

**CHENNAI – PONDICHERRY**

**Achieving Fairness-aware Two-level Scheduling for Heterogeneous Distributed Systems**

**Abstract**

In a heterogeneous distributed system composed of various types of computing platforms such as supercomputers, grids, and clouds, a two-level scheduling approach can be used to effectively distribute resources of the platforms to users in the first-level, and map tasks of the users in nodes for each platform in the second-level for executing many-task applications. When scheduling heterogeneous resources, service providers of the system should consider the fairness among multiple users as well as the system efficiency. However, the fairness cannot be achieved by simply distributing an equal amount of resources from each platform to every user. In this paper, we investigate how to address the fairness issue among multiple users in a heterogeneous distributed system. We present three first-level resource allocation policies of a provider affinity first policy, an application affinity first policy, and a platform affinity based round-robin policy, and two second-level task mapping policies of a most affected first policy and a co-runner affinity based round-robin policy. Using trace-based simulations, we evaluate the performance of various combinations of the first and second level scheduling policies. Our extensive simulation results demonstrate that the first-level policy plays a crucial role to achieve relatively good fairness.