CoRAVEN: Modeling and Design of a Multimedia Intelligent Infrastructure for Collaborative Intelligence Analysis*

P. M. Jones[1], C. C. Hayes[2], D. C. Wilkins[1], R. Bargar[3], J. Sniezek[1], P. Asaro[1], O. Mengshoel[1], D. Kessler[1], M. Lucenti[1], I. Choi[3], N. Tu[2], and MAJ J. Schlabach[4]

[1]Beckman Institute, University of Illinois at Urbana-Champaign
[2]Dept. of Mechanical Engineering, University of Minnesota
[3]NCSA, University of Illinois at Urbana-Champaign
[4]United States Army Intelligence and Security Command

ABSTRACT

Intelligence analysis is one of the major functions performed by an Army staff in battlefield management. In particular, intelligence analysts develop intelligence requirements based on the commander's information requirements, develop a collection plan, and then monitor messages from the battlefield with respect to the commander's information requirements.

The goal of the CoRAVEN project is to develop an intelligent collaborative multimedia system to support intelligence analysts. Key ingredients of our design approach include (1) significant knowledge engineering activities with domain experts, (2) representation of an explicit model of reasoning and activity to drive design, (3) the use of Bayesian belief networks as a way to structure inferences that relate observable data to the commander's information requirements, (4) collaborative graphical user interfaces to provide flexible support for the multiple tasks in which analysts are engaged, (5) sonification of data streams and alarms to support enhanced situation awareness, (6) detailed psychological studies of reasoning and judgment under uncertainty, and (7) iterative prototyping of candidate designs with domain experts for both formative and summative evaluation. This paper will discuss our current progress on all these fronts.

1. INTRODUCTION

Cognitive systems engineering research studies human activity in context, identifies problems in human-machine interaction, and designs and evaluates technology solutions to those problems. The CoRAVEN ("Collaborative RAVEN") project is an example of such research. The goal of this project is to study the process of Army intelligence collection management and analysis, define issues, and apply advanced technology towards solutions. In so doing, we plan to make substantive contributions to the theories and design methodologies of cognitive engineering as well as provide a proof-of-concept prototype that is a promising tool for the Army.

Currently, we have analyzed Army doctrine and other archival materials and conducted informal interviews and collaborative verbal protocol sessions with domain experts. Our analysis so far has highlighted (1) the problem of data overload and information filtering, (2) the importance of flexibility in collaborative support for teams, (3) the importance of maps in reasoning and planning, and (5) significant individual differences in reasoning under uncertainty and concomitant design approaches to support different cognitive styles.

To address these problems, the CoRAVEN project is engaged in developing a proof-of-concept tool in which analysts are able to view spatial data (maps), temporal data (the *synchronization matrix* which represents the schedule of collection assets) and graph-based models for fusing evidence (Bayesian networks).

* Appears in *1998 IEEE International Conference on Systems, Man, and Cybernetics*, October 11-14, 1998, San Diego, California.

The coordination of multiple views by a single user, and collaboration among multiple users, are also key issues. Furthermore, viewing rich dynamic data is not merely a visual process, but auditory as well; we are exploring methods for data sonification to better support situation assessment. Finally, we continue to engage domain experts in knowledge elicitation and design feedback sessions with our evolving prototypes as well as conduct more rigorous psychological experiments to validate portions of our approach and gain insight into the individual variability of our target population.

The above topics are explored in this paper as follows. Section 2 describes the application domain of intelligence collection management and analysis. Section 3 is an overview of the CoRAVEN user interface concept. Section 4 focuses on Bayesian networks. Section 5 describes sonification, Section 6 collaboration, and Section 7 presents the CoRAVEN architecture. Section 8 describes plans for the first set of psychology experiments.

2. THE APPLICATION DOMAIN: INTELLIGENCE COLLECTION MANAGEMENT AND ANALYSIS

In the military, “intelligence” refers to knowledge of the enemy: the collection, management, and analysis of data and information about enemy locations, forces, and so on. Intelligence is thus a standard function represented by an officer on a commander’s staff.

The work of Intelligence Collection Management (CM) and Analysis has a precisely defined role in military operations. Doctrinally, CM consists of several sub-functions including Requirements Management, Mission Management, Asset Management, Analysis, and Dissemination. The earlier planning sub-functions turn the intelligence needs of the commander's operational plan into formalized *Intelligence Requirements* (IRs) and *Priority Intelligence Requirements* (PIRs), a *Collection Plan* for investigating those areas in which significant activity is most likely to be observed called *Named Areas of Interest* (NAIs), and a *Synchronization Matrix* for allocating limited collection resources to NAIs such that the PIRs and IRs can be satisfied in a timely manner. CM begins while operational preparations are still being made and is repeated as necessary during the course of operations. The later stages of the Intelligence Cycle consist of analyzing, communicating and presenting the intelligence gathered by collection assets. It is the responsibility of the Intelligence Officer, and the central organizing principle of this officer's staff, to present the commander with a full, timely, and organized account of the intelligence collected and its significance for the commander’s operational decision making.

3. CoRAVEN USER INTERFACE CONCEPT

CoRAVEN is intended to be a flexible resource for intelligence analysis by providing easy navigation among three interrelated models and views: information abstractions about how observable evidence maps to PIRs and IRs, spatial abstractions such as NAIs that are used to organize planning and analysis, and temporal view of the synchronization and operations matrices. Providing flexible ways for analysts to map among these three interrelated models/views is a critical feature of CoRAVEN.

Currently, CoRAVEN has been designed to support the analysis of collected information and the communication of its significance among the analysis staff and to decision makers, and in particular addresses the following challenges: 1) identifying all of the information relevant to a given decision, 2) efficiently and reliably assessing the significance of all of the relevant information, and 3) effectively communicating the significance and relevance of information to a given decision. CoRAVEN seeks to address these issues by using Bayesian networks (BNs) to structure the relationship of evidence to PIRs and IRs and providing a collaborative audio-visual environment for the visualization and sonification of BNs, their evidential sources, and their relationship to the geographic and temporal structure of the situation.

4. BAYESIAN NETWORKS

The name CoRAVEN comes from “Collaborative RAVEN”. RAVEN is a research project on using Bayesian networks as a reasoning tool for intelligence analysts (Mengshoel and Wilkins, 1997, 1998a, 1998b]. Bayesian networks (BNs) are an important knowledge representation that are used for reasoning and learning under uncertainty [Pearl, 1988] [Lauritzen & Spiegelhalter, 1988]. Probability theory and graph theory form their basis: random variables are nodes and conditional dependencies are edges in a directed acyclic graph. Edges typically point from cause to effect. Consider a simple Bayesian network consisting of five nodes A, B, C, D, and E. In addition to the graph, there are conditional probability tables associated with each node V and its parents Pa(V), expressing the conditional probability Pr(V | Pa(V)). If the node D has two parents B and C, assuming discrete binary nodes with values {0,1}, Pr(D=0 | B=1, C=0)) is one of the entries in D's conditional probability table. Static and temporal BN variants can be used to model static and dynamic environments such as battlefields [Mengshoel & Wilkins, 1997].

Inference in Bayesian networks is one focus of our research. The inference task of belief updating amounts to the following: Given evidence at node E and query node Q, infer posterior probability Pr(Q | E=ei). Any nodes in the network can be evidence or query nodes. For the example BN, this leads to different types of inference: *diagnostic* as in Pr(A | E=ei); *causal* as in Pr(E | A=aj), and *mixed* as in Pr(D | E=ei, A=aj). A variety of approaches to Bayesian network belief updating and belief revision have been investigated [Pearl, 1988; Lauritzen & Spiegelhalter, 1988]. These inference algorithms vary in many respects: they are exact, approximate, or heuristic; work on singly or multiply connected graphs; and are used for different inference tasks. Computational hardness has been shown both for belief updating and belief revision. Research into non-exact algorithms for solving these tasks approximately or heuristically is therefore important.

A commercial BN tool, HUGIN, uses an exact algorithm known as cluster propagation [Lauritzen & Spiegelhalter, 1988]. For sparse BNs this algorithm works well; however, for large and highly connected BNs, it can become too slow for practical use. For this reason, a heuristic approach to belief revision in BNs is also investigated [Mengshoel & Wilkins, 1998a]. More specifically, we consider a BN as encoding a genetic algorithm (GA) fitness. This is a restriction on the fitness function, but probability theory in general and BNs in particular have proven sufficiently rich to make this an interesting restriction. Part of our research has focused on GA selection and BN abstraction, and we have shown promising results for GA-based belief revision [Mengshoel & Wilkins, 1998b] as well as integrating BN abstraction and refinement into GA-based belief revision [Mengshoel & Wilkins, 1998a].

CoRAVEN currently relies on the standard HUGIN implementation of BNs. (In later versions of CoRAVEN, we hope to incorporate advanced algorithms of the type described above to have a more efficient implementation). Our hypothesis is that a BN is a good normative model of the intelligence analysis process; that is, it expresses how good intelligence analysts should reason about evidence to answer PIRs and IRs. In particular, our approach is that each PIR and IR has an associated Bayesian network, with the top node being the PIR or IR itself, and the leaf nodes representing observable evidence. Thus, in our demonstration, analysts using CoRAVEN must navigate among a number of BNs (currently eight), where each BN can be fairly large (the largest networks in our demonstration are about 650 nodes). Hence, one critical issue is how analysts will be able to monitor dynamic updates to all these networks as messages are received from intelligence assets, thus triggering state changes in leaf nodes with inferences propagating throughout the networks. Part of our answer to this is data sonification, which is the subject of the next section.

5. SONIFICATION

Sonification is the transformation of numerical data into sound for purposes of observing that data. The essential research task of sonification is to identify and construct an intuitive perceptual space for the auditory display of data. This task includes the assimilation of technological, creative and scientific advances in sound synthesis and signal processing and in human perception and cognition (see Brady et al., 1996; Choi, 1997; http:// www.ncsa.uiuc.edu / ~audio).

The NCSA Sound Server (VSS) (Bargar et al., 1994) facilitates the application of sonification in scientific research by providing a distributed system and language for ubiquitous sound production in computational environments. The VSS supports both sound computation and sound authoring; the latter is the process of establishing automated relationships between objects or events in a silent computing application, and algorithms for sound production which operate in parallel to the silent application.

In CoRAVEN, sound authoring is applied to the Bayesian network display in two different ways: (1) as a way of monitoring the dynamic evolution of weights on the nodes and (2) as a means of users setting alarms related to certain nodes. The complexity of the BN is difficult to visualize, particularly the relative contribution of internal nodes to the final outcome. We apply sound authoring in layers of musical patterns that represent the probabilities at internal nodes. The use of musical patterns facilitates the ability to maintain coherence when information from many nodes is presented at the same time. Temporal patterns provide a high-dimensional space for differentiating elements in a complex state. Gradient alarms may be configured to report the onset of special conditions at a node. A gradient alarm informs a listener continuously as a system approaches or recedes from a designated alarm state, by the degree of onset of a notable change in the auditory texture.

Sonification supports the background monitoring of parallel processes in eyes-busy, hands-busy scenarios. The system architecture supporting data-driven sound with distributed sound synthesis engines allows efficient transmission of sonification among remote collaborators, and can assist the rapid exchange of alternative versions of a PIR representation during the decision-making process.

6. COLLABORATION

Intelligence analysis is a multi-person process, and thus, CoRAVEN also needs to address issues in collaboration and cooperative problem solving. Jones et al (1998) describe basic theoretical issues in cooperative work, such as sharing an information space, articulation work, and presence.

Our analysis has revealed a large repertoire of collaboration support features necessary. For example, analysts may want to share their current Bayesian networks with colleagues for comments, or may want to collaborate synchronously on setting alarms, or simply have shared displays of the networks as they are updated.

Thus, in CoRAVEN, we are designing two complementary demonstrations of collaboration technology. One demonstration is Java-based and will support shared displays and simple generic collaboration mechanisms such as real-time chatting and shared drawing on a whiteboard. The other demonstration in Visual C++ exploits the Watch-and-Notify feature of the POET™ commercial object-oriented database management system to support synchronous collaboration. Importantly, the POET implementation supports a flexible collaboration policy at runtime.

7. CoRAVEN ARCHITECTURE

All the technical ingredients described thus far are integrated into the CoRAVEN architecture shown in Figure 1. The architecture is rather like the Model-View-Controller paradigm, where the top four components (Graphical User Interface (GUI) and VSS message groups) are Views of the Models implemented in POET, HUGIN, and spatial data files or a geographic information system (GIS). The Controller (GUI-IO connectivity) manages the relations between the displays and the underlying models.

The user navigates among four different models and views of data: spatial (maps and overlays), temporal (operations and synchronization matrices), evidence (Bayesian networks), and sonic (VSS). Each of these views has edit capabilities so that users can flexibly set up mappings between the views and can also configure their displays as desired.

The controller/connectivity layer mediates interactions among the user displays and the deep models from which the displays are generated. The intention of this approach is to support flexible relations among views and models. The models include objects in POET, the Bayesian networks in HUGIN, spatial data in ESRI shape files or in a complete GIS system, and a VSS proxy that communicates with the VSS server to generate sound. To simulate battlefield messages, the Temporal Simulator sends time-stamped SALUTE reports to POET, which interfaces with the HUGIN inference engine and a memory management daemon.

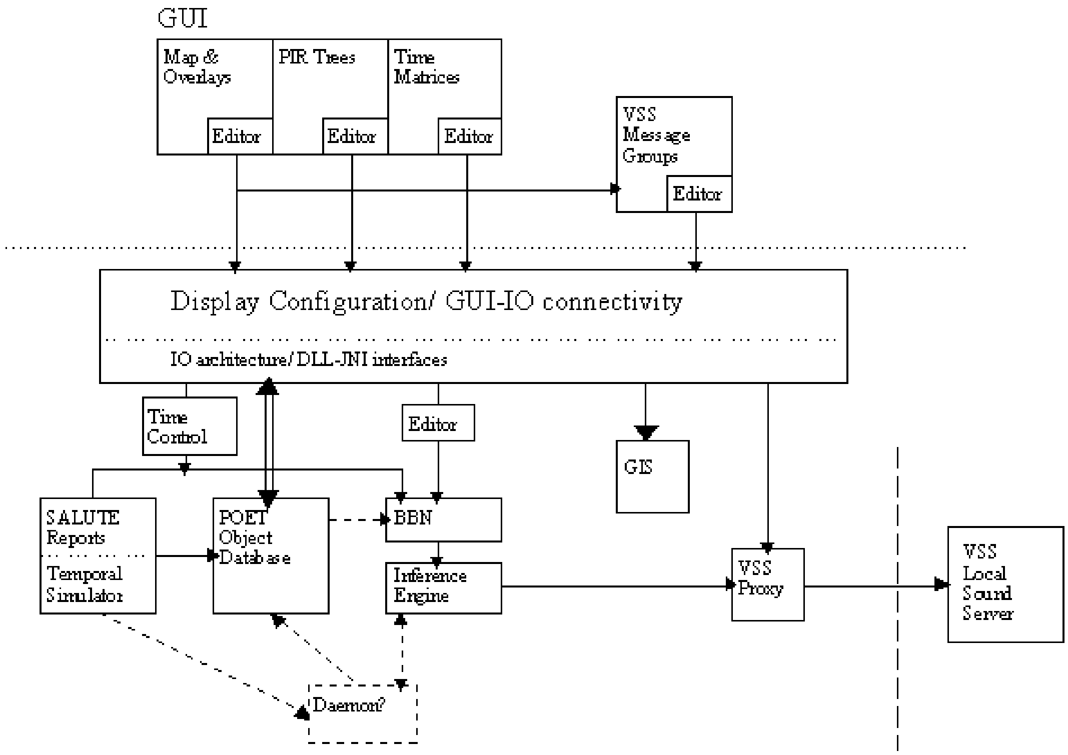

Figure 1. CoRAVEN Architecture.

8. PSYCHOLOGICAL STUDIES

In the context of developing and evaluating CoRAVEN, a number of basic psychological research questions exist, in addition to the validation of the actual BNs used in CoRAVEN with other domain experts. These questions include (1) the effects of common vs. unique knowledge (as distributed in a collaborative interface such as CoRAVEN) on expert uncertainty assessments (2) individual differences in efficiency and accuracy in using CoRAVEN (3) the impact of person perception as represented in CoRAVEN on expert uncertainty assessments, and (4) biases in expert uncertainty assessments via CoRAVEN.

Chao & Salvendy (1995) and Lehner & Zirk (1987) have proposed that there are individual differences in cognitive abilities of experts which subsequently affect the effectiveness of various knowledge elicitation techniques. The cognitive abilities in question are: associational fluency,

expressional fluency, figural fluency, ideational fluency, integrative processes, general reasoning, logical reasoning, verbal comprehension, flexibility of use, and induction. Though primarily serving to alleviate information overload in battlefield reasoning situations, the CoRAVEN interface can be alternatively conceptualized as a knowledge elicitation tool which is extracting probability assessments from users. In an extension of Chao & Salvendy's idea, we will investigate whether subjects exhibit variability on ten different cognitive abilities, and then whether such differences may affect the probability data elicited by the scenarios.

The first study is planned for July 1998 and will focus on how experts generate probability assessments given the Bayesian network structures used in CoRAVEN and on how these estimates are modulated by demographics (particularly expertise and years of experience) and cognitive style. In particular, subjects will first generate probability values in the context of a portion of one of the CoRAVEN Bayesian networks, and then will be shown the actual values used in CoRAVEN and will be asked to resolve and explain discrepancies between their own estimates and those embodied in CoRAVEN.

ACKNOWLEDGMENTS

This research is supported by the Army Research Laboratory and the Army Intelligence and Security Command via the Advanced and Interactive Displays Consortium. Consortium sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory or the U.S. Govemment.

REFERENCES

Bargar, R. and Choi, I. (1998). Sonification of probabilistic belief networks for a distributed environment. *Proceedings of the IEEE 1998 International Conference on Systems, Man, and Cybernetics*, to appear.

Bargar, R., I. Choi, S. Das, and C. Goudeseune. (1994). Model-based interactive sound for an immersive virtual environment. *Proceedings of the 1994 International Computer Music Conference*, 471-474. Aarhus, Denmark: International Computer Music Assn.

Brady, R., Bargar, R., Choi, I. and Reitzer, J. (1996). Auditory Bread Crumbs for Navigating Volumetric Data. *Proceedings of IEEE Visualization '96*. San Francisco: IEEE.

Chao, A. and Salvendy, G. (1995). Impact of cognitive abilities of experts on effectiveness of elicited knowledge. *Behavioral & Informational Technology*, 14(3), 174-182.

Choi, I. Sound synthesis and composition applying time scaling to observing chaotic systems. Proceedings of the Second International Conference on Auditory Display, ICAD '94, pp. 79-107. Santa Fe Institute, Santa Fe, NM.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley.

Jones, P. M., Hayes, C. C., Fiebig, C. and Dunmire, C. (1998). Cooperative problem solving: A cognitive engineering and distributed cognition view. *Proc. ARL Advanced Displays and Interactive Displays Fed. Lab. Second Annual Symposium*, February 1998, College Park MD.

Lauritzen, S. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society series B*, 50(2):157-224.

Lehner, P.E. & Zirk, D.A. (1987). Cognitive factors in user-expert system interaction. *Human Factors*, 29(1), 97-109.

Mengshoel, O. J. and Wilkins, D. C. (1997). Visualizing Uncertainty in Battlefield Reasoning Using Belief Networks. *Proceedings ARL Advanced Displays and Interactive Displays Fed. Lab. First Annual Symposium*, pages P-15 - P-22, Jan. 1997, Adelphi, MD.

Mengshoel, O. J. and Wilkins, D. C. (1998a). Abstraction for Belief Revision: Using a Genetic Algorithm to Compute the Most Probable Explanation. *Proc. AAAI Spring Symposium Series on Satisficing Models*, March 1998, Stanford, CA.

Mengshoel, O. J. & Wilkins, D. C. (1998b). Genetic Algorithms for Belief network inference: The Role of Scaling and Niching. *Proc. Seventh Annual Conference on Evolutionary Programming*, March 1998, San Diego, CA.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.