CAMUS: AUTOMATICALLY MAPPING CYBER ASSETS TO MISSIONS AND USERS

John R. Goodall, Anita D’Amico and Jason K. Kopylec
Secure Decisions Division of Applied Visions, Inc.
Northport, NY

ABSTRACT

This research advances Cyber Situation Management by proposing methods for automated mapping of Cyber Assets to Missions and Users (Camus). To enable accurate and efficient cyber incident mission impact assessment, a Camus ontology that defines entities, relationships and attributes (ERAs) associated with them has been drafted. Methods for fusing data from multiple data sources have been developed alongside an ontology-based system to populate the model using existing network data sources. The Camus system demonstrates how commonly available data sources can be rapidly collected, correlated, and fused to automatically map cyber assets to the users who depend on them, to the missions they support, and to the services they provide. Also discussed are the technical architecture and challenges to such an approach.

INTRODUCTION

To effectively remediate a cyber asset compromise, analysts need to clearly understand the relationships between the compromised asset and the affected missions and users. If all that is known about a compromised host is its IP Address, there is no evidence to project the cascading effects. Today’s network analysts have a limited view of the roles cyber assets play in the overall enterprise. Without this information, an analyst cannot accurately prioritize and assign resources to perform remediation.

The objective of this research is to improve cyber situation management by developing an automated mapping of Cyber Assets to Missions and Users (Camus) that facilitates accurate and efficient Cyber Incident Mission Impact Assessment (CIMIA) [9]. Central to the effort is the development of the Camus system, capable of automated relationship discovery between cyber assets, missions and users. To derive needed contextual information in an automated way, semantic web concepts are applied to model and automatically fuse the needed information. The Camus system integrates a number of common network feeds demonstrating how existing data sources can be used in new ways to provide contextual mission information.

RELATED WORK

Much of the grounding for Camus comes from Salerno’s [16][15] Air Force Situation Awareness Model (AFSAM). This model describes the path that data takes to become information that can be consumed by analysts for improved situation management. Of most interest to Camus is the portion of the AFSAM labeled as “knowledge of us,” which provides contextual information about the operational environment. Tadda et al. [17] refined the general AFSAM and applied it directly to the cyber domain, resulting in the Cyber Situation Awareness (SA) Model. Within the Cyber SA Model, the “knowledge of us” required for situation management is an accurate understanding of how operations are impacted when there are degradations and compromises in the cyber infrastructure.

The aim of Camus is to provide this information in a continually up-to-date, automated and scalable way that is usable by both human analysts as well as other Cyber SA systems. The work of Holsopple et al. [11], also grounded in the Cyber SA model, develops a Virtual Terrain that models the network and, to a limited extent, takes mission context into account. Problematically, their mission-related information must be manually added by analysts. It is this manual data entry that Camus attempts to automate.

Grimaila and Fortson [9] shift the focus on situation management away from *cyber* assets and instead to *information* assets. They discuss that what is truly valuable is the information that resides on the hardware and its confidentiality, integrity and availability. They propose a Cyber Damage Assessment Framework that requires the manual definition and prioritization of both operational processes and information assets. Bryant and Grimaila [3] show that there are a number of pitfalls when collecting information-centric data and that much of it is unavailable electronically. The Camus approach can greatly aid the collection and automation of information assets, although this is currently left to future work.

Work by Gomez et al. [8] in the domain of sensor-mission assignment applied a similar approach to Camus in the area of automated assignment of intelligence,

978-1-4244-5239-2/09/$26.00 ©2009 IEEE

surveillance and reconnaissance (ISR) assets to specific military missions. Their Missions and Means Framework (MMF) ontology closely parallels the Camus ontology including concepts such as missions, operations, tasks, capabilities and systems. Lewis et al. [13] propose their own mission reference model and are tackling the mapping of cyber assets to missions from a mathematical constraint satisfaction approach. What Lewis et al. does not comment on is the practical matter of collecting and fusing the data needed to support their mathematical models.

MOTIVATION

Bargar [1] describes a need for improved cyber situation management that is based on a shared understanding between mission commanders and network analysts about how compromises to cyber assets will affect mission essential functions (MEFs). Network security management systems do little to facilitate this common operating picture. Today's cyber defenders often have little contextual information about a compromised asset beyond its IP address and an Intrusion Detection System (IDS) alert description. Knowing only an IP Address, the affected machine could be the desktop belonging to a building manager that maintains an inventory of cleaning supplies. On the other hand, it could be a file server that supports time-critical communication between commanders in theater. In order to properly respond to cyber compromises, analysts need to know who uses the compromised asset and what it is used for. Only then can its criticality be determined and appropriate countermeasures be taken. In the case of the building manager's laptop, perhaps no action is needed, whereas for the critical file server, the information that was lost might have critical effects to the success of supported missions.

The primary operational obstacle is a lack of existing data sources that accurately map cyber assets to the missions they support. Even if cyber assets' functions are documented when initially put on the network, that information quickly becomes obsolete as the network is reconfigured over time. In current operations, mapping cyber assets to missions and users is a manual, time-consuming, error-prone, and expensive process, so it is rarely attempted.

Even if manual methods are employed, often the actual use of the network in operation is much different than its original architecture. Adding to the difficulty is that the networks' interdependencies are so numerous and complex that comprehending the mappings is impossible without proper formatting and display.

An optimal solution to these problems should provide the needed information to assess the mission impact of cyber incidents. The cyber asset to mission mappings should be trusted and accurate, maintaining provenance to trace back to original sources. Moreover, the information should be targeted to the particular role of the user. For example, the information needed by a commander to evaluate the go/no-go status of his missions is very different from the picture needed by a network analyst to determine how to improve redundancy and resiliency of the enterprise network. The commander needs a deep understanding of the missions he oversees and is less interested in the bits and bytes of the underlying computer network. Conversely, the network analyst needs a detailed knowledge of how the network is configured and running; mission and task-related information is only used to determine asset criticality and to ensure that the supporting infrastructure is in place and working properly. An optimal automated solution should be flexible to this variation in role-based granularity.

An even tougher challenge is to assign dependencies and criticality metrics to the relationships. It is one thing to say that a particular file server is used during a mission by a particular person. A much deeper knowledge about mission requirements is needed to determine automatically that the file server is *critical* and *depended on* for successful execution.

TECHNICAL APPROACH

The primary goal of the Camus technical solution is to meet these challenges and provide the needed context to support automated mission impact assessment. Armed with such a technology, the critical role that cyber assets have in mission success can be better understood. Beyond these research ideals, there are also practical requirements that Camus should be relevant and operationally feasible in today's large and dynamic networks. The Camus approach is grounded in the idea that the needed data does exist in digital format, but is in disparate locations and formats. Hence, much of the exercise to derive asset and mission relationships becomes a data mining, inference, and fusion task.

The Camus system relies on an ontology-based semantic approach to data integration and fusion, similar to the concepts discussed in Yoakum-Stover and Malyuta [19]. The ontologies were designed with subject matter experts in terms of *entities*, *relationships* and *attributes* (ERAs). The resulting ERAs were then translated into semantic ontologies, using the methodology in Fahad [6].

To build the Camus technology solution, a system architecture has been developed along with a software platform based on concepts from the semantic web. The semantic web uses ontologies as a structured representation of ERAs of a domain. The Camus system uses common semantic web tools Protégé and the XML-based Web Ontology Language (OWL) to represent its ontologies. Figure 1 graphically represents the high-level core of the Camus ontology.

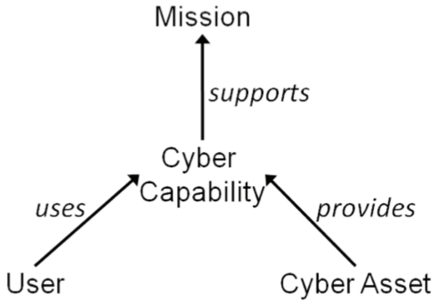


Figure 1. High-level Camus ontology mapping cyber assets to missions and users.

The core Camus ontology depicts the semantic relationships between missions, cyber capabilities, users and cyber assets. To implement the Camus system, this core ontology is extended to encompass the level of data granularity needed for a particular operator role. For example, a network analyst needs details about cyber assets, including the applications, ports and usernames associated with devices, but may need only basic descriptions of missions and their criticality. Conversely, a commander may need only basic information about assets, but require details about the essential and specified mission tasks that are under his control.

Camus' core ontology must be populated by data sources that provide the needed information. However, instead of populating the ontology all at once, the Camus technical approach involves translating data feeds that supply basic relationships for small portions of the Camus ontology, piece-by-piece. Even if portions of the ontology cannot be populated from available data, the logical reasoning capabilities implicit in OWL can infer indirect relationships and fill in this missing information.

To illustrate the approach, suppose a network administrator needs to know how users' workstations are associated with an organization's departments, and he has access to only two data sources: FTP logs and an LDAP server. The FTP logs and LDAP query results contain items that look like the sample in Table 1.

Table 1. Sample data records for FTP Logs and DNS Dumps, which include user and cyber asset information.

FTP Log	LDAP query
… jsmith@100.10.20.4 … … sjones@100.10.20.6 … … llaurel@100.10.20.9 … …	…jsmith|Logistics… …llaurel|Adminstrative… …sjones|Finance… …

To model this information, a *base ontology* is created that contains entities [User], [Department] and [IPAddress]. Added to that are semantic relationships to connect the entities: [User *isMemberOf* Department], [Workstation *isUsedBy* User] and [Workstation *supports* Department]. The relationship that the analyst really wants is the last one, [Workstation *supports* Department], but it is not explicit in the data sources. To derive this information, the data sources are used to instantiate the relationships they do represent explicitly, such as [100.10.20.4 isUsedBy jsmith] and [jsmith isMemberOf Logistics]. Note that each of the data sets is only responsible for the relationships it can provide. Using [User] as an *alignment point*, i.e. an entity that is common between relationships, the semantics of ontology can now infer the relationship that the analyst is really interested in, [100.10.20.4 supports Logistics], as shown in Figure 2.

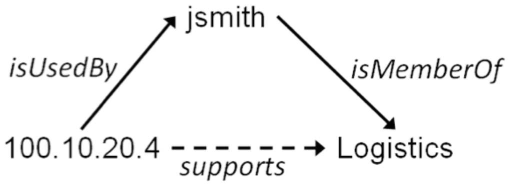


Figure 2. A sample instantiation of an ontology using two separate data sources. The files have been directly translated into semantic relationships (solid lines), allowing for automated inference of the relationship of interest [100.10.20.4 isUsedBy Logistics] (dashed line).

This capability is powerful because many data sources can be used to populate small portions of a base ontology and shared concepts become alignment points for fusing the data. The translation from raw data sources into the ontology model can be done in one of three ways:

- *Direct Translation* - Relationships can be drawn directly from the data; for example *...jsmith@100.10.20.4...* can be directly converted into [jsmith uses 100.10.20.4].
- *Inferred Translation* - Relationships can be assigned from data using heuristics or statistical methods. For

example, to assign the relationship [User dependsOn Workstation] if there is no data source that represents this dependency explicitly, a heuristic rule can be applied to infer that if a user always logs on to the same workstation, then that user *dependsOn* that workstation.

- *Ontology-to-Ontology Translation* - If the data source has its own model or schema, e.g. Microsoft Operations Framework (MOF) or the Universal Joint Task List (UJTL), entities can be aligned at the ontology level by defining alignment points; instances of one ontology are then automatically treated as corresponding instances in the second.

These techniques define a process for *ontology fusion*, bringing together disparate network data sources to define and infer mappings between cyber assets, mission and users. By modeling the needed information in an ontological format, data fusion happens automatically (see Boury-Brisset [2]). The results can then be coupled with other SA systems to provide programmatic access to the mission mappings and provide role-based information visualization views that depict the needed information.

ARCHITECTURE

Much of the development effort has been to build a system that implements the semantic functionality and situation awareness described in the Camus technical approach. The Camus system integrates a number of technologies that are used in the biological sciences and digital content management domains, including OWL, Jena, Lucene, Protégé, Servlet-based APIs and web based visualizations. Base ontologies are modeled in Protégé and exported as XML-based OWL files. These base ontologies are then coupled with easily understood JavaScripts that define where to find and how to translate available data sources into an instantiation of the base ontology. Within Camus, the base ontologies and their corresponding scripts are referred to as *ontology fuselets*. Multiple ontology fuselets can be combined into a master fusion ontology that defines the alignment points among the base ontologies. Figure 3 shows the conceptual system architecture, which consists of three main components: Data Integration, Information Fusion and Knowledge Management.

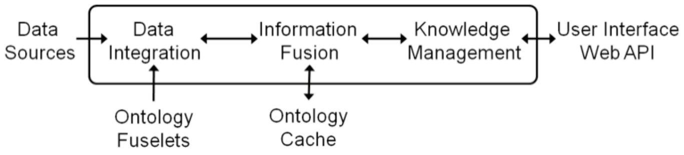


Figure 3. Camus system architecture overview.

The Data Integration module takes the ontology fuselets as input and uses them to parse raw data sources into ontology instances. The data sources can come in an array of formats such as semi-structured text files, databases or processes that publish alert or log information over the network. Once data sources are translated into an ontology representation, they are passed to the Information Fusion module which uses the alignment points in the master fusion ontology. It semantically couples the individual pieces together, performing automated data fusion. The resulting fully instantiated and fused ontology contains a complete mapping of relationships between cyber assets, missions and users. To handle issues of ontology scalability, the Camus system implements ontology caching mechanisms based on the concepts of Karnstedt et al. [12]. The complete ontology can be extremely large, so it is persisted using high-performance external Resource Description Framework (RDF) stores and indexes. This cache provides fast and robust querying mechanisms, which are accessed by the third module, Knowledge Management.

The Knowledge Management module consists of two parts, web APIs to access the ontology programmatically, and visualization capabilities for displaying mission context information directly to operators. The programmatic APIs are web based for easy coupling to outside systems, so that the mission context information contained in the ontology cache can be made available to external Cyber SA systems. The user interface provides a point-and-click visual interface to Camus information. When an item of interest is clicked, such as an IP Address, a client-side query is created. The client-side query is parsed by the Knowledge Management component and passed to the Information Fusion engine. The cache is consulted and updated, if needed, by the Data Integration module, which uses the ontology fuselets to retrieve the needed information from the original data sources.

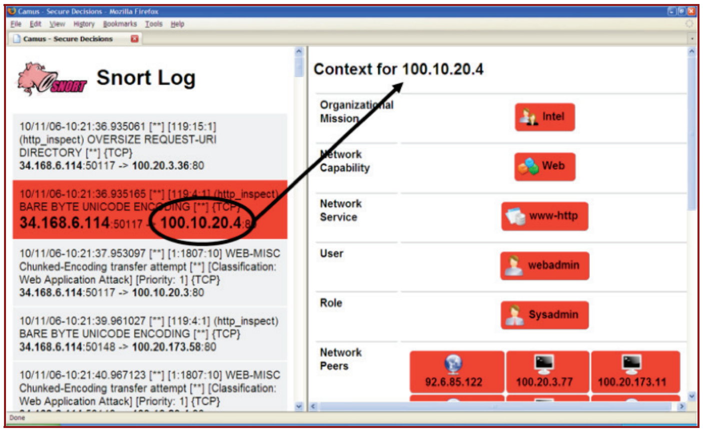
The mission context results are finally packaged up as OWL or GraphML and returned to the client via HTTP for further manipulation and/or display to the operator. The returned ontologies can also include visual attributes to aid visualization, similar to Rahman et al. [14]. Carroll [4] presents a number of applicable role-focused views for displaying visual mission hierarchies and the supported infrastructure. A number of performance enhancements have been made to the fusion engine and cache to ensure that user queries are returned within reasonable time to keep the user interface running smoothly and to parse large data sources rapidly.

APPLICATION

To illustrate the approach, a Camus prototype was developed that displays mission context information coupled to an existing intrusion detection system. If an organization has access to detailed electronic mission planning specifications, they can be easily integrated into the Camus system. In practice though, the network defense community may not have access to mission specifications and tasks, due to accessibility restrictions and classification. So the Camus system uses the enterprise organizational chart as an adequate baseline for representing users' roles and organizational missions. The organizational chart is an easily accessible, usually unclassified and regularly updated document which maps people to their roles, departments and superiors. However, the organization chart does not show, nor does any other common network data source, direct mappings of cyber assets to specific organizational missions. To determine this information, we use network data, like user logs and network traffic, to infer cyber asset-to-mission relationships. For example, if the system can deduce how users support the organization and also which machines they regularly access, it can show a reasonable approximation of how a compromise of those machines may affect the organization. Here [Users] are an alignment point for inferring cyber asset-to-mission dependencies.

To demonstrate the Camus system capabilities, an existing network security data set of network traffic and system logs were mined for asset, mission and user related data. LDAP provided a list of users, their roles and departments. FTP and Unix logs were processed to determine the logical network topology and user social network. These host-to-host communication networks provided information such as which machines regularly use a particular mail server. Armed with these basic data sources – LDAP, NetFlow traffic and user logs – fuselets were created for each, as well as a fusion ontology to align the common features, such as username and IP Address.

The ontology fuselets automatically parse the source data to populate and store the mission ontology. With the populated model mapping cyber assets to missions and users, the next step was to demonstrate how that information can be used to provide improved mission context to analyze and remediate cyber asset compromises. A web application was built to display Snort intrusion detection alerts, the IP Addresses referenced in the alerts, and the links of those IP addresses to the mission context identified by Camus. When the user clicks on an IP Address of interest, the Camus fusion engine consults the cache and parses any needed raw data files on the fly. The results are formatted in HTML and returned back to the client as easily understandable graphics. These are displayed to the user through the browser, providing on-demand mission context information. This is shown in Figure 4, in which the user has clicked on an IP (100.10.20.4, circled) in the list of alerts on the left pane and the contextual information is displayed on the right, including the operational mission, capability, network service, users and roles, and network peers associated with the selected IP address. This system is primarily a web server using the http network service that supports the Intel mission. The webadmin user has a relationship to this system; this user is likely the administrator (role Sysadmin) of the web server. There are several peer systems, some of which are other internal servers, internal clients, or external systems.


Camus - Secure Decisions - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Camus - Secure Decisions
SNORT
Snort Log
Done

```
10/11/06-10:21:36.935061 [**] [119:15:1]
(http_inspect) OVERSIZE REQUEST-URI
DIRECTORY [**] {TCP}
34.168.6.114:50117 -> 100.20.3.36:80

10/11/06-10:21:36.935165 [**] [119:4:1] (http_inspect)
BARE BYTE UNICODE ENCODING [**] {TCP}
34.168.6.114:50117 -> 100.10.20.4:8

10/11/06-10:21:37.953097 [**] [1:1807:10] WEB-MISC
Chunked-Encoding transfer attempt [**] [Classification:
Web Application Attack] [Priority: 1] {TCP}
34.168.6.114:50117 -> 100.10.20.3:80

10/11/06-10:21:39.961027 [**] [119:4:1] (http_inspect)
BARE BYTE UNICODE ENCODING [**] {TCP}
34.168.6.114:50148 -> 100.20.173.58:80

10/11/06-10:21:40.967123 [**] [1:1807:10] WEB-MISC
Chunked-Encoding transfer attempt [**] [Classification:
Web Application Attack] [Priority: 1] {TCP}
```

Context for 100.10.20.4

Organizational Mission	Intel
Network Capability	Web
Network Service	www-http
User	webadmin
Role	Sysadmin
Network Peers	92.6.85.122 100.20.3.77 100.20.173.11

Figure 4. Camus user interface displaying mission context information for an attacked cyber asset.

In preparing the system, there were a number of run-time performance challenges. These were focused in two areas: 1) scaling the amount of data that can be parsed and cached and 2) ensuring that the user experience was fast enough for normal browsing. The corpus of NetFlow traffic used in the demonstration has over ten thousand unique IP Addresses and is over one gigabyte in size. A number of high performance indexing and custom filtering mechanisms were implemented to increase throughput. The Camus fusion engine was able to parse and fuse all of this data efficiently using a commodity laptop and small memory footprint. The web interface reacts within normal web browsing response times that range between near instantaneous for simple queries to less than ten seconds for very complex queries.

From the Camus user interface, it is easy to see how attacked assets support specific users, portions of the organization and other cyber assets. For example, it was immediately apparent which departments authenticate to a particular domain server and which web servers were most widely used by external hosts. The system successfully meets the requirements of providing flexible and rapid integration of a number of disparate data sources to automatically map cyber assets to the missions and users they support.

FUTURE WORK

The next steps are to augment the existing Camus system with more sophisticated network management data sources, such as configuration management databases (CMDB) and cyber security monitoring systems. In addition, our research team is working on an expanded mission ontology that uses military planning standards, such as UJTL, to better model tasks and missions.

The Camus system will also store additional provenance information for the inferred relationships. This provenance enables dynamic drill-in to see original data sources that were used to create an asset-mission mapping. This provenance will be provided so an analyst using Camus can corroborate findings and improve trust and reliability in the system.

Capabilities to assign metrics to the relationships will also be added. This will facilitate any future dynamic computation of metrics such as mission criticality and redundancy. Chew et al. [5] discuss how network security metrics should be aligned closely with the missions of an agency. Once these metrics are captured for portions of the mission ontology, they can be propagated throughout the model using conditional probability methods. These metrics can include calculations of risk, similar to the work of Watters et at. [18], which proposes methods for calculating risk based on cyber asset mission dependencies.

Finally, continued improvements will be made to the Camus system API and user interface. Adequate visualization of large ontologies is an open and active area of research and development throughout the semantic web community.

CONCLUSION

Camus advances the state-of-the-art in situation management by providing essential ‘knowledge of us’ to the Cyber SA Model. Methods for automatically mapping cyber assets to the mission and users that rely on them were discussed. This capability is essential for assessing the mission impact of cyber incidents and will play an increasingly crucial role as cyber operations become more pervasive. The Camus technical approach uses ontology fusion and emerging semantic web tools to parse disparate data sources into a unified model of domain entities, relationships and attributes. Three methods were explained for converting available data sources into a semantic ontology representation, namely direct, inferred and ontology-to-ontology translation. The Camus software solution builds on these methods to bridge the gaps between data, information and knowledge. The resulting system provides context directly to analysts or to other cyber situation awareness systems. In addition, it displays mission critical information to users, improving overall situation awareness. The Camus system has been demonstrated with readily available data sources and found it to be both operationally grounded and reasonably scalable. Overall, Camus illustrates that practical, accurate, and automated cyber mission impact information is within reach throughout the cyber defense and network management community.

ACKNOWLEDGEMENTS

This research and development effort is supported by OSD and AFRL through a Small Business Innovative Research grant, under contract no. FA8750-08-C-0166.

REFERENCES

[1] Bargar, A. *DoD Global Information Grid Mission Assurance*. CrossTalk: The Journal of Defense Software Engineering, July, 2008.

[2] Boury-Brisset, A-C. *Ontology-based approach for information fusion*. Information Fusion, 2003.

[3] Bryant, A. and Grimaila, M. *Developing a Framework to Improve Information Assurance Battlespace Knowledge*. International Conference on Information Warfare and Security (ICIW), 2007.

[4] Carroll, S. *Mission Impact Analysis Visualization for Enhanced Situational Awareness*. MA Thesis, Air Force Institute of Technology, 2008.

[5] Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A. and Robinson, W. *Performance Measurement Guide for Information Security*. NIST Special Report 800-55, 2008.

[6] Fahad, M. *ER2OWL: Generating OWL Ontology from ER Diagram*. Intelligent Information Processing. Springer, 2008.

[7] Fortson, L. and Grimaila, M. *Development of a Defensive Cyber Damage Assessment Framework*. International Conference on Information Warfare and Security (ICIW). 2007.

[8] Gomez, M. et al. *An Ontology-Centric Approach to Sensor-Mission Assignment*. Knowledge Engineering: Practice and Patterns. Springer, 2008.

[9] Grimaila, M., Mills, R. and Fortson, L. *Improving the Cyber Incident Mission Impact Assessment Processes*.

4th Annual Workshop on Cyber Security and Information Intelligence Research, 2008.

[10] Grimaila, M. and L. Fortson. *Towards an Information Asset-Based Defensive Cyber Damage Assessment Process*. Computational Intelligence in Security and Defense Applications (CISDA), 2007.

[11] Holsopple, J. and Yang, S. *FuSIA: Future Situation and Impact Awareness*. Information Fusion, 2008.

[12] Karnstedt, M., Sattler, K., Geist, I. and Höpfner, H. *Semantic Caching in Ontology-based Mediator Systems*. Berliner XML Tage, 2003.

[13] Lewis, L., Jakoboson, G. and J. Buford. *Enabling Cyber Situation Awareness, Impact Assessment, and Situation Projection*. Situation Management (SIMA), 2008.

[14] Rahman, M., Pakstas, A. and F. Wang. *Towards Communications Network Modeling Ontology for Designers and Researchers*. Intelligent Engineering Systems. International Conference on Intelligent Engineering Systems (INES), 2006.

[15] Salerno, J. *Measuring Situation Assessment Performance through the Activities of Interest Score*. Fusion, 2008.

[16] Salerno, J., Hinman, M. and D. Boulware. *A Situation Awareness Model Applied to Multiple Domains*. Multisensor, Multisource Information Fusion, 2005.

[17] Tadda, G., Salerno, J., Boulware, D., Hinman, M. and Gorton, S. "Realizing Situation Awareness within a Cyber Environment", Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, 2006.

[18] Watters, J., Kertzner, P. and A. Hahn. *RiskMap: Finding Your Corporate Risks*. I3P White Paper, 2009.

[19] Yoakum-Stover, S. and T. Malyuta. *Unified Data Integration for Situation Management*. Situation Management (SIMA), 2008.