Title: **Identifying the Enemy – Part II: Algorithms versus Human Analysts**

Suggested Tracks: 
*Information Operations/Assurance, C² Modeling and Simulation*

Authors:

**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

**Rebecca Grier**  
Aptima Inc.,  
1726 M Street, N.W., Suite 900  
Washington, DC 20036  
Phone: (202) 842-1548  
Fax: (202) 842-2630  
e-mail: rgrier@aptima.com

**Tyrone Jefferson**  
Aptima Inc.,  
1726 M Street, N.W., Suite 900  
Washington, DC 20036  
Phone: (202) 842-1548  
Fax: (202) 842-2630  
e-mail: tjefferson@aptima.com

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

Correspondence: 

**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
Abstract
This paper is Part II of a 2-paper submission. Part I discusses the automated network identification model and focuses on the problem setup and provide the description of the computational algorithms at the core of the technology system called NetSTAR. In Part II we describe how our identification of adversarial organizations stem from our analysis of command and control (C^2) organizations and our analysis of what a model/algorithm must accomplish to identify and describe an adversarial organization. We then summarize the human lab experimentation and concomitant comparison of the accuracy of adversarial organization discovery obtained by a team of human analysts versus the automated C^2 identification process. Briefly, the findings show that the unaided human can perform organization identification at level far better than chance however the NetSTAR system is 100% better than the humans at this task. When the humans attempt to describe and map attributes of the identified organization they faired poorly, however the NetSTAR system showed little diminution in performance even under noise level (amount of error present) that severely decreased human’s performance. The implications of these findings are addressed in this paper.

Introduction
Knowledge of an enemy’s organization, objectives, and procedures are essential to successfully anticipate enemy actions, identify high-value targets, and develop effective countermeasures. Current approaches to threat analysis are labor intensive. Intelligence analysts have to deal with and filter through huge amounts of data, most of which has little to do with the specific threat. Information gaps abound complicated by missing data, deceptions, and errors, that must be addressed and all too often analysts only have their experience to fill the gaps which may or may not be up to the task. Moreover, such actions can introduce decision biases such as the confirmatory biases where the first seemingly valid hypothesis is accepted and used throughout the rest of the analysis relatively unchecked (Adelman, Tolcott, and Bresnick, 1993). This issue is compounded by the huge amounts of data and complexity of the problem analysts need to analyze, influencing what data is used and which is filtered out and never studied. All these factors can negatively impact the ability of the intelligence team to recognize enemy activities and further results in decreased efficiency of countermeasures leading to unintended consequences.

Presently, only a limited set of tools are available to intelligence operators to analyze, correlate and visualize the data. No tools have been developed with automated threat prediction and assessment capabilities that can reason from multi-source data and support the decisions about the enemy’s command and control organization. This was due in part to the inability to bring all data sources together for common analysis. However, new tools and data collection techniques, as well as, new technologies to automate threat prediction are in development and this paper is focused on one such tool and technology.

The Problem and C^2 Structure
This paper is Part II of a 2-paper submission describing a project to develop and validate the NetSTAR technology for automated threat identification. NetSTAR will perform a kind of reverse engineering. Based on detected enemy actors, activities, and communications the NetSTAR technology will attempt to describe the organizational structure that produced these detections and map the actors into appropriate nodes of the organization’s command and communication hierarchies. Below we describe how our identification of adversarial organizations stem from our analysis of command and control (C^2) organizations and our analysis of what a model/algorithm must accomplish to identify and describe an adversarial organization. We then summarize the human laboratory experiment and concomitant comparison of the accuracy of adversarial organization discovery obtained by teams of human analysts versus the automated C^2 NetSTAR identification process.

The threat analysis is based on understanding the decision-making processes in the general C^2 organization. C^2 refers to procedures used to effectively organize and direct armed forces to accomplish a mission. The
**command** function is oftentimes referred to as an art of an individual to set the initial conditions and providing the overall intent for mission execution (Alberts and Hays, 2006). The **control** is referred to as those structures and processes devised by command to enable it and to manage risk and other entities in the organization. The commander in a C² organization issues instructions to subordinates, makes suggestions to commanders of adjacent units, and makes requests from and reports to supporting units and superiors. He develops and maintains situational awareness of his area of operations through reports presented by other people or by electronic systems (Coakley, 1991). The basic premise of C² organizations is the ability to distribute the responsibilities among its elements and coordinate these seemingly independent entities for joint operations to achieve objectives (Alberts and Hays, 2006). The fundamental need for communications significantly constrains the options for C² making the communications infrastructure a critical feature of a C³ system. However, describing the communications links and nodes of a fighting force does not suffice to explain, understand, or predict successes and failures in C² organizations. We need to be able to represent, model, and identify the functions and objectives of the individual elements of the C² organization.

As illustrated in Fig. 1 a C² **organization** is a collection of C² nodes and resources connected via command, control, communication, and task structures. The roles, responsibilities, and relationships among C² nodes and resources constrain how the organization is able to operate. C² nodes are entities with information–processing, decision–making, and operational capabilities that can control the necessary units and resources to execute mission tasks, provided that such an execution does not violate the concomitant capability thresholds. C² node can represent a single commander, liaison officer, system operator, or a command cell with its staff. A set of physical platforms and assets, C² nodes, and/or personnel can be aggregated to form a resource (e.g., squad, platoon, weapons system, etc.). A resource is considered a physical asset of an organization that provides resource capabilities and is used to execute tasks. The level of aggregation depends on the problem at hand. For example, in cordon and search missions executed by a company–size force, the squads are the resources. The roles and responsibilities of the C² nodes and resources identify possible operational and tactical policies: decisions they can make and actions they can perform.

**Figure 1:** Example of C² Organization

Command structure, represented as a network with directed links, defines superior–subordinate relationships among C² nodes of the organization, thus specifying who can send commands to whom. Communication structure is a network between the decision makers of the organization, that defines “who can talk to whom”, the information flow in the C² organization, the communication resources that decision-makers can use (communication channels), as well as the security of the communication channels. A control structure is an assignment of resources to C² nodes, and specifies which commanders can send tasking.
orders to what assets. A task structure is a network among resources, where each link corresponds to operations jointly executed by these resources.

Fig. 1 depicts an example of an enemy C2 military team consisting of 5 command elements and 14 units/resources. The commanders of this organization make decisions to manage assigned resources in a cooperative manner to achieve team objectives. Commanders are executing mission tasks and prosecuting the desired targets via allocating their resources (military assets and weapons) and synchronizing their mission task execution and target engagements. Fig. 1 also describes the set of resources – military units and assets controlled by commanders. The assets include bomb making teams, sniper teams, mortar units, intelligence and reconnaissance teams, and trucks. This figure shows as well the functional or resource capabilities (Levchuk et al., 2002) of the units and resources in terms of bomb making, strike and small-arms attack, intelligence and monitoring, and transportation. As shown if Fig. 1b the authority structure among the 5 commanders is a flat hierarchy with a single commander (“BLACK”) being a main commander of enemy forces. The assignment of assets and units to commanders depicted in Fig. 1;c determines the control structure of the C2 organization. Note that in the hypothetical example of Fig. 1 the main commander (“BLACK”) does not control any resources directly. The communication structure (who can talk to whom) of the organization is illustrated in Fig. 1.d along with the direction of unit reporting observed events (information flow) beyond the control structure (we assume that units controlled by commanders also report their observations to these commanders). A partial task structure – a network between resources – is shown in Fig. 1.e. The task structure is due to the joint task execution by resources; therefore, it evolves throughout mission execution and depends on how the commanders manage their resources to assign and execute tasks.

The meaning of organizational discovery is the ability to recognize the C2, communication, and task structures of the organization. However, the challenge is that most of the time we cannot observe the elements of the structures of the organization. Instead, we can obtain the intelligence due to the actions and activities of the organization. The specific actions depend on the structure of enemy C2 organization and are derived from the goals of the team. Before we outline our methodology to relate the observations to the structural elements, we discuss the structure of the observational data available from intelligence gathering sources.

For threat analysis, we assume that the intelligence (observations, or data) given to us as shown in Fig 2 includes the set of tracked (monitored) individuals whose positions in the organization we need to determine, the resources of the enemy (including physical military, economic, and political resources), information about individuals (such as attributes of the individuals and resources – e.g., expertise of individuals, training, background, affiliation, family ties, roles and responsibilities, etc.), and information about transactions that involves these entities – communications among individuals (including some knowledge of its content), involvement in activities (such as individuals committing the same crime, or meeting among each other, or performing financial or business transactions, or using the resources in covert or open operations). The outcome of threat analysis is the prediction of the adversary’s organization – that is, the roles and responsibilities of individuals, and their command, communication, control, and information networks.
The NetSTAR Model

A more detailed discussion of the NetSTAR technology can be found in Part I of this 2-part presentation (Levchuk et al., 2007) – we summarize here. NetSTAR is a hybrid model-based structure and process identification methodology, developed to automate the identification of the acting organizational network and facilitate validation of network hypotheses developed by analysts during adversary analyses (Levchuk et al., 2005, 2006). NetSTAR performs network state/pattern recognition from multi-source uncertain data based on probabilistic attributed graph matching principles. The outcome of this process is finding the mapping between nodes of the observed graph and library graphs and rank-orders the library graphs in terms of their likelihood (probability that the observed data was generated by the library network). The node mapping corresponds to finding the roles of the observed nodes (actors, individuals, cells, resources) and mapping the command, control, communication, information and task networks of the enemy organization.

The graph matching problem has many complicating aspects. First, there exist many mappings from individuals/actors to command nodes (there are N*M mappings from N actors to M command nodes). Second, we need to explore many different hypotheses about enemy organization – that is, many organizational structures. Third, even if the organization is known, we still need to determine what goals/mission it has, and how far along this organization is in finishing the mission. Other issues, such as transcribing the communications to identify the content, constructing feasible organization and mission representations, and determining the most efficient intervention strategies must be addressed.

The NetSTAR’s automated $C^2$ identification process is aimed at reducing the complexity of organizational discovery. This will allow analysts to focus on information most essential for decision making and explore in detail only a limited number of most likely hypotheses. The goal of this evaluation is to assess the NetSTAR’s capability and evaluate whether the solutions produced can significantly increase capabilities to make inferences regarding enemy command structures and explore how discovered information can be used by friendly forces to disrupt adversarial activities.

Evaluation Strategy

Our evaluation strategy schematized in Fig. 3 leverages many years of model-based experimentation conducted under the Adaptive Architectures for Command and Control (A2C2) research program (Diedrich et al., 2003; Entin et al, 2003; Kleinman et al, 2003; and Levchuk et al, 2003).
This work studied the capability of modeling to develop optimized military organizational structures for different missions and to encourage organizational adaptation. The A2C2 program involved iterative cycles of experimentation to evaluate and validate the modeling approaches. These experiments were conducted using Distributed Dynamic Decisionmaking (DDD) medium fidelity simulation (Kleinman, Young, and Higgins, 1996). DDD is a distributed real-time simulation platform implementing a complex synthetic team task that includes many of the behaviors at the core of almost any C\textsuperscript{2} team: assessing the situation, planning response actions, gathering information, sharing and transferring information, allocating resources to accomplish tasks, coordinating actions, and sharing or transferring resources. Successive DDD generations have demonstrated the paradigm’s flexibility in reflecting different domains and scenarios to study realistic and complex team decision-making. An outcome of A2C2 program that directly feeds our validation strategy was the creation of DDD-based scenarios and organizational structures. The A2C2 experiments have catalogued a diverse set of outcomes for various teams, organizations, and mission conditions.

A DDD experiment trial includes a team of participants playing roles of commanders in a predefined C\textsuperscript{2} team and performing the mission tasks in the DDD simulation environment using kinetic and non-kinetic assets/resources. Of particular interest to our validation approach are A2C2 experiments with Joint Task Force organizations, which explored the range of possibilities to assign the C\textsuperscript{2} relationships, resource ownership, and individual responsibilities among commanders. Under the A2C2 program both traditional and non-traditional C\textsuperscript{2} structures were tested, thus providing rich data for the validation experimentation. Each A2C2 experiment trial produces a history file which includes task execution logs (who does what, with what, where, and when) and the communication interactions among team players. The latter information was coded into distinct categories corresponding to several types of formal and informal interactions in a C\textsuperscript{2} organization. This data was directly used by our validation strategy, graphically outlined in Fig. 4, with the addition of the uncertainty model component that can take the task execution and communication logs from the experiment trials and make the data noisy. That is, introduce deceptive events (false alarms), create missing data (e.g., misdetections), and add noise and errors to other data elements. The introduction of noise into the data produced a realistic analog to the data intelligence analysts must deal with.

The validation effort had two distinct parts: an experiment that observed the NetSTAR tool process the noisy observation data and an experiment that observed teams of human analysts process the same noisy observational data. In both experiments the goal was the same - reconstruct the acting enemy’s C\textsuperscript{2} organizational structure. The results generated by the two experiments created a third effort: a comparison...
between the human analyst teams’ performance and the performance of the NetSTAR algorithms to evaluate the value added by the NetSTAR technology. The comparison is a major focus of this paper.

**Method for the Experiment Involving Teams of Human Analysts**

A detailed reporting of the methodology and results of the experiment involving teams of human analysts can be found in Entin et al. (2007). A brief summary will suffice here.

The experiment was conducted at the Naval Postgraduate School allowing us to draw a sample of 18 active duty officers to serve as participants. The sample was organized into nine 2-person teams. The experiment the participants were engaged in involved the manipulation of two independent variables: organization type and fogging level (i.e., percentage of noise or error). Organization type was operationalized by varying organizational structures along a continuum, ranging from functional to divisional organizations. Following Diedrich et al. (2003) the functional organizational structure was organized such that each commander specialized in one or two aspects of a mission such as Strike or Air Warfare, where the specific assets controlled were distributed across multiple platforms (ships). In contrast, in the divisional organizational structure, each commander had control over a single multifunctional platform that was able to process a variety of functional tasks in a given location. An intermediate, or hybrid, organizational structure was a type of organization in which some commanders controlled assets functionally and other commanders comprising the team controlled assets divisionally. Stimuli data representing three types of organizations were presented to the participants: Functional, Divisional and a Hybrid. For the experiment, the hybrid organizational structure was defined as a structure where four commanders comprising the six person team controlled assets divisionally while the remaining two controlled assets functionally.

The second independent variable, fogging level, referred to the amount of noise or error injected into the tables and illustrations describing an organizational structure. Using the uncertainty model component three levels of fogging were produced: one with 10% noise or errors, one with 30% noise or errors, and one with 50% noise or errors.

Teams came to the lab for two 2-hour sessions. During each one hour trial, participants were provided one stimulus data set (i.e., the data sets to which participant teams were trying to match to the hypothesis organizational structures) and 7 hypothesis organizational structures. They were told that one of the 7 hypothesis organizational structures was a match to the stimulus data they were given and their goal was to find the matching organizational structure. They were also told that only the data in the stimulus set was noisy; the 7 hypothesis organizational structure data sets were noise free. The 7 hypothesis organizational structures ran the continuum from functional to divisional structures (i.e., one functional, one divisional, and five intermediate or hybrid structures). Description of each organizational structure was presented in nine spreadsheets and nine diagrams that, in turn, described the interaction among nodes comprising the organization. Nodes represented the items of interest in the organizations including commanders, leaders, assets, and areas. The stimulus data also consisted of nine descriptive spread sheets and nine corresponding diagrams in the same format as the hypothesis sets. Participants had 50 minutes to select the correct hypothesis organizational structure and get as far in the actor node (i.e., commanders and combatant platform owners) mapping as possible. After teams completed the first trial, the stimulus data set and response sheets were collected and participants were provided with a different stimulus data set and the process was repeated. Each team worked with four different stimulus data sets while the 7 hypothesis organizational structures were always the same.

**Contrasting Human Teams and NetSTAR Performance**

Overall, the human teams identified the organizational structure correctly in 17 of the 36 trials or 47.2% of the time (Entin et al., 2007). If just chance were operating, that is, the teams were just guessing they would have only a one in seven chance (14.3%) of selecting the correct hypothesis organization on any given trial. Clearly the human teams performed better than chance - more than 230% better than chance. The NetSTAR
tool, however, correctly identified the organizational structure 100% of the time, thus besting the human teams by more than 110% (Fisher Exact Test, p < .001). Holding the fogging (noise) level constant (at 30%) we can see from Fig. 5 that the human teams’ identification performance varied over the three organizational types. The human teams correctly identified the organizational structure 56% of the time when the organization was divisional, but the Fisher’s Exact Test showed this performance to be significantly below NetSTAR’s performance (p < .05). The human teams performed less well when the organization was functional, where correct identification of structure s only occurred on 44% of the trials, which was also significantly below the performance of the NetSTAR algorithms (p < .02). Turning to the hybrid structure we see that the human teams performed well and were correct 67% of the time which proved to be not significantly different from NetSTAR’s performance level (p > .2). Organizational structure identification performance for the hybrid structure with the lowest fogging was also good as teams were correct on 50% of the trials which once again was not significantly different from that of the NetSTAR algorithms (p > .09). When fogging was at its highest level (i.e., the 50% fogging level) the human teams did not fair well as indicated by only one correct identification and this performance was significantly below the performance of the NetSTAR tool (p < .01). Where as, organizational type and fogging level did not appear to effects the NetSTAR algorithms’ ability to correctly identify organizational structure these factors did effect the human teams’ identification performance.

![Figure 5: Organizational Structure Identification Performance For Human Teams And Netstar By Stimulus Organizational Type And Fogging Level](image)

The human analyst teams and the NetSTAR tool were also tasked to map the actor nodes (commanders and combatant platform owners). That is, to match the commanders of the selected hypothesis organizational structure correctly to the commander nodes of the stimulus organizational structure and to do this respectively for combatant platform owners’ nodes. There were 14 actor nodes - six commander and eight combatant platform owners’ - nodes to map. Fig. 6 shows the percent correct mappings across all human teams and the percent correct mappings achieved by the NetSTAR algorithms for the five stimulus organizations. Fisher Exact Tests showed that for every stimulus organization NetSTAR significantly out performed the human teams (all ps < .001). Chi squares were also computed for each of the five comparisons and the magnitude of the chi square value is an indication of the strength of the out come. The magnitude of the chi squares were largest for the hybrid-10% and hybrid-50% stimulus organizations (chi square = 84.0 and 85.7, respectively), indicating that for these two structures the human teams departed most from NetSTAR performances levels. The smallest chi square value occurred for the functional-30%
stimulus organization meaning it was on this structure the human teams came closest to the NetSTAR algorithms’ performance.

![Figure 6. Percent Correct Mapping of Commander and Leader Nodes For Human Teams And Netstar By Stimulus Organizational Type And Fogging Level](image)

**Discussion and Conclusion**

The results showed that the NetSTAR algorithm significantly outperformed the human analyst teams when it came to identifying the stimulus organization structures by a ratio of more than two to one. Moreover, the NetSTAR algorithms were unaffected by organizational type or amount of fogging when performing the identification task. This was not true for the human teams. When the stimulus organizational structure was functional or when fogging was high (i.e., 50%) team performance fell considerably. It is not hard to understand why high fogging levels diminish human performance. The noise injected by the high fogging level obscured the true structure, mislead the analysts, and increases the cognitive load for the analysts. But, explaining why the functional structure is more difficult to identify than other structures is more involved.

In their analysis of the human teams’ performance Entin et al. (2007) note that overlap among different hypothesis organizational structures used was not uniform. The functional structure used was closer to its distracters than either the divisional or hybrid structures were to its related structures. Thus, we hypothesize that is was more difficult for the analyst teams to differentiate the functional structure from other hypothesis organizational structures than it was for them to differentiate the divisional and hybrid structures from the other structures.

The NetSTAR algorithm also out performed the human analysts at mapping actor nodes. At this level a fair amount of information must be considered and a large number of possible match combinations evaluated. These are activities that are inherently better performed by computer based algorithms and performed notoriously poorly by humans. This was no more evident then in the results showing the NetSTAR algorithm’s ability to correctly identify 70% of actor-role mappings in the 50% fogging condition, whereas, human teams could only correctly map 12%. Unlike the stimulus organization structures identification task, the NetSTAR algorithms appeared affected by organizational type when performing actor node mapping. The human teams’ performance was relative uniform across the stimulus organizational structure at about 32% correct mapping except for high fogging condition where performance dropped. NetSTAR’s performance was lowest for the functional and hybrid-30% organizational structures and perfect for the hybrid-10% structure. It is not immediately obvious what in the NetSTAR algorithms and the organizational
structures interacted to create this pattern of results. But, it provides at least one point of departure to examine and improve the algorithms.

The comparisons between human teams and the NetSTAR algorithms demonstrated that while human analysts are capable of working with “noisy” observed data and discerning from a set of hypothesized organizational structures the organizational structure that produced the observed data, and to do so well above chance, they were still far below the performance accuracy provided by the NetSTAR algorithms. We also observed the inherent limitations of human decision makers: to handle large amounts of data and organizational networks with high network complexity. However, the experience of observing human analysts grappling with these problems helped us gain insights into the variables that influence their success. These invaluable insights will help us to further improve our ability to scale solutions to real domain problems. Moreover, by studying the way human analysts solve problems helps us to develop and improve NetSTAR-based threat assessment decision support tools that will be usable and trusted by the analysts. We believe that NetSTAR is a significant step forward to provide intelligence analysts tools to identify and analyze adversarial organizations.

References.


