2) In implementation that works now in OpenVZ you can notice the effect of prioritization much more, if you will run not one "disk-reader" (as in your test) in VE, but several of them. It is not the feature, but a drawback, and we're working on improvements in this area.

you wrote:
Quote:Can I do better than this in brutally prioritising one VE over another? My rationale for this is to protect my production webserver (two heavily used phpbb websites) from my other VEs. Currently (linux-vserver, Ubuntu 6.06), simply copying a file of moderate size (more than 500MB) brings the prod webserver to its knees for minutes, which makes me unpopular.

I understand you rationale very good. The I/O prioritization in Linux is on the blooding edge, so there is not perfect solution now. If current OpenVZ prioritizations is not enough for you, you can move your production VE (or all other VEs) to a separate hard drive.

HTH!