Subject: D-state processes on i2o servers
Posted by iurly on Sun, 29 Oct 2006 23:48:28 GMT
View Forum Message <> Reply to Message

Hi,

we are running Virtuozzo on a quad-core 3.20 GHz Xeon with 8 GBs of RAM and a SmartRAID V i2o controller.
The system hosts about a dozen VPSes used as individual software factories which we use to edit and compile a software package.

I don't know if simultaneous compilations of a project way bigger than the Linux kernel are a likely scenario of a Virtuozzo/OpenVZ
installation, but given the available resources I would expect reasonable performance.

What happens is that (apparently when the workload is high) the machine freezes for 5-10 seconds, then starts working again, then freezes again, and so on. These hiccups render the system unusable (and/or people frustrated) for several minutes.

I have been trying to trace this, and it seems like when this happens there are several processes stuck in the "D" state, waiting on "__wait_on_buffer". Somehow, this seems related to intense I/O activity on the disk.

However, even the most innocent processes (vi for instance) suffer from this problem, which is again very frustrating because one would expect at least lightweight programs to be responsive, but this is not the case.

We are running a 2.6.8-022stab070.4-enterprise kernel with i2o_block v.0.9 (don't know about the version of other i2o modules).

I also noticed the following lines in dmesg:

mtrr: type mismatch for dd000000,1000000 old: uncachable new: write-combining
i2o: could not enable write combining MTRR

Could this be the cause of the problem?

Could this behavior be related to the fault pointed out in
(http://forum.openvz.org/index.php?&t=msg&th=914)?

Note that in our case the machine does not hang forever, it just freezes for a few seconds.
Has this been solved in recent kernels?

If anyone has any idea on how to fix (or improve) this, it would be highly appreciated!

Thank you,
Gerlando

Subject: Re: D-state processes on i2o servers
Posted by Vasily Tarasov on Mon, 30 Oct 2006 06:46:17 GMT
View Forum Message <> Reply to Message

Hello,

if you're using Virtuozzo, not OpenVZ it's much more efficient to ask SWsoft official support, than ask it here.

Subject: Re: D-state processes on i2o servers
Posted by vaverin on Thu, 02 Nov 2006 10:04:13 GMT
View Forum Message <> Reply to Message

I would note that we have updated i2o drivers in 022stab078.20 kernel, I've back-ported drivers from latest mainstream. According to our customers new drivers works much better than old ones and I would like to recommend you to update the kernel.

thank you,
Vasily Averin

Subject: Re: D-state processes on i2o servers
Posted by iurly on Thu, 02 Nov 2006 18:47:37 GMT
View Forum Message <> Reply to Message

Hi,

thanks for your support.
I tried upgrading and switch to cfq scheduling policy, but
I noticed that the MTRR feature is not enabled anymore.
Is this correct?

My general feeling is that write-caching on the disk is disabled, so when the I/O pressure is high the impact on machine responsiveness is noticeable.

How could I check if this is the case or not? I mean, how can I measure disk I/O throughput when there are several concurrent accesses?

Also, could the overhead of vzfs be (at least partially) responsible?

As a side note, on the hardware node we are using an ext3 filesystem, and our VPSes make heavy use of LOTS of small files and LOTS of symlinks. Should we use a different filesystem to improve performance?

Thanks again!
Gerlando

Subject: Re: D-state processes on i2o servers
Posted by vaverin on Wed, 08 Nov 2006 09:13:02 GMT
View Forum Message <> Reply to Message

Hello Gerlando,
sorry for a long delay, but I do not have enough time for OpenVZ forum, I would like to recommend you to access our support.
Quote:
I noticed that the MTRR feature is not enabled anymore.
Is this correct?

I've seen this behavior too, but I don't know the correct answer on your question. I've found in google that i2o developer Markus Lidel asked the question about MTRR, but I don't found any answers.

However I do not think that it may lead to the some problems. Drivers were taken from mainstream kernel, and plain mainstream kerenl work by the same manner. We have tested new driver well and did not noticed any troubles. And we have a positive customers feedback.
Quote:
My general feeling is that write-caching on the disk is disabled, so when the I/O pressure is high the impact on machine responsiveness is noticeable.

How could I check if this is the case or not? I mean, how can I measure disk I/O throughput when there are several concurrent accesses?

As far as I understand write-caching on the disk should have a very low effect to the disk I/O throughput in case of heavy disk IO, just because of the cache will not used in this situation, new data will replace old cache content without any data-reusing.

As far as I know there is some tests for various filesystem operations (bonnie?), you can use it for measurements. However I would note that the IO performance depends vastly on where the data placed physically on the disk. Therefore it is very hard to analyze the test results.
Quote:
Also, could the overhead of vzfs be (at least partially) responsible?

As a side note, on the hardware node we are using an ext3 filesystem, and our VPSes make heavy use of LOTS of small files and LOTS of symlinks. Should we use a different filesystem to improve performance?

vzfs is a cow-like filesytem and it should not have a noticeable overhead.

Also I would note that ext3 performance is not so bad, but it's stability is much better than any other alternatives.

Therefore if you wish to improve filesystem performance i would like to recommend you think about disk IO-sysbsytem upgrade instead of the filesystem change.

Also you can store VPS date on the dedicated disks or disk partitions, it can decrease interference between various VPSes.

Thank you,
Vasily Averin