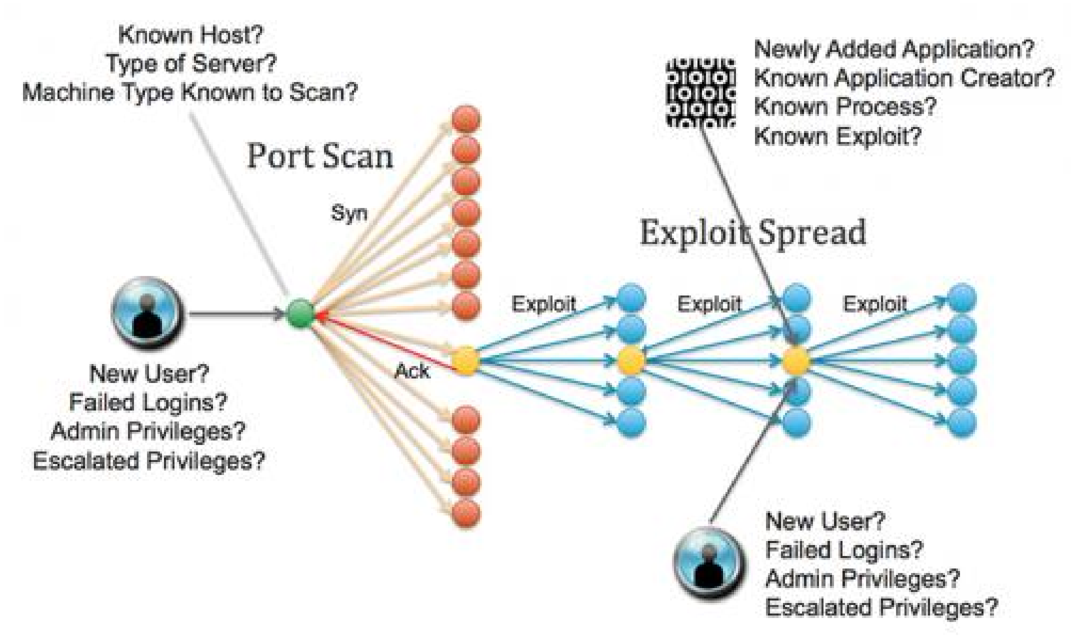
Novel graph method detects cyber-attack patterns in complex computing networks

March 27 2015



Example patterns searched by StreamWorks.

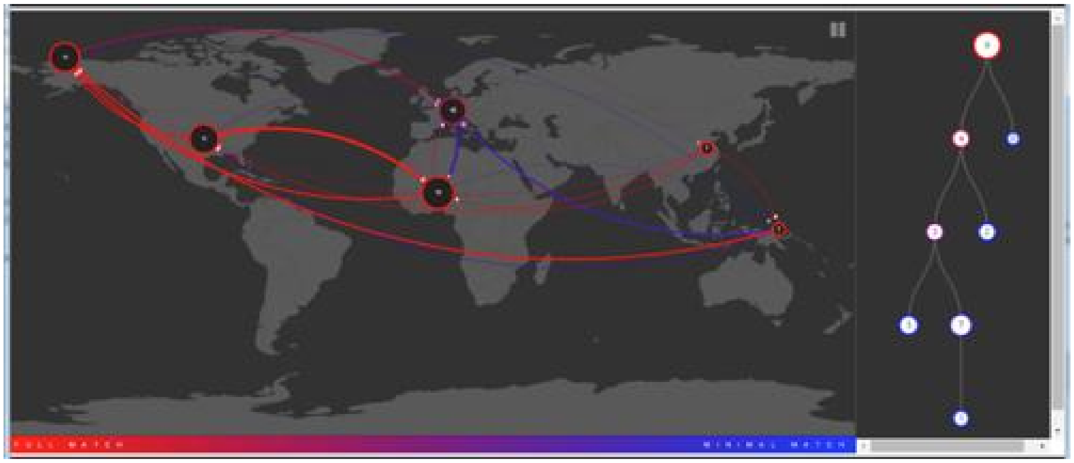
As the perils of cyber security breaches continue to plague industries, governments, and citizens throughout the world, the need to detect these infiltrating events, as well as identify their attack patterns, in complex computing networks as they emerge in real time remains a paramount concern and growing challenge. In their work involving streaming graphs, scientists at Pacific Northwest National Laboratory and Washington State University, devised a novel framework called StreamWorks that categorizes cyber attacks as graph patterns, which then can be examined using a continuous search (query) on a single, large streaming dynamic graph. "Continuous Query" focuses on finding matches for queries in a data stream as soon as they happen, which is in contrast to ad hoc querying supported by databases such as MySQL or Neo4J that aim to efficiently query a large, non-changing data set.

Using two real-world data sets—an online news stream from The New York Times and Internet network traffic from the Center for Applied Internet Data Analysis—their method produced efficient continuous queries on dynamic graphs with speedups up to 100x greater than current methods that do not support real-time subgraph pattern matching.

As the volume and throughput of network traffic or data sets rise exponentially, the inability to detect adversarial actions in real-time provides an asymmetric advantage to attackers. Meanwhile, the need to protect intellectual properties and customer data only continues to grow in priority for corporations and government agencies worldwide. Identifying precursor events and patterns as they emerge will go a long way in evading and mitigating the computer network intrusions and threats that have potentially criminal, even dangerous, consequences and have made cyber security a multi-billion dollar industry.

"In the high-stakes cyber security domain, processing streaming updates to a dynamic graph database accumulating multiple data sources, such as network flow firewall logs, is important for realizing real-time situational awareness," said Sutanay Choudhury, a computer scientist with PNNL's Data Sciences group and the paper's lead author.

For their work, the researchers needed to register patterns in data sets as a graph query and continuously perform said query on the graph as it evolved over time. Notably, the graphs are heterogeneous, or multi-relational, meaning a single data stream can have diverse characteristics. For example, in graph terms, this would mean different edge types for cyber security and varied node and edge types in social media. Rather than looking for matches within an entire graph, the selective search method exploits the data's heterogeneity by searching smaller subgraphs ordered by their selectivity, obtained using the single-edge level and 2-edge path distribution from the graph stream. The resulting decomposition is stored in a data structure, or subgraph join tree (SJ-tree), that tracks matching subgraphs and combines them to produce larger matches.



A map-based visualization interface to associate traffic patterns with origins and destination.

From there, the join order provides a heuristic technique for assembling matches without seeking every possible way to do so, thereby curtailing some of the cost associated with more broad searches. The team's "lazy search" graph algorithm, triggered only if a new edge is the most selective subgraph in the query or if one of the vertices in that edge participates in a match with the preceding subgraph in the join order, also limits the amount of partial matches that are tracked to subgraphs.

The researchers used the online news stream and network traffic from CAIDA, as well as a synthetic social media stream, data sets to examine how selectivity distribution of subgraphs played out. They compared the performance of multiple combinations of query decomposition execution methods. They also provided a "Relative Selectivity" rule for engaging an optimum search strategy. Their efforts showed that the distribution of 2-edge subgraphs is heavily skewed and a query decomposition strategy that exploits this skew using the "lazy search" algorithm will be consistently efficient.

The subgraph-selectivity-based approach proposed in this work is an initial start for tackling the problem of continuous pattern detection. More investigation is needed to provide accurate selectivity estimation. The team actively is working on a distributed implementation of the Continuous Query engine, which will afford the capability to scale in an HPC cluster. Additional research efforts can leverage work on query performance modeling and approximate algorithms. Supporting expressive, rich queries requires computations with high time and space complexity. A continuous stream processing system cannot afford to spawn computations that exceed a time or memory bound. Predicting query performance from the knowledge of stream statistics and using approximate algorithms provide a natural path toward that goal.

More information: Choudhury S, L Holder, G Chin Jr, K Agarwal, and JT Feo. 2015. "A Selectivity-based Approach to Continuous Pattern Detection in Streaming Graphs." In 18th International Conference on Extending Database Technology (ICDT 2015). March 23-27, 2015, Brussels, Belgium.

Provided by Pacific Northwest National Laboratory

Citation: Novel graph method detects cyber-attack patterns in complex computing networks (2015, March 27) retrieved 19 April 2024 from https://phys.org/news/2015-03-graph-method-cyber-attack-patterns-complex.html