Allocation Algorithms Maximize Throughput in Parallel Supercomputers

Contacts: Vitus Leung, Computer Science & Informatics, *vjleung@sandia.gov*
Cindy Phillips, Discrete Math & Complex Systems, *caphill@sandia.gov*

The CPA (at less than one percent of the cost of a parallel computer) is an example of how a relatively small investment in computer algorithms can dramatically leverage the return on a large investment in computer hardware.

In collaboration with researchers from Stony Brook University and the University of Illinois, Urbana-Champaign, CCIM has developed innovative solutions to resource allocation for parallel processing on supercomputers [1, 2] (Figure 1). Remarkably, the allocation algorithms use space-filling curves to reorder a network of processors so that locations remain close in the physical network of processors after reordering (Figure 2). They also use bin packing (Figure 3) for contiguous allocation and span minimization for noncontiguous allocation to find open processors that are close to each other. In experiments, this optimized node allocation strategy increased throughput by 23%, thus processing five jobs in the time it normally takes to process four.

The first production uses of some of these algorithms were in Cplant system software released in 2002 and 2003.

One of the current production uses of some of these algorithms is in the Compute Process Allocator (CPA). After extensive prototype development, the CPA's innovative solution was carried to the commercial sector in 2005 when CPA was licensed to Cray, Inc. The breadth of impact has been extended through software licensing to numerous laboratory and research centers that bought XT3/4 systems from Cray. Sites with supercomputers using CPA include SNL's Red Storm. The CPA (at less than one percent of the cost of a parallel computer) is an example of how a relatively small investment in computer algorithms can dramatically leverage the return on a large investment in computer hardware.

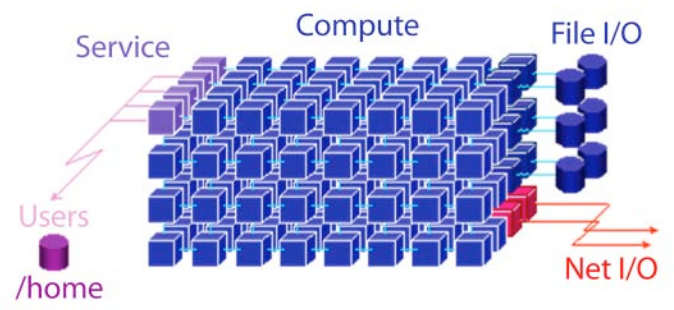

Figure 1. Parallel supercomputer.

With CPA, Sandia has leveraged research sponsored by the Laboratory Directed Research and Development (LDRD), Computer Science Research Foundation (CSRF), and Advanced Scientific Computing Research (ASCR), filed two patent applications on the technical advances underlying CPA, and affected the external community through software licensing. For its superior strategy and scalability over other allocators, the CPA won a prestigious 2006 R&D 100 Award.

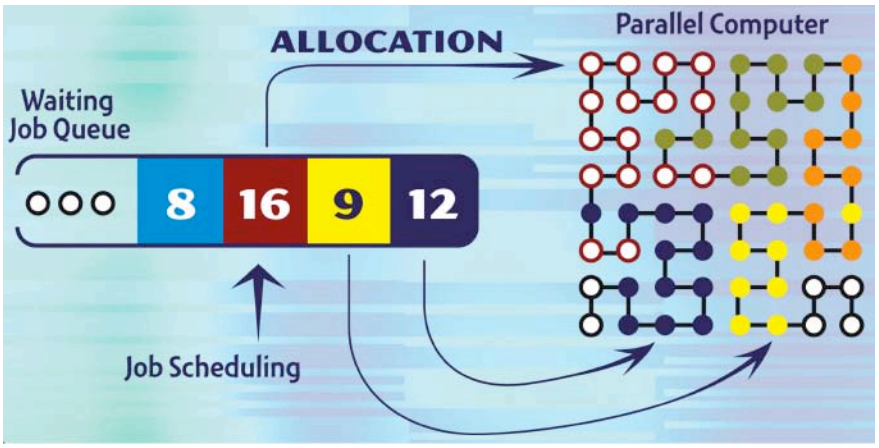

Figure 2. Allocation Algorithm with Hilbert Space-Filling Curve and Span Minimization.

The other current production use of one of these algorithms is in LLNL's Simple Linux Utility for Resource Management (SLURM). SLURM provides an Applications Programming Interface (API) for integration with external schedulers such as the Moab Cluster Suite. SLURM provides resource management on about 1000 computers worldwide, including some of the most powerful computers in the world.

References

1. Bender, M. A., Bunde, D. P., Demaine, E. D., et al. (2008). Communication-Aware Processor Allocation for Supercomputers: Finding Point Sets of Small Average Distance," *Algorithmica*, Vol. **50**, No. 2, pp. 279-298.

2. Leung, V. J., Arkin, E. M., Bender, M. A., et al. (2002). Processor Allocation on Cplant: Achieving General Processor Locality Using One-Dimensional Allocation Strategies, *Proceedings IEEE International Conference on Cluster Computing*.

This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.

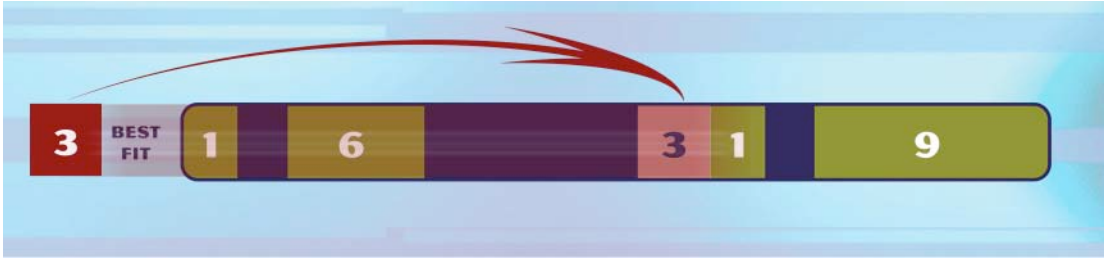

Figure 3. Best-fit packing.