

---

Subject: Re: seems to be a flaw in cfq

Posted by [Vasily Tarasov](#) on Wed, 20 Dec 2006 07:09:57 GMT

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

---

Hello, Jens.

Jens Axboe wrote:

>

> Back after dinner, the fresh energy served it's purpose - I think I know  
> what the issue is. We allow merging across process queues, which will  
> effectively serialize some io if they are sync (like this case). I'll  
> hack up a fix for current git and give it a test spin, to verify that  
> this is the problem here.

>

As far as I understand merging is an often event in this situation only  
because of sequential  
reading, Below is a job-file with rw=randread and direct=false. However  
we got the same bizarre results.

Job-file:

; cfq bug reproduce

[global]  
rw=randread  
direct=0

timeout=200

[reader1]  
filename=file1

[reader2]  
filename=file2

[reader3]  
filename=file3

; --- this two reads the same file ---  
[reader4]  
filename=file4

[reader5]  
filename=file4

; -----

[reader6]  
filename=file6

[reader7]  
filename=file7

[reader8]  
filename=file8

[reader9]  
filename=file9

[reader10]  
filename=file10

[reader11]  
filename=file11

Results:

```
# fio job1.file
reader1: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader2: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader3: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader4: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader5: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader6: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader7: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader8: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader9: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader10: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
reader11: (g=0): rw=randread, odir=0, bs=4K-4K/4K-4K, rate=0,
ioengine=sync, iodepth=1
Starting 11 threads
Threads running: 11: [rrrrrrrrrr] [100.00% done] [ 528/ 0 kb/s]
[eta 00m:00s]
reader1: (groupid=0): err= 0:
  read : io= 10MiB, bw= 53KiB/s, runt=200006msec
```

```

slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 0, max= 988, avg=75.54, dev=244.04
bw (KiB/s) : min= 25, max= 70, per=10.87%, avg=53.18, dev=54.17
cpu       : usr=0.00%, sys=0.07%, ctx=2888
reader2: (groupid=0): err= 0:
read : io= 10MiB, bw= 54KiB/s, runt=200067msec
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 0, max= 1034, avg=74.57, dev=241.93
bw (KiB/s) : min= 29, max= 90, per=11.04%, avg=54.01, dev=55.09
cpu       : usr=0.00%, sys=0.07%, ctx=2925
reader3: (groupid=0): err= 0:
read : io= 10MiB, bw= 53KiB/s, runt=200039msec
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 0, max= 965, avg=75.64, dev=243.64
bw (KiB/s) : min= 29, max= 86, per=10.91%, avg=53.34, dev=54.48
cpu       : usr=0.00%, sys=0.07%, ctx=2885
reader4: (groupid=0): err= 0:
read : io= 1MiB, bw= 5KiB/s, runt=200049msec
<<<<<
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 3, max= 1705, avg=716.52, dev=894.16
bw (KiB/s) : min= 2, max= 35, per=1.14%, avg= 5.59, dev= 7.47
cpu       : usr=0.00%, sys=0.00%, ctx=369
reader5: (groupid=0): err= 0:
read : io= 1MiB, bw= 5KiB/s, runt=200049msec
<<<<<<
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 3, max= 1705, avg=716.52, dev=894.17
bw (KiB/s) : min= 2, max= 35, per=1.14%, avg= 5.59, dev= 7.47
cpu       : usr=0.00%, sys=0.00%, ctx=366
reader6: (groupid=0): err= 0:
read : io= 10MiB, bw= 52KiB/s, runt=200055msec
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 1, max= 969, avg=77.21, dev=246.03
bw (KiB/s) : min= 22, max= 97, per=10.74%, avg=52.50, dev=53.67
cpu       : usr=0.00%, sys=0.08%, ctx=2832
reader7: (groupid=0): err= 0:
read : io= 10MiB, bw= 53KiB/s, runt=200011msec
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 0, max= 1047, avg=76.68, dev=245.52
bw (KiB/s) : min= 18, max= 77, per=10.76%, avg=52.60, dev=53.61
cpu       : usr=0.00%, sys=0.09%, ctx=2852
reader8: (groupid=0): err= 0:
read : io= 10MiB, bw= 53KiB/s, runt=200072msec
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00
clat (msec): min= 0, max= 1033, avg=76.68, dev=245.69
bw (KiB/s) : min= 7, max= 71, per=10.71%, avg=52.37, dev=53.40
cpu       : usr=0.00%, sys=0.08%, ctx=2851

```

reader9: (groupid=0): err= 0:  
read : io= 10MiB, bw= 52KiB/s, runt=200022msec  
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00  
clat (msec): min= 0, max= 973, avg=76.90, dev=245.66  
bw (KiB/s) : min= 5, max= 188, per=10.76%, avg=52.63, dev=54.71  
cpu : usr=0.00%, sys=0.06%, ctx=2841  
reader10: (groupid=0): err= 0:  
read : io= 9MiB, bw= 52KiB/s, runt=200028msec  
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00  
clat (msec): min= 0, max= 972, avg=78.06, dev=247.65  
bw (KiB/s) : min= 5, max= 106, per=10.58%, avg=51.72, dev=53.05  
cpu : usr=0.00%, sys=0.09%, ctx=2807  
reader11: (groupid=0): err= 0:  
read : io= 10MiB, bw= 52KiB/s, runt=200084msec  
slat (msec): min= 0, max= 0, avg= 0.00, dev= 0.00  
clat (msec): min= 0, max= 979, avg=77.57, dev=247.05  
bw (KiB/s) : min= 4, max= 73, per=10.60%, avg=51.82, dev=52.90  
cpu : usr=0.00%, sys=0.07%, ctx=2822

Run status group 0 (all jobs):

READ: io=93MiB, aggrb=489, minb=5, maxb=54, mint=200006msec,  
maxt=200084msec

Disk stats (read/write):

sda: ios=26139/56, merge=0/20, ticks=2000184/30587, in\_queue=2030770,  
util=100.00%

Thank you,  
Vasily.

---