Visually Fusing Contextual Data for Situational Understanding

Homeland Security
Science and Technology

Cyber Security Division
2012 Principal Investigators' Meeting

October 10, 2012

John Goodall
Team Lead, Visual Analytics Research Team
Oak Ridge National Laboratory
jgoodall@ornl.gov
865-446-0611

TTA6: Modeling of Internet Attacks

Situational Understanding and Attack Attribution

Homeland Security
Science and Technology

The challenge is the path from massive data to information to understanding

(a) Collect and store relevant data: Understand how to identify, collect and ultimately store data appropriate to the form of situational awareness desired.

(b) Analysis on heterogeneous data sources at large scale, using effective filtering techniques to remove irrelevant data, and supporting a variety of granularities, repeated patterns of interaction over long periods, and unexpected connections between companies and individuals.

(c) Novel approaches to presentation of large scale data, including scalable visualization, visualization with accurate geolocation, and zoomable visualization at varying levels, maintaining the ability to delve into the original data as well as broaden out to a high level people-aware view.

(d) Collaborative collection of non-open data, and the subsequent vetting, archiving, correlation, and generation of useful metadata.

(e) Cross-organizational boundary sharing of situational understanding.

(f) Situational understanding at multiple time scales, from autonomic response in the millisecond range to attack and threat scenarios over years.

Team

- Oak Ridge National Laboratory
  - John Goodall, PI
  - Erik Ferragut, Co-PI
- Pacific Northwest National Laboratory
  - Bill Pike
- Stanford University
  - John Gerth
- Research and Education Networking ISAC
  - Doug Pearson

The problem

1. Difficult to put security events into context
   - Events are information impoverished – only convey data about the event itself
   - Analysts manually gather and synthesize relevant data to add context to events
2. Difficult to perceive the threat landscape
   - Macro views are too high or low level to glean much about current threats trends
   - Analysts manually gather and synthesize data to determine which threats are important for resource allocation and policy decisions

Lack of context

What kind of attack?

How much should I care?

```
[**] [1:418:11] GPL SNMP request tcp [**]
[Classification: Attempted Information Leak]
[Priority: 2]
04/05-21:47:44.779542
172.23.240.156:39944 -> 172.23.0.1:161
TCP
TTL:48 TOS:0x0 ID:[illegible] IpLen:20 DgmLen:44 ******S*
Seq: 0x1A655FBF Ack: 0x0 Win: 0x400
TcpLen: 24 TCP Options (1) => MSS: 1460
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2002-0013]
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=2002-0012]
[Xref => http://www.securityfocus.com/bid/4132]
[Xref => http://www.securityfocus.com/bid/4089]
[Xref => http://www.securityfocus.com/bid/4088]
```

Who are the players?

Where can I get more info?

Provides a starting point and references, but analyst needs to manually identify, collect and synthesize relevant data

Lack of context

- Log aggregation systems can add some temporal context
- SIMs integrate some context by correlating other security logs
- CVEs can be examined for additional data about a vulnerability

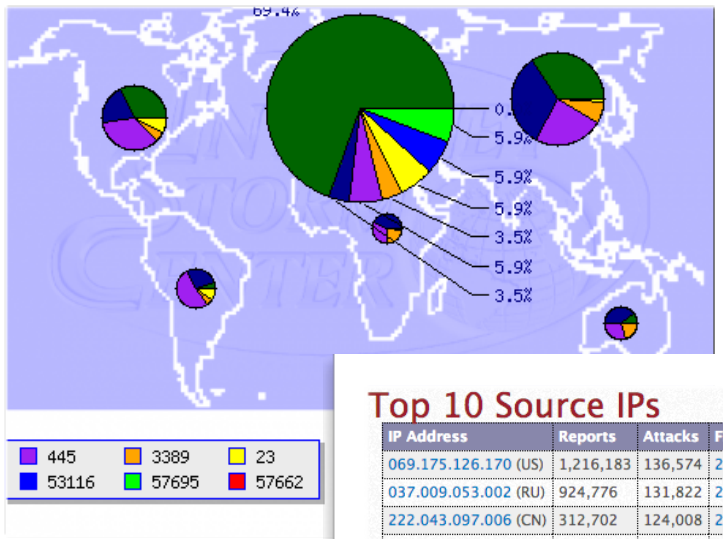
- Contextual data sources are largely from **within an enterprise**
- Identifying, gathering, and synthesizing disparate data is largely a **manual process**

The problem

1. Difficult to put security events into context
   - Events are information impoverished – only convey data about the event itself
   - Analysts manually gather and synthesize relevant data to add context to events

2. Difficult to perceive the threat landscape
   - Macro views are too high or low level to glean much about current threats and trends
   - Analysts manually gather and synthesize data to determine which threats are important for resource allocation and policy decisions

Lack of threat information



Top 10 Source IPs

IP Address	Reports	Attacks	First Seen	Last Seen
069.175.126.170 (US)	1,216,183	136,574	2012-07-11	2012-10-08
037.009.053.002 (RU)	924,776	131,822	2012-09-12	2012-10-08
222.043.097.006 (CN)	312,702	124,008	2012-06-27	2012-10-08
061.147.103.098 (CN)	250,935	99,039	2012-10-07	2012-10-07
060.174.198.082 (CN)	264,951	93,439	2012-10-08	2012-10-08
221.195.083.181 (CN)	85,372	82,488	2012-09-14	2012-10-08
200.085.122.011 (AR)	140,193	79,636	2012-08-19	2012-10-07
069.175.054.106 (US)	1,302,837	78,732	2012-07-14	2012-10-08
183.063.031.122 (CN)	138,324	77,814	2012-09-04	2012-10-08
220.225.004.021 (IN)	428,021	76,396	2012-09-14	2012-10-08

TA12-265A : Microsoft Releases Patch for Internet Explorer Exploit
TA12-262A : Microsoft Security Advisory for Internet Explorer Exploit
TA12-255A : Microsoft Updates for Multiple Vulnerabilities
TA12-251A : Microsoft Update For Minimum Certificate Key Length
TA12-240A : Oracle Java 7 Security Manager Bypass Vulnerability
TA12-227A : Microsoft Updates for Multiple Vulnerabilities
TA12-192A : Microsoft Updates for Multiple Vulnerabilities
TA12-174A : Microsoft XML Core Services Attack Activity
TA12-164A : Microsoft Updates for Multiple Vulnerabilities
TA12-156A : Microsoft Windows Unauthorized Digital Certificates
TA12-129A : Microsoft Updates for Multiple Vulnerabilities
TA12-101B : Adobe Reader and Acrobat Security Updates and Architectural Improvements
TA12-101A : Microsoft Updates for Multiple Vulnerabilities
TA12-073A : Microsoft Updates for Multiple Vulnerabilities
TA12-045A : Microsoft Updates for Multiple Vulnerabilities
TA12-024A : "Anonymous" DDoS Activity
TA12-010A : Microsoft Updates for Multiple Vulnerabilities
TA12-006A : Wi-Fi Protected Setup (WPS) Vulnerable to Brute-Force Attack

Provides an overview of trends, but hard to determine which of those trends are the most important

Lack of threat information

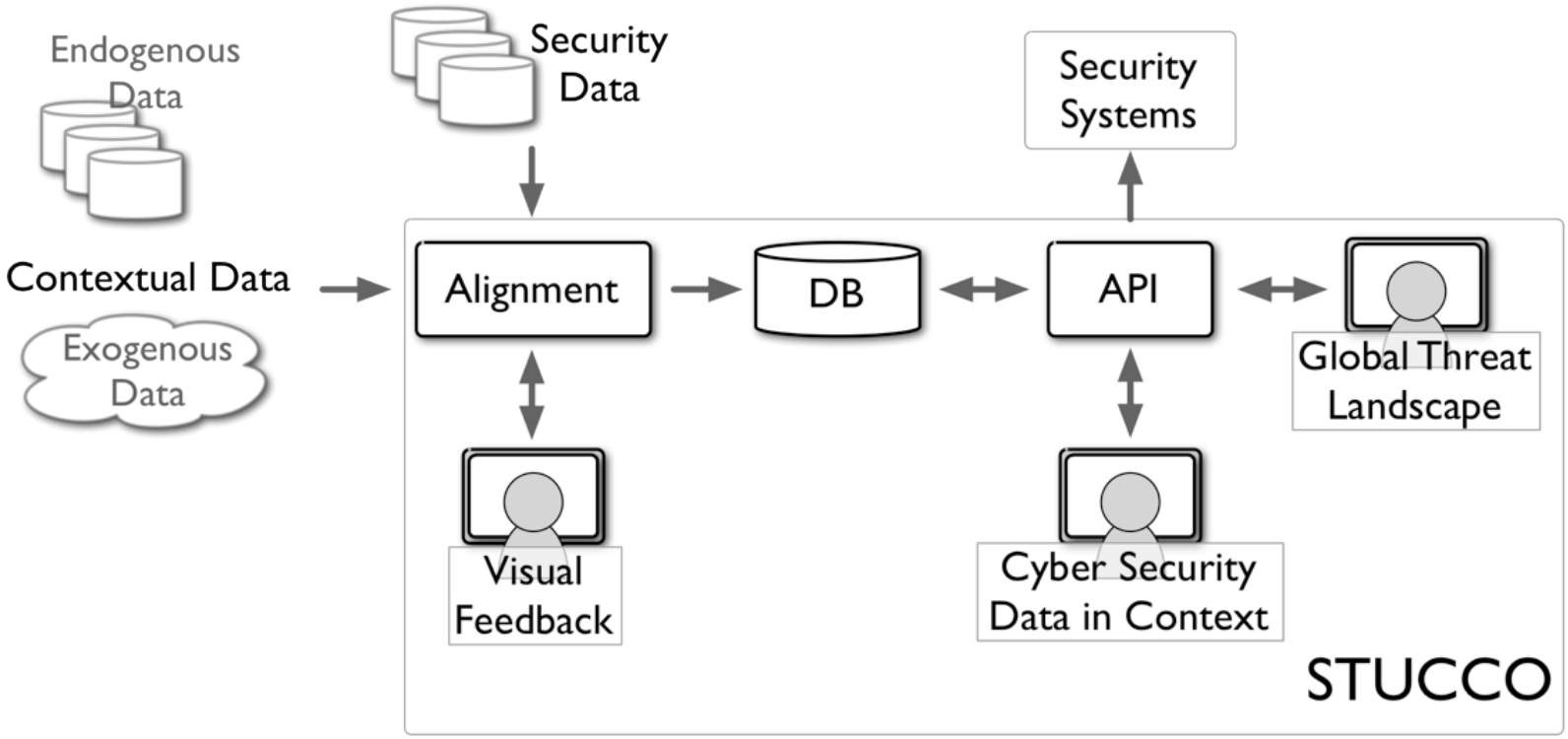
- SANS ISC/DShield provide summary data at a very high level
- SANS and other security organizations create lots of *top N* lists
- US-CERT and others publicize detailed lists and feeds of vulnerabilities and exploits

- Data relevant to threats is often high- or low-level with little support for moving between levels to **understand trends**
- Identifying, gathering, and synthesizing disparate data is largely a **manual process**

Goal

- Identify, collect, align, and provide visual and programmatic access to data sources not typically integrated into security systems to facilitate **situational understanding**

Technical approach



Situation & Threat Understanding by Correlating Contextual Observations

Technical approach

- Identify and evaluate non-traditional security data sources
  - Non-cyber security data sources within an enterprise
  - Open source data that is external to the enterprise
- Develop data alignment algorithms and visual feedback
  - Automated alignment of heterogeneous data sources to automate the synthesis process
  - Visualizations for integrating domain expertise into alignment process
- Develop APIs and data visualizations
  - APIs to provide contextual information for other systems
  - Threat landscape visualizations to explore security trends
- Demonstrate prototypes with operational users

Data / Use cases

Triggering Security Event

Data Resource		
	Internal Event Within Enterprise	External Event Within Enterprise
	Internal Event External to Enterprise	External Event External to Enterprise

Data / Use cases

Triggering Security Event

Data Resource

Internal Event Within Enterprise	External Event Within Enterprise
Internal Event External to Enterprise	External Event External to Enterprise

Was the attack successful? What is the impact?

Triggering Security Event: snort alert

Data Resource: splunk logs

Data / Use cases

Triggering Security Event

Data Resource

Internal Event Within Enterprise	External Event Within Enterprise
Internal Event External to Enterprise	External Event External to Enterprise

How does an event pertain to me?

Triggering Security Event: US-CERT announcement

Data Resource: splunk logs

Data / Use cases

Triggering Security Event

Data Resource	Internal Event Within Enterprise	External Event Within Enterprise
	Internal Event External to Enterprise	External Event External to Enterprise

How to remediate? How to prevent similar attacks?

Triggering Security Event: snort alert

Data Resource: slashdot, twitter, sans

Data / Use cases

Triggering Security Event

Data Resource		
	Internal Event Within Enterprise	External Event Within Enterprise
	Internal Event External to Enterprise	External Event External to Enterprise

What trends are most important?

Triggering Security Event: US-CERT announcement

Data Resource: slashdot, twitter, sans

STUCCO focus areas

Triggering Security Event

Data Resource

Endogenous Event Endogenous Data	Exogenous Event Endogenous Data
Endogenous Event Exogenous Data	Exogenous Event Exogenous Data

Endogenous – originating within enterprise

Exogenous – originating outside the enterprise

Benefits

- Contextual information about alerts can help security analysts **make better decisions**
- Global threat landscape can help security analysts allocate their resources and adjust policies proactively
- Public API can be consumed by other security systems to provide context to events
- Security research and practitioner communities can leverage the identification and categorization of relevant data sources
- Data alignment methods and visualization tools may be useful to other domains outside of cyber security

Milestones [year 1]

Task 1. Research data sources, alignment approaches, and visualization methods

Task	Milestones	Month Completed
Task 1.1. Identify and evaluate data sources for the correlation with cyber security behavior.	Data sources identified	6
Task 1.2. Evaluate alignment approaches with selected data sources.	Alignment algorithms identified	9
Task 1.3. Design mockups for visualization of cyber security data in context.	Mockups	10
Task 1.4. Design mockups for visualization of global threat landscape.	Mockups	10
Task 1.5. Evaluate visualization mockups and designs with operational users.	Design issues identified	12

Milestones [year 2]

Homeland Security
Science and Technology

Task 2. Develop system prototype

Task	Milestones	Month Completed
Task 2.1. Develop scalable system architecture and APIs for communication.	Prototype APIs implemented	15
Task 2.2. Develop data alignment algorithms and visual feedback displays.	Alignment algorithms implemented	21
Task 2.3. Develop visualization displays of cyber security data in context.	Visualizations implemented	21
Task 2.4. Develop visualization displays of global threat landscape.	Visualizations implemented	21
Task 2.5. Evaluate prototype components with operational users.	Implementation issues identified	22
Task 2.6. Design and conduct scalability and usability tests.	Scalability and usability issues identified	23
Task 2.7. Develop user documentation and manuals.	User documentation	24

Milestones [year 3]

Task 3. Demonstrate system prototype

Task	Milestones	Month Completed
Task 3.1. Identify operational environment for demonstration.	Demonstration site identified	26
Task 3.2. Prepare for technology demonstration in an operational environment.	Demonstration procedure finalized	28
Task 3.3. Deploy and support technology demonstrations in an operational environment.	Demonstration results in Final Report	36

Milestones [ongoing]

Task 4. Manage and document project

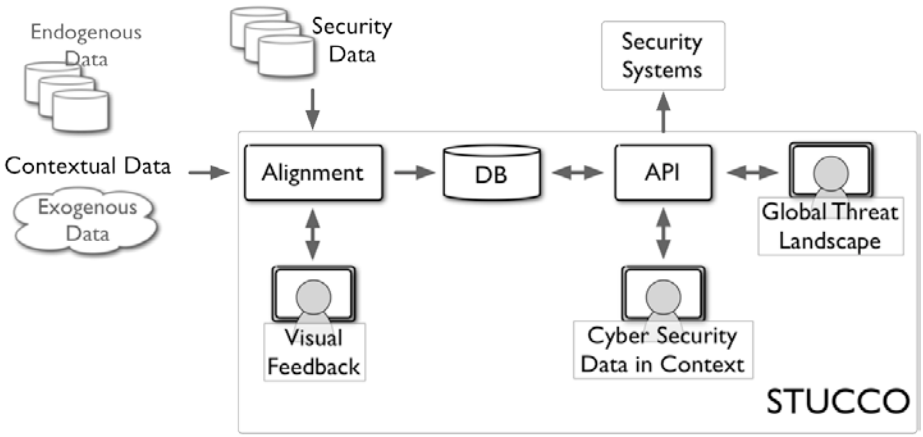
Task	Milestones	Month Completed
Task 4.1. Prepare for and participate in program reviews.	Periodic	Ongoing
Task 4.2. Document research results in scientific publications.	Publication-quality paper(s)	TBD
Task 4.3. Manage project and document results in regular progress reports and final report.	Monthly / Final Report	Ongoing / 36

Technology Transition Plan

- Testing
  - Non-synthetic data from PREDICT will be useful
  - DETER may be useful for testing alignment, but not for testing the entire system, for which we will need users and real data
- Operational demonstration
  - Research and Education Networking Information Sharing and Analysis Center (REN-ISAC)
  - ORNL Information Technology Services Division
  - Tennessee Valley Authority
- Community
  - All software will be open-sourced and released on github
  - We will publicize and start community building in year 2

STUCCO

Situation & Threat Understanding by Correlating Contextual Observations



Operational Capability

Visual fusion of heterogeneous data with security data

- Non-cyber security data sources within an enterprise
- Open source data that is external to the enterprise

Visual analytics for Situational Understanding

- Visual interface links context with cyber security events
- Broadly available service synthesizes threat indicators from non-cyber data

Open source software

- Low cost, customizable and modifiable open solution

Technical Approach

Develop methods to integrate and visualize social media, new articles and other indicators in context with traditional cyber security data

Tasks:

- Identify and evaluate non-traditional data sources
- Develop data alignment algorithms and visual feedback
- Develop visualization of cyber security data in context
- Develop visualization of global threat landscape
- Demonstrate prototypes with operational users

Related Efforts: visual integration of textual context from news content with cyber data in NUANCE software (PNNL)

Schedule, Deliverables, & Contact Information

Period of Performance: 3 years (Type I)

Milestones: prototype (mo 24), demonstration (mo 28)

Deliverables: all developed open source software, end-user documentation and reports.

Teaming Organization Information:

John Goodall, Oak Ridge National Laboratory
865-574-9778 / jgoodall@ornl.gov
William Pike, Pacific Northwest National Laboratory
John Gerth, Stanford University
Doug Pearson, REN-ISAC