

**Minimal Cost Server Configuration for Meeting Time-Varying Resource Demands in Cloud Centers**

**Abstract:**

We consider the minimal cost server configuration for meeting resource demands over multiple time slots. Specifically, there are some heterogeneous servers. Each server is specified by a cost, certain amounts of several resources, and an active interval, i.e., the time interval that the server is planed to work. There are different overall demands for each type of resource over different time slots. A feasible solution is a set of servers such that at any time slot, the resources provided by the selected servers are at least their corresponding demands. Notice that, a selected server can not provide resources for the time slots out of its active interval. The total cost of the solution is the summation of the costs of all selected servers. The goal is to find a feasible solution with minimal total cost. To solve our problem, we present a randomized approximation algorithm called partial rounding algorithm (PRA), which guarantees O(log(KT))-approximation, i.e., ηlog(KT)-approximation, where η is a positive constant. Furthermore, to minimize η as much as possible, we propose a varied Chernoff bound and apply it in PRA. We perform extensive experiments with random inputs and a specific application input. The results show that PRA with our varied Chernoff conclusion can find solutions closing to the optimal one.