

**CHENNAI – PONDICHERRY**

**PREFETCHING ON STORAGE SERVERS THROUGH MINING ACCESS PATTERNS ON BLOCKS**

**ABSTRACT:**

Distributed file systems have been widely deployed as back-end storage systems to offer I/O services for parallel/distributed applications that process large amounts of data. Data prefetching in distributed file systems is a well-known optimization technique which can mask both network and disk latency and consequently boost I/O performance. Traditionally, data prefetching is initiated by the client file systems, however, conventional prefetching schemes are not well suited for client machines that have limited memory and computing capacity. To offer an efficient prefetching approach for resource-limited client machines, this paper proposes a novel server-side prefetching mechanism. Specifically, we propose to piggyback client identification to I/O requests so that server side block access history can be put into context. On the server side, we utilize the horizontal visibility graph technique to transform per-client time series of block access sequences into a connected graph for which we employ Tarjan's algorithm to disclose cut points in the connected graph. We express these patterns with feature tuples and we propose the X-step pattern matching algorithm to find a matching access pattern (i.e., a feature tuple) for a given block access history. Experimental results indicate that our newly proposed prefetching mechanism can ease client machines and their applications from the process of data prefetching, boosting client performance accordingly, and that it yields an attractive increase in data throughput as well.