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

Automata Processor

CHALLENGE

Many new cyber-analytic applications must identify known interaction patterns in rich Semantic data. This requires finding labeled subgraphs in massive graphs, the problem of subgraph matching, which is known to be computationally complex for von Neumann architectures. Therefore, solutions on existing von Neumann architectures, such as central processing units (CPUs), do not scale to massive graphs. The objective of this project is to explore methods for solving this difficult problem using a newly developed non-von Neumann architecture, the Micron Automata Processor (AP). The AP can approach some computationally intractable problems more efficiently than classical processors.

CURRENT PRACTICE

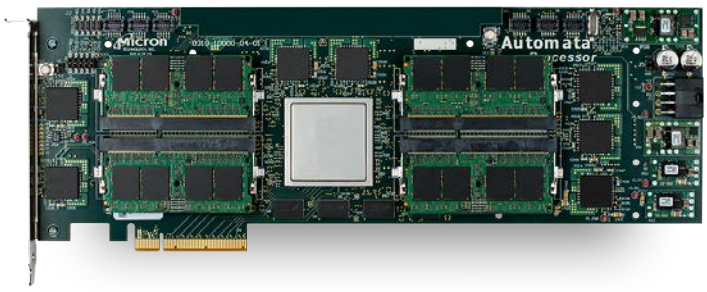
Existing solutions for the subgraph matching problem are built around batch processing on traditional CPUs. However, the problem is widely believed to be intractable, and the solutions do not scale well to big graphs. Some current approaches can handle massive graphs using large multi-core systems, but they require long indexing and search times. A non-von Neumann architectures, like the AP, may dramatically accelerate these methods and result in the development of a new tool.

TECHNICAL APPROACH

The AP can be programmed to compute a large set of user-defined nondeterministic finite automata (NFA) in parallel against a single data stream. One of the key insights of our approach is to represent all of the nodes and edges in the given graphs using strings that can be streamed to the processor. Then, the graph problem is transformed into a string-pattern-matching problem, which can be solved by computing NFA over streaming data.

Exploring methods to identify known interaction patterns in rich Semantic data.

As part of our proposed solution, for each motif, we identify incrementally bigger neighborhoods in the motif with respect to an anchor node. We refer to the neighborhoods as “motif-neighborhoods.” Then, our algorithm proceeds in iterations to identify nodes in the query graph, which can be used to get subgraphs of increasingly larger sizes that are identical in topology to



Automata Processor development board.

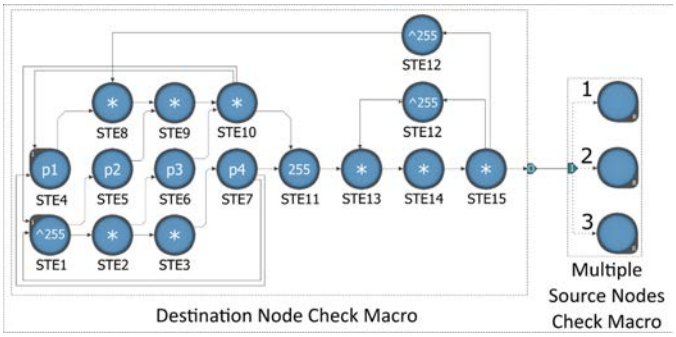
the corresponding motif-neighborhoods to be matched. In each iteration, all of the subgraphs of the query graph with identical topology as motif-neighborhoods of a particular size are programmed on the AP as automata, and all of the edges in the query graph are streamed to identify the subgraphs identical to the motif-neighborhoods with one more node. The iterations continue until the complete motif-neighborhoods have been explored. The subgraphs reported at the end of the iterations are reported as solutions.

The acceleration of the solution stems from the identification of all the edges that extend a subgraph in parallel. In addition, the stream-based solution does not suffer from poor data locality typical of graph applications. Further, the AP provides a greatly simplified execution model without the need to handle communication delays or race conditions common in other contemporary processors.

The ultimate aim of this project is to refine the solution proposed and developed new algorithms for subgraph matching using the AP. We will evaluate the algorithms for determining current performance and expected advantages based on projected future AP hardware improvements.

IMPACT

Accelerating subgraph matching would immediately benefit cyber-analytic applications. Rapidly finding patterns representative of illicit activity in computers and financial transactions could enable new analysis and monitoring algorithms. Extending these results to streaming graphs could provide rapid response to real-world changing conditions.



Automaton to extend streamed paths by chosen edges.

Contacts

Srinivas Aluru
Principal Investigator
(404) 385-1486
aluru@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