

Proudly Operated by **Battelle** Since 1965

"Pacific Northwest National Laboratory's current research and development agenda in dynamic network analysis serves its scientific mission in strengthening U.S. foundations for innovation. An important goal of this research work is to predict and engineer the behavior of complex systems to solve energy, environment and security challenges."

 Pacific Northwest National Laboratory Task Lead George Chin Jr.



Dynamic Network Analysis

Researching, developing, and evaluating dynamic network algorithms on multithreaded architectures

At a glance

At Pacific Northwest National Laboratory (PNNL), CASS-MT researchers are developing advanced approaches and methods for efficiently analyzing large-scale dynamic networks through the development and optimization of dynamic network algorithms running on multithreaded architectures. Dynamic networks represent evolving processes and systems composed of dynamic components such as nodes that appear and disappear, links that break and reconnect, network loads and bandwidths that fluctuate, and services or capabilities that move across nodes.



A distributed denial of service attack pattern may be decomposed into a set of precursor events or subpatterns and placed into a subgraph join tree for monitoring and tracking.

What we do

Dynamic network analysis (DNA) draws from and extends concepts, models, and techniques from a wide range of traditional network analysis areas including social network analysis, graph theory, and multi-agent systems. It often involves large-scale time-dependent data and the simultaneous analyses of multiple networks. In DNA, nodes are regularly treated as probabilistic with the ability to change, adapt, and learn over time. DNA studies the dynamic behavior of entities within and across networks as well as the evolution of entire networks.

DNA applications generally represent dynamic networks using dynamic graph data structures. They are likely to be irregular applications with random memory access patterns that exhibit poor temporal and spatial locality. Furthermore, DNA applications often generate dynamic networks from high-throughput streaming data arriving from near real-time data sources. Consequently, the original data and the associated network may be terabytes or more in size for particular problems and domain areas.

How we do it

Currently, we are developing and evaluating important classes of DNA algorithms on multithreaded architectures to tackle large-scale, dynamic scientific network problems as well as test the capabilities and capacities of multithreaded systems. Specific interests are in dynamic Bayesian networks and subgraph pattern matching or isomorphism.

A Bayesian network (BN) is a representation of a joint probability distribution over a set of random variables. A dynamic Bayesian network (DBN) represents sequences of variables, which are often time-series or symbol sequences. It is mainly used to model the stochastic evolution of a set of random variables over time. We are optimizing BN algorithms to support the evolving, dynamic nature of DBNs, while reducing their computational complexities and improving their performances on multithreaded architectures.

In recent efforts, we have been adapting our DBN algorithms to inference on large-scale probabilistic attack graph models used for computer network vulnerability analysis. Given a computer network configuration with a list of vulnerabilities for each node, one can generate an attack graph that models knowledge about how multiple vulnerabilities across machines may be combined into attack vectors. Such an attack graph may be further transformed into a probabilistic attack graph DBN by converting vulnerability scores (e.g., from Common Vulnerability Scoring System) into probabilities that may be assigned to exploits appearing in the attack graph. A probabilistic attack graph DBN would be polynomial in size with respect to the number of machines in the network, and thus, requires scalable DBN algorithms for inferencing.

Identifying emerging subnetwork patterns within massive networks is another common data analytics problem. We are developing advanced graph algorithms and a network analysis framework whereby an analyst may detect and identify precursor events and patterns as they emerge in complex networks. The framework is intended to be used in a dynamic environment where network data is streamed in and is represented as a largescale evolving dynamic graph. Specific graphical query patterns are collected in a library and are continuously and efficiently matched against the dynamic graph as it is updated. Each graph query is captured as a subgraph join tree which decomposes the query graph into smaller search subpatterns. These smaller subpatterns signify precursor events that emerge early before the full query pattern is complete.

We are applying our emerging subgraph pattern algorithms to computer netflow data to identify emerging computer network intrusions and threats. As shown in the above figure, a cyber attack such as a distributed denial of service attack may be described and represented as a graphical pattern. This pattern may be decomposed into smaller subpatterns that signify precursor events and then placed into a subgraph join tree. As the precursor events are detected in data streams, they are matched to the nodes of different subgraph join trees. Matching that occurs higher in a join tree indicates a higher probability that a specific type of attack is occurring. By identifying and tracking precursor events, one can mitigate or act upon the threat before it is fully realized.

Applications

- Threat detection in computer networks
- Distributed sensor networks
- Link analysis networks
- Geospatial networks and models
- Electric power grids

CASS is studying challenging irregular problems in search, knowledge discovery, cybersecurity, complex network, and natural language understanding. It is driving development of next generation software platforms, programming models, runtime systems, and high-performance computing systems that support global shared memory, hardware multi-threading, and fine-grain synchronization.

The Center manages a variety of computer systems with the potential to substantially accelerate data analysis and predictive analytics beyond the limitations of traditional computing. Our systems allow multiple, simultaneous processing, helping researchers find solutions to the world's most complex challenges faster. For example, our 128 processor, 2 TB Cray XMT can process at scale irregular, data-intensive applications that have random memory access patterns. 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.

We provide user accounts on all our systems, and can help you port and optimize your application. We seek collaborations in all our research areas of interest, and look forward to working with internal and external research partners.

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

George Chin Jr. Task Lead High-Performance Computing 509-375-2663 george.chin@pnnl.gov http://cass-mt.pnnl.gov/research/ dynamicnetwork.aspx



Proudly Operated by Battelle Since 1965