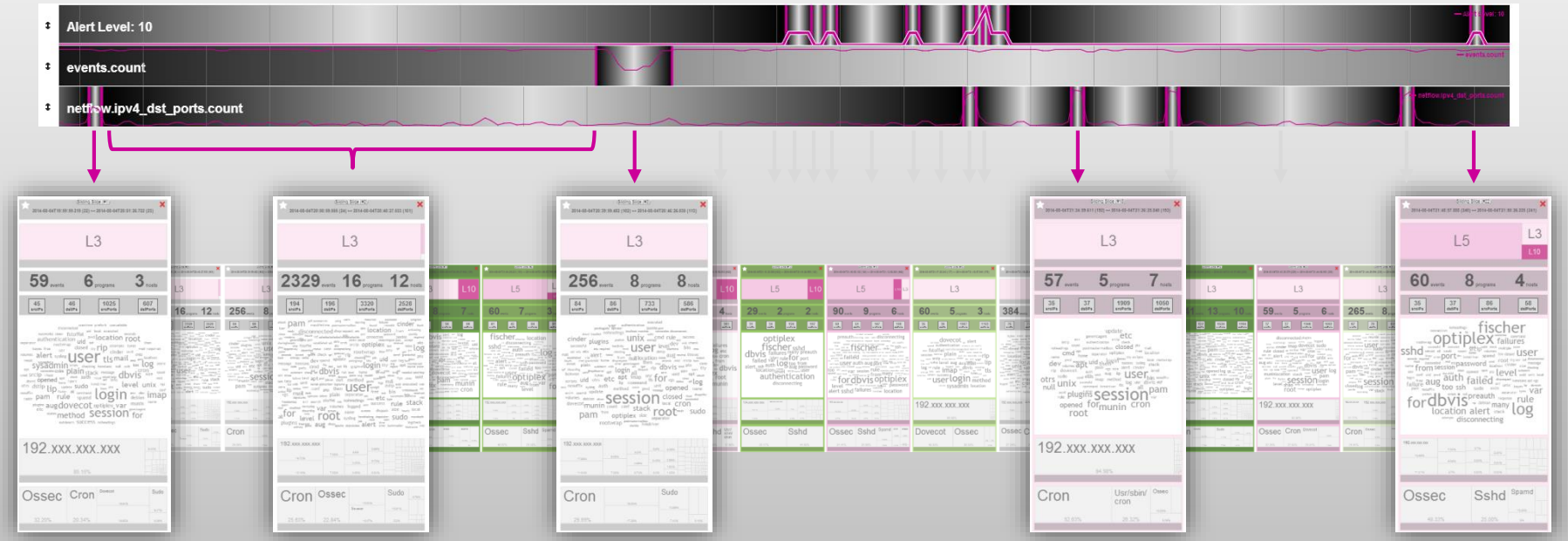
Symposium on Visualization for Cyber Security (VizSec 2014)
10th November 2014, Paris, France

Universität
Konstanz



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

Fabian Fischer and Daniel A. Keim

Fabian Fischer | Data Analysis and Visualization Group | University of Konstanz

Motivation: Heterogeneous Data Streams

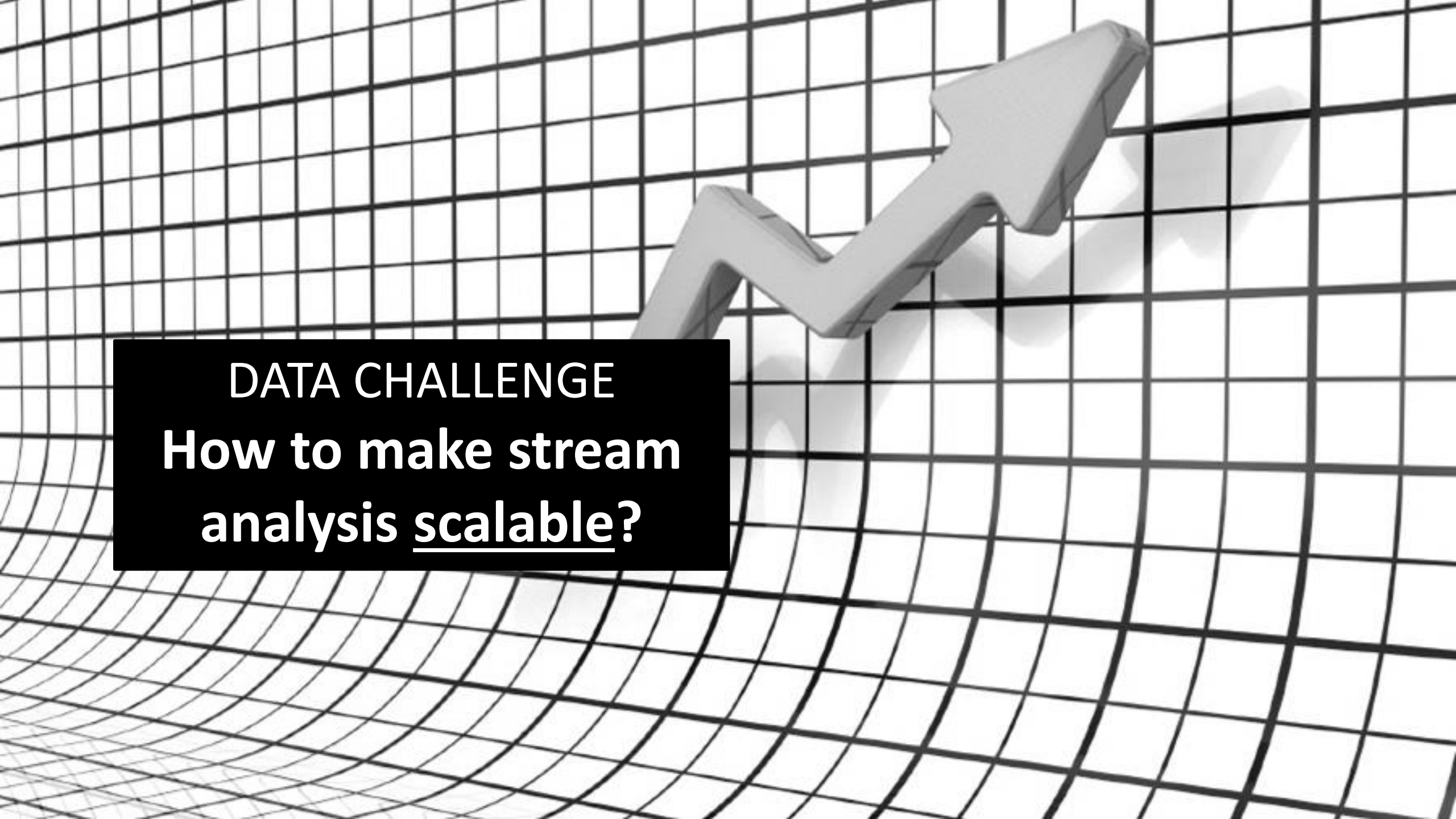
- **Network Alerts** (e.g., OSSEC)
- **Syslog Messages**
- **NetFlow Data**

Analyzing Data Streams =
Crucial for security in your network!

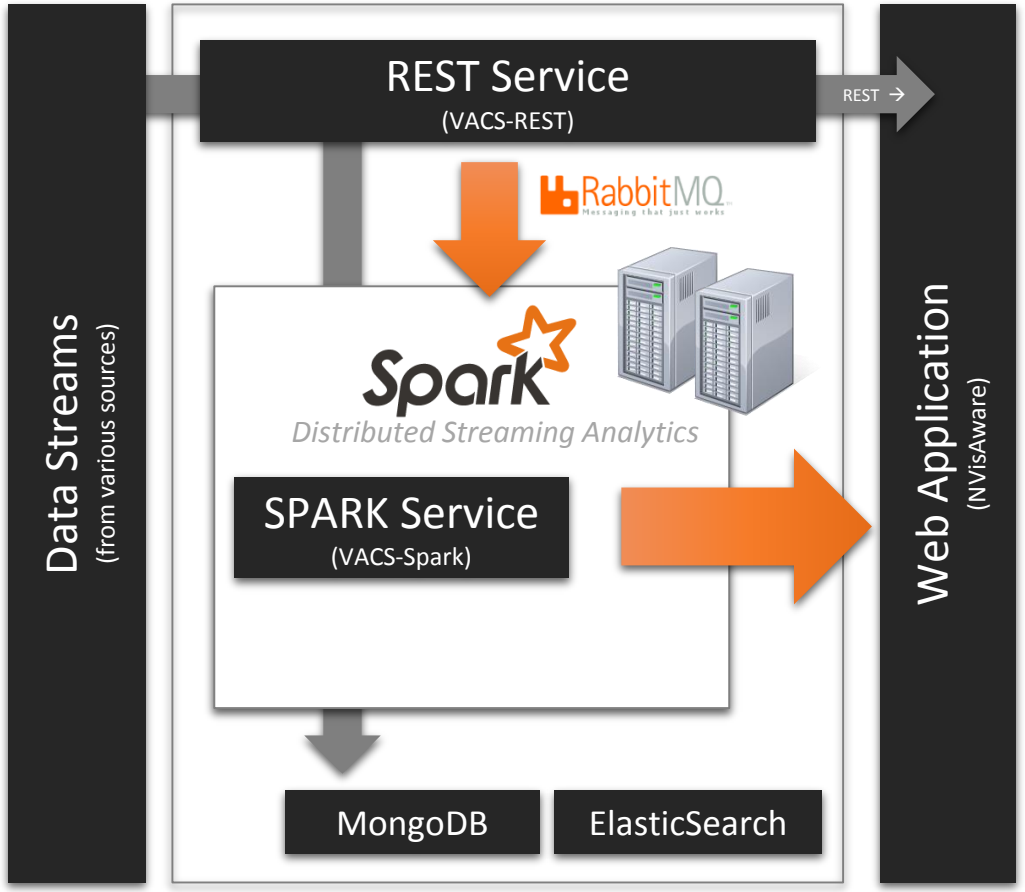
→ **REAL-TIME VISUAL ANALYTICS**

Monitoring & Exploration

Crucial for situational awareness (SA)!


DATA CHALLENGE
How to make stream analysis scalable?

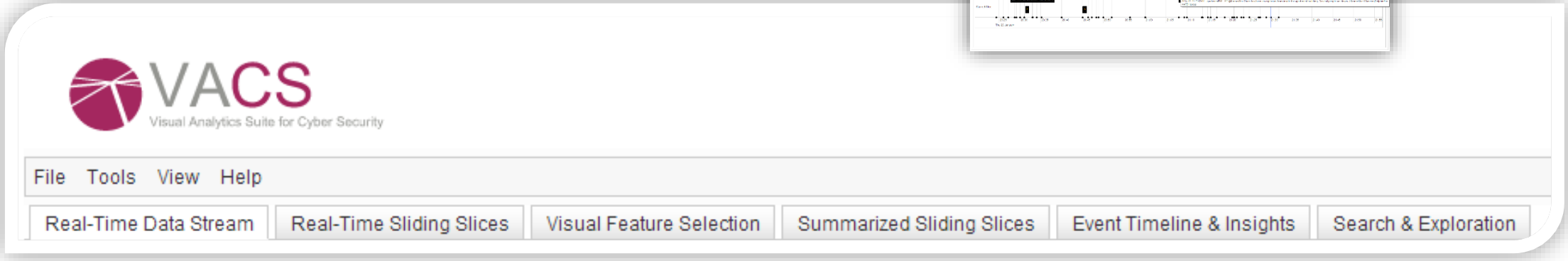
NStreamAware: Scalable Infrastructure



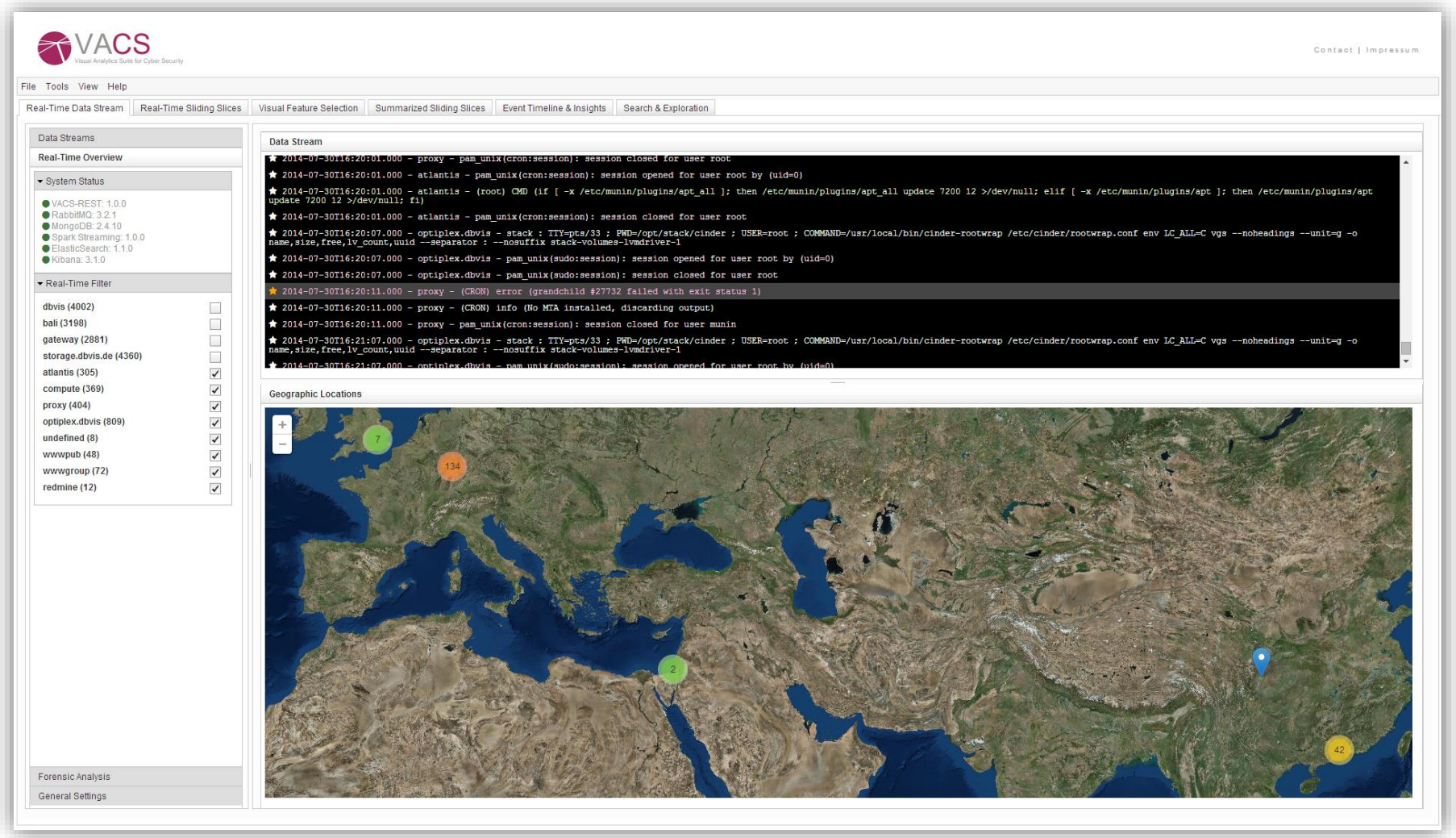
Apache Spark™ is a fast and general engine for large-scale data processing which can run on a distributed computer cluster.

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

VACS
Visual Analytics Suite for Cyber Security

File Tools View Help

Real-Time Data Stream | Real-Time Sliding Slices | Visual Feature Selection | Summarized Sliding Slices | Event Timeline & Insights | Search & Exploration

Data Streams

Real-Time Overview

System Status

- VACS-REST: 1.0.0
- VACS-Spark: 1.0.0
- RabbitMQ: 3.2.1
- MongoDB: 2.4.10
- Spark Streaming: 1.0.0
- ElasticSearch: 1.1.0
- Kibana: 3.1.0

Real-Time Filter

- gateway (7)
- dbvis (70)
- bali (52)
- optiplex.dbvis (28)
- wwwgroup (3)
- wwwpub (3)
- storage.dbvis.de (16)
- compute (6)
- proxy (16)
- atlantis (9)

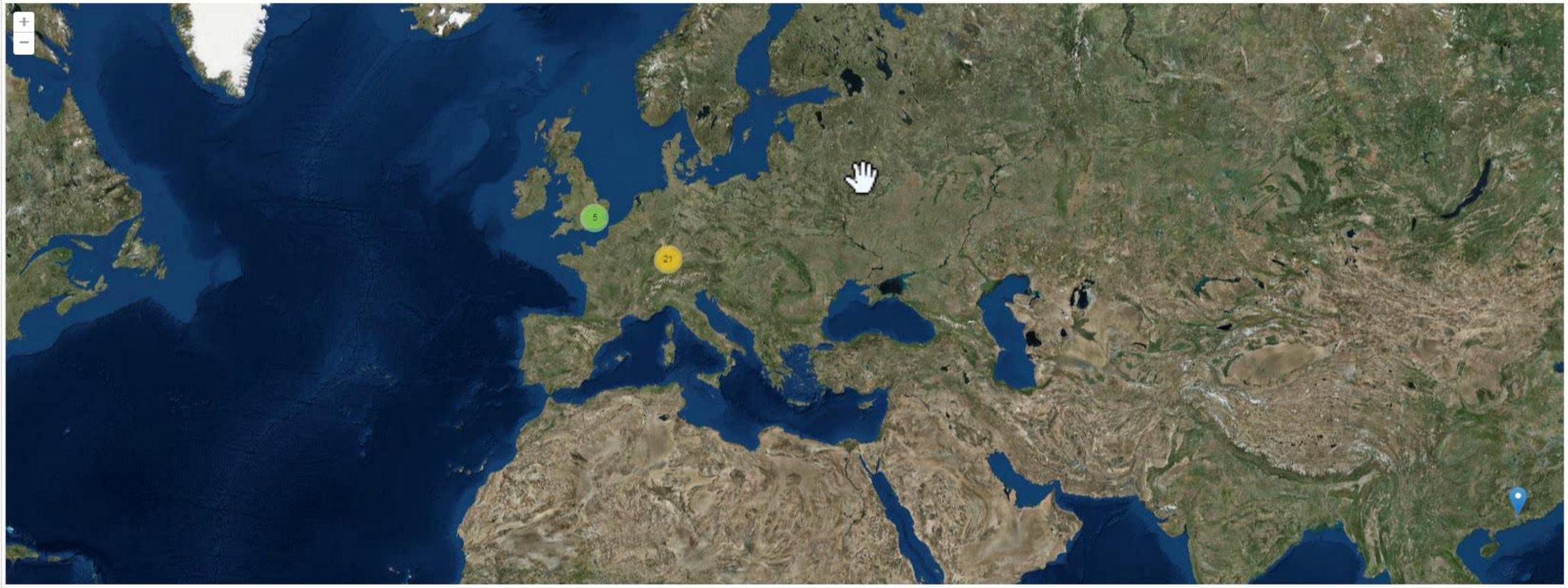
Forensic Analysis

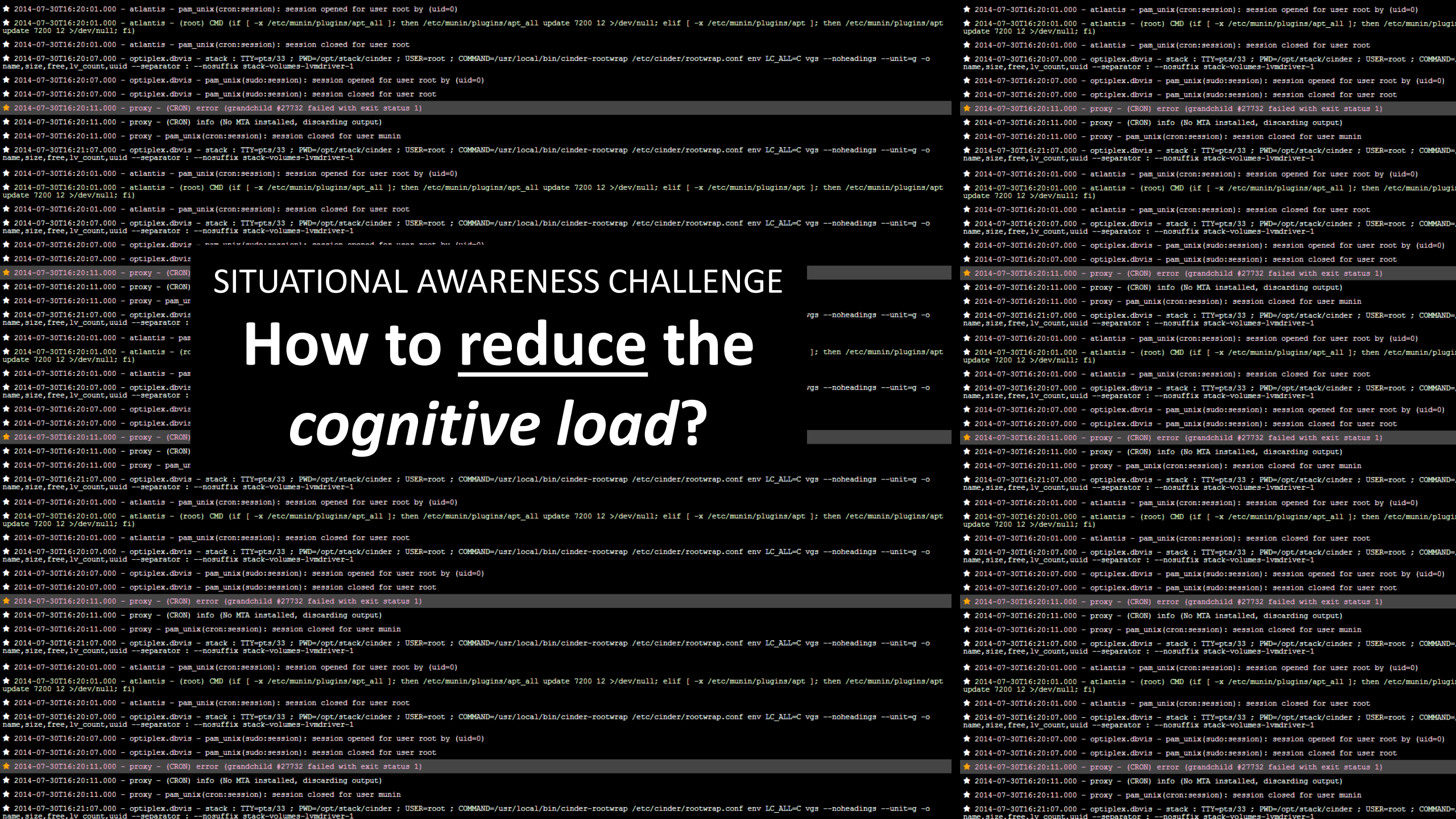
General Settings

Data Stream

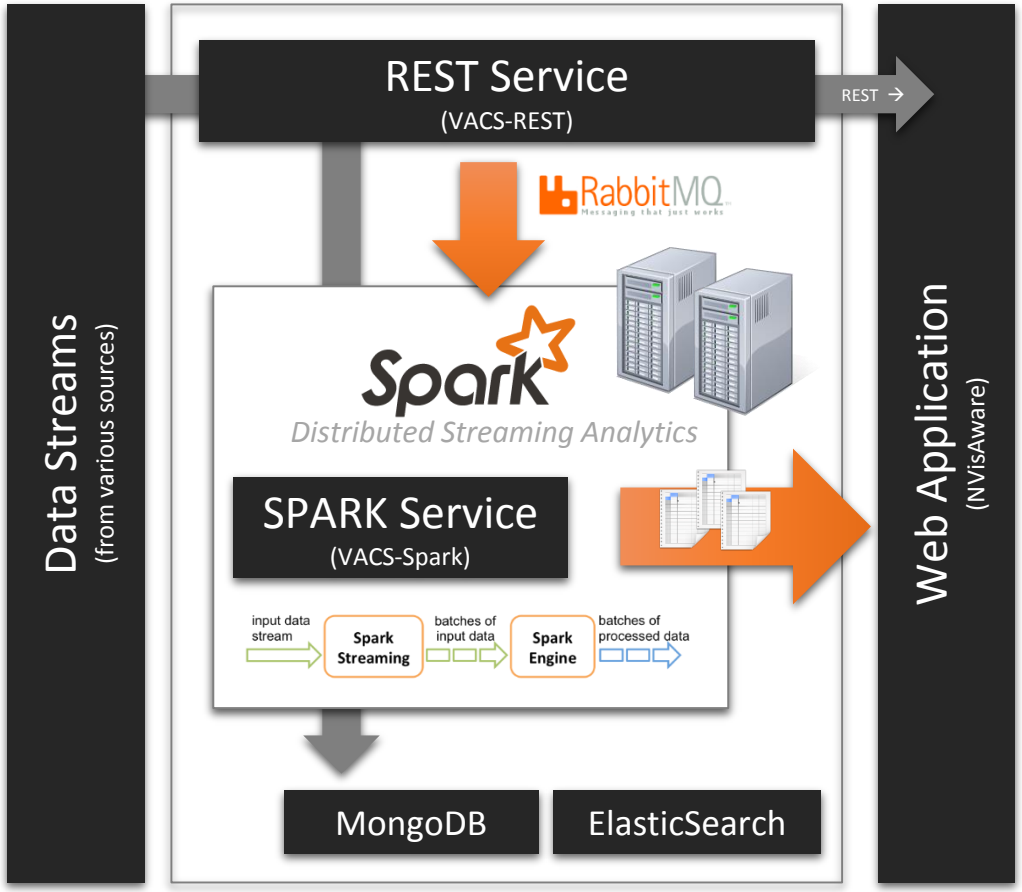
```
★ 2014-08-04T21:51:24.000 - bali - pam_unix(cron:session): session closed for user XXXX
★ 2014-08-04T21:46:01.000 - dbvis - pam_unix(cron:session): session opened for user XXXX by (uid=0)
★ 2014-08-04T21:46:01.000 - dbvis - (root) CMD (/root/scripts/check-dovecot.sh 2>&1 > /dev/null)
★ 2014-08-04T21:46:01.000 - dbvis - pam_unix(cron:session): session closed for user XXXX
★ 2014-08-04T21:46:22.000 - dbvis - IMAP(XXXXX): Disconnected for inactivity bytes=42/2782
★ 2014-08-04T21:46:22.000 - dbvis - IMAP(XXXXX): Disconnected for inactivity bytes=127/1253
★ 2014-08-04T21:46:27.000 - dbvis - imap-login: Login: user=<XXXXXX>, method=PLAIN, rip=95.XXX.XXX.XXX, lip=134.XXX.XXX.XXX, TLS
★ 2014-08-04T21:46:28.000 - dbvis - imap-login: Login: user=<XXXXXX>, method=PLAIN, rip=95.XXX.XXX.XXX, lip=134.XXX.XXX.XXX, TLS
★ 2014-08-04T21:51:58.000 - bali - Alert Level: 3; Rule:XX:XX:XX:XX user: jaeckle; Aug 4 21:46:XX:XX:XX:XX method=PLAIN, rip=95.XXX.XXX.XXX, lip=134.XXX.XXX.XXX, TLS
★ 2014-08-04T21:51:58.000 - bali - Alert Level: 3; Rule:XX:XX:XX:XX user: jaeckle; Aug 4 21:46:XX:XX:XX:XX method=PLAIN, rip=95.XXX.XXX.XXX, lip=134.XXX.XXX.XXX, TLS
★ 2014-08-04T21:46:41.000 - optiplex.dbvis - stack : TTY=pts/33 ; PWD=/opt/stack/cinder ; USER=root ; COMMAND=/usr/local/bin/cinder-rootwrap /etc/cinder/rootwrap.conf env LC_ALL=C vgs --noheadings --unit=g -o name,size,free,lv_count,uuid --separator : --nosuffix stack-volumes-lvmdriver-1
★ 2014-08-04T21:46:41.000 - optiplex.dbvis - pam_unix(sudo:session): session opened for user XXXX by (uid=0)
★ 2014-08-04T21:46:42.000 - optiplex.dbvis - pam_unix(sudo:session): session closed for user XXXX
★ 2014-08-04T21:52:20.000 - bali - Alert Level: 3; Rule:XX:XX:XX:XX optiplex.dbvis sudo: stack : TTY=pts/33 ; PWD=/opt/stack/cinder ; USER=root ; COMMAND=/usr/local/bin/cinder-rootwrap /etc/cinder/rootwrap.conf env LC_ALL=C vgs --noheadings --unit=g -o name,size,free,lv_count,uuid --separator : --nosuffix stack-volumes-
lvmdriver-1
```

Geographic Locations




SITUATIONAL AWARENESS CHALLENGE
How to reduce the cognitive load?

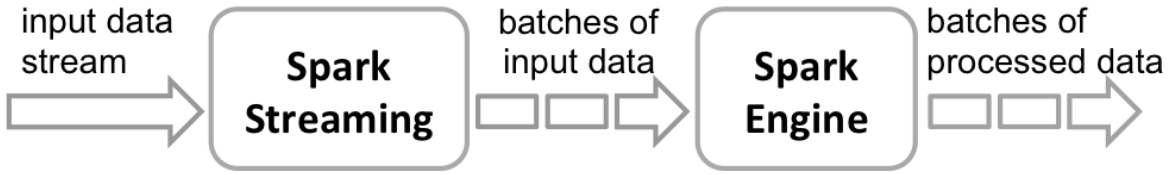
NVisAware: Analytics



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

Feature	Type	Stream
#events	count	Syslog
timestamps	set	Syslog
#programs	count	Syslog
#hosts	count	Syslog
#frequentWords	count	Syslog
programs	key-value list	Syslog
hosts	key-value list	Syslog
frequentWords	key-value list	Syslog
newHosts	new-set	Syslog
newPrograms	new-set	Syslog
srcAddr	key-value list	NetFlow
dstAddr	key-value list	NetFlow
srcPorts	key-value list	NetFlow
dstPorts	key-value list	NetFlow
topTalker	key-array list	NetFlow
#srcAddr	count	NetFlow
#dstAddr	count	NetFlow
#srcPorts	count	NetFlow
#dstPorts	count	NetFlow
ossecAlerts	key-value list	OSSEC

$slice_t$

- **Interactive Widgets**
  - Treemaps
  - Counters
  - Node-link diagrams
- **Interactions**
  - Star/Annotate slice
  - Remove slice
  - Retrieve data
- **Color Encoding**
  - Background for similarity
  - Importance of alerts

Demo

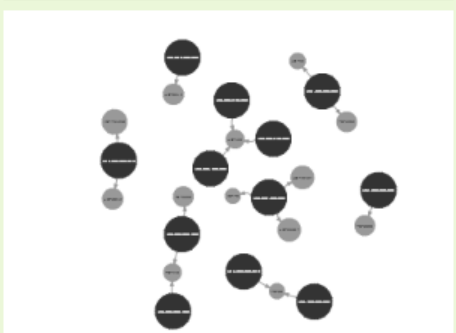
Sliding Slice (#151)
2014-08-04T21:04:29.221 ↔ 2014-08-04T21:05:27.572

L5 | L10 | L3

30 events | 3 programs | 3 hosts

32 srcIPs | 33 dstIPs | 52 srcPorts | 38 dstPorts

Ossec 50.00% | Sshd 40.00%



Sliding Slice (#190)
2014-08-04T21:23:59.420 ↔ 2014-08-04T21:24:57.605

L3 | L5

30 events | 7 programs | 3 hosts

29 srcIPs | 30 dstIPs | 125 srcPorts | 106 dstPorts

Ossec 33.33% | Sudo 20.00% | Cron

Sliding Slice (#192)
2014-08-04T21:24:59.611 ↔ 2014-08-04T21:25:55.979

L3

29 events | 5 programs | 3 hosts

31 srcIPs | 31 dstIPs | 1884 srcPorts | 1039 dstPorts

192.xxx.xxx.xxx
94.82%

Cron 48.28% | Ossec | Sudo

Sliding Slice (#200)
2014-08-04T21:28:59.192 ↔ 2014-08-04T21:29:57.368

L3

30 events | 5 programs | 7 hosts

25 srcIPs | 24 dstIPs | 50 srcPorts | 33 dstPorts

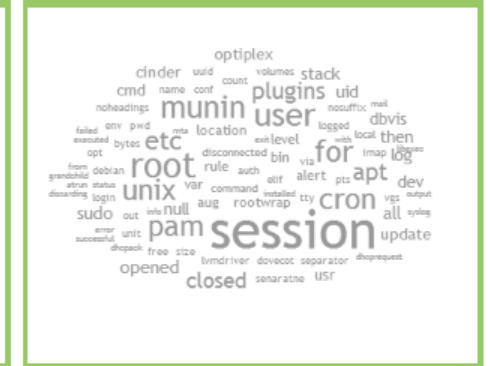
Cron 50.00% | Usr/sbin/cron 26.67% | Ossec | Sudo

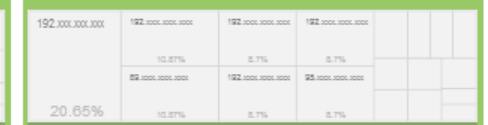
Sliding Slice (#201)
2014-08-04T21:29:29.239 ↔ 2014-08-04T21:30:27.451

L3

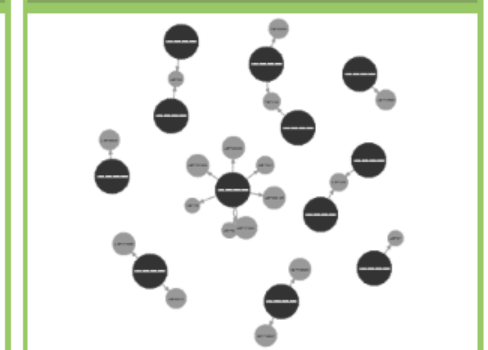
30 events | 6 programs | 8 hosts

18 srcIPs | 18 dstIPs | 46 srcPorts | 28 dstPorts





Cron 40.00% | Usr/sbin/cron 26.67% | Ossec | Dhcpd



VACS
Visual Analytics Suite for Cyber Security

Contact | Impressum

File Tools View Help

Real-Time Data Stream | Real-Time Sliding Slices | Visual Feature Selection | Summarized Sliding Slices | Event Timeline & Insights | Search & Exploration

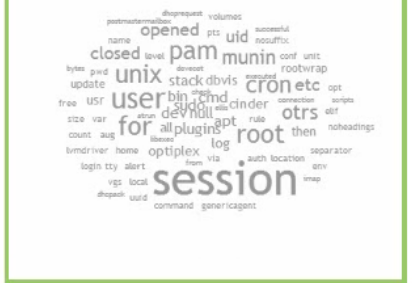
Sliding Slice Stream

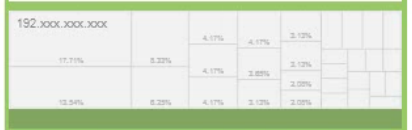
Sliding Slice (#227)
2014-08-04T21:42:28.786 ↔ 2014-08-04T21:43:27.373

L3

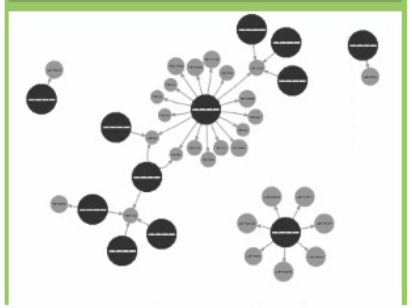
30 events 6 programs 6 hosts

30 srcIPs | 31 dstIPs | 73 srcPorts | 56 dstPorts





Cron 50.00% | Usr/sbin/cron 20.00%

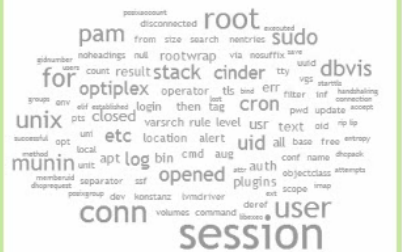


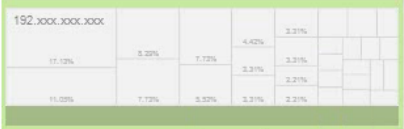
Sliding Slice (#234)
2014-08-04T21:45:59.325 ↔ 2014-08-04T21:46:57.510

L3

30 events 7 programs 5 hosts

31 srcIPs | 32 dstIPs | 70 srcPorts | 49 dstPorts





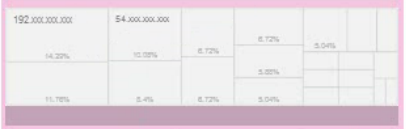
Slapd 43.33% | Cron 23.33%

Sliding Slice (#236)
2014-08-04T21:46:59.506 ↔ 2014-08-04T21:47:57.625

L3

30 events 5 programs 7 hosts

21 srcIPs | 23 dstIPs | 54 srcPorts | 36 dstPorts

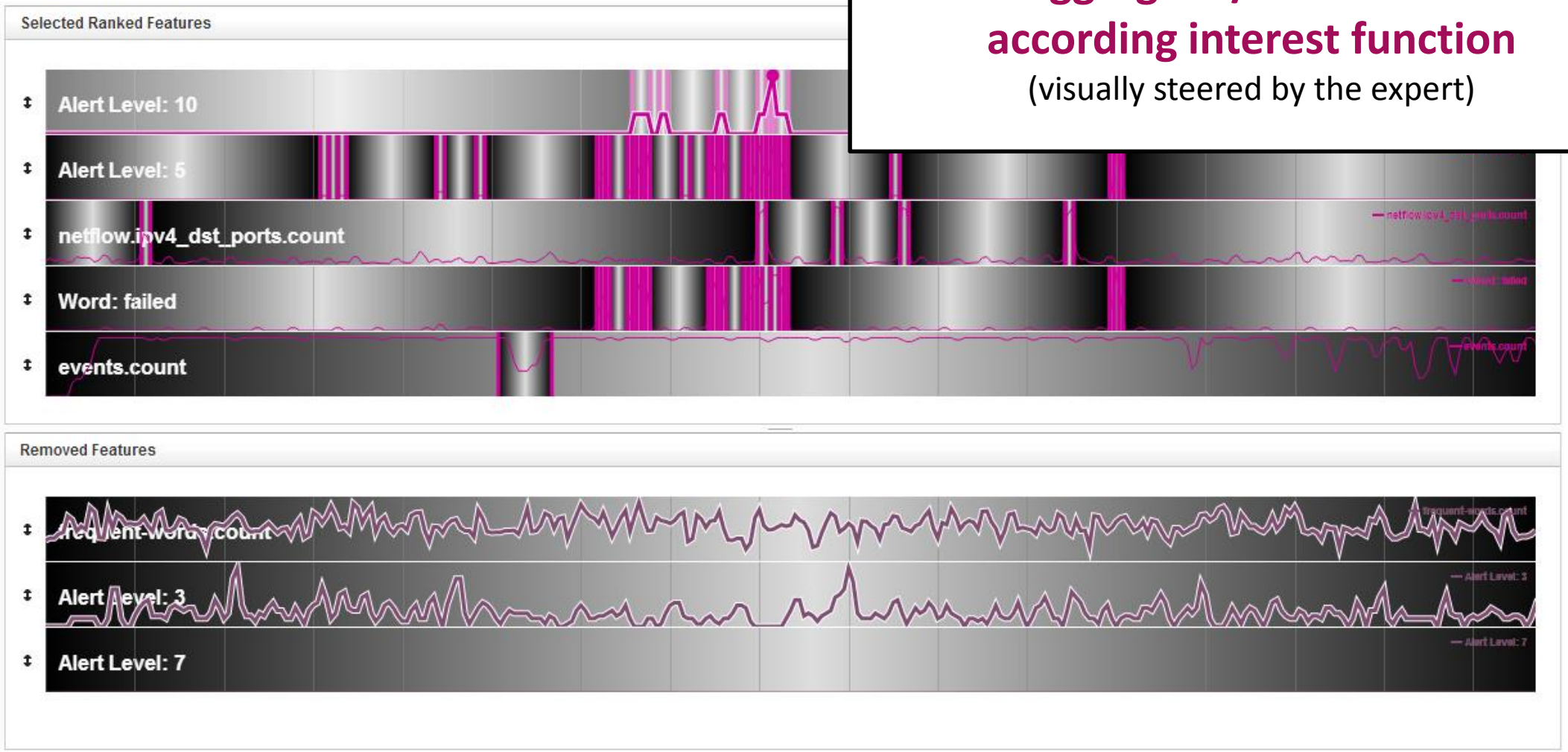


Cron 43.33% | Usr/sbin/cron 33.33% | Dovecot 13.33%

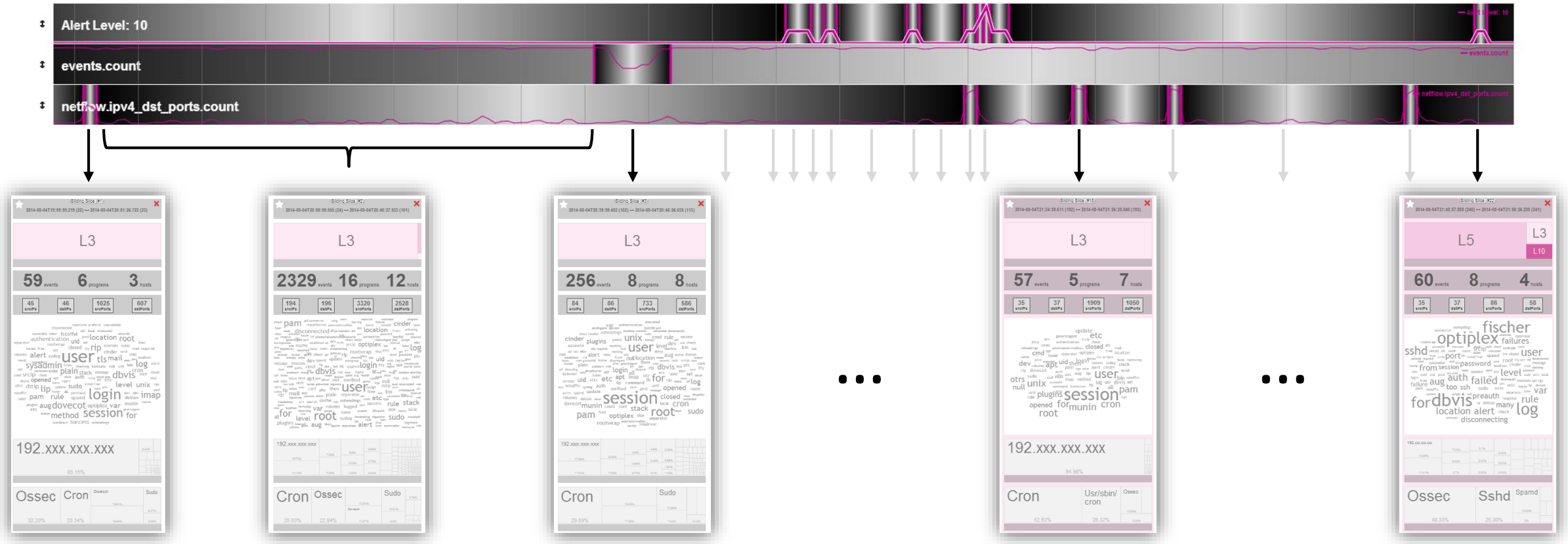

EXPLORATION CHALLENGE
How to explore
many *sliding slices*?

Visual Feature Selection

Visual Analytics Approach:
Aggregate / Summarize according interest function
(visually steered by the expert)



Example: Using Visual Analytics for Interactive Summarization



Demo

VACS
Visual Analytics Suite for Cyber Security

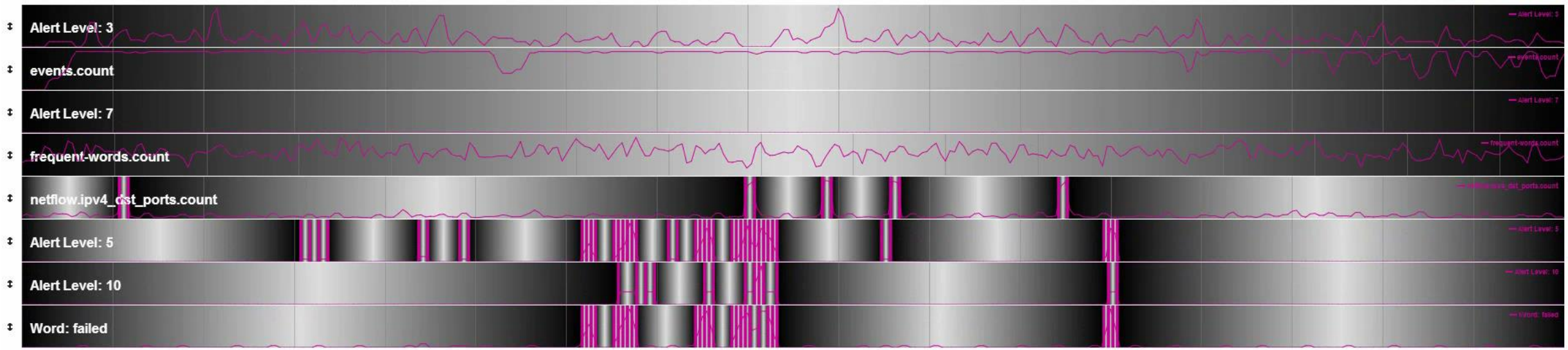
Contact | Impressum

File Tools View Help

Real-Time Data Stream | Real-Time Sliding Slices | Visual Feature Selection | Summarized Sliding Slices | Event Timeline & Insights | Search & Exploration

Load Feature Set: Feature Sets | Load Feature: Features | Create Feature | Summarize Stream

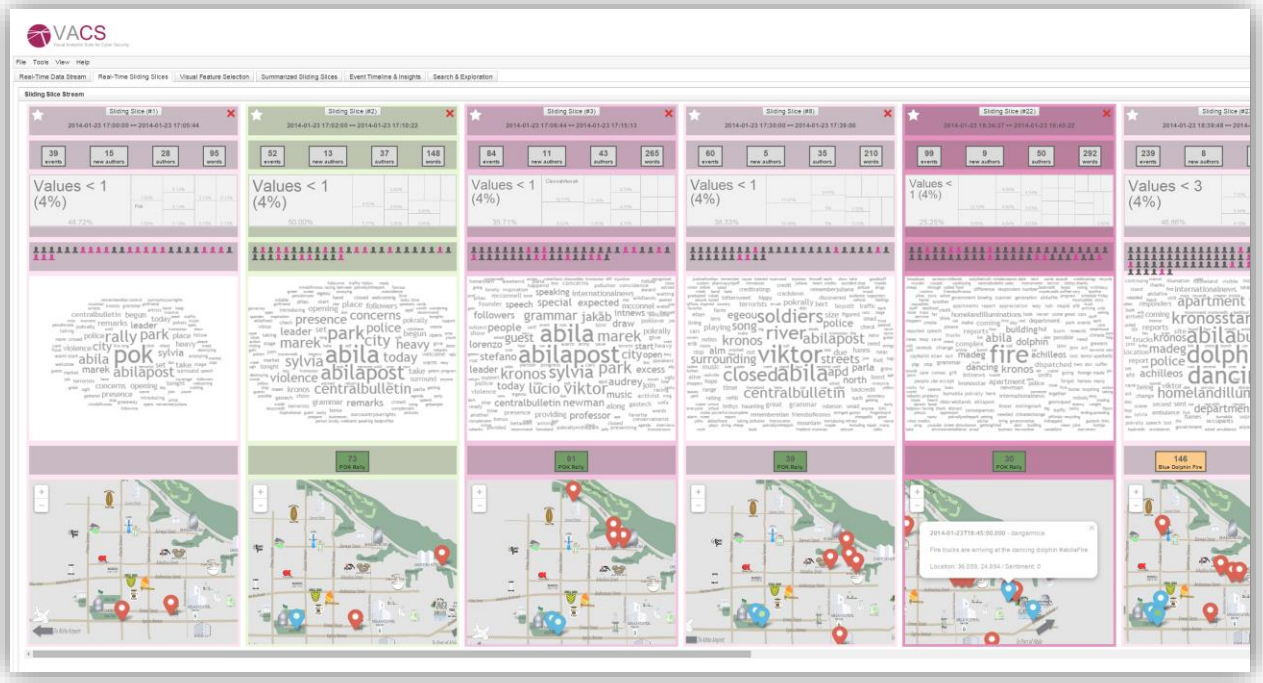
Selected Ranked Features

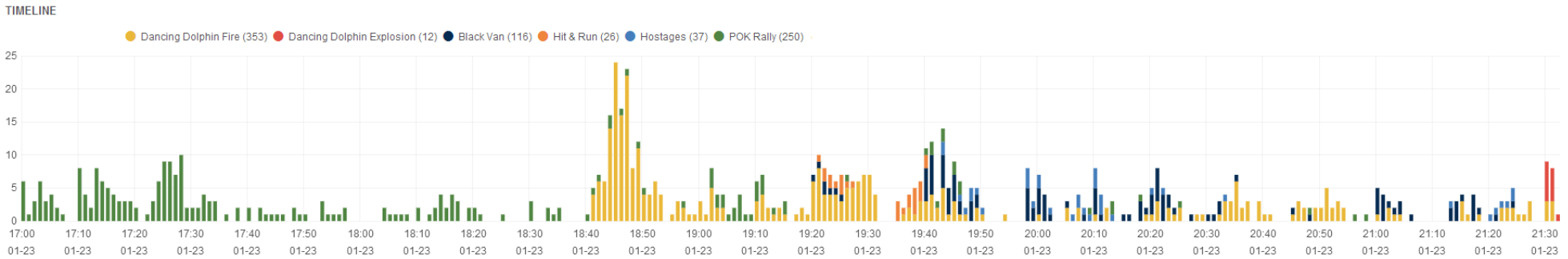


Removed Features

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

twitter
@f2cx

http://ff.cx/