

# Week 8 results

We try to use the “Ratio of connections” to help the placement in previous weeks, and it didn't work well.

So I am trying to answer a more fundamental problem, put the algorithms aside, **is it possible for us to use this metric to help the placement.**

## Parameter settings:

Trace file: LASR trace - 95 days long system call trace.

```
2476 UID 16470 PID 455 ?? A 990385736.276707 open("/proc/net/tcp", O_RDONLY, 0) = 5
2580 UID 16470 PID 455 ?? A 990385736.277108 open("/proc/net/tcp", O_RDONLY, 0) = 10
2684 UID 41754 PID 474 ?? A 990385736.450065 open("/var/run/atd.pid", O_CREAT|O_EXCL|O_RDWR, 0) = 3
3272 UID 16470 PID 460 ?? A 990385789.189085 read(5, 1024) = 1024
3304 UID 16470 PID 675 ?? A 990385789.189384 open("/etc/passwd", O_RDONLY, 3931) = 12
3348 UID 16470 PID 675 ?? A 990385789.189479 read(12, 4096) = 3931
3380 UID 16470 PID 675 ?? B 990385789.189500 close(12, 3931) = 0
4008 UID 62936 PID 871 ?? A 990385798.253842 open("/usr/X11R6/lib/X11/fs/config", O_RDONLY, 885) = 3
4072 UID 62936 PID 871 ?? A 990385798.253900 read(3, 28672) = 885
4104 UID 62936 PID 871 ?? A 990385798.253904 read(3, 28672) = 0
```

Edge adding window: 10/100s

Edge expire window: 1 day

Update period: 1day

Using fixed ssd capacity ratio to control the data placement instead of using threshold.

Metrics for placement:

1. access frequency in the future

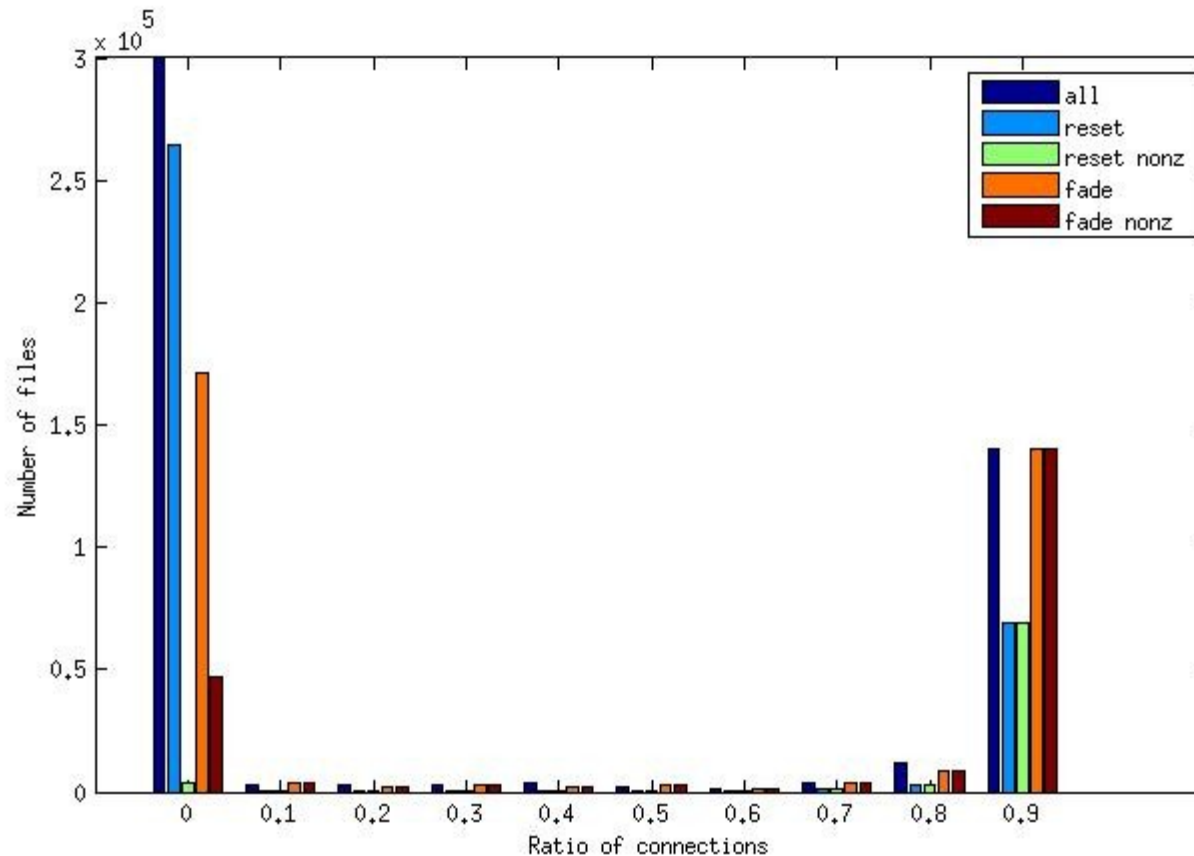
## Results:

To help answer the question above, I use the GOD mode, with which I can see the future and make perfect placement based on future's information.

So, the results I want to exam is that, with this perfect placement, **what are the Ratio of connections of files in SSD.**

For example, on 19<sup>th</sup> day, if I can see that file 1, 3, 5 will be the most popular (with highest access frequency) file on 20<sup>th</sup> day, then I can put file 1, 3, 5 into the SSD, which is the best placement I can do on 19<sup>th</sup> day. Now I want to see, what is the Ratio of connections of file 1, 3, 5 on 19<sup>th</sup> day.

Figure 1 - histogram of ratio of connections (zoom in version)



The figure above is a combined histogram of five separate experiments. The x axis is the ratio of connections of each file, while the y axis is how many files have this ratio of connections.

For each experiment, the result is the sum of all 95 days.

- all: the ratio of connections of all files, not only the popular files
- reset: edges will be removed each day, that is, the weight for each edge will be set to 0 at the beginning of each day
- reset nonz: remove the result of nodes without any edge that has a weight greater than 0
- fade: edges will fade instead of being removed each day, that is, the weight for each edge will be halved at the beginning of each day
- fade nonz: remove the result of nodes without any edge that has a weight greater than 0

Figure 2 – access size of files with nonzero ratio (all)

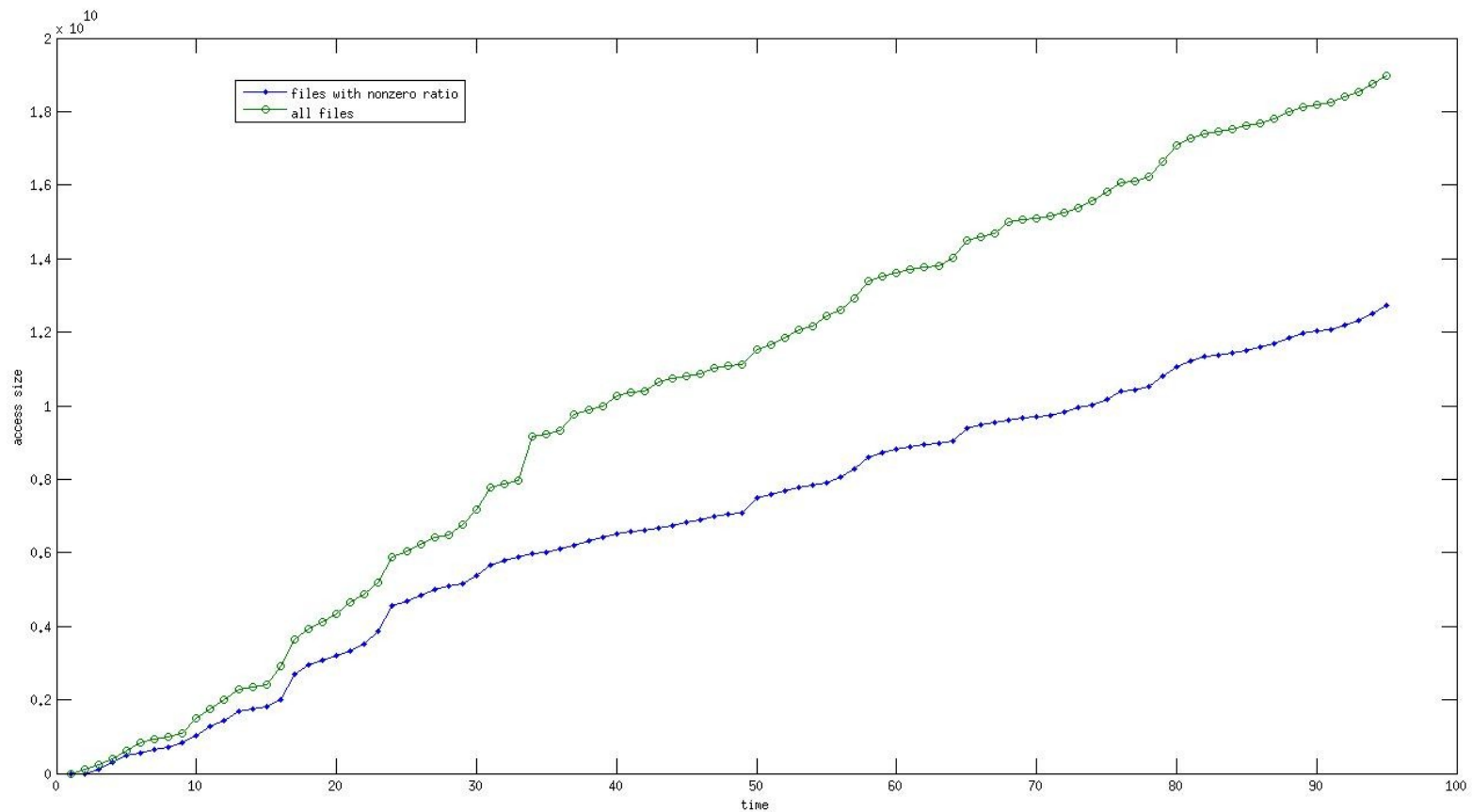
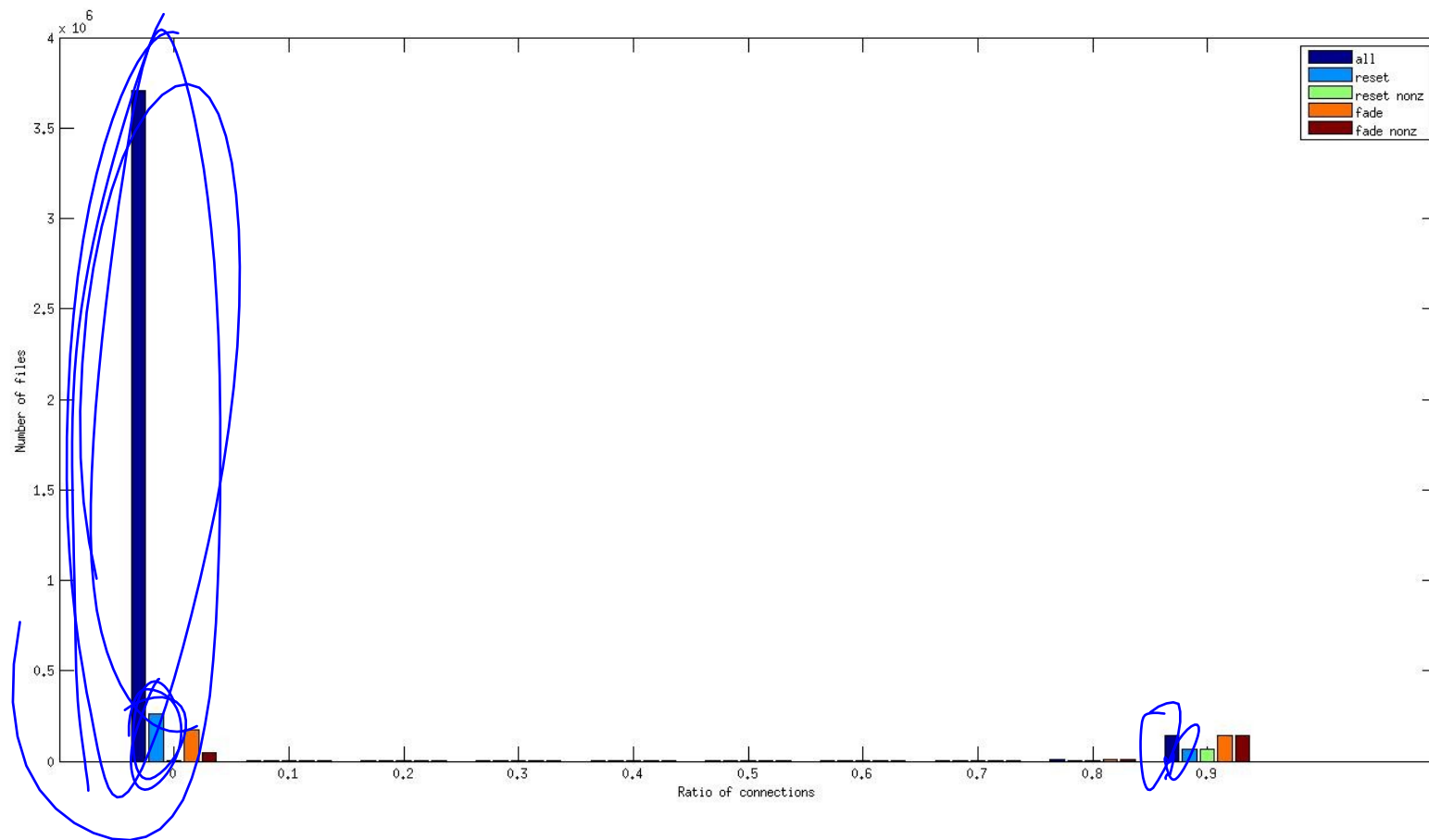


Figure 3 - Original histogram



Thoughts about the results:

- The ratio of connections are great for help choosing files with high ratio of connections, but can do nothing to help choosing files with low ratio of connections (yet they are portion of the files that should be placed on SSDs)

Problem remaining:

- The ratio of connection are based on files that have been put in SSDs, so how much will an unperfect placement affect the distribution of ratio of connections is still a problem

- Even if ratio of connection can help place the files with high ratio of connection. It is still possible that use the access frequency only can pick out these files too.
- Why the histogram looks like this, we have already found out why there is a peak at zero. But why there is a peak at near 1? And why so few files have a ratio between 0.1 to 0.9?
- Will the histogram looks the same on HPC traces?
- The performance difference of perfect placement and using access frequency alone, are there a lot of space for us to improve?
- If all the above question have positive answers, how can we take advantage of the information such a metric provide for placement.