



North Atlantic Treaty Organization  
RESEARCH AND TECHNOLOGY AGENCY

# ***Visualising Network Information***

IST-063/RWS-010

a workshop organized under the  
INFORMATION SYSTEMS TECHNOLOGY (IST) PANEL



Royal Danish Defence College  
**Copenhagen, Denmark**  
**17-20 October 2006**

**Pre Proceedings**



# Timetable/Program

## Tuesday, 17 OCT

08.00 Bus pickup at hotel  
08.30-09.00 Registration

### Session 1: General Introduction and Orientation

09.00-09.15 Welcome General Karsten Møller, Royal Danish Defence College, DNK  
09.15-09.20 Welcome Ms. [Lisbeth M. Rasmussen](#), Workshop Chair., Danish Defence Research Establishment, DNK  
09.20-09.30 RTA Intro Ms. [Ayşegül Apaydin](#), RTA, TYR  
09.30-09.40 IST-059/RTG-025 Intro Mr. [Vincent Taylor](#), RTG Chair. Defence R&D Canada, CAN  
09.40-09.45 Goals and challenges for the workshop Ms. [Lisbeth M. Rasmussen](#), DDRE, DNK  
09.45-09.55 The break out group concept Ms. [Lisbeth M. Rasmussen](#), DDRE, DNK  
09.55-10.10 Participant Introductions  
**10.10-10.30** Coffee break

### Session 2: Keynotes

10.30-11.00 [A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action](#) Prof. [Kathleen Carley](#), Carnegie-Mellon University, USA  
11.00-11.30 [Visualizing Cyberspace](#) Col. (Ret'd) [Randy G. Alward](#), CAN  
11.30-12.00 **Plenary session on keynotes**  
**12.00-13.00** LUNCH

### Session 3: Security & Defence I

13.00-13.20 [Automating the Presentation of Computer Networks](#) Mr. [Grant Vandenberghe](#), Defence R&D Canada, CAN  
13.20-13.40 [Knowledge Discovery in Databases - Overcoming massive data streams for intelligence tasks.](#) Dr. [Vera Kamp](#) & Dr. [Joachim Stamm](#), PLATH GmbH, DEU  
13.40-14.00 [Network Visualization of Social Data Extracted from Messages.](#) Ms. [Annette Kaster](#), FGAN-FKIE, DEU  
14.00-14.30 **Plenary discussion on session 3**  
**14.30-14.50** Coffee break  
14.50-15.00 Establishing break out groups  
15.00-16.30 **Break out groups**  
16.30-17.00 Initial reports from groups  
17.00 Bus for hotel  
17.30 Arrive at hotel  
18.30 BrewPub København. Tour of the brewery, Dinner, Address: Vestergade 29

## Wednesday, 18 OCT

08.00 Bus pickup at hotel

### Session 4: General/Theory I

08.30-09.00 [Reference Framework for Network Visualisation](#) Dr. [M. Martin Taylor](#), Martin Taylor Consulting, CAN

09.00-09.30 [Characterisation and Showcasing of Network Visualisation Approaches for Command and Control](#) Mr. [Alain Bouchard](#), Defence R&D Canada, CAN

09.30-10.00 **Plenary discussion on session 4**

**10.00-10.20**

Coffee break

### Session 5: Security & Defence II

10.20-10.50 [Scalable HAIPE Discovery](#) Mr. [Shu Nakamoto](#), MITRE Corporation, USA

10.50-11.10 [Custom Ontologies for Expanded Network Analysis](#) Dr. [Amy K.C.S. Vanderbilt](#), Wave Technologies, Inc., USA

11.10-11.30 [Real-Time Extraction of Course Track Networks in Confined Waters as Decision Support for Vessel Navigation in 3-D Nautical Chart](#) Mr. [Thomas Porathe](#), Mälardalen University, SWE

11.30-12.00 **Plenary discussion on session 5**

**12.00-13.00**

LUNCH

13.00-17:00 **Break out groups**

**14.30-14.50**

Coffee break

14.50-17.00 **Break out groups**

17.00 Depart for Little Mermaid

17.45 Tour of Citadel

18.30 Dinner at Citadel

## Thursday, 19 OCT

08.00 Bus pickup at hotel

### Session 6: General/Theory II

08.30-09.00 [A three pronged approach for improved data understanding: 3-D visualization, use of gaming techniques, and intelligent advisory agents](#) Dr. [David Hall](#), The Pennsylvania State University, USA

09.00-09.30 [Open Networks: Generalized Multi-Sensor Characterization](#) Mr. [Alan Steinberg](#), CUBRC, Inc., USA

09.30-10.00 [Reduction of complexity: an aspect of network visualization](#) Prof. [Jan Terje Bjørke](#), Norwegian Defence Research Establishment, NOR

**10.00-10.20** Coffee break

10.20-10.50 [Research Methodology to Study the Cognitive Responses to Complex Network Visualization Designs](#) Ms. [Sonya McMullen](#), Tech Reach, Inc., USA

10.50-12.00 Plenary discussion on General/Theory

**12.00-13.00**

LUNCH

### Session 7: Medical

13.00-13.20 [Epidemiologic Considerations in Network Modeling of Theoretical Disease Events](#) Dr. [Marcus Lem](#), Health Canada, CAN

13.20-13.40 [Application of Network Visualisation in Infectious Disease Management](#) Dr. [Margaret Varga](#), QinetiQ, GBR

13.40-14.00 [Modeling Influenza Pandemic Response Effectiveness in Canada](#) Dr. [Zachary Jacobson](#), Health Canada (Mr. [Benjamin Houston](#), Neuralsoft Corporation), CAN

14.00-14.30 Plenary discussion on Medical

**14.30-14.50**

Coffee break

14:50-16:00 **Break out groups**

16.00 Departs for hotel

16.30 Arrive at hotel

18.00 Tivoli

18.30 Dinner in Valhal (in Tivoli)

## Friday, 20 OCT

08.00 Bus pickup at hotel

### Session 8: Security & Defence III

08.30-09.00 [Practical Approaches for Analysis, Visualization and Splitting of Terrorist Networks](#) Mr. [Nasrullah Memon](#), Aalborg University Esbjerg, PAK/DNK

09.00-09.20 [Information Fusion and Visualisation in Anti Asymmetric Warfare](#) Dr. [Felix Optiz](#), EADS Deutschland GmbH, DEU

09.20-09.40 [Runtime Simulation for Post-Disaster Data Fusion Visualization](#) Dr. [T. Kesavadas](#), University at Buffalo, USA

09.40-10.10 Plenary discussion on Security & Defence

**10.10-10.30** Coffee break

10.30-12.00 **Break out groups**

**12.00-13.00** LUNCH

13.00-14.10 **Break out reports**

**14.10-14.30** Coffee break

14.30-15.00 Discussion on workshop

15.00-15.30 **TER Report** Col. (Retd) [Tom Johansen](#), Nordic Defense, NOR

15.30-16.00 Workshop wrap-up

16.00 Depart for station/hotel





# Abstracts



## ***Keynotes***



# **A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action**

**Kathleen M. Carley**

Institute for Software Research International, SCS  
Carnegie Mellon University  
Pittsburgh, PA 15213, USA

[kathleen.carley@cs.cmu.edu](mailto:kathleen.carley@cs.cmu.edu)

## ***ABSTRACT***

Dynamic network analysis (DNA) is an emergent field centered on the collection, analysis, understanding and prediction of dynamic relations among various entities such as actors, events and resources and the impact of such dynamics on individual and group behaviour. DNA facilitates reasoning about terrorist groups as complex dynamic networked systems that evolve over time. An interoperable DNA toolchain for collecting data on, assessing the network of, and forecasting changes in that network is presented. The use of these DNA tools to assess a terrorist group is then demonstrated using open source data. Key techniques are demonstrated using a dataset collected on terror networks. Techniques demonstrated include those for identification of an actor's sphere of influence, emergent leaders, and paths among critical actors, and metrics for assessing the potential immediate and near term impact of various courses of action.

## **1.0 DYNAMIC NETWORK ANALYSIS**

Complex socio-technical systems are furthermore dynamic. The dynamics result from multiple change processes such as natural evolutionary processes including learning, birth and aging as well as intervention processes such as altering the set of individuals who lead a system. Data on these systems is often incomplete, replete with errors, and difficult to collect, which further complicates the understanding and evaluation of these groups which can be extremely large. Traditional analysis approaches, such as Social Network Analysis (SNA) (Wasserman and Faust, 1994), forensic science (Saferstein, 2001) and link analysis (Thelwall, 2004), are limited in their ability to handle data the type of multi-plex, multi-mode, large-scale dynamic data that are needed to characterize terror networks. Consequently, tools that go beyond traditional SNA and link analysis are needed.

In response to these needs, a new sub-field, Dynamic Network Analysis (DNA), has emerged (Carley, 2003). In a DNA approach, computational techniques, such as machine learning and artificial intelligence, are combined social network analytic and link analytic techniques to provide an approach to assessing multi-mode, multi-plex dynamic large-scale systems under varying levels of uncertainty. Dynamic network analysis (DNA) facilitates the collection, analysis, understanding and prediction of dynamic relations among diverse entities and the impact of such dynamics on individual and group behaviour.

## **2.0 DNA TOOLCHAIN**

The integrated CASOS dynamic network analysis toolkit is an interoperable set of scalable software tools. These tools form a toolchain that facilitates the dynamic extraction, analysis, visualization and reasoning about key actors, hidden groups, vulnerabilities and changes multi-mode multi-plex dynamic networks, such as those formed by terror networks. This toolchain enables analysts to move from raw texts to meta-networks to the identification of patterns in networks to analysis of possible effects of alternative courses of action designed to disrupt or contain these networks. This is

accomplished by using tools for text mining, statistical network analysis and computer simulation. Secondary tools such as those for web-scraping, visualization, database management and data editing provide support capabilities. This toolchain is illustrated in Figure 1. Using this toolchain the kinds of questions that can be addressed by the analyst include:

- Who to target (vulnerabilities)
- What groups or individuals stand out
- How to influence
- Are there important connections among actors or groups
- What is the “health” of the organization
- Where might there be missing data
- How different are two groups – or two sources – or the same group at two different times
- What is the immediate and near term impact of various courses of action

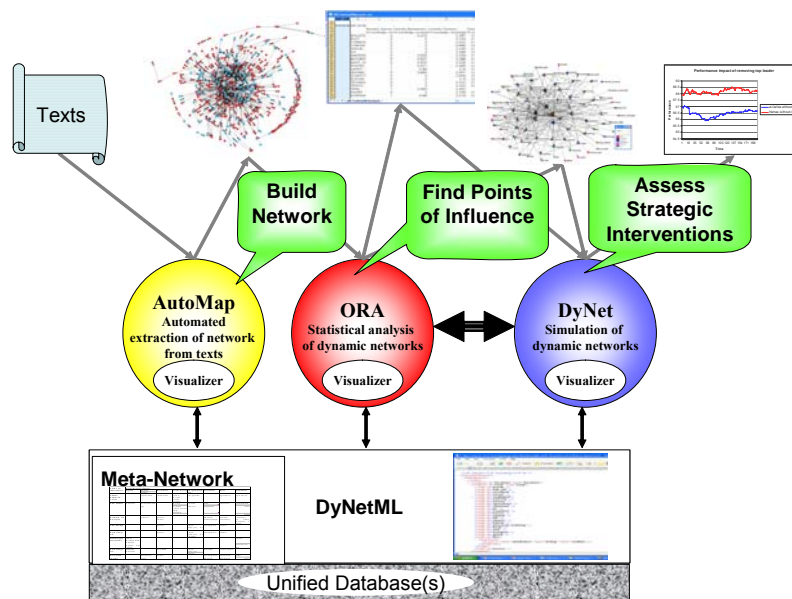


Figure 1: Toolchain for Dynamic Network Analysis

### 3.0 APPLICATION TO TERROR NETWORKS

Thousands of open source documents, such as newspaper accounts were scraped and then processed with AutoMap to create a database of information on the actors, organizations, resources, events, and locations. The resultant data was then analyzed with ORA. The network that emerges shows a decidedly cellular structure with 5-12 members per cell. Drill-down visualizations on this structure reveal the changing roles of key actors. Over time results show a decrease in the density and communication levels in the terror network but an increase in its congruence. This suggests a movement to a more distributed and more efficient organizational form, possibly with larger cells. Course of action evaluation shows that removal of highly central actors in the communication network will be less effective than removal of key emergent leaders, even though from a purely visualization stand-point the central actors would be thought to be more critical.

## 4.0 SUMMARY

Network oriented toolchains facilitate better analysis by reducing the time spent in repetitive tasks where little analyst insight is needed. To further support the analyst, however, additional effort needs to be invested in building parsers between the various databases, enabling automated data collection, and identifying common analyses and providing those in a single automated report. With this in mind, it is critical that future generations of DNA tools take into account factors such as the confidence in the data, automated estimates of robustness, data-farming, and tools for user-in-the-loop data testing.

The approach used herein affords analysts with greater analytical power and facilitates assessing large, dynamic and complex socio-technical systems such as terror networks. By taking into account not the networks among people and organizations as well as their relations with resources, knowledge, events, and locations etc., insights into diverse behaviours are gained. If we look only at the social network then the focus of attention is on hierarchies, communication and other social relations. The addition of events and locations facilitates course of action analysis and enables linkage to various strategic planning tools. The over-arching meta-network approach used in DNA has the promise of enabling effects based operation in areas as diverse as diplomacy, information, military and the economy to be assessed in a relational context.

## 5.0 REFERENCES

- [1] Carley, Kathleen M. 2003, "Dynamic Network Analysis" in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145.
- [2] Saferstein, R.E. (2001) *Forensic Science Handbook*, Prentice Hall, NJ.
- [3] Thelwall, M., 2004, *Link Analysis: An Information Science Approach*. Academic Press.
- [4] Wasserman S. and K. Faust, 1994, *Social Network Analysis*. New York, Cambridge University Press.



## ABSTRACT

### VISUALIZING CYBERSPACE

by

Colonel (retired) Randy Alward, CF

The 'Information Age', which in my view we are just entering, has forever changed how we live, work and exercise military force. Futurists, like Toffler, foretell of a revolution like the Industrial Revolution of the 17<sup>th</sup> century, the Information Revolution. It is characterized by an exponential growth in information and knowledge in all fields of human endeavour; by paradigm changes in social and economic activity, by a transformation of our militaries to small, highly mobile, informed, surgical like strike teams. The underpinnings of this 'Information Age' are computers, networks, software and immediate access to relevant information. How do we manage and maintain the health of these underpinnings?

A former American Vice President, Al Gore, characterized these underpinnings as 'the Information Highway'. A Canadian science fiction writer, William Gibson, coined the phrase 'Cyberspace'. We have a new infrastructure, an environment that needs to be maintained and protected. I will present a case for viewing this infrastructure as a battlespace, not unlike land, air and sea, over which we need to exercise 'Command and Control'. This new space is 'Cyberspace'.

Command and Control is the means by which we implement decisions. Consider the decision cycle of 'observe, orient, decide and act'. Visualization is the means by which we observe and understand. How do we visualize Cyberspace effectively; the network, the flow of information, the exercise of control? This is a primary domain of your working group and of this particular workshop.

As a former Commander responsible for Information Security, and guiding related research and development, for the Canadian Department of National Defence and Canadian Forces, I hope to provoke thought and discussion with my views on exercising Command and Control in Cyberspace, including how we should observe the network. I will outline current network operational capabilities and discuss what is needed to understand the network. I will also comment on current development and research activities that support this understanding.

24 April 06



## ***Security & Defence I***



# AUTOMATING THE PRESENTATION OF COMPUTER NETWORKS

**Grant Vandenberghe**

Defence R & D Canada - Ottawa  
3701 Carling Avenue  
Ottawa, Ontario  
K1A 0Z4

[Grant.Vandenberghe@drdc-rddc.gc.ca](mailto:Grant.Vandenberghe@drdc-rddc.gc.ca)

10 March 2006

## 1. ABSTRACT

A well-drawn computer network diagram tells a story. When the information is well organized the structure and relationships contained therein can be conveyed at a glance. System administrators use network diagrams as a basis for documenting, planning, and controlling computer networks. The user-friendly presentation of network information for a reasonably sized enterprise computer network involves a considerable degree of complexity. Defence R&D Canada has been developing the ability to automatically generate network diagrams that are relatively easy to interpret. A study was done which compared a series of automatic layout algorithms to understand their strengths and weaknesses, leading to the development of an improved layout algorithm to address shortcomings.

There is more than one way to present the information on a computer network diagram. In general, most network diagrams are based on either a physical or logical frame of reference. Although both frames of reference are implemented, this presentation will only discuss the layout of logical computer network diagrams. When considering logical diagrams it is frequently desirable to only consider a fraction of the complete network. Presenting information on thousands of network devices simultaneously creates a picture for an administrator or analyst that is difficult to interpret and leads to confusion.

To summarize the network content, a series of hierarchal abstractions called visualization levels is employed. A series of layout algorithms is then applied to both the highest visualization level and the children of each summary element (called an expanse). The final result of each individual layout operation is a mini-diagram or "patch". These patches are sewn together to form customized network diagrams. Eventually such a system should be able to support the recombination of multiple layout algorithms into a single network diagram.

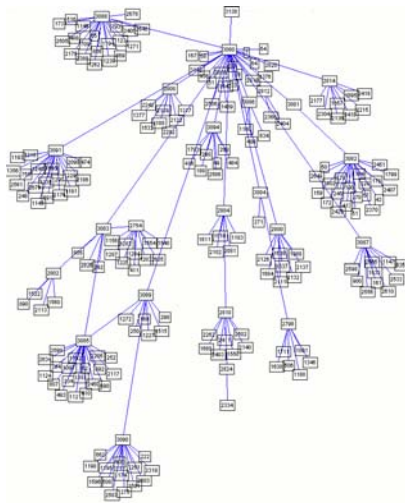
There are many types of layout algorithms described in the open literature. Some algorithms are more suited than others for producing computer network diagrams. In order to answer the question of which algorithms are best suited, a benchmark criterion was established. A survey of existing computer diagrams was conducted to compare the similarities and differences between manual and automatic drawings. A list of sixteen characteristics was established describing the qualities of a good network drawing.

Using the evaluation criterion as a base, six standard layout approaches were applied to the layout of a small corporate network. These existing algorithms performed well but fell short in a number of areas in terms of network presentation. To compensate for the inadequacies, two custom algorithms were devised which were specifically tuned to the layout issues surrounding the layout of computer network drawings.

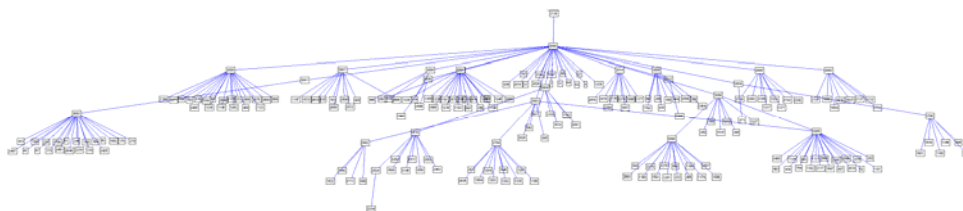
The XY Control System algorithm is intended for computer networks that are best represented using force or energy directed layout models. The technique uses a set of five linear and nonlinear controllers that have been isolated in the X and Y directions. The XY algorithm offers a stronger grid-like appearance than other existing force directed models. Samples of the results from the best existing algorithm is shown in Figure 1 while the results from the XY controller for the same graph are shown in Figure 2.

The Voting algorithm is intended for networks that can be represented by sparsely connected hierarchal directed graphs. The algorithm "combs" through the graph hierarchy splitting the network up into a series of vertical threads. The algorithm then proceeds to sequence the threads to minimize edge crossings. There can be multiple vertices present along the length of each thread. To sequence the vertices in the thread a unique four factor voting process is used. Finally groups of nodes that are clustered around a single device are placed using a detailed rectangular grid placement algorithm. The algorithm produces good results because it tries to simplify a problem before solving and because it uses multiple decision criteria in place of one for critical layout decisions. A sample of the results obtained from the voting algorithm is shown in Figure 3.

The overall layout algorithm has been integrated into a visualization tool that is part of a computer traffic exploration tool for network security. A demonstration of the visualization component of the tool is available to interested parties.



**Figure 1 Magnetic Spring Force Direct Model**



**Figure 2 XY Control Algorithm**

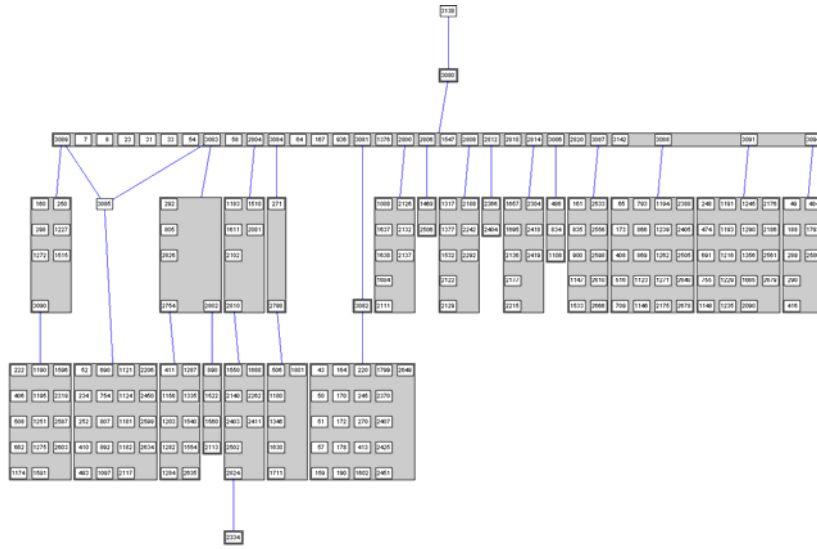


Figure 3 Voting Algorithm

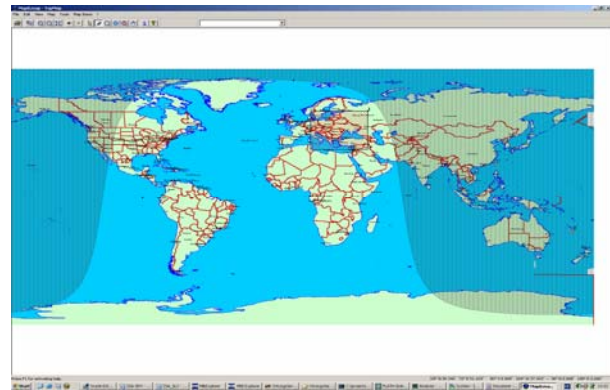


## KDD<sup>1</sup> - Overcoming massive data streams for intelligence tasks

Actually very interesting IT systems promise to reveal connections between apparently harmless and unrelated information pieces. An article from the New York Times in February 2006<sup>2</sup> makes clear that common data mining techniques were not successful in general. Despite huge investments, correlating data from different sources did not yield satisfactory results. Transforming low-level data by aggregation to meaningful events is nevertheless the key to building the basis for succeeding decisions in the context of situation reports.

More realistic and manageable is an approach that includes interactions with the user along with domain specific knowledge. Gaining security relevant messages should be based on an iterative multi-level process. This process represents the core element of intelligence analysis systems which play an important role for supporting decisions in management information systems<sup>3</sup>.

The following example illustrates the principal automated process for discovering communication structures in the context of radio reconnaissance: A crucial part of this process is the analysis and visualisation of communication structures, or more generally, of network information. This should be embedded in spatio-temporal data analysis with geo-oriented data access and the integration of domain-specific analysis functions.



The intelligent analysis of radio emission data is based on data mining techniques, cluster visualisations to validate the results, a model based communication detection (including domain-specific knowledge) and the visualisation of communications. The following use case of a simple simplex communication clarifies the problems and the applied methods. Module coupling is realised by a distributed architecture. Given are a huge amount of radio emissions which are arbitrarily distributed. Each emission is described by the attributes ID, frequency, modulation type, starting time, end time, latitude and longitude. It has to be considered that the data quality of single emissions depends on propagation conditions. Because these can vary, it can happen that single emissions or attributes are missing or on the other hand different classification level information are available. Furthermore, with a broadband collection of emissions the amount of information is extremely large and requires massive data handling which can not be processed in main memory.

### Use-Case Simplex-Communication

The use case is looking for a simplex communication chain with two stationary partners – a central station and a substation. Both are using the same constant nominal frequency and the same transmission mode. The partners are communicating alternating one after the other. The problem lies

<sup>1</sup> Knowledge Discovery in Databases

<sup>2</sup> J. Markoff, The New York Times, nytimes.com, 25.2.2006: "Taking Spying to Higher Level, Agencies Look for More Ways to Mine Data"

<sup>3</sup> Jan Herring, "What is Intelligence Analysis?", Competitive Intelligence Magazine, Vol. 1. No.2, July-Sep., 1998, p.14 .

in the amount of possible communication structure instances. Although the communication can be easily described in an informal way it is necessary to find an exact, formal specification in order to perform a computer-supported analysis. It should not be realised by a specific static algorithm but should be interactively and exploratively changeable by the user. The core concept includes the following steps:

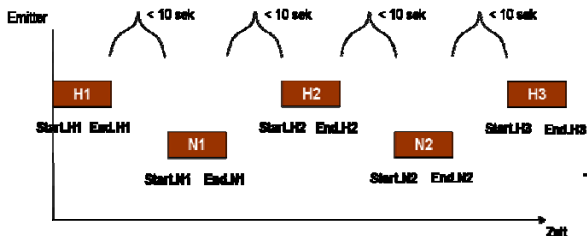
**(a) Data Mining**

During the first step emissions are assigned to clusters. These subsume emissions concerning the spatial, temporal or frequency criteria. In this way significant data reduction is achieved. By spatial clustering special emitter station could be determined. Besides when processing of extremely huge data amounts the main problem to solve is how to choose the best method and parameters.

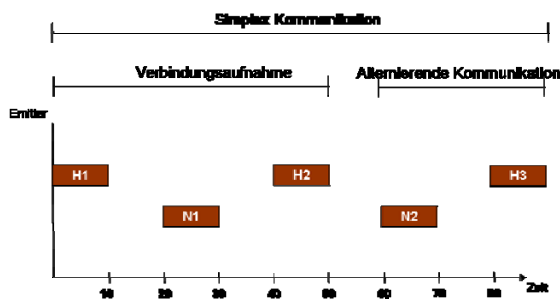
**(b) Cluster Visualisation**

The next step serves the validation of the data mining results and already provides a possibility to manually discover communication structures by the user relying on the presented visualisation, for example the presentation of spatial clusters. Emission can appear as single instances or as temporal ordered parts of a cluster. It is difficult to visualise the emissions and clusters clearly arranged in order to focus on the actual interesting data. Additionally different attributes have to be integrated.

**(c) Model based communication detection**

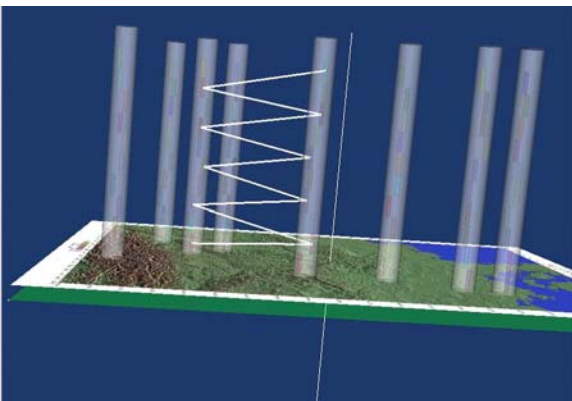


Computing communication structures from clusters is the next step. This is done by using typical communication models. A domain specific modelling language provides the possibility to represent the communication models. By this language the simplex communication can be formally specified. The model distinguishes between connection constitution and alternating communications.



The connection constitution consists of three emissions: the central station is sending, the substation replies. The alternating sequence consists of emissions of the central and the sub station. All emissions occur with the same frequency and modulation type. The distance between the emissions is flexible by a delay parameter. A graphical notation of such a model is illustrated by the two adjoining pictures.

**(d) Visualisation of discovered communications**



This step provides a presentation of the discovered communications and allows by this a validation of the model based communication detection. It has to deal with many composite events. A simple textual visualisation does not meet the needs. The graphical visualisation offers a better overview and manifold interaction possibilities.

The emitters are on the spatial level, time is the third dimension. The connection lines indicate the communications.

Overcoming massive data streams for intelligence tasks is a challenge which should involve the analysis process with a seamless data access and the intelligence analyst. The acceptance of the

results depends on the possibility to validate the results. The sustainability of results has to be guaranteed by flexible extension of actual domain specific analysis methods.

Author:

**Dr. Vera Kamp**

Head of intelligent analysis system development

**PLATH GmbH**

Gotenstrasse 18, 20097 Hamburg, Germany

[www.plath.de](http://www.plath.de)

[vera.kamp@plath.de](mailto:vera.kamp@plath.de)



# Network Visualisation of Social Data Extracted from Messages

**Annette Kaster and Oliver Witt**

FGAN - Research Institute for Communication, Information Processing and Ergonomics  
Ergonomics and Information Systems Department  
Neuenahrer Strasse 20, 53343 Wachtberg-Werthhoven, Germany  
Email: [a.kaster@fgan.de](mailto:a.kaster@fgan.de) and [witt@fgan.de](mailto:witt@fgan.de)

an abstract submitted in response to a Call for Contributions for the IST063/RWS010  
"Visualisation Network Information"

Organised by the NATO Research and Technology Agency – IST-063 / RTG-010  
**to be held in Copenhagen, DNK, 17-20 October, 2006**

In the area of Intelligence, Surveillance & Reconnaissance (ISR) the identification, analysis, visualisation and evaluation of contextual relations contained in acquired messages are of particular importance. The deployment of improved sensor technologies and the utilisation of innovative sources like the world wide web result in a high but also sophisticated message accumulation. These continuously increasing quantity and variety of information require extensively automated but also transparent support for the analysts. He has to be assisted in recognizing and analysing relevant information and interrelations from a comprehensive inventory of acquired and stored single messages. The support should cover, on the one hand, a nearly entirely automated identification and, on the other hand, an ergonomically designed visualisation of discovered relations and relation objects. An important aspect is the transparency of the automated processes since an entire automation is not conceivable due to incomplete, contradictory and incorrect messages. Subject of this contribution is a concept describing the procedure from acquiring a message, identifying and extracting relevant information up to the visualisation of contained conceptual relations. Several prototypes/tools have been developed for realizing and demonstrating the proceeding [1]. Basis is the "J2-database mission" (DBEins [2]). It is used for structured acquisition, processing and evaluation of J2-messages from conflict and operational areas. It contains tools for making documents available to participating services in a continuous information network within a command and control information system. Various document types are managed according to a common concept covering the complete spectrum from free text to formatted messages.

The messages archived in the DBEins are automatically analysed by an information extraction tool [3]. The tool-box GATE is used as the development environment [4]. It operates on accumulated text indicted in English. It allows to combine tools for information retrieval, information extraction or machine translation. It contains three logical components: a database for storing information about text, a graphical user interface for monitoring the processes on data as well as inspection and evaluation of results and a collection of algorithms and data interacting with the database and the user interface. This tool provides predefined transducers for recognising varying English proper names and English verbal phrases a gazetteer, a POS-Tagger, etc.. For the social data analysis this tool is used to recognise and to categorise names (of people) and to identify their relations.

Since network-like structures are to be presented graphical elements like graphs, nodes and edges are suggested for visualisation purposes. Each node symbolizes a person. The undirected edge between two nodes/persons represents the message where both persons are mentioned or the kind of relation respectively (an example network is depicted in Figure 1). Multiple edges indicate how often these persons appear together in messages.

A prototype has been developed using visualisation libraries of Tom Sawyer Inc. for representation, editing and automatic layout of graphs called experimental graph creation and representation

(xGERDA [5]. The bidirectional linkage between xGERDa and the DBEins allows access to the original messages coming from the relationship network.

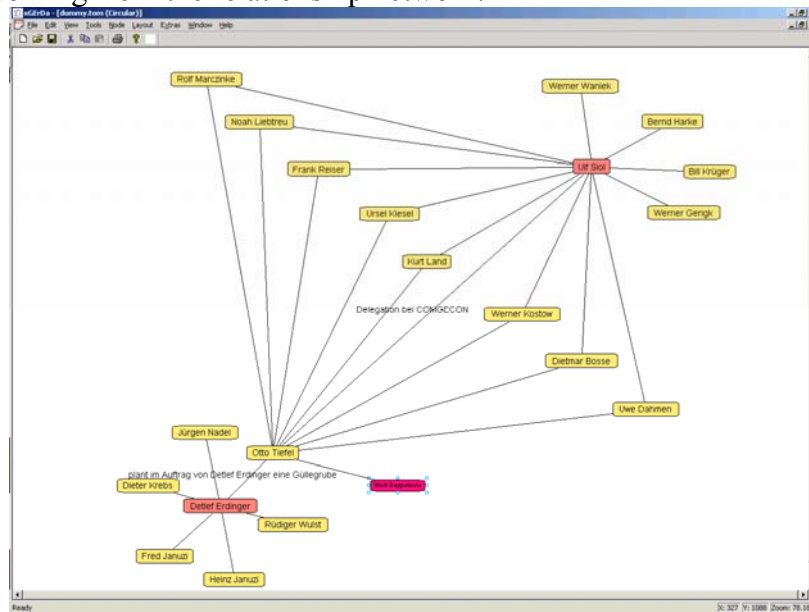


Figure 1: Social network represented by xGERDa

Graphs comprised of a multiplicity of nodes and edges might quickly become voluminous and complex so that they cannot be presented completely and clearly laid out at the screen. Layout algorithms are used to arrange graphical elements according to certain criteria and consequently can lead to an easier understanding of big complex graphs. Depending on the task on hand one algorithm might be superior than another. xGERDa provides various algorithms for the layout of graphs. Furthermore algorithms for network analysis have been integrated and can be used by the evaluator. An optimized search function simplifies the locating of targeting objects in large networks. The human machine interface of xGERDa is bilingual (German and English) in order to be used in international operations (e.g. it will be used during the international Coalition Warrior Interoperability Demonstration (CWID) in 2006 in Lillehammer, NOR). In the future it has to be analysed how a (optimal) layout algorithm can be determined automatically according to a given network and which further ergonomic functionalities might be useful for the analyst. This presentation will show the path from receiving a message, storing it, analysing it, extracting information up to representing relations of objects (people, infrastructure, vehicles, etc.) and the functionalities of the visualisation tool xGERDa. Furthermore, possibilities are discussed how semantical context might be visualized in the network.

#### References:

- [1] Kaster, A. and Witt, O. (2005), *J2-Database (DBEins) – From the J2-Message to the Graph of Relations*, NX-Workshop on „Social Network Analysis and Visualisation for Public Safety“, Germany.
- [2] Kaster, J. (2003), *Unterstützungskonzept „Datenbank Einsatz*, Project Report, FGAN, FKIE, EFS, Wachtberg-Werthhoven, Germany (*in German*).
- [3] Hecking, M. (2004), *How to Represent the Content of Free-form Battlefield Reports*. In: Proceedings of the 2004 Command and Control Research and Technology Symposium, San Diego, California, USA
- [4] Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002), *GATE: A framework and graphical development environment for robust NLP tools and applications*. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics, Philadelphia, USA.

- [5] Witt, O. (2006), *xGERDa-Handbuch*, Technical Report, FGAN, FKIE, EFS, Wachtberg-Werthhoven, Germany (*in German*).

## ***General/Theory I***



# Reference Framework for Network Visualisation

## Framework Working Group of IST-059

### Contents

- [1 Need for a Reference Framework for Network Visualisation](#)
  - [1.1 Other Approaches to a Reference Framework](#)
- [2 Starting points](#)
  - [2.1 VisTG Reference Model](#)
  - [2.2 Local and Global Network Properties](#)
    - [2.2.1 Broadcast and Stigmergic Networks](#)
  - [2.3 Mathematical Aspects of Networks](#)
    - [2.3.1 Fuzzy versus Crisp Nodes and Links](#)
    - [2.3.2 Fuzzy Chains and Cycles](#)
  - [2.4 Embedding Fields](#)
    - [2.4.1 Embedding and Inheritance](#)
    - [2.4.2 Embedding field for display](#)
    - [2.4.3 Dimensionality of Embedding Field](#)
  - [2.5 Dynamical Behaviour](#)
    - [2.5.1 Traffic Dynamics and Network Dynamics](#)
    - [2.5.2 Traffic dynamics and loop behaviour](#)
  - [2.6 Information Theoretic aspects](#)
    - [2.6.1 Network Dynamics and Display as an Information-theoretic Issue](#)
  - [2.7 Taxonomies of Data and Display types](#)
- [3 Summary: Bringing the threads together](#)

## 1. NEED FOR A REFERENCE FRAMEWORK FOR NETWORK VISUALISATION

Representations of networks come in a great variety of forms, each designed to show off some aspect of the network, and each based on the thoughts and intuitions of some designer. Some two hundred very different examples produced for a wide range of applications are illustrated at the "[Visual Complexity](#)" Web site.

Networks have many different properties that the user might find important for the purpose of the moment, and many different approaches have been taken to creating representations that support the users ability to visualise these properties. Clearly, the designer of one representation will have been influenced by having seen earlier ones, but there is no obvious way for someone who wants to design a representation supporting a particular application to generalize from earlier examples to the new case, unless the kind of network and the needs of the user are clearly analogous in the two situations.

The intention of any representation is to aid the human to visualise some aspect of the thing represented. The computer presents the data in some form, perhaps pictorial, perhaps not. Based on that presentation and his or her background memory and skill, the human user, as one of the routes to understanding the data, visualises its implications. The computer's presentation of the data is an aid to the user's visualisation, not its content. The content of the visualisation incorporates not only the data displayed by the computer, but also the user's memory and imagination. This important point is often lost in the design of displays that are designed to show as much of the data as is reasonably possible, including things the user might be expected to know already.

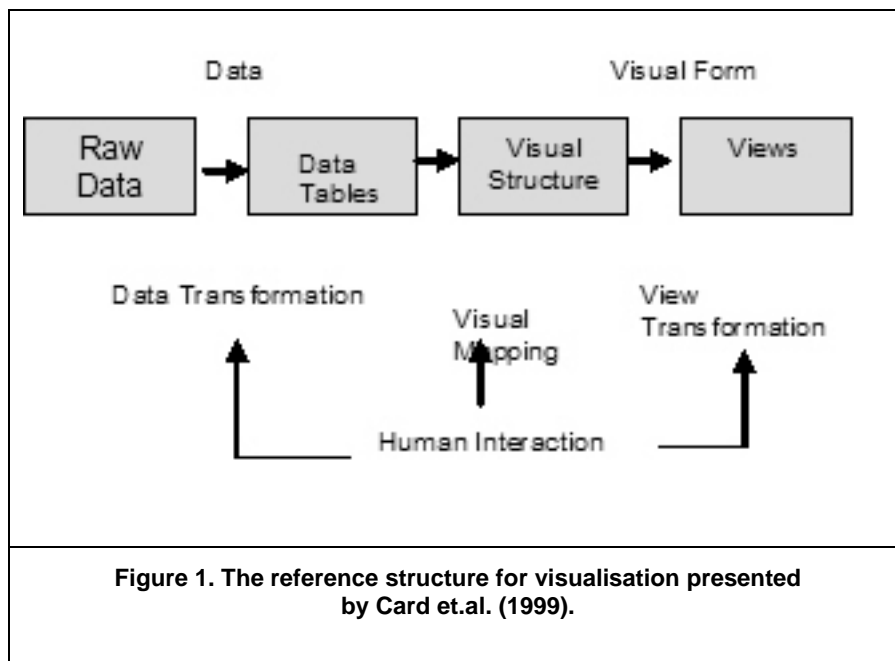
The user has some purpose in wanting the data to be displayed. Perhaps the purpose is no more than idle curiosity, but more commonly it is in support of some task of the moment. At any particular moment, the user's purpose probably does not require very much of the available data to be presented, but whatever data are presented must have context, whether it be supplied in the presentation or by the user's memory and imagination.

The objective of a Reference Framework is to provide a guide to assist generalization from one situation to another, and perhaps to assist in both the design process and the evaluation of a completed design. A good Reference Framework should also assist a user to determine whether some particular software is likely to be useful for a particular purpose, and to guide the selection of the most useful out of a collection of visualisation system.

At the Spring meeting in 2005, IST-059/RTG-025 initiated a Working Group to investigate the possibility of developing a Reference Framework for Network Visualisation. A preliminary report was presented at the Autumn meeting of 2005. This is a further status report that delineates some of the issues that must be addressed. It does not present a finished Reference Framework for Network Visualisation.

### 1.1 Other Approaches to a Reference Framework

Card et al. [Card, S. K, Mackinley, J. D., Shneiderman, B, in *Readings in information visualization: Using vision to think*. Morgan Kaufmann Publishers, Inc. San Francisco, 1999] define a general reference model for information visualization. They show schematically in the model the flow from raw data to the visualization component as well as the various manipulations that are possible to change the visual representation of the data.



In this reference model raw networked data can be stored in various repositories, i.e. any useful format. Data transformation maps raw data into a characteristic format, e.g. into data tables which are a relational description of data perhaps extended to include meta data. In the case of networked information this could, for example, be a topic map. The data tables are transformed by the visual mapping into a visual structure that combines the spatial substrates, marks and graphical properties. Views are then created by specifying the graphical parameters, e.g. position, scaling, viewing angle of the data etc. User interaction controls the parameters of these transformations.

In the case of information visualization of a network, the most commonly used mapping is to represent the network structure using nodes and arcs. While an arcs and nodes structure is often logical, it poses several problems. For example, overview of the entire network can be difficult and searching a large network can be difficult or even infeasible. Furthermore, the user may not be interested in the network structure, as such, but in some properties such as clustering or temporal dynamics, which the "point and arc" representation does not address.

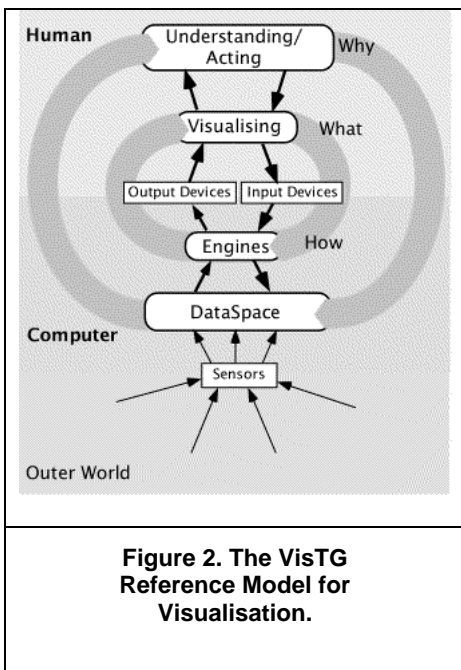
## 2. STARTING POINTS

The Working Group began its work with a Strawman document that was discussed and revised at the Spring 2005 meeting of IST-059/RTG-025. That document took as starting points several different concepts developed under the aegis of IST-059/RTG-025 and its predecessors. The generic VisTG Reference Model and the taxonomies and mappings of Data Types and Display Types were all published in the so-called "HAT Report" (Visualisation of Massive Military Data Sets: Human Factors, Applications, and Technologies, [NATO RTO-TR-030](#)) the Final Report of IST-013/RTG-002. Yet another starting point is the draft set of Properties of networks produced by one of the syndicate working groups at the 2004 Toronto Workshop, IST-043/RWS-006. These, and other subsequently introduced starting points will be developed together as part of the process of constructing a Reference Framework for Network Visualisation.

### 2.1 VisTG Reference Model

The VisTG Reference Model for visualisation was developed in 1996 by a predecessor of IST-059/RTG-025, as a generic framework within which any question of representation for display could be addressed. Although it was developed before the Card et.al. model, it rationalizes aspects of that model that are confounded in the representation of Figure 1.

The VisTG Reference Model for Visualisation is based around the fundamental concept that all display is to support some purpose of the user, and is not simply contingent on the data that must be displayed. The same data may be displayed in an indefinite range of ways, depending on what the user wants to do. This point is worth emphasising, since many approaches to data display seem to start with the notion that there is a certain set of data, and the data themselves determine a best form of display. The VisTG Reference Model starts from the position that this common assumption is simply false.



**Figure 2. The VisTG Reference Model for Visualisation.**

The basic structure of the VisTG Reference Model consists of three nested sets of feedback loops, each usually drawn as one "thick" loop. An outer set of loops (drawn as a single loop) links the user's purpose to the data: the user requires certain data to support the purpose, and the computer supplies those data. However, the connections between the user's purpose and the data cannot be done telepathically, so a middle set of feedback loops connects the user's visualisation of the data with "engines" in the computer that extract the relevant data and manipulate it into forms suited for presentation to the user. (These, of course, are the two central stages of the later Card et.al structure). The user communicates with those engines to control the extraction, manipulation, and presentation. Again, those communications cannot occur telepathically. They are performed through the innermost loop, which consists of the presentation and input hardware systems with their supporting software.

Each of the three sets of feedback loops is potentially complex, consisting of many independent or interacting simple pathways, but each individual simple loop contributes to the design or analysis of a visualisation system through a set of six or seven questions, which boil

down to the question of whether the user is able to perceive from the computer what is needed at the moment, and whether the user has the means to influence what the computer presents so that it does provide what is needed. These simple, and perhaps self-evident, requirements for any visualisation system provide a skeleton on which the specifics of a particular application build the flesh.

## 2.2 Local and Global Network Properties

A Local property of a network is a property that can be attributed to a single node or to a single link (including the mechanisms that make the connections in the initiating and receiving nodes). Examples of local properties might include the bandwidth of a cable carrying message packets (a link), the likelihood that an infected person will exhibit symptoms of a disease (a node), the directivity of a link, the distribution of delay times between the receipt of traffic and its consequence on emitted traffic (a node), or the functions that transform incoming to outgoing traffic (a node). Over the network, such local properties may be attributed to any or all links and nodes. The distribution over the nodes and links is a global property of the network or of a subnet.

A Global property cannot be attributed to a single node or link, but involves at least two nodes and the link between them or two links and a node to which they both connect. Most of the interesting mathematical properties of a network, such as the density of connections or whether it contains cycles, are global.

### 2.2.1 Broadcast and Stigmergic Networks

When a network is drawn as a set of nodes and links, the hidden implication is that traffic that originates at a node will be received by the node at the other end of the link. For two kinds of network, broadcast and stigmergic, this is true only in retrospect. In a broadcast or a stigmergic network, existence of a link is determined retrospectively, by the fact that a node did in fact receive traffic that originated at another node.

A node may broadcast its traffic, and any of a large number of other nodes may have the capability of receiving it, but only a (possibly null) subset of those potential recipients actually receive the traffic. If the traffic is not received when it is sent, it is not thereafter available to be received. The traffic is transient. The classic example is of a radio transmitter, but examples occur in many fields. A broadcast network is one in which some or all of the links are broadcast.

A stigmergic network depends on a related effect that occurs when a node acts so as to alter some facet of the embedding field (see below) of the network, in such a way that the alteration may later be detected by one or more nodes. One classic example occurs when a vehicle drives along a muddy road, leaving a rut that induces later traffic to follow the same rut; the earliest published example was of the pheromone trails left by ants foraging for food, trails which guide other ants toward profitable locations and away from unprofitable ones. Another example might be the transmission of disease through infective agents left on surfaces such as door knobs or drinking vessels. This phenomenon is called "Stigmergy". A stigmergic network is one in which some or all of the nodes act in a stigmergic manner.

The feature common to broadcast and stigmergic networks is that the recipient of the traffic is not known *a priori*. Nor is it known whether a particular element of traffic emitted by a node will be received at all, or if it is, by how many recipients. The dynamics of broadcast and stigmergic networks are therefore stochastic rather than deterministic.

Broadcast and stigmergic networks differ in the same way that an electrical signal differs from its time integral. In a broadcast network, the traffic is ephemeral, existing only when the "wavefront" of its transmission reaches each potential recipient node, whereas in a stigmergic network, the traffic, once emitted, retains its effect over time and may be received by different other nodes at different later times. Stigmergic traffic may, however, decay or be erased or overlain by later traffic from the same or other nodes, as might happen if a grader came along and removed the ruts on the once-muddy road.

## 2.3 Mathematical Aspects of Networks

Crisp networks (those in which nodes and links either exist or do not exist) have been well studied over the year, and there is much literature on their mathematical properties.

It is possible to take any set of sub-units that interact with one another and define a suitable network to represent such a system [Caldarelli, G., The structure of biological and social systems, The Mathematics of

Networks, SIAM News, Volume 37, Number 3, April 2004]. Graph theory has been used to analyze the mathematical properties of many such networks. In Erdos and Renyi's Random Graph Theory, it is assumed that each pair of vertices (or nodes) in their network model is connected with probability  $P$ , which results in a structure whose number of connections (or edges) per vertex follows a binomial distribution. In most real-world networks, however, the number of edges per vertex decays at much lower rates than in the random graphs. Such real networks are termed scale free.

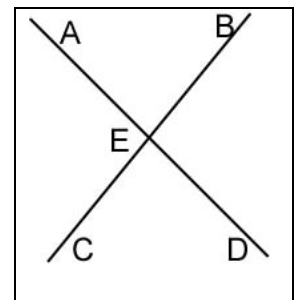
A scale free network is characterized by a few highly connected nodes that link the rest of the nodes to the system. This characteristic can be explained by the observation that the number of vertices in naturally evolving real world structures tends to grow constantly, but new vertices tend link to the system through existing vertices that already have large numbers of edges per vertex. This kind of growth and preferential attachment produce a scale-free network. In some cases the scale-free property of the network can be due to the intrinsic variety and randomness of the different vertices.

Real-life networks, however, are not always so cleanly defined.

### 2.3.1 Fuzzy versus Crisp Nodes and Links

Networks are often shown in print and on computer screens as nodes that are represented by dots, connected by links that are shown as lines. Nodes A and B either are or are not connected by a link. This either-or, "yes-no", dichotomy is called "crisp". Things are different in the real world that interests us. There, the status of a link or a node may not be crisp. It might be fuzzy. Whether it is crisp or fuzzy, and the membership of a possible connection in the class "link" if it is fuzzy, may depend on the intentions of the user of the network as much as on the physical structure being represented. Consideration of networks as fuzzy shows promise as a way to link the properties of the network to the visualisation needs of the user.

Consider a road network as an example. We can define towns as nodes (and ignore for the moment the fuzziness of the status of hamlets or rural service stops). Also we can define that two towns are linked by road if a traveller can get from one to another along the road network without passing through a third town. To make this concrete, if there are four towns, A, B, C, and D, arranged in a square, and straight roads connect the corner towns  $A \leftrightarrow D$  and  $B \leftrightarrow C$ , with a crossroad in the middle of the square, then every pair of towns is linked, not just the diagonally opposed ones. If there were a fifth town, E, at the crossroad, then no pair of the original four would be linked, but all four would be linked to E.



**Figure 3. Four towns (A, B, C, and D), linked by two roads that cross at the crossroad E.**

In the foregoing, the role of "the traveller" is overlooked, and it should not be. Whether two towns are linked by particular roads depends very much on the use to which a user wants to put them. Take two extreme examples: (1) a logistics officer who needs to transport large volumes of heavy traffic quickly between A and B, and (2) a hiker who wants to walk pleasantly between A and B. If A and B are linked only by a footpath, there is no link for the logistics officer, but a very good link for the hiker; if, however, they are connected only by a 6-lane expressway, there is no link for the hiker but a good link for the logistics officer. The apparent network that should be represented in any display is different in the two cases.

The interesting situation is the case in which A and B are connected by a two-lane highway. In this case, the logistics officer might consider there to be a link, but not a very good one, and so might the hiker. This situation is best represented by asserting that the road has a fuzzy membership in the class "link", and that the degree of membership depends on the user's intentions for the network. This membership function therefore ought to be importantly represented in any display of the network for that user at that time. It is, in a way, shown on conventional road maps, in that expressways are shown differently from two-lane highways, gravel roads, and footpaths, and scenic roads are marked with green. (Topographic Maps are discussed in <http://www.physicalgeography.net/fundamentals/2d.html>, and various road symbols are shown in <http://www.ordnancesurvey.co.uk/oswebsite/education/pdf/25knewEGMlegend.pdf>.) To provide this variegated

symbology is probably about as well as can be done if the map maker does not know whether the user will be driving with intent to go fast, driving with interest in scenery, or will be hiking. But there are more possibilities than those for what the user might want of the road link. For instance, travel time might be of interest to both the hiker and the logistics officer, but their actual time expectations would be very different, and hard to enter on a map without inducing unwanted visual clutter.

Returning to the four-towns example, it seems clear that the notion of a "node" ought to be fuzzy, too. Consider the possibility that a town grows up around the crossroad (town E). If the roads are 6-lane expressways, the growth of the town increases traffic in its region of the expressway, thereby reducing the quality of the link for the logistics officer. Likewise, if the roads are footpaths, the quality of the link for the hiker might at first be improved (offering refreshments at the newly built pub) and then deteriorate, until as town E grows, it might present a block for the hiker wanting to go from A to C. At that stage, the hiker sees no link between A and C, but there are links between A and E and between C and E.

The connection between A and C does not lose its membership in the class "link" all at once. It does so smoothly over time, while the memberships of AE and EC (which use the same physical roads) increase their memberships in the class. There is a stage in the development of the five-town array when the connections AE, EC, and AC all have memberships between 0 and 1 in the class "link." In the definition, therefore, the notion of "passing through a third town" also is fuzzy. An entity has a fuzzy membership in the class "node". As Town E grows from a crossroads pub to an industrial powerhouse, its membership in "Node" increases from zero to unity, and indeed, as the town grows, it could split into several nodes, the split itself being fuzzy.

Networks being representations of relationships, the same argument can be extended to apply to many different kinds of physical and conceptual networks. Both the relationships and the entities that may be related can be fuzzy, and that fuzziness may well depend on the momentary interests of the user. For example, an intelligence analysis seeking potential terrorists by examining patterns of communication may be better served by a display that shows the likely degree of influence of one person on another than by one that shows crisply whether they communicated.

### **2.3.2 Fuzzy Chains and Cycles**

In a crisp network, if there is a link between nodes A and B, and another between B and C, a path exists along the chain between A and C. The situation is less clear if the network is fuzzy. Suppose the connections between A and B, and between B and C have respective memberships of 0.4 and 0.8 in the class "link", what is the membership of the path between A and C in the class "chain"? It may depend on the user's requirements.

One possibility for the membership of a path in the class "chain" is that it is the minimum of the memberships of the individual connections in the class "link". In the foregoing example, the path A-B-C would have a membership of 0.4 in the class "chain". Another possibility would be that the path would have a membership in "Chain" equal to the average of the connector memberships in the class "link". The first approach follows the idea that a chain is as strong as its weakest link, and would be suitable if the link membership had to do with its passability for traffic, whereas the second might be more appropriate if the link membership were related to the time it might take traffic to pass and the question interesting the user were the time to traverse the path.

## **2.4 Embedding Fields**

The mathematical analysis of a network or graph considers it to be an abstract object with a certain pattern of edges and vertices that have no properties other than being edges and vertices. The networks that people ordinarily want to visualise are abstractions from the real world, such as roads, friendships, packet traffic among computers, infection contacts. The mathematical properties of abstract networks apply to these representational networks, but there is much more to be said about the representational networks than about the abstract graphs. Roads have load limits, traffic volume capacities, and speed limits, none of which apply to an edge in the abstract graph that symbolizes the road connections, although even in a crisp network, the concept of link length or strength may be used.

A real network exists in some supporting context. A road network exists in a landscape; the World-Wide Web is based on a packet-switched network that is itself based on computers that run the TCP-IP protocols; a network of social contacts is supported by geographic proximity and communication possibilities; an infective network is the same, without the communication support, but with the addition of infective vectors (e.g. insects, birds). This kind of supporting context for a network can be encapsulated in the concept of an "Embedding field".

An embedding field is often a physical substrate for a network, as in the examples above but it need not be. One can generate a network of the syntactic relations among the words of a text or a speech, for which the embedding field is not the marks on the paper or the vibrations in the air, but the conceptual identities of the words. The words provide the embedding field for the syntactic network.

The same embedding field can support more than one network. The words of a discourse can be linked not only into a syntactic network, but also into a semantic network that not only omits many of the function words, but also has links between words too far separated for them to have any syntactic connection whatever. Likewise, the same landscape can be an embedding field for a road network and for a network of river drainage or for a network of vegetation-prey-predator relations in an ecology.

#### **2.4.1 Embedding and Inheritance**

Embedding fields can be hierarchically related quite analogously to the hierarchic relations of the inheritance structure of software objects. Properties and processes that belong to a parent are likely to be inherited by the child, whether the child be a software object or an embedded network. Consider a wired computer network. A network of wires is (trivially) embedded in the geographic 3-space of the world; a network of TCP/IP-capable computers is embedded in the wired network but is not identical to it, since some of the connected computers may not be running the TCP-IP protocols; a message-passing network is embedded in the network of TCP/IP-capable computers; the World-Wide Web is embedded in the message-passing network, and also embedded in the message-passing network are many networks of users who communicate frequently to each other by e-mail.

For software objects, each level of inheritance builds on the capabilities and may loosen the restrictions defined by the parent; for embedding fields restriction loosening does not happen, though capability enhancement does. A packet-switched network cannot transmit packets faster than is allowed by the bandwidth of the underlying hardware. What embedding does is to rearrange the elements of the embedding field, to correlate them and thereby to reduce the inherent information-theoretic capacity of the embedding field while enhancing the specificity of the child with respect to the parent. This enhancement provides additional capabilities to the child that were not properties of the parent.

#### **2.4.2 Embedding field for display**

The concept of an embedding field applies not only to the networks themselves, but also to the displays on which they may be represented. Mapping between the hierarchy of embedding fields of the display and those of the network to be displayed can suggest useful methods of displaying the properties of interest to the user.

For example, the hardware of a computer screen permits every pixel to vary over the whole gamut of colour 30 times per second. That very large potential information rate is unusable by humans, who rely on spatial and temporal correlations to make sense of the world. Treating the display surface as an embedding field, a 3-D representation may use temporal correlation (changes of apparent viewpoint), spatial correlation (objects spanning more than one pixel), spatial correlation again (fading of "distant" objects, and so forth. The display of objects in 3-D is embedded in the computer screen display. A display of a network using blobs connected by lines specializes the more general correlations, further restricting the possibilities for display variation, but at the same time enhancing the human's ability to make sense of what is displayed.

The concept of embedding, at least when applied to display, leads naturally to an information-theoretic analysis of the display possibilities and limitations. The same information-theoretic analyses may be applicable both to the networks being displayed and to the abilities of the human to make sense of the display.

When the specificity of a display is well matched to the background knowledge and the physiological capability of the human, the information transmission through the display about the thing displayed is optimized. When the embedding of the thing displayed corresponds to the human's conceptual understanding of the embedding field (e.g. the human understands what "packet switching" means), the human's understanding of the embedded network is optimum. When both conditions hold, the display is "transparent" or "natural".

### **2.4.3 Dimensionality of Embedding Field**

An embedding field may have any dimensionality, from zero upwards. If the embedding field for a network has dimensionality zero, the links of a network cannot be within it. They must be entirely conceptual, having no physical representation. Two examples of this situation can be given on the same zero-dimension embedding field, a continuous piece of text. The words of this text form nodes in a syntactic network that relates their formal dependencies. The syntactic links among the words have rich properties, but those properties are entirely conceptual, having no representation within the text itself. A second, quite independent network with this same embedding field is the network of semantic relationships among the words. Whereas the syntactic network includes all the words, and most of the links are between words nearby in the text, the semantic network omits many function words (such as "the" and "and"), and has long-distance links that can relate words many pages apart in the text. The words may have physical dimension as marks on paper or sounds in air, but they have zero dimension considered simply as nodes of the syntactic or semantic networks.

If the whole network is physically embodied, it must have an embedding field of dimensionality greater than zero, in order to accommodate the links. In that case, the properties of the network are constrained by the properties of the embedding field, as, for example, the rate of message passing along a wire connection is constrained by the bandwidth of the communication link over the wire.

## **2.5 Dynamical Behaviour**

### **2.5.1 Traffic Dynamics and Network Dynamics**

The dynamics of networks has two distinct aspects: the dynamical evolution of network structure, and the dynamical behaviour of traffic on a network of given structure. Of course, the dynamics of traffic on a network of variable structure may be of considerable interest, too, but each aspect needs to be addressed as a problem of its own. Traffic dynamics is of primary interest when considering the problems of detecting intrusions in computer networks, or of analyzing road traffic patterns, whereas the changing structure of a network may be of equal interest in counter-terrorist intelligence applications, or considering the construction or widening of roads.

### **2.5.2 Traffic dynamics and loop behaviour**

Traffic dynamics depends not only on external traffic sources but also on the feedback loops implicit in the network. A feedback loop exists when an event that propagates from one location in the network has an effect later back at the same point in the network. If the return effect augments the influence of the initiating event, the loop has "positive feedback", whereas if the returned effect diminishes the effect of the initial event, the loop shows "negative feedback." Networks that contain positive feedback loops are ordinarily unstable in the absence of nonlinearities that dampen the influence of extreme values. Loops with positive feedback will ordinarily show behaviour that becomes exponentially more extreme, whereas loops with negative feedback will tend to stabilize.

The detailed behaviour of a network depends on the particulars of its components and its structure, in particular the time-delays imposed by its nodes and links, the external influences that inject events at certain points in the network (such as the advent of new traffic onto a road network), and the effects of non-linearities. Even quite

simple networks with nonlinearities may exhibit a variety of distinct behaviours with attractors that range from fixed-point, through oscillators of different lengths, to fully strange (chaotic); such a network may sometimes be switched among these modes simply by the injection of traffic at critical moments.

Most networks that have nonlinearities are liable to display chaotic behaviour under suitable conditions. These considerations are quite important in social networks, in which positive feedback can come from opinions disseminated by one person that influence the the opinions of others, whose opinions then return to influence the first person. Sometimes such feedback loops in social networks engender outbreaks of dangerous mob behaviour that can result in great destruction.

## **2.6 Information Theoretic aspects**

The information-theoretic aspects of networks, of network dynamics, of embedding inheritance, and of displays for network visualisation by humans are likely to be important, but have not yet been addressed by the Working Group. We keep the headings as place-holders.

### **2.6.1 Network Dynamics and Display as an Information-theoretic Issue**

Bjørke, at the Toronto Workshop (IST-043/RWS-007), presented an information-theory based approach to the display of uncluttered road network maps. This work will form part of this section.

## **2.7 Taxonomies of Data and Display types**

The Final Report of IST-013/RTG-002 (*Visualisation of Massive Military Data Sets: Human Factors, Applications, and Technologies*, NATO RTO-TR-030, known as the HAT Report) provided a taxonomy of data types, reproduced in Table 1 from Table 3.1 of the report.

Table 1. Summary of Data Types				
Acquisition	Streamed	regular		
		sporadic		
	Static			
Sources	Single			
	Multiple			
Choice	User-selected			
	Externally imposed			
Identification	Located			
	Labelled			
Values	Analogue	scalar		
		vector		
	Categoric (classic)	symbolic	linguistic	
			non-linguistic	
		non-symbolic	linguistic	
			non-linguistic	
	Categoric (fuzzy)	symbolic (non-linguistic)		
		non-symbolic (non-linguistic)		
Interrelations	User-structured			
	Source-structured			

This taxonomy identifies six characteristics of data, each of which may take on different kinds of value. The particular values for different attributes often affects the choice of ways to display the data. For example, if the data has symbolic values, it will probably be best displayed in some way involving symbols, whereas if its values are analogue, graphs or continuum displays may be more appropriate. Likewise, streamed data often implies a different kind of display than static data, if only because streaming carries the implication that a significant event may occur at any moment, and the user should be able to detect the occurrence of the event when it happens. In connection with networks, the topology of a network is likely to be static within the time frame of the analysis, but traffic on the network is more likely to be streamed, which suggests that a different kind of display might be suited for analyzing network structure as opposed to network traffic.

The [HAT Report \(RTO-TR-030\)](#) considered a taxonomy of display types, shown in the Table (Table 3.2 of the HAT Report).

<b>Table 3. Summary of Display Types</b>		
Display Timing	static	
	dynamic	
Data Selection	User-selected	
	Algorithmically directed	
Data Placement	Located	
	Labelled	
Data Values	Analogue	scalar
		vector
	Categoric	linguistic
		non-linguistic

Some of the data types map naturally onto the display types. For example, symbolic linguistic data may well be displayed in a categoric linguistic manner, such as text or tables. The HAT Report listed a few of these natural mappings, as shown in Table 2 (Table 3.3 of the HAT Report).

**Table 4. Some examples of mapping data types onto display types**

Data type	Display type	Comment
Streamed	<b>Dynamic</b>	The user ordinarily wants to act when some event occurs
Located 2-D or 3-D	<b>Located</b>	The display is a 2-D or 3-D map of some attribute(s) of the data. If the location identification of the data are in higher dimensional spaces, there is no such natural mapping. Tricks must be used.
Labelled	<b>Labelled</b>	The display is likely to be tabular, or some kind of a graph such as a histogram or pie chart
Analogue scalar	<b>Analogue scalar</b>	Even if the data are identified by label, its analogue values map naturally to analogue display variables such as the length of a line or the brightness of a pixel.
Analogue vector	<b>If 2-D or 3-D, Analogue vector</b>	A 2-D attribute can be mapped onto an area display, a line with length and orientation, a colour hue, a sound location, a sound intensity and pitch, and so forth, all analogue vector attributes of the display. A 3-D attribute can similarly be mapped into a volumetric display. Higher dimensional analogue attributes can be displayed, but the mapping is less obviously "natural."
Categoric	<b>Categoric</b>	Categoric data values have no natural relation to analogue display values, and must be displayed categorically. The categoric display attributes may or may not map "naturally" onto the categoric data attributes. This kind of mapping is usually considered to be "cognitive metaphor."

### 3. SUMMARY: BRINGING THE THREADS TOGETHER

One of the uses of a Framework is to help designers and users of displays to ask the right questions before settling on a design or evaluating the usefulness of a design already created. The first thing that must be determined when designing or evaluating a display, for a network or for anything else, is to ask what the user's purpose is in wanting the display. What is it about the network that, if visualised by the user, will help realize that purpose? Only after that question is answered can the value of any particular method of representing the network be even considered. Accordingly, the Framework should commence with a suitable way to ask that question.

The VisTG Reference Model offers a structured representation in which the user's requirements for visualising the data, the user's requirements for control of the display, and the interface itself are interdependently considered. For any particular usage, such as the visualisation of networks, the VisTG Reference Model must be specialized, and that is what the Reference Framework for Network Visualisation is intended to do.

Networks have properties that are independent of the user's visualisation requirements. Some of these properties relate to the structure of the network, some to individual elements, and some to the traffic on the network. In the abstract, a network may be only a collection of nodes, some of which are connected by links, but in the real world nodes may act as transformers, have delays and probabilistic behaviour, and may even be fuzzy in the sense that an entity may only poorly be what the user might think of as a node or a link. Links have traffic limitations, offer probabilistically variable transit times for traffic from the originating node to the receiving

node, and may be fuzzy in the same sense that nodes may be fuzzy. The degree to which an entity in the real world being represented satisfies the user's criterion for "node" or "link" will depend on the user's purposes. What is a node or a link for one user at one moment may not be one at all for another user at another time.

The notion of "fuzzy" links and nodes offers a potential connection between the mathematical world of the abstract network, the world of real networks (whether their existence is conceptual as in a social network, or overtly physical as in a wired computer network), and the user's requirements. The status of a potential connection as a "link" depends on the use to which it will be put. A normal telephone line is capable of transmitting gigabytes of data, but does it only slowly. A person interested only in interactive conversation may see this as a perfect link (membership 1.0 in the class "link") while someone interested in transmitting high-definition movies would consider it not a link at all (membership 0.0 in the class "link"). Between these, might be a person interested in occasionally transmitting moderate size files, which the connection will accommodate, though slowly. For that user, the connection might have a membership 0.5 in the class "link". The same person might have all of these requirements, at different time. So, the membership function of a connection in the class "link" depends largely on the user's requirements of the moment, and may vary over time. Similar considerations apply to the class "node".

The importance of the foregoing is that it implies that although a network may have well-defined physical properties, those properties are insufficient to determine its properties from the viewpoint of a person who has specific interests in the network. This has always been known, but consideration of the fuzzy idiosyncratic properties of the network elements may permit a mathematical approach to describing the requirements for displays that aid the user's visualisation.

Networks in the real world exist in a context. They are supported by, and are affected by their context, to which we give the name "embedding field". Displays also can be treated as being supported by and existing in the context of embedding fields. Embedding fields have an inheritance hierarchy analogous to the inheritance hierarchy of software objects, but unlike the inheritance of software, the inheritance of embedment is restricted by information-theoretic considerations. Information-theoretic considerations apply also to the interpretation of displays by humans. The implication is that the concepts of embedding fields, like those of fuzzy representation, may find a useful place in the development of a Reference Framework for Network Visualisation.

We have suggested some of the issues that need to be considered in developing a Reference Framework for Network Visualisation, without presenting such a framework. We hope that collaborative work over the lifetime of IST-059 will allow us to construct a framework that will be valuable to future developers and users of network visualisation products.



# Characterisation and Showcasing of Network Visualisation Approaches for Command and Control

(Extended Abstract)

<p><b>Alain Bouchard, M. Sc.</b> Defence Scientist Defence R&amp;D Canada – Valcartier 2459 Pie-XI Blvd. North Val-Bélair, Quebec G3J 1X5 Canada <a href="mailto:Alain.Bouchard@drdc-rddc.gc.ca">Alain.Bouchard@drdc-rddc.gc.ca</a></p>	<p><b>Dr. Rudi Vernik</b> Research Leader Defence Science and Technology Organisation PO. Box 1500 Edinburgh SA 5111 Australia <a href="mailto:Rudi.Vernik@dsto.defence.gov.au">Rudi.Vernik@dsto.defence.gov.au</a></p>
---	---

## ABSTRACT

*Network Visualisation technologies are becoming relevant in Command and Control environments to help cope with the increased complexity of defence operations. A significant number of Network Visualisation technologies are currently available so the choice of a particular approach to support specific C2 tasks can be difficult. This paper describes an approach for characterising Network Visualisation tools in terms of domains of use, what they describe and how they present information using a reference model for visualisation. The paper also introduces a novel distributed system, named Imago, which supports the characterisation, discovery, showcasing, and evaluation of Information Visualisation approaches. We argue that the methods proposed in this paper could aid the process of selecting and deploying Network Visualisation tools for Command and Control activities.*

## 1.0 INTRODUCTION

In the ever more complex world of networked and asymmetric military operations, command staff needs to rapidly assimilate and understand a vast range of additional information including social and communication networks, resupply and logistics, belligerents profiling, etc. One would imagine that, given the vast array of visualisation approaches that have been developed and the critical and complex nature of defence operations, military decision-makers would be the early adopters of such technologies. Yet, relatively few command support applications use Information Visualisation approaches, in particular Network Visualisation approaches.

In recent years, surveys of Information and Network Visualisation technologies have been performed (e.g., Bouchard [1], Herman et al [7]), identifying over 70 different products, to facilitate the discovery of Network Visualisation technologies. However, most of these surveys are functionality based. This makes it difficult to identify technologies to support specific contexts of use. Also, limited support can be offered from these surveys for evaluating and transitioning various approaches.

Instead of using a taxonomy of Information Visualisation technologies such as provided by Card et al [3], the Information Visualisation Action Group (AG3) of The Technical Cooperation Program (TTCP) developed a reference model (cf. [5, 6]) for visualisation (RM-Vis) to characterise and showcase visualisation approaches. The use of a reference model such as RM-Vis allows the rapid identification of relevant approaches for particular activities, as well as providing support for evaluation and transitioning of the approaches in operational environments.

This paper first introduces the TTCP RM-Vis reference model and describes how it has been used for characterising network visualisations technologies in terms of domains of use, the descriptive aspects (i.e. what they describe) and the approaches that they use for presenting the information. We discuss our experiences in using these approaches for the development and use of a system called C2NetVis which characterises and showcases network visualisation technologies. The contexts of use and descriptive aspects of network visualisation approaches used in Command and Control environments are provided.

Finally this paper discusses how we are extending on the work done by TTCP as part of Project Imago. This project is developing a web-based distributed environment that can be used to collaboratively define, prototype, evaluate, and transition visualisation approaches for C2.

## 2.0 REFERENCE MODEL FOR VISUALISATION

The TTCP Action Group on Information Visualisation developed a Reference Model framework (cf. [6]) for the application of Visualisation approaches (RM-Vis), which defined to support the characterisation, identification and showcasing of visualisation approaches in the C3I domain. This framework has been used to characterise visualisation solutions in terms of their context of use, the representation and presentation techniques used, and key features of tool support provided such as types of user interactions and deployment support.

The RM-Vis has three key dimensions:

- The **Domain Context** is a model that defines the focus for the application of visualisation approaches i.e. *where* visualisation approaches will be applied, *who* will be supported, and *why* the approaches are needed.
- **Descriptive Aspects** (DA) define what needs to be described for particular domain contexts. For example, DAs could be defined in terms of the various elements (or things) that are of importance, the relationships between those elements and particular attributes that describe the elements and relationships.
- The **Visualisation Approach** dimension defines how the required information can be provided through computer-based visualisation. Approaches are characterised in terms of the visual representations used (e.g. graphs, charts, maps), visual enhancements (e.g. use of overlays, distortion, animation), interaction (direct manipulation, drag and drop, haptic techniques etc), and deployment which includes the computing environment (display devices, COTS software) and advanced deployment techniques such as intelligent user support and enterprise integration.

In parallel to the development of the reference model, the members of AG-3 created three instantiations of a database containing views referencing the model. C3I-Vis, MIL-Vis, and G-Vis were created to characterise and showcase visualisation approaches in the C3I, Military, and general domains. This paper discusses how these approaches have been used in the development and use of a system called C2NetVis which characterises and showcases network visualisation technologies.

## 3.0 USE OF NETWORK VISUALISATION FOR COMMAND AND CONTROL

Using Network Analysis techniques, especially Social Network Analysis, has become increasingly relevant approaches in recent years, particularly with the advent of Network Enabled Operations and the current focus on counter-terrorism. The battlespace is becoming more and more networked. Intel analysts are using these techniques to depict relationships among people and their behavioural activities (eg. travel, banking, phone

calls, transfer of illicit material) to unearth suspicious behaviours and prevent unfortunate events. The same techniques are used to prosecute orchestrators of terrorist events, as described in Bouchard [2]. These graphs of relationships are often shared among staff officers and commanders for decision-making and situational awareness.

Social Network Analysis (SNA) is the most commonly used Network Analysis technique in Command and Control. SNA consists of representing people in graphs with their formal or informal relationships. Dekker [4] presents many applications of SNA in a C4ISR context, such as delays analysis and identifying best routes in logistics operations. Similarly, Network Analysis techniques are used in urban operations to identify shortest paths, safest paths, centre of gravities, vulnerabilities, and so on.

As with decision-makers in other enterprises, defence command staff require information to support activities related to planning, analysis, and synchronised action. This information includes aspects such as the disposition and status of force elements such as ships, planes, and people.

Most of the tasks involved in creating these graphs of networked entities are manual but sometimes automated or semi-automated algorithms are used to extract information from unstructured documents.

The list of potential uses of network analysis and visualisation approaches in a C2 context is extensive. These techniques can be used to provide new insights into complex arrangements which are represented in graphs showing various entities, their attributes, and relationships. Since there are a large number of available network visualisation approaches available, it becomes time-consuming to identify an approach that is appropriate to a particular domain context, user, and representing the appropriate information. In order to address that problem, the following section presents a characterisation of network visualisation approaches to help the discovery and evaluation of network visualisation approaches

#### **4.0 CHARACTERISATION OF NETWORK VISUALISATION APPROACHES**

A fourth instantiation of the reference model database has been developed, called C2NetVis, to contain Network Visualisation approaches for the Command and Control domain context. The first step was to describe Network Visualisation approaches in term of the reference model. The reference model consists of a hierarchy of domain contexts and descriptive aspects. Characterising a Network Visualisation approach in term of the reference model requires selecting attributes from the hierarchy. As an example, an approach could have the characteristics shown in Table 1.

**Table 1. Characterisation of NetMiner approach**

Domain Contexts	<ul style="list-style-type: none"><li>• Category: Defence</li><li>• Area: Intelligence</li><li>• Activity: Monitoring, Alerting, Reporting</li><li>• Role: Intel Analyst</li><li>• Task: Find belligerent</li></ul>
Descriptive Aspects	<ul style="list-style-type: none"><li>• Location, Time, Sex, People, Relationship (friends, enemies), Assets</li></ul>
Visualisation Approach	<ul style="list-style-type: none"><li>• Product name: NetMiner</li><li>• Representations: Graph/Network, Tree, 2D, 3D, text</li><li>• Visual enhancements: Nested nodes</li><li>• Interaction: Direct manipulation, DnD, CnP, zoom, pan</li><li>• Deployment: Windows environment</li></ul>

The database has been populated with products surveyed from different studies and characterised using the framework. This C2NetVis database may then be used to identify and showcase Network Visualisation approaches relevant to a particular context and also to evaluate the relevance in this context.

Even though scientists and analysts may populate and navigate the C2NetVis database itself, and similarly the other three database instantiations, the main interest in the database is the collaboration among many individuals to share approaches, evaluations, and for showcasing purposes. As the number of users and operations on the database increases it becomes difficult to synchronise remotely located instances. An ideal configuration would consist of a unique database instance shared by many users, distributed over an enterprise bus. The project Imago has been designed to address this issue.

## **5.0 FUTURE WORK**

Project Imago is a project being lead by the Defence Science and Technology Organisation (DSTO) to support the development, evaluation and transitioning of Information Visualisation approaches to support of military Command and Control (C2). Imago is a distributed environment that can be used to rapidly prototype, evaluate, and transition visualisation approaches for C2. The platform will provide a means of integrating and sharing the output of visualisation tools, storing, accessing and managing showcase examples of visual representations via an underlying reference model, and providing access to underlying data sources provided through simulation, representative data, and/or operational data.

Using the C2NetVis as a reference model and its related database, the Imago system might then be used to characterise, showcase, and evaluate Network Visualisation approaches for the Command and Control domain.

## 6.0 REFERENCES

- [5] Bouchard, A. (2004). Link Analysis & Visualization Product Review. Technical Memorandum. DRDC Valcartier TM-2004-175.
  - [6] Bouchard, A. and Gouin, D. (2004). Using Link Analysis to Track Terrorist Leaders. NATO IST-043/RWS-006 Workshop on Visualisation and the Common Operational Picture (COP). Toronto, Canada
  - [7] Card, S.K., Mackinlay, J.D., and Shneiderman, B., (1999). Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers.
  - [8] Dekker, A., (2002). Applying Social Network Analysis Concepts to Military C4ISR Architectures. *Connections*, 24(3), 93-103.
  - [9] Gouin, D., Evdokiou, P., and Vernik, R. (2002). A Showcase of Visualization Approaches for Military Decision Makers. NATO's Research & Technology Organisation (RTO) Information Systems Technology panel Workshop: Massive Military Data Fusion and Visualisation: Users Talk with Developers. Halden, Norway.
  - [10] Gouin, D., Vernik, R. J., Evdokiou, P., Houghton, P., Barnes, M., Monk, D., and Rosenblum, L. (2002). Information Visualization in C3I: Final Report of TTCP C3I AG-3. DOC-C3I-AG3-1-2002. TTCP C3I AG-3.
  - [11] Herman, I., Melançon, G., and Marshall, M.S., (2000). Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 24-43.
  - [12] Vernik, R. (2000). A Proposed Reference Model Framework for the Application of Computer-Based Visualisation Approaches. Presentation at the NATO Workshop on Visualisation of Massive Military Datasets, Quebec, Canada.
- The work described in this abstract is cleared for presentation to NATO audiences at a NATO/PfP UNCLASSIFIED level.
  - If work is sponsored by a Government Agency, identify the organization and attest that the organization is aware of submission.
  - The abstract is technically correct.
  - The abstract is NATO/PfP UNCLASSIFIED.
  - The abstract does not violate any proprietary rights.



## ***Security & Defence II***



# Scalable HAIPE Discovery

**Shu Nakamoto**  
202 Burlington Road  
Bedford, MA 01730, USA

[nakamoto@mitre.org](mailto:nakamoto@mitre.org)

## ***EXTENDED ABSTRACT***

With the recent advent of IP-based encryption and the introduction of High Assurance Internet Protocol Encryption (HAIPE), the transition to IP encrypted networks has begun in earnest. Legacy encryption using link level encryption operates below the IP layer and has no impact on issues such as peer discovery or routing. However one pair of encryptors is needed for every communication link and all communications and network equipment has to be managed by cleared personnel within appropriately protected facilities. Use of IP encryption permits the possibility of using commercial network services and significantly reduces the number as well as the total cost of these encryption devices. It also allows for statistically multiplexing different types of traffic at different classification levels, making better use of available bandwidth.

However, with those advantages, some new services are needed to regain network transparency in the presence of IP encryption. The first of these services (and the subject of this paper) is a peer discovery function that is scalable to global levels. In addition to discovery, it is also important that segmented (red) enclaves (as part of a “striped” environment) also know how to route packets to their ultimate destination within their own interior gateway protocol (IGP) domain. This is important since segmented red enclaves may be part of the same autonomous system (AS) but may be unable to participate in the IGP without deliberate (and sometimes inefficient) measures (such as tunneling). There are many reasons why “segmentation” may occur: use of a commercial satellite link, traversing from one secure facility to another on a base network, operating over a radio frequency network, and so on. Each of these network segments need to be secured (in this case, via IP encryption, which causes the segments to exist). The term “striped” in this context refers to traversing from a red (or classified) network to a black (or unclassified) network to a red network in a multiple concatenated fashion (i.e., red-black-red-black-red ...). The boundary between red and black sides is assumed to be protected via a HAIPE device. As we transition from a single (link level encrypted) red network (our current legacy environment) to a potentially segmented (IP level encrypted) striped network, the peer discovery problem becomes complex. The basic question needing to be answered is “how does one find the ciphertext or black IP address of the fronting IP encryptor that ‘sits in front of’ the destination workstation that I am trying to reach?” This information is needed to establish the proper security association (SA) between the associated IP encryptors. Now extend this scenario such that the data packet must transition from a red to black to red to black to red networks (i.e., the striped network) prior to reaching its destination.

The discovery mechanism was first addressed by examining the discovery process through a single black “cloud” and extended to a “striped” environment. Version 3 of the High Assurance Internet Protocol Encryption Interoperability Specification (HAIPE IS) is used as a foundation capability throughout this paper. In v3, a Routing Information Protocol (RIPv2 or RIPv6) capability has been added to permit the HAIPE device to “discover” the networks that it protects (passive RIP participation). This alleviates the need to manually enter the network prefixes protected by this HAIPE device. Upon retrieving these prefixes, the HAIPE device can register (another v3 function) with a local Peer Enclave Discovery (PED) server (of which an experimental prototype has been developed at MITRE). Each PED server can “manage” approximately 10,000 networks (and possibly more). A HAIPE device could be fronting for one workstation or 200+ prefixes (the number of networks typically in a medium size military base). The sizing for the PED server is a function of the quantity of prefix/HAIPE associations, the frequency of queries being serviced (i.e., rate of queries – one per new security association or SA), and frequency of change (prefix additions or losses as a result of changes or unit mobility).

The PED server, as currently prototyped and performance tested, supports a group of HAIPE devices collectively fronting as many as 10,000 networks (per server). There are no assumptions with regards to address ranges being serviced by a given PED server, i.e., no address summarization is required at the HAIPE (enclave) level. Each PED server will, in turn, register its set of prefixes to one or more Referral Servers (similar to the .com or .org level in DNS terms). These Referral servers (RS) will “own” some percentage (say 10% to 20%) of the total IP address space (currently, implying 5 to 10 servers total on a global scale – not counting redundancy and server mirroring). The allocation of prefixes to server can either be manually managed based on performance monitoring or dynamically balanced. A single Referral server (not counting backups or mirror sites) holds peer discovery information for a given prefix (organized along prefix groupings). The HAIPE devices keep sync with their PED server which, in turn, keeps sync with their Referral servers. In this respect, the model more closely resembles reverse DNS rather than regular DNS. Each PED server will register with its corresponding Referral server the fact that “it knows about” a given IP address range (prefix). The PED server also summarizes this information to further minimize the volume of data being sent. The actual number (of Referral Servers) would be a function of administrative control and performance. A single Root Server would keep track of all Referral Servers and permit dynamic management as well as other potential discovery services. The local PED server is also being designed in 2006 to support local peer-to-peer discovery without access to a Referral or Root server – to support mobile units that can “come together” dynamically even when no “reachback” communication is available.

While outside the scope of this paper, this project is exploring other “black core” issues such as implementing QoS without “advertising” DiffServ values, full support for multicast routing, supporting traffic engineered “paths” that meet minimum protection attributes (quality of protection), anonymous routing for hiding source and destination IP addresses of HAIPE devices, different types of “red gateway” concepts (for computer network defense or performance enhancing proxy support), and other security related mitigation concepts.

We believe our approach provides a realistic way to use planned (version 3) HAIPE devices in a true scalable networked environment while addressing many of the issues that IP encryption introduces. These concepts will work for any IP encrypted network including IPSEC-based networks. From a performance viewpoint, our design exhibits enhanced DNS-like behaviour with three queries (maximum) for initial conditions (estimated time <300 ms including “typical” latency) for a global access model (currently being confirmed in lab testing and simulation analysis). Conventional caching techniques makes discovery virtually transparent (except at start up). From a scalability viewpoint, the design is globally scalable to many millions of networks. The hierarchical architecture permits growth from one to three tiers in a top down or bottom up fashion (and permits “disjointed” evolution similar to “real world” DNS implementation). Our approach keeps HAIPE and PED/Referral/Root server design simple, extensible, and scalable (with near zero configuration requirements). The PED, Referral, and Root servers have been prototyped at MITRE and are currently undergoing testing and scalability analysis. A HAIPE (ver. 3.0) emulator has been developed and tested with the discovery service with full striped network support. We are also studying deployment concepts of DNS as a possible model for HAIPE discovery deployment. The simple design is highly survivable due to its distributed architecture (based on lessons learned from DNS implementation). Finally, the design takes special consideration for mobility and permits rapid updates to take place on a global level.

# Custom Ontologies for Expanded Network Analysis (Extended Abstract)

Amy K. C. S. Vanderbilt, Ph.D. and George Strauss

4465 Brookfield Corporate Drive, Suite 200A,  
Chantilly VA 20151, USA

[avanderbilt@wvtec.com](mailto:avanderbilt@wvtec.com), [gstrauss@wvtec.com](mailto:gstrauss@wvtec.com)

## INTRODUCTION

*This paper discusses a new approach to answering Requests for Information (RFIs) from military commanders, intelligence analysts, individual soldiers and others received by reach-back information and intelligence collection repositories. This new approach avoids the previous ideals of either searching out a set of a thousand documents or building one large all-encompassing ontology and instead embraces the concept of custom ontologies based on each users query and returns to that user a concise and organized knowledge set along with visualizations that invite exploration and facilitate assimilation.*

## 1.0 PREVIOUS APPROACHES TO RFI RESPONSE AND KNOWLEDGE DISSEMINATION

When searching a large corpus of free text documents for just the right information, the first thing many try is to do a web-type search using a search engine such as Google, Alta Vista, or other such document search and retrieval algorithm. These search engines are queried with a short word string that is terse by necessity and therefore fairly general in concept. The simplest of these searches are based on scoring functions and return an ordered stack of documents that have a certain percent match to the query. The result is a set of one hundred to ten thousand or more documents that match the query to some degree and may or may not contain the specific type of information the user seeks. This satisfies the need to believe that we have taken in all available information on a topic, but forces the user to spend time paging through each document in the hopes of finding the dozen or so kernels of knowledge, and associated relationships that they seek. Essentially, the user is performing information extraction in their heads on every document that may contain the targeted information. The time-consuming nature of this process has led to another method at the other end of the spectrum.

Instead of extracting information document by document in the users head, an alternate is to pre-extract all possible information from all possible documents related to a subject or set of subjects and form a large-scale ontology to house the information and relationships. This approach was exemplified in the Cyc research project that began in 1984. This project and others continue to evolve into ever larger editions of large-scale living ontologies [1].

Although this approach returns more precise information, the problem with this approach is that the ontology must grow and be constantly modified to allow for new concepts and to update old ones. This leads to a number of problems from size to manageability, ease of use and others.

## 2.0 CUSTOM ONTOLOGIES AND THE URBAN WARFARE ANALYSIS CENTER

At the beginning of a recent project, we were forced to choose which method would be the basis of our technical approach. Dissatisfied with both extremes, a middle ground was devised. Building on the idea of information fusion for common operational understanding [2], the concept of custom user ontologies was derived and detailed. At the cornerstone of the concept is the idea of creating a populated ontology complete with entities, relationships and appropriate visualizations for each user and each query. The initial absurdity of the idea may seem to be rooted in the perceived onslaught of ontologies that would result, or the potentially long wait time required to search a large corpus for the specific entities and relations. While it is true that many

ontologies will result, each one is small. For any one user, a single query is looking for relatively little information: events in a certain location over the last month, the social network of a person in a photograph, new techniques for new tactical situations, and so on. The user craves not the drink from an informational fire hose yielded by the above methods, but a small bowl of concise knowledge that answers their curiosity of the moment. The issue of computability time can be solved in a number of ways from converting documents to text only, removing filler words, etc.

The concept of custom ontologies is the basis for the Urban Warfare Analysis Center (UWAC). The challenge is to provide situational awareness, not by a snapshot in time, but as a continual understanding by the Warfighter of his environment, the players in it and how it is changing. The process is as follows. Users submit a Request for Information (RFI) to the UWAC analysts by phone or email or via the UWAC website. Users submit requests in their own words...in their own natural language. The UWAC Analyst then enters the query into the system, and oversees the software as it builds the custom ontology and initiates information extraction based on that ontology. Carnegie Mellon University's AutoMap and Ora software tools [3,4], integrated with organic UWAC algorithmic components perform the search space reduction, ontology construction and information extraction necessary to make custom user ontologies a reality.

Given the types of information in the ontology, appropriate visualizations of the ontology and its contents are rendered. Visualizations may be GIS or 3D landscapes, timelines or timelines blended with GIS, or architectural views of the ontology itself. Interactive visualizations that feed off of the rendered ontology are posted to a web page for the user to view and manipulate.

The relevance of this method to social network and other network analysis is of particular importance. When analyzing a social network for person X, more than a tree of associate persons is required. Users will likely need information about the events, groups, people and locations at what times associated with that person and those with which they are associated. Such an expanded social network may reveal far more about a person or social network, than would a people-only network.

## **4.0 REFERENCES**

- [13] Lenat, D. and Guha, R. 1990. Building Large Knowledge Based Systems. Reading, Mass.: Addison-Wesley.
- [14] Amy K. C. S. Vanderbilt, Robert I. Desourdis Jr.. Information Fusion for Common Operational Understanding. Extended Abstract. NATO RTO Workshop on Visualization and The Common Operating Picture (VizCOP) (IST-043/RWS-006). Toronto Canada, September 14-17, 2004.
- [15] Jana Diesner and Kathleen M. Carley. AutoMap 1.2 – Extract, Analyze, Represent and Compare Mental Models from Texts. CASOS Technical Report CMU-ISRI-04-100, January 2004, Carnegie Mellon University, Pittsburgh, PA.
- [16] Kathleen M. Carley and Jeff Reminga. ORA: Organization Risk Analyzer. CASOS Technical Report CMU-ISRI-04-106, January 2004, Carnegie Mellon University, Pittsburgh, PA.
  - The work described in this abstract is cleared for presentation to NATO audiences at a NATO/PfP UNCLASSIFIED level.
  - If work is sponsored by a Government Agency, identify the organization and attest that the organization is aware of submission.
  - The abstract is technically correct.
  - The abstract is NATO/PfP UNCLASSIFIED.

- The abstract does not violate any proprietary rights.

## Real-Time Extraction of Course Track Networks in Confined Waters as Decision Support for Vessel Navigation in 3-D Nautical Chart

Thomas Porathe, Information design, Dep. of Innovation, Design and Product Development, Mälardalen University, Sweden

### Extended abstract

In presentations at the 2002 and 2004 workshops, “Massive Military Data Fusion and Visualization: Users Talk with Developers” (IST-036/RWS-005), in Halden, Norway, and “Visualisation and the Common Operational Picture” (IST-043/RWS-006 ) in Toronto, Canada, a decision support system for nautical navigation was suggested. [1] [2]

From an information design point of view, the 3-D chart aims at reducing the cognitive workload of the navigator by presenting map information in a more effective way. This is done with the help of three methods: *the bridge perspective*, *No-Go Area polygons* and the *road metaphor*.

*The bridge perspective* allows the map data base to be presented in an egocentric 3-D “natural view” of the world surrounding the ship. The dynamic *No-Go area polygons*, colors water too shallow for the vessel at the current tidal level thus dividing the water area in a free and a forbidden area. Finally the *road metaphor* uses red and green “carpets” to present a network of traffic separated sea ways to the navigator.

The suggested sea ways visualises the existing navigational channel infrastructure. These are drawn by maritime authorities based on existing and preferred patterns of sea transportation. In a military context they could be alternative routes plotted with other objectives in mind. Common for both is that they are defined for a minimum draft of some kind based on some specific water-level. This means that for a particular vessel (with a draft deeper or shallower than the general case) on a particular point in time (with a water-level different from the specified) the possible track alternatives might differ. In deep water on the open sea this will be of minor importance as there is always a free set of possibilities. But in the confined waters of a complicated archipelago like the Scandinavian, the situation will be entirely different.

The factor that is determining the possibility to travel in a certain area will be the depth of the water. Improving bathymetrical sensing techniques will in the future lead to better resolution of underwater databases. By using the 3-D technique described in [1] a dynamic high resolution depth curve can be produced in real time. A detailed knowledge of depth, the own ships draught and the water level will safely open new tracks in the archipelago; this will be of particular importance for military units trying to avoid detection from different types of sensors or navigating autonomously.

The decision support system can help the navigator by suggesting different alternative routes to a given goal. The dynamic No-Go area polygon transposes a given region into free and forbidden areas. By making a geometrical computation from the present position to a given destination point the system can suggest different routes to the goal. The alternatives can be computed based on preferences such as shortest route, most weather protected route given wind speed and direction, most sheltered route from a stealth and radar detection point of view based on the height of surrounding islands or underwater topography, etc.

After deciding on which track to follow the track can be presented as for example a white carpet in front of the vessel and dynamically attached to the vessel so that if an evasive maneuver is made the track will compensate for the maneuver and continue to show the way from the new position.

For the less experienced civilian leisure domain such way finding guidance can be of great interest.

### **References:**

[1] Porathe, T. & Sivertun, A. (2002). *Real-Time 3D Nautical Navigational Visualisation*, presented at NATO's Research and Technology Organization's Workshop IST-036/RWS-005 Massive Military Data Fusion and Visualisation in Halden, Norway 10-13 Sep. 2002

[2] Porathe, T. (2004). *Visualizing the Decision Space of a Ship's Maneuverability in a Real-Time 3-D Nautical Chart*, presented at the NATO's Research and Technology Organization's Workshop IST-043/RWS-006 Visualisation and the Common Operational Picture (VizCOP) at the Canadian Forces College in Toronto, Canada 14-17 September 2004

## ***General/Theory II***



# **A three pronged approach for improved data understanding: 3-D visualization, use of gaming techniques, and intelligent advisory agents**

David L. Hall  
Michael McNeese  
John Yen

Magy Seif El-Nasr

College of Information Sciences and Technology  
The Pennsylvania State University  
University Park, PA 16802-6823  
[dhall@ist.psu.edu](mailto:dhall@ist.psu.edu)

## **Abstract**

A fundamental paradox exists in information fusion. The paradox is that information analysts are drowning in a sea of data but unable to obtain the knowledge that they need to address difficult problems. This has often been referred to as the data overload dilemma or more recently framed "cogmenutia fragmentosa". On one hand, an unprecedented capability exists to collect data via distributed sensors, commercial information providers, human sources, or Internet resources. Smart micro-scale sensors, wireless communications, and global Internet accessible resources enable the entire earth to be a potential information resource. Such information is available literally at the fingertips of the analysts. However, the wealth of data has not produced a commensurate improvement in analyst abilities. Analysts are literally swamped with data. They have a wide variety of choices to make as to what is *useful* and *usable*, given the context of what they are trying to *understand*.

This paper describes a three-pronged approach to improve information understanding including; (1) use of 3-D visualization and interaction techniques, (2) role-playing gaming (RPG) concepts, and (3) use of team-based intelligent advisory agents (cyber-advisors). The environment promotes rapid development and evaluation of hypotheses regarding evolving complex situations in an environment in which enormous amounts of data and information are available, but for which there is no clear mapping between observables and underlying threat conditions or activities. Use of advanced visualization techniques and gaming concepts assist in focusing the analysts' attention and promotes an interactive, creative analysis process in which hypotheses are formulated, evaluated, criticized, modified, and changed. The use of gaming techniques leverages the skills of new analysts, already experienced in gaming technologies.

For the application of crisis management and related applications, we have developed cognitive models (using a knowledge elicitation approach on experienced analysts), implemented software prototypes to demonstrate the concepts, and evaluated the effectiveness of the tools and techniques using human subjects in a living laboratory environment. The living laboratory environment links observations of analysts working in real environments, to development of cognitive models using a formal knowledge elicitation approach, and development and evaluation of prototype tools. The tools are tested in a simulation environment using teams of human subjects to quantitatively evaluate the effectiveness of the tools. A special laboratory environment and simulation tool (neoCITIES) is used for the evaluation. The implementation of these tools leverages several special laboratory facilities including: a synthetic environment applications laboratory containing a full immersion, 3-dimensional visualization facility, a laboratory for intelligent agents, and the user science and engineering laboratory. In addition gaming tools and visualization creation tools are used from our real-time aesthetics and experience laboratory.

This paper will introduce these three main concepts, describe examples of implemented prototypes and simulations, and present preliminary results.



## **Open Networks: Generalized Multi-Sensor Characterization**

**Alan N. Steinberg**

CUBRC, Inc.

This paper presents a novel method for characterizing the performance of information sources as necessary for data fusion and coordination in a net-centric environment. The concept, called Open Networks, explicitly estimates the performance, trustworthiness and allegiance of agents as a part of the general multi-sensor/multi-target state estimation process. The information network is construed as unbounded; with node agents having various degrees – and possibly time-varying degrees – of allegiance, common purpose, cooperativeness, information fidelity, controllability, etc. Agents share information with friends, foes and innocent bystanders alike, with varying degrees of cooperativeness and openness.

The approach builds on our recent work in developing systems for Situation and Threat Assessment. The work involves developing the theoretical and ontological foundations to enable the representation and recognition of relationships and of threat situations. Such applications require the system to exploit a wide range of evidence and a wide range of entity and aggregate behavior models. This is certainly the case in the many situations of interest that involve estimating and predicting human individual and group behavior.

This concept extends the notion of a Service-Oriented Architecture to include cases of partial and conditional cooperation. It extends the notion of Network-Centric Operations to include partially and conditionally participating agents; and, indeed, to non-cooperative information sources. It extends the conventional NCW notion of “power to the edge” to include networks without edges.

A sensor system’s characterization includes a random vector of bias terms to include – besides the usual platform and sensor state terms – terms for state estimation by the sensor system (e.g. target classification, composition, location, activity, capability, or intent estimation).

The estimation and prediction of states of targets and other external entities, as well of one’s own platform, its sensors and other system, is a single problem, amenable to a unified self-consistent solution. The evaluation of the system’s models of the characteristics and behavior of all of these external and organic entities is likewise a component of the single problem of estimating the actual world state.

The problem of characterizing information sources is present even in systems that have been designed and integrated as a unit. The problem is made more difficult when the performance of information sources cannot be assumed. This is the case in network-centric operations, in which calibration and registration are not easily performed.

It is even more the case when the cooperation or common purpose of information sources cannot be assured; e.g. when sources are agents with varying degrees of autonomy, raising the possibility of private agendas. The same factors can affect the estimation and controllability of assets having some degree of autonomy. Cases extend to the use of non-cooperating agents; e.g. enemy radars that “report” to our ELINT systems who and where they are; generally with no attempt to deceive.

Other cases include purposeful deception, to include such Information Warfare techniques as decoys, deceptive jamming and propaganda.

Finally, there are third-party agents – e.g. commercial news sources, reference texts, and the like – that may or may not be unbiased.

We formulate the problem as that of simultaneously estimating states of multiple targets and of the values of systematic biases in the information sources that corrupt the reported observables. Constrained problems of this sort have been addressed in the area of multi-sensor/multi-target geolocation.

We denote the entire collection of bias states across all of the collection platforms by a random vector. A multi-dimensional vector of bias terms can include measurement mis-calibration and biases in target state estimation by the sensor system (e.g. target location, velocity, classification, composition, activity, capability or intent estimation).

This is straight-forward in the case of continuous reporting elements such as the sensor's reports of target position and kinematic state. In the case of discrete reporting elements – the sensor's estimate of target type or activity, for example – an ontology-based vector of semantic can be applied, by which systematic biases can be captured.

Many different measures of similarity or dissimilarity are found in the classification literature. Many such techniques involve describing entities in terms of a set of attributes variables, which generally are assumed to be mutually independent. The problem of selecting a set of attributes and ensuring independence has led to a consideration of classification and attribute semantics. Furthermore, the problem of defining distance on discrete attributes is not particularly easier than on discrete entity classes. This intrinsically requires a consideration of ontological, or semantic, metrics.

It is not necessary that we select unequivocally among such techniques for the present purpose. However, for purposes of design and evaluation, we propose to use the Jiang-Conrath metric, which has been experimentally shown to perform well in comparison with other metrics.

# Reduction of complexity: an aspect of network visualization

Jan Terje Bjørke  
Norwegian Defence Research Establishment  
P O Box 115, NO-3191 Horten, Norway  
Email: [jtb@ffi.no](mailto:jtb@ffi.no)

## **Abstract**

### **Introduction**

Networks are topological structures composed of nodes and arcs. Often, networks are visualized by point symbols and lines illustrating the nodes and the arcs, respectively. These simple visualization techniques need to be refined when complex real world phenomena are to be visualized. For example, how to deal with the possible visual overload of nodes and arcs, or how to illustrate uncertainty or fuzziness in the information. The purpose of the present research is to identify strategies for the visualization of network information.

Several models of visual communication can be set up. Distinction can be made between visual communication and visual exploration. Visual communication deals with how to visualize results of different kinds of analysis, i.e., visualization in the case that the message is well defined. Another view is covered by visual exploration. In this case the message is not well defined and the analysis of structure or detection of important information is left to the map reader, i.e., tools for exploration is addressed.

### **Review of literature**

Etien et al. (2004) present an approach that combines visual and computational analysis to deal with large volumes of geospatial data. Their approach is based on the application of computational algorithms to extract patterns and relationships in the data. The visual representations are applied to portray the extracted patterns.

Lai et al. (2004) assess the effectiveness of dynamic symbols in cartographic communication. Their experiments confirmed that dynamic symbols tend to attract the attention of users and even in situations when foreground-background contrast were poor. On the question of how many flashing or animated symbols a display could sustain before the flickering became distracting, they conclude that 15 blinking symbols on a single display were acceptable.

Fabrikant et al. (2004) show how spatialization is used to generate an information display of non-spatial data. They point out that information spatialization is inspired by the intuition that spatial displays as maps, charts or diagrams can help to amplify cognition. Spatialization typically rely on dimension reduction techniques and layout algorithms to project relatedness in non-spatial data content onto distance, such that semantically similar documents are placed closer to one other than less similar ones in an information space. Their empirical study suggest that the distance-similarity metaphor applies to network spatializations by equating metric distance along network lines to similarity. They also find that line size, colour value and hue modify the distance-similarity metaphor in subtle yet logical ways.

Perceptual conflicts in the visual representation of networks can be solved by different cartographic generalization operators. Harrie (1999 and 2003) shows how the least squares method can be applied to move

map objects from its original position to a new in order to reduce the amount of visual conflicts. This method is certainly also applicable to the visualization of networks. Højholt (2000) presents a similar displacement algorithm based on the finite element method. In the case that the arcs of the network not are straight lines, but are sinuous, line smoothing techniques can be applied to increase their visual clarity. Different methods can be applied from the simple ones to more sophisticated techniques based on energy minimization (Burghardt 2005), for example. Bjørke and Isaksen (2005) show how information theory can be used to select an optimal number of arcs in a network. The methodology is demonstrated on road networks in Norway.

MacEachern et al. (2005) give a review of visualization of uncertainty in geographical information and identify key research challenges in the field considered. Slocum et al. (2003) developed a tool for a wall-size display to visualize global water balance together with the uncertainties associated with the model. They report that all the participants in their investigations were enamored with the wall-size display, but to prove its effectiveness compared with more traditional displays, needs more investigations. An interesting aspect of their investigation is the observation that decision makers are likely to feel uncomfortable with the notion of uncertainty. This fact implies that we should not merely present uncertainties, but also suggest how to deal with uncertainty.

Plewe (2003) demonstrates how various sources produce uncertainties that are different in their nature and in their appearance. Two types of uncertainty have been recognized, ambiguity and fuzziness (Klir and Wierman 1998, p.105). Ambiguity is an acknowledgement that an observation is only one of many possible measurable values for a phenomenon. In contrast, fuzziness is an inherent characteristic of the phenomenon being observed. Plewe (2003) points out that ambiguity and fuzziness need to be treated as separate for three reasons: (1) those using the data must be able to differentiate them, (2) both types may be present simultaneously, and (3) one type can be transformed into the other type.

## **Information theory applied to the evaluation of map complexity**

A perception study was carried out in order to find the optimal number of coloured depth intervals in seafloor maps. Thirty subjects, approximately, participated in the study. From the study we derived the conditional probabilities that the coloured intervals were misinterpreted. Then the difference between the map entropy and the equivocation was computed. The map with seven colours on the scale from light to dark blue did correspond to the channel capacity. One can ask to which degree this experiment can be generalized to network visualization. Is seven nodes optimal? Probably not in general, since this kind of generalization is too simplistic, but to claim that there is an optimal number of nodes, is sensible. This recognition leads to the emphasis of the concept of hyper-nodes.

Each hyper-node consists of several sub nodes. Say that the optimal number of nodes in a map is  $n$ . Then, we can imagine a zoom into the nodes so that for each node there are  $n$  sub nodes. This gives a zoom in and out of the nodes. This kind of zoom requires that the arcs of the network are grouped into hyper-arcs corresponding to the hyper-nodes. How to derive this structure can follow to principles: (1) the communication view and (2) the exploration view. In the first case the structure is predefined to the user, but in the latter the structure can be generated and manipulated by the map reader.

## **Conclusion**

An interesting application of information theory is in the computation of an optimum number of nodes in a display. A consequence of the optimum number principle, is the confirmation of the principle of reduction of map complexity. An application of this principle leads to the concept of hyper-nodes and hyper-links, i.e., zoom in and out of the network. Distinction between communication and exploration should be done. Visual and computational analysis should be integrated in the analysis of large volumes of nodes and arcs. Dynamic symbols represent a tool to highlight information. Moreover, the distance-similarity metaphor is an interesting concept used in the transformation of the network to the display space.

## References

- Bjørke, J. T. and E. Isaksen (2005). Map Generalization of Road Networks: Case study from Norwegian small scale maps. *Proceedings XXII International cartographic Conference*, A Coruña, Spain, 11-16 July 2005.
- Burghardt, D. (2005). Controlled line smoothing by snakes. *Geoinformatica* 9:3, 237-252.
- Etien, L. K. and M.-J. Kraak (2004). Alternative visualization of large geospatial datasets. *Cartographic Journal*, 41:3, 217-228.
- Fabrikant, S. I., D. R. Montello, M. Ruocco, and R. S. Middleton (2004). The distance-similarity metaphor in network-display spatializations. *Cartography and Geographic Information Science*, 31:4, 237-252.
- Harrie, L. (2003). Weight-setting and quality assessment in simultaneous graphic generalization. *The Cartographic Journal*, 40:3, 221-233.
- Harrie, L. (1999). The constrained method for solving spatial conflicts in cartographic generalization. *Cartography and Geographic Information Science*, 26, 55-69.
- Højholt, P. (2000). Solving space conflicts in map generalization: using a finite element method. *Cartography and Geographic Information Science*, 27, 65-73.
- Klir, G. J. and M. J. Wierman (1998). *Uncertainty-based information: elements of generalized information theory*. Heidelberg, Germany: Physica-Verlag.
- Lai, P.C. and A. Gar-On Yeh (2004). Assessing the effectiveness of dynamic symbols in cartographic communication. *Cartographic Journal*, 41:3, 229-244.
- MacEachren, A. M., A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler (2005). Visualizing geospatial information uncertainty: what we know and what we need to know. *Cartography and Geographic Information Science*, 32:3, 139-160.
- Plewe, B. S. (2003). Representing datum-level uncertainty in historical GIS. *Cartography and Geographic Information Systems*, 30:4, 319-334.
- Slocum, T. A., D. C. Cliburn, J. J. Feddema, and J. R. Miller (2003). Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30:4, 299-317.

## Litterature

- Graph glossary can be found at <http://www.utm.edu/departments/math/graph/glossary.html>  
or <http://www.math.fau.edu/locke/GRAPHTHE.HTM>
- References to graph theory books can be found at <http://www.math.fau.edu/locke/Graphstx.htm>
- Graph visualization tools can be found at <http://www.caida.org/tools/visualization/walrus/>  
<http://www.informatik.uni-bremen.de/uDrawGraph/en/>



## Research Methodology to Study the Cognitive Responses to Complex Network Visualization Designs

### Purpose:

An evolving problem in intelligence gathering and military situation and threat assessment is the challenge of accessing and ingesting the enormous amount of collected data. The problem is due in part to the rapid increase in sensing capability via national resources, networks of nano and micro-scale sensors (e.g., smart dust), unmanned aerial vehicles (UAVs), utilization of individual soldiers as data collectors and sensors, and utilization of web resources (e.g., open source information). The amount of collected data is increasing rapidly, with little regard to the actual ability to ingest or understand the data. Users are faced with a situation in which they are “drowning in a sea of data” but “thirsting for knowledge”. In addition, multiple analysts may be working on parts of a particular issue and may not have direct interaction or any interaction at all. A traditional view is that automated systems must be developed to process all of the incoming data and somehow abstract or represent the data for human consumption. This concept is illustrated in Figure 1.

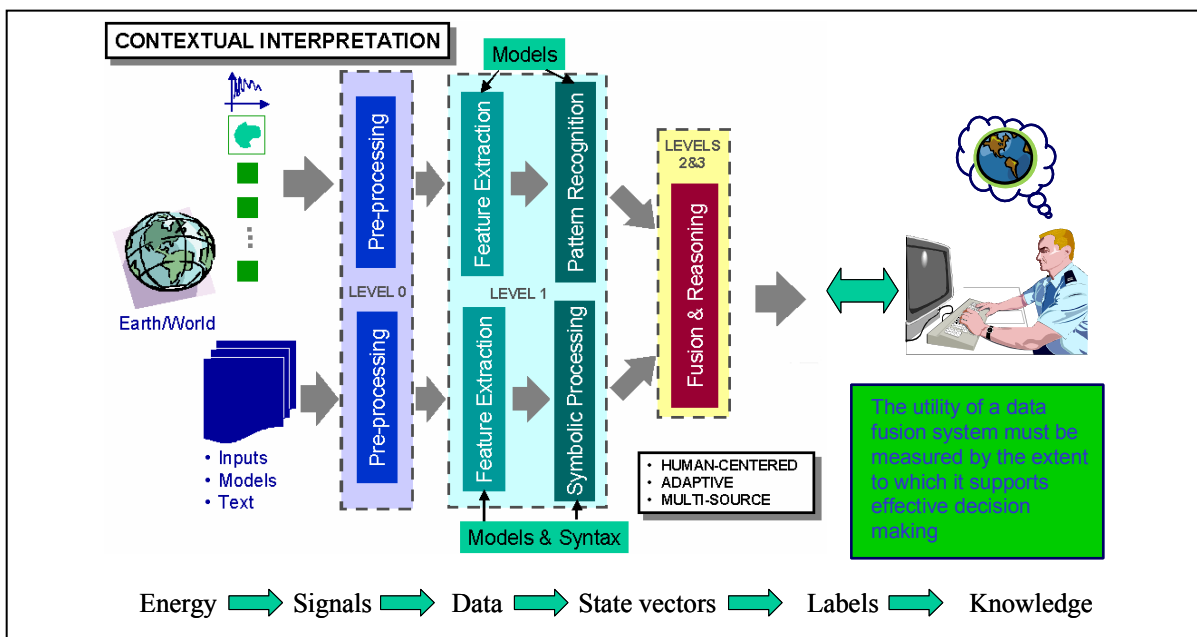


Figure 1: Traditional View of Processing Intelligence Data

In this view, data collected from multiple sources are processed via multiple levels of data fusion (using the Joint Directors of Laboratories model) to provide results to be viewed or analyzed by users. Processing techniques are drawn from statistical estimation; pattern recognition, automated reasoning, and optimization for performing Level-1 through Level-3 fusion (see [1] and [2]).

The interpretation of data by the analyst or decision-makers may be influenced by social and cognitive psychological factors. These factors include heuristics and biases in decision making at the individual level that are below the awareness level of the person, such as confirmation bias and the general inability of humans to make good estimates of probability. Additionally, color psychology suggests that color may have a temporary effect on mood, which suggests that the visualization techniques and tools themselves may effect how people interpret information presented. Finally, social psychological study indicates that people are influenced by the behavior of others around them indicating that a combination of personality and behavioral factors in a meeting-room setting in addition to the other factors described may have a significant impact on decision making and data interpretation.

### Problem Statement:

Network visualization is a rapidly evolving technique employed to assist in complex data analysis problems and as a decision-making tool. These new techniques, however, require an understanding how of how the human will interpret the data using the network visualization. From the analysts' perspective, questions regarding the impact of the visualization tool on the analysis process include:

1. How does the use of color impact the comprehension of the presented data?
2. How does the organizational pattern of the network links and nodes impact the analysis?
3. How does the mood of the analyst prior to and after exposure to network visualization impact interpretation of presented data?
4. How does color and organizational pattern of network links effect the analysts interpretation of presented data?

Further from the decision-makers' perspective, questions regarding the impact of the visualization tool on the decision process include:

1. How does the use of color impact the comprehension of the presented data?
2. How does the organizational pattern of the network links and nodes impact the final decision?
3. How does the mood of the individuals affect the decision-making process?
4. How does the behavior of the presenter influence the final decision and group process?
5. How does the behavior of others present in the meeting influence the final decision?

#### Proposed Research Study:

Phase I: Using a 2 X 3 factorial, within-subjects design, our research team will examine the effects of two classifications of colors (warm and cool) and three types of network organizational formats (hierarchical, radial, and random) on mood and threat assessment. Research participants will simulate an analyst perspective as they will be presented with information at a computer screen individually and asked to make assessments and decisions based on what is presented. Mood and threat assessment will be measured using a questionnaire, Likert scale format for continuous and more objective data.

Phase II: After interpreting and analyzing data gathered from Phase I, the decision-makers' perspective will be investigated using the same variables and design, color and network organizational format. In order to simulate the decision-makers' perspective, research participants will engage in a group meeting in which a PowerPoint presentation, a presenter (trained by the research team for consistency and neutrality), and group discussion. After this process, research participants will be measured on mood and threat assessment as in Phase I.

Phase III: In the final phase, our research team will investigate the influence of the presenter on interpretation of data. In a small group, meeting-room environment, research participants will be exposed to two versions of the same presentation, with the same presenter behaving in one of two ways: calm and reasonable or irritable and dramatic. Our research team will train an actor to play the two roles in order to create consistency and to control for variance. The use of deception and a confederate (a trained part of the research team who plays various roles) is common in social psychological study and helps to isolate social variables. After assessment of mood and threat level, research participants will be debriefed appropriately.

#### **References**

1. Hall, D. and S. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, 2<sup>nd</sup> edition, Artech House, Inc., 2004.
  2. Hall, D. and J. Llinas, editors, *Handbook of Multisensor Data Fusion*, CRC Press, 2001.
  3. Sternberg, Robert J., *Cognitive Psychology*, 2<sup>nd</sup> edition, Harcourt Brace College Publishers, 1999.
  4. Brehm, S. S., Kassin, S. M., Fein, S., *Social Psychology*, 4th ed. Houghton Mifflin Company, 1999.
- The work described in this abstract is cleared for presentation to NATO audiences at a NATO/PfP UNCLASSIFIED level.

- If work is sponsored by a Government Agency, identify the organization and attest that the organization is aware of submission.
- The abstract is technically correct.
- The abstract is NATO/PfP UNCLASSIFIED.
- The abstract does not violate any proprietary rights.

***Medical***



# **Epidemiologic Considerations in Network Modeling of Theoretical Disease Events**

**Marcus Lem, MD, MHSc, FRCP(C)**

Communicable Disease Control Division,  
Primary Health Care and Public Health Directorate,  
First Nations and Inuit Health Branch, Health Canada,  
A.L. 1915C, Tunney's Pasture,  
Ottawa, Ontario, Canada K1A 0K9

email: [marcus\\_lem@hc-sc.gc.ca](mailto:marcus_lem@hc-sc.gc.ca)

Social network modeling is a relatively new addition to the armament of public health and epidemiology. Network analysis has shown utility in the study of a range of communicable disease outbreaks affecting both health and commerce, including SARS, tuberculosis, syphilis and foot and mouth disease, and may have applications in automated disease surveillance systems.

Although network analysis of disease transmission has been used to analyze outbreaks retrospectively and as they progress, greater utility may be found in prospectively analyzing potential and theoretical disease events and developing intervention options a priori.

One of the greatest strengths of network models is the ability to visually represent complex and heterogeneous distributions of interactions in an intuitively comprehensible and interactive fashion. However, some epidemiologic characteristics related to the host, the disease agent, and the environment are less easy to represent, but should still be considered in theoretical network models. This is especially important when modeling new or emerging diseases with unknown or changing characteristics.

Considerations may be divided into characteristics of the host, the disease agent, and the environment:

- ◆ The host factor of susceptibility to infectious disease exists on a spectrum which may or may not have definable threshold. Disease susceptibility may be non-modifiable (e.g. genetic, co-morbidities), modifiable (e.g. by immunization or natural sub-clinical infection, lifestyle, such as smoking), or even intermittent in cases of waning immunity in the absence of wild-type disease challenge. Co-factors for disease, where a single infective or toxic agent or genetic predisposition is necessary, but not sufficient to cause disease may also exist.
- ◆ The pathogenicity of a disease-causing organism also may not remain constant either, independent of changes in host immunity. Evolutionary pressures favor increasing infectiousness and decreasing pathogenicity of an organism with successive passes through a population. "Successful" organisms tend not to kill their hosts. Conversely, genetic shifts and interspecies jumps, such as can occur with influenza, may lead to dramatically more virulent organisms.
- ◆ Environmental factors may profoundly affect the spread of a disease, regardless of whether the spread is person-to-person, food-borne, water-borne, or by fomites (contaminated objects). Diseases such as malaria or lyme disease are spread by arthropod vectors which have their own habitat and dispersion characteristics. Social network modeling of tuberculosis and sexually transmitted illnesses have shown that a geographic location, rather than a specific individual, may be the critical vertex between infected persons.

Combining network analysis with other public health tools such as traditional epidemiologic statistics and geographic information systems may lead to better characterization of disease events. Some possible implications for network analysis research are as follows:

1. Network analyses of the flow of humans, information, and commerce need to be performed before crises occur and the effects of vertex deletion assessed, with respect to unintentional/accidental breakdowns in infrastructure, criminal and terrorist acts, and interventions by public health authorities to control disease.
2. Network analyses of different aspects of the same issue may yield insight into more appropriate interventions. Thus analyses of both disease spread and utilization of health care or other services may identify optimal intervention points or locations.
3. Continued consideration should be given by the research community to interventions and investigations which mimic the behavior of the disease, especially in cases where network analysis cannot identify appropriate control points. Such interventions are only warranted when the danger of the disease is greater than the risk associated with the intervention, as is the case for oral polio vaccine in endemic countries.
4. Network analyses of disease events should consider the possibility of situations where vertices have different natures (eg. person, place, time), may appear, disappear or otherwise change (ie. changes in immunity or susceptibility, presence of disease co-factors), or have intermediaries (eg. vectors or fomites).
5. Innovative ways to visually represent aspects of disease epidemiology in network models need to be explored.

# ***APPLICATIONS OF NETWORK VISUALISATION IN INFECTIOUS DISEASE MANAGEMENT***

**Margaret Varga and Zack Jacobson**

QinetiQ, St. Andrew's Road, Malvern, WR14 3PS, U.K.

Health Canada, Jeanne Mance Building, 5th Floor

Ottawa, ON, K1A 0K9, Canada.

[m.varga@signal.qinetiq.com](mailto:m.varga@signal.qinetiq.com) and [zack\\_jacobson@hc-sc.gc.ca](mailto:zack_jacobson@hc-sc.gc.ca)

## ***ABSTRACT***

The SARS outbreak during the spring of 2003 showed that new cases can trigger rapid community and non-social cluster transmission of SARS-CoV. This created substantial health, social and economic consequences. Improved outbreak management requires that case-identification of such diseases must trigger an immediate public health response. Rapid detection and management of cases of diseases such as SARS and their contacts together with the prompt implementation of the appropriate control measures can arrest and contain the transmission and spread of the disease.

Given the likelihood and observed frequency of outbreaks of SARS and other infectious diseases, there is a need to be able to detect, manage and control the disease transmission as early as possible and thus minimise the impact of the outbreaks. Effective public health surveillance can act as an early warning system by detecting microbial, environmental, behavioural, occupational, and other health threats. It can also concentrate resources, focus interventions and facilitate future projections.

The issues and technological solutions for infectious disease management are applicable to Chemical, Biological, Radiological and Nuclear (CBRN) threat domains as well; these share many common factors.

This paper aims to discuss the use of and need for visualising the network of interconnected information for situation awareness to assist in taking effective control measures during infectious disease outbreaks.

## **INTRODUCTION**

The outbreak of SARS triggered the current, critical examination of information system. Questions arise regarding the quality, type and availability of critical data. The ability of organisations to generate and communicate information has also come under discussion. The level of preparedness and the ability of the health community to respond to unexpected large or even small events remain topics of concerns.

The relevant information to be managed for disease control can be said to form a Network of Information. That is a large number of different but related or connected items of information from a variety of sources and in a variety of locations or systems.

There is a need for a framework that is able to manage comprehensively and cohesively the network of disease information, provide facilities to analyse the available data and information and communicate the resultant information to relevant bodies so that effective control measures can be defined and implemented to control or contain transmission and spread.

To be fully effective this requires a single network of information that functionally and organisationally integrates all relevant partners across the country.

In summary, the goal of a disease surveillance programme is to be able to monitor and assess disease trends, i.e. provide **situation awareness**, and use this information to guide prevention and intervention programmes. In addition, to inform public health policy and policy makers and protect confidentiality as well as providing/presenting/**visualising the network of information** to those who need to know.

## **SITUATION AWARENESS**

Situation awareness is essential for effective decision making.

There are many definitions of situation awareness, different domains and communities define it differently. The following are some definitions:-

*“The understanding of the situation gained from the sum total of the relevant information provided to make a correct decision regarding the allocated objectives and/or the desired end state.” [1]*

*“Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.” [2]*

Situation awareness has been studied for many years in many domains. In the dangerous environment that is common to:

- Outbreak of infectious disease,
- Chemical, biological, radiological and nuclear threats;
- Emergency;
- Nuclear power;
- Aviation;
- Military and other domains;

there is a requirement to make decisions in high stress environments with significant uncertainty, time pressure and often with lives at stake [3].

In complex environments strong situation awareness can greatly improve the rate and the quality of human decision making.

The results of the situation assessment assist in developing a mental model which represents our situation awareness. This model is a representation of the environment and its expected behaviour. The model is used to provide the conceptual framework for describing, explaining and predicting future system states.

True situation awareness will be made possible by a diverse set of information and telecommunications technologies interoperating in a coordinated fashion. It can only be achieved through major advances in the national health/disease information infrastructure [3].

## **APPLICATION OF NETWORK VISUALISATION**

The above shows that a complex network of data is collected for various purposes and from which it is important that relevant information is extracted, analysed and collated with other information to assist in decision making. Often the information items are dynamic, diverse and can be tightly or loosely correlated, however, together they provide the essential support to make informed decisions, for example, to control the spread of the infectious disease or to halt any CBRN threats. This complex information network presents an information overload to the user/decision maker [4].

Formally, network data refers to items called nodes that have relationships (links) to an arbitrary number of other items/nodes. Nodes and links can have a variable number of attributes.

The attributes and relationships between items can be very complex. The internet, for instance, consists of tens of thousands of servers with numerous possible paths between them. Attempting to understand internet traffic

patterns, peaks and troughs of usage, and alternative paths between nodes by looking at tables and statistics is only possible for very limited tasks. Visualisation provides a means of addressing this problem.

The concept of network visualisation is to present (or visualize) the structure of a network and its associated data in an effective way so that vital information can be extracted, analysed, understood and applied in a timely manner. Thus the goal of network visualisation is to enable the user to gain insight into a structure that may consist of many data items.

There are many categories of networks, e.g. acyclic, lattice, rooted versus un-rooted, directed versus undirected. This categorisation enables researchers to develop algorithms to perform tasks on these structures such as finding the shortest or least cost paths connecting two items or traversing the entire network.

Large sets of data tend to have much of their information hidden. Finding a structure hierarchy among a set of data points is not easy. Much of the early work on visualization came from graph drawing. A graph with few points or vertices is easy to draw and to comprehend visually, but tools are needed to manage large data sets.

The most straightforward way to illustrate a network is a 2-dimensional or 3-dimensional diagram. Some visualisation techniques add animation, distortion and tightly coupled overview windows to reveal additional information about a network.

The effectiveness of the network visualisation usually depends on the size of the network, understanding of the information required from the network and the user's expectation. In the case of small networks it is possible to focus on detailed elements of the graph structure while in the case of large networks only gross topology can be captured. Visualisation of networks with thousands of nodes requires abstraction.

The aesthetics of the presentation of the topology features of the graph help to facilitate comprehension and usability. For example, edge crossing, running edges 'under' nodes that are not connected to each other and stacking nodes on top of each other all hinder comprehension and use of the graph [7]. Colour, size and shape are always useful ways to increase comprehension and add additional information on relational attributes.

There are a number of on-going and active network visualisation efforts and programmes (academic and commercial), many of which focus on the internet and related components.

Graphical network maps have also been used for many years in commercial network management products designed for enterprise networks. Network managers use these graphical interfaces to monitor an enterprise's computer network and access software functions to troubleshoot and fix network problems. However, network visualisations can also be used for grocery purchases, or the many interconnected components of software code.

## **SOCIAL NETWORK ANALYSIS AND INFECTIOUS DISEASE**

There are healthcare-related tasks that could benefit from the use of network visualisation; such as social network.

Social networks are used to study a variety of different social aspects, for example, effects on the spread and possible containment of multiple, interacting disease causing organism. The network can be used to model how social contacts might change in response to an outbreak or to intervention strategies. Many diseases spread through human populations between infective individuals and susceptible individuals. The pattern of these disease causing contacts (physical or otherwise) forms a network. This contact network can be used as a mathematical model for the spread of an infectious disease and provides a possible means of quantifying important spreading information such as the conditions under which a large outbreak is possible, probability of such outbreak and the size of the outbreak if it occurs.

A useful model for social networks is to represent social distance as physical distance. This representation allows users to get a spatial understanding of social relations; nodes with many relations in common are placed

close together in a representation (visualization). An intuitive impression of network structure then emerges from the proximities in the visualization image. For example, one can easily see racial segregation in network visualization as the clustering of races into distinct spaces on the display [8-15]. A social distance-based representation of network structure is facilitated when edge lengths are equated to relational strengths. There are various force-directed layout techniques which are usually successful at this.

## **NETWORK VISUALISATION TOOLS**

There are various commercial and free software tools available for the analysis of social networks. Some of the tools are targeted at visualization of networks only, while others contain analysis procedures only: there are also packages which integrate network analysis and visualization. The software available include: UCINET, Pajek, NetMiner, STRUCTURE, NetDraw, MultiNet, StOCNET, SNAFU, InFlow, NetVis, Visone, PGRAPH, etc.

## **CONCLUSIONS AND RECOMMENDATIONS**

The visualisation work reported in this paper show that there are many ways information visualisation can help people access information more effectively. Visualisation systems enable users to examine large amounts of complex data and quickly find the required information, navigate and interact with data more easily, recognise patterns and trends and to obtain a better understanding the information.

A wide variety of network visualisation tools have been developed specifically or can be adapted for health related tasks. However, none of them provide the capabilities that encompass all the requirements for managing the dynamic and unpredictable scenario of detecting, monitoring and managing the possible outbreak of infectious disease. Much work is still required to develop an infectious disease network visualisation tool that can manage the dynamic and unpredictable nature of any potential infectious disease outbreak or CBRN threats.

## **REFERENCES**

- [1] Situation Awareness, Defence Situation Awareness Seminar, DSTO Salisbury 1998., Peter Briggs.
- [2] Network Centric Warfare, second edition, September 1999. D. S. Alberts, J. J. Garstka, F. P., Stein.
- [3] Towards a Theory of Situation Awareness in Dynamic System, M. R. Endsley, Human Factors, 1995, 37(1), pp 32-64.
- [4] GeneVis: Visualization tools for genetic regulatory network dynamics, Baker, C.A.H., Carpendale, M.S.T., Prusinkiewicz, P., Surette, M.G.
- [5] Card, S.K. Robertson, G. G. & York W. The WebBook and the Web Forager: An information Workspace for the World-Wide Web, Proceedings of the ACM Human Factors in Computing Systems, pp, 111. 1996.
- [6] Plaisant, C. Facilitating Data Exploration: Dynamic Queries on Health Statistics Map, Proceedings of the Annual Meeting of American Statistical Association, 1993.
- [7] Sneiderman, B. Information Visualization Designing the User Interface: Strategies for Effective Human-Computer Interaction, 3rd Edition, pp 522-541/ Reading, M.A. Addison Wesley Longman, Inc. 1998.
- [8] Cummings, J.N. (2003). NetVis module -Dynamic visualization of social networks. Cambridge: Massachusetts Institute of Technology. Corander, J., Dahmstrom, P. 2003.
- [9] Brandes, U., and Wagner, D. visone. Analysis and visualization of social networks. In Junger, M., and Mutzel, P. (eds.), Graph Drawing Software, pages 321–340. New York: Springer. 2003
- [10] White, D.R., and Skyhors. PGRAPH: Representation and analytic program for kinship and marriage networks. Version 2.7 for Windows. Irvine: University of California. 1997

- [11] Harary, F., and White, D.R. P-Systems: a structural model for kinship studies. *Connections*. 24, 22–33. 2001
- [12] White, D.R., Batagelj, V., and Mrvar, A. (1999). Analyzing large kinship and marriage networks with PGRAPH and Pajek. *Social Science Computer Review*. 17, 245–274.
- [13] De Nooy, W., Mrvar, A., and Batagelj, V. *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press 2002
- [14] Freeman, L.C. Graphical techniques for exploring social network data. In Carrington, P.J., Scott, J., and Wasserman, s. (eds.), *Models and methods in social network analysis*. Cambridge: Cambridge University Press. 2004.
- [15] Davidson, R. & Harel, D., Drawing graphics nicely using simulated annealing, *ACM Transactions. Graph*, 15:301-331, 1996.



## Modeling Influenza Pandemic Response Effectiveness in Canada

Ben Houston  
Exocortex Technologies  
[ben@exocortex.org](mailto:ben@exocortex.org)

Zack Jacobson  
ARAD Health Canada  
[zack.jacobson@hc-sc.gc.ca](mailto:zack.jacobson@hc-sc.gc.ca)

NOTE: This abstract is written with the perspective that this project is relatively complete. While this is the plan for October 2006, currently development of the simulator is in its early stages.

### Introduction

As the risk of a global influenza pandemic increases there is growing response preparedness efforts within Canada. One question that governmental decision makers have in this context is what is the most effective distribution of anti-virals, such as oral oseltamivir, within the population of first responders, health care workers, administrators and the general public in addition to what extent should the anti-virals be used as prophylactics. To provide an answer to this question, we have developed a Canada-wide influenza pandemic simulator and visualization system that allows for the modeling of various patterns of anti-viral distribution and use.

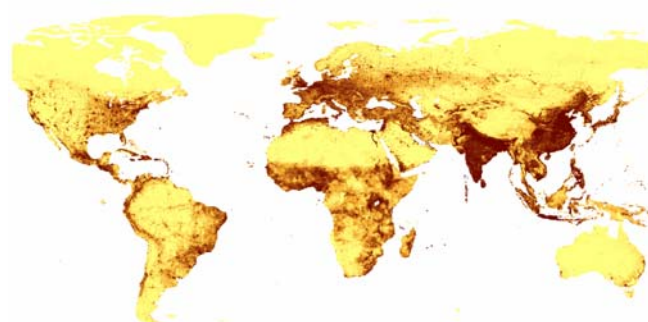
### Background

The seminal paper on modeling the effects anti-viral distribution and use in response to an influenza pandemic is that of [Ferguson *et. al.* 2005]. This paper focused on an indigenous outbreak of influenza from within Thailand and its eventual spread to neighboring countries within South East Asia.

We have presently presented a simple agent-based disease simulation model. This presentation represents the continued development of this on-going project.

### Adapting to Canada

Unlike South East Asia, it is very unlikely that pandemic influenza will spontaneously develop internally. entry into Canada of pandemic influenza is likely to occur at either Canada-US border crossings or at one of Canada's international airports. Thus in our simulation modeling we have spent considerable effort to these unique points of entry.



unlikely  
Rather, the  
influenza  
US border  
simulation  
modeling

Also, unlike South East Asia, Canada has a different socioeconomic distribution, broader domestic travel routines in addition to having a significantly more geographically distributed population with relatively few major urban centers. This presents another modeling challenge that was not faced by [Ferguson *et. al.* 2005].

## Visualization

The end result of our simulation work is massive amounts of population, disease and response pattern data which represent the dynamics of the disease spread over time. We have developed a geographically-based visualization system that displays the incremental spread of the disease in tandem with the dynamic response in anti-viral usage and distribution.

We will compare our visualization model with that of [Ferguson *et. al.* 2005].

## References

[Ferguson *et. al.* 2005] “Strategies for containing an emerging influenza pandemic in Southeast Asia” *Nature*, 2005. Available online at:

<http://www.nature.com/nature/journal/v437/n7056/full/nature04017.html>

## ***Security & Defence II***



# Practical Algorithms for Destabilizing Terrorist Networks

Nasrullah Memon

Henrik Legind Larsen

Software Intelligence Security Research Center  
Department of Software and Media Technology  
Aalborg University  
Niels Bohrs Vej 8  
DK- 6700, Esbjerg, Denmark

## Abstract

This paper uses centrality measures from complex networks to discuss how to destabilize terrorist networks. We propose new algorithms, implemented in newly developed investigative data mining tool, i.e. iMiner. These algorithms will provide assistance to law enforcement agencies, indicating when the capture of a specific terrorist will likely destabilize the terrorist network.

## 1 Introduction

In general, the network studied in this paper can be represented by an undirected and un-weighted graph  $G = (V, E)$ , where  $V$  is the set of vertices (or nodes) and  $E$  is the set of edges (or links). Each edge connects exactly one pair of vertices, and a vertex pair can be connected by (a maximum of) one edge, i.e., multi-connection is not allowed. A terrorist network consists of  $V$  set of actors (nodes) and  $E$  relations (ties or edges) between these actors. The nodes may be individuals, groups (terrorist cells), organizations, or terrorist camps. The ties may fall within a level of analysis (e.g. individual to individual ties) or may cross levels of analysis (individual-to-group ties).

## 2 Destabilizing Terrorist Networks

In this paper we propose three new practical approaches for destabilizing terrorist networks.

### 2.1 Using Centrality measures and constructing hierarchy

We propose to use centrality measures from SNA literature i.e. degree centrality and Eigen vector centrality. The algorithm1 is used to convert undirected graph to a directed graph. While algorithm2 we constructed hierarchy from the directed graph using Dependence Centrality discussed at section 2.5. Hierarchy shows who is in the power of whom, i.e. who is central or peripheral in the network. We applied the algorithms on the network of alleged 9-11 hijackers and their affiliates [1] and results of the Algorithms are depicted in figure 1.

### 2.2 The efficiency of a network

The network efficiency  $E$  is a measure to quantify how efficiently the nodes of the network exchange information [2]. To define efficiency of  $G$  first we calculate the shortest path lengths  $\{d_{ij}\}$  between two generic points  $i$  and  $j$ . Let us now suppose that every vertex sends information along the network, through its edges. The efficiency  $\epsilon_{ij}$  in the communication between vertex  $i$  and  $j$  is inversely proportional to the shortest distance:  $\epsilon_{ij} = 1/d_{ij} \forall i, j$  when there is no path in the graph between  $i$ , and  $j$ , we get  $d_{ij} = +\infty$  and consistently  $\epsilon_{ij} = 0$ .  $N$  is known as the size of the network or the numbers of nodes in the graph. Consequently the average efficiency of the graph of  $G$  can be defined as (Latora V., 2004):

$$E(G) = \frac{\sum_{i \neq j \in G} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \quad (1)$$

The above formula gives a value of  $E$  that can vary in the range  $[0, \infty]$ , while it be more practical to normalize  $E$  in the interval of  $[0, 1]$ .

### 2.3 The Critical Components of a network

Latora V. et al recently proposed a method to determine network critical components based on the efficiency of the network briefly discussed in the previous subsection. This method focuses on the determination of the critical nodes. The general theory and all the details can be found in Ref. [2].

The main idea is to use as a measure of the centrality of a node  $i$  the drop in the network efficiency caused by deactivation of the node. The importance  $I(\text{node}_i)$  of the  $i$ th node of the graph  $G$  is therefore:

$$I(\text{node}_i) \equiv \Delta E = E(G) - E(G - \text{node}_i), \quad i = 1, \dots, N, \quad (2)$$

Where  $G - \text{node}_i$  indicates the network obtained by deactivating  $\text{node}_i$  in the graph  $G$ . The most important nodes, i.e. the critical nodes are the ones causing the highest  $\Delta E$ .

### 2.4 Position Role Index (PRI)

The position role index (PRI) is the other proposed measure which highlights a clear distinction between followers and gatekeepers. It depends on the basic definition of efficiency as discussed in equation (1).

It is crystal clear fact that efficiency of a network in presence of followers is low as compared to their absence in the network. This is because they are usually less connected nodes and their presence increases the number of low connected nodes in a network, thus decreasing its efficiency. To find  $E(G - \text{node}_i)$ , whatever the nodes be deactivated, we take the number of active nodes as the size of the graph. So we use  $M$  (number of active nodes) in the denominator of formula (1) instead of  $N$  (total number of nodes). If we plot the values on the graph, the nodes which are plotted below x-axis are followers, whereas the nodes higher than remaining nodes with higher values on positive y axis are the gatekeepers. While the nodes which are on the x-axis usually central nodes, which can easily bear the loss of any node. The leaders tend to hide on x-axis there.

### 2.5 Dependence Centrality (DC)

The dependence centrality of a node  $D_{mn}$  shows how much node  $m$  is dependent on node  $n$ . In simple words, we can say that how much node  $m$  is useful to node  $n$  in order to communicate with other nodes of graph.. Mathematically it can be written as:

$$DC_{mn} = \sum_{p \in G} \frac{d_{mp}}{N_p} + \Omega \quad (3)$$

Where  $m$  is the root node which depends on  $n$  and  $N_p$  is the number of geodesic paths coming from  $m$  to  $p$  through  $n$ , and  $d_{mn}$  is geodesic distance from  $m$  to  $n$ . The  $\Omega$  is taken 1 if graph is connected and 0 in case it is disconnected. In this paper we take  $\Omega$  as 1, because we consider that graph is connected (at this initial stage of our work we do not consider disconnected graphs)

## 3. Constructing Hierarchy from Non-hierarchical Terrorist Networks

Using new introduced measure and newly introduced algorithms (will discuss in detail in the paper) we constructed the hierarchy shown in figure 6, of the hijackers involved in 9/11 terrorist attack and their affiliates (from the publicly available dataset as shown in figure 1.) The hierarchy clearly suggests that *Muhammad Atta* (33) was the key leader of the plot. While *Marvan al shehhi* (40) was assisting him as he is below in the hierarchy. They both were suggested as potential leaders in 9/11 attack and led their respective groups. They were also both members of Hamburg Cell. *Fayez Ahmed* (31) and *Mohand Al Shehri* (42), who were in the same hijacked plane with *Marvan Al Shehhi* are below *Marvan Al Shehhi* in same level of hierarchy. *Abdul Aziz alomari* (39), *Wail Al Shehri* (38) and *Wail Al shehri* (32) are in 3rd level in the hierarchy.

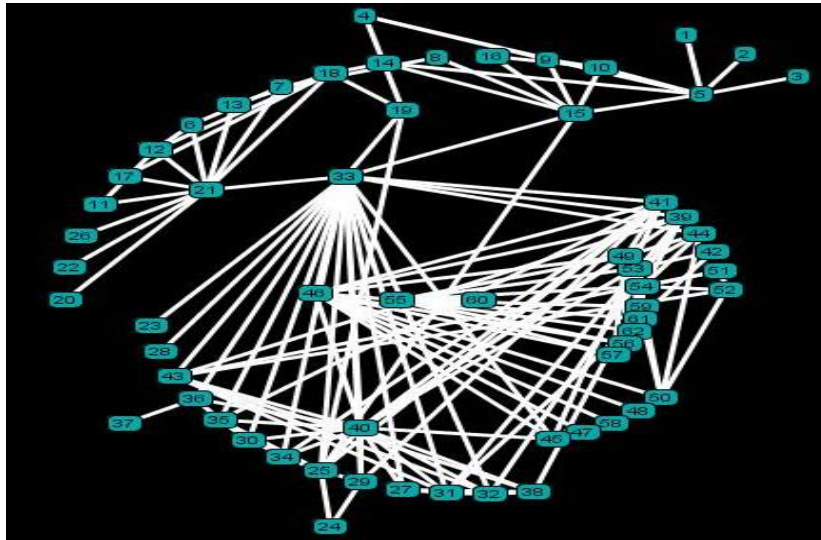


Figure 1. 9/11 hijackers and their affiliates

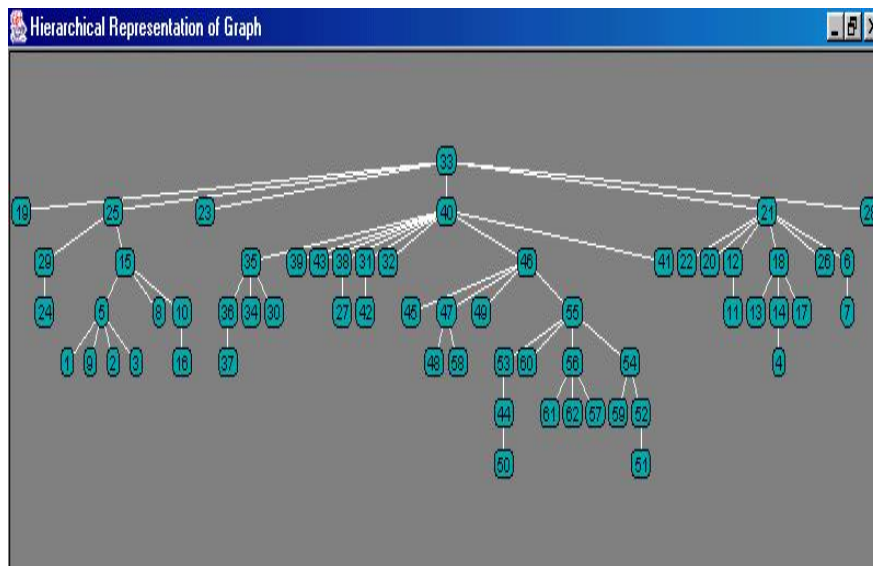


Figure 2. Hierarchy discovered by *iMiner* from graph shown in Figure 1.

The intelligence agencies can easily detect who are potential *leaders/gatekeepers* and even peripheries by using these new algorithms. The dependence centralities of the hijackers and their affiliates as shown in figure 1 can be seen at [3].

## Conclusion

In this paper we presented practical algorithms for destabilizing terrorist networks. The algorithms can easily find leaders/gatekeepers, dependency of the nodes on each other, important nodes in a network, in order to assist law enforcement in destabilizing the network.

## References

- [1] Krebs, V. (2002) Mapping Terrorist Networks, *Connections* 24(3).
- [2] Latora V. et al. 2004. How the science of complex networks can help developing strategies against terrorism. *Chaos, Solutions and Fractals*. 20. 69-75.
- [3] Dependence centralities of the 9/11 hijackers and their affiliates.  
[http://cs.aue.aau.dk/~nasrullah/DC\\_9\\_11.h](http://cs.aue.aau.dk/~nasrullah/DC_9_11.h)



# Information Fusion and Visualisation in Anti Asymmetric Warfare

Felix Opitz, Kaeye Daestner, Thomas Kausch  
Defence and Communications Systems /  
ADBM 64 EADS Deutschland GmbH  
Wörthstr. 85, 89077 Ulm, Germany  
[felix.opitz@eads.com](mailto:felix.opitz@eads.com)

Thilo Trapp  
Defence and Communications Systems /  
ADBM 61 EADS Deutschland GmbH  
Wörthstr. 85, 89077 Ulm, Germany  
[thilo.trapp@eads.com](mailto:thilo.trapp@eads.com)

*Abstract – Information Fusion is a key factor for ensuring information superiority in various military and civil surveillance system. These may be found in combat management, ground based air defence or in coastal surveillance systems addressing the more civil area. A new aspect in all this applications is the issue of new types of threats, e. g. anti asymmetric warfare and low intensity conflicts. Considering these new types of asymmetric threats in the information fusion process leads to new requirements and very new concepts for design, implementation and integration of the information fusion process. This paper presents some of the aspects and concepts related to the information fusion and its visualisation within coastal surveillance, combat management and air defence systems realizing anti asymmetric warfare aspects.*

*Keywords: Information fusion, visualisation, maritime domain awareness, anti asymmetric warfare, coastal surveillance, combat management, air defence system.*

## 1 Extended Abstract

In modern applications of information fusion systems new requirements related to the so called asymmetric warfare or low intensity conflicts have to be addressed. Examples of such applications may be found in the civil and military area: coastal surveillance and defence systems, the surveillance in the urban environment, the newest generation of combat management systems of navy ships, or future air defence systems. Their operational frame may be e.g. counter-terrorism, peace keeping and peace support operations.

The classical situation of symmetric conflicts was determined by uniformed forces. The utilisation of weapons was assumed to comply with international treaties and conventions and with national laws. All targets should comply with the principles of the laws of armed conflict: military necessity, discrimination, proportionality, and minimization of unnecessary suffering. The tactics uses numbers, mobility, weapons/ systems effects, etc. to bring the right capabilities to bear to achieve desired outcomes.

Different from this classical scenario is the asymmetric one. Here, the objectivity of an adversary is to create instability using irregular forces. These may include prohibited weapons, improvised devices, the use of civilian facilities and equipment as weapons, or the use of legitimate weapons in an unlawful way. Also, civilian and protected targets (both inside the conflict area and elsewhere) may be attacked by the adversary if such actions serve his objectives. So one may think about the utilisation of low cost platforms against expensive own platforms. These may be sports/micro light plane, hang glider, airliner, diver, mine, UAV/UUV, dinghies, car/floating bomb, bazooka, grenade, artillery fire sniper instead of the potential threats carry out by high tech platforms in the symmetric warfare. The tactics may take advantage of the choose of times, places, and targets. It is lead by the willingness to use irregular forces and unconventional weapons to advantage.

However even in this situations of asymmetric threats one is restricted to legal responses, which takes care of

- The principle of military necessity: It justifies those measures not forbidden by international law, which are indispensable for securing the complete submission of the enemy as soon as possible.
- The prohibition of superfluous injury or unnecessary suffering.
- The principle of distinction: The parties to a conflict must at all times distinguish between combatants and non-combatants, and between military and non-military targets. Neither the civil population as a whole nor civilians in particular may be subjected to attacks.
- The principle of proportionality: It is forbidden to attack a military target if such an action would cause excessive damage to civilians and civilian objects.

The answer to this threat scenarios is different. It consists both on operational concepts and technical solutions.

The search for operational concepts lead already to different national and multi-national tactics, e.g. [1], [2], [7].

Technical solutions addresses the issues of information superiority and intervention capability by realising situation awareness and decision support even in stress situations of the operator. Further one has to take into account the uncertainty and reliability of the information in the anti asymmetric warfare area. These are covered by:

- An adapted and integrated information fusion
- Adapted Visualisation and HMI Concepts
- Specialized sensor suite
- Specialized effectors suite

To describe these points the JDL model is very useful (fig. 1). Level 0 fusions or sub “object assessment” is related to the sensor hardware environment. It deals with extraction and closely hardware related signal processing. Therefore it is performed within the sensor itself. Level 1 considers the object assessment The overall aim of this process level is to find a unique representation of all objects in the environment considered. Therefore, the real objects within the surveillance area are described by so called tracks, which are built by data association and state estimation techniques. The 2nd and 3rd level deals with situation and impact assessment. Within these parts the relationships between several objects are studied. This may be of different types, e.g. spatial-temporal, organizational, causal, and similarity. The process refinement in level 4 tries to optimise the ongoing fusion process itself. This may be executed e.g. by adaptive data acquisition and processing. Through a 5th level called “cognitive refinement” and a human machine interface the result of data fusion is presented to the operator.

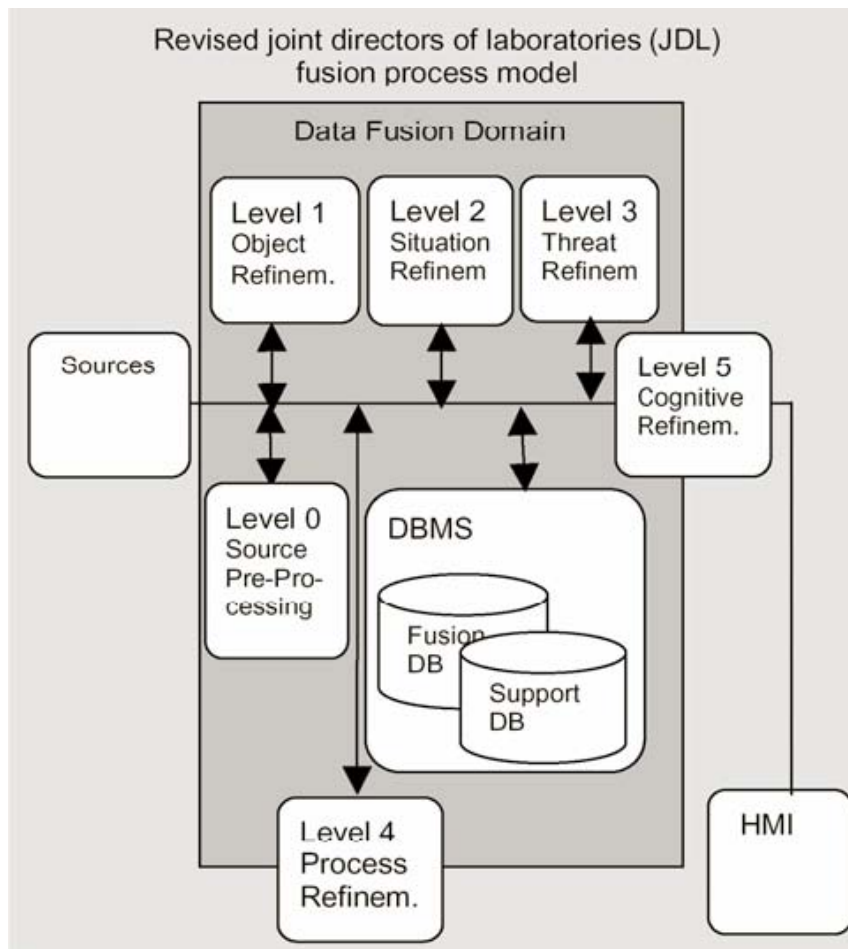


Figure 1 : The JDL model.

The paper presents possible realisations of these fusion levels most applicable to the anti asymmetric warfare framework presented above: The applicable sources of information are analysed and evaluated. The specifics of the fusion levels and the benefit of their interaction in integrated multi sensor systems are discussed. Finally possible concepts of visualisations suitable for anti asymmetric warfare are demonstrated

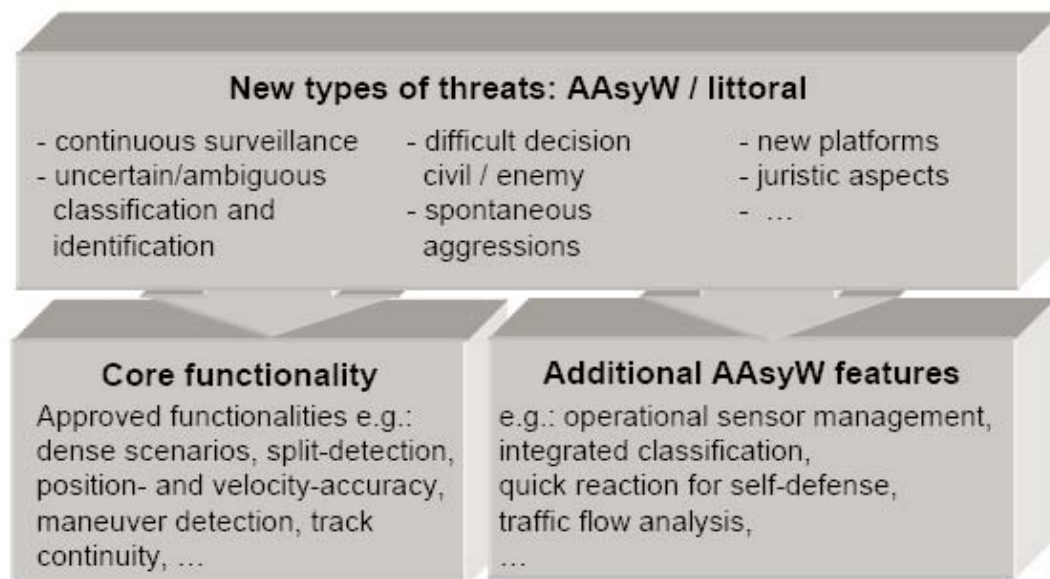


Figure 2 : Level 1 Fusion and AASyW.

## References

- [1] *Non-Lethal Weapons and Future Peace Enforcement Operations*, RTO TECHNICAL REPORT TR-SAS-040.
- [2] Mark A. Gammon, *Maritime Concept Development Approach for a Force Protection In-Harbour Transit of a Canadian Patrol Frigate*, Defence R&D Canada – Atlantic Technical Memorandum, TM 2005-148, September 2005.
- [3] Felix Opitz, Thomas Kausch, Werner Henrich, *Data Fusion Development Concepts within complex Surveillance Systems*, Proc. 7th Int. Conf. on Information Fusion, Stockholm, Sweden, 2004.
- [4] Bedworth, M. and O'Brien, J., *The Omnibus model: a new model of data fusion?*, Proc. 2nd Int. Conf. on Information Fusion, 1999.
- [5] Alan N. Steinberg, Christopher L. Bowman, Revisions to the JDL Data Fusion Model, in David L. Hall and J. Llinas: *Handbook of Multisensor Data Fusion*, CRC Press, Boca Raton, 2001.
- [6] F. Opitz and G. Kouemou: *An Unified Approach to Radar System Simulation & Development*, Proc. International Radar Symposium IRS 2005, Berlin, Germany, 06-08 September 2005
- [7] F. Opitz and G. Kouemou: *An Unified Approach to Radar System Simulation & Development*, Proc. International Radar Symposium IRS 2005, Berlin, Germany, 06-08 September 2005
- [8] D. S. Alberts, J. J. Garstka, R. E. Hayes and D. A. Signori: *Understanding Information Age Warfare*, CCRP, 2001.
- [9] M. G. Vickers and R. C. Martinage: *The Revolution in War*, CSBA, 2004.
- [10] K. F. McKenzie: *The Revange of the Melians: Asymmetric Threats and the Next QDR*: Institute for National Strategic Studies, 2000.



# RUNTIME SIMULATION FOR POST-DISASTER DATA FUSION VISUALIZATION

Kesavadas T.\*

kesh@eng.buffalo.edu

Virtual Reality Laboratory and Center for Multisource Information Fusion (CMIF)

The State University of New York at Buffalo

Buffalo, NY 14260, USA

**Abstract** - With regard to the threats of recent natural and man-made disasters, it is important for a control center to be aware of the situation and be able to assess the threat. However, simulating a large amount of post-disaster fused data is a complicated task, and its visualization is even more difficult to achieve with the paradigm of common geo-referencing systems. We have developed a post-disaster monitoring interface that runs in a fusion-based simulation with High Level Architecture/Run Time Infrastructure (HLA/RTI). In our visualization system, damage and recovering activities are presented in a fast GIS vector map with convenient data and display manipulation. All data that comes from the data fusion federates is displayed at run-time and stored for further analysis. In addition, the pattern of time-aggregated data has enabled dynamic visualization, which includes the morphing of the casualty clusters. This feature provides an effective way to keep track of a region so that a user can easily be aware of the emerging trends. A unique approach to multiple views by the integration of 2D and 3D displays of the fused data is also described.

**Keywords** - Disaster, emergency response, situation awareness, fusion, integrated architecture, network visualization.

## 1. Introduction

The time immediately following a natural or man-made disaster can be a chaotic experience to any individual or community. This is evident with regards to the natural and man-made disasters, which have occurred, in recent years. A prior study has shown that in an earthquake situation, the information collected and dispersed in the first 72 hours is the most crucial, since most people still severely injured after this time are not likely to survive [1]. When a disaster spreads over an area, and causes thousands of casualties in a short time, it is nearly impossible to manage the disaster by human observation alone because of the massive amount of incoming information. In fact, for large-scale disaster management, the first and most imperative step is the awareness of the situation in order to optimize the allocation of available resources. Therefore, the situation awareness is an essential role of a disaster monitoring or visualization system. With the paradigm of conventional geo-referenced display, this is difficult to achieve because its implementation is limited to the positioning of the corresponding graphic images. This usually results in thousands of scattered and cluttered icons on the display. The present research provides a monitoring environment through efficient data management and a user-friendly graphics interface that deals with the massive influx of data from the data fusion process. Furthermore, our technology adds time-aggregated data management that contributes to the visualization of more abstract and comprehensible graphics. This approach encourages tactical thinking and strategic control for a severely attacked area [2].

### 1.1 Previous Works

Recently, researchers have been involved in data fusion for dealing with complex natural phenomena or algorithmic problems. A few visualization projects exist involving information fusion, such as a weather visualization application for emergency planning [3], NASA's wind tunnel simulation [4], and seismic activity visualization by the University of California at Irvine [5]. A research team from the University at Buffalo has also been working to create a battlefield visualization scenario using fused data [6].

---

\* Author for communication

Without fused data, there have been studies on emergency response, such as decision making aids by interaction through voice/gesture recognition completed by Pennsylvania State University [7], a framework for incorporating the many emergency response models for a simulation by Jain and McLean [8], and satellite/airborne image or video processing for the Kocaeli earthquake in Turkey by Ozisik and Kerle [1]. The works mentioned above implemented data processing and mapping in a two or three dimensional space, but do not provide abstract information from which a user could comprehend a situation. The work of Kim and Kesavadas considered effective icon/symbol generation regarding the viewer's visual recognition, which has been a cognitive issue in the military community since the advent of the digital display in the 1960's [9]. They have suggested *Automated Dynamic Symbology* by parameterization of graphic components connected to fused data. Their methodology was considered as an appropriate feature for visualization of strategic information and remote-networked implementation, and served as a basis in the current research.

## **1.2 Issues and Our Approach**

The present work is an achievement of a large scale fusion-based post-disaster simulation project. For simulation test, it specifically takes the post-earthquake situation data which occurred in Northridge, California on January 19, 1994. This was considered to be one of the worst earthquakes in the Los Angeles area in recent memory [10]. Our simulation model relies on the output from HAZUS [11] developed by the Federal Emergency Management Agency (FEMA). The fused data is currently being produced by a multidisciplinary group at our institution [12, 13]. The fusion output includes data of low-level fusion (identification) which covers roadway damage, casualties, hospitals, and ambulance routing and police information. As mentioned earlier, this approach usually causes information overflow to a viewer, such as different types of icons cluttering and overlapping on top of each other. Therefore, our system includes high-level fusion data (situation awareness and threat assessment), such as casualty clusters and its trends and prediction.

A challenge for the visualization of a post-disaster simulation is to deal with the substantial and complex data interface so a user can manipulate and retrieve desired information. Unlike the previous works, our application provides a visual display of emergency response data at run-time and through a networked simulation environment. The advantage is that it displays information as it is received and without delay. Commercially available Geographic Information System (GIS) software has been used for the data construction of urban terrain and traffic network. However, they are only capable of low-level fusion output, which is the functionality of location and identification. They do not have the functionalities for high-level fusion that demands many complicated tasks, such as large data-set manipulation, dealing with time aggregated data, and the capability of putting depth and height cues to the display.

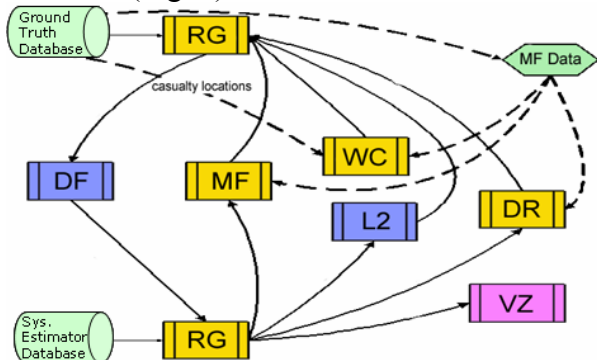
To deal with such enormous and complicated data, we have developed the common class interface for sharing and synchronization of 2D and 3D graphics. It allows a user to see what resources are available, and where casualties are located in 2D, as well as 3D, which give a better understanding of the spatial relation amongst the resource objects. Each visual mode adopts different application programming interfaces (API) and rendering environment, Windows MFC and OpenGL for example. Our work achieved implementation of both in one application combined in sync, which is described in the following sections.

## **2. Run-time Federation Interface**

In this project, information communication is implemented with the HLA/RTI composed of several federates. A federate simply being one piece of the RTI which carries out a specific task, such as information regarding walk-in casualty or medical facility. The RTI allows for common variables to be changed by one federate and then updated in another based on the concept of "publishing" and "subscribing" to variables [14]. The following section explains the interface of the post-disaster simulation and describes how the data is passed to the visualization system.

## 2.1 The Federation Interface of HLA/RTI

The importance of HLA/RTI in the current simulation is that all simulation data is being generated at run-time. It should also be noted that for each federate it is crucial to know the current state of the situation at all the times. This is achieved through the report flow, or exchange interface, between federates (Fig. 1).

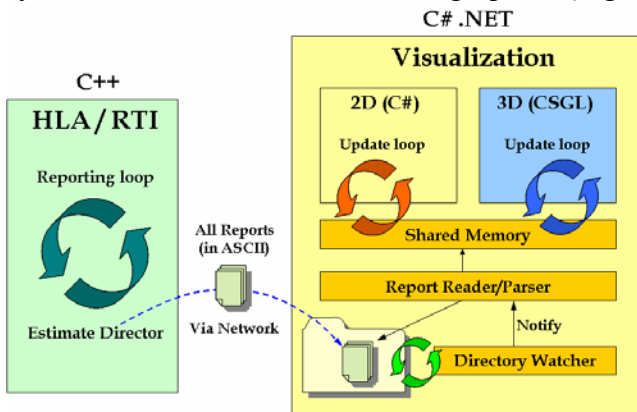


**Figure 1.** Post-earthquake simulation data flow in the HLA/RTI network.

The Report Generator federate (RG) generates all reports for the simulation based on the output from the HAZUS earthquake model [10]. Data Fusion (DF, also called Level-1 fusion), then decides which reports are not repeated and fuses them into one report. This information provides the core to the simulation which the rest of federates use to carry out their own tasks. Level-2 fusion (L2) determines the time-stamped formation of casualty clusters from the casualties reported by RG. The figure shows the interaction between other federates, such as Walk-in Casualty (WC), Medical Facility (MF), Dispatcher and Router (DR), and Visualization (VZ).

## 2.2 Data Interface of the Visualization Federate

Unlike the data relay in other federates, the visualization federate currently only takes information from the rest of the federation via Report Generator/Estimate Director. To create a highly abstract and robust runtime performance, we adopted the C# (C sharp) programming language on the Microsoft .NET platform [15]. Because the HLA/RTI is designed to support only C++ objects, it was critical to come up with a technique in order to integrate the two different programming codes. A solution was found for bridging the two executions not on the programming level, but the OS level. A system was devised for a directory to be setup where the C++ process stores all the report messages in ASCII format which can then be read by the C# process. Running in a call back loop, the directory watcher notifies the parser immediately after it gets the reports from the estimate director federate. The stored ASCII text is then parsed and stored in a shared memory for synchronization of the 2D and 3D graphics (Fig. 2).



**Figure 2.** Data interface of the visualization federate between the HLA/RTI architecture.

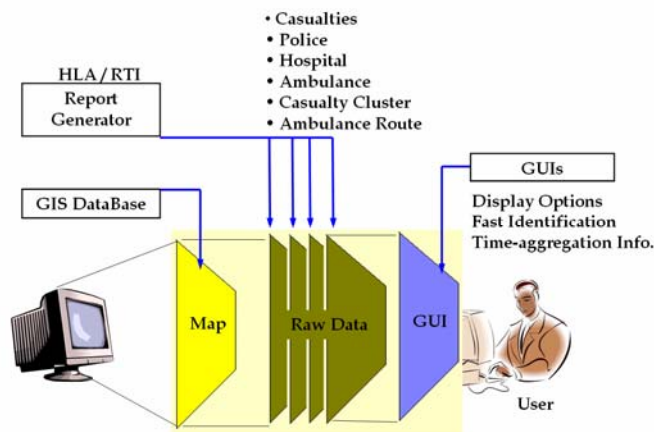
### 3. Modeling of Visualization framework

The following section addresses how we achieved efficient data manipulation of a large data set and integration of two different kinds of graphics (2D and 3D) to provide sufficient options to a viewer.

#### 3.1 Data Layers: The Visualization Pipeline

Layering is a useful way to organize massive GIS data. The United States Geological Survey (USGS) [16] offers an accurate depiction of the Northridge area which was used for our vector map. In addition to map viewing capabilities, a monitoring capability has also been developed. All the GIS data is layered at the bottom of the pipeline and rendered first. Built on top of this information is the federate data from the HLA/RTI. Placed on top of the federate information is the graphical user interface (GUI) developed for the simulation (Fig. 3).

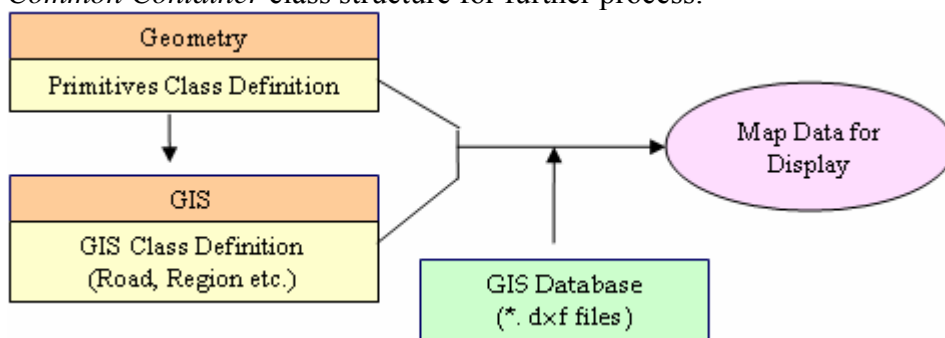
Data layering and implementation not only provided an easy-to-debug environment to programmers, but also produced better rendering performance.



**Figure 3.** Data layers of the software architecture  
 3.1.1 *Map Layer: Generation of Fast Vector Map*

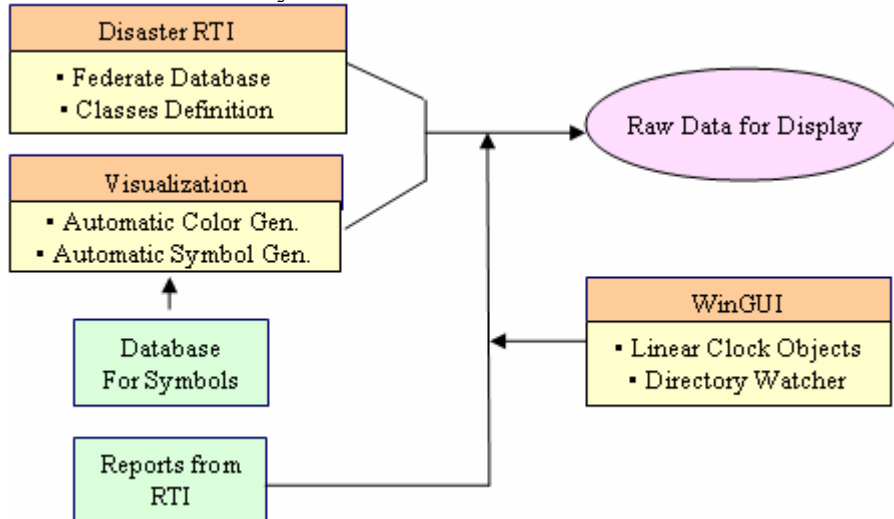
Two C# namespaces were created to handle the generation of the GIS map database for the display (Fig. 4). First, the *Geometry* namespace deals with the definition of all geometrical entities in a hierarchical structure. Another namespace, *GIS*, stores structured information of GIS objects for display. It uses a primitive class definition from the *Geometry* namespace with real GIS data. While all other GIS data was used from \*.dxf files, a detailed road network is generated from a Tele-Atlas database file [17] which is helpful in determining ambulance routes and retrieving street information. Each of the links is represented by a 256 character line containing information like Link ID, Length, Street Name etc., which is further divided in a database of start and end points. Each end point is known as a node, which is stored with a unique ID and Universal Transverse Mercator (UTM) geographic coordinate in the database.

A *road link* class then stores all the information in runtime memory. Finally this map is stored in the *Common Container* class structure for further process.



**Figure 4.** Pipelines to Map Data Generation.  
 3.1.2 *Raw Data Layer: Data Extraction from Incoming Reports*

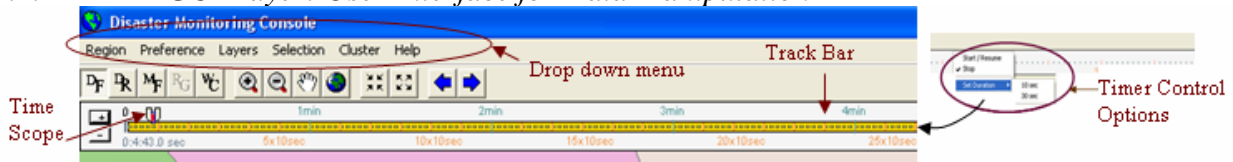
The Disaster-RTI defines all types of federate class definitions including casualties, police, ambulance, medical facility, roadway damage, ambulance route etc., in terms of their properties and functions (Fig. 5). Dynamic arrays have been built on top of this information to store their objects and reports in a hierarchical order. A separate class called *Visualization* is then defined for automatic generation of different color data which is used for each zip-code region in the Northridge area. This class is also used for identifying proper symbols for each federates' database which is stored in a symbol database.



**Figure 5.** Pipelines to Report Data Generation

The directory watcher class and its functions are defined in the *WinGUI* namespace which keeps a track of new reports in the directory folder. For every report from the RTI, raw data is generated for display by combining these files and having the data stored in the *Common Container* class for further processing.

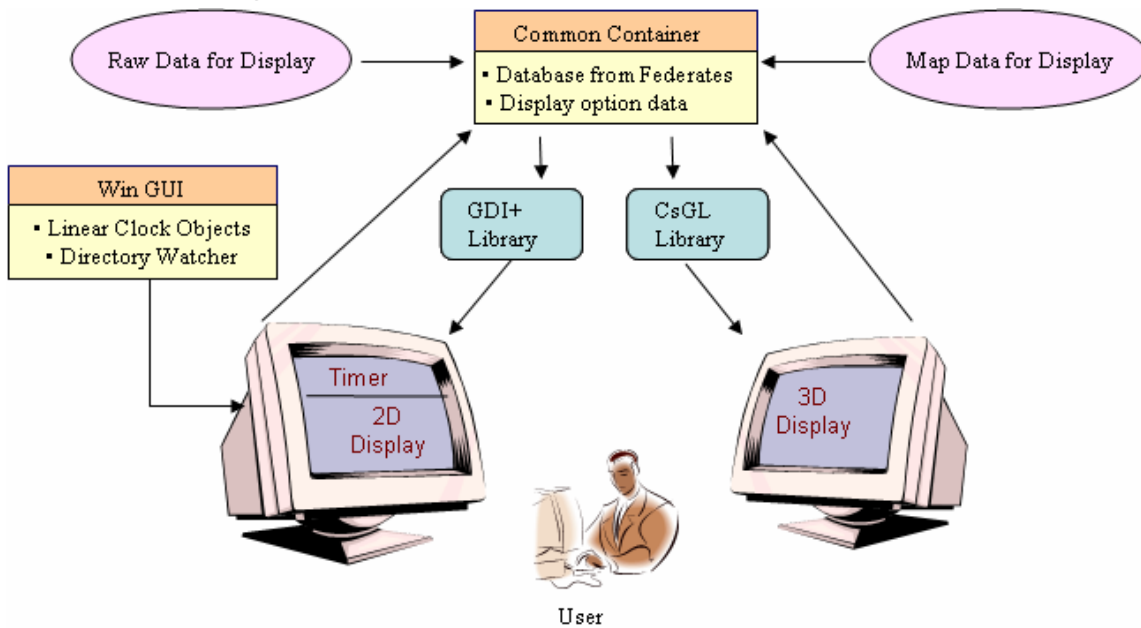
### 3.1.2 GUI Layer: User Interface for Data Manipulation



**Figure 6.** Linear clock and track bar for the manipulation of time-aggregated data.

What is controlled by the GUI is directly related to what is being captured in the federate layer underneath. The GUI includes a menu and tool bar control system with the capabilities of mouse and keyboard interactions. Since the simulation has been designed to run over a period of time, it is important to store accumulated federate data for further usage. The track bar shown in the GUI pane gives the user the ability to go back in time (Fig. 6). This way a person can see what progress is being made in specific areas over a period of time. It also has an option for time scope expansion, which allows for a longer duration of time to be rendered at once [18].

### 3.2 Integrated Simulation Architecture



**Figure 7.** Integrated simulation architecture for multiple and different display of 2D and 3D. We took advantage of multiple windows of two different graphics modes: two dimensional Windows MFC-based graphics, and three dimensional OpenGL-based graphics. For the synchronization in these multiple windows, display variables in either the 2D or 3D window need to be updated in the other window. Since it is impossible to swap references on two files due to restriction on circular dependency in C#, we developed a common class to access the shared memory and update the value (Fig. 7). All the corresponding objects for the federate report database (raw data) and map data are stored in the *Common Container* which can be passed to the 2D and 3D windows for display as needed. Using the Graphical Device Interface Plus (GDI+) library for creating graphics in C#, all 2D graphics could be created in an effective manner [19]. The 3D display objects and navigation controls were then defined in the *3D display* class using the CsGL graphics library, which is simply an OpenGL wrapper for the C# programming language.

## 4. Run-Time Fusion Data Visualization

The following sections report on the user interface and visual displays for low level fusion functionality (position and identification) and high level fusion functionality (situation awareness).

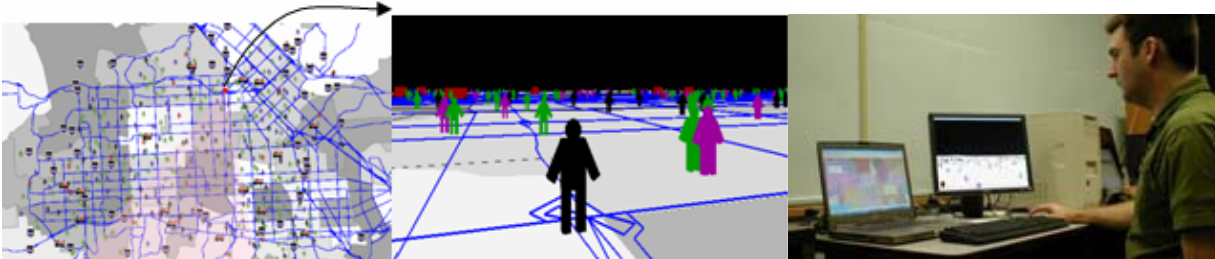
### 4.1 Multiple Display: 2D Fast Vector Map & 3D Visualization

The present research includes graphics controls and data manipulation found in common geo-referencing system. The zip code areas, road data and even the colors used, can be altered as needed by the user. This allows for the simulation to be viewed in a way that is least visually distracting so more attention can be focused on the resources (Fig. 8).



**Figure 8.** Manipulation of map environment: color (left), grayscale (middle), and single color.

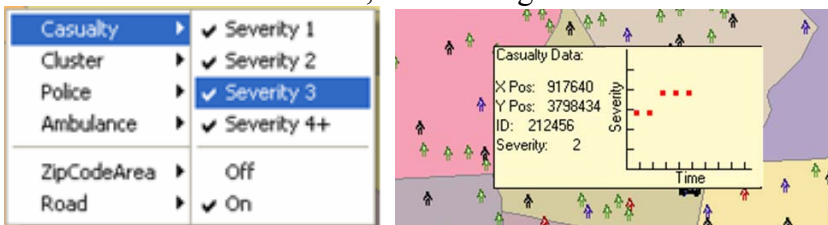
The image of Fig. 9 (left) shows raw data from the report generator, such as casualties, police and ambulances. The image in the middle shows the corresponding view in 3D space. Even though 3D simulation has some disadvantages, such as unfamiliar interface and viewpoint control, the height and depth cues are an invaluable source of knowledge to a user [20]. This is especially true in a natural or man-made post-disaster simulation for an urban area. Height and depth cues are crucial for buildings or other volumetric spaces. As mentioned, this 3D view works in sync with the 2D graphics, so that viewers can easily switch their attention at their own preference. In the future the 3D view could be integrated with 3D building, structures and landmarks and to provide more comprehensive urban casualty visualization.



**Figure 9.** 2D view of Northridge Area (left), corresponding 3D view (middle), and actual implementation of multiple display visualization system.

## 4.2 Display Control

The requirement of fast identification is essential for visualization of low level fusion. In order to be able to view only federate data which is desirable, a menu bar system was developed to switch certain data layers ‘on’ and ‘off’. As mentioned, this avoids the problem of cluttering that occurs if too many reports are coming into the visualization federate at one time. A context menu has been included which can be accessed by right clicking in the display window and performs the same function as the menu bar (Fig. 10, left). In the 3D view, an independent viewing position was created to allow for an outside viewer to navigate throughout the 3D environment to give a better sense of realism. We took advantage of the ability to change the scale and size of the icons in 3D, so based on the size of the data, a user can get a better understanding of the disaster scenario.



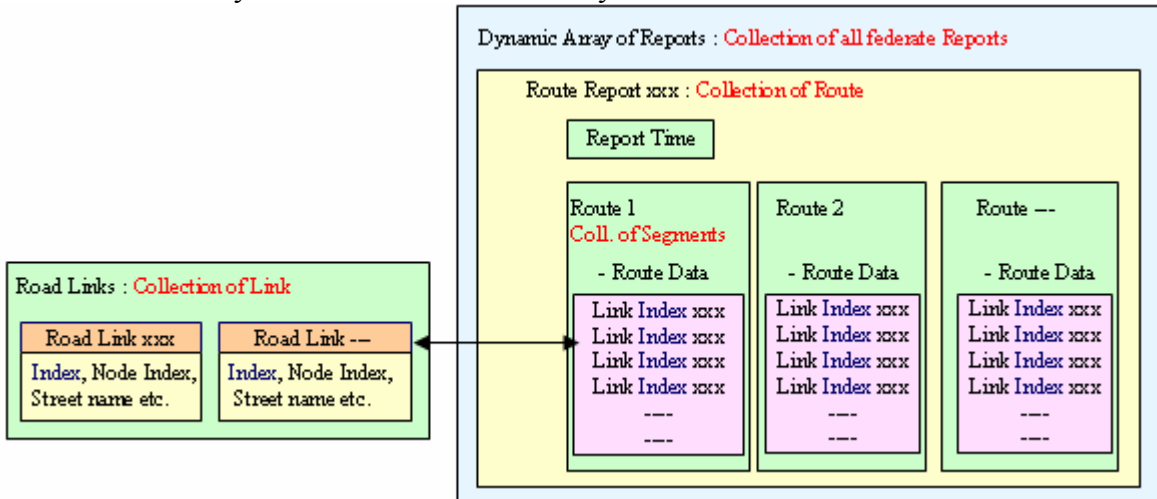
**Figure 10.** Fast data manipulation and identification: resource context menu (left), and pop-up window for object identification and the visualization of its trend (right).

## 4.3 Quick Identification

The information specific to each casualty is stored over time and is available to the user by simply hovering over a casualty icon on the screen. A pop-up window then appears which shows the pertinent information (Figure 10, right). Since it would be impossible to track each casualty if thousands were present the information regarding casualty clusters will be more useful for time aggregation analysis (section 4.4). Therefore, specific areas can be monitored during the simulation.

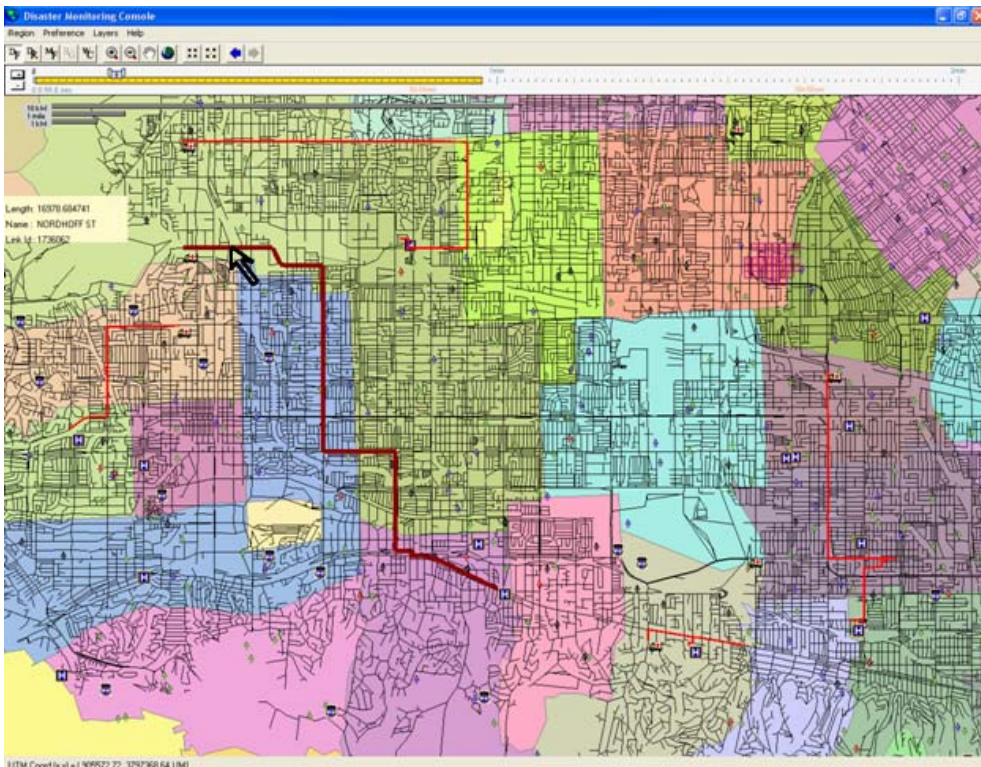
All ambulance route reports in the simulation come with road links Id’s, which define the routes for the ambulances. It is possible to store all road link Id’s in a report database and use that Id with a search algorithm to retrieve all related information by accessing a link database. However, this will increase runtime computation, which further lowers the performance. An easy and compatible solution to the problem is to store the object index in a container of road links objects. In this way, a search algorithm only needs to be run once when a report is received. The data can then be retrieved very quickly by just accessing the container by object index. We have used the binary search

algorithm to find index of road link object in a large container. This algorithm needs  $O(\log N)$  iterations in worst case to find an object in a container, which is acceptable for database of about 30,000 objects. To make efficient use of runtime computer memory it is desired that all the information regarding road link data be stored only once so that it can be retrieved as needed by accessing its unique Id (Fig. 11). Each file from the RTI contains multiple routes at a time which show a separate route for each ambulance. A hierarchical dynamic array structure has been created to store a report time and multiple route information at that time for each report. Each route can then be accessed by road link index in that array database.



**Figure 11.** Database design for information storage.

As shown in Fig. 12, ambulance paths can then be highlighted on the screen by merely hovering over them with the mouse. This allows a user to get a clearer understanding of the path and the specific road it may be located on. In order to optimize performance for graphical output, only the current path with the mouse tip is refreshed during mouse hovering. Therefore, the entire display does not have to be refreshed which can distract the user. To catch this mouse event, a small rectangular region is defined around each road link by considering its spatial orientation in the 2D window. As soon as any region catches the mouse tip the searching algorithm will exit with the current route index from the main report index. Based on the current ambulance route index a graphical region is created by joining small oriented rectangular regions of road links which will make the ambulance route. Finally, the system invalidation function is called to refresh only that region in the graphical output window. A backup route index is also stored in run-time memory which is helpful in refreshing the same route region again when the mouse tip moves out of that region.



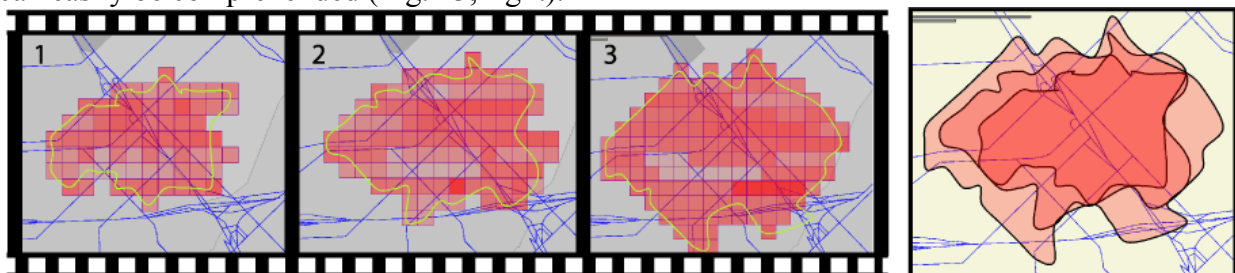
**Figure 12.** Ambulance route display in an urban area of Northridge region.

#### **4.4 Clusters: Regional Information**

A cluster can be defined as the extraction of large amounts of data and displaying it as a group in a vast database system. In a post-disaster simulation the need for dealing with large sets of data and comprehending abstract information is a key to understanding situation awareness. Work on clustering with fused data can be found in applications for a battlefield [21] and the visualization of large datasets [20]. In our simulation, the casualty cluster information is produced by the Level – 2 fusion federate [22].

##### *4.4.1 Formulation of Cluster and Boundary Formation*

As an approach to cluster visualization, the entire Northridge area was divided into cells with horizontal and vertical grid data. Fig. 13 (1) shows a cluster with a group of cells which defines the cluster. The outlier information is helpful in the identification of the area the cluster covers, as well as, interpolating intermediate shape information while morphing. Cell information is also effective in the quick identification of areas with the highest emergency within a cluster. In the figure, the opacity of a cell represents the number of casualties present and allows for quick identification of the most troubled spots. It is also possible to see clusters when only considering one level of severity by using the context menu. In addition to the cluster, a boundary can help a viewer recognize the shape of the cluster. With only boundary information, the trend of the regional shape can easily be comprehended (Fig. 13, right).



**Figure 13.** Situation awareness by cluster morphing. The casualty clusters between two discrete time steps (1 & 3) are generated by interpolation (2). The expanding trend of such cluster boundary is shown for better understanding of the situation over time (right).

##### *4.4.2 Dynamic Visualization by Morphing*

It is difficult to capture the trend of a situation from one specific time to another. In most cases, the user has to depend on his/her memory to relate the states. In order to provide a way of seeing how the clusters change with respect to time, we incorporated morphing into the simulation. All the visualization elements, such as position, color and shape were stored at each discrete time step. Corresponding discrete states were then interpolated to create the morphing. However, the associated cluster ID may change from one point to the next which makes it difficult to morph the clusters directly. Therefore, morphing was carried out on a cellular level. This way individual cells, which comprise the clusters, may appear and disappear from subsequent states to help inform the user. Using the morphing interface, a user can keep track of a cluster and cells, comprehend the trend, and have effective awareness of the situation (Fig. 13).

## 5. Conclusion and Future Work

The main attribute of the present work is to achieve a user-friendly and intuitive visualization methodology for massive post-disaster data fusion in a run-time HLA/RTI network. We developed an integrated GUI framework that provides the capability of fast identification of target objects and manipulation of data layers (low level fusion visualization). It also achieves dynamic visualization of time aggregated data for situation awareness (high level fusion visualization). This has been done through the use of 2D and 3D graphics displays in sync to provide relevant perspectives to the user.

We envision that the framework will be the basis for an actual emergency response center which relies on visualization to help allocate resources in the event of a disaster. For this, a series of performance test will be implemented in the future to verify the usability of the interface. Future plans also include visualization of hazardous material, such as chemical spill and toxic plume, and inclusion of tactical or strategic information for the assessment of future threat caused by a natural or man-made disaster.

## Acknowledgements

This work was supported by the AFOSR under award F49620-01-1-0371. The authors gratefully acknowledge valuable input from James Scandale, Dr. Peter Scott and Dr. James Llinas.

## References

- [1] D. Ozisik and N. Kerle, Post-earthquake damage assessment using satellite and airborne data in the case of the 1999 Kocaeli earthquake, Turkey. *Proc. of the XXth ISPRS congress: Geo-imagery bridging continents*, Istanbul, Turkey, 2004, 686-691.
- [2] J. Llinas, Information Fusion for Natural and Man-Made Disasters, *Proc. 5th International Conference on Information Fusion*, Annapolis, MD, USA, 2002.
- [3] L. A. Treinish, Visual Data Fusion for Applications of High-Resolution Numerical Weather Prediction, *Proc. of IEEE 2000 Visualization Conference*, Salt Lake City, UT, USA, 2000, 477-480.
- [4] K. Severance, P. Brewster, B. Lazos, and D. Keefe, Wind Tunnel Data Fusion and Immersive Visualization: A Case Study, *Proc. of IEEE 2001 Visualization Conference*, San Diego, CA, USA, 2001.
- [5] T.-J. Hsieh, F. Kuester, and T. C. Hutchinson, Visualization of Large-Scale Seismic Field Data, *Proc. of 2003 High Performance Computing Symposium*, Orlando, FL, USA, 2003.
- [6] G. Srimathveeravalli, N. Subramanian, and T. Kesavadas, A Scenario Generation Tool for DDF Simulation Testbeds, *Proc. of 2004 Winter Simulation Conference*, Washington D.C., 2004.
- [7] I. Rauschert, P. Agrawal, S. Fuhrmann, I. Brewer, R. Sharma, G. Cai, A. MacEachren, and H. Wang, Designing a Human-Centered, Multimodal GIS Interface to Support Emergency

- Management, *Proc. of ACM International Symposium on Advances in Geographic Information Systems*, McLean, VA, USA, 2002.
- [8] S. Jain and C. McLean, A Framework for Modeling and Simulation for Emergency Response, *Proc. of 2003 Winter Simulation Conference*, New Orleans, Louisiana, USA, 2003.
- [9] Y. Kim and T. Kesavadas, Automated Dynamic Symbology for Visualization of High Level Fusion, *Proc. of 7th International Conference on Information Fusion*, Stockholm, Sweden, 2004, 944-950.
- [10] Northridge Earthquake, Southern California Earthquake Center. Retrieved on March 20, 2005 from [http://www.data.scec.org/chrono\\_index/northreq.html](http://www.data.scec.org/chrono_index/northreq.html).
- [11] HAZUS, Multihazard Loss Estimation Methodology, FEMA. Retrieved on January 11, 2005 from <http://www.fema.gov/hazus/>.
- [12] P. D. Scott and G. L. Rogova, Crisis Management in a Data Fusion Synthetic Task Environment, *Proc. of 7th International Conference on Information Fusion*, Stockholm, Sweden, 2004, 330-337.
- [13] Q. Gong, A. Jotshi, and R. Batta, Dispatching/Routing of Emergency Vehicles in a Disaster Environment using Data Fusion Concepts, *Proc. of 7th International Conference on Information Fusion*, Stockholm, Sweden, 2004, 967-974.
- [14] S. Xiaoxia and Z. Quihai, "The Introduction on High Level Architecture (HLA) and Run-Time Infrastructure (RTI)," *Presented at SICE Annual Conference*, Fukui, Japan, 2003.
- [15] K. Watson, D. Espinosa, Z. Greenvoss, J. H. Pedersen, C. Nagel, J. D. Reid, M. Reynolds, M. Skinner, and E. White, *Beginning Visual C#*. Indianapolis, IN: Wiley Publishing, Inc., 2003.
- [16] USGS, U.S. Geological Survey. Retrieved on January 11, 2005 from <http://www.usgs.gov>.
- [17] TeleAtlas. Retrieved on March 20, 2005 from <http://www.teleatlas.com>.
- [18] G. M. Karam, Visualization using Timelines, *Proc. of the 1994 ACM SIGSOFT international symposium on Software testing and analysis*, Seattle, Washington, USA, 1994, 125-137.
- [19] E. White, C. Garrett, and S. Robinson, *GDI+ Programming: Creating Custom Controls Using C#*. Birmingham, UK: Wrox Press Ltd., 2002.
- [20] A. Schilling and A. Zipf, Generation of VRML city models for focus based tour animations: integration, modeling and presentation of heterogeneous geo-data sources, *Proc. of the eighth international conference on 3D Web technology*, Saint Malo, France, 2003.
- [21] C. M. Hoffman, Y. J. Kim, R. P. Winkler, J. D. Walrath, and P. J. Emmerman, Visualization for Situation Awareness, *Proc. of the 1998 workshop on New paradigms in information visualization and manipulation*, Washington D.C., United States, 1998.
- [22] K. Chen and L. Liu, ClusterMap: Labeling Clusters in Large Datasets via Visualization, *Proc. of ACM Conference on Information and Knowledge Management*, Washington D.C., USA, 2004, 285-293.

