Family of sporadic prefetching algorithms based on the distance function

Voevodkin Vadim Sergeevich

Student

HSE University, Nizhny Novgorod, Russia

YADRO, Nizhny Novgorod, Russia

E-mail: v.voevodkin@yadro.com

Научный руководитель — *Sokolov Andrey Pavlovich*

Modern storage systems intensively utilize data prefetching algorithms while processing sequences of the read requests. Performance of the prefetching algorithm (for instance increase of the cache hit ratio of the cache system – CHR) directly affects overall performance characteristics of the storage system (read latency, IOPS, etc.).

There are widely known prefetching algorithms [1, 3] that are focused on the discovery of the sequential patterns in the stream of requests. This study examines a family of prefetching algorithms that is focused on mining of the pseudo random (sporadic) patterns between read requests – sporadic prefetching algorithms. The key contribution of this paper is that it discovers a new, lightweight family of distance-based sporadic prefetching algorithms (DBSP) that outperforms the best previously known results on MSR traces collection.Another important contribution of this paper is a thorough description of the procedure for comparing the performance of sporadic prefetchers.

The key performance characteristics of data storage systems are latency and IOPS. The latency is the total delay between the arrival of the input read/write request and receiving data requested by the host. IOPS value is the number of Input/Output operations per second.

Comparing prefetching algorithms with each other is a non-trivial task. Various methodologies for comparison of the prefechers were proposed in other works [1, 2, 4]. Unfortunately, comparison methodologies provided in these works lack many important details. For example, which datasets and algorithm parameters were used and other important particulars. In this paper, we will try to eliminate this drawback and present a detailed description of the methodology of comparing sporadic prefetching algorithms.

An appropriate algorithm was found for comparative analysis, which also targets sporadic data streams. Mithril [2] is currently one of the best known prefetching algorithms focused on the sporadic read request sequences. We consider this algorithm as a main baseline for comparison.

The key contributions of the paper are:

- Firstly, this paper introduces the new family of read prefetching algorithms – Distance Based Sporadic Prefetcher (DBSP). These algorithms after optimization of its parameters on MSR traces [3] collection succeeds to overcome Mithtil algorithm near 2% in terms of CHR and 3% in terms of the Precision.

- The second, in this paper, we introduce the detailed methodology for evaluating the quality of the prefetching algorithms.

Литература

1. Gill B.§. and Bathen,Ł.. AMP: Adaptive multi-stream prefetching in a shared cache. // In 5th USENIX Conference on File and Storage Technologies (FAST 07), San Jose, CA. USENIX Association, 2007.
2. Yang,., Karimi,., Sæmundsson,., Wildani,., and Vigfusson,. Mithril: Mining sporadic associations for cache prefetching, 2017. arXiv: 1705.07400.
3. Narayanan,., Donnelly,., and Rowstron,. Write Off-Loading: Practical power management for enterprise storage. // In 6th USENIX Conference on File and Storage Technologies , San Jose, CA. USENIX Association, 2008.
4. Griffioen,. and Appleton,. Reducing file system latency using a predictive approach. // In USENIX summer, P.97–207, 1994.