DE LA SALLE UNIVERSITY

A Heuristic Approach to File Placement

in Distributed Systems

1 8000

A Thesis
Presented to
the Faculty of the College of Computer Studies
De La Salle University

In Partial Fulfilment
of the Requirements for the Degree of
Master of Science in
Computer Science

by

Huang, En-Hsin

DE LA SALLE UNIVERSITY

ABSTRACT

In the Distributed Computing System environment, an efficient and effective means of resource allocation and monitoring is difficult to achieve. A highly available and reliable system rarely exist. Since each processing element is independent from each other and can be geographically distributed, the redundancy of tasks and resources are permissible and widely observed. It is therefore important to let the processing element access the nearest available resources so as to minimize communication cost.

When physical link failure occurs, system could consequently be partitioned and resources could be inaccessible to the processing elements. It is therefore important to locate the still possible accessible resources to ensure the successful execution of processing elements.

This paper presents a conceptual model with algorithms for computing a near optimal and a reliable solution and performing reliability analysis of distributed task and distributed system based on graph-theoretic formalism, matrix and set operation. Formal proof of correctness of algorithm is conducted and sensitivity analysis is performed to examine various relationship among availability, degree of duplication and network topologies in the Distributed Systems environment.

First, a model is constructed to represent the respective types of topologies, and the initial maximum degree of redundancy of data file is determined. A set of feasible file allocation with corresponding query cost is

computed. To determine the distributed task reliability and the distributed system reliability of the computed solutions, a set of minimum spanning tree is constructed and a terminal algorithm is applied.

With the algorithm presented here, the ability for quick access to resources and maintenance of the system's reliability will greatly improve the overall system performance.