Taylor & Francis
Taylor & Francis Group

# A Rule-Based Prefetching Approach for Object-Based Storage Device

## SHUIBING HE, DAN FENG, CHUNHUA LI, and YANLI YUAN

Key Laboratory of Data Storage Systems, Ministry of Education of China School of Computer, Huazhong University of Science and Technology, Wuhan, China

*Capturing the advantages of NAS and SAN, Object-based Storage System (OBSS) is becoming an active research area. OBSS is composed of clients, metadata server, and Object-based Storage Devices (OSDs). Similar to the prefetching of file and block, object prefetching is also an effective method to reduce the object access latency in OSD. However, the existing methods focus on intra-object prefetching [1] and they limit the prefetching capability among objects to a certain extent.*

*In this paper, we present a novel inter-object prefetching approach which considers the characteristics of the object access in OSD. This approach adopts a kind of data mining technique called sequential pattern mining to find the object access sequential patterns [2], then produces the object sequential rules according to the patterns, and finally prefetches objects according to the rules and the object access context. The whole process includes the following three modules:*

*1. Object Request Service Module. It maintains a current object access context, which is a l-length recently accessed object sequence. Furthermore, it appends new object information to the historic object access stream when an object is accessed.*
*2. Sequential Rules Mining Module. It is composed of the object sequences preprocessing, the sequential pattern mining and the sequential rules producing modules. The preprocessing module divides the object access stream into many subsequences. Base on the discovered object sequential patterns gained by the sequential pattern mining module, object access sequential rules table is constructed.*
*3. Prefetching Engine Module. When the current requested object is not in the object cache, the prefetching engine selects a matched rule with the highest confidence from the rules table according to the context, then chooses the corresponding objects and preloads them into the object cache.*

*Simulated experimental results show that both prediction accuracy and object hit rate can be improved comparing with the Group-based prefetching method [3] which can also be used for object prefetching.*

## References

1. D. Nagle, D. Serenyi, and A. Matthews. ''The Panasas ActiveScale Storage Cluster – delivering scalable high bandwidth storage,'' *Proceedings of the 2004 ACM/IEEE Supercomputing Conference*, November 2004.
2. J. Pei, J. Han, B. Mortazavi-Asl, ''PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth,'' in *Proceedings of 2001 Int. Conf. Data Engineering*, pp. 215–224, Heidelberg, Germany, April 2001.
3. A. Amer, D. Long, and R. Burns. ''Group-based management of distributed file caches,'' in *Proceedings of the 17th Int'l Conf. on Distributed Computing Systems*, pp. 525–534, July 2002.

Address correspondence to Shuibing He, Key Laboratory of Data Storage Systems, Ministry of Education of China School of Computer, Huazhong University of Science and Technology, Wuhan, China. E-mail: hesbingxq@163.com