Pacific Northwest NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

"We have discovered genes implicated in breast cancer using a massively multithreaded algorithm on the Cray XMT. It's like finding a needle in a haystack. The algorithm searches for genes whose removal quickly causes networks and pathways in the cell to breakdown."

- Georgia Tech computational scientist and engineer David A. Bader.

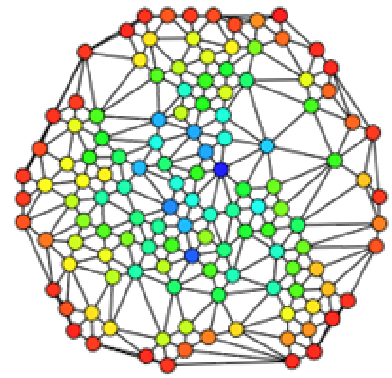
U.S. DEPARTMENT OF ENERGY

Center for Adaptive Supercomputing Software-Multithreaded Architectures (CASS-MT)

STING: Spatio-Temporal Informatics for Networks and Graphs

At a glance

Spatio-Temporal Informatics for Networks and Graphs (STING) refers to the mechanism by which semantic data—data or entities that are related in various ways—are processed for analysis. Understanding the complex dynamics of entity networks can provide the key to unlocking critical information about behaviors in various types of communities.



Dynamic spatio-temporal graph

What we do

At PNNL's Center for Adaptive Supercomputing Software-Multithreaded Architectures (CASS-MT), researchers are developing methods for analyzing massive complex semantic networks. More thoroughly understanding social networks, for example, can help tackle challenges such as influencing change and understanding critical trends in behaviors and customs. Similar computational capabilities can be applied to finding vulnerabilities in the power grid and monitoring important protein interactions in cancer research.

In the context of social networks, for instance, Facebook represents a network of over 200 million active users. Within this overarching framework, people are connected to other people, who are connected to other people, and so on, and underlying communities develop based on other parameters such as age, geographic location, interests, or lifestyle indicators. As people are removed from an extended social network, or their other parameters change, their associated communities are also transformed—constantly evolving the network landscape.

Analysts study this information to:

- Detect communities and predict community structures
- Better understand group dynamics
- Understand paths between two entities, such as how—or how closely—they are related or how the path between them shortens and lengthens based on changes in their communities
- Identify anomalies, such as when an entity changes communities
- Look for centrality, or entities in a network that at first do not appear critical, but without which the network breaks down

Developing STING algorithms for use with the multithreaded computing architecture provided by the Cray XMT will allow analysts to use a more massive, steady stream of dynamic, unstructured data, and conduct more highly complex analyses than traditionally possible using conventional approaches.

How we do it

Currently, researchers are designing and implementing a new complex multithreaded algorithm for use on the Cray XMT supercomputer based on community structure identification, and incorporating advanced algorithmic features to process dynamic spatio-temporal graphs. The Cray XMT architecture has the unique ability to process massive volumes of irregular, data-intensive applications, and is ideal for graphs—data that connects to other data in varying ways.

Researchers are continuing to scale computing mechanisms to meet the supercomputer's processing capability, and developing methodologies to understand its performance under the new parameters and identify memory bottlenecks that arise in the system during testing. The ability to scale existing approaches to analyze millions to billions of entities will enable much more complex analyses not only of social networks—which are the primary test case for this research—but also of other real world applications including electric power grids and biological sciences.

Applications

Current Applications

- Understanding relationships and social networks
- Understanding community structure and formation
- Collaboration on challenges with the electric power grid
- Work in biological sciences, such as understanding protein interaction networks

Future Applications

- Answering more dynamic social network-related queries
- Scale to massive problem sets using larger supercomputers

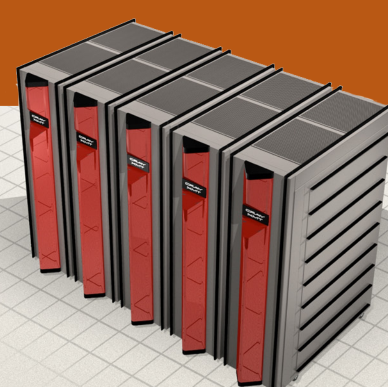
CASS-MT is dedicated to research on systems software, programming environments, and applications in a High-Performance Computing (HPC) multithreaded architecture environment.

We offer the only Open Science Cray XMT system, a one-of-a-kind supercomputer consisting of 128 multithreaded processors, 1 TB RAM, and a 7.7 TB Lustre parallel filesystem.

The Cray XMT supercomputer has the potential to substantially accelerate data analysis and predictive analytics beyond the limitations of traditional computing. Multithreaded processors allow multiple, simultaneous processing, helping researchers find solutions to the world's most complex challenges faster. The XMT can process irregular, data-intensive applications that have random memory access patterns. Unlike many applications where data delivery is dependent on memory speed, the Cray XMT's multi-threaded architecture tolerates memory access latencies by switching context between multiple threads that work continuously, overlapping the memory latency and preventing the processor from being held up while it waits for data to arrive.

The multithreaded technology powering our Cray XMT is ideally suited to perform pattern matching, scenario development, behavioral prediction, anomaly identification, and graph analysis.

Try it for yourself. We seek to create collaborations and provide expertise for porting and optimizing applications. The opportunity to use our Cray XMT system is available to internal and external research partners.



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

David A. Bader
Task Lead
Georgia Tech
http://www.cc.gatech.edu/~bader

Georgia Tech College of Computing

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

June 2009
PNNL-SA-66759