Cost Minimization Algorithms for Data Center Management

Objective:

The aim to optimally adjusting servers and dispatching workload to active servers.

Abstract:

Due to the increasing usage of cloud computing applications, it is important to minimize energy cost consumed by a data center, and simultaneously, to improve quality of service via data center management. One promising approach is to switch some servers in a data center to the idle mode for saving energy while to keep a suitable number of servers in the active mode for providing timely service. In this paper, we design both online and offline algorithms for this problem. For the offline algorithm, we formulate data center management as a cost minimization problem by considering energy cost, delay cost (to measure service quality), and switching cost (to change servers's active/idle mode). Then, we analyze certain properties of an optimal solution which lead to a dynamic programming based algorithm. Moreover, by revising the solution procedure, we successfully eliminate the recursive procedure and achieve an optimal offline algorithm with a polynomial complexity.

Introduction:

Recent years, many researchers devote themselves into the area of data center management for cloud computing applications. The goals of data center management may include minimizing energy cost and improving quality of service.

Energy cost is a major part of a center's budget which should be minimized to decrease service provider's cost, and more importantly, to keep our Earth green. One approach to minimize energy cost is to switch some servers from active mode to idle mode whenever possible These switching decisions are made based on environment of the servers, such as network state or storage state Meanwhile, we want to achieve good service quality, which can be measured by the average delay of serves' responding time. For this purpose, there should be enough active servers in order to process tasks initiated by clients in time. To achieve both goals of data center management, we should maintain a suitable number of active servers and then distribute jobs to these active servers

Research efforts on data center management can be classified as inter-center or intra-center management. Inter-center management, e.g., considers how to distribute workload (jobs) among data centers while intra-center management, deals with how to distribute workload among servers within a data center. In this paper, we focus on intra-center management and consider dynamic workload (measured by the number of jobs) over a period of time. In the minimization problem, the cost includes energy cost, delay cost and switching cost. Energy cost at an active server can be modeled as a function of its assigned workload, or a function of the percentage of its processing capability used to finish the assigned workload. Delay cost at an active server can be used to measure the quality of service, which is also a function of assigned workload. We call the combination of these two costs as the operating cost. Switching cost is the cost spent when changing a server from active mode to idle mode or reversely. In the cost minimization problem, we aim to optimally adjusting servers' status and dispatching workload to active servers.

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032Ph:080-40969981, Website:www.technofist.com. E-mail:technofist.projects@gmail.com