World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:17, No:07, 2023

Optimization of Topology-Aware Job Allocation on a High-Performance Computing Cluster by Neural Simulated Annealing

Authors : Zekang Lan, Yan Xu, Yingkun Huang, Dian Huang, Shengzhong Feng

Abstract : Jobs on high-performance computing (HPC) clusters can suffer significant performance degradation due to inter-job network interference. Topology-aware job allocation problem (TJAP) is such a problem that decides how to dedicate nodes to specific applications to mitigate inter-job network interference. In this paper, we study the window-based TJAP on a fat-tree network aiming at minimizing the cost of communication hop, a defined inter-job interference metric. The window-based approach for scheduling repeats periodically, taking the jobs in the queue and solving an assignment problem that maps jobs to the available nodes. Two special allocation strategies are considered, i.e., static continuity assignment strategy (SCAS) and dynamic continuity assignment strategy (DCAS). For the SCAS, a 0-1 integer programming is developed. For the DCAS, an approach called neural simulated algorithm (NSA), which is an extension to simulated algorithm (SA) that learns a repair operator and employs them in a guided heuristic search, is proposed. The efficacy of NSA is demonstrated with a computational study against SA and SCIP. The results of numerical experiments indicate that both the model and algorithm proposed in this paper are effective.

Keywords : high-performance computing, job allocation, neural simulated annealing, topology-aware
Conference Title : ICCC 2023 : International Conference on Cluster Computing
Conference Location : Tokyo, Japan
Conference Dates : July 17-18, 2023