NStreamAware: Real-Time Visual Analytics for Data Streams to Enhance Situational Awareness

Fabian Fischer and Daniel A. Keim
Motivation: Heterogeneous Data Streams

- Network Alerts (e.g., OSSEC)
- Syslog Messages
- NetFlow Data

Analyzing Data Streams = Crucial for security in your network!

Monitoring & Exploration
Crucial for situational awareness (SA)!
DATA CHALLENGE
How to make stream analysis scalable?
NStreamAware: Scalable Infrastructure

Apache Spark™ is a fast and general engine for large-scale data processing which can run on a distributed computer cluster.
Integrated Perspectives

- Real-Time Data Stream Monitoring
- Real-Time Sliding Slices (NVisAware)
- Visual Feature Selection
- Summarized Sliding Slices
- Event Timeline & Insights
- Search & Exploration
Real-Time Data Stream Monitoring
Demo
SITUATIONAL AWARENESS CHALLENGE

How to reduce the cognitive load?
NVisAware: Analytics

Visual Analytics Approach:
Calculate and visualize sliding slices.
(based on sliding windows)

- Calculate **Sliding Slice Summary** for each sliding window.
- **Push slice**$_t$ to web application.
Real-Time Sliding Slice

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type</th>
<th>Stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>#events</td>
<td>count</td>
<td>Syslog</td>
</tr>
<tr>
<td>timestamps</td>
<td>set</td>
<td>Syslog</td>
</tr>
<tr>
<td>#programs</td>
<td>count</td>
<td>Syslog</td>
</tr>
<tr>
<td>#hosts</td>
<td>count</td>
<td>Syslog</td>
</tr>
<tr>
<td>#frequentWords</td>
<td>count</td>
<td>Syslog</td>
</tr>
<tr>
<td>programs</td>
<td>key-value list</td>
<td>Syslog</td>
</tr>
<tr>
<td>hosts</td>
<td>key-value list</td>
<td>Syslog</td>
</tr>
<tr>
<td>frequentWords</td>
<td>key-value list</td>
<td>Syslog</td>
</tr>
<tr>
<td>newHosts</td>
<td>new-set</td>
<td>Syslog</td>
</tr>
<tr>
<td>newPrograms</td>
<td>new-set</td>
<td>Syslog</td>
</tr>
<tr>
<td>srcAddr</td>
<td>key-value list</td>
<td>NetFlow</td>
</tr>
<tr>
<td>dstAddr</td>
<td>key-value list</td>
<td>NetFlow</td>
</tr>
<tr>
<td>srcPorts</td>
<td>key-value list</td>
<td>NetFlow</td>
</tr>
<tr>
<td>dstPorts</td>
<td>key-value list</td>
<td>NetFlow</td>
</tr>
<tr>
<td>topTalker</td>
<td>key-array list</td>
<td>NetFlow</td>
</tr>
<tr>
<td>#srcAddr</td>
<td>count</td>
<td>NetFlow</td>
</tr>
<tr>
<td>#dstAddr</td>
<td>count</td>
<td>NetFlow</td>
</tr>
<tr>
<td>#srcPorts</td>
<td>count</td>
<td>NetFlow</td>
</tr>
<tr>
<td>#dstPorts</td>
<td>count</td>
<td>NetFlow</td>
</tr>
<tr>
<td>ossecAlerts</td>
<td>key-value list</td>
<td>OSSEC</td>
</tr>
</tbody>
</table>

- **Interactive Widgets**
  - Treemaps
  - Counters
  - Node-link diagrams

- **Interactions**
  - Star/Annotate slice
  - Remove slice
  - Retrieve data

- **Color Encoding**
  - Background for similarity
  - Importance of alerts
Demo
EXPLORATION CHALLENGE

How to explore many sliding slices?
Visual Feature Selection

**Visual Analytics Approach:**
Aggregate / Summarize according interest function (visually steered by the expert)
Example: Using Visual Analytics for Interactive Summarization
Demo
Application to Real-Time Social Media Analysis (VAST Challenge 2014 MC3)

- **Real-Time Monitoring Task:**
  Discover major events in the stream to support an ongoing police operation.

- **Available Data Stream:**
  Real-time feeds of microblogs and emergency calls.

- **Successful participation:**
  "Award for Outstanding Comprehensive Mini-Challenge 3 Submission"
Further Challenges and Future Work

• **Challenge:** Parameter adjustment for sliding slices and clustering.

• **Automated merging of sliding slices** based on the interest function.

• **Performance Evaluation for a large network** using security operational data stream.

• **Responsiveness issues** when increasing the number of complex interactive visualizations.

• **Data retention and rotation** for the visualization interface.
Contributions

DATA CHALLENGE
How to make stream analysis scalable?

• **NStreamAware** – Building a *web-based visual analytics system* using scalable technologies.

SA CHALLENGE
How to reduce the cognitive load?

• **NVisAware** – *Sliding Slices Visualization* with embedded visualization widgets.

EXPLORATION CHALLENGE
How to explore many sliding slices?

• **NVisAware** – *Summarized Sliding Slices* steered using interactive visualizations.
Thank you very much for your attention!

Questions?

For more information about this work please contact

Fabian Fischer
Tel. +49 7531 88-2780
Fabian.Fischer@uni-konstanz.de

http://ff.cx/