Enhancing the Dynamic Adaptive Replication Strategy for P2P Systems Using Time-Based Decaying Function

A Project Report
Presented to
The Faculty of the College of
Engineering

San Jose State University
In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Software Engineering

By
Govind Vineeth Reddy, Madupu Sai Kiran, Mandava Abhiteja, Yamsani Sai Santhosh
May 2023

APPROVED

Dr. Haonan Wang, Project Advisor

ABSTRACT

Enhancing the Dynamic Adaptive Replication Strategy for P2P Systems using Time-Based Decaying Function

By Govind Vineeth Reddy, Madupu Sai Kiran, Mandava Abhiteja, Yamsani Sai Santhosh

Availability is one of the most important aspects of a peer-to-peer (P2P) network. To maintain the high availability of data, replication strategies are used to create multiple copies of the data file. As the number of replicas increases, the availability and performance of the system will improve. For each node in the network, the data replication decision is made based on a) which file to replicate, b) when to replicate, and c) where to place the replicated file. Most replication strategies use fixed threshold values on a node's current load of requests as criteria to replicate. However, this approach may lead to over-replication as it does not utilize timing information. For example, Dynamic Adaptive Replication Strategy (DARS) is one of the strategies that can achieve high availability and create a relatively small number of replicas. It addresses the problem of when to replicate the file by recognizing potential overload with overheating similarity (i.e., node's proximity to overload threshold) and finds the optimal placement node using fuzzy clustering analysis. However, when deciding which files need to be replicated, DARS simply picks the file with the highest number of accesses, as it is more likely to cause an overload (i.e., exceed the node's available bandwidth) without considering the timestamps of those requests. Thus, DARS may falsely pick a file that was frequently used in the past but miss a file that is more frequently used now, leading to suboptimal performance. To overcome this problem, we propose to use a Time-Based Decaying Function (TBDF) to rank the candidate files for replication. TBDF takes the temporal locality of file accesses into consideration by utilizing timestamps of requests. It also reduces the weight and relevance of file accesses as time passes. To evaluate our approach, we developed a highly efficient event-driven P2P network simulator and configured it to several network topologies. The result shows that our approach archives a 37.65% improvement in node availability, 2.14% available bandwidth, and 3.35% latency compared to DARS.

Acknowledgments

The authors are deeply indebted to Professor Haonan Wang for his invaluable comments and assistance in the preparation of this study.