Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

STINGER Optimizations for High-performance Computing Platforms

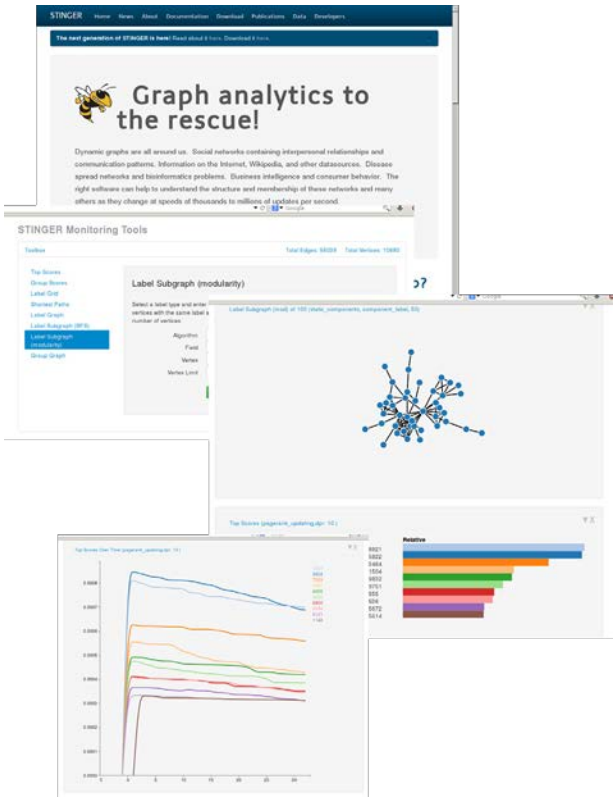
CHALLENGE

Streaming graph data represent the changing relationships in computer networks and data transfers. Detecting unusual behavior in these rich data sets is a common requirement for a variety of cyber analytic applications. Georgia Tech's framework and data structure for streaming graph analysis, Spatio-Temporal Interaction Networks and Graphs Extensible Representation (STINGER), is a successful research vehicle for developing streaming graph algorithms. STINGER is the basis for the first high-performance community and seed set expansion maintenance algorithms, the fastest streaming connected components and triangle counting algorithms, and very low-latency updated PageRank. STINGER's architecture allows keeping one massive, updating graph in memory while analysis kernels and users attach as separate processes, ensuring that kernels in development do not disrupt the graph data structure. Currently, STINGER supports tens of billions of edges and at least 50 simultaneous analysis kernels with low latency response times. However, STINGER's current internal architecture does not integrate well with many high-performance computing (HPC) and high-performance data analytics toolkits and workflows.

CURRENT PRACTICE

Widely available graph analysis systems on HPC platforms rely on explicit message passing through message passing interface (MPI). This entails message packing overheads and synchronization delays, limiting the ability of these

Improving the performance and ease of using STINGER opens rich, new capabilities to prevent network attacks, stop illicit data transfers, or identify disease epidemics.



STINGER provides high-performance analysis tools for streaming graph analysis in cyber analytics.

systems to respond quickly to changing situations, but permitting horizontal size scalability. Recent and upcoming memory technologies greatly increase single-system memory capacity and reduce the need for scaling through explicit message passing. The availability of accelerators with specialized memory systems, such as graphics processing units (GPUs) with high-bandwidth memory technologies, shows promise for rapid response, but they only support an order-of-magnitude-less graph storage capacity. Almost no available graph frameworks for accelerators or distributed memory support the streaming graph requirements of cyber analytics.

TECHNICAL APPROACH

With care, partitioned global address space (PGAS) abstractions can map across distributed systems, as well as across multiple in-node accelerators while also optimizing for current single-node memory architectures. STINGER relies on C structures and explicit pointers, making interfaces with partitioned memory painful. The first step is altering the core data structures to use more uniformly blocked arrays with explicit indices. Pointers are indices into a "global memory" array and are difficult to analyze and optimize. Explicit array indices are easier to optimize because they are bound to the specific data structures' partitioning and layout. A PGAS architecture could support quickly written, one-off queries, along with carefully optimized, long-running distributed queries.

Array layout poses interesting questions regarding data locality. Vector-like architectures, including typical GPUs and cell updates per second (CUPS), want similar data (neighboring vertices) packed together to optimize memory read bandwidth. The graph updating process would need to scatter incoming data carefully, reducing write bandwidth utilization. Fast, high-density non-volatile memory also requires careful consideration. Writes must be ordered properly to prevent the "tearing" that occurs when only part of the data is updated before a failure.

IMPACT

Rapid analysis of streaming data benefits growing cyber-analytic uses in computer security, along with more traditional uses in biological and medical informatics. Improving the performance and ease of using STINGER opens rich, new capabilities for analysts looking to prevent network attacks, stop illicit data transfers, or identify disease epidemics.

In the short-term, every aspect of the STINGER framework will become more flexible. The system will map onto different, novel HPC architectures more easily. Deployments can choose between horizontal scalability for utterly massive graphs across, or low-latency response for *in situ* applications. More uniform STINGER structures also will improve ease of use through simpler binding to different programming and analysis systems. Interquery communication can use the same infrastructure, leading to rapid response pipelines built for even more simultaneous analyses.

Contacts

Jason Riedy
Principal Investigator
(404) 385-4075
jason.riedy@cc.gatech.edu

David A. Bader
(404) 385-4785
bader@cc.gatech.edu

John R. Johnson
Program Director
(509) 375-2651
John.Johnson@pnnl.gov

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Georgia Institute of Technology