12th ICCRTS-2007 “Adapting C2 to the 21st Century”

Title: **Identifying the Enemy – Part I: Automated Network Identification Model**

Suggested Topics: **Modeling and Simulation, Network-Centric Experimentation and Applications, Organizational Issues**

Authors:

Georgiy M. Levchuk
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-935-3966x267
Fax: 781-935-4385
e-mail: georgiy@aptima.com

Feili Yu
Storrs, CT
Phone: 860-486-2890
Fax: 860-486-5585
e-mail: yu02001@engr.uconn.edu

Haiying Tu
Qualtech Systems, Inc.
Putnam Park, Suite 603
100 Great Meadow Road
Wethersfield Connecticut 06109
Tel: (860) 257-8014
Fax: (860) 257-8312
e-mail: tu@teamqsi.com

Krishna R. Pattipati
Professor, ECE Dept., UCONN
Storrs, CT
Phone: 860-486-2890
Fax: 860-486-5585
e-mail: krishna@engr.uconn.edu

Yuri Levchuk
Aptima Inc.,
1726 M Street, N.W., Suite 900
Washington, DC 20036
Phone: (202) 842-1548x323
Fax: (202) 842-2630
e-mail: levchuk@aptima.com

Elliot Entin
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-935-3966
Fax: 781-935-4385
e-mail: entin@aptima.com

Correspondence:

Georgiy M. Levchuk
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-935-3966x267
Fax: 781-935-4385
e-mail: georgiy@aptima.com

Identifying the Enemy – Part I: Automated Network Identification Model

Abstract

To successfully predict the actions of an adversary and develop effective counteractions, the knowledge of the enemy organization, objectives, and the modus operandi are needed. Current approaches to analyzing the threat are manual, labor-intensive, and require significant amount of time. In empirical studies, humans exhibited decision and confirmatory biases, which negatively impacted their assessment of the adversary. Compounded by huge amounts of data, large information gaps, and significant complexity of the problem people need to analyze, the ability of the intelligence teams to recognize an active enemy is reduced, further resulting in decreased effectiveness of counteractions and unintended consequences.

In this 2-part paper, we discuss a project that focused on developing an automated adversarial organization identification technology. Used as decision support system, this technology, called NetSTAR, promises to result in significant manpower, decision time and error reductions during threat analyses tasks. In our experiments, the NetSTAR system has significantly outperformed unaided human analysts. In part I of this 2-part paper, we discuss the problem setup and provide a description of the computational algorithms at the core of the NetSTAR system. A computational experiment is provided to assess the capabilities and robustness of the NetSTAR algorithms to data uncertainty and problem complexity.

1. Motivation: Adversarial Analysis Problem

The U.S. Army conducts operations using doctrinal military decision-making process (MDMP) (Wade, 2005). One of the important steps in the MDMP process, intelligence preparation of the battlefield (IPB), requires the assessment of enemy’s command and control structure to predict the actions of the adversary, identify high-value targets, and develop effective counteractions. Currently, the intelligence operations officer provides input to help the planning officer develop the IPB templates, databases, and other products that portray information about the adversary and other key groups (Figure 1) in the area of operations and area of interest. These products contain information about each group’s leaders and decision makers, size and location of enemy forces, and linkages among groups and leaders. The linkage information is produced manually from the data on activities (using activity matrix template) and intelligence on the relationships between individuals. Using this information, a link diagram is developed to show the interrelationships of individuals, organizations, and activities.

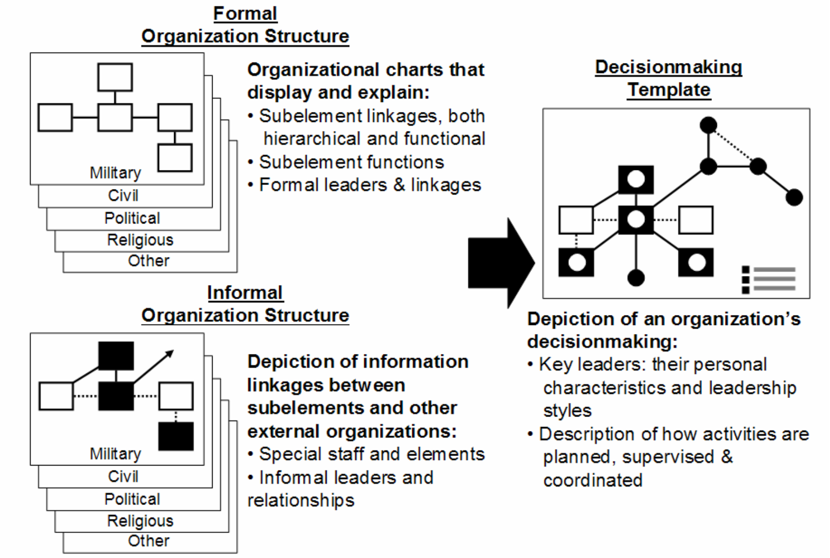

Formal
Organization Structure
Military
Civil
Political
Religious
Other
Organizational charts that display and explain:
• Subelement linkages, both hierarchical and functional
• Subelement functions
• Formal leaders & linkages
Informal
Organization Structure
Military
Civil
Political
Religious
Other
Depiction of information linkages between subelements and other external organizations:
• Special staff and elements
• Informal leaders and relationships
Decisionmaking
Template
Depiction of an organization’s decisionmaking:
• Key leaders: their personal characteristics and leadership styles
• Description of how activities are planned, supervised & coordinated

Figure 1: Sample Information Operations Doctrinal Template (adapted from (FM 3-13, 2003))

Knowledge of connections (e.g., communication, command) between individuals and specific roles of individuals in the covert organization are needed because of the following effects. First, connections provide a means to share information and resources, and coordinate execution of operations. Second, captured individuals can share information about those to whom they are connected. Since it is a given that members of a cell share information and can compromise one another, the relevant question might be how interconnected are the cells that make up the organization? And third, capturing individuals, destroying enemy’s mission-critical resources, or disabling organizational connections would allow the disruption of enemy’s operations and decision–making processes for preemptive actions. Therefore, identifying an enemy objectives and command and control (C2) organization – command hierarchy, communication networks (formal and informal), control structure (amount, distribution, and access to resources), and roles of individuals – are the key elements of current U.S. Army information operations.

Generally speaking, an *organization* is a group of people intentionally brought together to accomplish an overall, common goal or a set of goals. Organizations can range in size from two people to tens of thousands. One of the most common ways to look at organizations is as social systems (McNamara, 2005). Simply put, a system is an organized collection of parts that are highly integrated in order to accomplish an overall goal. Organizations exist in many domains – military, business, civic, political, religious, as well as virtual. These organizations have different decision-making principles, levels of decentralization, formalization and adherence to strict organizational rules and doctrines. There are many difficulties associated with identifying the organizations that have many informal relationships among their members and change dynamically over time. In our research, we focus on the command and control (C2) organization, which is designed to manage personnel and resources to accomplish the mission requiring their collective employment. Such organizations are distinguished by relatively formal structures and limited variability over time, and are common to both friendly and adversary military forces. Given specific functions and principles of individuals together with the structural form in which they are organized, myriads of the different potential organizations can be constructed. All of them are based on the underlying C2 principles defining how individuals interact in the organization and what actions they perform (Alberts and Hayes, 2006). These interactions can be utilized to detect and understand organizational relationships.

Currently, only a limited set of tools is available to intelligence operators to analyze, correlate and visualize the data. The two most commonly used network analysis tools are StarLight[1] and AnalystNotebook[2]. These tools are often used together with technologies performing data mining and automated entity and link discovery from text sources (Miller et al., 2000; Grishman, 2003; Stolfo et al., 2003) or manually using HUMINT and other data sources. They rely on domain understanding (Krebs, 2001; Sageman, 2004) or applied social network analyses (Van Meeter, 2001; Dombroski and Carley, 2002; Dombroski, Fischbeck, and Carley, 2003; Skillicorn, 2004). However, these tools merely present and visualize the networks formed by observations and do not solve the network identification problem of “cleaning uncertain observations” in the presence of missing, irrelevant, deceptive, and mislabeled attributes and links. It is evident that none of the existing tools provide automated threat prediction and assessment capabilities that can reason from multi-source data and that support the decisions about the enemy command and control organization.

As a result, current approaches to analyzing the threat are manual: the intelligence analysts rely on their experience to make sense of visualized structural and temporal data. Large information gaps, including missing data, deceptions, and errors, have to be dealt with, and analysts often fill the gaps with their experience, which may not be applicable to the problem they need to solve, thereby resulting in *decision biases*. In addition, people tend to exhibit *confirmatory biases* when the first seemingly valid hypothesis is selected and further relied upon during the analysis. This issue is compounded by huge amounts of data and complexity of the problem people need to analyze, influencing what data is used and which is filtered out

[1] http://starlight.pnl.gov/

[2] http://www.i2.co.uk/Products/Analysts_Notebook/default.asp

and thus never studied. All these factors negatively impact the ability of the intelligence team to recognize an active enemy and further result in decreased effectiveness of counteractions and unintended consequences.

At present, there are no tools that can help the analysts reduce the number of hypotheses that need to be analyzed, or focus their attention on only the most critical information, thereby filtering out the information that is not critical to identification of adversarial roles and relationships. None of the existing tools can utilize previous experiences of adversarial network analyses in making cross-references from current situation to previous case studies. In addition, only experts in organizational theory can take full advantage of the existence of topological constraints on the organizational structures of the adversary and effects of structures on individual and team behavior.

In our 2-part paper, we discuss a project that focused on developing an automated adversarial organization identification technology. Used as a decision support system, this technology, called NetSTAR, promises to result in significant manpower, decision time and error reductions during threat analyses tasks. In particular, this decision support system will offer the following benefits:

- Detect and classify the groups and individual actors’ roles;
- Identify the adversaries’ objectives and predict their next actions;
- Allow users to define hypotheses about the adversarial networks;
- Construct and store network models from previous adversarial analyses;
- Match hypotheses and model networks against currently observed data to rank-order hypotheses and offer the users a limited set for further analysis;
- Focus the user on analyzing or collecting the data elements most critical to adversary’s identification.

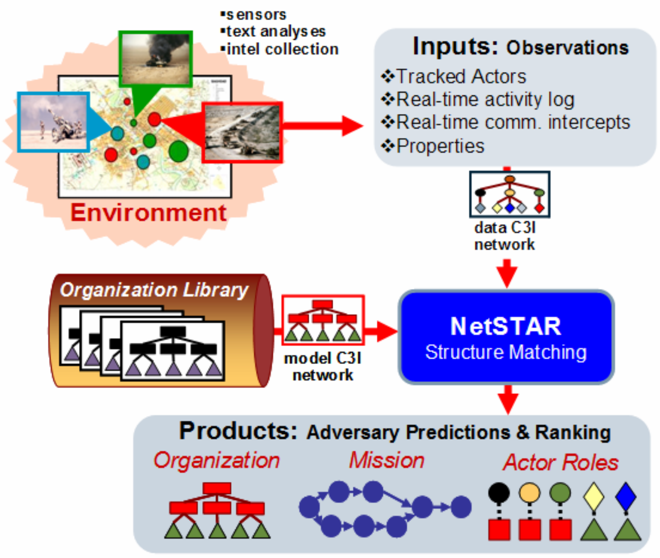


Figure 2: NetSTAR Adversarial Identification Process

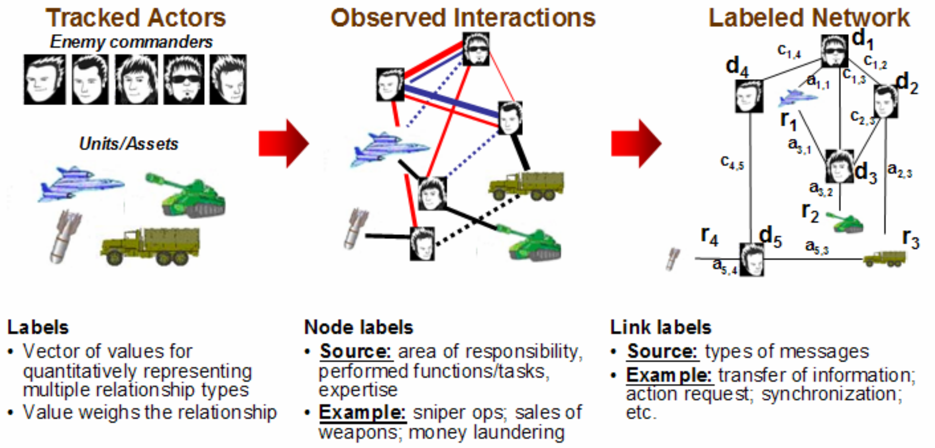
As the first phase of system validation, we have compared the ability of NetSTAR to automatically identify observed (data) networks against unaided human analysts. The set of potentially true (model) networks was given, and both human analysts and NetSTAR system had to pick a model network to correspond to observed data network and specify the match between the nodes of model and data networks. In our experiments, the NetSTAR system has significantly outperformed unaided human analysts. In part I of this 2-part paper, we discuss the problem setup and provide a description of the computational algorithms at the core of the NetSTAR system. Computational experiment results are provided to assess the capabilities and robustness of the NetSTAR algorithms to data uncertainty and problem complexity.

2. Method: Automated Organization Identification Model

NetSTAR is an optimization-based model to identify an adversarial organization and mission using observations about actors’ actions and interactions (Levchuk and Chopra, 2005; Levchuk, Levchuk, and Pattipati, 2006). The NetSTAR (Figure 2) model performs hypothesis testing using a probabilistic attributed graph matching algorithm. The algorithm finds a mapping of observed actors (nodes in observed data network) to organizational positions (nodes in model networks from organizational library) and rank-orders the organizational network hypotheses based on their likelihood values.

The data used by NetSTAR for identifying adversarial organization and mission consists of partially categorized interactions and relationships among tracked actors (e.g., communication transactions, such as “members of a militant wing engaged in a meeting with weapons suppliers at 11:35 am for 35 min to procure explosives”) and their individual actions (e.g., individual and joint operations, such as “BLUE team discovered a safe house and apprehended RED operatives attempting to manufacture weapons”). Such data is very noisy and sparse due to challenges in data collection, e.g. limited sensors and/or human intelligence, security of adversary communications, uncertainty in voice capturing and text translation, data association uncertainty, etc. Therefore, our framework has to rely on **probabilistic association** between tracked actors and the nodes in the model (hypothesized) networks.

As a result of observation pre-processing, we are observing a network of relationships of different types among the enemy actors (individuals, groups, physical resources), tasks, goals, etc. This network must be mapped to the network of command, control, and communications of the hypothesized organization. Given this type of data, we pose the problem as one of finding the mapping between the nodes of two graphs: observed (also termed *data*) network of adversary actors and their interactions/relationships, and a hidden network corresponding to the hypothesized (also termed *model*) network (Figure 3). The mapping is found by maximizing a match score, which could be a likelihood function or a posterior probability. The mapping must account for the attributes or features of both nodes and links that are mapped, and the models of attribute uncertainty (the probability of observing the attribute(s) correctly). Node attributes can include areas of responsibility, performed functions and/or tasks, expertise of the node (e.g., sniper operations; weapons sales; money laundering; etc.), while the link attributes may correspond to types of interactions and relationships between nodes in the adversary C2 organization (e.g., communication messages may be of the following type: request for or transfer of information, resource, action; acknowledgement; etc.).



Labels
- Vector of values for quantitatively representing multiple relationship types
- Value weighs the relationship

Node labels
- **Source:** area of responsibility, performed functions/tasks, expertise
- **Example:** sniper ops; sales of weapons; money laundering

Link labels
- **Source:** types of messages
- **Example:** transfer of information; action request; synchronization; etc.

Figure 3: Problem Setup

Formally, the NetSTAR model is the following. We represent a *hypothesized organization* as a graph $G_M = (V_M, E_M)$, – a **model network** where V_M is a set of C2 and resource nodes and E_M is a set of edges among them. Without loss of generality we assume that we deal with a single network structure of the enemy organization. The edges can also be expressed in the form of an adjacency matrix: $M = \|M_{\alpha,\beta}\|$, where $M_{\alpha,\beta} = 1$ if and only if $(\alpha, \beta) \in E_M$. Observed data is aggregated to a **data network** – a graph $G_D = (V_D, E_D)$ with adjacency matrix $D = \|D_{\alpha,\beta}\|$. Here, V_D is a set of observed individuals and resources, and E_D is a set of observed relationships among them. We need to discover the ***mapping*** from actors to their roles in the organization – that is, from the nodes of data graph to the nodes of model graph. This is accomplished by finding an assignment matrix $S = \|s_{a,\alpha}\|_{a\in V_D,\alpha\in V_M}$, where $s_{a\alpha} = 1$ if data node a is mapped to model node α.

In our previous research (Levchuk et al., 2006), we have presented an algorithm for finding an assignment matrix S that maximizes the likelihood function $P(G_D | G_M, S)$, which is equal to the probability that the observation (data network) has been generated by the hypothesized organization (model network) given the roles of tracked individuals (mapping between nodes of data and model graphs). In this model, the uncertainty of observing relationships between the network nodes is modeled using false alarm probability for observed, but deceptive, activities and probability of a miss for unobserved secure/covert activities. While direct optimization of the likelihood function is infeasible, an approximate solution can be obtained by relaxing a structural consistency measure to consider subgroup matches, and then employing expectation maximization algorithm to find the mapping iteratively. Not only do we obtain the correspondence of tracked individuals to specific nodes in each hypothesized organization, but we can also rank-order these associations for each organization using values of likelihood functions $P(G_D | G_M, S)$.

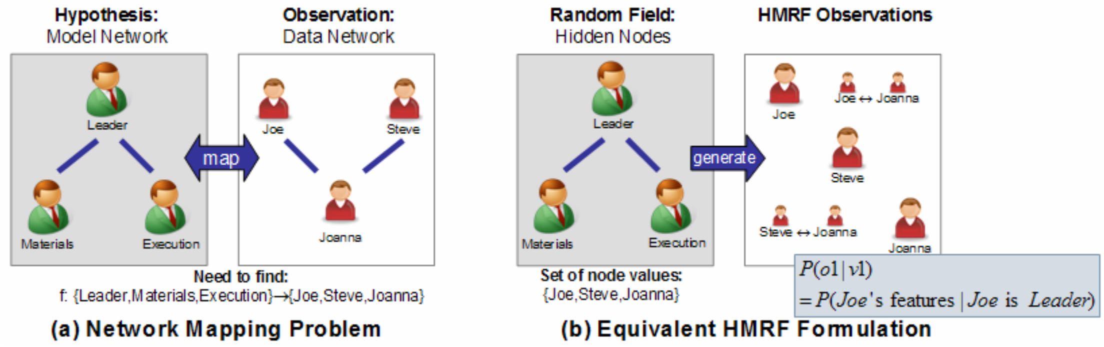


Figure 4: Application of HMRF to Network Mapping

An alternative to likelihood function estimation is to find the mapping that maximizes the a-posteriori probability $P(G_M, S | G_D)$. A corresponding maximum a-posteriori (MAP) estimator (Yu et al., 2007) is the basis for the results described in this paper. This estimator uses the Hidden Markov Random Field (HMRF) theory (Sutton and McCallum, 2006) to approximate the posterior probability using energy functions. In HMRF, a hidden field is defined by specifying the set of nodes, a finite set of values that those nodes could take, and a neighborhood structure and the concomitant probabilistic influence among nodes with Markov property. The observation about a point in the random field is obtained, and the problem is to find a node in the field that corresponds to this observation. The HMRF theory states that MAP-based solution can be found by minimizing an energy function equal to the sum of clique potentials of the HMRF graph.

Figure 4 shows how the HMRF is applied to finding the mapping between the nodes of data and model networks. The graph structure of the model network is used to construct the neighborhood structure of the HMRF graph. Each node of the model network is thus a node in the random field, and these nodes can take values from a set of names of tracked actors. The attributes of the observed network nodes and links then constitute the observation for HMRF. The outcome of the HMRF is equivalent to a model-to-data node mapping, because HMRF produces a point in the random field that assigns the names of tracked actors to the nodes in the model network.

3. Example of the Problem: How to Quantify Organizational Networks?

The NetSTAR system evaluation leveraged many years of similar model-based experimentation cycles executed for the Adaptive Architectures for Command and Control (A2C2) research program (Diedrich et al., 2003; Entin et al., 2003, 2004; Levchuk et al., 2003; Kleinman et al., 2003). This work studied the ability to use models to develop optimized military organizational structures for different missions and to encourage organizational adaptation. The A2C2 program included iterative cycles of human-in-the-loop (HIL) experimentation to evaluate and validate different command and control team structures. The A2C2 experiments have catalogued a diverse set of outcomes from HIL runs for various organizations and mission conditions. For each HIL run from an A2C2 experiment, the data logs have been captured which include task execution logs (who does what, where, and when) and the communication interactions among team players. The latter information has been coded into distinct categories corresponding to several types of formal and informal interactions in a C2 organization.

For our validation, we have inverted the problem to study the ability to recognize the U.S. military command and control Joint Task Force organization. The JTFs from A2C2 experiments under consideration consisted of 6 commanders, 8 regional leaders representing the commanders of ship platforms and bases, and 62 field assets (including helicopters, boats, special forces, UAVs, and weapon systems). The scenarios contained 36 classes of operations (including search and rescue, seize and capture ops, mine clearing, SAM sites, etc.) and a total of over 100 engagements. To geo-locate the engagements and define the areas of responsibility of commanders and leaders, we defined 6 geographical areas. The interactions among human and simulated entities have been captured from interface logs, voice communications among commanders have been recorded and manually tagged with one of the 12 communication categories (Entin, Diedrich, and Rubineau, 2003), and the links among simulated entities have been defined from simultaneous engagement in attacks on the same target (similarly to activity templates used currently in information operations).

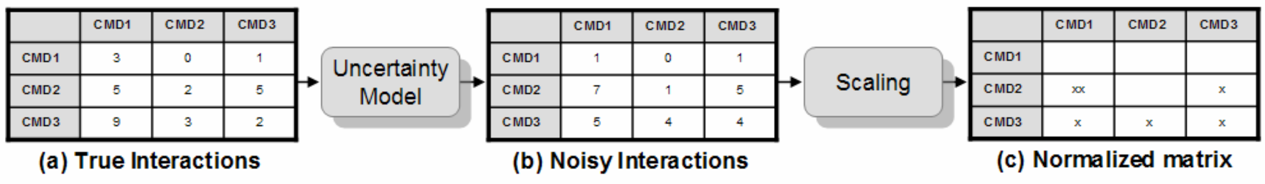
Table I: Messages extracted from A2C2 Experiments

Matrix ID	From-To	Message Types	Events from A2C2 scenarios	Role in modeling (attribute of)
1	Commander-Commander	Command	Voice communications	Link
2	Commander-Leader	Command	Launch messages	Link
3	Commander-Commander	Coordination	Voice communications	Link
4	Asset-Asset	Coordination	Attacks on same target	N/A
5	Commander-Asset	Control	Attack/Detect log messages	Node
6	Leader-Asset	Control	Attack/Detect log messages	Node
7	Commander-Area	Responsibility	Target selection and information request	Node
8	Commander-Task	Responsibility	Target selection and information request	Node
9	Asset-Task	Responsibility	Attack/Detect log messages	N/A
10	Asset-Area	Responsibility	Target selection and information request	N/A

Various events and activity logs have been translated into messages of four major types: command, control, coordination, responsibility. Table I explains how these messages have been generated. For each HIL simulation, based on the main four message types, we have constructed on the order of 1000-4000 messages and organized them into a set of 10 matrices, each corresponding to messages among specific entity classes. These matrices were specified using the counts of corresponding messages. To further simplify the problem, we defined the network nodes as consisting of commanders and regional leaders (14 in total). Nodes had attribute vectors assigned to them which quantified the amount of messages of the class corresponding to the attribute type. Commander nodes had asset control, task responsibility, and area of responsibility messages as attributes, while leader nodes had only asset control messages as attributes.

Links among commanders had command and coordination messages as attributes, while links between commanders and leaders had command messages used as attributes. As a result, only 7 out of 10 matrices have been presented to analysts and the NetSTAR model.

The baseline 7-matrix data has been stored and normalized, and both a 2-person human analyst team and the automated NetSTAR model received a noisy version of it (Figure 5) as a set of observations with deceptive messages (which did not exist in the original matrices), missing data (obtained by deleting messages from matrices), and noise and errors to other data elements (obtained by removing the entries in the matrix and randomly moving it to another entry of one or the other matrix). For the human team, we have generated network drawings based on the message matrices.



	CMD1	CMD2	CMD3
CMD1	3	0	1
CMD2	5	2	5
CMD3	9	3	2

(a) True Interactions

	CMD1	CMD2	CMD3
CMD1	1	0	1
CMD2	7	1	5
CMD3	5	4	4

(b) Noisy Interactions

	CMD1	CMD2	CMD3
CMD1			
CMD2	xx		x
CMD3	x	x	x

(c) Normalized matrix

Figure 5: Example of Data Setup

The outcomes of human analyst team and automated identification model were then compared to judge the benefits of NetSTAR in terms of identification accuracy (of organization identification and node mapping) and the time required to identify the adversary (which could be equated to manpower needs).

4. NetSTAR Validation: Computational Experiment

A set of 7 sample datasets from HIL runs for different organizational C2 structures has been selected for testing. In order to properly evaluate the NetSTAR benefits, we needed to answer the following two questions:

(1) Is it possible to judge the impact of *uncertainty* on the quality of the organization identification and node mapping solution?

(2) Is it possible to judge the impact of *problem domain* and *complexity* on the quality of the organization identification solution?

To address the first question, our study included exploring various levels of uncertainty in the data. To address the second question, we conducted comparisons according to the type of organization that needs to be recognized. Different types of information are needed to recognize different types of organizations. In our pilot studies, we found that when the low-noise commander-to-subordinate intercepts can be obtained, a functional organization, where a single commander controls resources of the same type distinct from other commanders, is easier to recognize than a divisional organization, where each commander controls a variety of resources but has similar capabilities to other commanders. The divisional organization, which is similar in nature to militia organizations and is a standard for current U.S. Army force structuring, is more complex than the functional organization (which is specialization-based and a doctrinal organization for U.S. Navy’s composite warfare command) in terms of resource control, but can be easily recognized given the low-noise data of commanders’ activity locations, since commanders’ geographic responsibilities in divisional organization are distinct. Both functional and divisional organizations have elements that are encountered in today’s command and control teams, and thus a study of such “hybrid” teams is essential to explore how difficult it is for human analysts to use multiple types of information for C2 organization discovery.

To simplify the analysis, the uncertainty in data was controlled using two parameters: (i) % of missing data (p_m), and (ii) % of deceptions or false alarms (p_f). The approximate signal-to-noise ratio was then found as $\frac{1-p_m}{p_f}$. More details about human experiment and results are presented in the companion part 2 paper. Here, we describe the results of the computational experiment that compared the impact of seven levels of uncertainty (Table II).

Table II: Levels of Data Uncertainty for NetSTAR Computational Experiments

Uncertainty Level	Low-1	Med-1	Med-2	Med-3	High-1	High-2	High-4
% missing data	10	30	40	50	55	60	70
% deceptions/errors	10	20	30	30	30	35	45
SNR = true messages/deceptive messages	9	3.5	2	1.6666667	1.5	1.1428571	0.6666667

For each of the uncertainty levels and the baseline organization, we have conducted 5 Monte-Carlo runs and tested 7 data sets distinguished by the underlying true C2 organization. As the result, we obtained 5 data points for each pair of true organization and data uncertainty level. We have compared the results to the accuracy of solution provided by human analysts and obtained during project’s table-top exercises (for more details – see part II of this paper (Entin et al., 2007)).

To answer the question about sensitivity of identification to uncertainty, we analyzed solutions for a single true organization over three levels of uncertainty – Low-1, Med-1, and Med-2. We selected organization D2 out of the 7 organizational networks from A2C2 experiments because it exhibited the characteristics of all other organizations in terms of resource control allocation and roles of command nodes. Figure 6 shows this analysis, with comparison of NetSTAR algorithms to human analysts and random solution in both organizational identification accuracy (ability to identify correct acting organization) and actor mapping accuracy (ability to correctly recognize individual actors in the organization). NetSTAR provided on average over 2.5X better accuracy in both organization identification and actor mapping compared to human analysts, and results were statistically significant ($p<0.001$). Also, results suggest that human analysts have performed well and statistically better than random solution in all but higher (Med-2) noise level.

To answer the second question about the impact of domain and complexity of baseline solution, we have analyzed separately the accuracy of predicting specific organizational forms. We have fixed the noise level at 30% of missing data and 20% of deceptions (Med-1) and varied the types of acting organizations in the data set. Three organizations have been used to compare accuracy of NetSTAR algorithms and human analysts. The results are depicted in Figure 7.

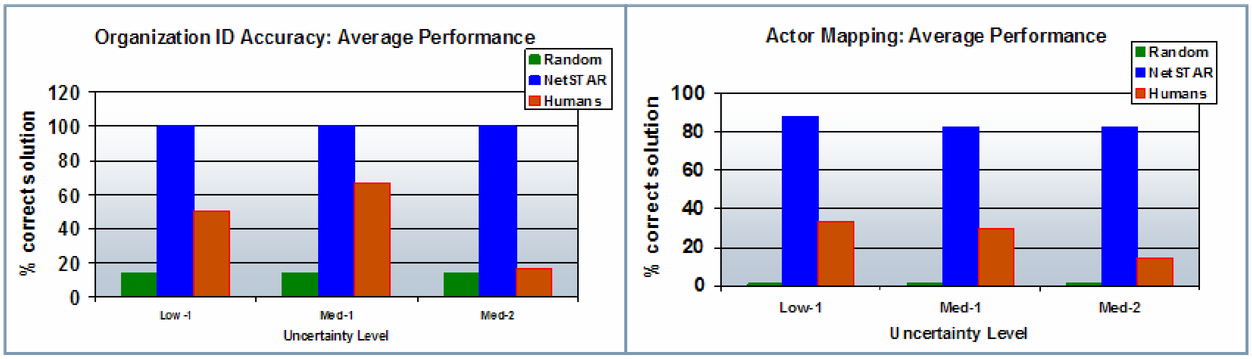


Figure 6: Sensitivity to Uncertainty for Fixed Organization Baseline = D2
(Random = results of random identification; NetSTAR = results of automated algorithm identification; Humans = results of threat identification by human analysts during table-top experiment)

Average results of detailed sensitivity analysis of NetSTAR algorithms are shown in Figures 8-9. We can see that NetSTAR provides >2.5X better detection than human analysts under same uncertainty level, and NetSTAR achieves the same performance as human analysts under 3X higher uncertainty level. NetSTAR algorithm also provided a robust solution by being able to correctly identify 70% of actor-role mapping for 50% of missing data and 30% detection rate.

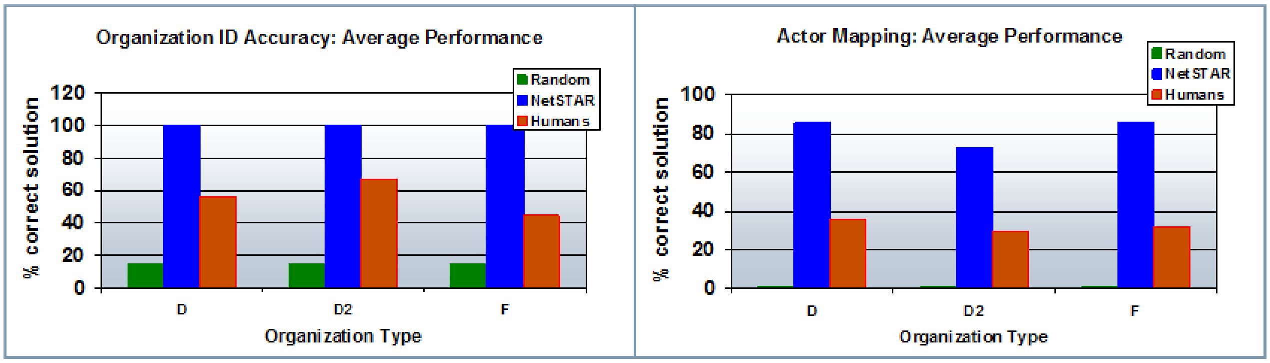


Figure 7: Sensitivity to Organization Type for Fixed Noise Level = Med-1
(D=Divisional; D2=Hybrid; F=Functional)

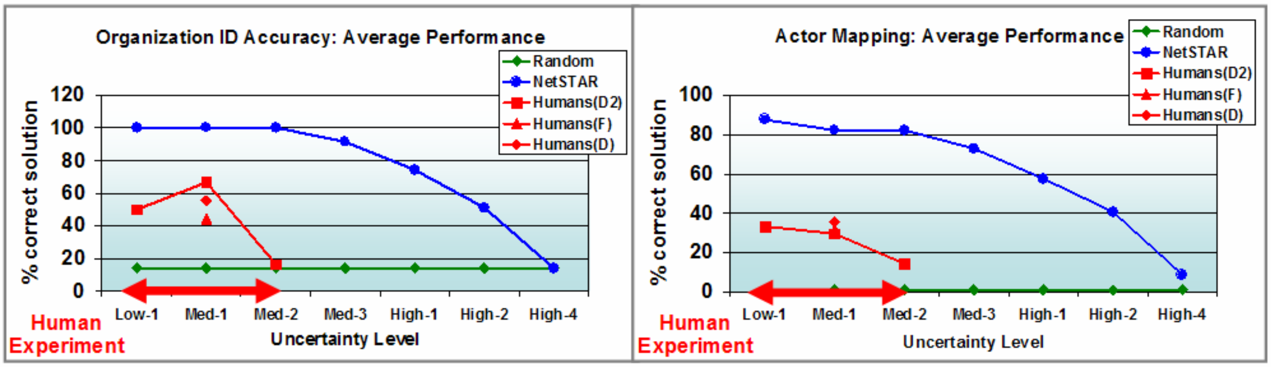


Figure 8: Sensitivity Analysis – Average Performance Results
(Random = results of random identification; NetSTAR = results of automated algorithm identification; Humans = results of threat identification by human analysts during table-top experiment)

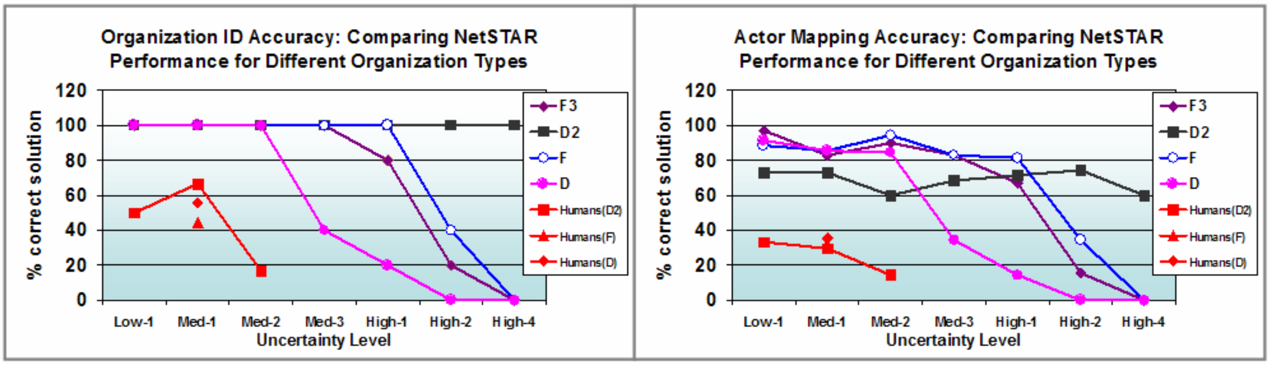


Figure 9: Sensitivity Analysis – Performance Results for specific Baselines (F=Functional; D=Divisional; D2=Hybrid; F3=Functional alternative)

Figure 9 shows the organization identification and node mapping accuracy for 4 out of 7 baseline cases. We have observed that there where no specific patterns in the data except for recognition of D2 (hybrid) organizational structures. Further analysis of the distance between organizational hypotheses networks revealed that this structure was significantly benefiting from the hypotheses set: all other model networks were far away from it (Figure 10). The distance is found using the a-posteriori energy function for a situation without uncertainty. We then concluded that NetSTAR’s performance is affected only by distinguishability of organizational hypotheses and not by experience biases. Since some hybrid organizations exhibit unique structural patterns, these patterns would improve the accuracy of NetSTAR in detecting non-traditional adversaries.

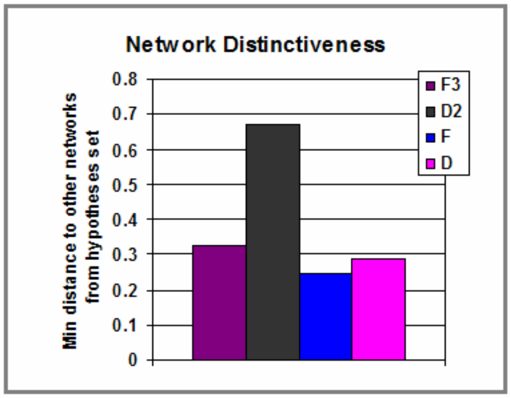


Figure 10: Minimum Distance from given Organization to Alternative Hypotheses Networks

5. Conclusions and Next Steps: Guided Information Collection for Improved Identification

In this paper, we have presented the NetSTAR system that can help information analysts to deal with complexity of adversarial characterization problem and improve the accuracy of their decisions. The experiments conducted during the NetSTAR program showed significant improvements that this technology can bring compared to unaided analysts. NetSTAR outperformed humans in both accuracy of organization identification and actor role identification, and was able to handle higher levels of uncertainty than unaided human analysts. Used as decision support system, NetSTAR promises to result in significant manpower, decision time and error reductions during threat analyses tasks.

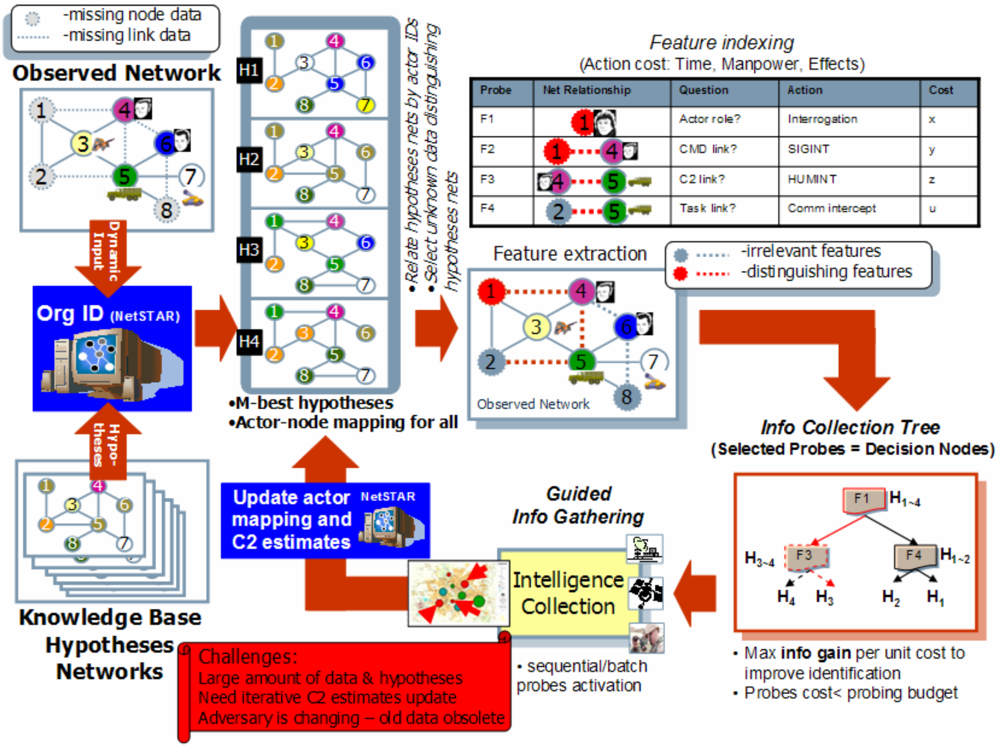
Current information operations and adversarial analysis require continuous situation monitoring and assessment. NetSTAR system is well-suited to handle online data collection due to the iterative nature of its algorithms and its ability to define how information elements can improve the ambiguity of current predictions. When additional information collection is possible, the ability to prioritize and plan these activities might be needed, especially when the data collection resources (sensors, human collection teams, reconnaissance units, interrogation facilities) are limited and the impact of collection efforts needs to be taken into account. Our core hypotheses-testing network identification approach can be extended to conduct cost-effective intelligence gathering to achieve maximum identifiability of the enemy network over time. The approach uses current network hypothesis ranking to come up with the most important missing information elements (features) that would facilitate the largest reduction in the ambiguity of organization identification. The data collection plan is then developed by ordering the data collection efforts for feature exploration in a collection tree. The construction uses the constraints on information collection resources and aims at maximizing the information gain from data collection efforts. As illustrated in Figure 11, intelligence collection planning can be integrated with network mapping and has the following steps:

Step 1 (Network Mapping Output): The data network derived is matched against all hypothesis networks. The most probable hypotheses can then be further explored. Each node from the hypothesized network is mapped to a data network node, or in other words is given an ID. This allows relating hypotheses to each other.

Step 2 (Feature Extraction): We compare the hypothesis networks with the current data network to see which of the potential data elements (attributes of nodes and links) are missing. The resulting data elements are then classified as relevant (ones that distinguish the hypotheses) and irrelevant. The latter subset is discarded, and the former is categorized according to the action that needs to be performed to collect the data element, the cost of this action (e.g., in terms of time, money, effects of the action such as making the enemy suspicious of our presence, etc.), the question it is supposed to answer, and the corresponding network relationship that it will resolve. Each action-data element is then called a “probe” to indicate that it is trying to explore the existence of a specific attribute and its value through active search in the environment.

Step 3 (Intel Plan Design): An intelligence collection plan is constructed in the form of a decision tree. The nodes of this tree correspond to probes to be conducted to reduce the ambiguity of predictions. Branches from a decision node correspond to feasible probe outcomes and lead to nodes defining the next probing actions. Organization of the probes into a data collection tree is based on maximizing the anticipated information gain which contributes to our ability to distinguish among hypotheses. The probes can then be merged together for integrated intelligence collection actions.

Step 4 (Guided Intel Collection): Guided intelligence collection is a dynamic execution of the intelligence collection plan, and is performed by taking actions based on a decision tree. After performing an action from a tree node, the process moves to the next node in the tree along the branch corresponding to the selected action’s outcome. At every step, the ranking of hypotheses is updated, which may result in a reduction of feasible hypotheses list. Guided intelligence collection can be conducted either sequentially or in a batch mode based on the availability of the budget (time, cost, resources, etc.).


-missing node data
-missing link data
Observed Network
Dynamic Input
Org ID (NetSTAR)
Hypo-theses
Knowledge Base
Hypotheses
Networks
H1
H2
H3
H4
•M-best hypotheses
•Actor-node mapping for all
•Relate hypotheses nets by actor IDs
•Select unknown data distinguishing hypotheses nets
Feature indexing
(Action cost: Time, Manpower, Effects)
Feature extraction
-irrelevant features
-distinguishing features
Observed Network
Info Collection Tree
(Selected Probes = Decision Nodes)
F1
$H_{1\sim4}$
F3
$H_{3\sim4}$
F4
$H_{1\sim2}$
H_4
H_3
H_2
H_1
• Max **info gain** per unit cost to improve identification
• Probes cost< probing budget
Guided
Info Gathering
Intelligence Collection
• sequential/batch probes activation
Update actor mapping and C2 estimates
NetSTAR
Challenges:
Large amount of data & hypotheses
Need iterative C2 estimates update
Adversary is changing – old data obsolete

Probe	Net Relationship	Question	Action	Cost
F1	1	Actor role?	Interrogation	x
F2	1 ···· 4	CMD link?	SIGINT	y
F3	4 ···· 5	C2 link?	HUMINT	z
F4	2 ···· 5	Task link?	Comm intercept	u

Figure 11: Guided Information Collection Process

The outcome of adversarial predictions would allow analysts to define the high-value targets and conduct counter-actions resulting in better effects against the adversaries. Automated tools are therefore needed that incorporate the knowledge of enemy C2 networks and mission into assessing the vulnerabilities of an adversarial organization and finding the impact of BLUE’s actions against a partially identified RED side. Since adversarial analysis often produces multiple predictions of similar rank about the adversary, effective vulnerability and impact assessment models should rely on stochastic and robust approaches. Our current research is focused on addressing these challenges.

References:

Alberts, D.S., and R.E. Hayes, *Understanding Command and Control*, Washington, DC: CCRP Publications, 2006

D.L. Kleinman, G.M. Levchuk, S.G. Hutchins, and W.G. Kemple (2003),“Scenario Design for the Empirical Testing of Organizational Congruence”, *Proceedings of the 2003 International Command and Control Research and Technology Symposium*, Washington, DC, June.

G.M. Levchuk, K. Chopra (2005), “NetSTAR: Identification of Network Structure, Tasks, Activities, and Roles from Communications”, *Proceedings of the 10th International Command and Control Research and Technology Symposium*, McLean, VA, June.

G. Levchuk, Y. Levchuk, and K. Pattipati, “Identifying Command, Control and Communication Networks from Interactions and Activities Observations”, *Command and Control Research and Technology Symposium*, 2006, San Diego, CA.

McNamara, C. *Field Guide to Consulting and Organizational Development: A Collaborative and Systems Approach to Performance, Change and Learning*, Authenticity Consulting, LLC, February, 2005

Wade, N.M., The Battle Staff SMARTbook: Doctrinal Guide to Military Decision Making and Tactical Operations, 2nd Edition, Lakeland: The Lightning Press, 2005

FM 3-13, “Information Operations: Doctrine, Tactics, Techniques, and Procedures”, *Headquarters, Department of the Army*, Washington, DC, 28 November 2003

Grishman, R., “Information extraction.” In R. Mitov (ed.), *The Oxford handbook of computational linguistics*, pp. 545-759, New York, NY: Oxford University Press, 2003

Stolfo, S.J., et al., “Behavior Profiling of Email,” *Proceedings of NSF/NIJ Symposium on Intelligence & Security Informatics*, 2003

Miller, D., S. Boisen, R. Schwartz, R. Stone, R. Weischedel, “Named entity extraction from noisy input: speech and OCR”, *Proceedings of the sixth conference on Applied natural language processing*, Seattle, Washington, 2000, pp. 316-324

Krebs, V.E., “Mapping Networks of Terrorist Cells”, *Connections*, Vol. 24, 2001, pp. 43–52

Sageman, M., *Understanding Terror Networks*, Philadelphia, PA: University of Pennsylvania Press, 2004

Skillicorn, D., “Social Network Analyses via Matrix Decompositions: al Qaeda”, *Report*, available from http://www.cs.queensu.ca/home/skill/alqaeda.pdf, Aug., 2004

Van Meeter, K.M., “Terrorists/Liberators: Researching and Dealing with Adversary Social Networks”, *Connections*, Vol. 24, 2001, pp. 66-78

Dombroski, M.J., and K.M. Carley, “NETEST: Estimating a Terrorist Network's Structure”, *CASOS 2002 Conference*, No. 8, 2002, pp. 235-241

Dombroski, M.J., P. Fischbeck, and K.M. Carley, “Estimating the Shape of Covert Networks”, *Proceedings of the 8th International Command and Control Research and Technology Symposium*, National Defense War College, Washington, DC, 2003

Yu, F., G. Levchuk, K. Pattipati, and F. Tu, “A Probabilistic Computational Model for Identifying Organizational Structures from Uncertain Message Data”, submitted to *IEEE Transactions on SMC*, 2007.

Sutton, C., and McCallum, A., “An Introduction to Conditional Random Fields for Relational Learning”, In *Introduction to Statistical Relational Learning*, Edited by Lise Getoor and Ben Taskar, MIT Press, 2006

Diedrich, F.J., et al., “When do Organizations Need to Change (Part I): Coping with Incongruence”, *Proceedings of the 2003 International Command and Control Research and Technology Symposium*, Washington, DC, 2003

Entin, E. E., et al., “When do Organizations Need to Change (Part II): Incongruence in Action”, *Proceedings of the 2003 International Command and Control Research and Technology Symposium*, Washington, DC, 2003

Entin, E. E., et al., “Inducing Adaptation in Organizations: Concept and Experiment Design”, *Proceedings of the 2004 Command and Control Research and Technology Symposium*, San Diego, CA, 2004

Levchuk, G.M., et al., “Congruence of Human Organizations and Missions: Theory versus Data”, *Proceedings of the 2003 International Command and Control Research and Technology Symposium*, Washington, DC, June, 2003

Kleinman, D.L., P. Young, and G.S. Higgins, “The DDD-III: A Tool for Empirical Research in Adaptive Organizations”, *Proceedings of the 1996 Command and Control Research and Technology Symposium*, Monterey, CA, June, 1996

Entin, E.E., F.J. Diedrich, and B. Rubineau, “Adaptive Communication Patterns in Different Organizational Structures”, *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, Denver, CO, 2003

Entin, E., R. Greer, T. Jefferson, and G. Levchuk, “Identifying the Enemy – Part II: Algorithms versus Human Analysts”, to appear in *Proceedings of the Command and Control Research and Technology Symposium*, New Port, RI, 2007.