Linux Filesystems benchmarks, part III

(Visit http://www.lourdas.name to read parts I and II)

System configuration

$ uname -a
$ Linux bench 2.6.32-gentoo #2 SMP PREEMPT Wed Dec 9 21:55:01 EET 2009 x86_64 Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz GenuineIntel GNU/Linux

4GB physical RAM

Kernel was compiled with all filesystem types included (not modules).

$ mkfs.ext3 -V
$ mke2fs 1.41.9 (22-Aug-2009)
Using EXT2FS Library version 1.41.9

$ mkfs.xfs -V
$ mkfs.xfs version 2.10.1

$ mkfs.jfs -V
$ /sbin/mkfs.jfs version 1.1.14, 06-Apr-2009

$ mkfs.reiserfs -V
$ mkfs.reiserfs 3.6.19 (2003 www.namesys.com)

$ hdparm -i /dev/sdb

/dev/sdb:

Model=WDC WD2500KS-00MJB0, FwRev=02.01C03, SerialNo=WD-WCANK2784629
Config={ HardSect NotMFM HdSw>15uSec SpinMotCtl Fixed DTR>5Mbs FmtGapReq }
RawCHS=16383/16/63, TrkSize=0, SectSize=0, ECCbytes=50
BuffType=unknown, BuffSize=16384kB, MaxMultSect=16, MultSect=off
CurCHS=16383/16/63, CurSects=16514064, LBA=yes, LBAsects=488397168
IORDY=on/off, tPIO={min:120,w/IORDY:120}, tDMA={min:120,rec:120}
PIO modes: pio0 pio3 pio4
DMA modes: mdma0 mdma1 mdma2
UDMA modes: udma0 udma1 udma2 udma3 udma4 udma5 *udma6
AdvancedPM=no WriteCache=enabled
Drive conforms to: Unspecified: ATA/ATAPI-1,2,3,4,5,6,7

* signifies the current active mode

The benchmark described here used iozone and 3 more “real-world” activities to measure each filesystem's performance. Details follow.

iozone

iozone was used as a benchmark tool against a 49GB partition of the above mentioned hard disk. After each filesystem creation the partition was mounted under /media/BENCH and unmounted for the creation of the next filesystem. The mount parameters for all filesystems where **noatime,nodiratime**, except for the **xfs**** filesystem, which was mounted using the additional parameters **logbufs=8,logbsize=262144**.
The command used for each filesystem benchmark was:
iozone -b result.xls -f /media/BENCH/test -i 0 -i 1 -i 2 -i 4 -i 6 -i 7 -i 8 -p -a -n 4k -g 64m -O -R
Explanation:
-b result.xls: output the result in an Excel file named result.xls
-f /media/BENCH/test: the temporary file that is used for the benchmark
-i 0 -i 1 -i 2 -i 4 -i 6 -i 7 -i 8: use tests with numbers 0=write/rewrite, 1=read/re-read, 2=random-read/write, 4=Re-write-record, 6=fwrite/re-fwrite, 7=fread/Re-fread, 8=random mix. The tests are described in the iozone website documentation (http://www.iozone.org).
-p: purges the processor cache before each file operation
-a: use the automatic mode for tests
-n 4k: minimum filesize for the automatic mode for tests (4KB)
-g 64m: maximum filesize for the automatic mode for tests (64MB)
-O: outputs the results in operations per second
-R: generates the report in an Excel file
iozone benchmarked the filesystems using a file size of 4KB up to including 64MB (in powers of 2, that is 4KB, 8KB, 16KB, 32KB, 64KB, 128KB, 256KB, 512KB, 1024KB, 2048KB, 4096KB, 8192KB, 16384KB, 32768KB and 65536KB) using different record size each time. The record size ranged from 4KB up to 16MB (in powers of 2, that is 4KB, 8KB, 16KB, 32KB, 64KB, 128KB, 256KB, 512KB, 1024KB, 2048KB, 4096KB, 8192KB and 16384KB). Because of this great variation of file and record sizes, I kept the measurements of all file sizes for record sizes of 4KB, 8KB, 16KB, 64KB, 512KB and 2048KB. There are no measurements for file sizes using larger record sizes of the file size itself (eg. no measurements for a 16KB file using record sizes greater than 16KB).

Custom benchmarks

For each filesystem, cp was used to copy a 1.9GB file (2041008128 bytes) from another disk partition to the 49GB test partition. The command used to measure the time needed for the copy was:

time (cp 8.0-RELEASE-i386-dvd1.iso /media/BENCH && sync)

This command was executed 3 times the average of the 3 results was taken as the final result.

For each filesystem, tar was used to uncompress the Linux kernel 2.6.32 to the test partition. The command used to measure the time needed for the uncompression was:

time (tar xjf linux-2.6.32.tar.bz2 -C /media/BENCH && sync)

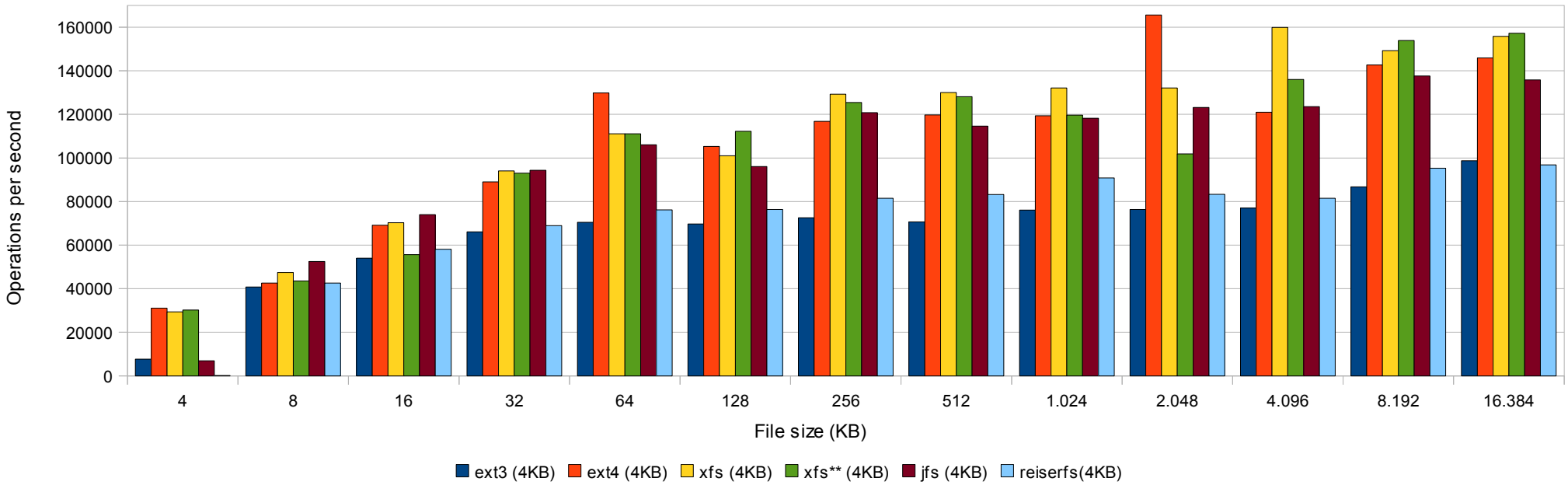
This command was executed 3 times the average of the 3 results was taken as the final result.

For each filesystem, rm was used to delete the directory structure of the uncompressed qt-everywhere-opensource-src-4.6.0.tar.gz file, which contained **30786 files** and **1657 directories**. The command used to measure the time needed for the removal was:

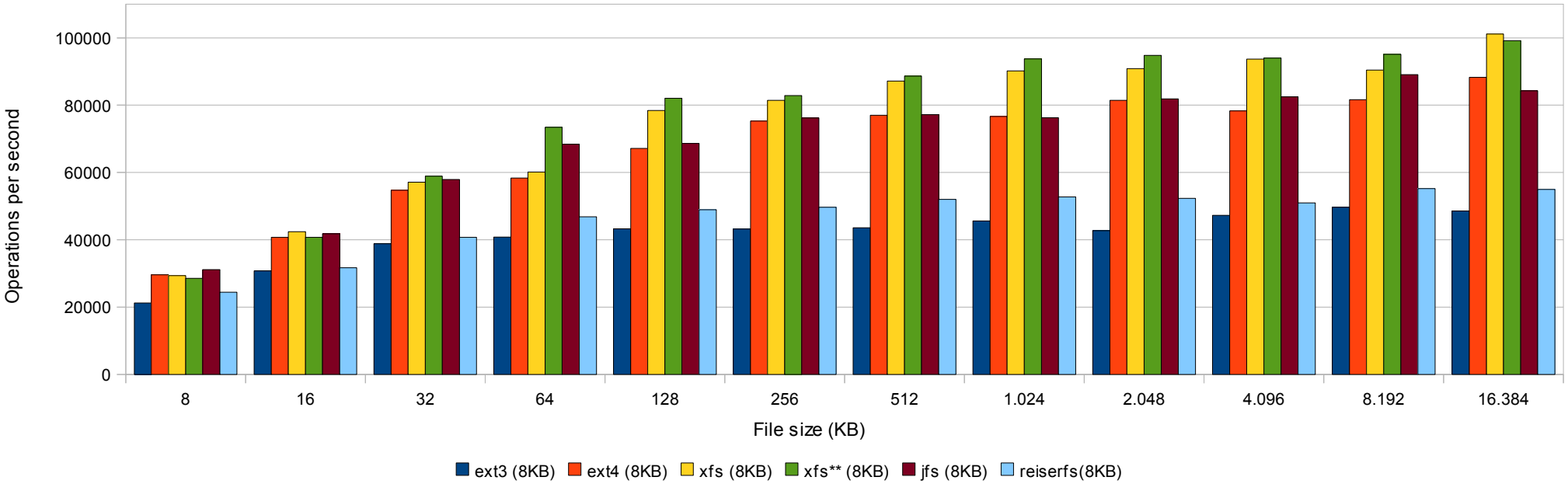
time (rm -rf /media/BENCH/qt-everywhere-opensource-src-4.6.0/ && sync)

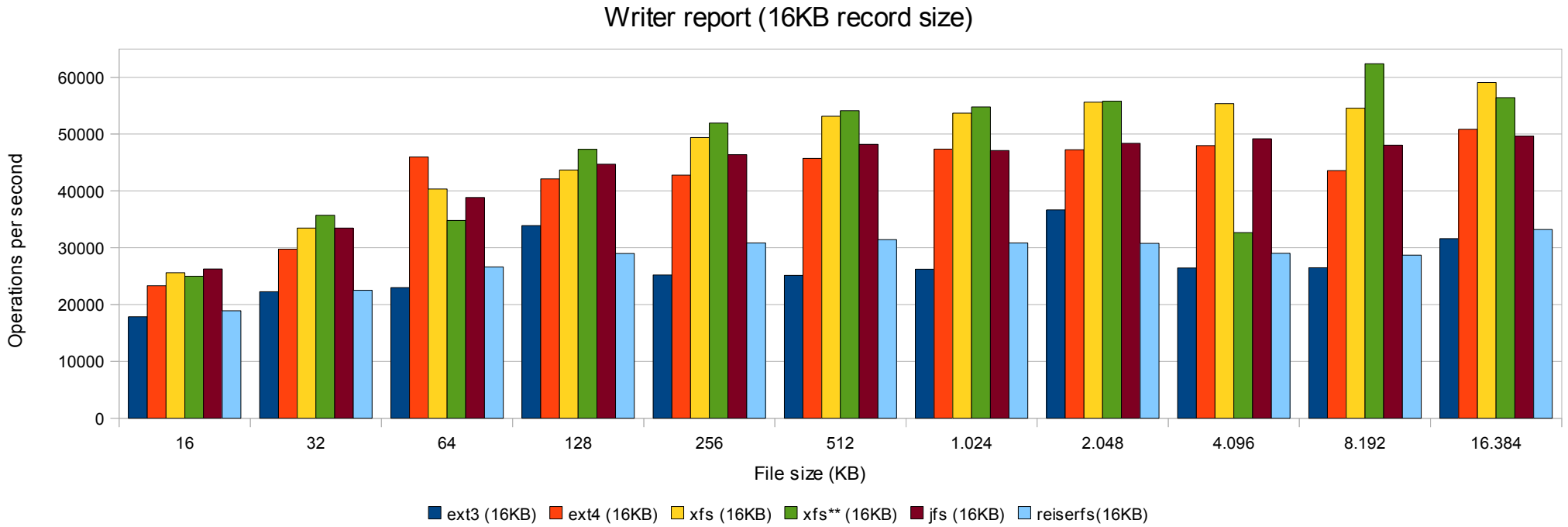
This command was executed 3 times the average of the 3 results was taken as the final result.

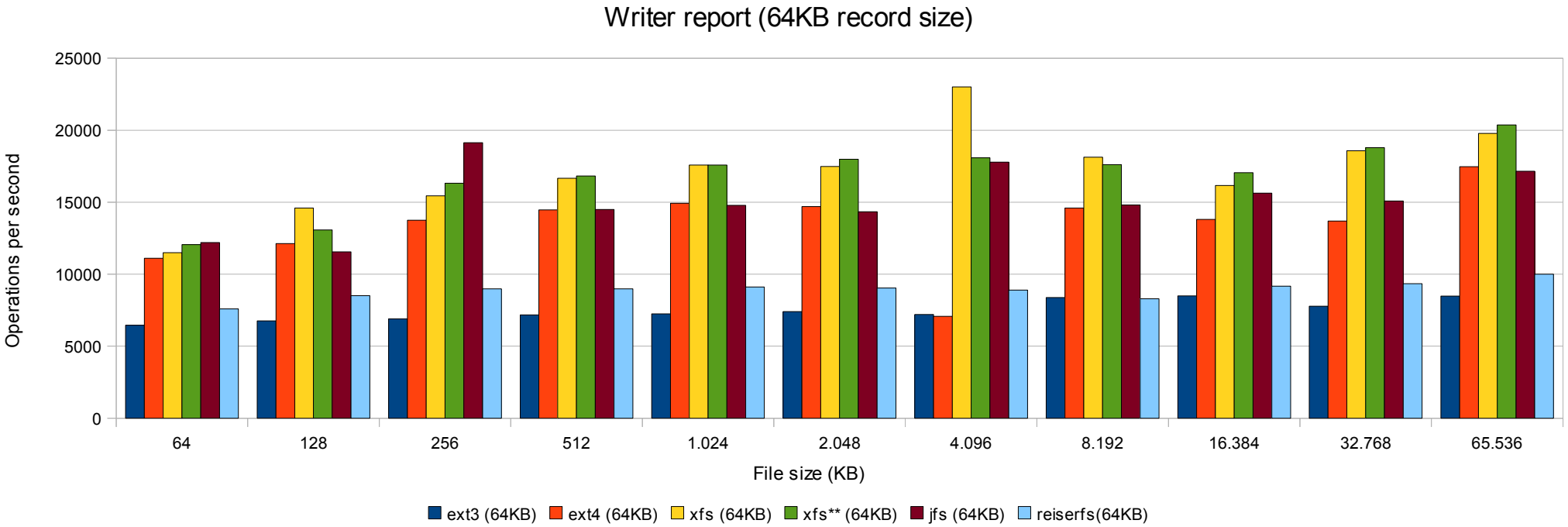
Writer report (4KB record size)



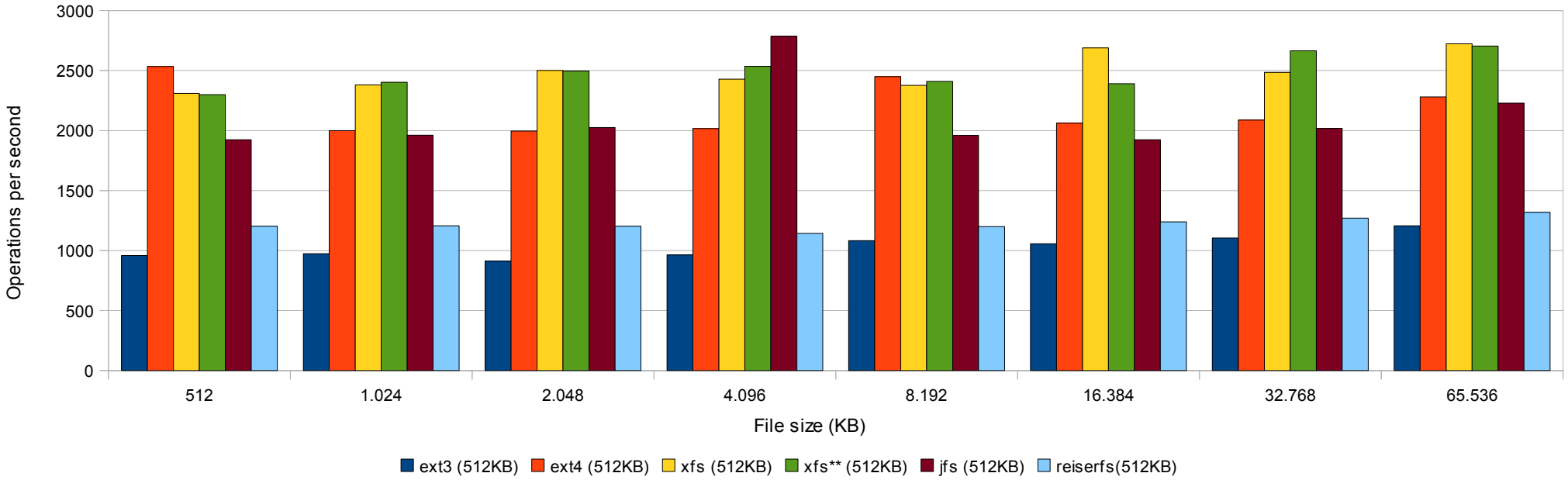
Writer report (8KB record size)



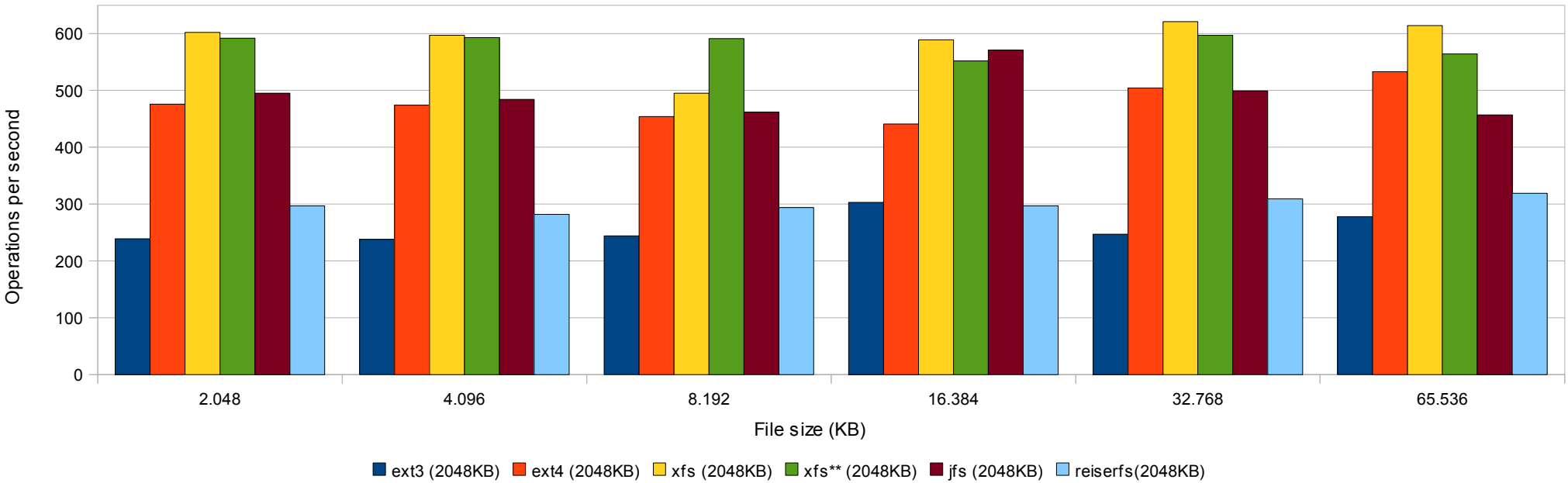

Writer report (16KB record size)
Operations per second
60000
50000
40000
30000
20000
10000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)


Writer report (64KB record size)
Operations per second
25000
20000
15000
10000
5000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

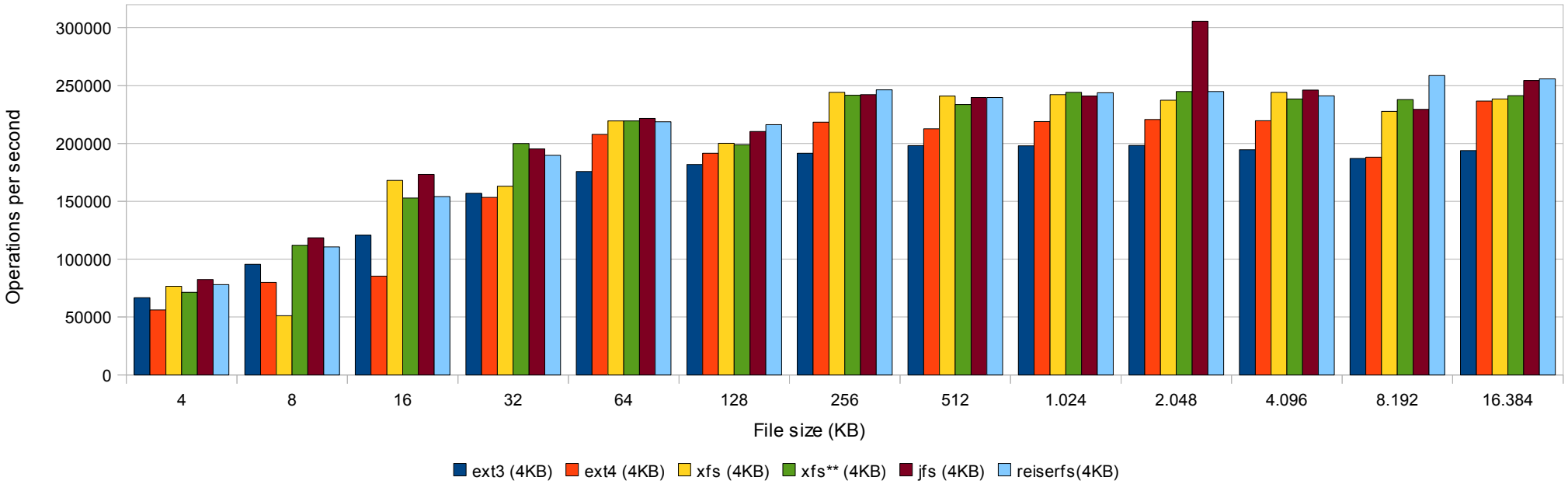
Writer report (512KB record size)



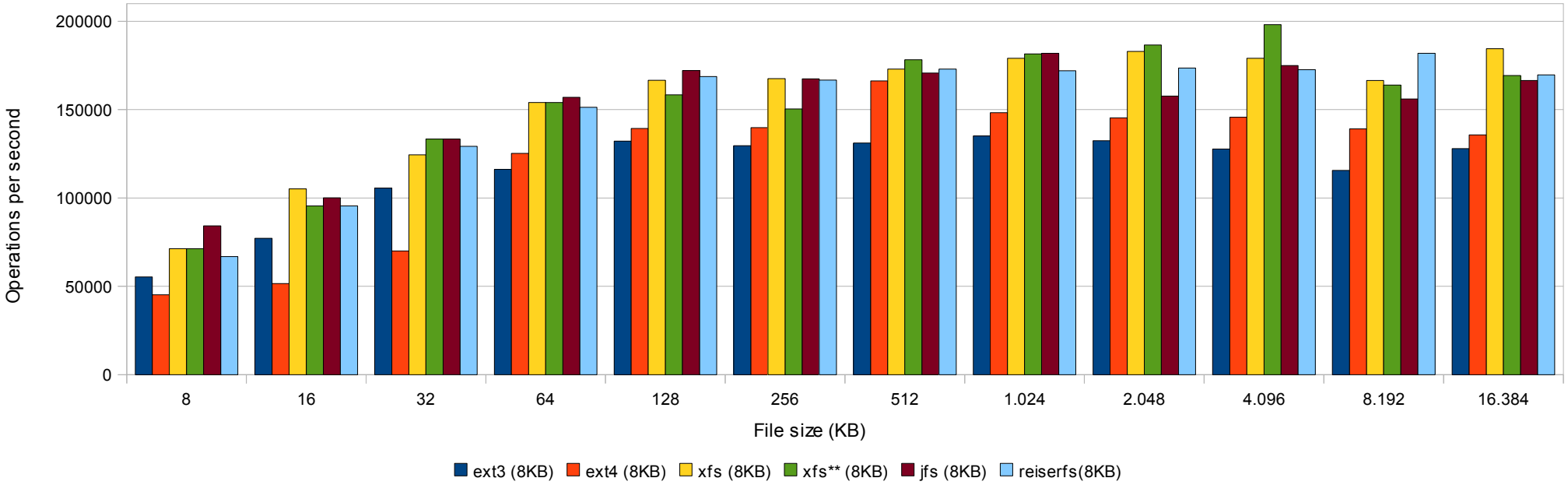
Writer report (2048KB record size)



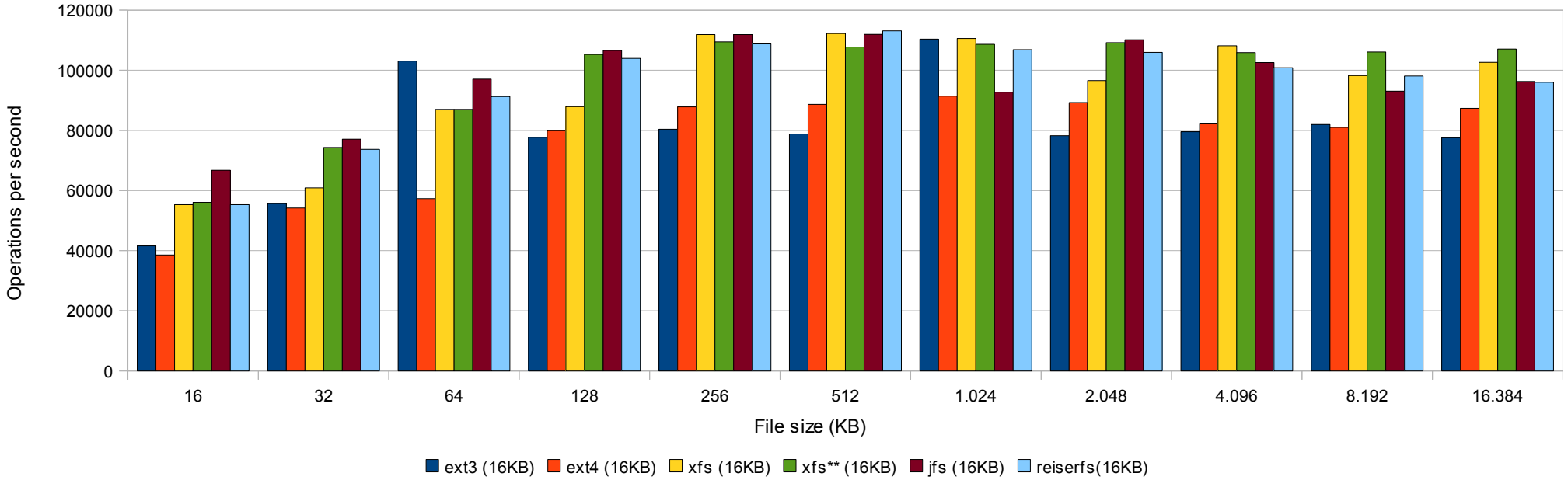
Rewriter report (4KB record size)



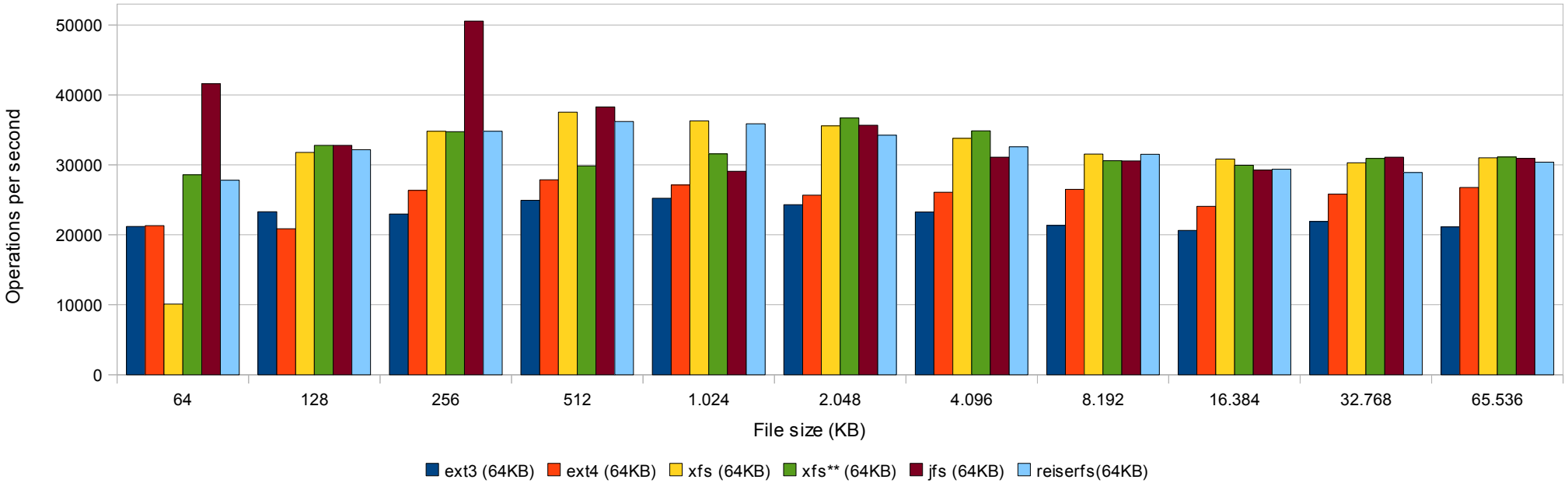
Rewriter report (8KB record size)



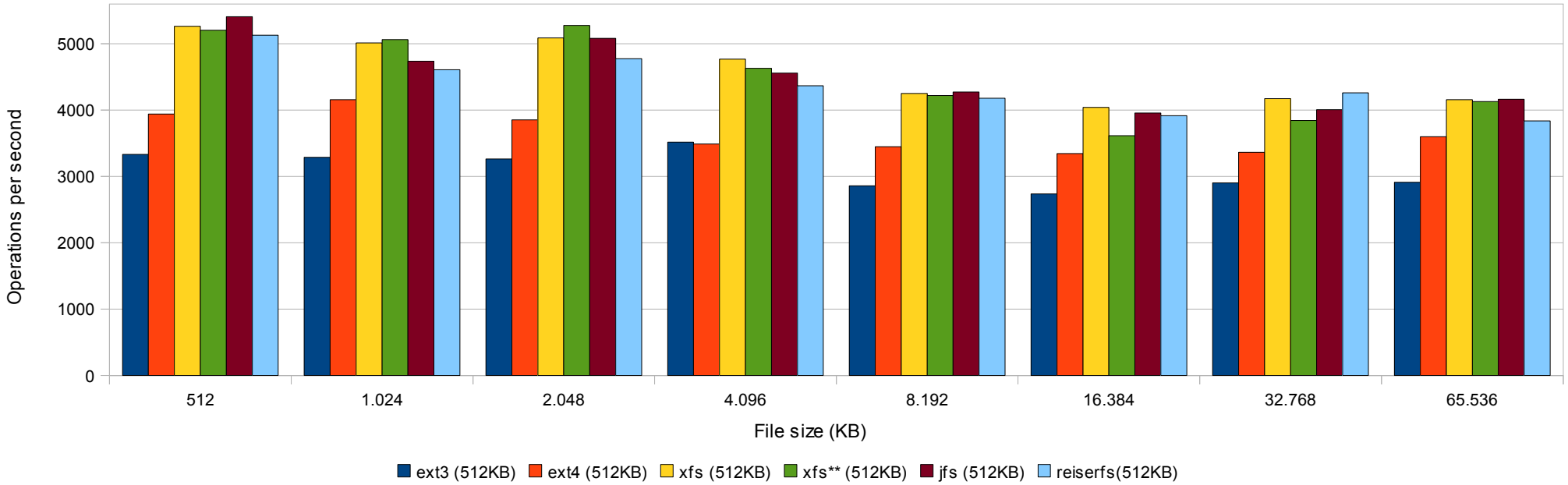
Rewriter report (16KB record size)



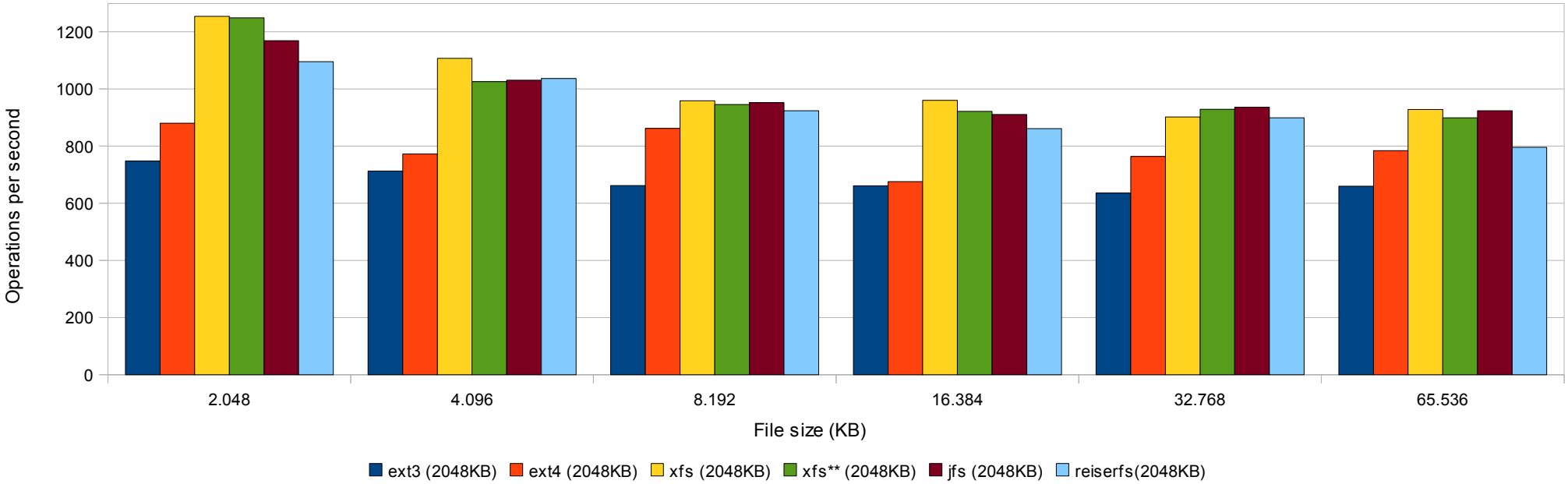
Rewriter report (64KB record size)



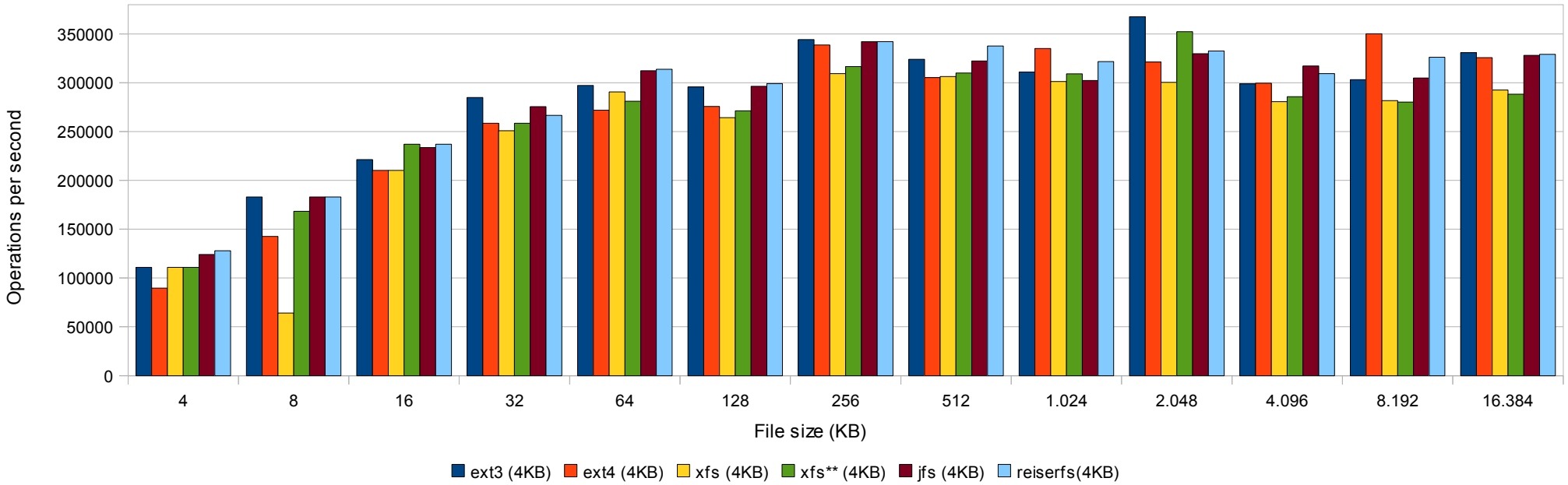
Rewriter report (512KB record size)



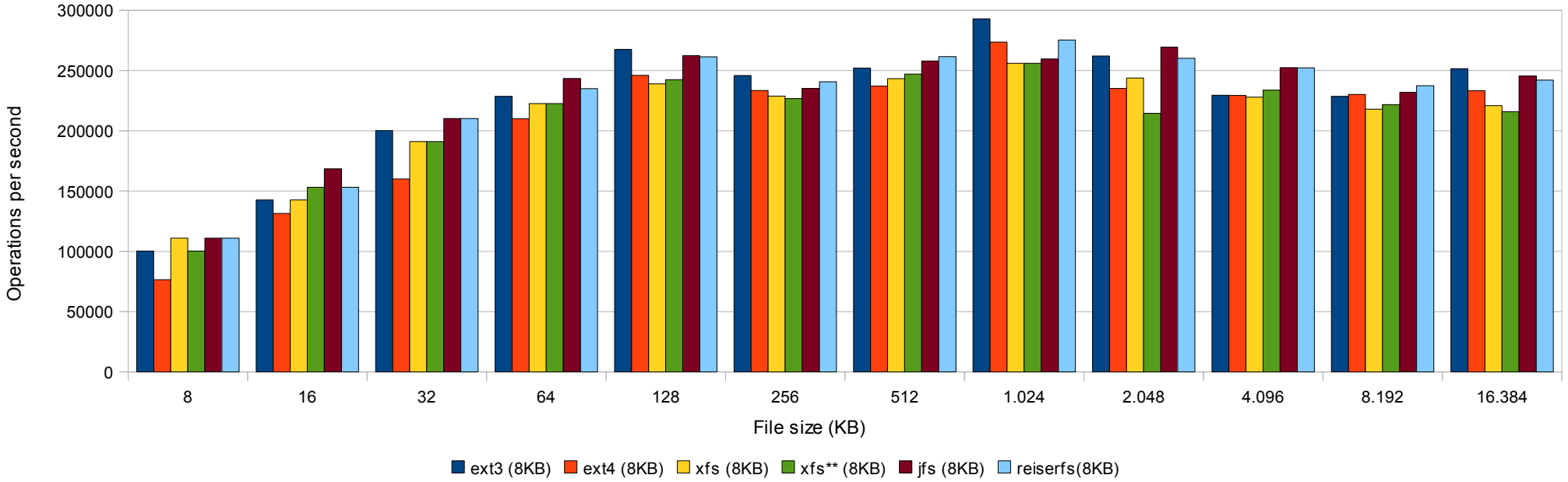
Rewriter report (2048KB record size)



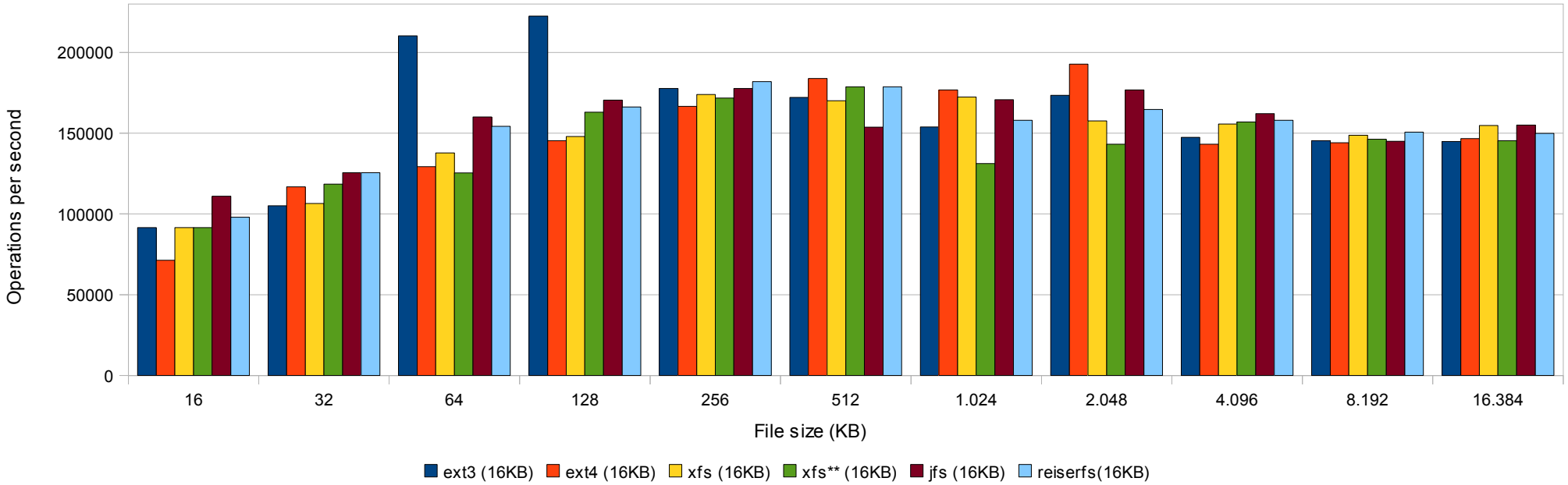
Reader report (4KB record size)



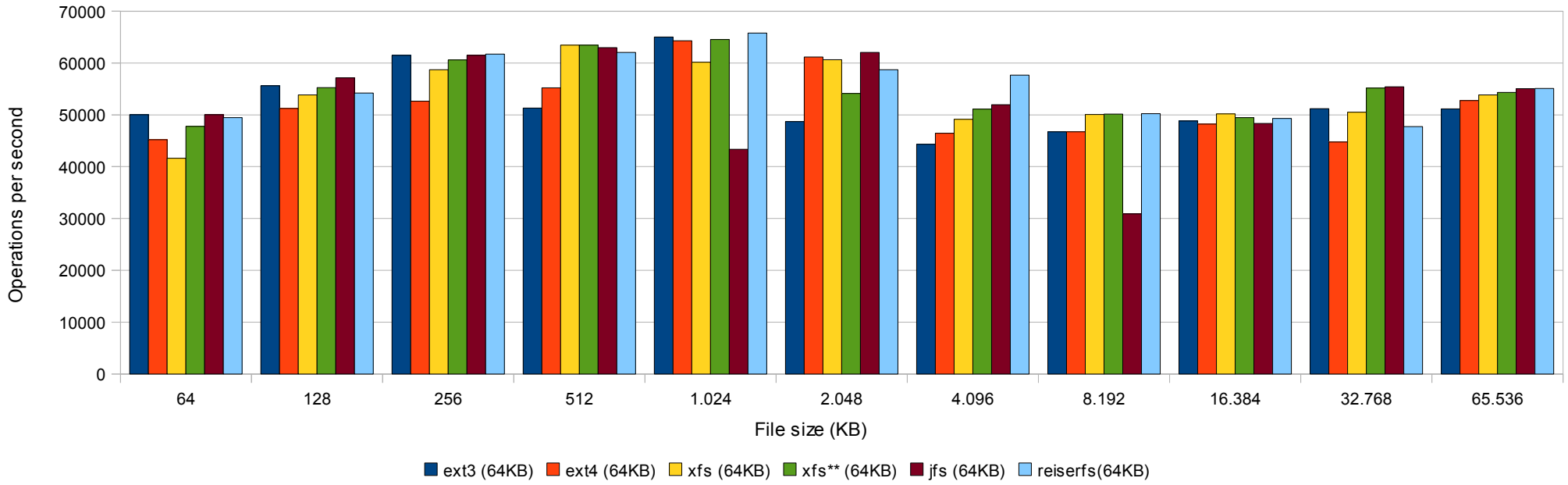
Reader report (8KB record size)



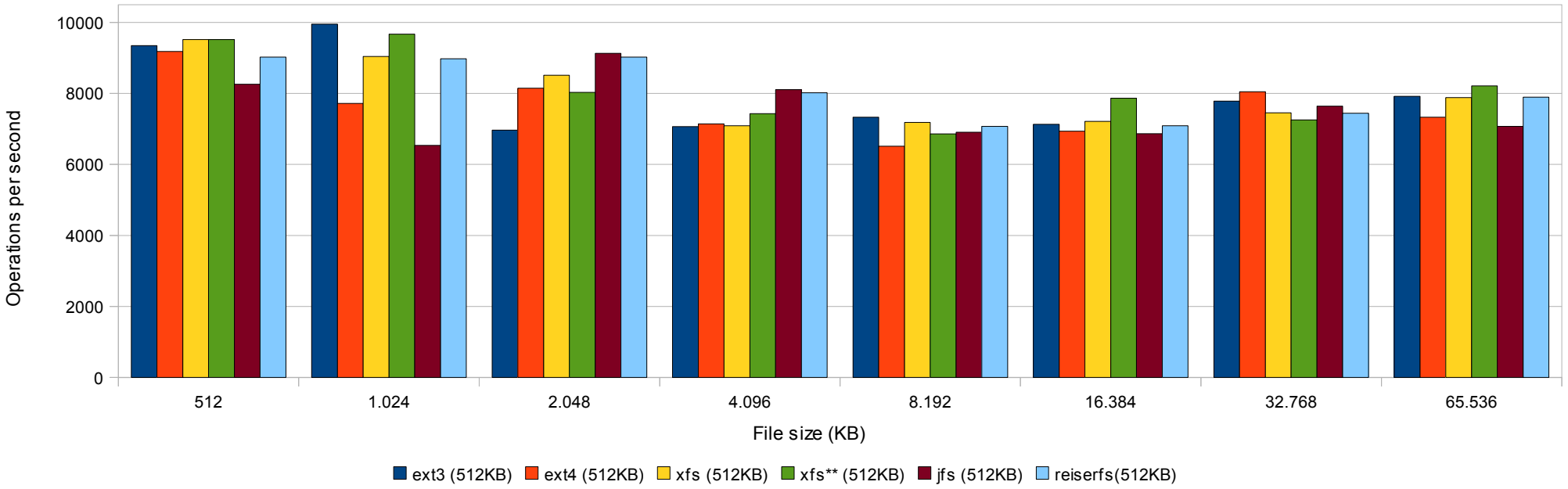
Reader report (16KB record size)



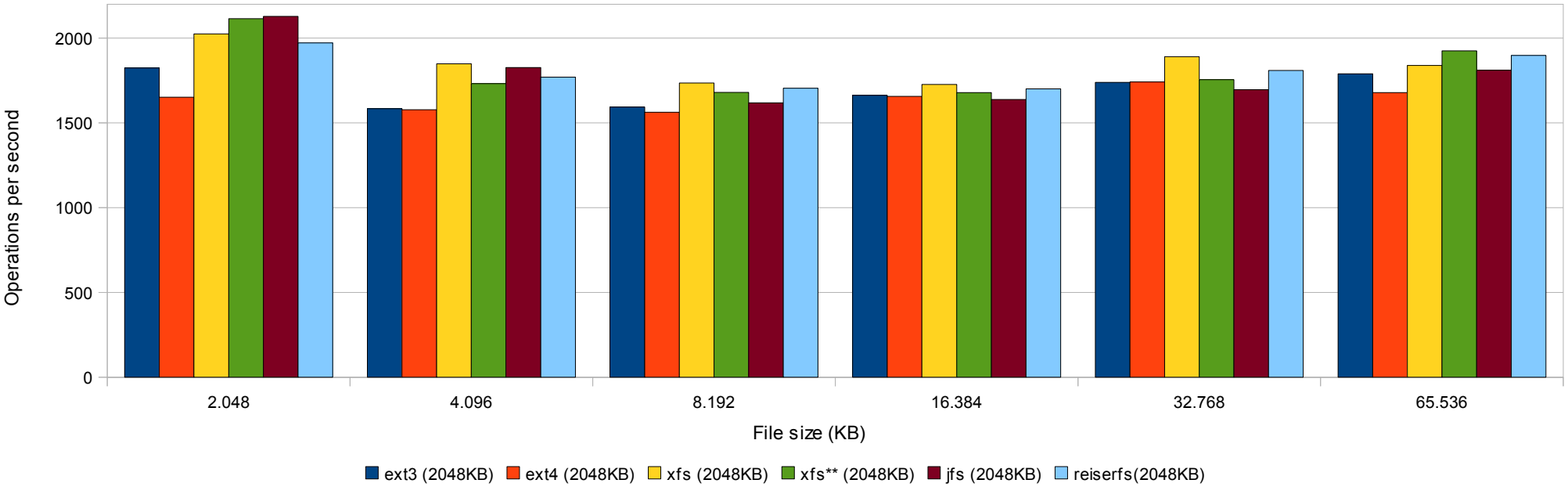
Reader report (64KB record size)

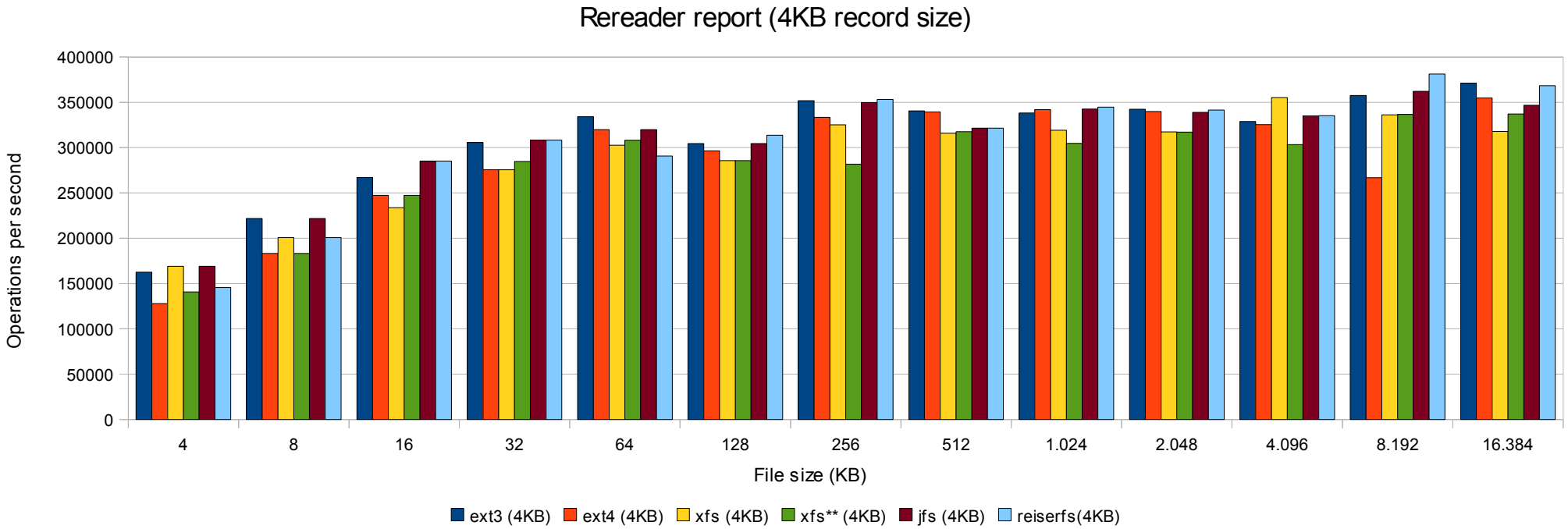


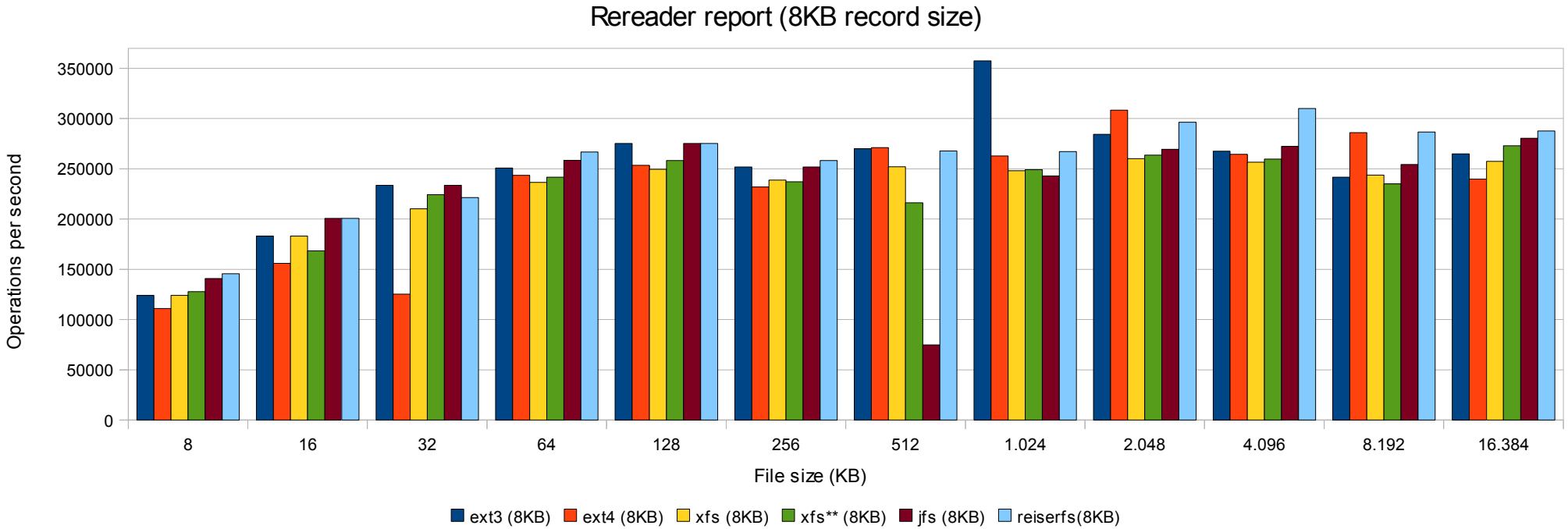
Reader report (512KB record size)



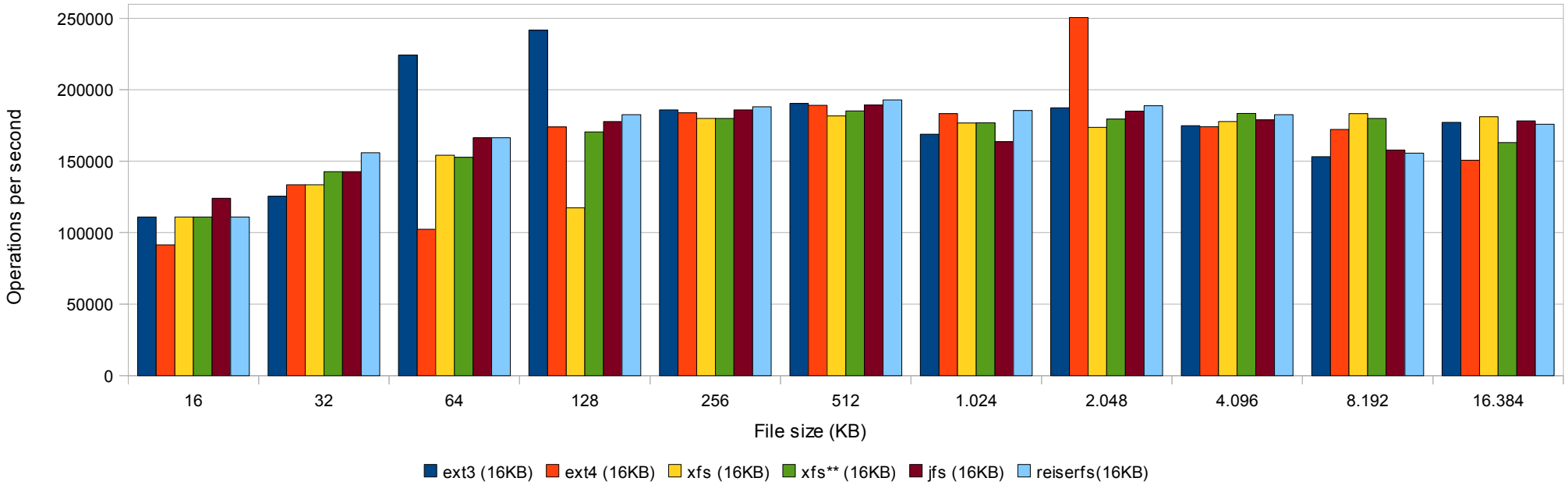
Reader report (2048KB record size)



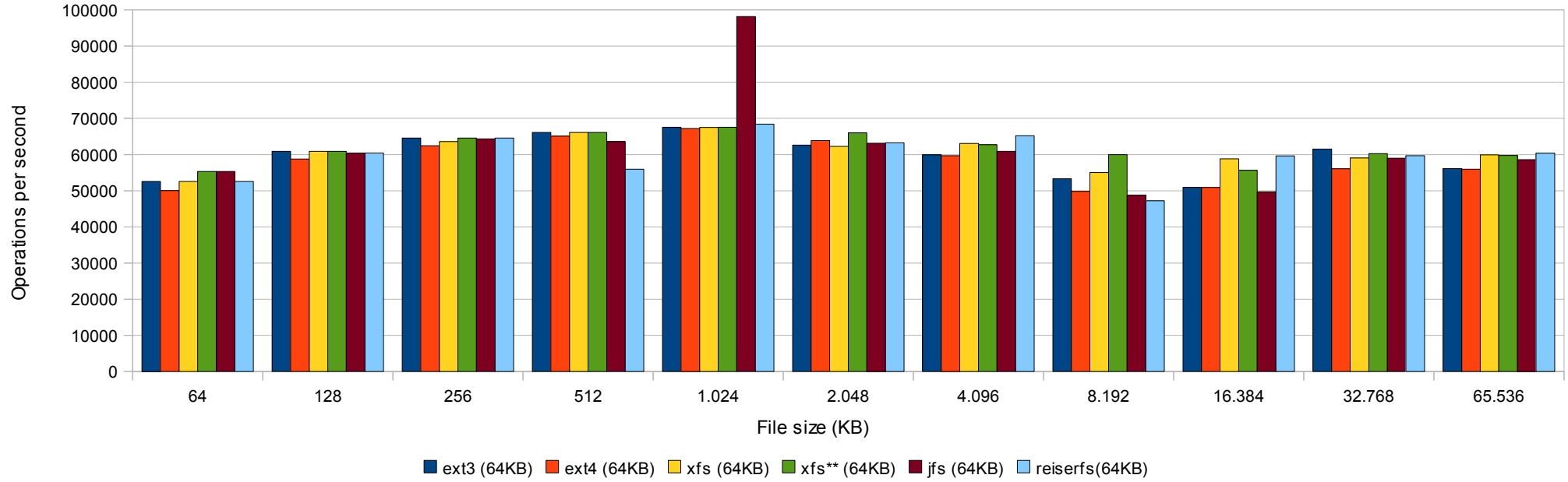

Rereader report (4KB record size)
Operations per second
400000
350000
300000
250000
200000
150000
100000
50000
0
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)


Rereader report (8KB record size)
Operations per second
350000
300000
250000
200000
150000
100000
50000
0
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

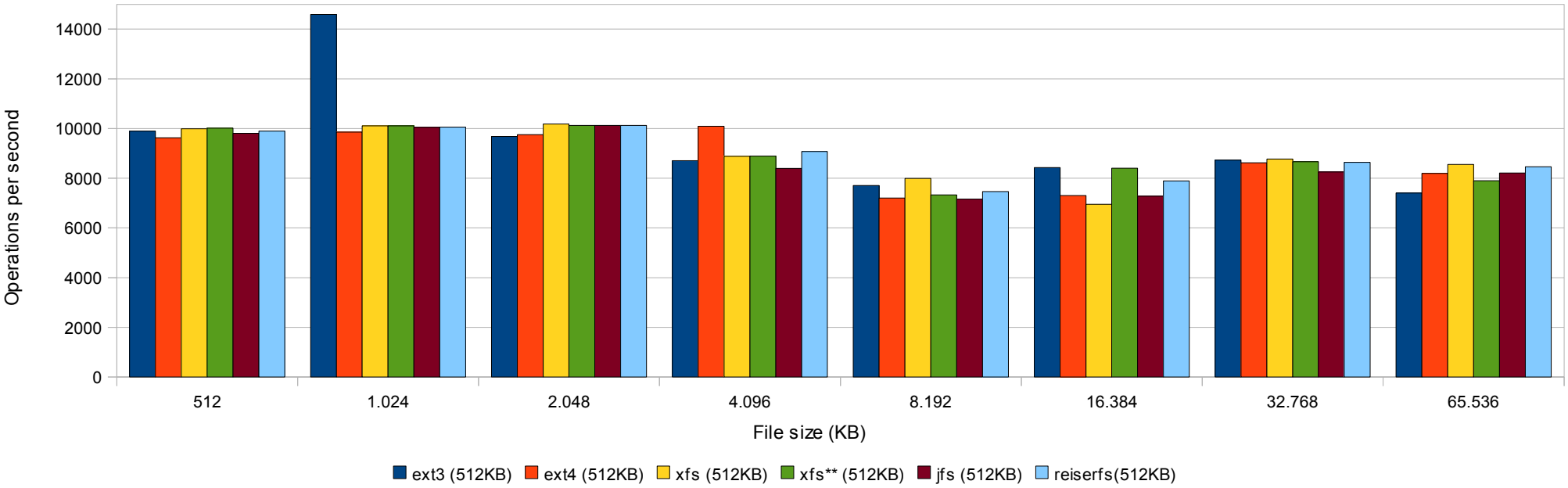
Rereader report (16KB record size)



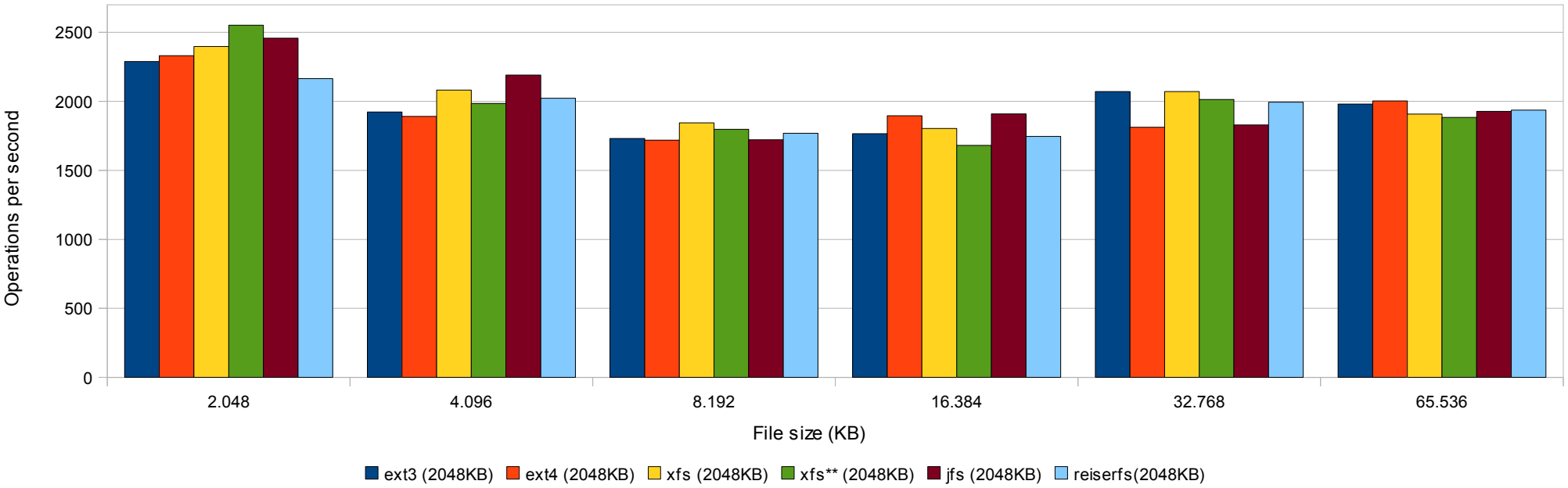
Rereader report (64KB record size)

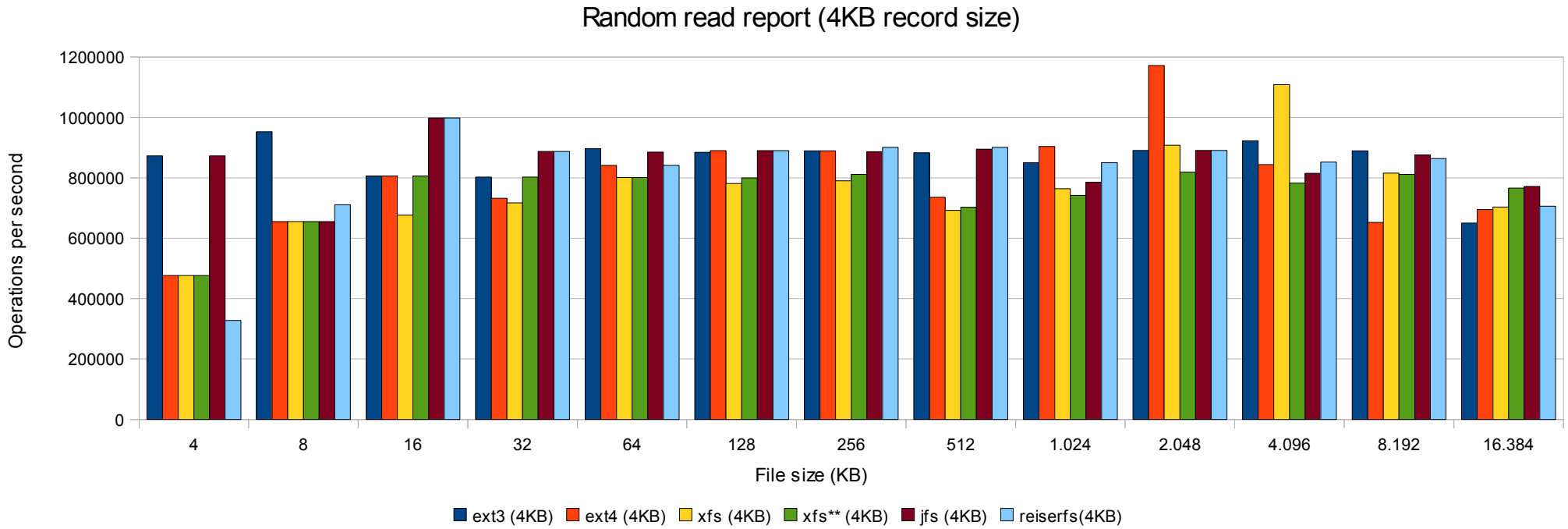


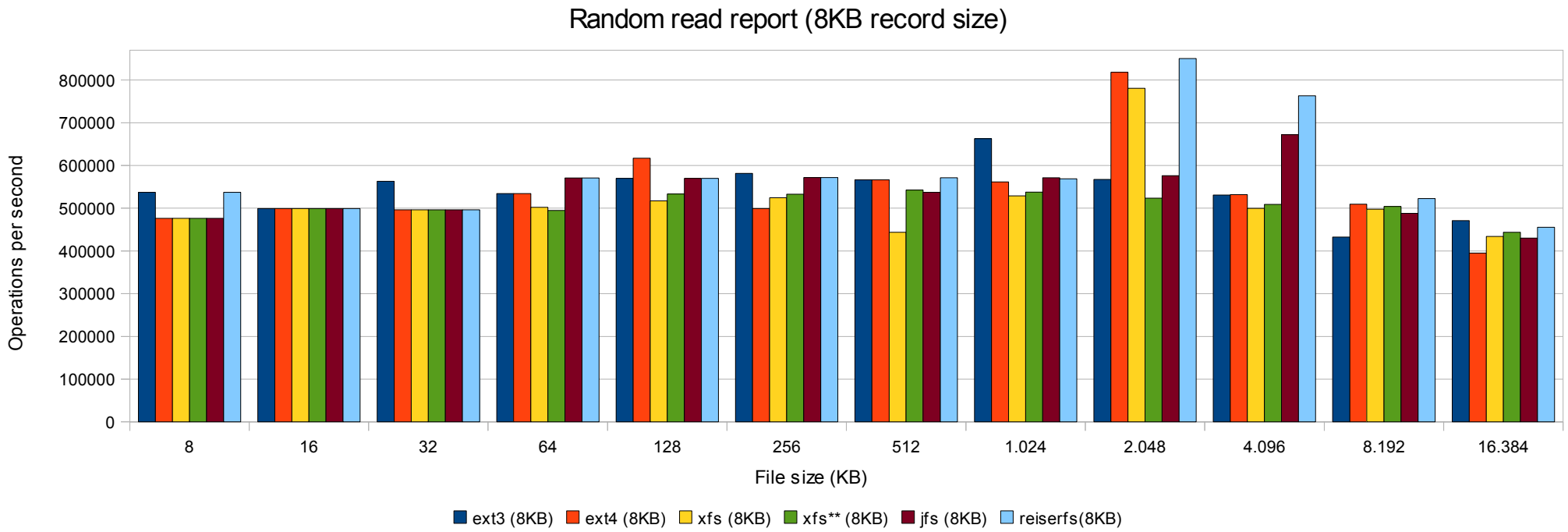
Rereader report (512KB record size)

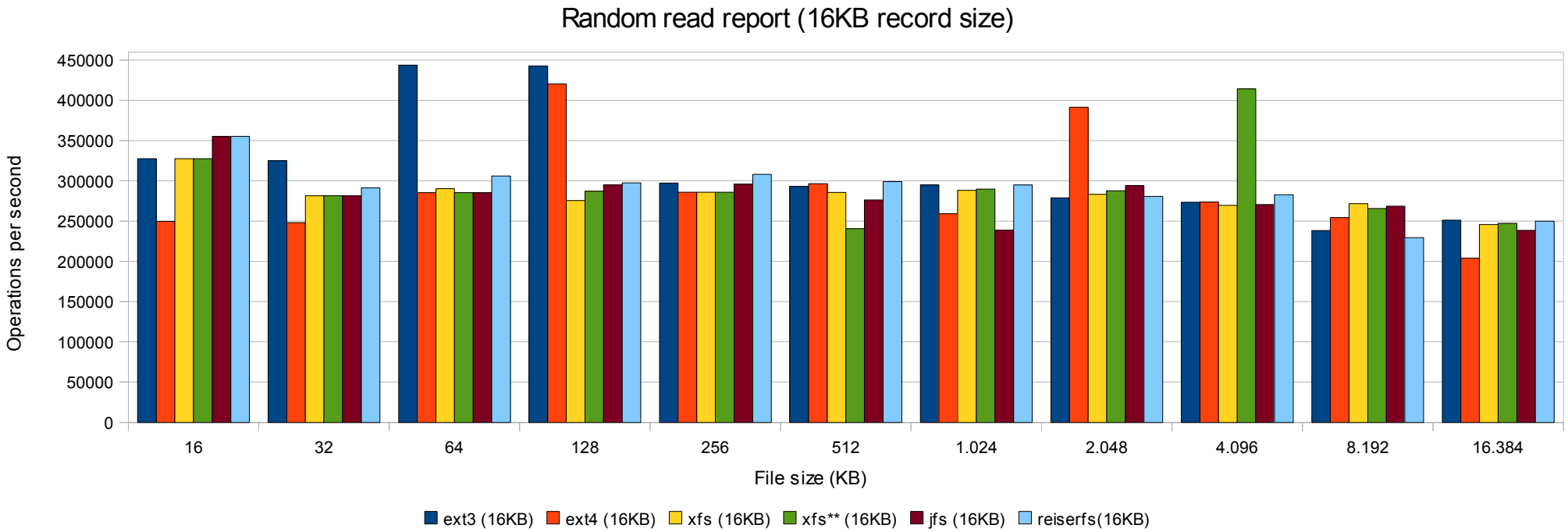


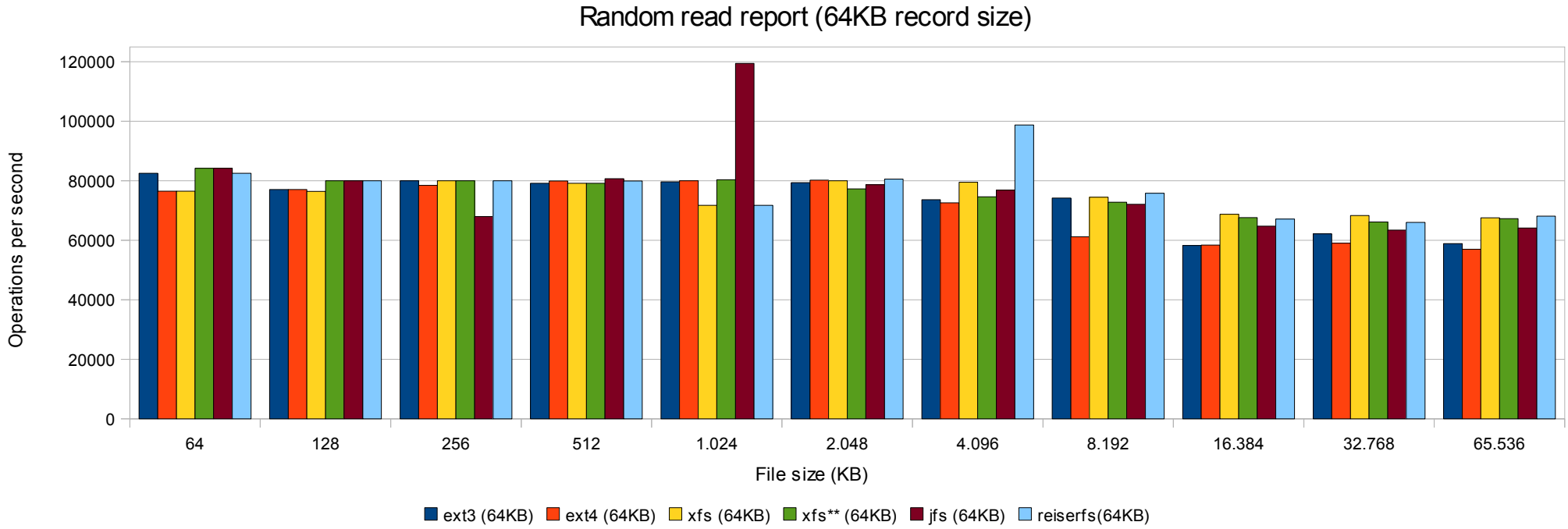
Rereader report (2048KB record size)



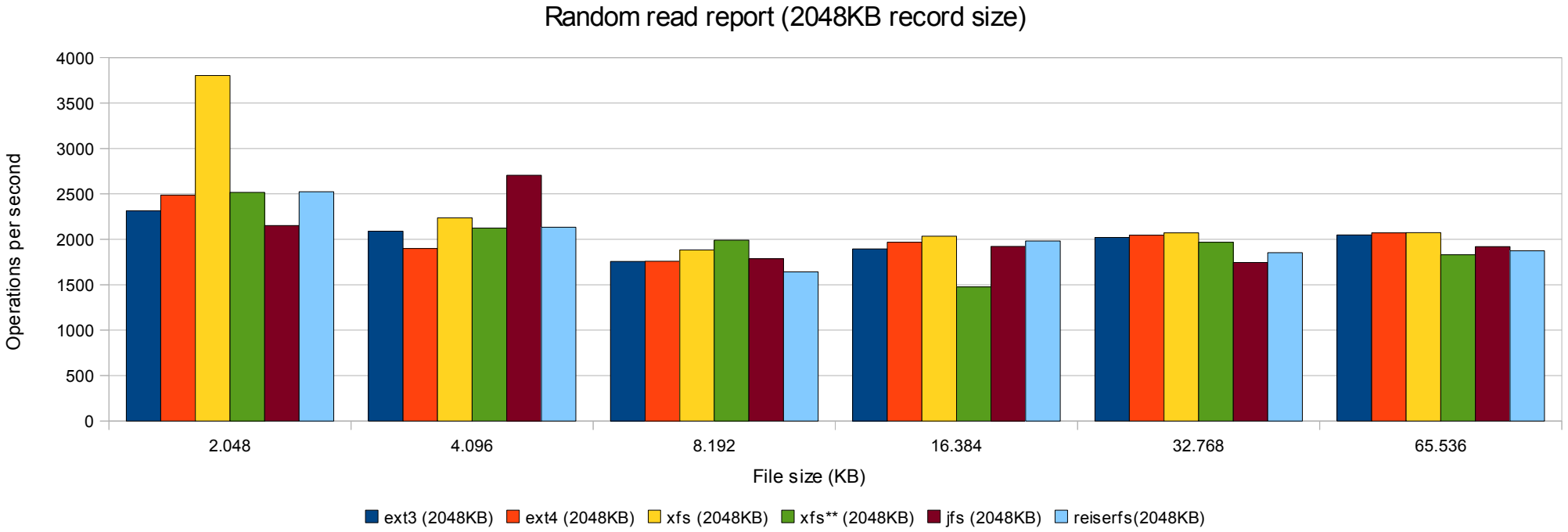

Random read report (4KB record size)
Operations per second
1200000
1000000
800000
600000
400000
200000
0
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)

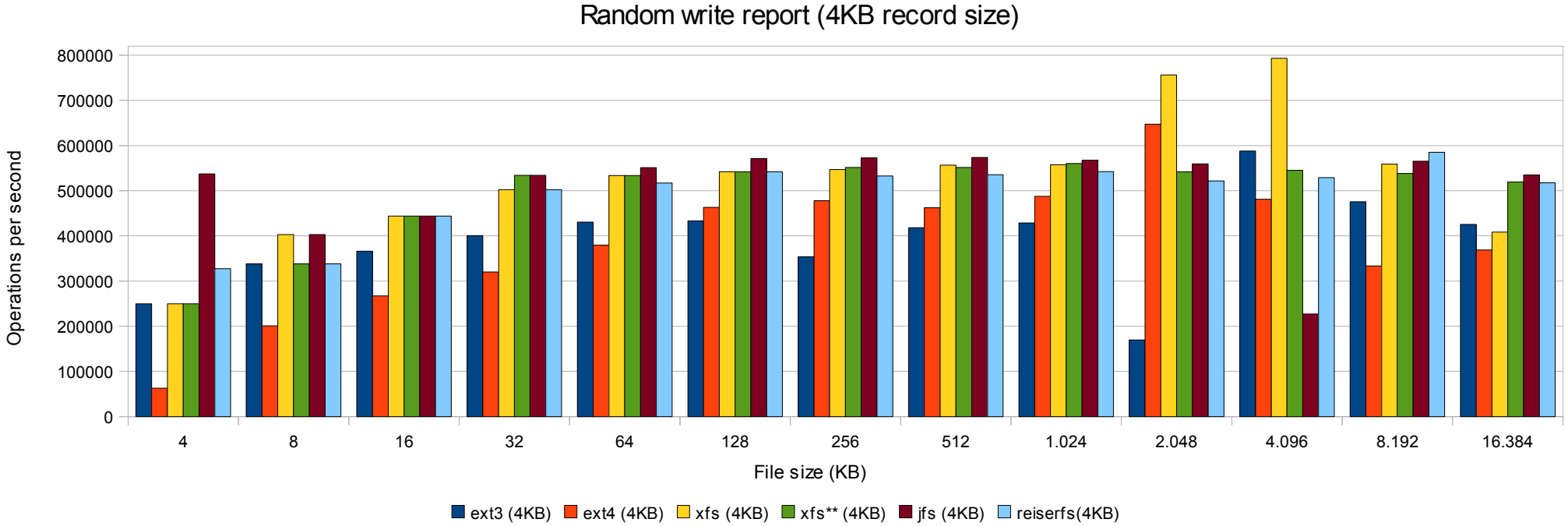

Random read report (8KB record size)
Operations per second
800000
700000
600000
500000
400000
300000
200000
100000
0
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

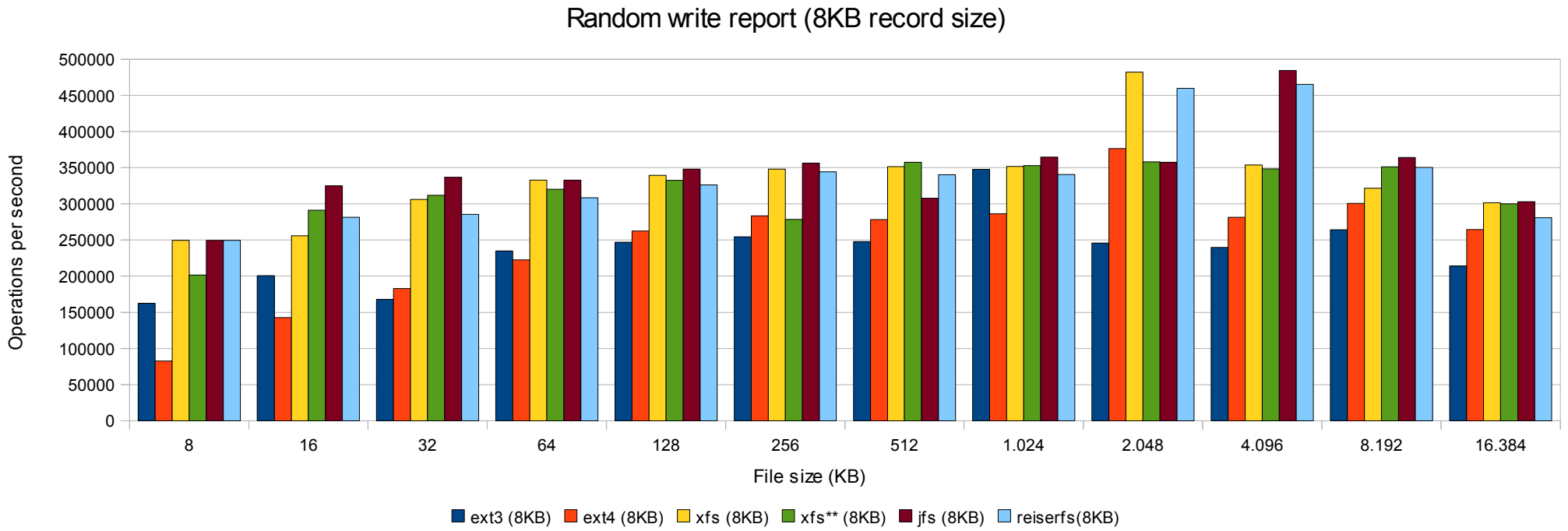

Random read report (16KB record size)
Operations per second
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)

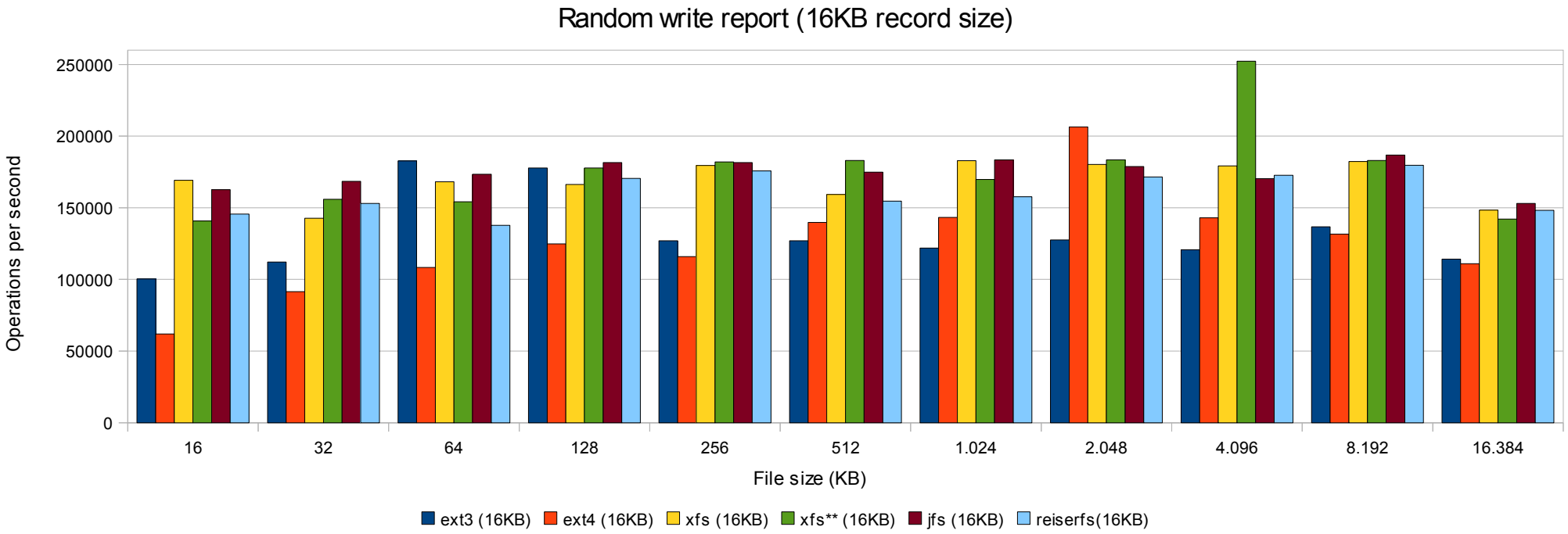

Random read report (64KB record size)
Operations per second
120000
100000
80000
60000
40000
20000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

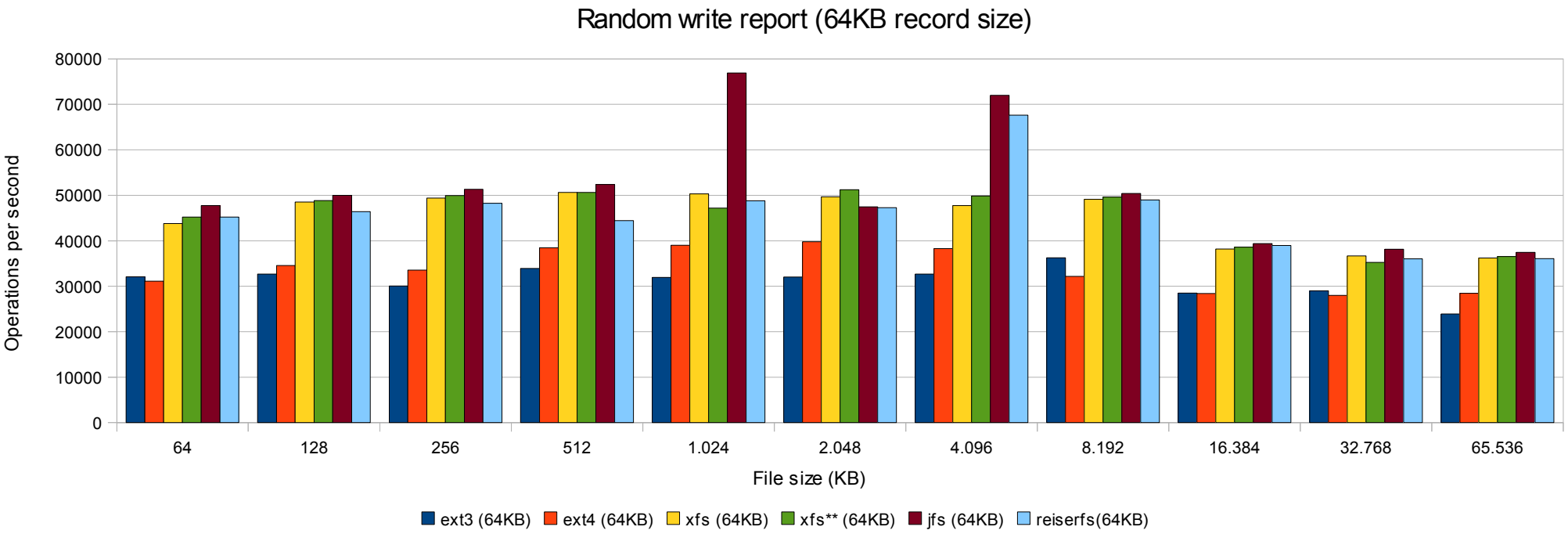

Random read report (512KB record size)
Operations per second
16000
14000
12000
10000
8000
6000
4000
2000
0
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (512KB)
ext4 (512KB)
xfs (512KB)
xfs** (512KB)
jfs (512KB)
reiserfs(512KB)

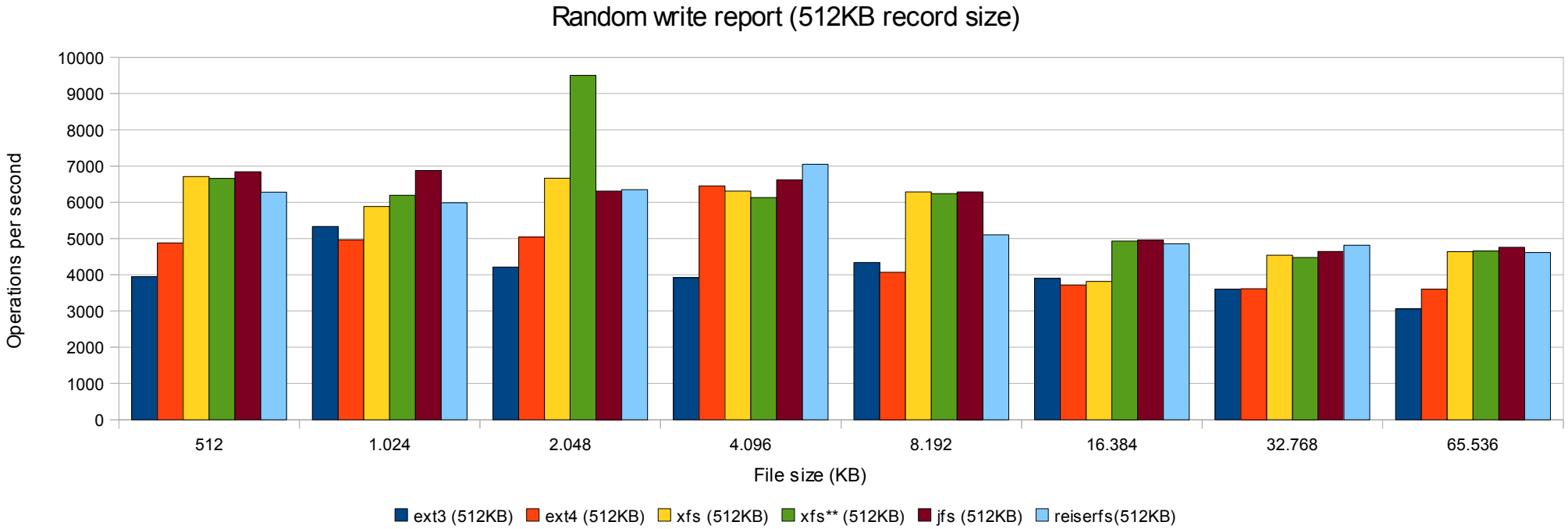

Random read report (2048KB record size)
Operations per second
4000
3500
3000
2500
2000
1500
1000
500
0
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (2048KB)
ext4 (2048KB)
xfs (2048KB)
xfs** (2048KB)
jfs (2048KB)
reiserfs(2048KB)

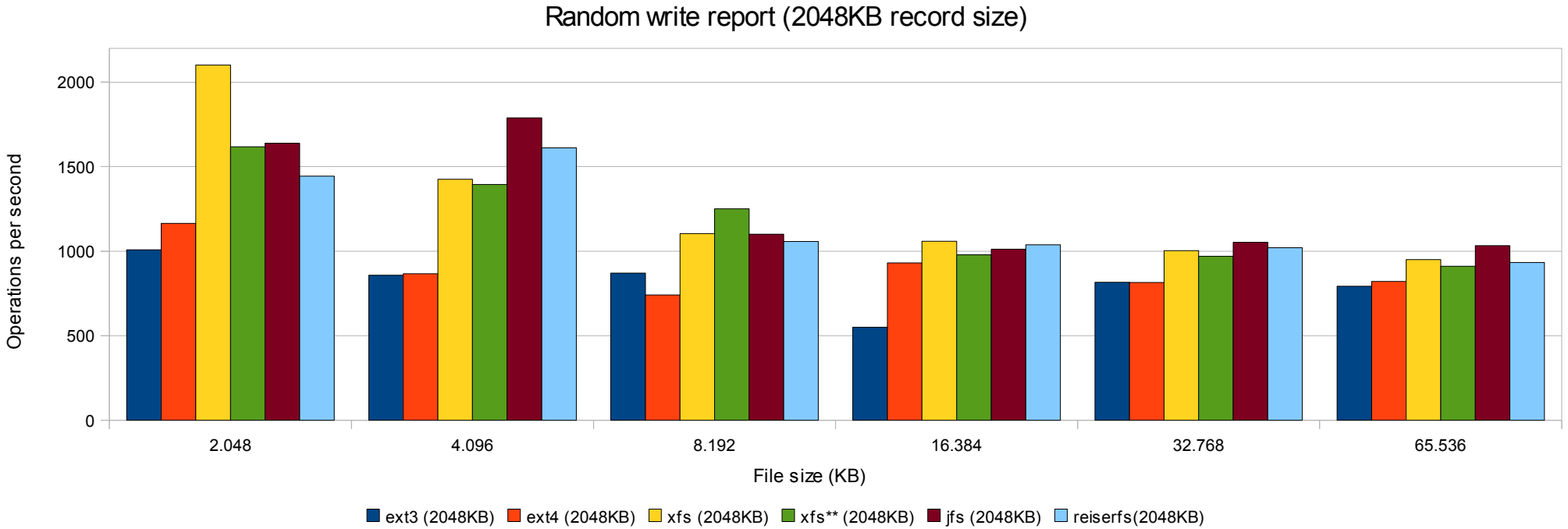

Random write report (4KB record size)
Operations per second
800000
700000
600000
500000
400000
300000
200000
100000
0
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)


Random write report (8KB record size)
Operations per second
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

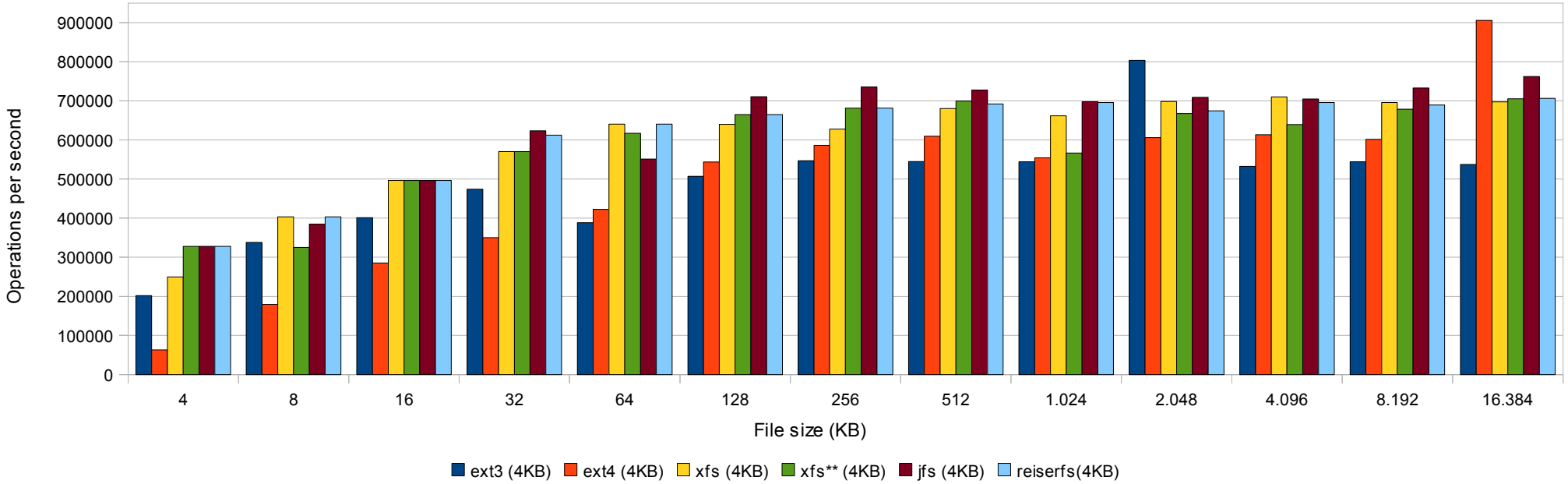

Random write report (16KB record size)
Operations per second
250000
200000
150000
100000
50000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)


Random write report (64KB record size)
Operations per second
80000
70000
60000
50000
40000
30000
20000
10000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

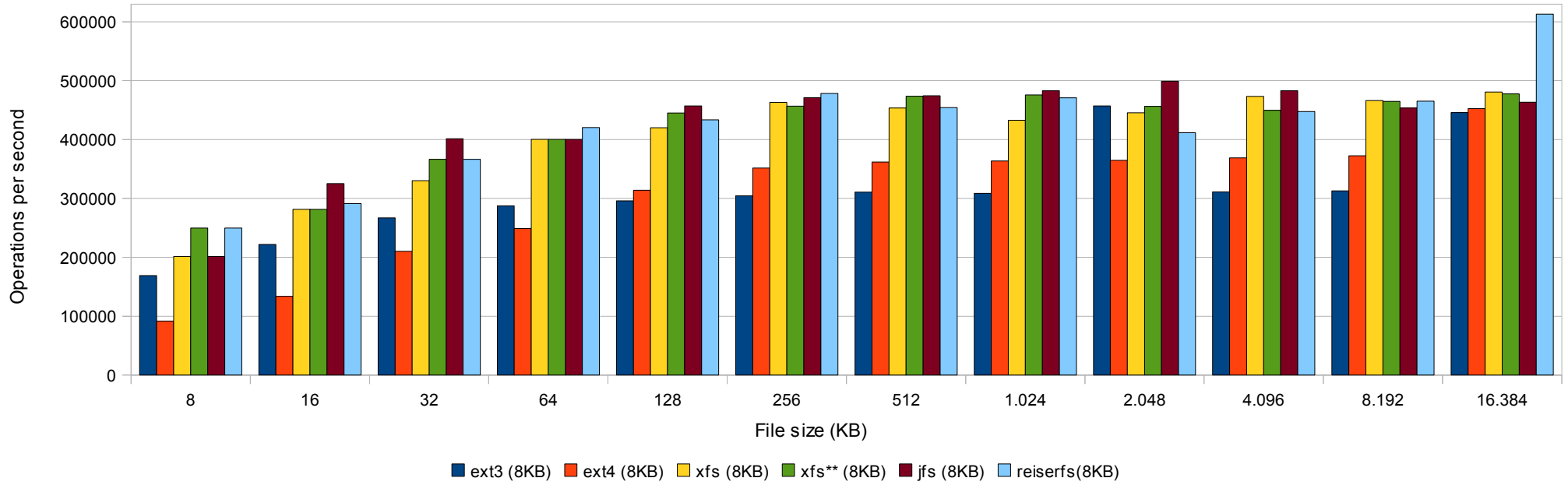

Random write report (512KB record size)
Operations per second
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (512KB)
ext4 (512KB)
xfs (512KB)
xfs** (512KB)
jfs (512KB)
reiserfs(512KB)


Random write report (2048KB record size)
Operations per second
2000
1500
1000
500
0
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (2048KB)
ext4 (2048KB)
xfs (2048KB)
xfs** (2048KB)
jfs (2048KB)
reiserfs(2048KB)

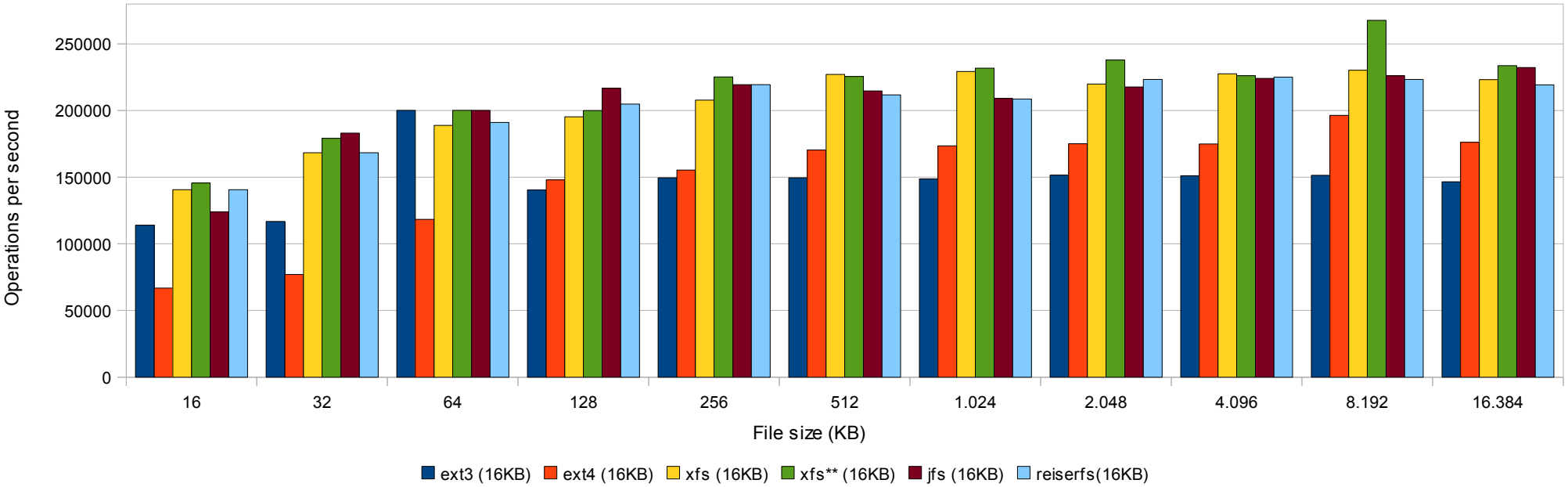
Record rewrite report (4KB record size)



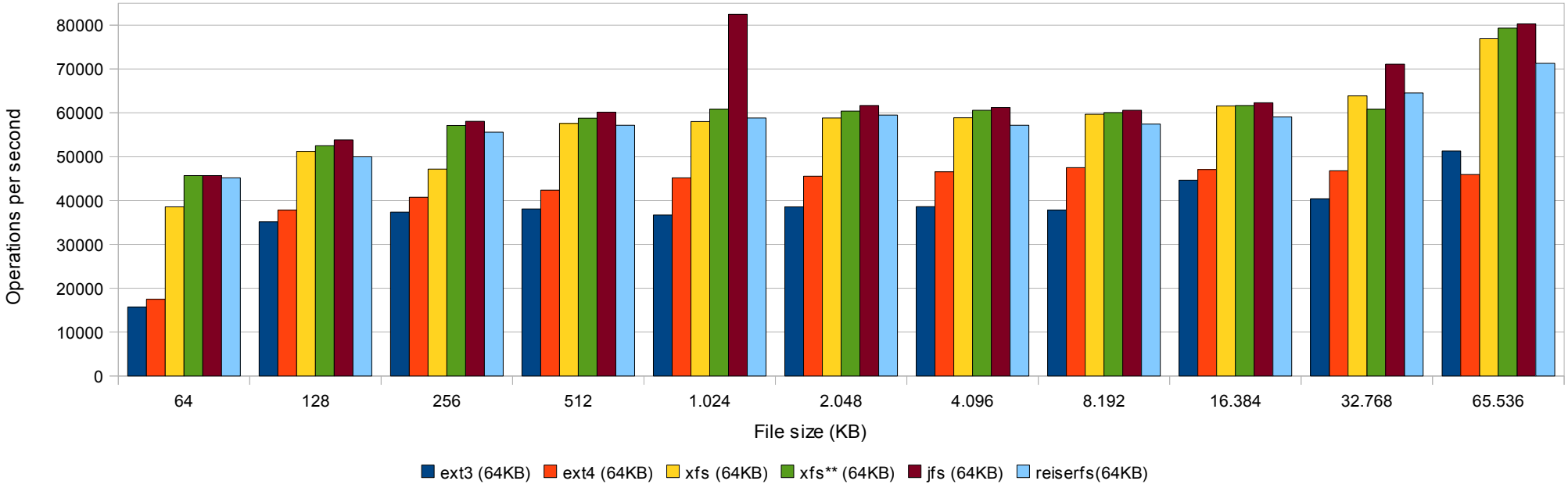
Record rewrite report (8KB record size)



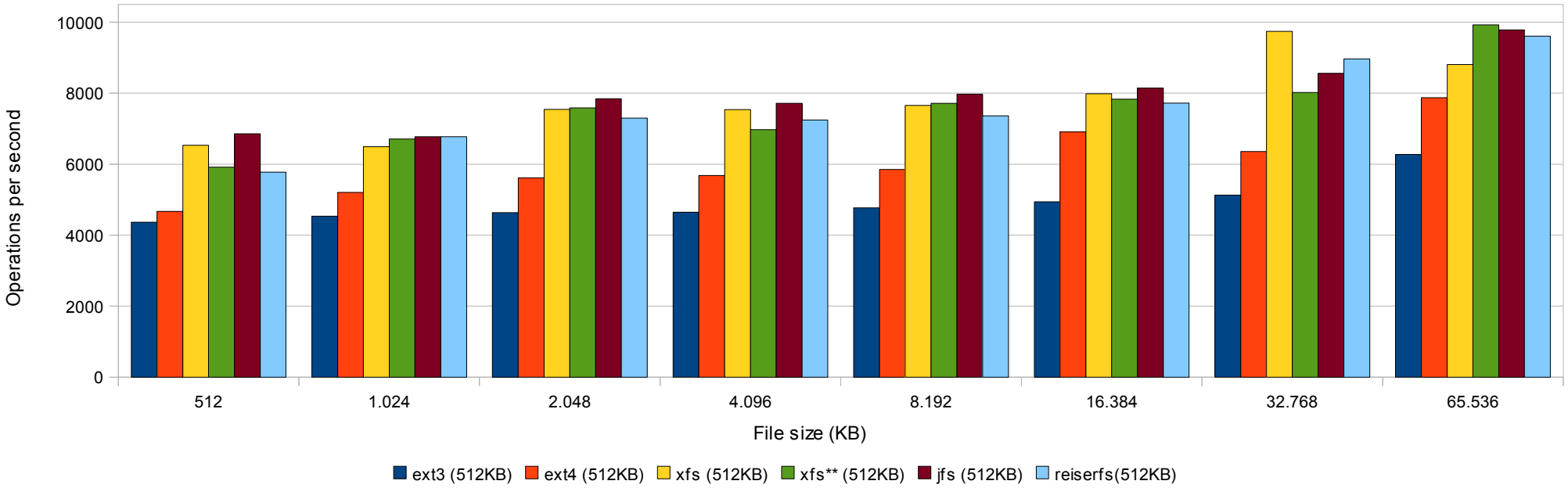
Record rewrite report (16KB record size)



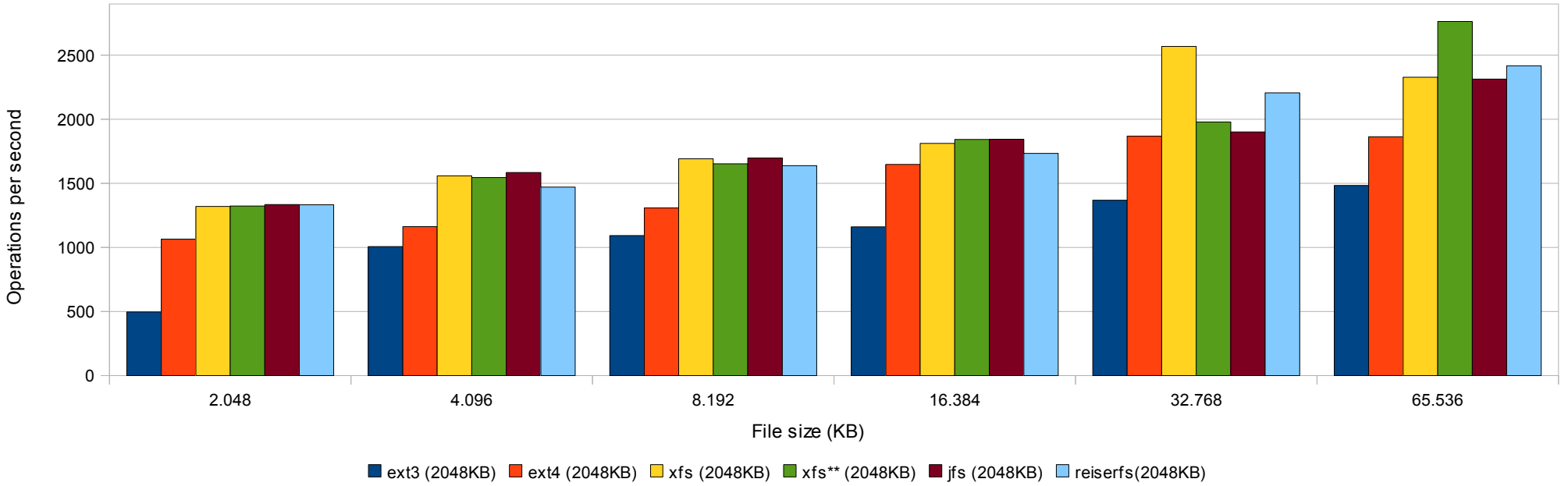
Record rewrite report (64KB record size)

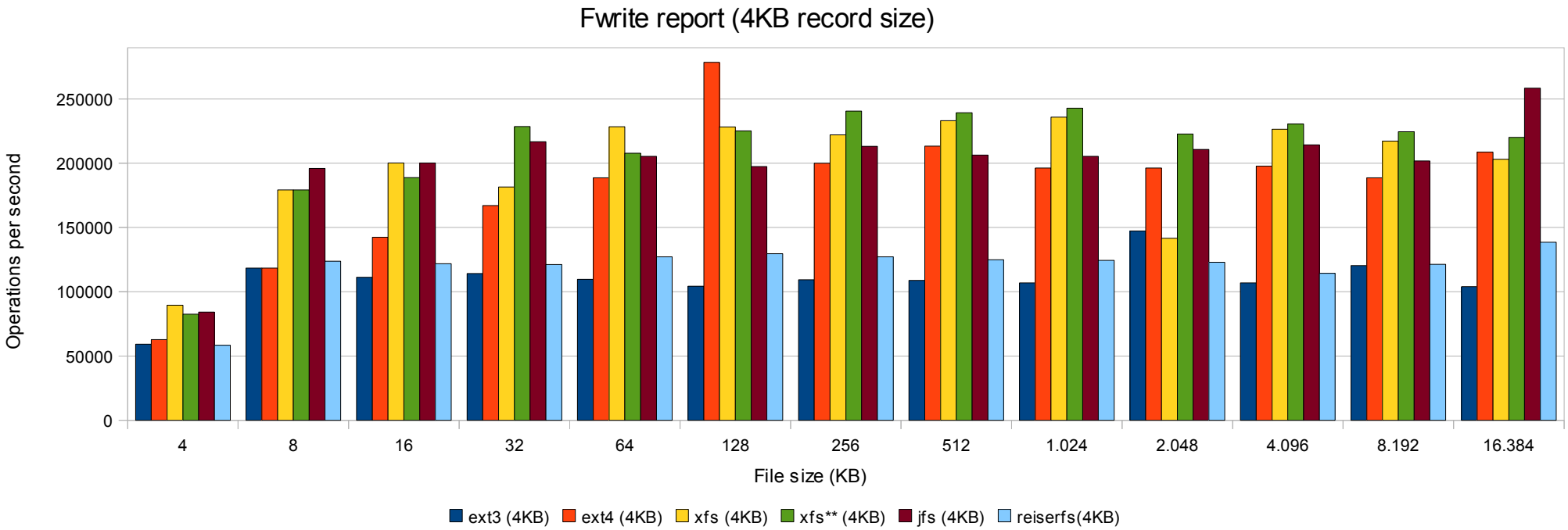


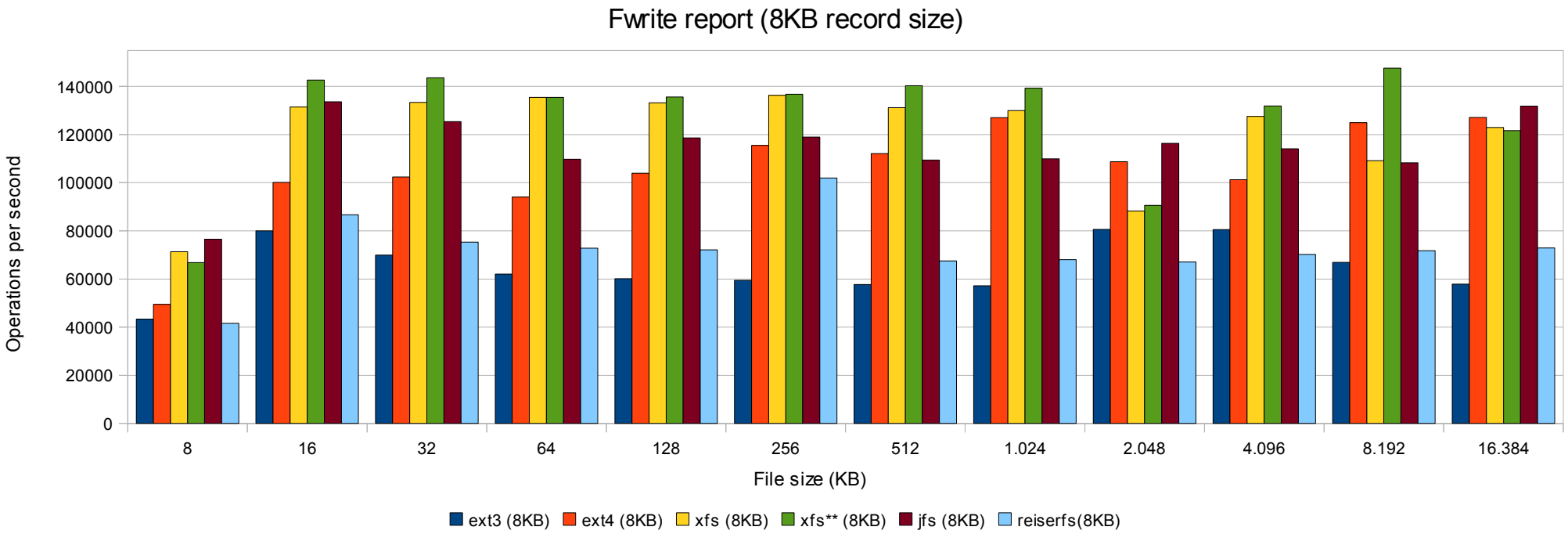
Record rewrite report (512KB record size)

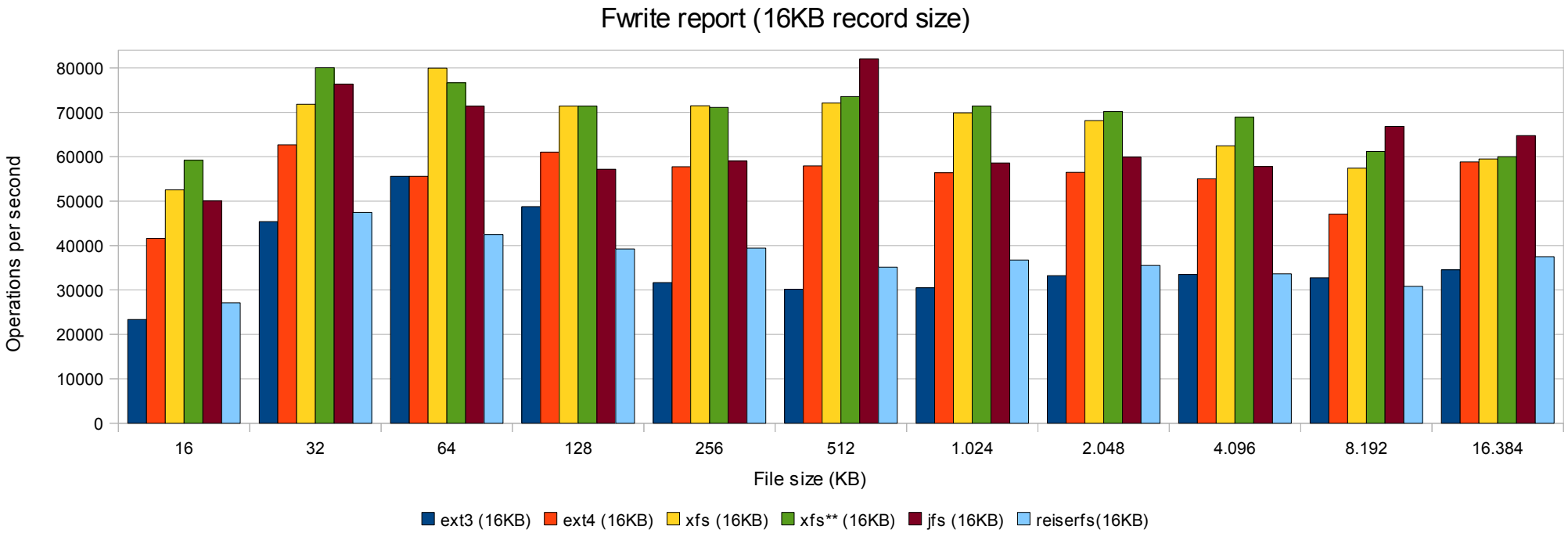


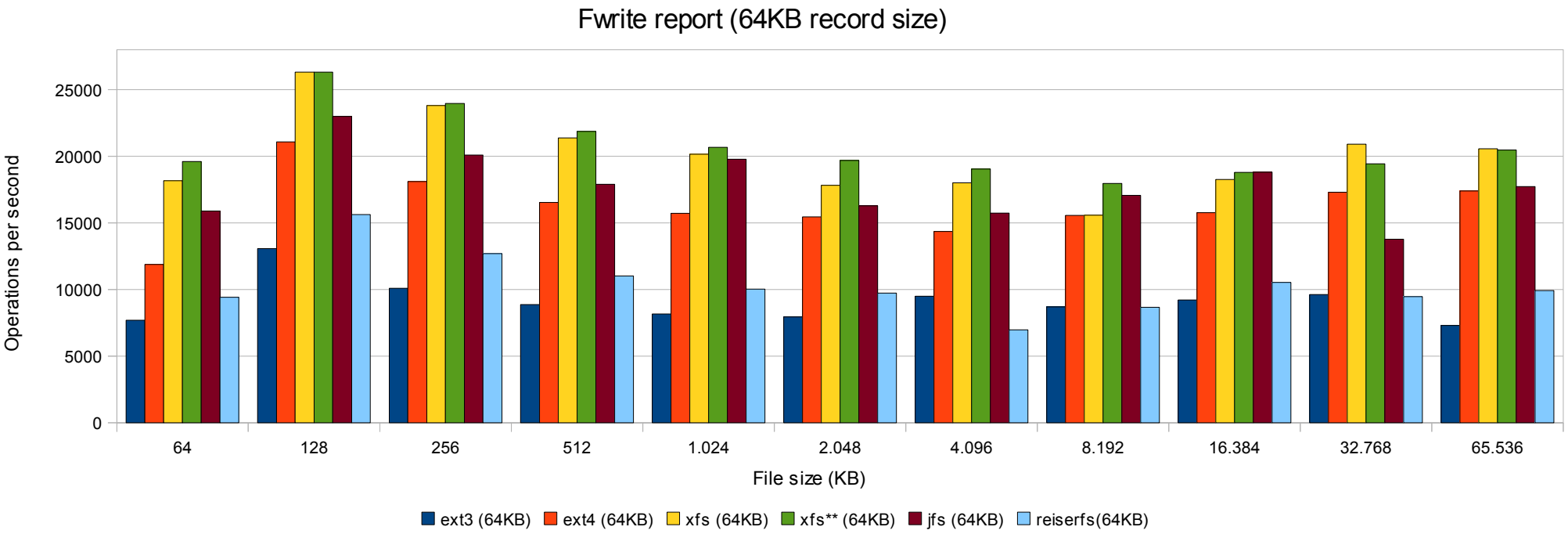
Record rewrite report (2048KB record size)



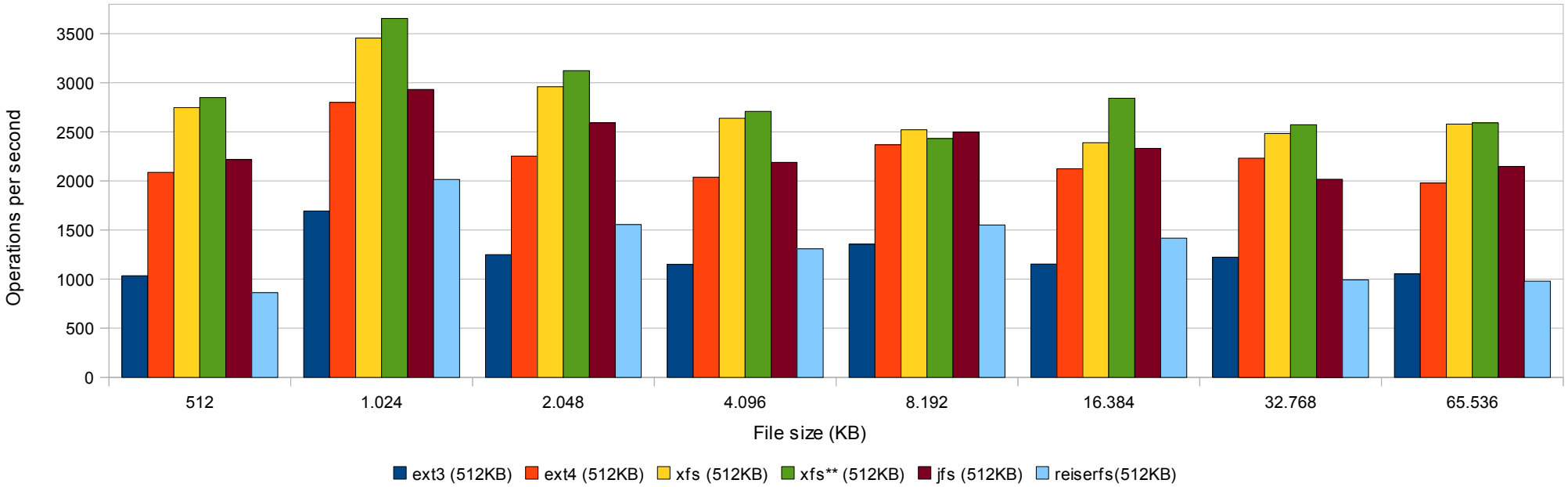

Fwrite report (4KB record size)
Operations per second
0
50000
100000
150000
200000
250000
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)


Fwrite report (8KB record size)
Operations per second
0
20000
40000
60000
80000
100000
120000
140000
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

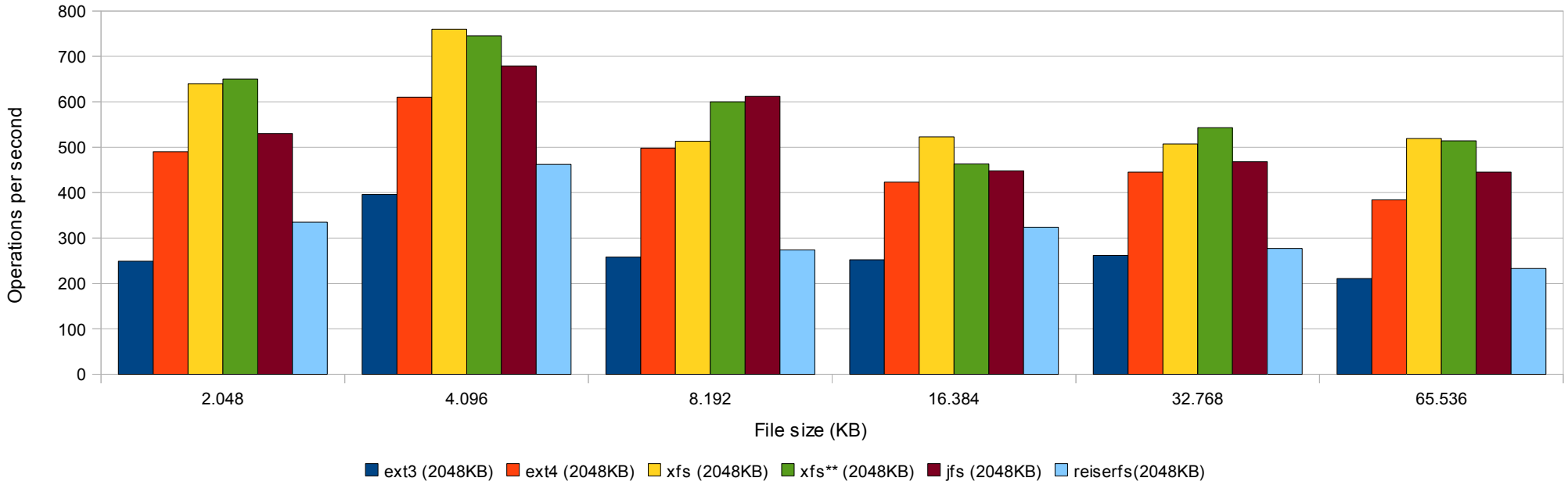

Fwrite report (16KB record size)
Operations per second
80000
70000
60000
50000
40000
30000
20000
10000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)

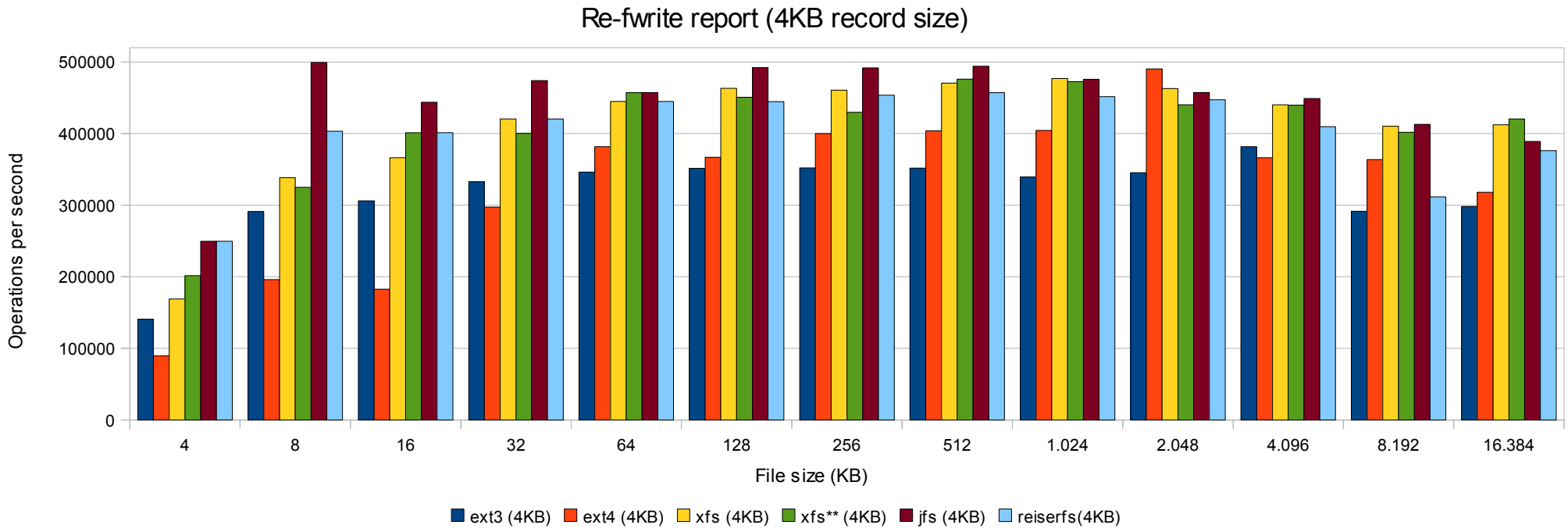

Fwrite report (64KB record size)
Operations per second
25000
20000
15000
10000
5000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

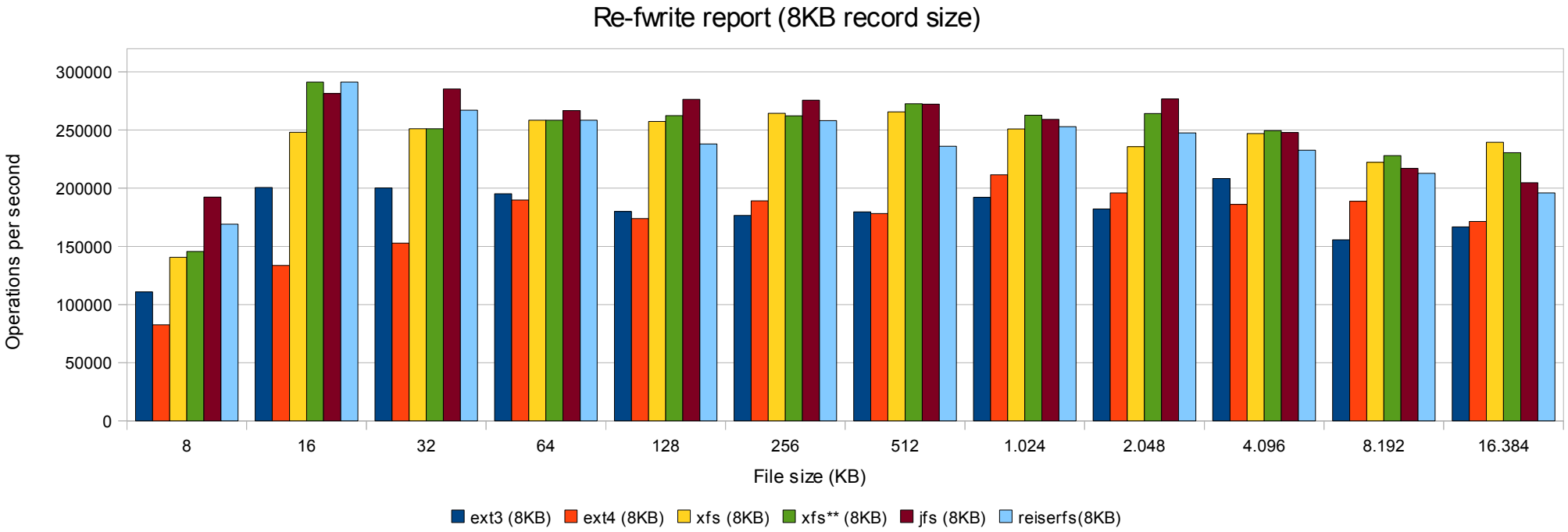
Fwrite report (512KB record size)

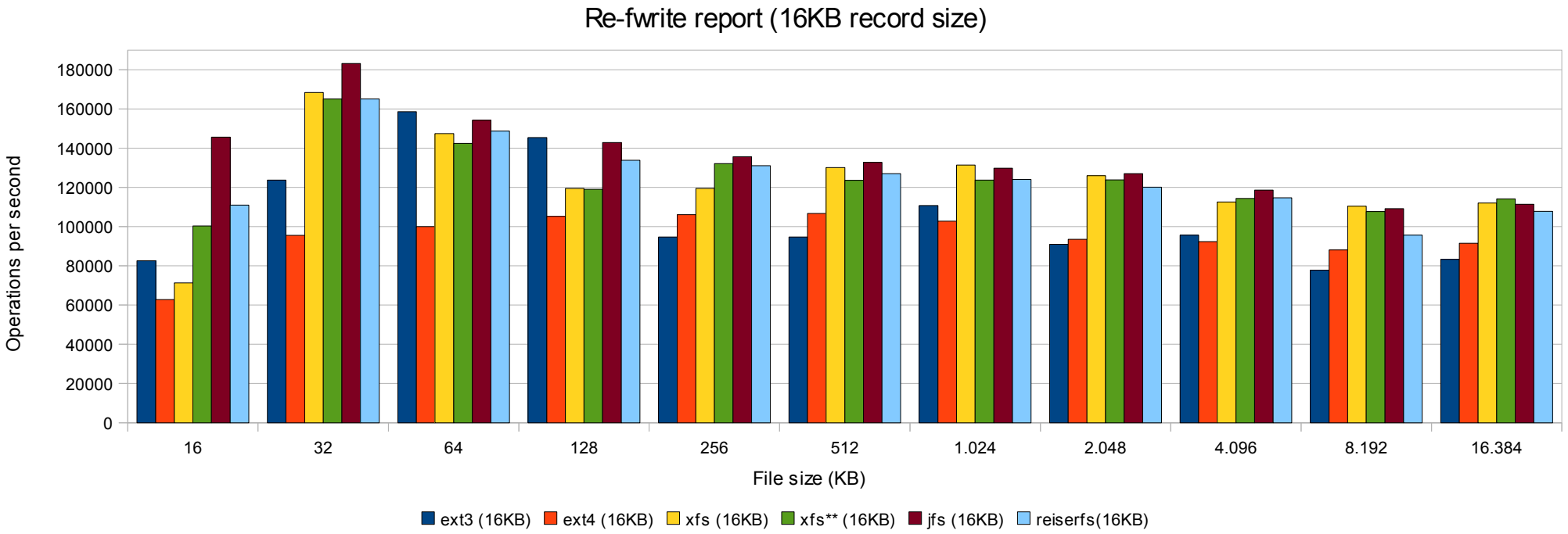


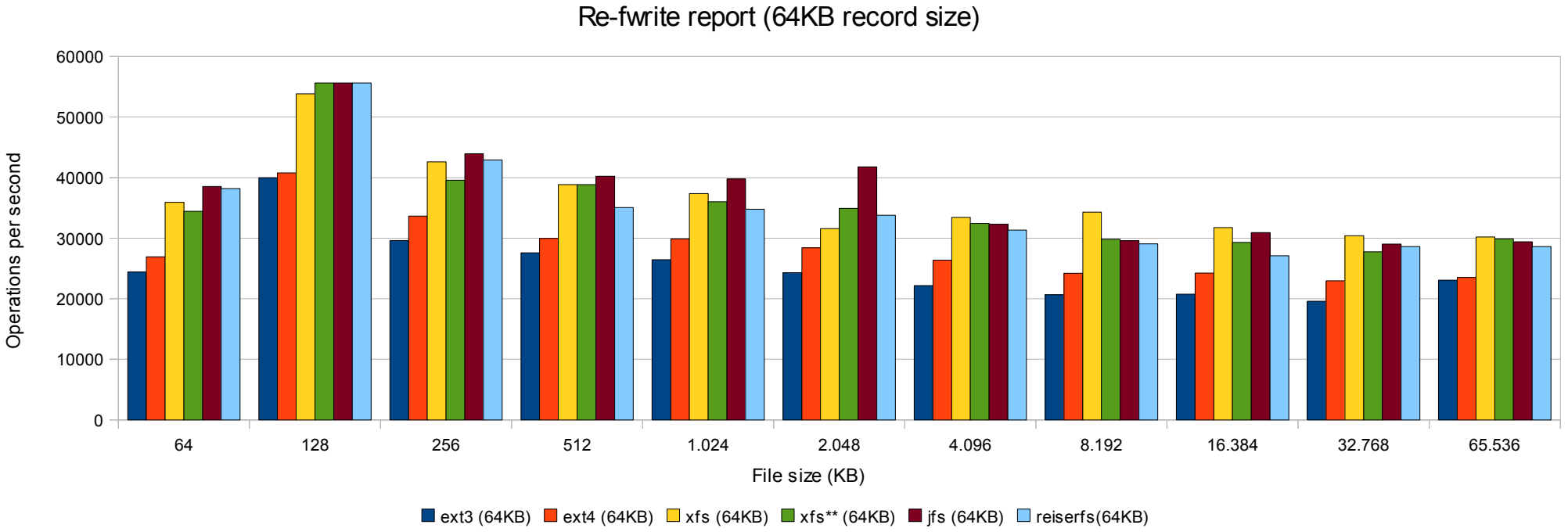
Fwrite report (2048KB record size)



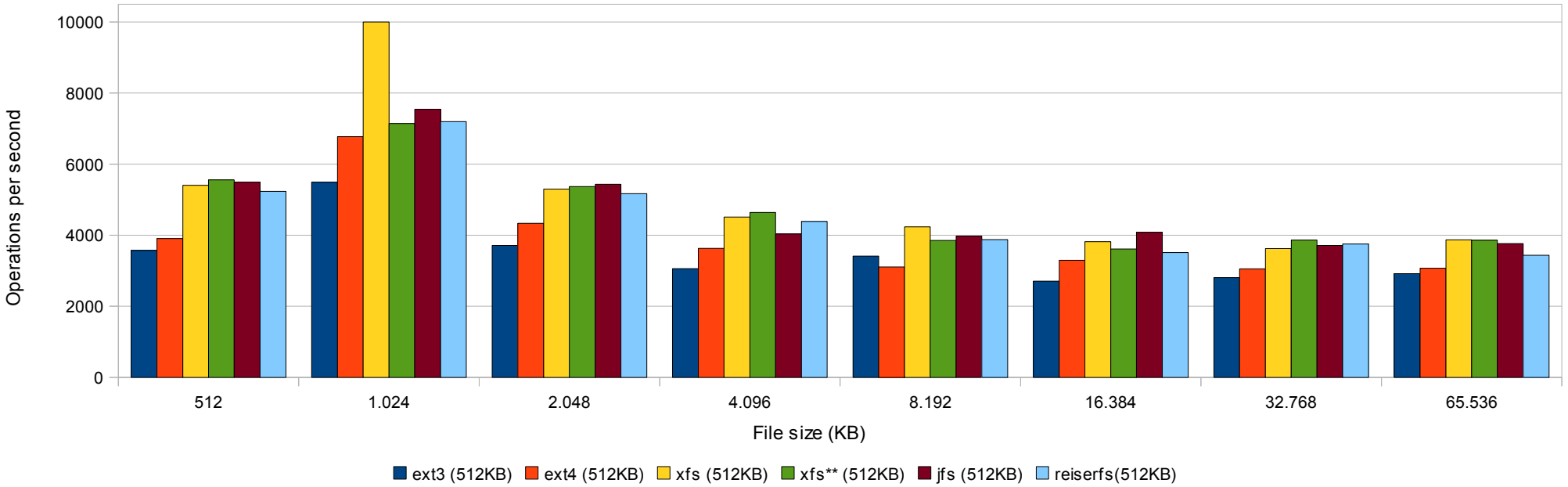

Re-fwrite report (4KB record size)
Operations per second
500000
400000
300000
200000
100000
0
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)


Re-fwrite report (8KB record size)
Operations per second
300000
250000
200000
150000
100000
50000
0
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

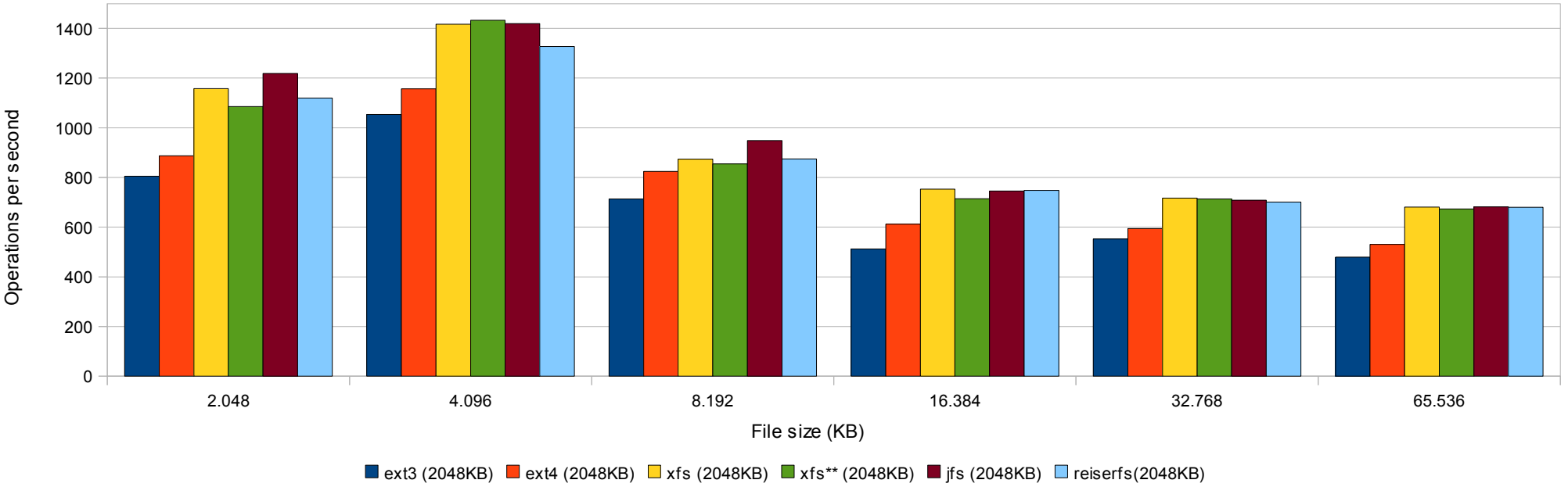

Re-fwrite report (16KB record size)
Operations per second
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)

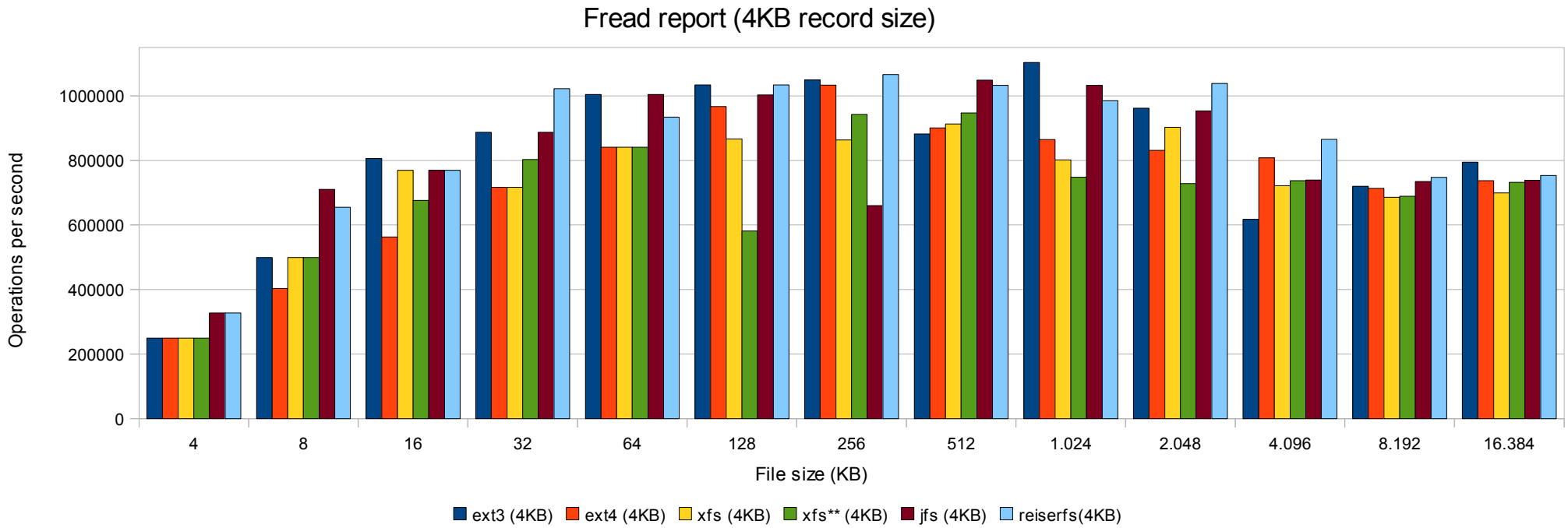

Re-fwrite report (64KB record size)
Operations per second
60000
50000
40000
30000
20000
10000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

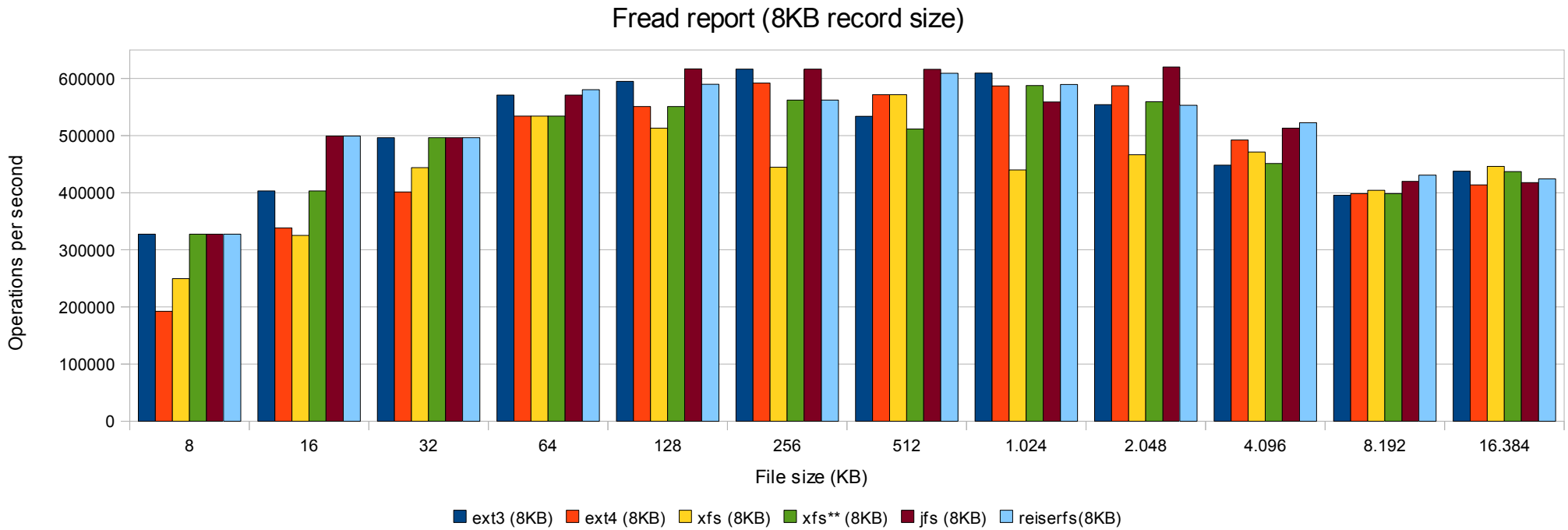
Re-fwrite report (512KB record size)

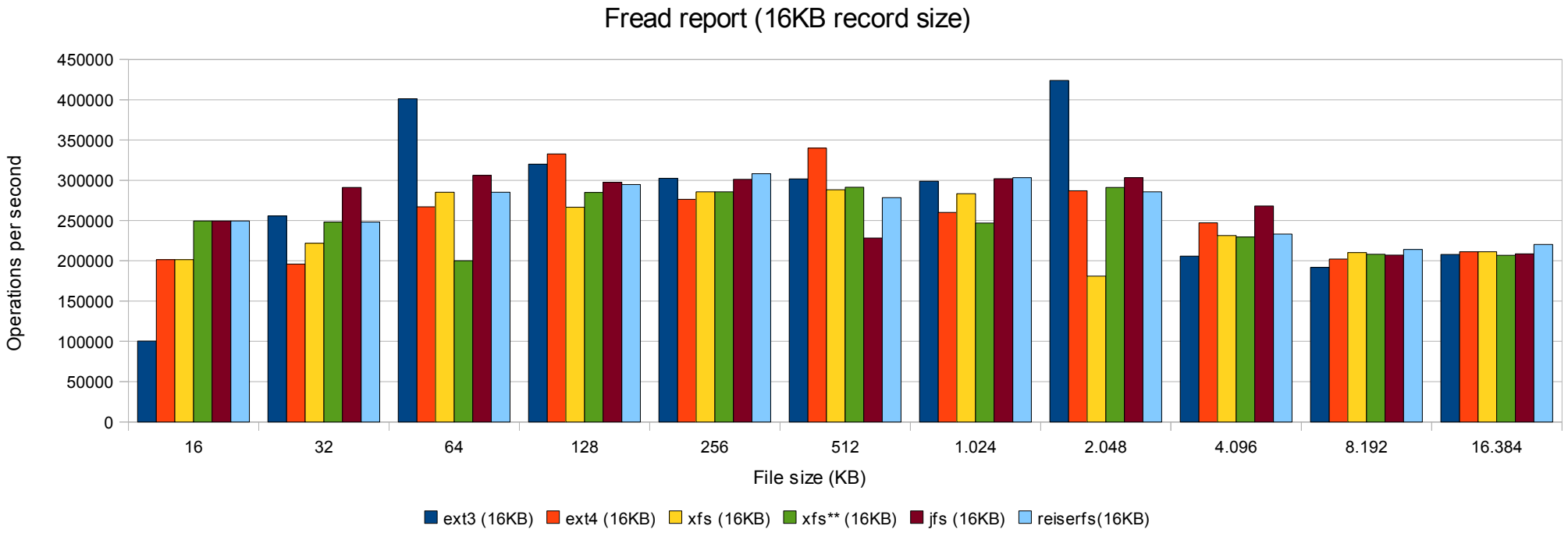


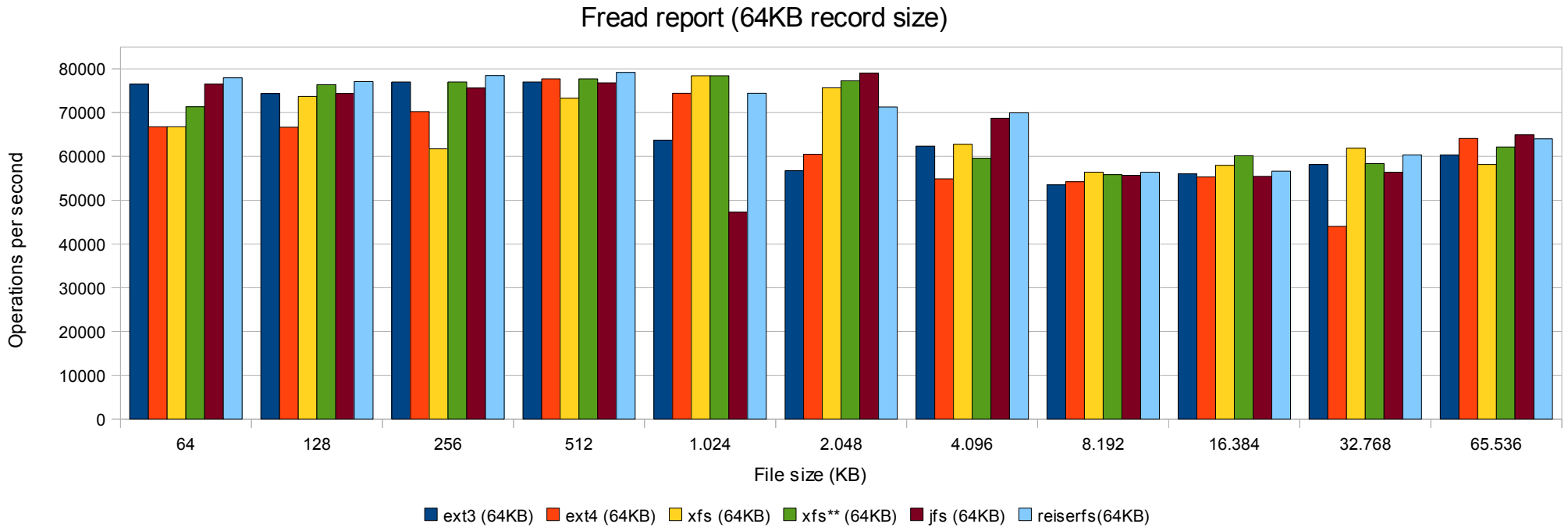
Re-fwrite report (2048KB record size)



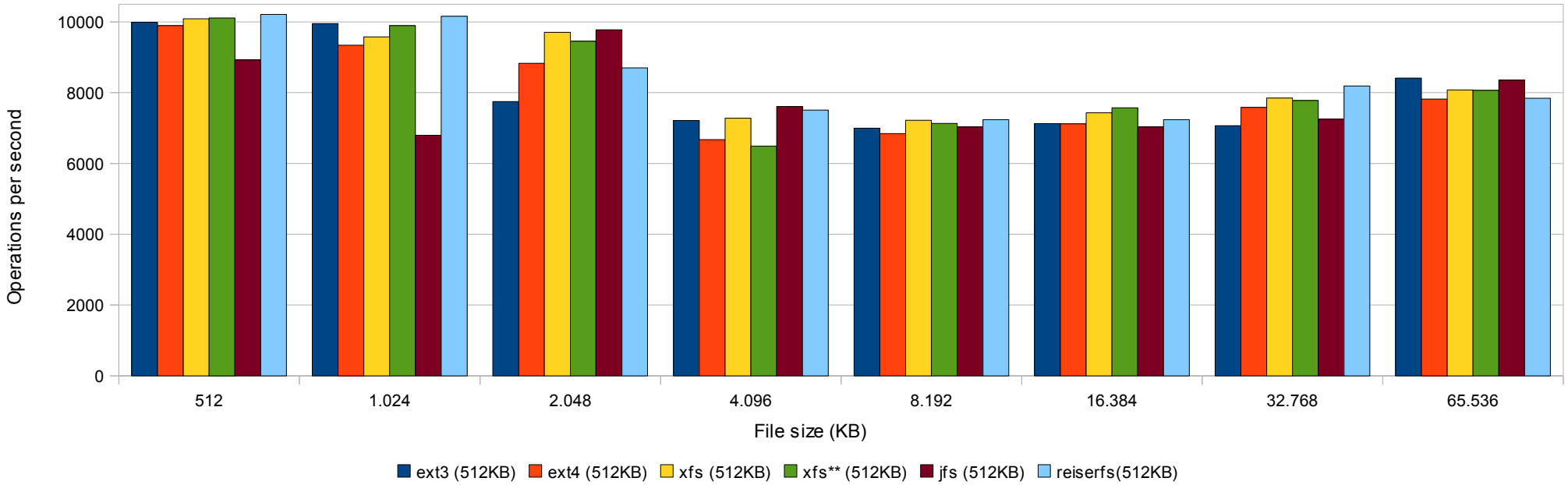

Fread report (4KB record size)
Operations per second
0
200000
400000
600000
800000
1000000
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)


Fread report (8KB record size)
Operations per second
0
100000
200000
300000
400000
500000
600000
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

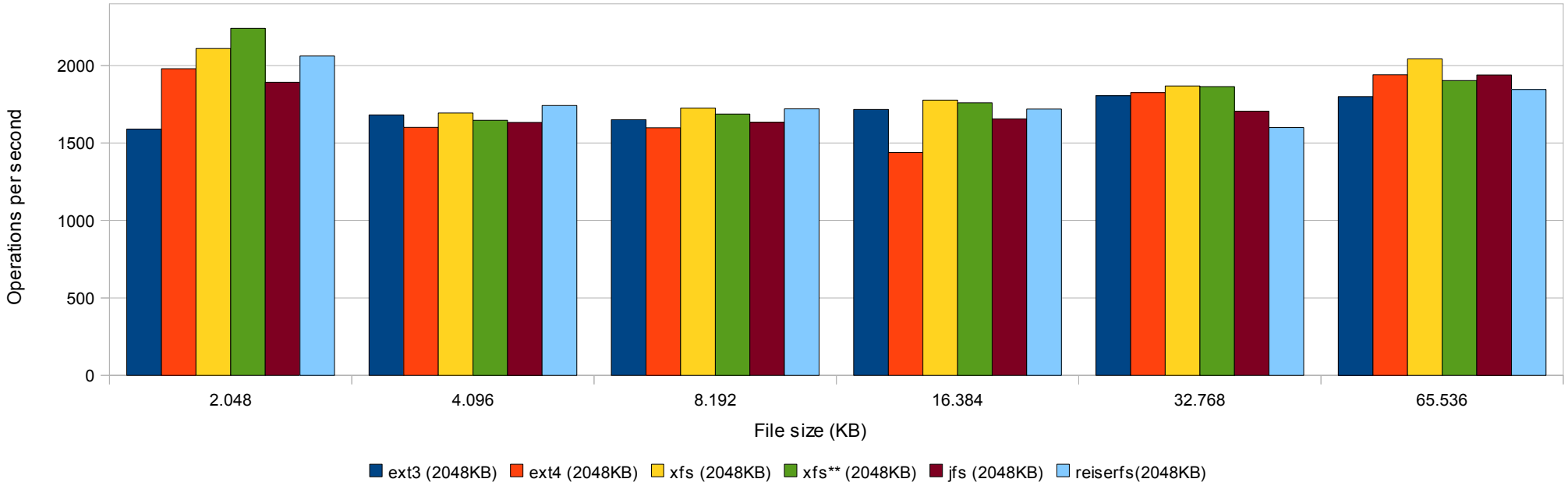

Fread report (16KB record size)
Operations per second
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (16KB)
ext4 (16KB)
xfs (16KB)
xfs** (16KB)
jfs (16KB)
reiserfs(16KB)

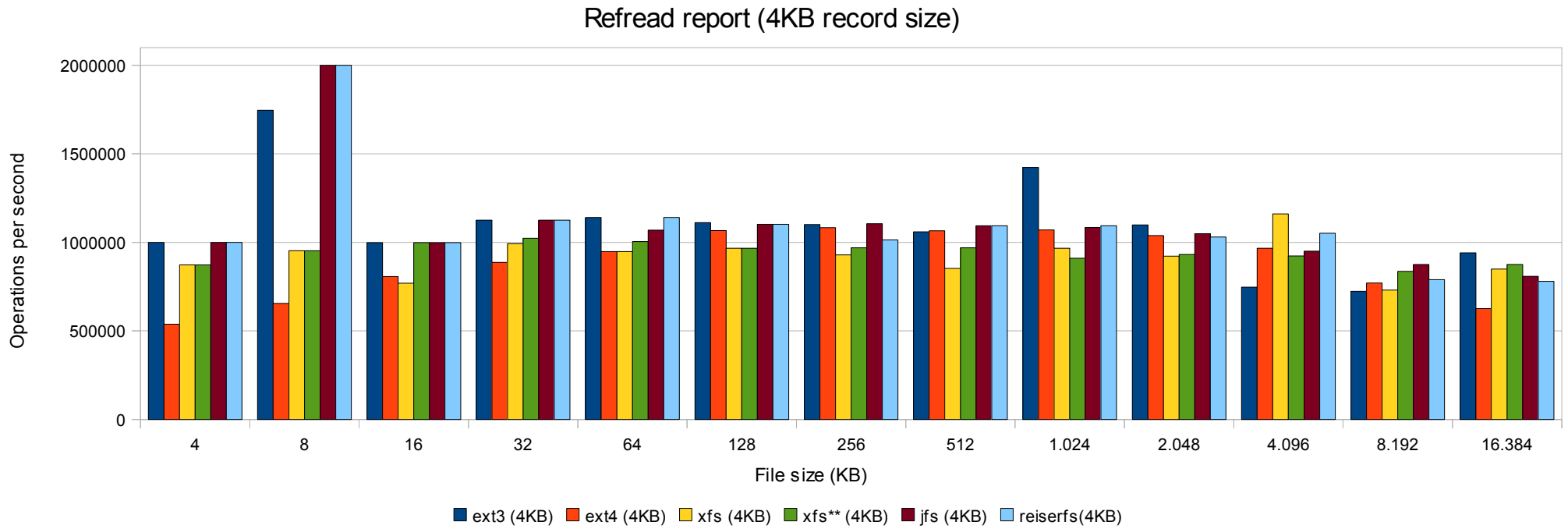

Fread report (64KB record size)
Operations per second
80000
70000
60000
50000
40000
30000
20000
10000
0
64
128
256
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (64KB)
ext4 (64KB)
xfs (64KB)
xfs** (64KB)
jfs (64KB)
reiserfs(64KB)

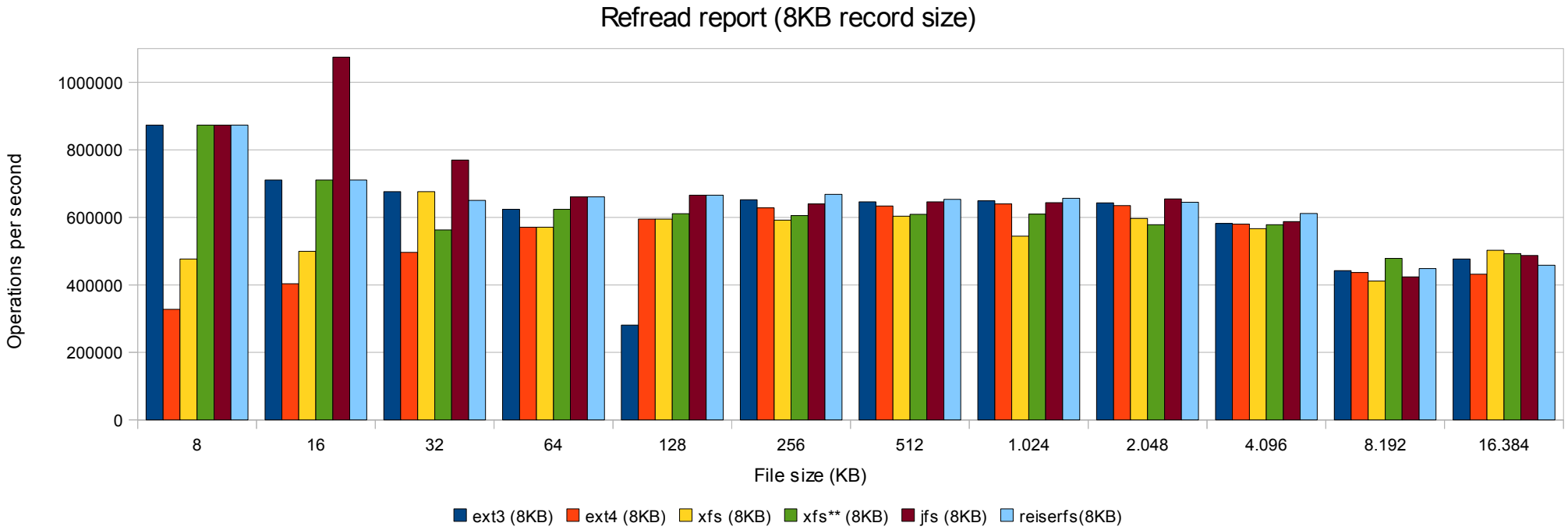
Fread report (512KB record size)



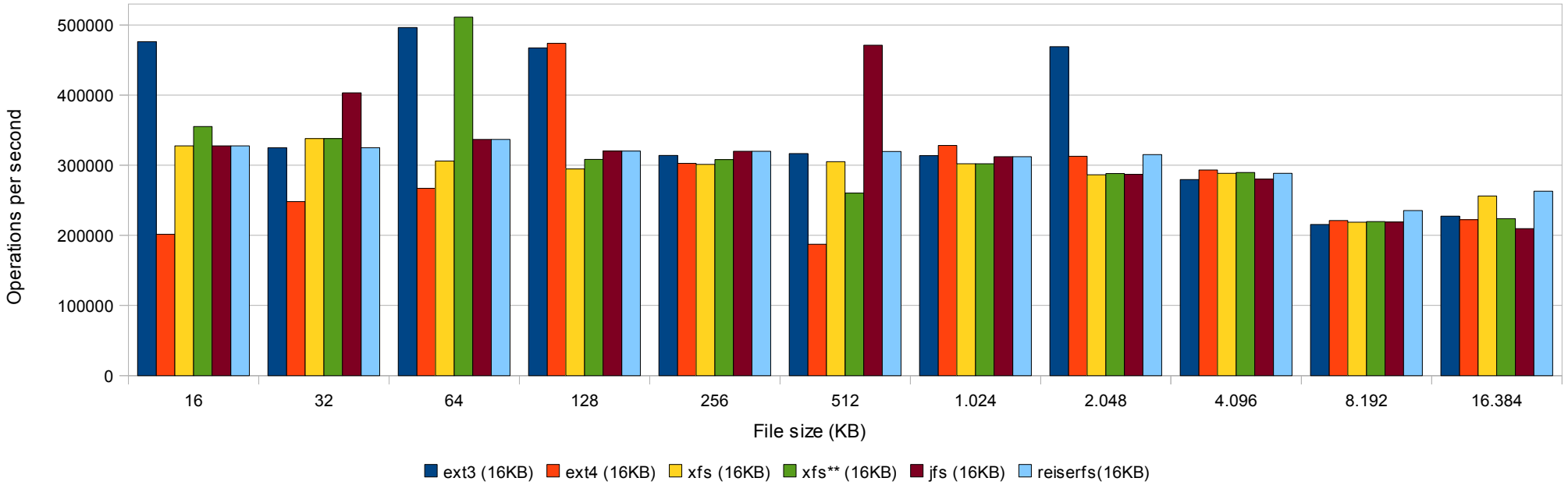
Fread report (2048KB record size)



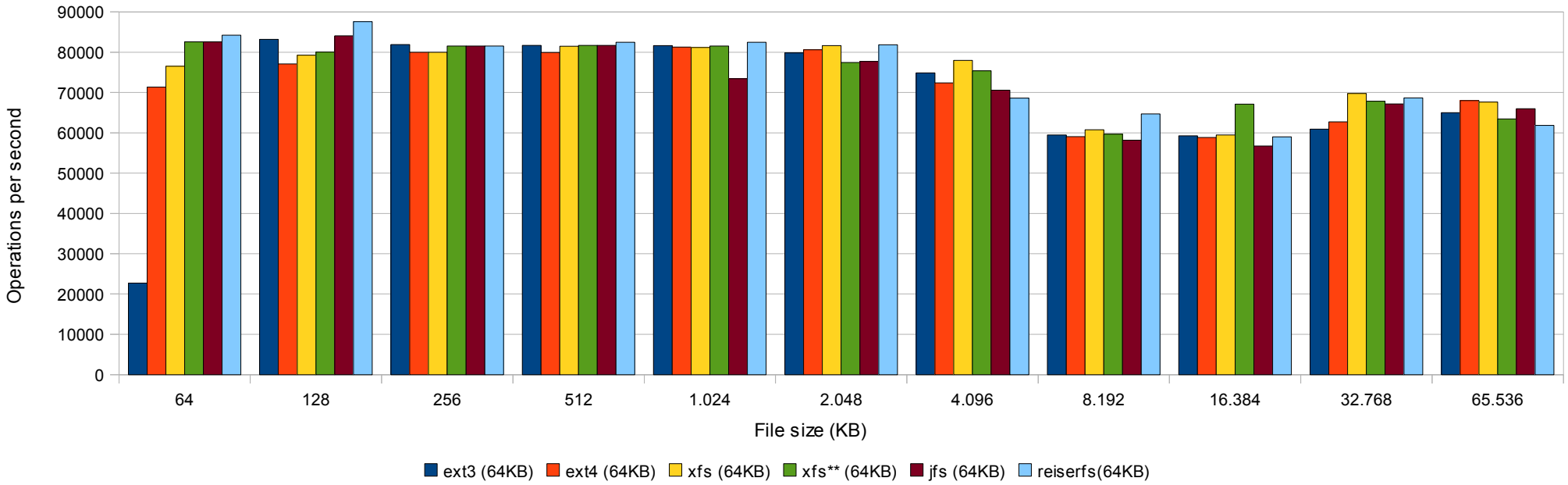

Refread report (4KB record size)
Operations per second
2000000
1500000
1000000
500000
0
4
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (4KB)
ext4 (4KB)
xfs (4KB)
xfs** (4KB)
jfs (4KB)
reiserfs(4KB)

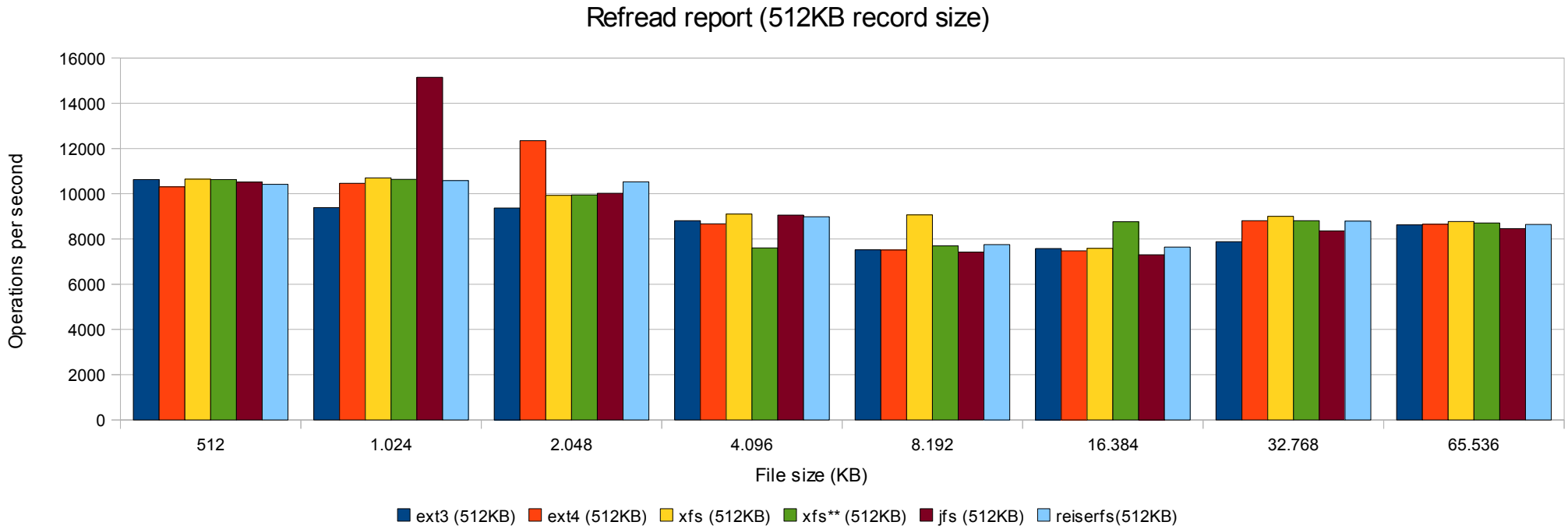

Refread report (8KB record size)
Operations per second
1000000
800000
600000
400000
200000
0
8
16
32
64
128
256
512
1.024
2.048
4.096
8.192
16.384
File size (KB)
ext3 (8KB)
ext4 (8KB)
xfs (8KB)
xfs** (8KB)
jfs (8KB)
reiserfs(8KB)

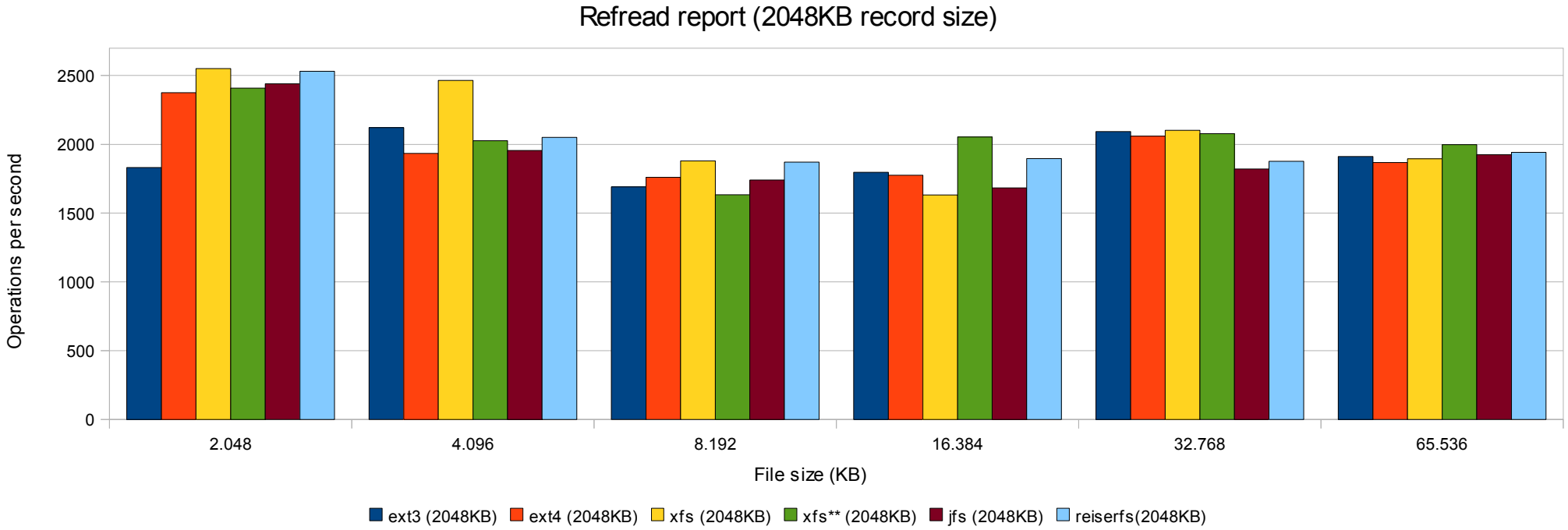
Refread report (16KB record size)



Refread report (64KB record size)

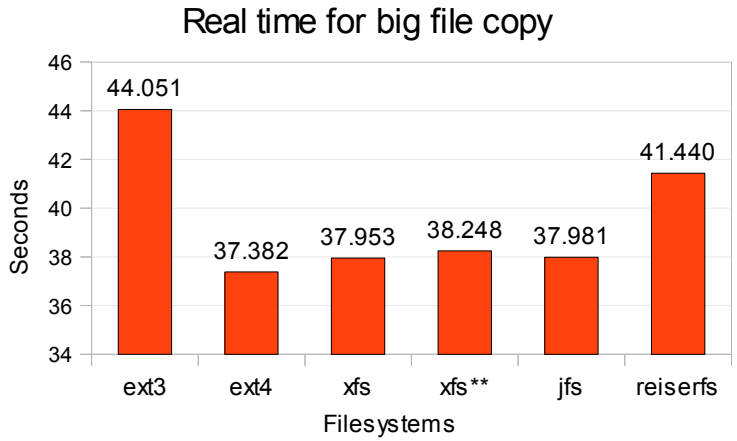


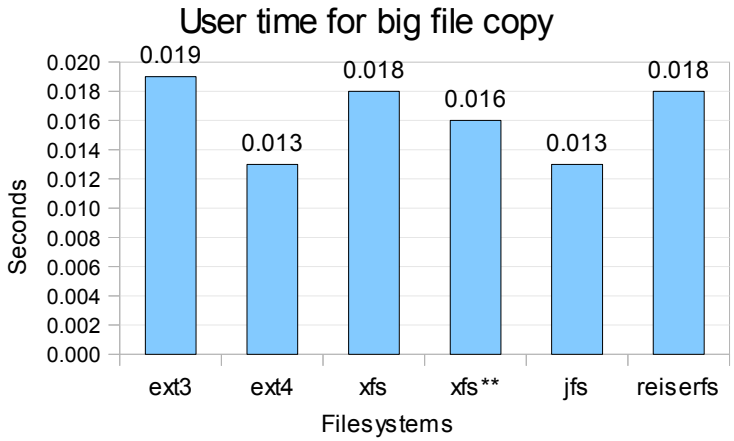

Refread report (512KB record size)
Operations per second
16000
14000
12000
10000
8000
6000
4000
2000
0
512
1.024
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (512KB)
ext4 (512KB)
xfs (512KB)
xfs** (512KB)
jfs (512KB)
reiserfs(512KB)

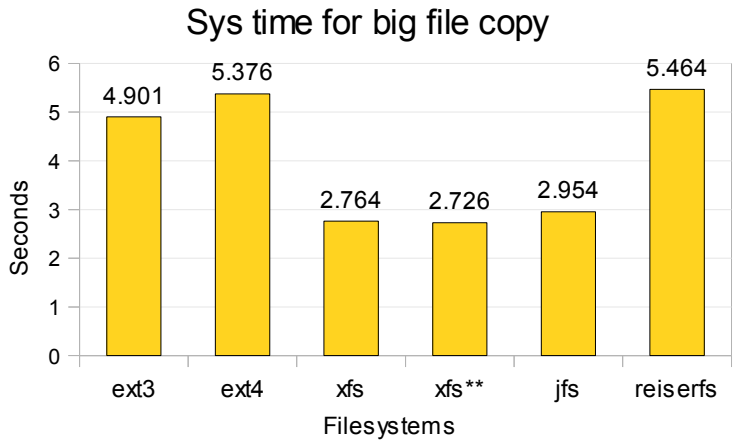

Refread report (2048KB record size)
Operations per second
2500
2000
1500
1000
500
0
2.048
4.096
8.192
16.384
32.768
65.536
File size (KB)
ext3 (2048KB)
ext4 (2048KB)
xfs (2048KB)
xfs** (2048KB)
jfs (2048KB)
reiserfs(2048KB)

Big file copy

	ext3	ext4	xfs	xfs**	jfs	reiserfs
Real	44.051	37.382	37.953	38.248	37.981	41.440
User	0.019	0.013	0.018	0.016	0.013	0.018
Sys	4.901	5.376	2.764	2.726	2.954	5.464

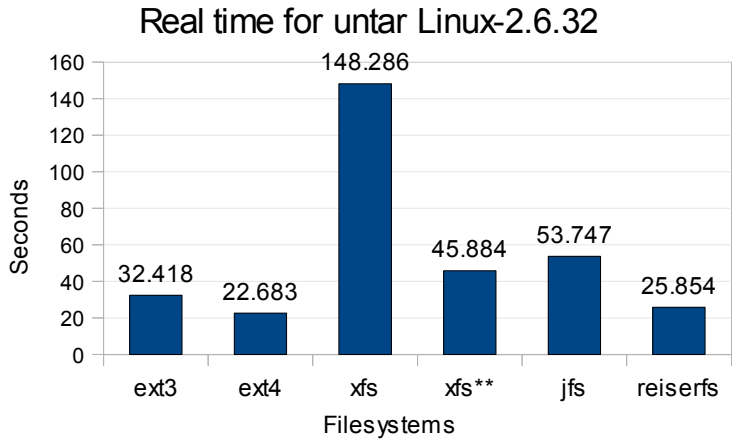


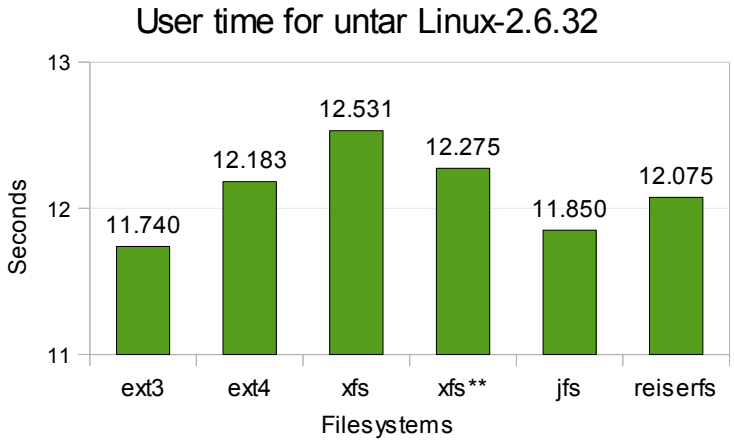


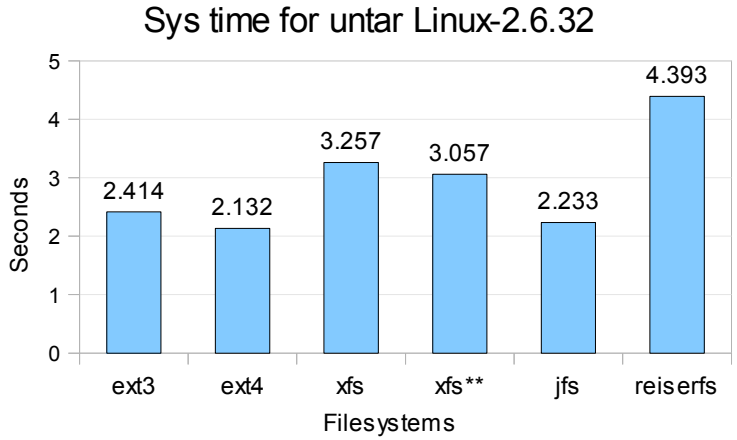


Untar linux-2.6.32.tar.bz2

	ext3	ext4	xfs	xfs**	jfs	reiserfs
Real	32.418	22.683	148.286	45.884	53.747	25.854
User	11.740	12.183	12.531	12.275	11.850	12.075
Sys	2.414	2.132	3.257	3.057	2.233	4.393







rm qt-everywhere-opensource-src-4.6.0/

	ext3	ext4	xfs	xfs**	jfs	reiserfs
Real	1.020	1.015	70.548	13.621	4.456	1.119
User	0.015	0.013	0.200	0.200	0.019	0.015
Sys	0.490	0.670	1.591	1.336	0.558	1.057