We evaluate the performance of our learning algorithm by replaying realistic workload traces. We use a long-term I/O traces, LASR traces [1], which are taken at system-call level in 2000 and 2001 as part of a security research project.

We track different files' access time during their lifetime. We divide the time into fixed length time slots and count how many times each file has been accessed during each time slot.

Through all the files that have been accessed in a trace, we select 1700 files of them which are more frequently accessed and study their access frequency and time relation.

By analyzing the access of these frequently accessed files, we find out that most of these files can be put into two categories according to the access patterns during their lifetime.

The first category is files which has constant access patterns. Files in this category have been frequently accessed during their whole lifetime, without too much difference between the maximum access point and minimum access point.

Fig. 1 shows a typical file falls into this category. The learning algorithm, especially the Markov chain approach can achieve a higher level of accuracy for this kind of files.

The second category is files with a burst access pattern. Files in this category have only been accessed at very few time slots, but the access count for each time slot that has been accessed is very large.

Fig. 2 shows a typical file falls into this category. Currently, it is pretty hard for any learning algorithm including Markov chain approach to make a good decision for this kind of files.

Reference:

[1] Geoff Kuenning, "LASR traces," http://www.lasr.cs.ucla.edu/seer/seer\_traces.html, 2001