Title: **Identifying Command, Control and Communication Networks from Interactions and Activities Observations**

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

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

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

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

Abstract

This paper presents a novel methodology that utilizes the observations of interactions and relationships among enemy actors, resources, and tasks to identify the acting C2 organization (i.e., decision hierarchy, allocation of resources and roles to decision–makers, communication and information structures, and expertise of decision–makers). The model is based on probabilistic attributed relational graph matching methodology that has been successful in computer vision and pattern recognition. The basis for our model, an expectation maximization algorithm to solve maximum-likelihood estimation of the node mapping, has successfully outperformed other graph matching methods and shown the ability to correctly map 70% of nodes under 50% of link (structural) errors (Luo and Hancock, 2002). Our proposed algorithm enhances the Luo's EM-SVD algorithm by accounting for both node and link attributes uncertainty, as well as multiple substructures of the interaction networks, to derive the mapping between observed nodes and model (hidden) nodes. Once the mapping between actors and the nodes in a given command organization is learned, it can be used to associate the observed transactions with hidden enemy activities and further improve the recognition and prediction of adversary's behavior.

1. Motivation

The knowledge of the principles and goals under which the adversary organization operates is required to predict its future activities. To implement successful counter-actions, additional knowledge of the specifics of organizational command, communication, control, and information access (C3I) networks, as well as responsibilities distribution among members of the organization, is required. In short, the attributes of and relationships among team members, assets, environment areas, and mission tasks are needed.

Analysis of the behavior of organizations, ranging from the more structured command systems of a conventional military to the decentralized and elusive insurgent and terrorist groups, suggest that the strong relationship exists between the structure, resources, and objectives of those organizations and the resulting actions. The organizations act in their missions by accomplishing tasks which may leave detectable events in the information space. The dynamic evolution of these events creates patterns of the potential realization of organizational activities and may be related, linked, and tracked over time (Pattipati *et al.*, 2004). The observational data, however, is very sparse, creating a challenge to connect relatively few enabling events embedded within massive amounts of data flowing into the government's intelligence and counter-terrorism agencies (Popp *et al.*, 2004). On one hand, to successfully detect the interaction patterns over time, the models need to associate the observations with true but hidden interactions in the adversarial organization. Such association is at the core of activity pattern identification of NetSTAR process (Levchuk and Chopra, 2005) and requires the knowledge of the mapping between the observed actors and the decision-making nodes in the hypothesized enemy organization. On the other hand, this mapping defines the role of the actors and their place in the organization, which is essential to directing the counter-measures against the most important enemy nodes or relationships among them. In this paper, we propose an enhancement to our NetSTAR process via finding the mapping between the observed actors and the decision-makers of the hypothesized enemy organization.

Ultimately, the best source of information about the enemy is the enemy himself. While getting into the mind of the enemy is impossible, thinking similarly is desirable to truly predict its next moves. When dealing with an enemy organization, we are not merely interested in learning about individuals, but in how they are organized as a team and what they can do together. The structure of an enemy organization defines its capabilities, while its goals define the mission(s). Just like the brain structure can be discovered using MRI scans, the structure of the enemy organization can be discovered from observations of interactions and activities of its members; these may be observed as part of normal activity monitoring, or they may be

induced with intentional probes. In the context of discovering a covert organization, the scope of the probes is very limited. Therefore, one needs to use real–life observations obtained as part of normal monitoring, which are tightly coupled with the intent of the adversary.

Connections (e.g., communication, command) between individuals in the covert organization have three effects. First, they provide a means to share information and resources, and coordinate task execution. 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 bottleneck resources, or disabling organizational connections would allow disruption of enemy's operations and decision–making processes for preemptive actions.

The relationships among adversaries and patterns of their activities change over time. This requires continuous/periodic updates of the knowledge about organizational structure to execute effective counter–measures. The accelerating technologies of communications and computers are overflowing the intelligence analysts with information at various levels of decision making. Current labor–intensive manual processes to discover enemy organization fail to keep up with dynamic environments. Therefore, analytical tools are needed to reduce the complexity of organizational discovery and to allow analysts to focus on information most essential for decision making and search through only a limited number of most likely hypotheses.

The problem of structural discovery is very complex: the observed data does not relate to the structure directly; instead, it relates to its manifestation in the form of activities and processes that are enabled by the organizational structure(s) and performed by the organization's members. Therefore, the algorithms to reconstruct the organization from observations alone would need to search through a very large space of possible structures. Given historic data and the availability of subject–matter experts, we can instead pose the problem as one of hypotheses testing, where the set of predefined hypotheses about the enemy organization and its subelements, albeit very large, is given. The problem then becomes one of rank–ordering these predefined hypotheses on the basis of how best they match (or explain) the observed data.

To form the hypotheses about the enemy organization structure to help the identification process, we first need to need to understand what is it that we need to discover. That is, we need to develop the **quantitative representation** of the various aspects of the adversary's organization, relationships between its members, behavior rules, and how this behavior can be structured, modeled, and related to the observations. This representation of enemy command and control organization is presented in Section 2. In Section 3, we outline the structure of the observations available for our use from intelligence sources. Section 4 describes the formal model for mapping the actors to specific positions in the given organization, with mapping algorithm details presented in the Appendix. We conclude the paper with the application discussion in Section 5.

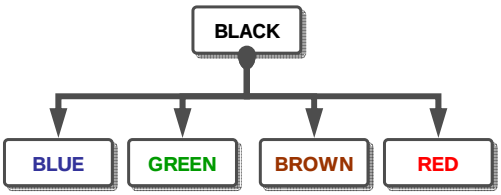
2. Command and Control Modeling Principles

Our adversary identification methodology is based on understanding and modeling the decision-making processes in the general command and control (C2) organization. While C2 organizations are designed to manage personnel and resources to accomplish the mission requiring their collective skills, the term is not isolated to only one type of the organization and is common to both friendly and adversary domain. Given a 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. Since one of the most important findings from the organization theories research is that there is no single "best" approach to (or philosophy of) command and control, many organizational constructs are possible.

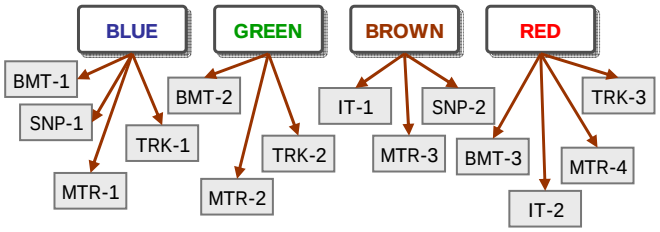
Command and control refers to procedures used in effectively organizing and directing armed forces to accomplish a mission. The *command* function is oftentimes referred to as an art of an individual to set the initial conditions and providing the overall intent for mission execution. The *control* is 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, suggestions to commanders of adjacent units, and requests and reports to supporting units and superiors. He develops and maintains a situational awareness of the area of his operations through reports presented by other people or by electronic systems (Coakley, 1991). The basic premise of command and control organization is the ability to distribute the responsibilities among its elements and coordinate these seemingly independent entities for joint operations to achieve the objectives. The fundamental need for communications significantly constrains the options for both command and control, making communications infrastructure a critical feature of a C2 system. However, describing the communications links and nodes of a fighting force does not suffice to explain, understand, or predict successes and failures in command and control. We need to be able to represent, model, and identify the functions and objectives of the individual elements of the C2 organization.

			functions			
Name	#	Description	Bomb	Strike	Intel	Transport
BMT	3	Bomb Maker Team	1	0	0	0
SNP	2	Snipers	0	1	0	0
MTR	4	Mortars	0	2	0	0
IT	2	Intel Team	0	0	2	0
TRK	3	Truck	0	0	1	1

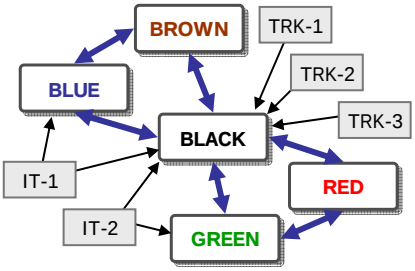
(a) Resources of C2 Organization



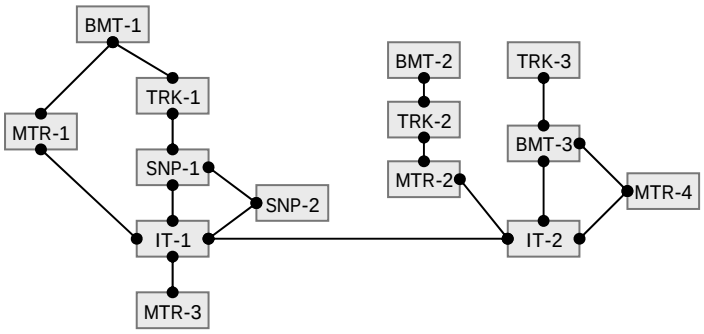
(b) C2 Nodes & Command Structure



(c) Control Structure



(d) Communication Structure (with resource reporting beyond control structure)



(e) Task Structure (example)

Figure 1: Example of C2 Organization

In our modeling, we describe the *command and control organization* as a collection of command and control (C2) nodes and resources connected via command, control, communication, and task structures

(Figure 1). 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 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 the company–size forces, we can consider resources being the single squads. 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.

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.

In Figure 1 we present an example of the enemy command and control military team consisting of 5 command elements and 14 units/resources. The commanders of this organization make decisions to manage assigned resources in 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. Figure 1.a describes the set of resources – military units and assets controlled by commanders. The assets include bomb maker teams, sniper teams, mortars, intelligence and reconnaissance teams, and trucks. This chart also shows the *functional* or *resource capabilities* (Levchuk *et al*., 2001) of the units and resources in terms of bomb making, strike and small-arms attack, intelligence and monitoring, and transportation. The authority structure among 5 commanders is a flat hierarchy (Figure 1.b) with a single commander ("BLACK") being a main commanders of enemy forces. The assignment of assets and units to commanders (Figure 1.c) determines the control structure of the C2 organization. Note that in hypothetical example of Figure 1 the main commander ("BLACK") does not control any resources directly. A communication structure (who can talk to whom) of the organization is depicted in Figure 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 Figure 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 the organization discovery is the ability to **recognize** the command, control, 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 command and control 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 observation data available from intelligence gathering sources.

3. Observations

What can be gleaned about the adversary organization? What data is most useful to discover individual's roles as well as important relationships within the organization, and which data is meaningless? What types of observations are more important to discovery, and thus should be the focus of data collection efforts? These questions are at the heart of organizational identification problem. While it is desirable to have full access to certain types of information about the enemy to improve detection capabilities, in practice, we are constrained with the types of information that *can* be collected.

In the normative representation of organizational structures and quantitative theories of organization identification, we restrict our models to several types of intelligence information that are currently feasible. We assume that the **observations** can include the set of tracked (monitored) ***individuals*** whose positions in the organization we need to determine, information about these individuals, and identified adversary's resources. Tractable information regarding the individuals encompasses their attributes and resources – e.g., expertise of individuals, training, background, affiliation, cultural characteristics, family ties, etc. Information about adversary's resources may include detection of enemy's military assets and their capabilities, communication means, political connections, and financial capabilities. In addition, the observed information may also include ***transactions*** that involve these entities; these comprise of intercepted partially deciphered *communications* among the individuals and the individuals' *actions* – the involvements in various observable activities. Communication observations (e.g., "*members of militant wing engaged in a meeting with weapons suppliers at 11:35 am for 35 min to procure explosives*") may include some classification of communication content, e.g. request for or transfer of information, resource, action; acknowledgement; direction; etc. Action observations (e.g., "*BLUE team discovered a safehouse and apprehended RED operatives attempting to manufacture weapons*") may include functions/tasks performed, such as individuals committing the same crime, performing financial or business transactions, or using the resources in covert or open operations. 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, attempts of the adversary to conduct deceptive actions, 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.

The observations can be related to true activities and communications of the adversary organization. They could also be related to the structural elements of the organization. For example, if we intercept the command message from one individual to another, it must be that the first individual is a commander of the second individual. If we intercept the information exchange between two individuals, it means that they are connected through an existing communication channel. However, we encounter problems when the uncertainty of the observations increases, and we cannot make abovementioned conclusions with confidence.

Adopting the hypotheses testing approach, where we need to test which of the organizations from the library of hypothesized C2 structures **best explains** the observations, addresses some of these problems. In this model, we need to relate a given structural link or an activity to the one that was observed. However, there are too many structural links and activities that seem to be similarly related to the observation. This is especially true when the content of the observation is highly uncertain and sparse. What we can and need to use to improve the discovery is the information about participants of the observed action and/or communication. In this case we encounter a new challenge: while the individuals of the adversary can be tracked over time and distinguished between one another, we do not immediately know how they related to the individual positions in the hypothesized organization. This problem is illustrated in Figure 2. For example, the intelligence can be obtained from intercepted communication that the "***individual X ordered individual Y to conduct operation***", but what operation is not known. This observation can be associated

with the command relationship between the nodes of adversary organization (on the left of Figure 2), as well as with some hypothesized actions in the behavior model (on the right). First, we need to understand what nodes in hypothesized C2 organization do the tracked individuals "X" and "Y" correspond to. If "X" is commander "BLACK", and "Y" is commander "BROWN", then this observation can be related to the structural relationship "*BLACK is commander of BROWN*". In this case, we are also able to associate this observation with hypothesized activity "*BLACK ordered BROWN to assure safe exit in the village*". If another association is made (e.g., "X" is commander "BROWN" and "Y" is unit leader of "SNP-2"), then other relations of this observation with structural links and with hypothesized activities are possible. Note that such relations are not necessarily unique. For example, if we had received intelligence that "*unit Z has been observed maneuvering in the village*", then if we assume that "Z" is unit "IT-1", then this could mean any of the actions "*IT-1 surveys village entrance*", "*IT-1 monitors roads*", or "*IT-1 sets positions for sniper fire with SNP-2*".

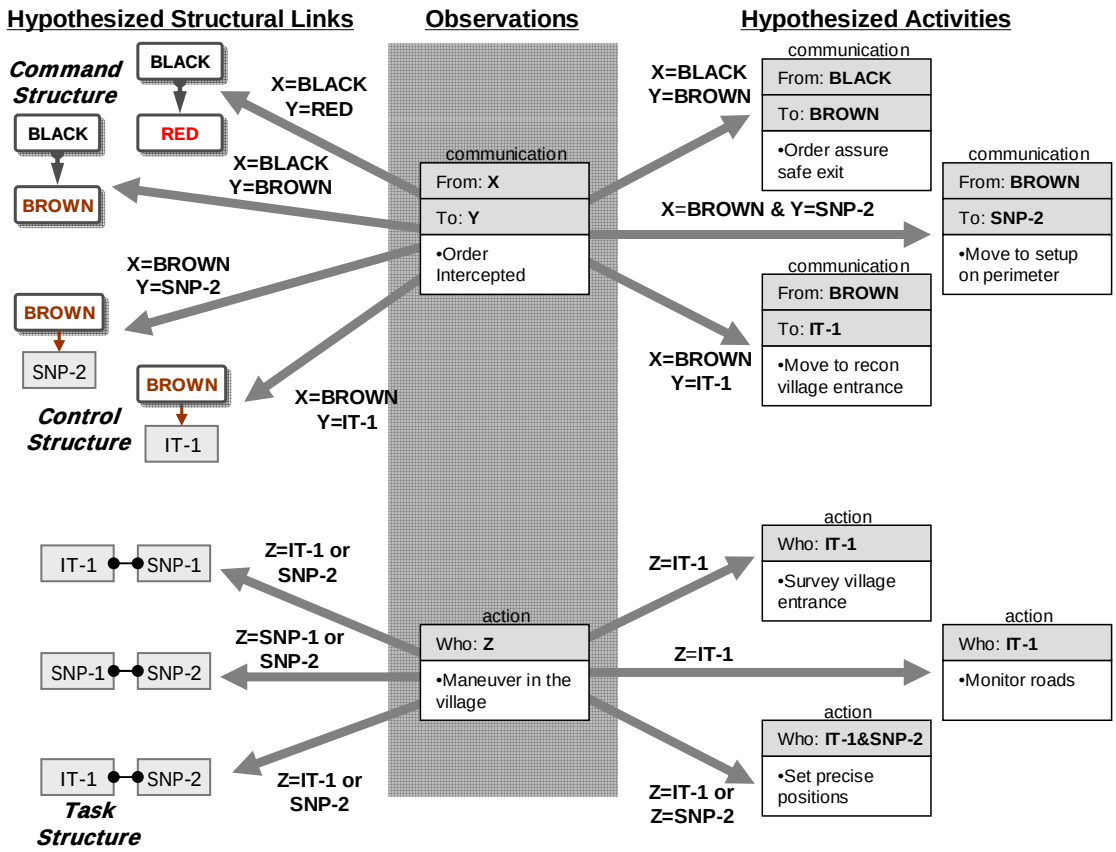


Figure 2: Example of Relating Observations to Structural Links and Activities

The associations of the tracked individuals "X" and "Y" with commanders and units/resources of hypothesized C2 organization is called *node mapping*: assigning observed individuals to the positions in a hypothetic organization. It might seem that being able to place a tracked individual at the specific position in the organization is enough. But discovering positions of individuals in the organization in this fashion lacks the notion of organizational activity *evolution*, and therefore cannot distinguish between the organizations and missions. Still, without the node mapping we would not be able to relate the observations with hypothesized activities. Even if such associations were possible, they would be too broad and would influence future relations. This influence would be due to implicit association between tracked individuals (such as "X" and "Y") and the corresponding participants in the hypothesized activities. This association

would prevent us from using hidden Markov models (HMMs) to identify activity patterns (Levchuk and Chopra, 2005), since HMMs assume independence between observations.

This example illustrates the need to obtain the mapping of tracked individuals to actual C2 nodes and resources of hypothesized organization. We address this challenge in next section using the formalism of probabilistic graph matching.

4. Mapping Actors to Organizational Nodes using Probabilistic Graph Matching

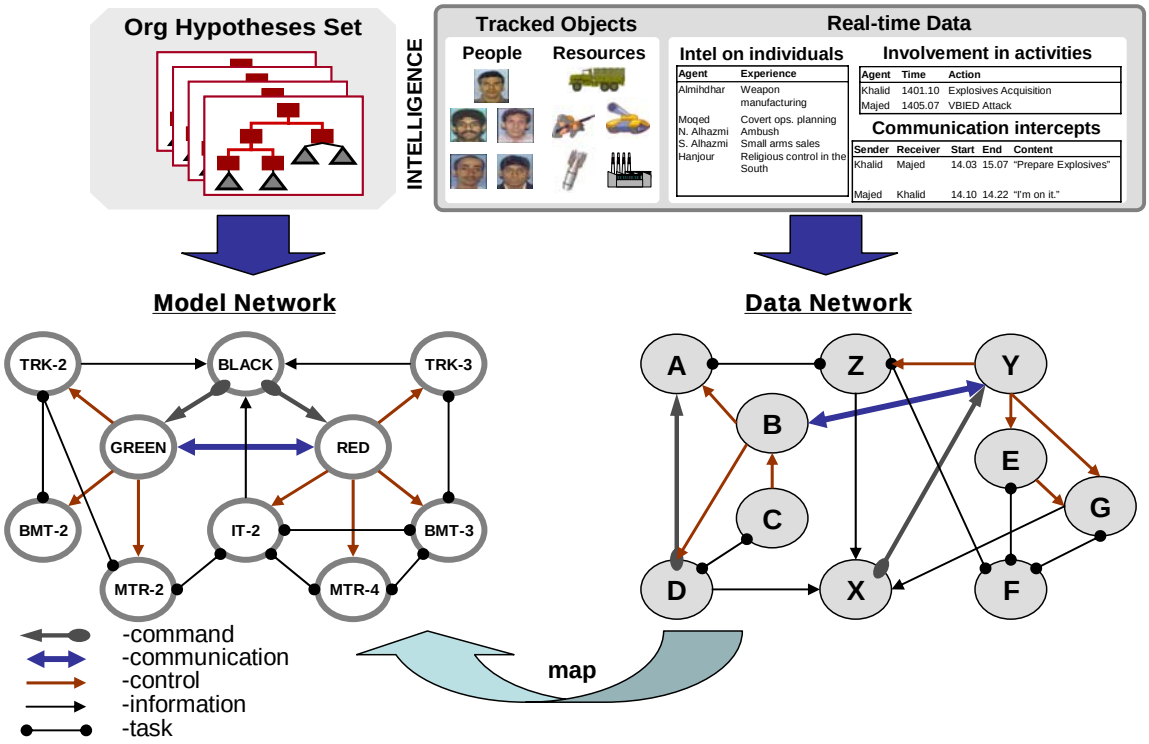
How can we know the true identify of a tracked individual? This is a main question of **node mapping** problem. The challenge is to relate the observed individual with a place in the organization – mapping him against C2 node and its concomitant command role, expertise, dedicated responsibilities, access to people, control over resources, etc. Just as law enforcement agencies discover this knowledge by collecting the pieces of data about an individual, his relationships to others, and participation in transpired events, we base our decisions on observed attributes of individuals, transactions between them, and attributes of those transactions.

The observations about communication exchanges among nodes can be augmented with discovering the linkages between the nodes and 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, etc. This network must be mapped to the network of command, control, and communications of the hypothesized organization. The mapping is accomplished via mapping between the nodes of corresponding networks.

The time component is temporarily disregarded, with all observations aggregated and used for a single estimation evaluation. All collected observations are linked together to form the **data (observed) network**. The nodes of this network are tracked individuals, units of enemy fight force, and other resources. The links of this network are the structural relationships perceived to be realized in observed communications between tracked individuals, commands sent from one individual to another or to the unit, information requests and information transfers, joint task executions (discovered from action observations), etc. In our modeling, we utilized the classification of communication into 12 classes of communications (Entin, Diedrich, & Rubineau, 2003). These classes were assumed to be fixed and no other information about the intercepted communications was used. The classes allowed determining what organizational relationship did the intercepted communication correspond to, thus creating the concomitant link in the data network. We need to find how to map the nodes of this data network to the **model network** – the hypothesized C2 organization with command, communication, control and task substructures. Figure 3 illustrates the problem, where 10 nodes and units of the adversary has been detected (A,B,C,D,E,F,X,Y,Z), and the communication intercepts and action observations of the adversary are aggregated to a data network. Matching the topology of this network to a hypothesized C2 model network produces the correct mapping A=MTR-2, B=GREEN, C=BMT-2, D=TRK-2, E=MTR-4, F=BMT-3, G=TRK-3, X=BLACK, Y=RED, Z=IT-2. That is, we say that tracked agent “X” is commander “BLACK”, tracked resource “A” is a mortar resource (MTR-2), agent “Y” is commander “RED”, etc.

To relate the nodes of data network to nodes of model network, we employ probabilistic node mapping (Wilson and Hanckock, 1997), (Luo and Hanckock, 2001) to maximize the likelihood that the data network has been generated by the model network over all possible mappings from nodes of data network to nodes of model network. The mapping does not have to be one-to-one. It could be one to many, which is equivalent to discovering the individual of the enemy organization that performs multiple roles, or many to one, which is equivalent to finding the enemy command cell.

To develop our models, we draw on the research in 2- and 3-D computer vision and structural pattern identification. The mapping is found by maximizing the likelihood function equal to the probability that the observed interaction/relationship network has been generated by the model network coupled with the mapping between the nodes of observed and model networks. This 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.).



Correct Mapping:
A=MTR-2, B=GREEN, C=BMT-2, D=TRK-2, E=MTR-4, F=BMT-3, G=TRK-3, X=BLACK, Y=RED, Z=IT-2

Figure 3: Example of Node Mapping Problem

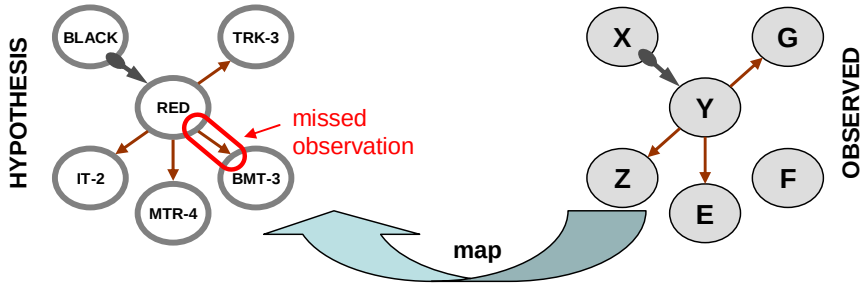


Figure 4: Example of Node Mapping Problem

To find maximum likelihood, we employ structural consistency approach. The mapping is obtained to maximize the topological closeness between data and model networks. In simplified terms, it is equivalent

to finding the consistency between the elementary substructures of the data and model networks. An example of elementary substructure is a clique – a node and all its neighbors (see Figure 4). The consistency is based on (i) correctly matching the nodes of data network to the nodes of model network (i.e., relating observed individuals to the C2 nodes in hypothesized C2 organization), (ii) maximizing the number of correctly identified relationships between nodes, and (iii) minimizing the number of incorrectly identified relationships and missed (unobserved) relationships. For a formal problem formulation and mathematical derivations for the node mapping algorithm, see the Appendix. The node mapping algorithm (illustrated in Figure 5) using expectation maximization iteratively updates the consistency between the topologies of the data graph, permuted with the mapping function, and the model graph. The consistency is based on (i) correctly matching the nodes of data network to the nodes of model network (i.e., relating observed individuals to the C2 nodes in hypothesized C2 organization), (ii) maximizing the number of correctly identified relationships between nodes, and (iii) minimizing the number of incorrectly identified relationships and missed (unobserved) relationships.

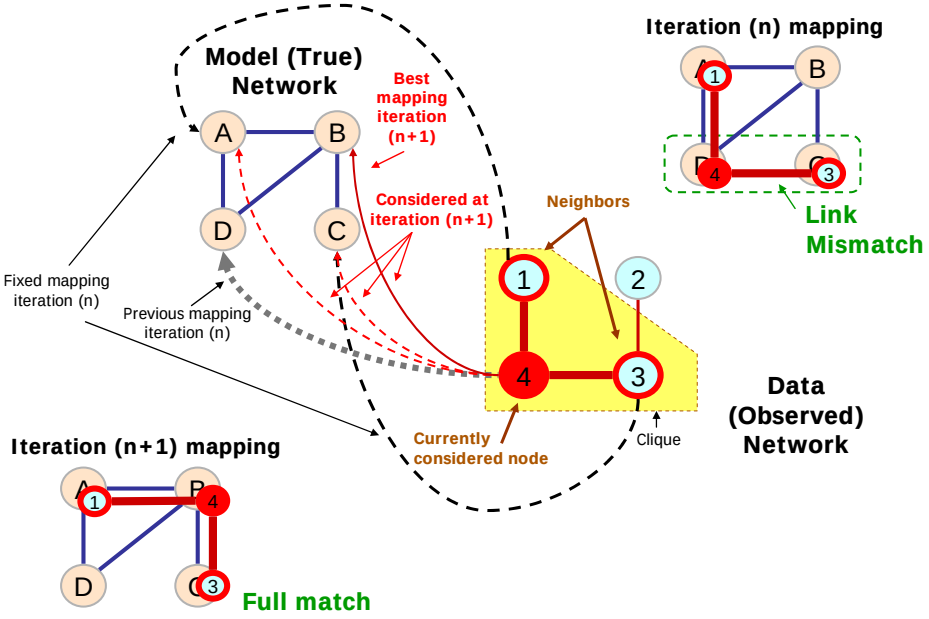


Figure 5: Illustration of Node Mapping Update

5. Model Application and Future Research

One of the important modeling aspects of this research is the ability to associate the observed transaction (communication, action) with the structural elements of the organization. More sophisticated models are needed to represent these relationships. One such approach is to discover the attributes that describe the observed data (both suspected enemy actors and their behaviors) and use them to relate observation to the structural relationship. This approach would require utilizing the heterogeneous error models, solutions for which will require resolving significantly higher complexity. Another challenge is deciphering the content of the communication and using its partial knowledge in the detection process. We have utilized fixed classification of the observed communications. The actual message classification is beyond the scope of this research.

It might seem that being able to relate a tracked individual to the specific position in the organization is enough. But discovering positions of individuals in the organization in this fashion lacks the notion of organizational activity *evolution*, and therefore cannot distinguish between the organizations and missions.

The mapping algorithm uses all the observed data together. Its complexity does not allow the dynamic re-mapping to be performed with every incoming observation, which is needed for organization monitoring to detect when the organization changes its structure or changes its mission. The mapping does, however, allow associating the observed transactions with activities in the behavior model of the adversary. This will be used to monitor and detect patterns of adversary activities and will serve as the final organization detection step.

The ultimate product of adversary organization research is the development of a system to disrupt enemy organization at the tactical and/or strategic level based on control–theoretic methodology. The information collected from various sources about enemy activities, communication intercepts, knowledge about enemy resources, relationships between goals and actions, hypotheses about the enemy command and mission, – all are the inputs into the system. The developed system can be integrated as a part of a larger OODA decision cycle, transformed into Observe, Identify, Assess, Plan, Execute process.

More specifically, the knowledge obtained with the help of adversary identification process can be used in two ways:

(a) Improving predictions of enemy's courses of action (COAs).

(b) Improving the efficiency of disrupting the enemy.

Current predictive algorithms oftentimes do not take the enemy organization into account when developing possible enemy COAs. Usually, the methods assume either one specific distribution of the resources control, information, knowledge, and command in the enemy organization, or the "worst–case" COAs are developed which imply joint planning and execution of enemy activities. This approach does not match to the real world. It has been proven in computational organization theories and has been validated in numerous experiments that the structure of the command, control, communication, and information access significantly impacts the actions selected by the organization.

The actions to disrupt the enemy developed by currently employed methods are often based on trying to eliminate the most critical enemy resources/assets. Most of the time, the criticality of the assets is judged outside of the enemy organization by simply considering the individual functions and capabilities of the assets. However, such an approach may end up with miscalculating the future impact of the assets since it can only be evaluated through understanding where the enemy is intending to use this asset. The latter depends on the responsibilities of the commander who controls the asset. As the result, if we can identify the future employments of the assets by the enemy, then we may be able to eliminate those assets and prevent the enemy's ability to coordinate supplementing the lost resources. Due to distributed nature of such operations, this will represent a more efficient approach to disrupt the enemy capabilities than traditional procedures.

References.

Coakley, T.P., *C3I: Issues of Command and Control*, Washington, D.C.: National Defense University, 1991, pp. 43–52.

Entin, E.E., Diedrich, F.J. & Rubineau, B. "Adaptive Communication Patterns in Different Or-ganizational Structures", *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, Denver, CO, 2003.

Kleinman, D.L., 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.

Levchuk, G.M., 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.

Levchuk, G.M., Y.N. Levchuk, J. Luo, K.R. Pattipati, and D.L. Kleinman, "Normative Design of Organizations - Part I: Mission Planning", *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 32, No. 3, May 2002, pp. 346-359.

Levchuk, G.M., Y.N. Levchuk, J. Luo, K.R. Pattipati, and D.L. Kleinman, "Normative Design of Organizations - Part II: Organizational Structure", *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 32, No. 3, May 2002, pp. 360-375.

Luo, B., and E.R. Hancock (2001), "Structural Matching using the EM algorithm and singular value decomposition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1120—1136

Luo, B., and E.R. Hancock (2002), "Structural Graph Matching using the EM Algorithm and Singular Value Decomposition", Asian Conference on Computer Vision 2002, pp. 75-80

APPENDIX: Node Mapping Problem Formulation and Solution Algorithm

Formally speaking, we represent a *hypothesis 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 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 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. We find an assignment matrix S that maximizes the likelihood function $P(G_D \mid G_M, S)$, which is equal to the probability that the observation (data network) has been generated by the hypothesis 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 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 is infeasible, an approximate solution can be obtained by relaxing structural consistency measure to consider subgroup match, and then employing expectation maximization algorithm to find iterative solution (Luo and Hanckock, 2001). 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 function $P(G_D \mid G_M, S)$.

To simplify problem formulation, we assume that there are two types of errors present. One type corresponds to the deletions of the links from the model graph, which occurs with probability p_e^L. This is equivalent to saying that probability of observing an existing link (probability of detection) in the model (of the hidden organization) is $1 - p_e^L$. For simplicity, we also assume that probability of "adding" the link (false alarm probability, i.e. having a noisy link – observation about relationship which is deceptive) is p_e^L. Another error generation is according to the random mislabeling of the actors to the roles/nodes in the hypothesized graph – that is, the initialization error of incorrectly mapping the node of the data graph with

the node of the model graph. This occurs uniformly with probability p_e^N, and is equivalent to saying that the probability to correctly observe the node's attributes (or mapping the node in a model to the node in a data graph) is equal to $1-p_e^N$.

To solve maximum-likelihood problem, we assume conditional independence and mixture distribution assumptions similar to (Luo and Hancock, 2001): $P(G_D \mid G_M, S) = \prod_{a\in V_D}\sum_{\alpha\in V_H} P(x_a \mid y_\alpha, S)$. Here, $P(x_a \mid y_\alpha, S)$ corresponds to the probability of observing the data node (tracked individual) x_a given a model node (an information about C2 node) y_α and a mapping from data to model nodes S. Our algorithm uses clique-based relaxation and both node and link uncertainty models. The relaxation techniques used in our algorithm are better representing structural consistency of matching graphs.

To find $\hat{S} = \arg\max_S P(G_D \mid G_M, S)$, we first use conditional independence and mixture distribution assumptions to write: $P(G_D \mid G_M, S) = \prod_{a\in V_D}\sum_{\alpha\in V_H} P(x_a \mid y_\alpha, S)$. Then, using Bayesian theory, we have:

$$P(x_a \mid y_\alpha, S) = \frac{P(S \mid x_a, y_\alpha)P(x_a, y_\alpha)}{P(S \mid y_\alpha)P(y_\alpha)} = \frac{\prod_{b\in V_D} P\left(\prod_{\beta\in V_M} s_{b,\beta} \mid x_a, y_\alpha\right)P(x_a, y_\alpha)}{\prod_{b\in V_D} P\left(\prod_{\beta\in V_M} s_{b,\beta} \mid y_\alpha\right)P(y_\alpha)}$$

$$= \frac{\prod_{b\in V_D} \frac{P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)P\left(y_\alpha \mid \prod_{\beta\in V_M} s_{b,\beta}\right)P\left(\prod_{\beta\in V_M} s_{b,\beta}\right)}{P(x_a, y_\alpha)}P(x_a, y_\alpha)}{\prod_{b\in V_D} \frac{P\left(y_\alpha \mid \prod_{\beta\in V_M} s_{b,\beta}\right)P\left(\prod_{\beta\in V_M} s_{b,\beta}\right)}{P(y_\alpha)}P(y_\alpha)} = \left[\frac{1}{P(x_a \mid y_\alpha)}\right]^{|V_D|-1} \prod_{b\in V_D} P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)$$

Assuming that $P(x_a \mid y_\alpha) = P(x_a)$, we get: $P(x_a \mid y_\alpha, S) = B_a \prod_{b\in V_D} P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)$. Here, $P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)$ corresponds to the probability of observing the data node (tracked individual) x_a given a model node (an information about C2 node) y_α and a mapping from some other data node x_b to model nodes (i.e., the mapping subset $\prod_{\beta\in V_M} s_{b,\beta}$). The probability $P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)$ is computed using the following observations:

1. If $(a,b) \in E_D$ (x_b is a neighbor of x_a in data network) and $\sum_{\beta \in V_M} s_{b,\beta} M_{\alpha,\beta} = 1$ (the node of model network which a data node x_b is mapped to according to mapping $\prod_{\beta \in V_M} s_{b,\beta}$ is a neighbor of y_α), then $P\left(x_a \mid y_\alpha, \prod_{\beta \in V_M} s_{b,\beta} \right) = (1 - p_e^N)(1 - p_e^L) + p_e^N p_e^L = \varepsilon$

   This condition implies that either the link was observed, in which case the correspondence is correct, or that the link was missed, in which case the correspondence is incorrect.

2. If $(a,b) \in E_D$ (x_b is a neighbor of x_a in data network) and $\sum_{\beta \in V_M} s_{b,\beta} M_{\alpha,\beta} = 0$ (the node of model network which a data node x_b is mapped to according to mapping $\prod_{\beta \in V_M} s_{b,\beta}$ is not a neighbor of y_α), then $P\left(x_a \mid y_\alpha, \prod_{\beta \in V_M} s_{b,\beta} \right) = p_e^N (1 - p_e^L) + (1 - p_e^N) p_e^L$

   This condition implies that either the link was not noisy (not "erroneously observed"), in which case the correspondence is in error because there is a link, or that the link was actually noisy (there was a false alarm error), in which case the correspondence is correct.

3. If $(a,b) \notin E_D$ (x_b is not linked to x_a in data network) and $\sum_{\beta \in V_M} s_{b,\beta} M_{\alpha,\beta} = 1$ (the node of model network which a data node x_b is mapped to according to mapping $\prod_{\beta \in V_M} s_{b,\beta}$ is a neighbor of y_α), then $P\left(x_a \mid y_\alpha, \prod_{\beta \in V_M} s_{b,\beta} \right) = (1 - p_e^N) p_e^L + p_e^N (1 - p_e^L)$

   This condition implies that either the link was observed, in which case the correspondence is in error, or that the link was missed, in which case the correspondence is correct.

4. If $(a,b) \notin E_D$ (x_b is not linked to x_a in data network) and $\sum_{\beta \in V_M} s_{b,\beta} M_{\alpha,\beta} = 0$ (the node of model network which a data node x_b is mapped to according to mapping $\prod_{\beta \in V_M} s_{b,\beta}$ is not a neighbor of y_α), then $P\left(x_a \mid y_\alpha, \prod_{\beta \in V_M} s_{b,\beta} \right) = (1 - p_e^N)(1 - p_e^L) + (1 - p_e^N)(1 - p_e^L)$

   This condition implies that either the link was not noisy (not "erroneously observed"), in which case the correspondence is correct, or that the link was actually noisy, in which case the correspondence is in error because there is no link here.

Therefore, we can write:

$$P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right)$$
$$= [\varepsilon]^{D_{a,b}\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta}} \times [1-\varepsilon]^{(1-D_{a,b})\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta}} \times [1-\varepsilon]^{D_{a,b}\left(1-\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta}\right)} \times [\varepsilon]^{(1-D_{a,b})\left(1-\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta}\right)}$$
$$= \varepsilon\left[\frac{\varepsilon}{1-\varepsilon}\right]^{2D_{a,b}\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta} - \sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta} - D_{a,b}}$$

Hence,

$$P(x_a \mid y_\alpha, S) = B_a \prod_{b\in V_D} P\left(x_a \mid y_\alpha, \prod_{\beta\in V_M} s_{b,\beta}\right) = B_a \varepsilon^{|V_D|}\left[\frac{\varepsilon}{1-\varepsilon}\right]^{\sum_{b\in V_D} 2D_{a,b}\sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta} - \sum_{\beta\in V_M} s_{b,\beta}M_{\alpha,\beta} - D_{a,b}}$$
$$= K'_a T_\alpha \exp\left\{\mu'\left(\sum_{b\in V_D}\sum_{\beta\in V_M} D_{a,b}s_{b,\beta}M_{\alpha,\beta}\right)\right\}$$

where $K'_a = B_a\varepsilon^{|V_D|}\exp\{-\mu\cdot c(a)\}, T_\alpha = \exp\{-\mu\cdot c(\alpha)\}, \mu' = 2\ln\frac{\varepsilon}{1-\varepsilon}$, and $c(x)$ is a connectivity index of the node: $c(a) = \sum_{b\in V_D} D_{a,b}, c(\alpha) = \sum_{\beta\in V_M} M_{\alpha,\beta}$

Hence, the log-likelihood function $L(S) = \log P(G_D \mid G_M, S)$ can be expressed as

$$L(S) = \sum_{a\in V_D}\log\left\{\sum_{\alpha\in V_M} K'_a T_\alpha \exp\left\{\mu'\left(\sum_{b\in V_D}\sum_{\beta\in V_M} D_{a,b}s_{b,\beta}M_{\alpha,\beta}\right)\right\}\right\}$$

This problem cannot be directly solved, but we can apply **expectation maximization (EM) algorithm**. The **maximization step** applies *weighted log-likelihood function*

$$\Lambda\left(S^{(n+1)} \mid S^{(n)}\right) = \sum_{a\in V_D}\sum_{\alpha\in V_M} P\left(y_\alpha \mid x_a, S^{(n)}\right)\ln P\left(x_a \mid y_\alpha, S^{(n+1)}\right)$$

to find an iterative solution: $S^{(n+1)} = \arg\max_S \Lambda\left(S \mid S^{(n)}\right)$. Denoting $Q_{a,\alpha}^{(n)} = P\left(y_\alpha \mid x_a, S^{(n)}\right)$, we get:

$$\Lambda\left(S^{(n+1)} \mid S^{(n)}\right) = \sum_{a\in V_D}\sum_{\alpha\in V_M} P\left(y_\alpha \mid x_a, S^{(n)}\right)\ln P\left(x_a \mid y_\alpha, S^{(n+1)}\right)$$
$$= \sum_{a\in V_D}\sum_{b\in V_D}\sum_{\alpha\in V_M}\sum_{\beta\in V_M} Q_{a,\alpha}^{(n)}\left\{\mu' D_{a,b}s_{b,\beta}M_{\alpha,\beta}\right\} + \sum_{a\in V_D}\sum_{\alpha\in V_M} Q_{a,\alpha}^{(n)}\left\{\ln K'_a + \ln T_\alpha\right\}$$
$$= \sum_{a\in V_D}\sum_{b\in V_D}\sum_{\alpha\in V_M}\sum_{\beta\in V_M} Q_{a,\alpha}^{(n)}\mu\left\{\left(2D_{a,b}-1\right)s_{b,\beta}M_{\alpha,\beta}\right\} + \ln K'_a\cdot|V_D|$$

To maximize weighted log-likelihood function, we then need to maximize

$$\sum_{a\in V_D}\sum_{b\in V_D}\sum_{\alpha\in V_M}\sum_{\beta\in V_M} Q_{a,\alpha}^{(n)}\left\{\left(2D_{a,b}-1\right)s_{b,\beta}M_{\alpha,\beta}\right\} = \sum_{a\in V_D}\sum_{b\in V_D}\sum_{\alpha\in V_M}\sum_{\beta\in V_M} Q_{a,\alpha}^{(n)}M_{\alpha,\beta}\left(s_{\beta,b}\right)^T\left(2D_{a,b}-1\right)^T$$
$$= tr\left\{Q^{(n)}MS^T\left(2D-\Omega\right)^T\right\} = tr\left\{\left(2D-\Omega\right)^T Q^{(n)}MS^T\right\}$$

where Ω is a matrix of all 1's and $Q_{a,\alpha}^{(n)} = P\left(y_\alpha \mid x_a, S^{(n)}\right)$.

Hence, we get to maximize: $tr\left\{(2D-\Omega)^T Q^{(n)} M S^T\right\}$

Hence, we need: $S^{(n+1)} = \arg\max_S tr\left\{(2D-\Omega)^T Q^{(n)} M S^T\right\}$

Let $Z = (2D-\Omega)^T Q^{(n)} M$ - this is a $|V_D| \times |V_M|$ positive-definite matrix. Then we can perform singular-value decomposition $Z = U_D \cdot \Delta \cdot U_M^T$, where U_D is $|V_D| \times |V_D|$ orthogonal and U_M is $|V_M| \times |V_M|$ orthogonal matrix, and Δ is a $|V_D| \times |V_M|$ matrix with zero non-diagonal elements and non-zero diagonal ones. If we let E be a matrix obtained from Δ by replacing diagonal elements with 1's, then a matrix $R = U_D \cdot E \cdot U_M^T$ maximizes $tr\left\{ZS^T\right\} = tr\left\{(2D-\Omega)^T Q^{(n)} M S^T\right\}$. Then, we choose the following rule:

$$s_{a,\alpha}^{(n+1)} = \begin{cases} 1, \text{ if } R_{a,\alpha} = \arg\max_b R_{b,\alpha} = \arg\max_\beta R_{a,\beta} \\ 0, \text{ otherwise} \end{cases}$$

The **expectation step** then updates the model probabilities, and the process is repeated. Using the Bayes rule, the iterative conditional probability update is as follows:

$$P\left(y_\alpha \mid x_a, S^{(n+1)}\right) = \frac{P\left(x_\alpha \mid y_a, S^{(n)}\right)\pi_\alpha^{(n)}}{\sum_{\alpha \in V_M} P\left(x_\alpha \mid y_a, S^{(n)}\right)\pi_\alpha^{(n)}},$$

where $\pi_\alpha^{(n)} = P\left(y_\alpha \mid S^{(n)}\right) = \frac{1}{|V_D|}\sum_{a \in V_D} P\left(y_\alpha \mid x_a, S^{(n)}\right) = \frac{1}{|V_D|}\sum_{a \in V_D} Q_{a,\alpha}^{(n)}$. We also update

$$Q_{a,\alpha}^{(n+1)} = \frac{\exp\left[\mu \sum_{b \in V_D}\sum_{\beta \in V_M}(2D_{a,b}-1)s_{b,\beta}M_{\alpha,\beta}\right]\pi_\alpha^{(n)}}{\sum_{\alpha' \in V_M}\exp\left[\mu \sum_{b \in V_D}\sum_{\beta \in V_M}(2D_{a,b}-1)s_{b,\beta}M_{\alpha',\beta}\right]\pi_{\alpha'}^{(n)}} = \frac{\exp\left[\mu(2D-\Omega)\cdot S^T \cdot M'\right]_{a,\alpha}\pi_\alpha^{(n)}}{\sum_{\alpha' \in V_M}\exp\left[\mu(2D-\Omega)\cdot S^T \cdot M'\right]_{a,\alpha'}\pi_{\alpha'}^{(n)}}.$$

Using the described approach, we can find the mapping $S_M{}^* = \arg\max_S P\left(G_D \mid G_M, S\right)$ from the data network to every model network from the hypotheses library. The solution to the adversary recognition problem is then to find a model network $G_M{}^* = \arg\max_{G_M} P\left(G_D \mid G_M, S_M{}^*\right)$. The final solution to the adversary identification problem is then a pair $< G_M{}^*, S^* >$ of the organization structure (network) and the mapping from the actors to the positions in this organization.