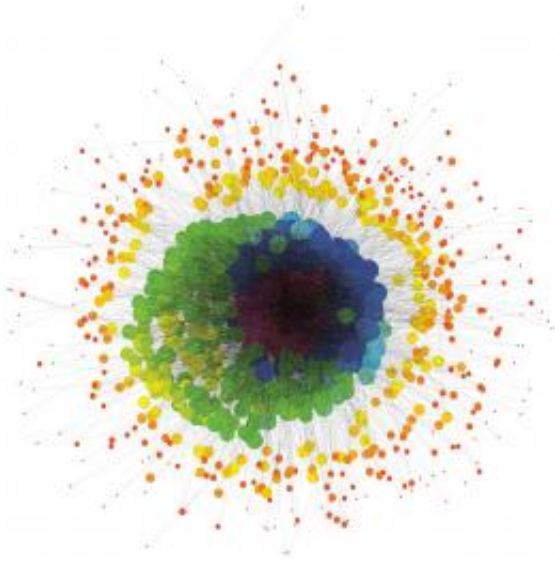


New standard proposed for supercomputing

November 15 2010



This small, synthetic graph was generated by a method called Kronecker multiplication. Larger versions of this generator, modeling real-world graphs, are used in the Graph500 benchmark. (Courtesy of Jeremiah Willcock, Indiana University)

A new supercomputer rating system will be released by an international team led by Sandia National Laboratories at the Supercomputing Conference 2010 in New Orleans on Nov. 17.

The rating system, Graph500, tests supercomputers for their skill in analyzing large, graph-based structures that link the huge numbers of data points present in biological, social and [security problems](#), among other areas.

“By creating this test, we hope to influence computer makers to build computers with the architecture to deal with these increasingly complex problems,” Sandia researcher Richard Murphy said.

Rob Leland, director of Sandia’s Computations, Computers, and Math Center, said, “The thoughtful definition of this new competitive standard is both subtle and important, as it may heavily influence computer architecture for decades to come.”

The group isn’t trying to compete with Linpack, the current standard test of [supercomputer](#) speed, Murphy said. “There have been lots of attempts to supplant it, and our philosophy is simply that it doesn’t measure performance for the applications we need, so we need another, hopefully complementary, test,” he said.

Many scientists view Linpack as a “plain vanilla” test mechanism that tells how fast a computer can perform basic calculations, but has little relationship to the actual problems the machines must solve.

The impetus to achieve a supplemental test code came about at “an exciting dinner conversation at Supercomputing 2009,” said Murphy. “A core group of us recruited other professional colleagues, and the effort grew into an international steering committee of over 30 people.” (See www.graph500.org)

Many large computer makers have indicated interest, said Murphy, adding there’s been buy-in from Intel, IBM, AMD, NVIDIA, and Oracle corporations. “Whether or not they submit test results remains to be seen, but their representatives are on our steering committee.”

Each organization has donated time and expertise of committee members, he said.

While some computer makers and their architects may prefer to ignore a new test for fear their machine will not do well, the hope is that large-scale demand for a more complex test will be a natural outgrowth of the greater complexity of problems.

Studies show that moving data around (not simple computations) will be the dominant energy problem on exascale machines, the next frontier in supercomputing, and the subject of a nascent U.S. Department of Energy initiative to achieve this next level of operations within a decade, Leland said. (Petascale and exascale represent 10 to the 15th and 18th powers, respectively, operations per second.)

Part of the goal of the Graph500 list is to point out that in addition to more expense in data movement, any shift in application base from physics to large-scale data problems is likely to further increase the application requirements for data movement, because memory and computational capability increase proportionally. That is, an exascale computer requires an exascale memory.

“In short, we’re going to have to rethink how we build computers to solve these problems, and the Graph500 is meant as an early stake in the ground for these application requirements,” said Murphy.

How does it work?

Large data problems are very different from ordinary physics problems.

Unlike a typical computation-oriented application, large-data analysis often involves searching large, sparse data sets performing very simple computational operations.

To deal with this, the Graph 500 benchmark creates two computational kernels: a large graph that inscribes and links huge numbers of

participants and a parallel search of that graph.

“We want to look at the results of ensembles of simulations, or the outputs of big simulations in an automated fashion,” Murphy said. “The Graph500 is a methodology for doing just that. You can think of them being complementary in that way — graph problems can be used to figure out what the simulation actually told us.”

Performance for these applications is dominated by the ability of the machine to sustain a large number of small, nearly random remote data accesses across its memory system and interconnects, as well as the parallelism available in the machine.

Five problems for these computational kernels could be cybersecurity, medical informatics, data enrichment, social networks and symbolic networks:

- * **Cybersecurity:** Large enterprises may create 15 billion log entries per day and require a full scan.

- * **Medical informatics:** There are an estimated 50 million patient records, with 20 to 200 records per patient, resulting in billions of individual pieces of information, all of which need entity resolution: in other words, which records belong to her, him or somebody else.

- * **Data enrichment:** Petascale data sets include maritime domain awareness with hundreds of millions of individual transponders, tens of thousands of ships, and tens of millions of pieces of individual bulk cargo. These problems also have different types of input data.

- * **Social networks:** Almost unbounded, like Facebook.

- * **Symbolic networks:** Often petabytes in size. One example is the human cortex, with 25 billion neurons and approximately 7,000 connections each.

“Many of us on the steering committee believe that these kinds of

problems have the potential to eclipse traditional physics-based HPC [high performance computing] over the next decade,” Murphy said.

While general agreement exists that complex simulations work well for the physical sciences, where lab work and simulations play off each other, there is some doubt they can solve social problems that have essentially infinite numbers of components. These include terrorism, war, epidemics and societal problems.

“These are exactly the areas that concern me,” Murphy said. “There’s been good graph-based analysis of pandemic flu. Facebook shows tremendous social science implications. Economic modeling this way shows promise.

“We’re all engineers and we don’t want to over-hype or over-promise, but there’s real excitement about these kinds of big data problems right now,” he said. “We see them as an integral part of science, and the community as a whole is slowly embracing that concept.

“However, it’s so new we don’t want to sound as if we’re hyping the cure to all scientific ills. We’re asking, ‘What could a computer provide us?’ and we know we’re ignoring the human factors in problems that may stump the fastest computer. That’ll have to be worked out.”

Provided by Sandia National Laboratories

Citation: New standard proposed for supercomputing (2010, November 15) retrieved 27 April 2024 from <https://phys.org/news/2010-11-standard-supercomputing.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.