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

Extended Abstract

Motivation: Adversarial Analysis Problem

To successfully predict the actions of the adversary, identify high-value targets 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: the intelligence analysts have to deal with huge amounts of data, most of which is irrelevant to the analysis being performed. Large information gaps, including missing data, deceptions, and errors, have to be dealt with, and analysts often fill the gaps with their experiences which might not be applicable to the problem they need to solve, thus 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 and thus never studied. All these factors negatively impact the ability of the intelligence team to recognize active enemy and further results in decreased efficiency of counteractions and unintended consequences.

Currently, only a limited set of tools is available to intelligence operators to analyze, correlate and visualize the data. No tools with 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 have been developed. In the past, this was due to the inability to bring all data sources together for common analysis. As new tools and data collection techniques become available, the feasibility of new technologies to automate threat prediction is increasing.

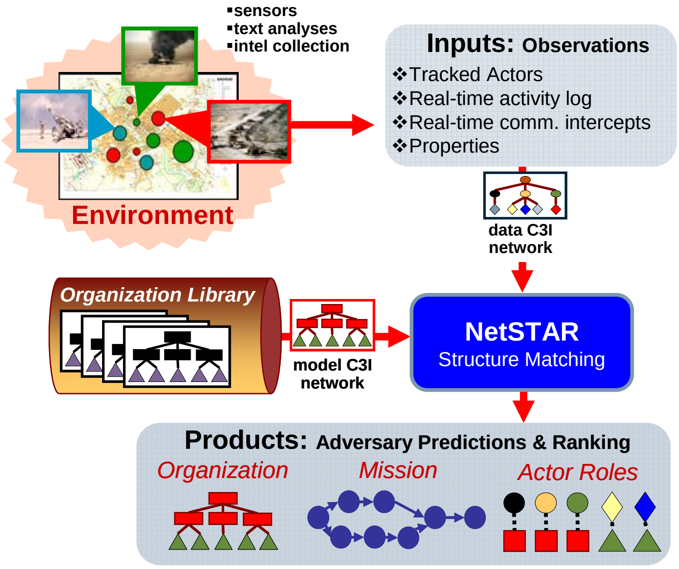


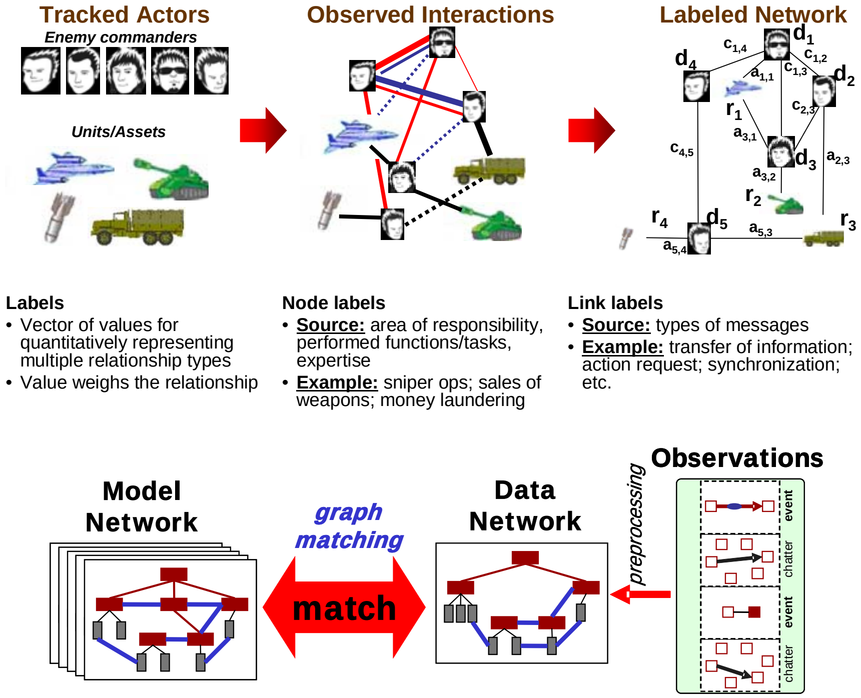
Figure 1: NetSTAR Adversarial Identification Process

Method: Automated Organization Identification Model

Recently, a new model to identify an adversarial organization using observations about actors’ actions and interactions has been proposed (Levchuk and Chopra, 2005; Levchuk, Levchuk, and Pattipati, 2006). The approach (Fig. 1) is based on the hypothesis testing principles, and the model uses probabilistic attributed graph matching algorithms to come up with a mapping of observed actors to organizational nodes and rank-ordering of the organizational network hypotheses.

The available data that can be used for identifying adversarial organization usually consists of partially classified communication transactions among tracked actors (e.g., “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., “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 communication networks, uncertainty in message translation, data association uncertainty, etc. Therefore, our framework has to rely on **probabilistic**

association between the actors and the decision-making nodes and between the observed assets and the resources of the adversarial organization.



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 2: Problem Setup

The observations about communication exchanges among nodes can be augmented by discovering the linkages among the commanders, resources, tasks, and other environment objects. In short, we are observing a network of relationships of different types among the enemy actors (humans), physical resources, tasks, goals, etc. This network must be mapped to the network of command, control, and communications of the hypothesized organization.

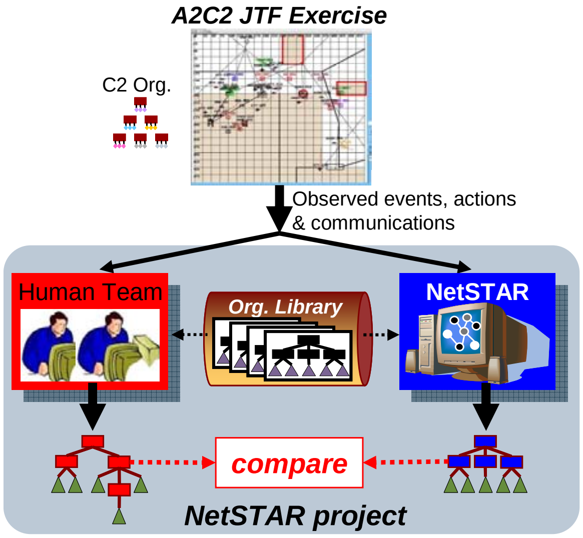


Figure 3: Problem Setup

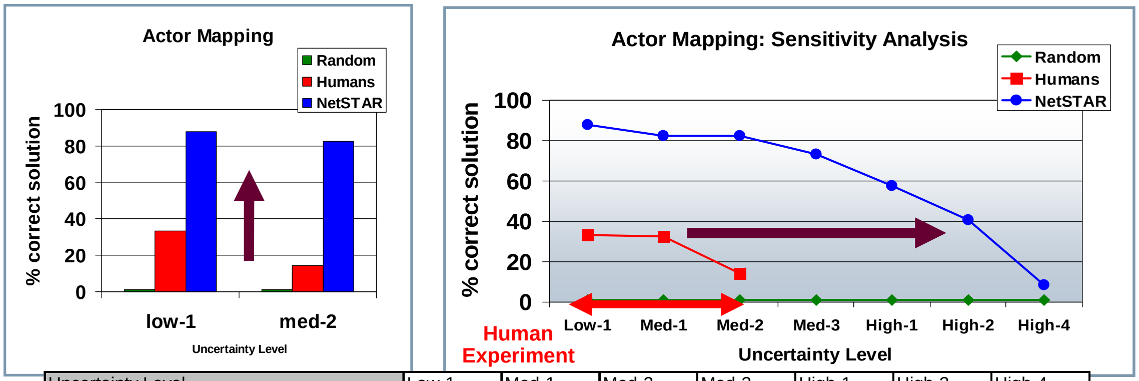
Given such type of data, we pose the problem as one of finding the mapping between nodes of two graphs: observed (also termed *data*) network of adversary actors and their interactions/relationships, and hidden network corresponding to the hypothesized (also termed *model*) network (Fig. 2). The mapping is found by maximizing the scoring function, which might be a likelihood function or a posterior probability. The mapping must account for the attributes or features of both nodes and links, 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 link attributes may correspond to types of interactions and relationships between nodes in the adversary C3I organization (e.g., communication messages may be of the following types: request for or transfer of information, resource, action; acknowledgement; direction; etc.).

Results: Comparison of Model-based Solution versus Human-based

To validate the NetSTAR model, we have conducted a project sponsored by Information Exploitation Office of DARPA. The project has implemented NetSTAR algorithms and conducted human table-top experiment to obtain the accuracy data for human analysts performing the same task as NetSTAR automated identification tool. The project has used the data from virtual command and control experiments of joint task forces mission exercises (Fig. 3) conducted in past several years under Adaptive Architectures for Command and Control (A2C2) program (Kleinman et al., 2003).

This paper is a part I of 2-part paper submission describing the NetSTAR project. In this paper, we describe the problem setup, the types of data that were used for analyses, the novel modeling approach for attributed graph matching that we have been developed during the project, and the outputs of the computational experiment assessing the sensitivity of the NetSTAR identification to data uncertainty and type of the enemy organization that needs to be identified. In the part II paper, we will describe the experimental setup, present experimental analysis, and compare results of NetSTAR algorithm to those of human analysts.

The NetSTAR project has showed that the automated threat identification algorithm outperformed unaided human analysts providing at least 2.5 times more accurate mapping between observed actors and their correct roles in the organization (Fig. 4). NetSTAR algorithm also provided a robust solution being able to correctly identify 70% of actor-role mapping for 50% of missing data and 30% detection.



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

NetSTAR provides >2.5X better detection than human analysts under same uncertainty level

NetSTAR achieves same performance as human analysts under 3X uncertainty level

Figure 4: Sensitivity to Uncertainty and Comparison of NetSTAR Algorithm to Human Performance

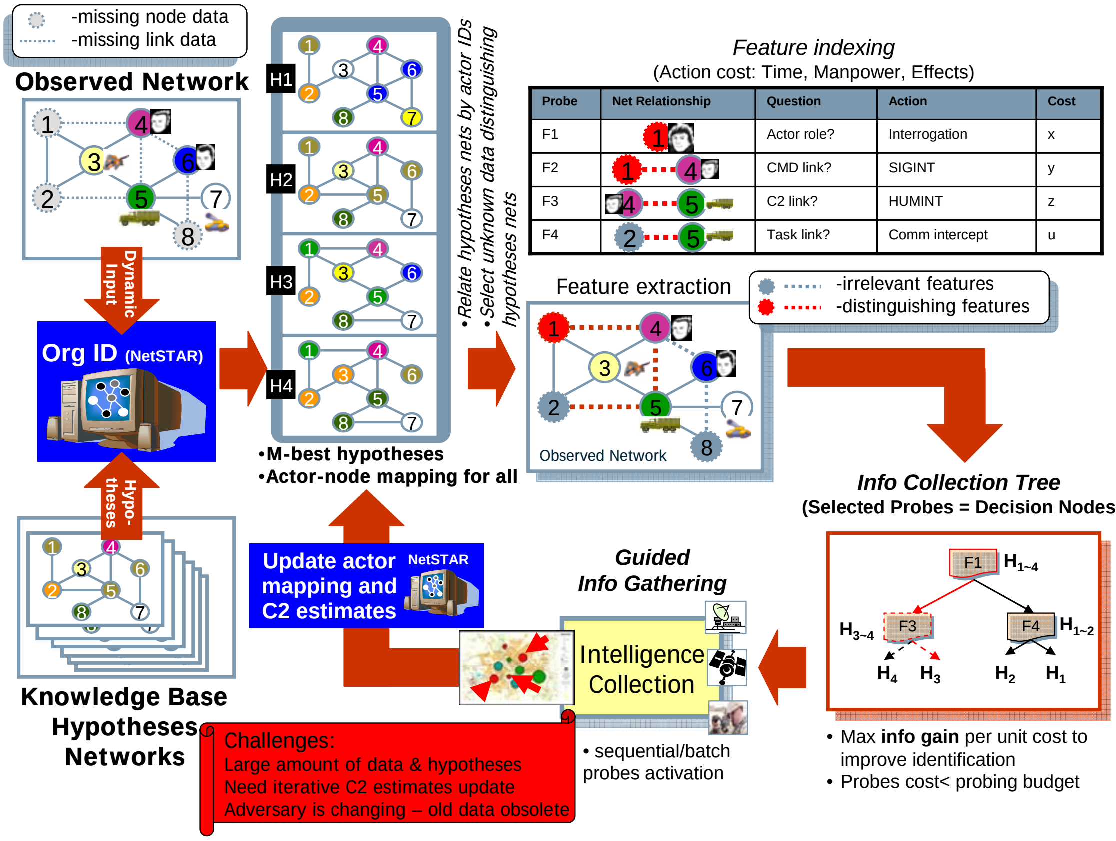
(Random = results of random identification; NetSTAR = results of automated algorithm identification; Humans = results of threat identification by human analysts during table-top experiment)

Network Identification Extension: Change Detection & State Evolution Tracking

As the enemy adapts, so should the output of threat identification system. Current NetSTAR models use all data aggregated without considerations for temporal evolution of the adversarial organization. In this paper, we will describe an extension to structure matching approach to detect evolving adversarial networks.

Next Steps: Guided Intelligence Gathering for Improved Identification

When additional information collection is possible, the ability to prioritize and plan these activities is essential 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 hypotheses ranking to come up with most important missing information elements (features) that would reduce the ambiguity the most for current organization identification. The data collection plan is then developed by ordering the collection actions 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 (Fig. 5).



-missing node data
-missing link data
Observed Network
1
2
3
4
5
6
7
8
Dynamic Input
Org ID (NetSTAR)
Hypo-theses
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
F3
F4
H_{1-4}
H_{3-4}
H_{1-2}
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
Knowledge Base Hypotheses Networks
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 4: Guided Information Collection Process

References:

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.