2015 7th International Conference on Cyber Conflict:
Architectures in Cyberspace
M.Maybaum, A.-M.Osula, L.Lindström (Eds.)
2015 © NATO CCD COE Publications, Tallinn

Visual Structures for Seeing Cyber Policy Strategies

Jennifer Stoll
Lehrstuhl für Philosophie und
Wissenschaftstheorie
Technische Universität München (TUM)
München, DE
j.stoll@tum.de

Rainhard Z. Bengez
Lehrstuhl für Philosophie und
Wissenschaftstheorie
Technische Universität München (TUM)
München, DE
bengez@web.de

Abstract: In the pursuit of cyber security for organizations, there are tens of thousands of tools, guidelines, best practices, forensics, platforms, toolkits, diagnostics, and analytics available. However according to the Verizon 2014 Data Breach Report: "after analysing 10 years of data… organizations cannot keep up with cyber crime—and the bad guys are winning." Although billions are expended worldwide on cyber security, organizations struggle with complexity, e.g., the NISTIR 7628 guidelines for cyber-physical systems are over 600 pages of text. And there is a lack of information visibility. Organizations must bridge the gap between technical cyber operations and the business/social priorities since both sides are essential for ensuring cyber security. Identifying visual structures for information synthesis could help reduce the complexity while increasing information visibility within organizations. This paper lays the foundation for investigating such visual structures by first identifying where current visual structures are succeeding or failing. To do this, we examined publicly available analyses related to three types of security issues: 1) epidemic, 2) cyber attacks on an industrial network, and 3) threat of terrorist attack. We found that existing visual structures are largely inadequate for reducing complexity and improving information visibility. However, based on our analysis, we identified a range of different visual structures, and their possible trade-offs/limitation is framing strategies for cyber policy. These structures form the basis of evolving visualization to support *information synthesis for policy actions*, which has rarely been done but is promising based on the efficacy of existing visualizations for cyber incident detection, attacks, and situation awareness.

Keywords: *cyber security policy, visualization, human-computer interaction, visual structures, organizations*

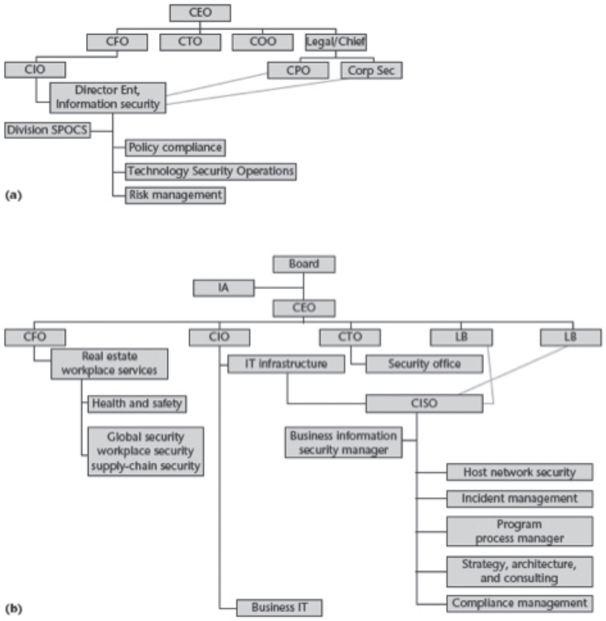
1. INTRODUCTION

A core task in making cyber policy actions is *seeing* the data that support them. In other words, decision-makers must take highly disparate data, many point of views, and synthesize them into a coherent and concise narrative that fits into a broader strategy. Yet seeing *cyber policy* remains difficult. With the growing Internet of Things, cyber policy is quickly becoming intractable for decision-makers for several reasons. One reason is the sheer complexity in terms of the volume, variety, and velocity of cyber data. To illustrate, in the Verizon 2014 Data Breach Report, over 100,000 different cyber incidents were identified in the analysis [19]. Also much of our information suffers from fragmentation. Information we need is often "trapped" in other organizations due to conflicting priorities because of privacy issues, funding issues, proprietary data and so forth. Technical concerns further exacerbate the fragmentation due to interoperability issues or inherent limitations in the design of databases and sensor systems for data collection. Additionally, much of the policy we need to see is encoded into text, because abstractions like cyber policy have not been spatialized so that they can be visualized beyond text.

Challenges for organizations: Complexity, fragmentation, interoperability issues, and lack of spatialization summarizes why cyber policy is hard to see. These four issues degrade information visibility in organizations. And the impact of these challenges is manifested in a range of organizational factors that undermine the security of organizations, while enabling challenges such as unintentional insider fraud [11]. One example is a tendency of organizational complacency towards cyber security based on erroneous perceptions of security risks. Critical information is obscured about the impact of not implementing a range of security controls to deter activities such as insider fraud or to prioritize based on areas of risk comparison. Additionally, interdependencies and the implementation of inappropriate controls result from the lack of information visibility between technical operations, managers, and non-technical staff within organizations. Basically, organizations struggle to see why certain solutions are needed are how they should fit into the broader organizational context, especially in light of other expenditures and allocation of resources.

A position paper by Johnson & Goetz [6] adds how organizational structure adds structural challenges that further hinder visibility. Figure 1 below shows two main organizational structures to highlight overlaps in responsibility and the multi-layered coordination that security tasks require. According to their study: "the security group's organizational structure is in flux and seems to undergo frequent change…It's difficult to pinpoint structural best practices because the security landscape changes so rapidly that further structural changes are likely in the coming years." [6]

FIGURE 1: "ORGANIZATIONAL STRUCTURE. (A) IN SOME ORGANIZATIONS, SECURITY MANAGEMENT REPORTS DIRECTLY TO THE CIO; (B) IN OTHERS, IT REPORTS INDIRECTLY TO THE CIO THROUGH OTHER IT EXECUTIVES." [REF]

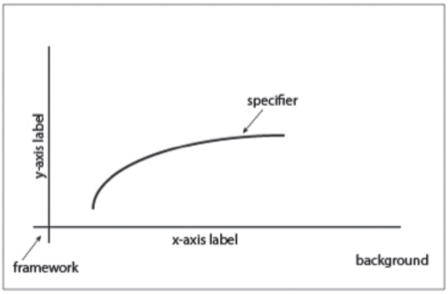


This constant shifting could indicate internal attempts by organizations to cope with the fact that security of organizations requires the cooperation and attention of all members. And the movement from area to area is a symptom of trying to find a home for security, which is a challenge because again, security needs to be part of the entire organization. The implication here is that visual structures that accommodate the multiple viewpoints present in an organization are critically needed in order to embed security within organizations and not solely IT systems.

2. VISUAL STRUCTURES

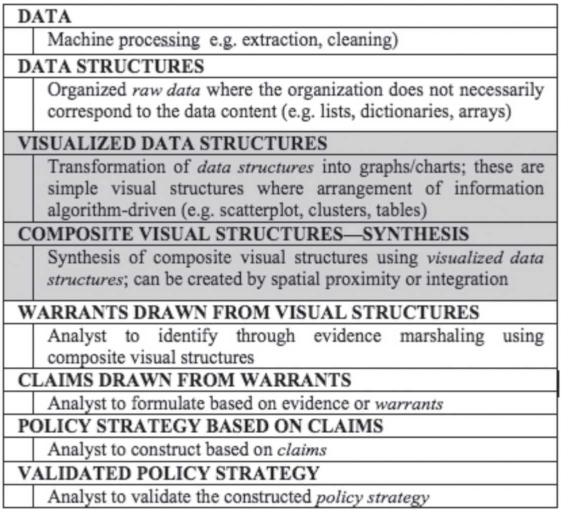
In other words, organizations must bridge the gap between technical cyber operations and the business/social priorities since both sides are essential for ensuring cyber security. Identifying visual structures for information synthesis could help reduce the complexity while increasing information visibility within organizations. This paper lays the foundation for investigating such visual structures by first identifying where current visual structures are succeeding or failing. We first conceptualize the notion of "visual structure" using the work of Kosslyn [7] who defined the components. Considered abstractly, a single visual structure such as a chart or graph according to Kosslyn, have four basic level constituent parts: 1) the background though not essential, can serve to highlight, emphasize or reinforce the information being conveyed; 2) the framework provides the mapping, the axes, or logic of the arrangement for the specifiers and labels; 3) the specifiers are elements such as lines, blocks, bars, points, and so forth, which represent the data; 4) the labels are letters, words, numbers or even pictures that help us to correctly interpret the specifiers or aspects of the framework. Figure 2 below provides a simple illustration of these parts.

FIGURE 2: VISUAL STRUCTURE COMPONENTS FOR SIMPLE GRAPH



We then extend this conceptualization to capture the structure of multiple visual structures used in conjunction, which reflects the actual core task of policy analysis where a wide-range of visuals and information are employed. We use the work of Toulmin's informal structure for building an argument, which includes the use of warrants (based on data) to marshal evidence to support claims that comprise a policy strategy or the overall "argument" [14]. Table 1 below incorporates this information structure and shows six in-between transformations of "data".

TABLE 1: CHAIN-OF-CONNECTIONS FROM RAW DATA TO POLICY STRATEGY



DATA	
	Machine processing e.g. extraction, cleaning)
DATA STRUCTURES	
	Organized *raw data* where the organization does not necessarily correspond to the data content (e.g. lists, dictionaries, arrays)
VISUALIZED DATA STRUCTURES	
	Transformation of *data structures* into graphs/charts; these are simple visual structures where arrangement of information algorithm-driven (e.g. scatterplot, clusters, tables)
COMPOSITE VISUAL STRUCTURES—SYNTHESIS	
	Synthesis of composite visual structures using *visualized data structures*; can be created by spatial proximity or integration
WARRANTS DRAWN FROM VISUAL STRUCTURES	
	Analyst to identify through evidence marshaling using composite visual structures
CLAIMS DRAWN FROM WARRANTS	
	Analyst to formulate based on evidence or *warrants*
POLICY STRATEGY BASED ON CLAIMS	
	Analyst to construct based on *claims*
VALIDATED POLICY STRATEGY	
	Analyst to validate the constructed *policy strategy*

These transformations capture where visual structures (highlighted in green) are currently being used in the process of formulating data-driven policy strategy—starting first from the raw data and culminating into the validated policy strategy. The Toulmin argument structure provides a flexible way of organizing the various information structures that could form the basis of policy. As a first step towards specifying visual structures for information synthesis in formulating policy, we identify two paths that can be taken, which are based existing visual systems used in the case studies:

- Synthesis by *proximity* where synthesis is accomplished by placing or combining individual visual structures in close spatial arrangements;
- Synthesis by *integration* where synthesis is accomplished through joining by using a common parameter to intersect the data represented by the visual structures.

An example of synthesis by proximity is the common "multi-view" visualization tools that place multiple windows of different graphs from scatterplots to timelines or clusters in close physical proximity. Often these graphs are created using the same source of data. However, they represent individual graphs only and primarily provide different views of the data. In contrast, the synthesis by integration may use the same data source, but different graphical approaches are combined into one view using common parameters. Examples of such seem to be less common but are illustrated in each of the case studies.

We use this extended conceptualization of visual structure synthesis and Kosslyn's notion of visual structure to analyze the case studies, which is presented in the following section.

3. CASE STUDIES: EPIDEMIC, CYBER ATTACKS, TERRORISM

We applied this extended notion of Kosslyn's visual structure to samples from the VAST 2011 Contest [16]. The contest involved three mini-challenges and one grand challenge where teams had to 1) characterize an epidemic spread, 2) identify cyber security issues in a corporate network, and/or 3) investigate terrorist activity in a document set. Teams were required to analyze the same raw data supplied to all teams and then using any visualization of their choice, construct a policy strategy by identifying a set of claims based on a range of evidence. The data supplied by the Challenge were synthetic, both computer and human-generated. The different datasets included: microblog messages collected from mobile GPS enabled devices, population statistics, observed weather, additional facts about geographic location, computer network architecture of the corporation, a list of security policy rules, a firewall log, an intrusion detection system log, an aggregated system logs for all hosts on network, a Nessus Network vulnerability scan report, and 4,400+ text documents. All datasets had anomalies, with only some of them being significant.

There were a total of 18 teams submitting correct solutions across the challenges. For our study, we selected eleven samples, excluding submissions with incorrect answers since our focus was to examine visualizations that support the framing analysts need to make between the raw data and the policy strategy. Our study differs from studies of argument-based systems in that we are not evaluating the soundness of an argument as in [11]. Instead, we seek to understand the relationship between the arguments formed and visualizations used for support. Our goal is not on the cognitive processes occurring inside the analyst's head, but we focus on the relationship between the visual structures and the resultant policy strategy generated from them.

Thus for each of the eleven samples, we analyzed the submissions to establish what we refer to as the chain-of-connections to go from raw data to policy strategy; and these chains identify the transformations involved in this process. After identifying a chain-of-connection from raw data to policy strategy for each submission, we compared and contrasted the Visual Data Structures and Composite Visual Structures used to generate the warrants and claims for the strategies.

A. Case #1: Epidemic

The first mini-challenge tasked teams with identifying the origin of an epidemic spread, outlining the affected area, and hypothesizing on how the epidemic is spreading. The task requires the following information to be derived from raw data: 1) three claims on origin, spread, and vector of the epidemic, and 2) the warrants or evidence to support the three claims. In the analysis, we identify a chain-of-connection for each of the three correct submissions. We refer to them as Team A, B, and C. All three teams used the same raw data provided to all teams and similar data structures: 1) thousands of microblog messages organized as a table, 2) population statistics and observed weather for specific days such as wind direction organized as a table, and 3) additional facts about the fictional city Vastopolis as well as 4) a geographic map showing landmarks.

The composite visual structure that Team A created (shown in fig. 3a) included all of the datasets. Team A correctly ascertained the origin and half of the epidemic spread by the wind to uptown Vastopolis, but failed to identify the other half spread down river.

FIGURE 3A: TEAM A'S
SPATIO-TEMPORAL MAP

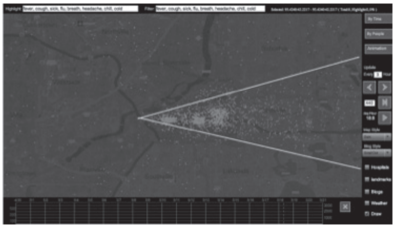
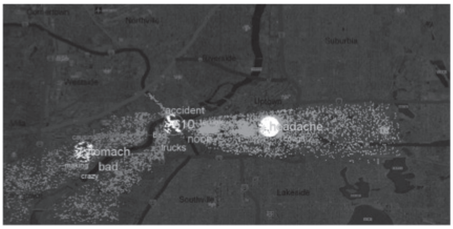

Highlight: fever, cough, sick, flu, breath, headache, chill, cold
Filter: fever, cough, sick, flu, breath, headache, chill, cold

FIGURE 3B: TEAM B'S
SPATIO-TEMPORAL MAP



Team B used clusters and graph-set operations to integrate the visualized data structures along the dimension of geographic coordinates, i.e., the scatterplots and terms extracted from the microblog texts. The placement of specifiers and labels was determined solely by examining density and proximity of microblog message clusters as shown in Figure 3b. Interestingly, Team B did not attempt to incorporate the time dimension, or the weather data structure. This approach helped Team B easily identify the origin and spread of the epidemic in two primary areas, but they did not identify the vectors for spreading the disease, nor any details of timing.

Team C used all data structures to create a synthesized view and preserved views of each visualized data structure using an implied compartmentalized approach. The terms from the text extraction of the microblog message were displayed as a tag cloud cluster. While the filter terms were displayed as bars on the right. The weather and wind were displayed below the map, and a layered stack to represent the messages over time. They additionally integrated all of the visualized data structures using geographic coordinates and cardinal directions to arrange them on the map background as shown in Figures 4 and 5 below.

FIGURE 4 AND 5: TEAM C'S SPATIO-TEMPORAL MAP

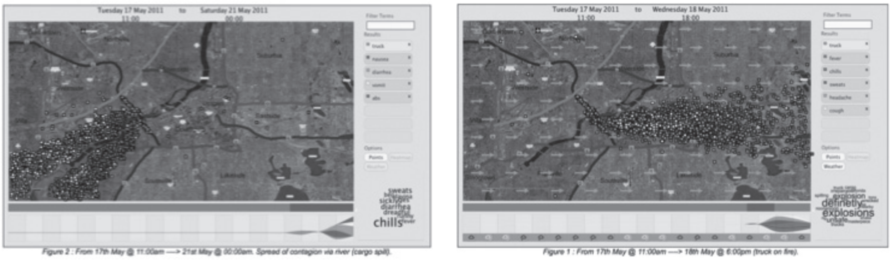
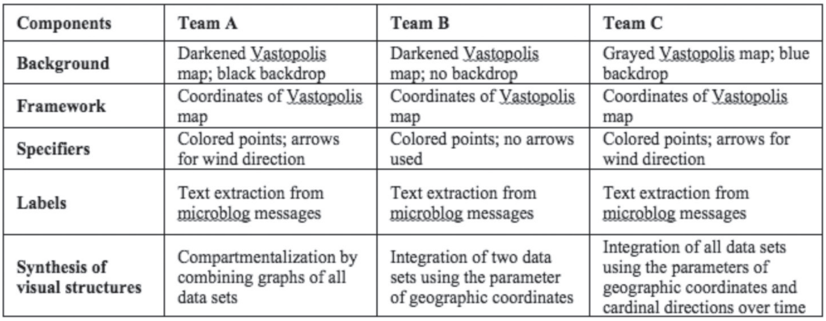


Figure 2 : From 17th May @ 11:00am ---> 21st May @ 00:00am. Spread of contagion via river (cargo spill).

Figure 1 : From 17th May @ 11:00am ---> 18th May @ 6:00pm (truck on fire).

In addition, interaction widgets in the implied compartments were used in the synthesized visual structure in the center. For example, selecting a specific filter term generated the geolocation as scattered points on the map; and selecting a term in the tag cloud highlighted the relevant colored dots. The background, framework, specifiers, and labels were effectively integrated into one view, including the arrows representing the wind pattern arranged on the map background. All three teams used similar visualized data structures but different composite visual structures, which are summarized in Table 2. For the background and framework, the teams used the map provided by the Challenge. For the specifiers, all three teams used colored dots to indicate the geo-location of each microblog entry. The labels utilized were also extracted from the same microblog data, indicating symptoms of illness and an unusual truck accident on fictional Highway 610 in Vastopolis. A critical difference here is that Team B did not use a visualized data structure for the weather, resulting in overlooking critical details for situation awareness such as the start date for the epidemic.

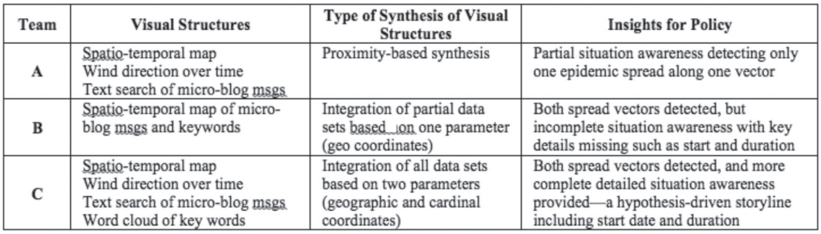
TABLE 2: VISUAL STRUCTURE COMPONENTS ACROSS TEAMS



Components	Team A	Team B	Team C
Background	Darkened Vastopolis map; black backdrop	Darkened Vastopolis map; no backdrop	Grayed Vastopolis map; blue backdrop
Framework	Coordinates of Vastopolis map	Coordinates of Vastopolis map	Coordinates of Vastopolis map
Specifiers	Colored points; arrows for wind direction	Colored points; no arrows used	Colored points; arrows for wind direction
Labels	Text extraction from microblog messages	Text extraction from microblog messages	Text extraction from microblog messages
Synthesis of visual structures	Compartmentalization by combining graphs of all data sets	Integration of two data sets using the parameter of geographic coordinates	Integration of all data sets using the parameters of geographic coordinates and cardinal directions over time

Despite using the same raw data, data structures, and similar visual structures, Team A supplied only a partially correct answer. Team B answered mostly correctly, but missed key details that would have facilitated a more complete hypothesis. However, Team C provided the most complete and correct answer that matched the posted solution for this task.

TABLE 3: CASE #1: VISUAL STRUCTURE SYNTHESIS FOR EPIDEMIC HYPOTHESIS



Team	Visual Structures	Type of Synthesis of Visual Structures	Insights for Policy
A	Spatio-temporal map Wind direction over time Text search of micro-blog msgs	Proximity-based synthesis	Partial situation awareness detecting only one epidemic spread along one vector
B	Spatio-temporal map of micro-blog msgs and keywords	Integration of partial data sets based on one parameter (geo coordinates)	Both spread vectors detected, but incomplete situation awareness with key details missing such as start and duration
C	Spatio-temporal map Wind direction over time Text search of micro-blog msgs Word cloud of key words	Integration of all data sets based on two parameters (geographic and cardinal coordinates)	Both spread vectors detected, and more complete detailed situation awareness provided—a hypothesis-driven storyline including start date and duration

As summarized in Table 3, the primary difference in the resultant policy insights generated the three teams seemed to be how the visual structures were synthesized. The integration of all data sets using multiple parameters resulted in the most complete hypothesis for the epidemic event, which would inform the situation awareness needed to take concrete policy actions for this case. This case illustrates how policy makers need to be aware that adopting different approaches for synthesizing the visual structures may result in varying degrees of hypothesis completeness.

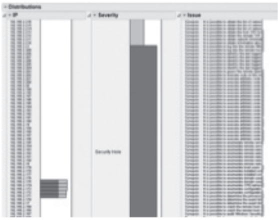
B. Case #2: Cyber Attacks on Corporate Network

For this case, we examined the submissions of five different teams using a range of visual structures to complete the task. As in the previous section, the team names are randomly assigned and do not correspond with any submission names on the VAST 2011 Challenge site. The cyber security mini-challenge task was to identify up to five security incidents of interest from the given data. The raw data supplied to and used by all teams were composite, unstructured format, and included 1) a text description of the computer network architecture, which identified priority computers, 2) a set of security policy rules, 3) firewall log data, 4) intrusion detection system log data, 5) aggregated syslogs for all the hosts on the network, and 6) a Nessus Network Vulnerability Scan Report. All teams used a range of visualized data structures and composite visual structures. In what follows, we detail the chain-of-connections for each team organized according the type of visual structure used by the team.

1) Simple Table

Team 1 imported the raw data supplied by the Challenge into a table structure and used different filter and sort functions to navigate the information as shown in Figure 6. A total of three separate tables were created for each type of log data. Using their three tables, Team 1 identified one incident of interest per table, which is described below.

FIGURE 6: TEAM 1 VISUAL STRUCTURE – SIMPLE MATRIX

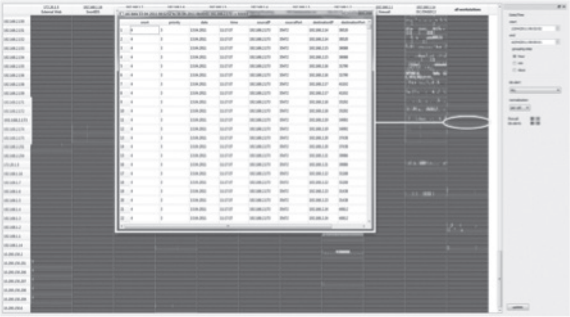


Impact of the simple table visual structure on policy strategy: These three claims with their associated warrants comprise the policy strategy of security events constructed by Team 1. Using the visual structure of a table to organize the data enabled Team 1 to easily identify incidents that generate a high frequency of the same data, e.g. message flooding. For these events, many relevant details were displayed directly, without any need to "drill in." However, infrequent, but highly important events, such as login attempts, were not found with this structure, though they were present in the data. The resultant policy strategy based on these three claims tended to focus on high-noise events and overlooked the quieter events that may be even more pernicious and difficult to detect.

2) Complex Table

Team 2 also utilized a complex table structure to organize the data by using a larger table to show the relations between each source and destination, although some machines were grouped together to reduce visual complexity. Each cell of their table contained a histogram of events that occurred between each pair of machines or groups of machines during the selected time window as shown in Figure 7. Their table also included additional sub-framework within the larger one. More specifically, analysts could select any of the histograms to drill down to a table of the raw data that it represented. They included a panel on the right to enable some basic filtering according to desired time ranges and alert types. Team 2 also used a commercially available data analysis tool (Tableau), to support some of their analysis. This was used to generate a few simple summary charts, which supported some of their warrants.

FIGURE 7: TEAM 2 VISUAL STRUCTURE – COMPLEX TABLE

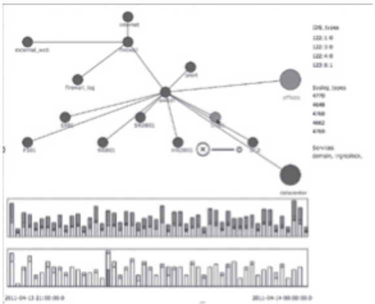


Impact of the complex table visual structure on policy strategy: The main limitation of Team 2's complex table as a visual structure is not utilizing visualization for representing overall network activity. Instead, their complex table organized the data by individual machine, and giving separate summaries of each combination of point-to-point connections. Thus, there was no chronological overview across all machines. Although summary charts generated with Tableau these summaries were not integrated with the rest of the visualization. This resulted in identifying attacks on individual machines but not when the attacks involved multiple disparate ones.

3) Graph and Histogram

As shown in figure 8, Team 3 utilized two visual data structures: a network topology graph to show locations of devices and their interrelationships, and two stacked histograms of SNORT and IDS log data.

FIGURE 8: TEAM 3 VISUAL STRUCTURE – GRAPH & STACKED HISTOGRAM



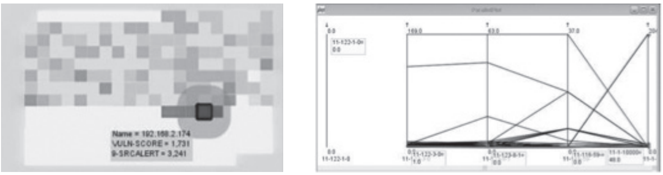
In the visual structure for the network topology, the background is implied. The framework or the logic of arrangement is dictated by how the computer network was actually set-up for the VAST 2011 Challenge data. The specifiers are the nodes and lines representing the devices on the network with corresponding labels. For the stacked histogram, the background is also implied. The framework has time on one axis and numbers of events by type on the other, with corresponding labels. The specifiers are the colored blocks of the histogram representing the total number of events by type over time with corresponding labels. Team 3 does not attempt to join the two visual structures to create a composite visual structure. Instead, Team 3 seems to use these to provide an initial overview of the data of leads for where to look at the raw data. However to find actual evidence or warrants to support their claims, they perform direct SQL queries against a database with the raw data. In other words, the chain-of-connection for Team 3 effectively bypasses the "visual data structure" and "composite visual structure" steps of the chain. This indicates that members of Team 3 relied primarily on their domain knowledge to navigate a way through the raw data.

Impact of the graph and stacked histogram visual structure on policy strategy: The visual structure of the network topology combined with the views of the stacked histogram, enabled team 3 to see some initial relevant information for both the whole network and the significant entities (machines, traffic, and events) over time. As their claims collectively demonstrate, this particular visual structure supports uncovering insights showing the impact of a machine on a network. However, one limitation is the difficulty seeing machine-specific issues within subnets: the visual structure obscured the presence of individual machines within the "offices" and "datacenter" subnets in both the topology as well as the histogram. This visualization served primarily as an overview and a starting point for constructing SQL queries. Thus these queries, rather than the visual structures, were then used in identifying 4 different attacks. Such visualization could be initially useful for domain experts, but less so for non-expert policy makers.

4) Simple Heat Map & Parallel Coordinate Plot

As indicated by figure 9a&b, Team 4 utilized two visual data structures: a simple heat map that presented traffic and alerts per machine, and a parallel coordinate plot that showed IDS log data on a per hourly basis. The granularity of the heat map was per machine, with a single block of the map representing one device on the network. The visual structure framework organizes IDS log data using time (per hour), source and destination nodes as the axes.

FIGURE 9A & 9B: TEAM 4 VISUAL STRUCTURE – SIMPLE HEAT MAP AND PARALLEL COORDINATE PLOT

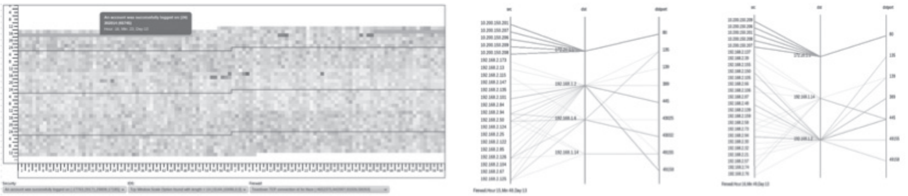


Impact of the heat map and parallel coordinate plot visual structure on policy strategy: The visual structure of a simple heat map combined with a parallel coordinate plot enabled Team 4 to easily see issues occurring on a per-machine and per-hour basis. This visualization provided an effective summary, but seemed to obscure infrequent but highly important events, such as the RDP login to the webserver. This visual structure did not easily reveal events that overlapped hours or machines, due to compartmentalization of the time slices to per-hour sections, and the relations between different machines were often not clear. This may have contributed to many of their claims lacking detail and specificity with regard to situation awareness.

5) Complex Heat Map & Parallel Coord. Plot

Team 5 also used two visual data structures: a complex heat map and a parallel coordinate plot, both at finer granularities as shown in figure 10a & b below.

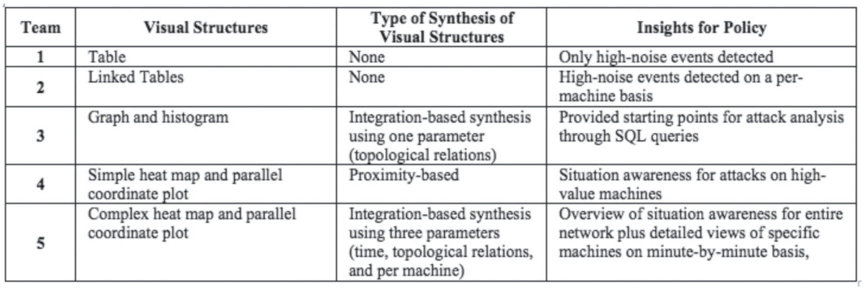
FIGURE 10A & 10B: TEAM 5 VISUAL STRUCTURE – COMPLEX HEAT MAP AND PARALLEL COORDINATE PLOT



For the complex heat map, the background was a bounded outline. The framework used time on the y-axis and event types on the x-axis, with corresponding labels. The specifiers were colored blocks for the entire network that changed colors depending on network traffic levels per event type per minute of each hour, with corresponding labels. For the parallel coordinate plot, the background is implied. The framework uses parallel axes of addresses of source (src) nodes on the network, destination nodes (dst), and the destination port (dstport). The specifiers are lines representing traffic from nodes to ports, with corresponding labels (e.g., src: 192.168.2.25, dst: 192.168.1.14, dstport: 445).

Impact of visual structure on policy strategy: Team 5's visualizations were particularly effective for showing time relationships between various events, which allowed causal sequences of events to be determined. This is often extremely important, such as identifying events where there were user logins to several machines immediately before they began scanning the rest of the network. The visualizations used by the other teams indicated the presence of scans, but were not able to convey important additional details such as these logins related to the scans. Team 5's complex heat map organized by time and machine remains a consistent visual structure throughout, providing continuous context, but also many useful filters to highlight various categories of events before relying on the parallel coordinates chart for still more additional details. Their visual structure enabled seeing issues related to the entire network using a fine-grained minute-by-minute representation, and well as going into the specific related data structure to identify related critical information.

TABLE 4: CASE #2: VISUAL STRUCTURE SYNTHESIS FOR CYBER ATTACK ANALYSIS



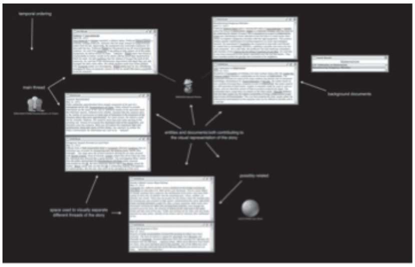
Team	Visual Structures	Type of Synthesis of Visual Structures	Insights for Policy
1	Table	None	Only high-noise events detected
2	Linked Tables	None	High-noise events detected on a per-machine basis
3	Graph and histogram	Integration-based synthesis using one parameter (topological relations)	Provided starting points for attack analysis through SQL queries
4	Simple heat map and parallel coordinate plot	Proximity-based	Situation awareness for attacks on high-value machines
5	Complex heat map and parallel coordinate plot	Integration-based synthesis using three parameters (time, topological relations, and per machine)	Overview of situation awareness for entire network plus detailed views of specific machines on minute-by-minute basis,

As summarized in Table 4, the different visual structures used resulted in the detection of different classes of attacks from high-noise, low-noise, machine-specific, and so forth. Also, the synthesis of the visual structures impacted the range of attacks that were detected using visualizations. Team 5, which integrated the visual structures using the most parameters, was able to provide both high-level and fine-grained analysis of cyber events in the network. This case again illustrates the need to consider how the use of proximity-based synthesis of visual structure results in significantly different situation awareness than using integration-based synthesis. Also, increasing the number of parameters for integration seems to result in more complete situation awareness.

C. Case #3: Terrorist Plot

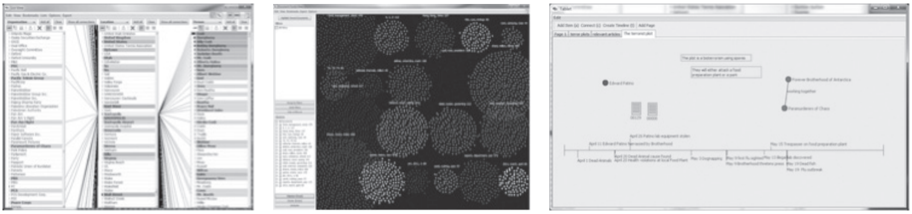
Teams were tasked to analyze a corpus of documents (n=4,474). Each team created different diagrams from the raw document data to support their policy strategy construction of possible terrorist activity. The goal was to identify all documents relevant to an actual terrorist plot (13 total). For their visual data structure, Team A created a node-link diagram that interconnected related clusters of documents as shown in Figure 11 below.

FIGURE 11: TEAM A VISUAL STRUCTURE – NODE-LINK DIAGRAM OF DOCUMENT CLUSTERS


temporal ordering
main thread
background documents
entities and documents both contributing to the visual representation of the story
possibly related
space used to visually separate different threads of the story

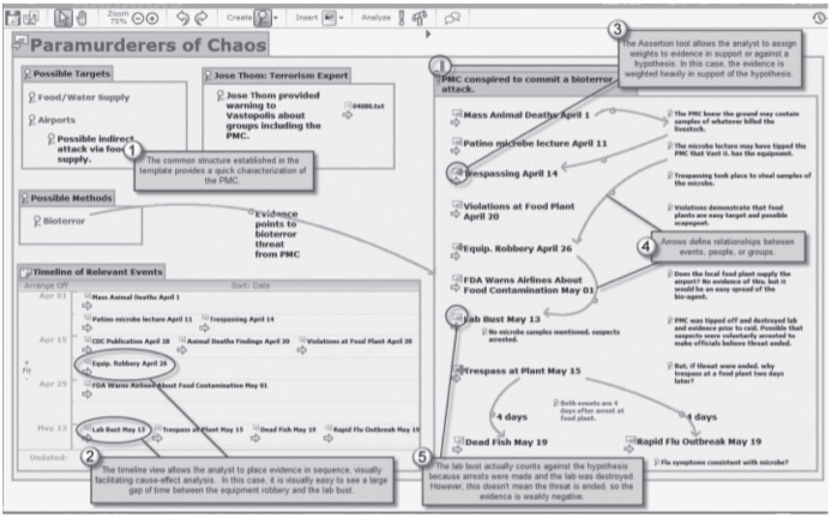
The document clusters were created using entity extraction and a vector-space model, to build graphs of both sentence-based and document-based co-occurrence, as well as document-neighbor discovery. Based on the extractions, the documents were then examined for items of interest. Key entities and phrases were temporally arranged based on related themes, entities and events for further analysis. Using a visualize structure which synthesized data using a compartmentalized approach, Team A correctly identified five out of thirteen documents needed for constructing a reliable policy strategy.

FIGURE 12A,B & 13: TEAM B VISUAL STRUCTURE – 3 OF 5 VIEWS



As partially illustrated in figs. 12a,b and 13, Team B created a total of five visual data structures from the raw document data after first extraction and manually cleaning: 1) a list view, 2) cluster view, 3) document view, 4) calendar view, and 5) timeline view. The list view grouped related entities, while the cluster view grouped related documents. The document view enabled detailed exploration of related documents using a tag cloud to navigate the document set. The calendar view ordered documents identified as suspicious according the dates associated with the documents, while the timeline ordered the notes from the analysis according to relevance and order of occurrence. Using a hybrid-synthesis of compartmentalized visual structures and simple, integrated structures based on the time parameter, Team B correctly identified 11 of 13 documents needed for constructing their policy strategy. However, they also included in the solution one false lead and three isolated incidents unrelated to the imminent threat.
In contrast, Team C used indented lists inside an integrated visual structure that laid out multiple timelines within different hypothesis-driven story lines, which resulted in a nested visual framework approach. Their synthesis of visual structure enabled them to organize the specifiers and labels of indented lists representing the raw data. They preprocessed the data using both custom and standard dimensions for extracting and clustering documents of interest. They manually reviewed these documents, and manually extracted information of interest to create an initial timeline view.

FIGURE 14: TEAM C VISUAL STRUCTURE SYNTHESIS – EXAMPLE NESTED FRAMEWORK


Zoom 75%
Create
Insert
Analyze
Paramurderers of Chaos
Possible Targets
Food/Water Supply
Airports
Possible indirect attack via food supply.
Jose Thom: Terrorism Expert
04080.txt
Possible Methods
Bioterror
Evidence points to bioterror threat from PMC
Timeline of Relevant Events
Arrange Off
Sort: Date
Apr 01
Apr 15
Apr 29
May 13
Fit
Undated:
Mass Animal Deaths April 1
Patino microbe lecture April 11
Trespassing April 14
CDC Publication April 18
Animal Deaths Findings April 20
Violations at Food Plant April 20
Equip. Robbery April 26
FDA Warns Airlines About Food Contamination May 01
Lab Bust May 13
Trespass at Plant May 15
Dead Fish May 19
Rapid Flu Outbreak May 19
PMC conspired to commit a bioterror attack.
4 days
4 days
1
2
3
4
5

Jose Thom provided warning to Vastopolis about groups including the PMC.

1 The common structure established in the template provides a quick characterization of the PMC.

2 The timeline view allows the analyst to place evidence in sequence, visually facilitating cause-effect analysis. In this case, it is visually easy to see a large gap of time between the equipment robbery and the lab bust.

3 The Assertion tool allows the analyst to assign weights to evidence in support or against a hypothesis. In this case, the evidence is weighted heavily in support of the hypothesis.

4 Arrows define relationships between events, people, or groups.

5 The lab bust actually counts against the hypothesis because arrests were made and the lab was destroyed. However, this doesn't mean the threat is ended, so the evidence is weakly negative.

The PMC knew the ground may contain samples of whatever killed the livestock.

The microbe lecture may have tipped the PMC that Vast U. has the equipment.

Trespassing took place to steal samples of the microbe.

Violations demonstrate that food plants are easy target and possible scapegoat.

Does the local food plant supply the airport? No evidence of this, but it would be an easy spread of the bio-agent.

No microbe samples mentioned, suspects arrested.

PMC was tipped off and destroyed lab and evidence prior to raid. Possible that suspects were voluntarily arrested to make officials believe threat ended.

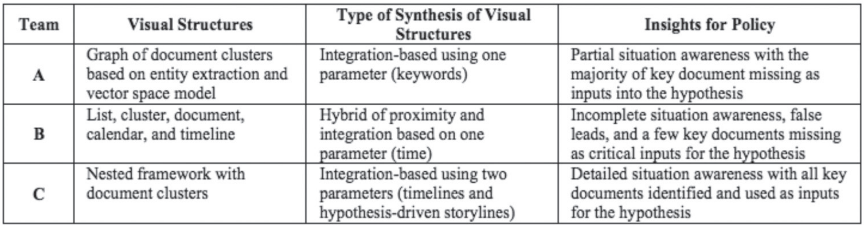
But, if threat were ended, why trespass at a food plant two days later?

Both events are 4 days after arrest at food plant.

Flu symptoms consistent with microbe?

Separate timelines were then created for several competing hypotheses coupled with related documents. One of the hypotheses of interest was selected for further development. The selected hypothesis was used as the framework to integrate timelines, warrants, sub-claims that supported the hypothesis, i.e., associated extractions and clusters from the processed raw data were tied to specific hypotheses, which was organized according to entities. Figure 14 shows such a synthesis for the entity "Paramurderers of Chaos." In other words, Team C synthesized the visual structure using entities as the primary parameter for integration, and timelines within hypothesis-driven storylines components (targets, expert perspective, methods, warrants) as additional sub-parameters for data integration. Team C correctly identified all thirteen documents needed for constructing their policy strategy.

TABLE 5: CASE #3: VISUAL STRUCTURE SYNTHESIS FOR POSSIBLE TERROR ATTACK

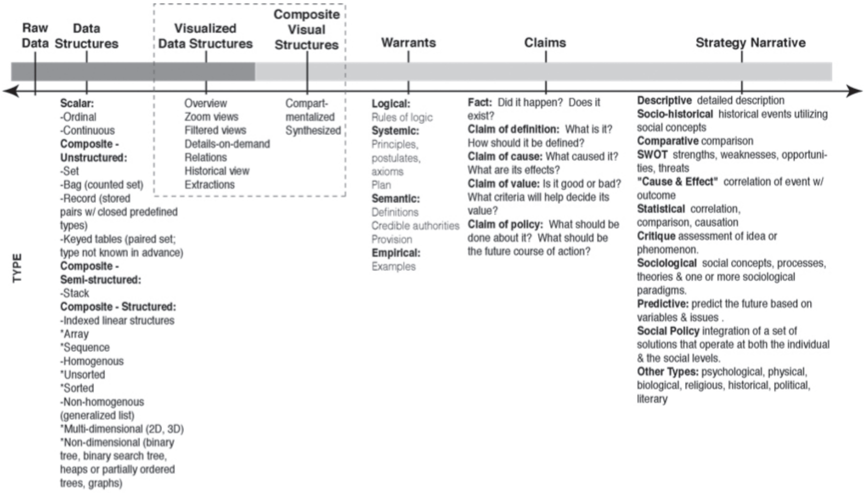


Team	Visual Structures	Type of Synthesis of Visual Structures	Insights for Policy
A	Graph of document clusters based on entity extraction and vector space model	Integration-based using one parameter (keywords)	Partial situation awareness with the majority of key document missing as inputs into the hypothesis
B	List, cluster, document, calendar, and timeline	Hybrid of proximity and integration based on one parameter (time)	Incomplete situation awareness, false leads, and a few key documents missing as critical inputs for the hypothesis
C	Nested framework with document clusters	Integration-based using two parameters (timelines and hypothesis-driven storylines)	Detailed situation awareness with all key documents identified and used as inputs for the hypothesis

4. IMPLICATIONS FOR CYBER POLICY INSIGHT

These case studies demonstrate a gap in our understanding of composite visual data structures, and how their synthesis can drastically reduces or illuminates the direction of policy strategy. As illustrated by the first case study, the *cyber policy* strategies we are able to see depends on how visual structures are used to synthesize data. E.g., for Case 1 Epidemic, the team used an *integrated* approach rather than a proximity approach, and was thus able to compile a more complete situation awareness to inform action. In Case 2 Cyber Attacks, the team using the *most parameters* for synthesizing the visual structures was able to identify the broadest range of attacks on the corporate network. And in Case 3 Terrorist Plot, the team used a nested framework to support a *narrative-based integration* parameter and was able to construct the most reliable hypothesis to inform situation awareness. The key implication for cyber policy is that these case studies point toward a critical need to further investigate how visual structures are synthesized and how they inform policy action.

We offer the following spectrum of information structures as a starting point in figure 15 below. Based on the Toulmin argument structure, this spectrum represents an initial chain-of-connections from data to policy strategy/narrative.



Raw Data
Data Structures
Visualized Data Structures
Composite Visual Structures
Warrants
Claims
Strategy Narrative
TYPE
Scalar:
-Ordinal
-Continuous
Composite - Unstructured:
-Set
-Bag (counted set)
-Record (stored pairs w/ closed predefined types)
-Keyed tables (paired set; type not known in advance)
Composite - Semi-structured:
-Stack
Composite - Structured:
-Indexed linear structures
*Array
*Sequence
-Homogenous
*Unsorted
*Sorted
-Non-homogenous (generalized list)
*Multi-dimensional (2D, 3D)
*Non-dimensional (binary tree, binary search tree, heaps or partially ordered trees, graphs)
Overview
Zoom views
Filtered views
Details-on-demand
Relations
Historical view
Extractions
Compart-mentalized
Synthesized
Logical:
Rules of logic
Systemic:
Principles, postulates, axioms
Plan
Semantic:
Definitions
Credible authorities
Provision
Empirical:
Examples
Fact: Did it happen? Does it exist?
Claim of definition: What is it? How should it be defined?
Claim of cause: What caused it? What are its effects?
Claim of value: Is it good or bad? What criteria will help decide its value?
Claim of policy: What should be done about it? What should be the future course of action?
Descriptive detailed description
Socio-historical historical events utilizing social concepts
Comparative comparison
SWOT strengths, weaknesses, opportunities, threats
"Cause & Effect" correlation of event w/ outcome
Statistical correlation, comparison, causation
Critique assessment of idea or phenomenon.
Sociological social concepts, processes, theories & one or more sociological paradigms.
Predictive: predict the future based on variables & issues .
Social Policy integration of a set of solutions that operate at both the individual & the social levels.
Other Types: psychological, physical, biological, religious, historical, political, literary

FIGURE 15: CHAIN-OF-CONNECTIONS FROM RAW DATA TO NARRATIVE

This chain-of-connection begins with "raw data" and enumerated "data structures" [1] because they form the basis of policy actions. Acting as a link, "visual structures"—both individual and composite—create a bridge between data and the "warrants" and "claims" that comprise "policy strategy/narrative". The dashed box outlining both types of visual structures highlight their importance in shaping our understanding of situation awareness for policy action. Currently, most cyber policy is informed by visualized data structures rather than composite visual structures that support higher-order information structures enumerated along the blue bar in Figure 15.

The implications for cyber policy are several. First, there is a critical to investigate how visual structures can help synthesize the information needed to inform policy decisions, which tend to fall into three categories: standard, irregular, and emergency. Decisions that are "Standard" are routine decisions where procedures are well-established, and historical data is likely available. In contrast, "Irregular" decisions that are outside the routine, but not urgent, while decisions that are "Emergency" are both irregular and time-sensitive. Identifying visual structures could help reduce the complexity of information for each of these three different types of policy decisions. That is, these patterns would facilitate both short and long-term analytics of policy actions based on data as well as provide alternate perspectives in understanding future decision-making. In other words, these visual patterns could also help streamline the information flow process in organizations by connecting policy strategy from the past with future decisions to be made.

However, there are caveats in pursuing these visual structures for information synthesis, which is illustrated by a case study for the Federal Chancellery of a European country [13]. The first caveat is that while visualization of information is important, it is only useful if it is integrated in an information flow process that is part of the decision-making. The second caveat is that simplicity in the visualization supported decision-making much more than complex ones.

Organizations need an innovative approach that 1) efficiently conveys policy prescriptions and 2) provides mechanisms for synthesizing these prescriptions with recommendations for policy actions in organizations [3]. One approach, which we will investigate in future work, is to develop patterns of cyber policy[1] in organizations, which can be visualized for the three different categories of decisions: standard, irregular, and emergency. Identifying and developing policy patterns would enable policy to be efficiently conveyed and provides a framework for synthesizing policy information in organizations.

A number of different patterns for cyber security have been developed for attacks, forensics, vulnerabilities, and user behavior. However, patterns of visual structures for cyber policy in organizations have not been the focus of cyber security research beyond complex text-based prescriptions. Visual policy patterns for organizations would be novel, but rely on the proven success of using visualization for cyber security. These patterns of visual structures could help organizations move beyond incremental security and towards innovative management of policy for issues like unintentional insider threats.

In future work, our plan is to develop a complementary framework to Kosslyn's visual structure to analyze the information content conveyed by visual structures. Having this dual-framework of visual structure and information content could enable policy makers to better assess the data foundation of their strategy and to consider alternate perspectives offered by differently structured visualizations.

REFERENCES

[1] Dale, N., and Walker, H. A classification of data types. Computer Science Education 3.3 (1992): 223-232.

[2] Harvey, M., Long, D., Reinhard, K. Visualizing NISTIR 7628, Guidelines for Smart Grid Cyber Security. Power and Energy Conference at Illinois (PECI), 2014.

[3] Herath, T., Rao, H.R. Encouraging Information Security Behaviors in Organizations: Role of penalties and perceived effectiveness. Decision Support Systems 47 (2009) 154-165.

[4] Hossain, M. et al. 2012. Storytelling in Entity Networks to Support Intelligence Analysts. Proceedings of the ACM Conference on KDD'12, Beijing, China.

[5] Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., Möller, T. 2013. A Systematic Review on the Practice of Evaluating Visualization. IEEE TVCG, Oct. 2013.

[6] Johnson, Eric; Goetz, Eric. 2007. Embedding Information Security into the Organization. IEEE Security & Privacy, May/June 2007.

[7] Kosslyn, S. M., 1989. Understanding Charts and Graphs." Applied Cognitive Psychology, Vol. 3, 185-225.

[8] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S. Empirical Studies in Information Visualization: Seven Scenarios. 2012. IEEE TVCG, 18(9):1520-1536, Sept. 2012.

[9] Lowrance, J. D., Harrison, I. W. & Rodriguez, Andres C. 2000. Structured Argumentation for Analysis. Proc. of 12th Int'l Conf. on Systems Research, Informatics, and Cybernetics, Baden-Baden, Germany, pp. 47-57.

[10] North, C. 2006. Toward measuring visualization insight. IEEE CGA, 26(3):6-9, May/June 2006.

[11] Reed, C. and Rowe, G. 2004. Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools 13.04: 961-979.

[12] Samonas, S. Originally presented at 12th Annual Security Conference, Keynote Lecture. 11 April 2013, Las Vegas, Nevada. This version available at: http://eprints.lse.ac.uk/50344/

[1] A simple example of a policy pattern is password generation required four different character sets. A more complex example is the "Baseline Security" provided by the German government or "Common Criteria", an international standard for cyber security.

[13] Stoll, J., Siemssen, J., Bengez, R. Visualization, Insider Cyber Threats & Legal Informatics. To be published in the Proceedings of the ACM Austria: Internationales Rechtsinformatik Symposion (IRIS 2015).

[14] Toulmin, S. E. (2003). The uses of argument. Cambridge University Press.

[15] Visual Analytics Science & Technology, 2011. www.visualanalytics.com

[16] Wickham, H., Cook, D., Hofmann, H., Buja, A. 2010. Graphical Inference for Information Visualization. IEEE TVCG, Vol. 16, No. 6.

[17] Yi, JS., Kang, Y., Stasko, J., Jacko, J. 2008. Understanding and Characterizing Insights: How do People Gain Insights Using Information Visualization? BELIV, Florence, Italy.

[18] Ziemkiewicz, C., Gomez, S., Laidlaw, D. 2012. "Analysis Within and Between Graphs: Observed User Strategies in Immunobiology Visualization." CHI'12, Austin, TX.

[19] Verizon. 2014. Data Breach Investigations Report. Accessed: www.verizonenterprise.com/DBIR/2014