Visual Analytics Sandbox: A big data platform for processing network traffic

Raju Gottumukkala, Ph.D.
Director of Research, Informatics Research Institute
Site Director, NSF Center for Visual and Decision Informatics
Assistant Professor, College of Engineering
University of Louisiana at Lafayette

2017 Internet2 Global Summit (04/26/2017)

Funded by NSF award No.1429526
Motivation

- Cyber environment is increasingly getting complex
- Existing tools do not support interactive analysis of dynamic graphs

Web graph [©GW3BI]
Ip Flow Graph [©CVDI]

- Graphs are complex data stores
**Application of TVG in cyber-security**

- Extraction of Traffic Dispersion Graph (TDG) from IP-Flows
- Graph structural properties indicate abnormal traffic patterns (malware)

<table>
<thead>
<tr>
<th>TDG Properties</th>
<th>Malware Indication (TDG Class)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average In-Degree &gt; 2.8</td>
<td>P2P traffic</td>
</tr>
<tr>
<td>In &amp; Out Edges &gt; 1</td>
<td></td>
</tr>
<tr>
<td>In &amp; Out &gt; 1</td>
<td>P2P traffic (BitTorrent)</td>
</tr>
<tr>
<td>Diameter &gt; 11</td>
<td></td>
</tr>
<tr>
<td>Sink Vertices &gt; 95%</td>
<td>MyDoom Malware</td>
</tr>
<tr>
<td>GWCC &lt; 5</td>
<td></td>
</tr>
<tr>
<td>Sink Vertices &gt; 95%</td>
<td>Bobax Malware</td>
</tr>
<tr>
<td>Average Out Degree &gt; 150</td>
<td></td>
</tr>
<tr>
<td>Sink Vertices &gt; 95%</td>
<td>Slammer Malware</td>
</tr>
<tr>
<td>Average Out Degree 50 to 150</td>
<td></td>
</tr>
<tr>
<td>Otherwise</td>
<td>Normal Traffic (HTTP, FTP, etc.)</td>
</tr>
</tbody>
</table>

A TDG for soribada P2P
What is Visual Analytics Sandbox?

- A unique analytics environment for processing high-volume, high-velocity data streams
  - IoT, IP flow graphs, click streams, social media, etc.
- Experimental infrastructure to develop next-generation decision support tools
Hardware

• 112 T Flops of computing
  – 22 Intel Xeon E5 processors (308 cores)
  – 12 NvDIA P100 GPU’s

• Storage
  – 1.8 TB RAM
  – 25.6 TB NvME SSD
    • 25X faster than HDD, 5X faster than SSD
  – 20 TB HDD

• Networking
  – 2X10G Ethernet Cards per Node
Benchmarking Graph Operations

<table>
<thead>
<tr>
<th>Type of Temporal Characteristic</th>
<th>Graph Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal network topology</td>
<td>Degree, connectivity, density</td>
</tr>
<tr>
<td>Reachability analysis</td>
<td>Paths, walks, trails</td>
</tr>
<tr>
<td>Detecting outliers</td>
<td>Node or edge clustering</td>
</tr>
<tr>
<td>Node neighborhoods</td>
<td>Persistent patterns &amp; motifs</td>
</tr>
</tbody>
</table>

Execution times for retrieving cumulative edge weights of a node (Neo4J vs MySQL)

Figure 9. Execution times for retrieving shortest path between two selected nodes (Neo4J schemes)

Execution times for retrieving weighted shortest path between two nodes (Neo4J with timestamp on node and caching vs no caching)
Touch interactions

(a) Node history graphs with timestamp browsing
(b) Graph filtering based on edge weight with multitouch controller (lower-bound highlighted yellow).
(c) Control widget (around finger) and zoom window with detached source.
(d) Local neighborhood of a selected node arranged and rendered in detached window (upper-right).
(e) Progress indicator circle provides feedback to users while analytics jobs are processed.
(f) Shortest paths between nodes are emphasized in red, and nodes along the paths are highlighted on the graph.
Connectivity

- Connected to LONI Network and Science DMZ
- Connected to Internet 2.0 through GENI switch
- 40 Gbps connections for faster data transfer
Expectations & Timeline

• How can you use this system?
  – Virtual big data environment for stream processing, graph analytics
  – Leverage stream processing, graph mining & visualization tools that are part of the sandbox

• Plan for deployment
  – Software developed in a pilot environment for (a) intelligent levee surveillance, (b) event detection in social media and (c) influenza forecasting
  – Looking for collaboration for one use case in cybersecurity use case for network analysis (Aug 2017)
  – System will be made available for users on Internet 2 (Spring 2018)
Thank You

Contact: raju@louisiana.edu
Homepage: www.ucs.louisiana.edu/~nrg0821
Real Time Event Detection