April 2010

Center for Adaptive Supercomputing Software

http://cass-mt.pnl.gov/

CASS Research Areas

Applications

- » Social Media Analysis
- » Contingency Analysis for the Electric Power Grid
- » High-Throughput Video Analysis
- » Understanding Text Documents

Architectures

- » Architectural Studies on Multithreaded Architectures

Languages

- » Chapel for hybrid systems

Methods

- » Dynamic Network Analysis
- » Social Network Analysis
- » Irregular Database Systems

Runtime Systems

- » Compiler and Runtime System
- » Performance Analysis and Tools
- » Communication software for hybrid systems

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

CASS research traces the evolution of swine flu

Shahid Bokhari, Laura Pomeroy & Daniel Janies
Department of Biomedical Informatics
The Ohio State University

A research project funded by the Center for Adaptive Supercomputing Software (CASS) is analyzing data pertaining to the pandemic strain of the H1N1 influenza virus known as swine flu to try and identify how the virus evolves.

The genome of the influenza virus has eight distinct RNA segments of lengths varying from 900 to 2500 bases. Influenza evolves by mutation as well as reassortment. Mutation is the modification of nucleotide bases (ACTG) in the genome, while reassortment refers to the interchange of entire segments between two strains of the virus. Reassortment takes place when a host cell is infected by two different viruses—the eight segments from each virus intermix and give rise to new combinations of segments. Reassortment can cause large changes in the genome of a virus and can lead to new strains against which humans or animals have little resistance. This is why the pandemic strain of H1N1 influenza, a product of reassortment, is of great concern to public health authorities.

Established tools for tracing the evolution of organisms, such as phylogenetic trees, cannot directly account for reassortment. We have developed the concept of Reassortment Networks that can be used to analyze the evolution of influenza and other organisms that evolve via mutation as well as reassortment. To analyze the evolutionary relationships between n viruses with s segments each over t stages of evolution requires the construction of a graph with stn2 nodes. A shortest tree in this graph gives the minimum cost evolutionary paths to a specified target virus and can be found in stn3 time. As the number of influenza viruses known is more than 5000, the time and space requirements of this problem are quite large and, furthermore, it is difficult to parallelize well on conventional parallel supercomputers because of the overheads of problem partitioning and interprocessor communications.

The Cray XMT is insensitive to partitioning issues and its massively multithreaded architecture hides memory latency. It is thus well suited to set up and solve reassortment networks. Over the past year we have created an implementation of our algorithm on "Cougar," the 128 processor XMT at CASS at Pacific Northwest National Laboratory. Our algorithm analyzes all known influenza-A viruses and takes about one hour per evolutionary stage using all 128 processors. We are currently fine-tuning our algorithm and comparing its outputs against known biological results.

A paper describing the theory behind Reassortment Networks has been published in the *IEEE Transactions on Computational Biology and Bioinformatics*, ***www.computer.org/portal/web/csdl/doi/10.1109/TCBB.2008.73***.

RECENT EVENTS

SIAM Conference – Seattle, WA

John Feo presented "Language Requirements for Large-Scale Shared Memory Programming" at the SIAM Conference in Seattle, Washington, February 24-26th. John's presentation was part of the larger session titled "Exploring Languages for Expressing Medium to Massive On-chip Parallelism" organized by John Shalf of Lawrence Berkeley National Laboratory. John noted that there is abundant fine-grain parallelism in many problems that can be expressed using current programming languages. The obstacle is bulk synchronous computer systems that support only coarse grain parallelism.

Mahantesh Halappanavar also presented at SIAM. His presentation "Parallel Graph Matching and Coloring for Petascale Scientific Computing" was part of the larger session titled "Parallel Algorithms and Software for Massive Graphs."

We are also proud of collaborator, **Kamesh Madduri**, of Lawrence Berkeley National Laboratory who was given a Junior Scientist Prize Presentation award for "High Performance Computing for Massive Graph Analysis" during the SIAM Conference.

LCI International Conference on High-Performance Clustered Computing – Pittsburgh, PA

The paper titled "MODA: A Memory Centric Performance Analysis Tool" was submitted by Andres Marquez and accepted by the 11th annual LCI conference that took place in March.

MODA: A Memory Centric Performance Analysis Tool

Manzano JB, A Marquez, and R Gao. 2009. "MODA: A Memory Centric Performance Analysis Tool." In *The 11th LCI International Conference on High-Performance Clustered Computing*

This paper introduces the MODA memory centric performance tool prototype. MODA is designed to instrument, collect and analyze streams of memory requests. The tool provides a dynamic application behavior view of memory requests as they access various memory organizational constructs. MODA enables the detection of algorithmic and architectural resource conflicts that can greatly degrade performance. Moreover, possible scalability deterrents can be found predictively at relatively small data set scale. Careful management of shared resource bandwidth will be one of the future multi-core technologies. The Cray XMT (at PNNL) is ideally suited to serve as a test platform to analyze shared resource contention at massive scale. The initial prototype tool has been implemented for the Cray XMT multiprocessor and it has been used to analyze the behavior of small benchmark examples.

UPCOMING EVENTS

IEEE International Parallel & Distributed Processing Symposium (IPDPS)

Friday, April 23 in Atlanta, Georgia

David Haglin (PNNL) and **Eric Goodman** (Sandia) will be co-presenting their paper, "Hashing Strategies for the Cray XMT".

This paper explores the performance and suitability of the XMT for hashing large amounts of data using the two most common hashing strategies of "linear probing" and "hashing with chaining." Some adaptations of the typical hashing with changing with chaining were requires to achieve good performance on the XMT. These techniques scale really well on the XMT.

Cray Users Group (CUG)

May 24-27 in Edinburgh, Scotland

Jace Mogill and **David Haglin** will be presenting their paper "A Comparison of Shared Memory Parallel Programming Models".

The paper is of interest to scientists who wish to use their programs on both conventional shared memory computers and the Cray XMT. Jace and David will compare different parallel algorithms in terms of the amount of concurrency they afford, and how that concurrency is implemented using different programming models and tools.

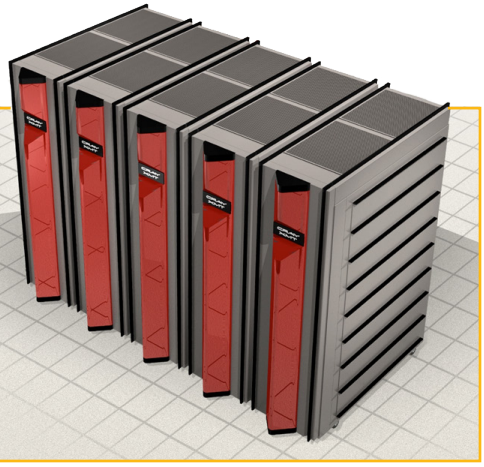
Computing Frontiers

May 17-19 in Bertinoro, Italy

Antonio Tumeo will be attending the conference and looking to learn new techniques that he can bring back and implement in the simulator.

During the trip, Antonio will be giving a talk at Dipartimento di Elettronica e Informazione, Politecnico di Milano (Milano, Italy, 20 May 2010) on CASS, with a particular focus on Task 11, presenting the challenges in managing irregular applications on current high performance architectures, discussing the current status of our Cray-XMT simulator for architectural studies and showing some of the preliminary results.

John Feo,
Director of CASS
(509) 375-3768
John.feo@pnl.gov
cass-mt.pnl.gov/



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

PNNL-SA-72372