Subject: Re: [PATCH 2/2] dm-ioband: I/O bandwidth controller v0.0.4: Document
Posted by Ryo Tsuruta on Mon, 28 Apr 2008 08:06:27 GMT
View Forum Message <> Reply to Message

Hi,

> Most writes are performed by pdflush, kswapd, etc. This will lead to large
> inaccuracy.
>
> It isn't trivial to fix. We'd need deep, long tracking of ownership
> probably all the way up to the pagecache page. The same infrastructure
> would be needed to make Sergey's "BSD acct: disk I/O accounting" vaguely
> accurate. Other proposals need it, but I forget what they are.

I also realize that some kernel threads such as pdflush perform actual
writes instead of tasks which originally issued write requests.
So, taka is developing a blocking I/O tacking down mechanism which is
based on cgroup memory controller and posted it to LKML:
http://lwn.net/Articles/273802/

However, the current implementation works well with Xen virtual
machines, because virtual machine's I/Os are issued from its own kernel
thread and can be tracked down. Please see a benchmark result of Xen
virtual machine:
http://people.valinux.co.jp/~ryov/dm-ioband/benchmark/xen-blktap.html

As for KVM, dm-ioband was also able to track down block I/Os as I
expected. When dm-ioband is used in virtual machine environment,
I think even the current implementation will work fairly.

But unfortunately I found KVM still had a performance problem that
it couldn't handle I/Os efficiently yet, which should be improved.
I already posted this problem to kvm-devel list:
http://sourceforge.net/mailarchive/forum.php?thread_name=20080229.210531.226799765.ryov%40valinux.co.jp&forum_name=kvm-devel

> Much more minor points: when merge-time comes, the patches should have the
> LINUX_VERSION_CODE stuff removed. And probably all of the many `inline's
> should be removed.

Thank you for your advice. I'll have these fixes included in the next
release.

Ryo Tsuruta