June 2007 Volume 2 Number 1 ISSN 1557-6418

JAIF 07

Journal of Advances in Information Fusion

A semi-annual archival publication of the International Society of Information Fusion

Regular Papers | Page

From the Editor-in-Chief

Are We Making Progress?

www.isif.org

ISIF

INTERNATIONAL SOCIETY OF INFORMATION FUSION

The International Society of Information Fusion (ISIF) seeks to be the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied INFORMATION FUSION technologies. Technical areas of interest include target tracking, detection theory, applications for information fusion methods, image fusion, fusion systems architectures and management issues, classification, learning, data mining, Bayesian and reasoning methods.

Manuscripts are submitted at http://jaif.msubmit.net. If in doubt about the proper editorial area of a contribution, submit it under the unknown area.

Journal of Information Advances in Information Fusion (ISSN 1557-6418) is published semi-annually by the International Society of Information Fusion. The responsibility for the contents rests upon the authors and not upon ISIF, the Society, or its members. ISIF is California Nonprofit Public Benefit Corporation at P.O. Box 4631, Mountain View, California 94040.

From the Editor-In-Chief:

June 2007

Are We Making Progress?

After nine years of success with the International Conferences on Information Fusion (ICIF) and a year of the Journal of Advances in Information Fusion (JAIF), the question arises: Are we, the community of researchers in the area of information fusion, making progress?

With all of the focus of today's corporate management on performance metrics, knowledge points, balanced scorecards, etc., I am surprised that this question has not been raised earlier. However, a natural question does follow. How does one define progress in information fusion? For an individual researcher, progress can be defined as movement toward the goal of an ideal solution to the problem under study. Defining progress with respect to the community of researchers in information fusion presents a challenge.

One measure of the progress in information fusion is the developmental activity with reference to the commercial opportunities created thereby or to the promotion of the material well-being of the public through the goods, techniques, or facilities created. In other words, progress in information fusion can be measured by the number of products that include information fusion and their impact on society.

A second measure of the progress in information fusion is the maturity of solutions for problems in information fusion. Mature solutions involve well established and accepted approaches for which the cost and benefits of the implemented solution are well understood. Documented design methods are also part of a mature solution. An example of a problem with a mature solution is the tracking of maneuvering targets. The Interacting Multiple Model (IMM) estimator is well-accepted as the best approach to tracking of maneuvering targets when the computational cost of the algorithm is considered [1, 2], and design methods for application of the IMM estimator are emerging [3, 4].

I am sure that the research community has knowledge of other successful examples of progress in information fusion. I encourage researchers to consider progress in information fusion and document the progress through papers and special sessions at the annual International Conference on Information Fusion. This might include a series of papers documenting the relative performances of competing solutions or a paper documenting the application of information fusion in a commercial product. Such examples might be appropriate for a special issue of the *Journal of Advances in Information Fusion*.

William Dale Blair
Editor In Chief

[1] Bar-Shalom, Y. and Blair, W. D.
Introduction.
In Y. Bar-Shalom and W. D. Blair (Eds.), *Multitarget-Multisensor Tracking: Applications and Advances*, vol. III, Dedham, MA: Artech House, 2000.

[2] Blair, W. D. and Bar-Shalom, Y.
Tracking maneuvering targets with multiple sensors: Does more data always mean better estimates?
IEEE Transactions on Aerospace and Electronic Systems, **32**, 1 (Jan. 1996), 822–825.

[3] Kirubarajan, T. and Bar-Shalom, Y.
Kalman filter versus IMM estimator: When do we need the latter?
IEEE Transactions on Aerospace and Electronic Systems, **39**, 4 (Oct. 2003).

[4] Blair, W. D.
Design of nearly constant velocity filters for tracking maneuvering targets.
Proceeding of the 11th International Conference on Information Fusion, Cologne, Germany, July 2008, 1452–1457.

INFERD and Entropy for Situational Awareness

MOISES SUDIT
MICHAEL HOLENDER
ADAM STOTZ
TERRY RICKARD
RONALD YAGER

As technology continues to advance, services and capabilities become computerized, and an increasing amount of business is conducted electronically, there is an interesting need for real-time decision-making systems with many capabilities in various domains. In this paper we introduce INFERD (INformation Fusion Engine for Real-time Decision-making), an adaptable information fusion engine which performs fusion at levels zero, one, and two to provide real-time situational assessment. The advantages to our approach are threefold: (1) The level of abstraction in which the analyst interacts with the engine, (2) the speed at which the information fusion is presented and performed, and (3) our ability to give the user the choice to disregard ad-hoc rules or a priori parameters, which have both advantages and disadvantages. We present both a parameterized approach founded in statistical mechanics theory and a non-parameterized approach using concepts in entropy as understood in the context of information theory.

Manuscript received October 12, 2005; revised June 29, 2006; released for publication January 4, 2007.

Refereeing of this contribution was handled by Stephane Paradis.

Authors' addresses: M. Sudit and M. Holender, Center for Multi-source Information Fusion, SUNY at Buffalo, Buffalo, NY; A. Stotz, Calspan-UB Research Center, 4455 Genesee Street, Buffalo, NY; T. Rickard, Lockheed Martin, 4637 Shoshone Drive, Larkspur, CO; R. Yager, Machine Intelligence Institute, Iona College, New Rochelle, NY.

1. INTRODUCTION

1.1. INFERD

INFERD was created in the context of cyber security [25] as a decision aid tool to improve the analyst understanding of the situation and ultimately expedite their processing. To cope with the volumes and data rates of current sensed environments such as cyber security and others, decision aid tools must provide their assessment of the situation in a very time efficient manner. In most cases, this time constraint eliminates the possibility of some non polynomial approaches such as optimal inexact graph matching and must instead rely on heuristics to provide *good* results in a timely manner. INFERD's hierarchical fusion approach was developed to do such a task. The two forms of input to INFERD are in the form of a *Guidance Template* (a priori), and sensor data (runtime). The actual fusion process, both a parameterized and unparameterized approach, and how these two forms of input produce valuable output will be addressed throughout the paper.

INFERD's unique approach to Information Fusion can arguably provide these basic advantages: (1) The flexibility of the system to be transitioned to different environments, (2) the level of abstraction of the output of the system compared to the specificity of the models, and (3) the rate at which INFERD can process data and produce results.

1.2. Parameterization v. Non-Parameterization

Parametric approaches are typically general enough to be applied to a variety of environments. Deploying a parametric system to networks of varying topologies usually consists of retraining the system on test data obtained for that environment. The problem of systems using parametric approaches based on training data sets is a sensitive one that can often lead to large numbers of false positives or inaccuracies when working on data not in the training set. The two classical cases of *overfitting* and *overtraining* can arise when a parameter vector v is obtained that configures the system to be very accurate on the training data but generalizes poorly to non-training data. The accuracy/generality tradeoff problem is a well-studied one in many academic areas such as statistics (known as the *bias-variance tradeoff* [15]), Bayesian inference (known as *penalized likelihood* [6], [19]), and in pattern recognition/machine learning (known as *minimum message length* [39]).

Rule-based approaches are expressive in the way that the security analyst provides system configuration. Rules are created or modified in accordance with the environment in which the system is running. This methodology, however, has arguable deficiencies in that every possible condition for the environment in which the system is running must be accounted for in its rule set.

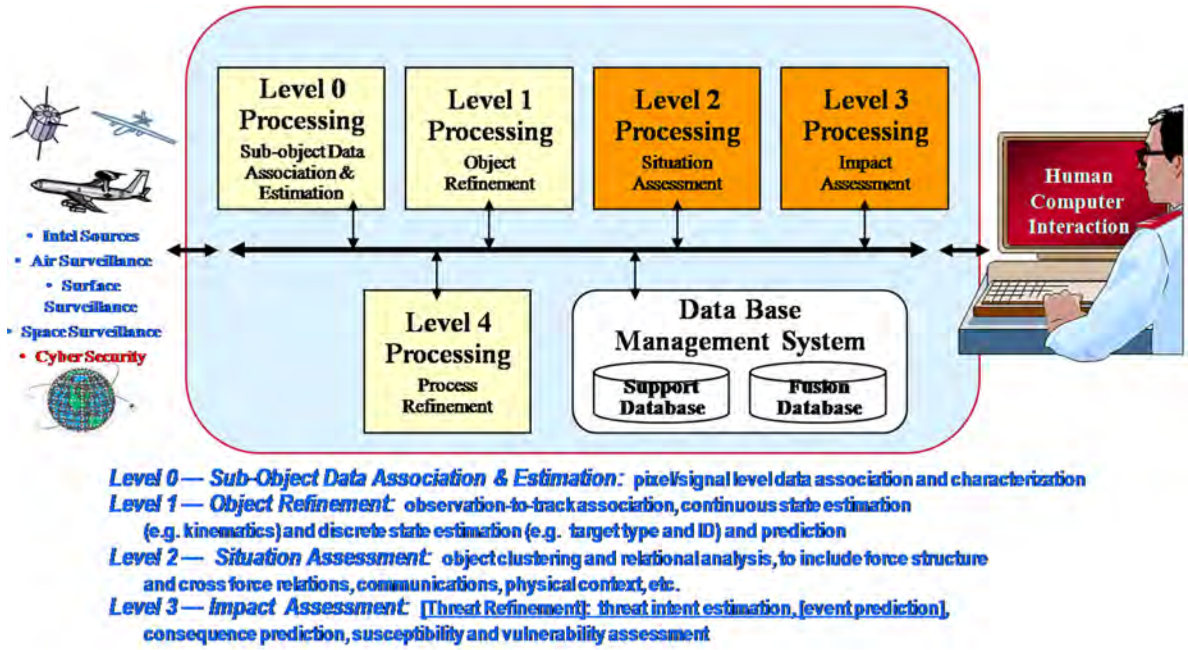


Fig. 1.3.1. JDL fusion model.

TABLE 1.2.1
Methodology Comparison

Methodology	Advantages	Disadvantages
Parametric	Portability Generality	Need for a priori training process Accuracy variance
Rule-Based	Expressiveness	Accuracy variance Rule sets become unwieldy

Many domains provide very dynamic systems; on any given day there may be topology changes in the templates (to be explained later), patches applied making certain vulnerabilities dissipate or even materialize as a side effect, discovery of new exploits, realization of new attacking strategies, the list goes on. With such frequent changes in the environment, the rule sets can quickly become too complex and unwieldy to remain synchronized with the latest changes. As rules are left out and changed the chances of system accuracy being maintained diminish and the system becomes legacy providing no benefit to the present.

See Table 1.2.1 for an overview of the advantages and disadvantages of parametric and rule-based systems. In summary, we wish to solve the problem of performing real-time detection of complex, multistage, systems in such a fashion that minimizes a priori settings and is sustainable across the breadth and frequency of changes that can occur within the deployment environment.

1.3. Information Fusion Overview

In order to address the problems in data fusion, the US Joint Directors of Laboratories (JDL) developed a five-level Data Fusion Model, shown in Figure 1.3.1 [33]. Level 1 on Object refinement seems to have received the most attention. Level 1 processing functions include: data alignment, association, tracking, and identification. Less mature are Level 2 processing [16] [30], situation assessment, which seeks a higher level of inference above Level 1 processing, and Level 3 processing which performs threat assessment. Threat assessment is an iterative process of fusing the combined activity and capability of enemy forces to infer their intentions and assess the threat that they pose. Level 1 is very often called "low-level" processing, and the others as "high-level" processing.

Higher level fusion problems are generally more difficult than Level 1 because they involve higher dimensionality corresponding to the relationships among entities identified at Level 1. Higher level fusion also concerns modeling behavior of aggregate entities, through the understanding of their individual behaviors and relationships. Some commonly recognized relationships are spatio-temporal relationships, part/whole relationships, organizational relationships, various casual relationships, semantic relationships, similarity relationships, etc.

- Level 0: (Sub-Object Data Association & Estimation)—This deals with signal level data association and characterization.
- Level 1: (Object Refinement)—This deals with track-to-truth and track-to-track association, kinematics estimation and target type and ID prediction.
- Level 2: (Situation Assessment)—This deals with object clustering and relational analysis, to include structure and relations, communications and physical context.

- Level 3: (Impact Assessment or Threat Assessment)—This deals with threat intent estimation, consequence prediction, susceptibility and vulnerability assessment.
- Level 4: (Process Refinement)—This is an adaptive search and processing step.

Level 0 is a special case of Level 1, where entities are signals/features. Level 3 is a special case of Level 2, where relations are cost impact. Level 4 is a special case of Resource Management. Here we will be looking at computational techniques applied in Level 2 and Level 3 data fusion.

1.4. Approaches to High Level Data Fusion

There have been many approaches to performing high level data fusion (L2+) which have been developed, extended, modified, and refined over the years. Many of these approaches which will be discussed shortly have seen success through modification to specific problems, but no single approach has proven to be a single solve-all solution. Every approach has its advantages and disadvantages and the key is to exploit these properties in an optimal fashion for the problem at hand. The various INFERD terms used within this section will be defined and discussed in Section 2 of this paper.

1.4.1. Knowledge Based Expert Systems

Knowledge Based Systems (KBS) are computer systems that contain stored knowledge and solve problems like humans would. KBSs are drawn from the broad discipline of artificial intelligence (AI) where a knowledge base is defined in terms of rules, facts and meta-knowledge. These systems are utilized for combining expert knowledge and sensor information to form a knowledge base which is used for reasoning about the current situation or threat. They are symbolic programs which solve problems by symbol manipulation. Base techniques of knowledge-based systems are rule-based techniques, inductive techniques, hybrid techniques, symbol-manipulation techniques, case based techniques, qualitative techniques, model-based techniques and temporal reasoning techniques.

There are many advantages of using knowledge based expert systems. In expert systems the changes in field of interest are well-tracked and increase the expert's ability and efficiency. In addition to advantages, there are some limitations to knowledge based expert systems. Their knowledge is from a narrow field of interest and they don't know the limits to which it can extend. There can be many exceptions and this can increase the size of knowledge base and eventually the running time of the algorithm. The answers from the expert systems are not always correct, hence the advice has to be analyzed before actually applying it. The expert systems don't have common sense and so all of the self-evident checking has to be predefined. Some examples of applied expert systems for decision support can be found in [3], [42], [1], and [2].

It is typical for expert based knowledge to be required in the classification of observables into detailed domain specific concepts. Otherwise, complex inference processes and a large ontology must be defined, making the solution intractable for time critical applications. The Guidance Templates in INFERD contain *Feature Nodes* that define a set of constraints which (when satisfied) map sensor data into events. This allows INFERD to take advantage of the speed efficiencies of classification in the same manner as KBS for low level fusion, but does not require the definition of complex and interrelated rules needed for higher levels of fusion.

1.4.2. Graph Based Matching Techniques

Graph based matching techniques [10] have been used as a powerful tool for a number of decades, but most notably in the early eighties. Graph based pattern recognition or graph matching is the process of finding a correspondence between the nodes and the edges of two graphs that satisfies some constraints ensuring semantic and syntactic relationships. Graph matching techniques are divided into two broad categories: (1) the exact graph matching method that requires stringent correspondence among the graphs to be matched and (2) the inexact graph matching method, where two graphs can be compared even though they are semantically or topologically different.

Graph matching has been used in high level fusion to abstract complex situations from large amounts of data. The ease of representation of graph patterns and the cognitive advantages of representing situations as a matching between graph based patterns has made the approach increasingly popular with the introduction of new high powered computers. The fundamental problem however, is the theoretical complexity of the graph matching problem. The matching problems mentioned above are all *NP-complete*, with the exception of attributed graph matching in which the nodes are guaranteed to have distinct attributes. In this case the problem becomes polynomial.

To take advantage of the expressiveness and ease of defining graph based patterns, the INFERD Guidance Template has adopted a graph based structure. The structure will be detailed in Section 2, but the similarities stop here in terms of INFERD's fusion process in comparison to graph matching techniques. Because of the theoretical complexities of the matching process, the research team investigated and developed an alternative approach. Remember that the motivation was to provide timely hypothesis generation. These high level hypotheses can very well be linked to graph matching patterns, effectively producing a ranked list of patterns to be matched. This linkage between INFERD and graph matching techniques makes the matching problem more time tractable when there are large numbers of patterns to be matched to a given data graph.

1.4.3. Bayesian Belief Networks

In recent years there has been a surge in use of Bayesian Belief Networks (BBNs) to solve the problems of situation and impact assessment. BBNs have become a popular knowledge inference scheme for probabilistically related evidence and inferences. Their attractiveness lies in the fact that BBNs provide both a sound theoretical framework and a conceptually simple interpretation for representing and manipulating knowledge graphically in a probabilistic domain. BBNs are directed acyclic graphs (DAG), which provide a framework for a structured representation of knowledge about uncertain quantities [12] where nodes and arcs represent conditional probabilistic dependency between variables.

The sound theoretical foundation of BBNs in Bayesian theory can be either an advantage or a disadvantage depending upon the application. In well known environments, BBNs can work very well, however this is not the case in highly dynamic or unknown environments. BBNs are highly dependent upon, and only as good as, the conditional probability tables which are defined. In unknown environments where some or all of these conditional probabilities are not known, or can only be grossly estimated, the accuracy of the BBN will suffer. INFERD does not rely on likelihood functions for this reason. An example of a BBN used in a decision support problem can be seen in [14].

1.4.4. Fuzzy Logic

Fuzzy Logic (FL) is an inferencing methodology that is directed toward vague relationships between evidence and assertions. Fuzzy inference is the process of formulating the mapping from a given input to an output using FL. Because of its multidisciplinary nature, fuzzy inference systems are associated with a number of names, such as fuzzy-rule based systems, fuzzy expert systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers, and simply (and ambiguously) fuzzy systems.

Fuzzy logic systems have the advantage of introducing more flexibility into the processing layer to symbolic manipulations or calculations through the definition of *fuzzy membership functions* which can be useful in making decisions in light of information that is imprecise and/or incomplete. Fuzzy logic techniques have become popular to address various processes for multisensor data fusion. Examples include the following: fuzzy membership functions for data association [29] [34], evaluation of alternative hypotheses in multiple hypothesis trackers, fuzzy-logic-based pattern recognition (target identification) [18], and fuzzy inference schemes for sensor resource allocation [23].

A future extension to INFERD could incorporate fuzzy logic into the mapping process of observables into events. Currently observables are mapped to events on a $\{0,1\}$ basis, this could be extended to allow multiple mappings in a fuzzy sense ([0,1]) relaxing INFERD's fusion process to be a multi-hypothesis evaluation system.

1.4.5. Genetic Algorithms

Genetic Algorithms (GAs) are a type of evolutionary algorithm which are the result of studying the natural adaptation of living organisms and are a way of incorporating a similar adaptation into computer systems. They try to mimic environmental factors such as reproduction, random variation, competition, and selection of competing individuals. Genetic algorithms are now widely applied in science and engineering as adaptive algorithms for solving practical search problems particularly suited to multidimensional data where global solutions are found within multiple local minima.

In the information fusion community, GAs are being utilized in many different applications relative to the threat assessment. One of the challenges in a GA based course of action (COA) optimization system is the ability to generate and evaluate thousands of candidate COAs in order to generate the best solution. This consists of two key aspects: the ability to encode the enemy COA into a set that comprises the GA population under evaluation and the ability to quickly evaluate each COA to determine which survives to the next generation. Because the key to success for a GA is evaluating many candidates, it is necessary to be able to abstract the battlefield in order to be able to both encode the situation as a solution string and to be able to rapidly war game each COA in order to evaluate it. Examples of GAs used in high level information fusion problems can be found in the following references: [32], [24], [8], [4], and [5].

As stated in Section 1.4.2, there will be a future need for generation of Template Graphs within certain problem domains. Genetic Algorithms along with Graph Matching could provide a means for creating such templates.

1.4.6. Neural Networks

Artificial Neural Networks (ANNs) are computational systems premised upon the principles of biological neural systems. In general, this means that ANNs are characterized by having many low-level processing units with a high degree of interconnectivity. It is difficult to characterize the field of ANNs succinctly, because the approaches and the results are so diverse. Recently fuzzy logic is been used extensively along with neural networks [20] [9]. Fuzzy logic uses approximate human reasoning in knowledge-based systems while the neural networks aim at pattern recognition, optimization and decision making. A combination of these two technological innovations delivers better results than when used independently.

The advantage of ANNs is that when trained appropriately they produce accurate results for similar problems without the need of any type of parameterization.

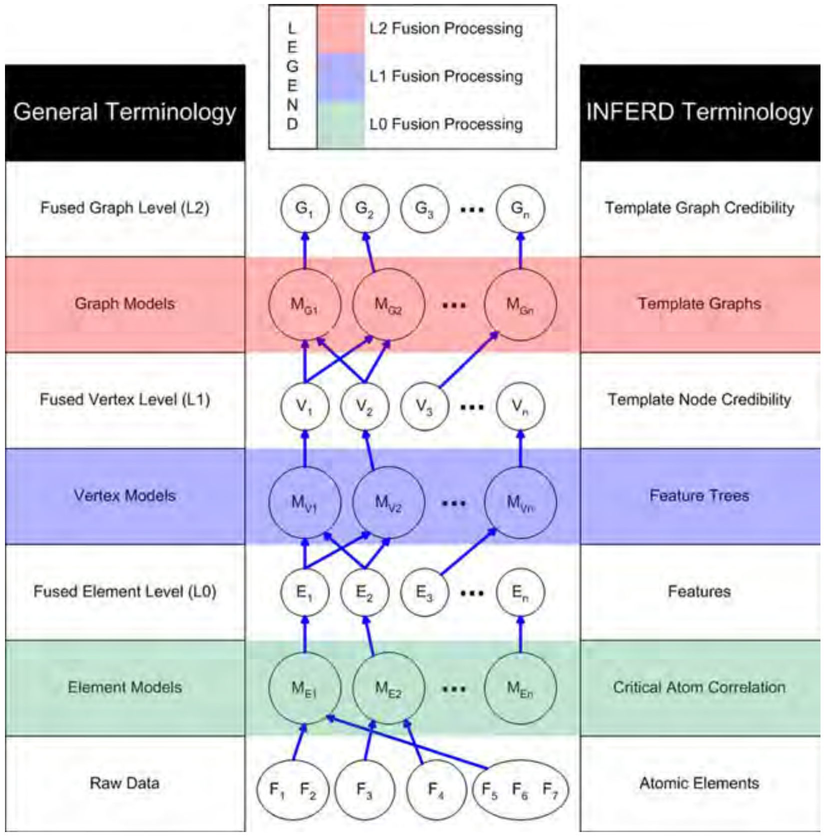


Fig. 2.1.1. General fusion framework.

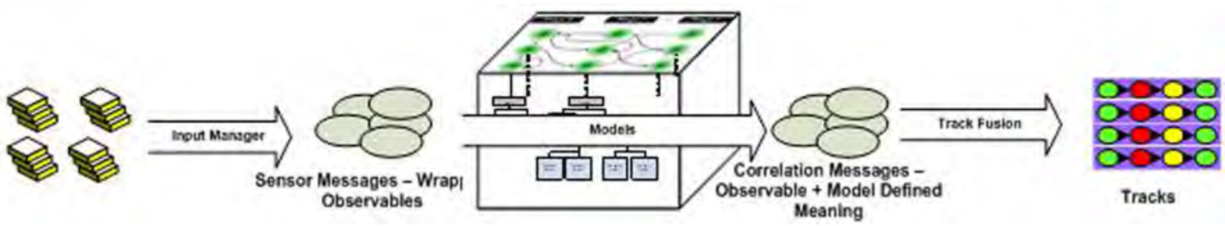


Fig. 2.1.2. INFERD high level information flow diagram.

The disadvantage of neural networks is that their results are contingent upon their level of training. Often in the information fusion community realistic data sets are unavailable or scarce at best. In these situations, neural networks will not be the best solution approach.

Wang and Archer [40] have proposed a neural network based fuzzy set model to support organizational decision making under uncertainty. The model makes use of single back propagation neural network to generate a crisp fuzzy membership function. The authors [41] have used a connectionist approach to multi criteria decision making.

2. THE INFERD ENGINE

2.1. General Fusion Methodology in INFERD

Great care has been taken in the processing structure of the INFERD engine to minimize necessary computation time. In many domains, data rates produced by sensors are computationally intensive to process, so there is not much overhead to spare. The fusion being performed in INFERD is bottom-up in a hierarchical fashion at Levels 0, 1, and 2 according to the JDL model for information fusion [17]. Figure 2.1.1 shows the general terminology and how our system and terminology maps. In the INFERD fusion framework, each subsequent level of fusion feeds off of the previous levels output. This is not a requirement of the JDL model, but suited our system and its network and sensor environment well.

For a summary of the overall fusion processes in INFERD consider the information flow diagram in Figure 2.1.2 as it would apply to the cyber security problem. In this diagram, we can see the flow of basic sensor information, to ultimately, a set of tracks of that information.

The first stage of processing, performed by the *Input Manager*, wraps incoming sensed observables (sensor

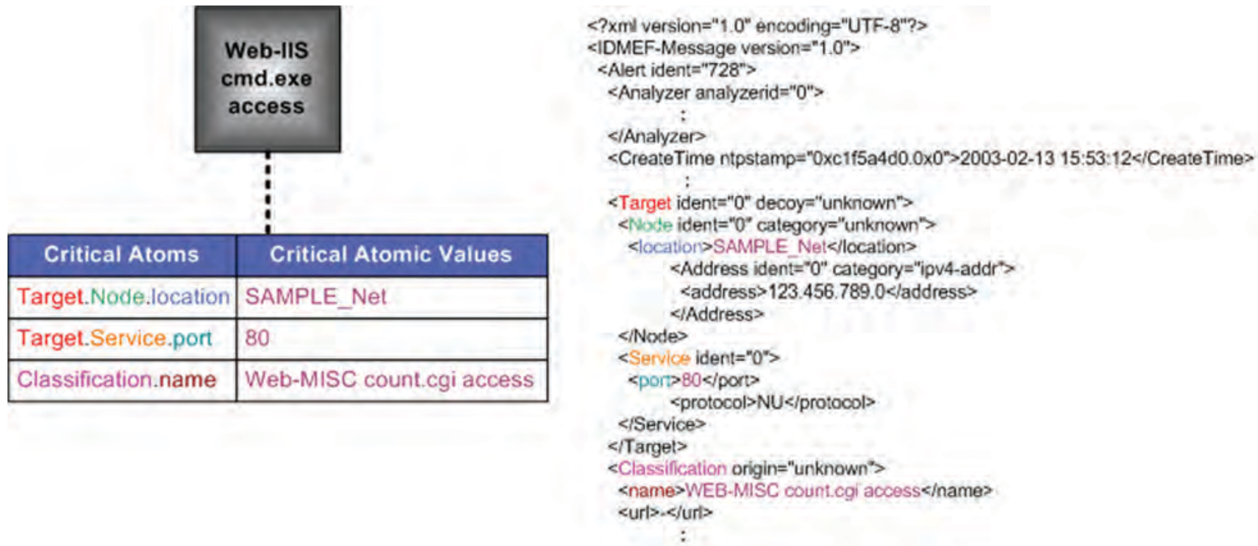


Fig. 2.1.1.1. Example feature node definition.

output) into *Sensor Messages*, a format which is understood by the *Model* and *Track Fusion* Processes. By isolating the fusion processing from the I/O architecturally, INFERD is able to fuse information from sensors of radically different formats and types but still define the *Guidance Templates* in a common language. In the case of cyber security the Input Manager would transform the sensor alerts into an XML object and provide a common referencing method to retrieve values from the object.

The second stage of processing, performed by the Model Fusion Process, assigns model-based meaning to the Sensor Messages. In this stage of processing, *Guidance Templates*, or a priori models, classify the Sensor Message into a higher level event type and expose valid relationships to other previously classified alerts. This newly added information to the Sensor Message, forming a *Correlation Message*, is then understood by and sent to the *Track Fusion Process*. In the cyber security example, this process might reference the target IP address, and signature found within the Sensor Message and classify it as a Scanning Reconnaissance attack on the corporate web server. It would also add the knowledge that this could be a predecessor step for a number of intrusion type attacks on that machine.

The third stage of processing, performed by the *Track Fusion Process*, takes the Correlation Message and fuses it to the existing runtime set of tracks already in existence, possibly resulting in a new track. By fusing piecemeal event steps into unified event tracks, INFERD offers similar advantages to ground target tracking systems, but in a multi-int environment and in new fusion application domains. By analyzing tracks instead of low level sensor events, the analyst is able to prune his search space much more efficiently and have a better understanding of the situation when it is time to make decisions.

Sections 2.1.1 through 2.1.3 will now detail the fusion process in more specificity.

2.1.1. Level 0 Fusion—The Atomic Element

Level 0 fusion is the first processing that occurs once a piece of information is accepted as input into the engine. This piece of information can be of any type such as numeric, text, or file based information. Input to the L0 process is taken in raw data form and then necessary information is extracted by generalized data objects which connect the abstracted data types to the actual sensor message data values. In many instances, more information is taken into the system than is required to analyze what is happening within the desired domain. These desired pieces of information are arranged into Feature Nodes in a tree-structure as understood in basic graph theory. This structure is described later in the fusion discussion.

Once a piece of information (discrete sensor message) is published to a Feature Node, the Critical Atomic Values contained in the message are checked against those specified in the Feature Node in the form of constraint satisfaction. These constraints can take a number of forms such as greater than, less than, equality, string equality, regular expression pattern matching, etc. If all of the defined constraints are satisfied then that Feature Node becomes asserted. The credibility values of Feature Nodes are binary $(0, 1)$ with respect to their assertion state.

In addition to specifying Critical Atoms and Critical Atomic Values, each Feature Node has a specified lifetime associated with it. These lifetimes indicate the maximum amount of time the Feature Node should stay in the asserted state since the time of the last piece of incoming information correlated to it. If Feature Nodes did not de-assert themselves in some fashion, the credibility values of the Template Graphs containing them would never decrease and there would be no temporal aspect to the INFERD engine. In many domains, it is important that the relative timing of incoming information be considered as to its relevant effects on the system at hand.

Whenever a Feature Node changes state, the parent nodes in the Feature Tree containing that node and subsequently the Template Node specified by that Feature Tree re-calculate their credibility values. This credibility calculation will be discussed in Section 2.1.2 as the L1 fusion process, but it is important to note the bottom up processing which occurs in INFERD. The publish-subscribe service for information input to the system saves a great deal of computation time by not performing the Critical Atomic Value comparisons for the possibly thousands of Feature Nodes which can ignore the alert.

2.1.2. Level 1 Fusion—The Feature Tree

Level 1 fusion processing or Template Node credibility calculation takes over once a Feature Node changes assertion state. The input to the L1 fusion process or Fused Element Level is the Feature Node which has changed assertion state, the Vertex Model is the Feature Tree containing that Feature Node, and the output or Fused Vertex Level of L1 fusion processing is the credibility value or estimated likelihood of occurrence of the Template Nodes who's Feature Tree contains that Feature Node which has changed assertion state. The calculation of L1 credibility values is inherent in the structure of the Feature Tree and the values of the Relation Nodes within that tree.

Every Relation Node specifies a function determining how its children relate to each other. We use Yager's Generalized Ordered Weighted Average (GOWA) function as a means of calculating the relation [43, 44, 45]. Assume $\{A_1, A_2, \ldots, A_n\}$ are n criteria of concern in this multi-criteria decision problem. These are the criteria described in the atomic elements above. Let us further assume that the values $a_1, a_2, \ldots, a_n$ represent credibilities associated with the above set A of n elements. We can then construct a function $F(a_1, a_2, \ldots, a_n)$ that will be used to aggregate its children at the relation node. Yager describes many properties of such a function. His OWA operators are designed by introducing two vectors B and W. Let B be an ordering vector that "rearranges" the credibilities $a_1, a_2, \ldots, a_n$ in descending order. Let W be a weighting vector such that $\sum_{i \in W} w_i = 1$, $w_i \geq 0$. In vector form, the OWA operator is expressed as $F(a_1, a_2, \ldots, a_n) = W^T B$. Numerical examples are shown in Yager's referenced papers. The theory is carried out further to describe a concept known as "attitudinal character" that describes the level of "ANDness and ORness" that the W vector takes on. The attitudinal character is described by the following: $AC(W) = \sum_{j=1}^{n} w_j (n-j)/(n-1)$. For example, if $W = [1, 0, 0, \ldots, 0]$, then $AC(W) = 1$ thus saying that we have the greatest possible "ORness" since this would give us a maximization function. This is true since we are multiplying W and B where only the first element of B would be considered (since $w_1 = 1$). Note the first element of B is $\max(a_i)$. Similarly, we have maximum "ANDness" when $W = [0, 0, \ldots, 0, 1]$; $AC(W) = 0$. Finally, we simply compute the average value when $W = [1/n, 1/n, \ldots, 1/n]$; thus $AC(W) = 1/2$. Such a general function has unlimited possibilities and can be applied to any domain using aggregation functions.

The Feature Tree used in INFERD consists of a Template Node at the "top" of the tree. Below it may be a series of child nodes. Each of these child nodes may be a series of child nodes to them (or grandchild nodes to the Template Node). The above GOWA function is used to describe the relationships between the child nodes and their respective parents. To calculate the values of the Template Node (or parent node as it is known in graph theory), INFERD begins with understanding of the child nodes at the very bottom of the tree, then it works its way upward. The "bottom-most" nodes of the tree are the pieces of information taken in via L0 fusion discussed in the above section. Once these binary values are obtained, we can apply the OWA function to obtain the probabilistic value of their parent nodes. This process continues up the tree structure until a value is figured for the Template Node and further used in the L2 fusion steps.

We will now introduce an example of a system that could be analyzed using INFERD. We will continue using this example throughout the paper. Airport Security is an increasingly important issue in today's society. There are many measures taken to prevent unsafe situations. We will present a somewhat simplified way to answer the question: Is this passenger of any danger to their fellow passengers? This question will be answered probabilistically through determining its Credibility Factor (discussed later in the paper). There are many different considerations in answering this question; to describe the Feature Tree, we will look into the verification of a passenger's identity. The node in the Template Graph is called "ID Verification." We will see later how this becomes a part of the Template Graph and how it interacts with other nodes. For now, let us look at its underlying Feature Tree such that we can obtain credibility for the node. For our understanding, let a higher credibility indicate a higher chance of this passenger being an immediate danger. Figure 2.1.2.1 provides a visual of the Feature Tree.

For simplified understanding of INFERD, we will consider only three measures taken to verify a passenger's identification prior to their boarding of a commercial aircraft. Applying the GOWA function on our Relation Nodes, we choose $W = [1/n, 1/n, \ldots, 1/n]$ to be our measure, hence we will take a weighted average of its immediate children, Risk Assessment, Photo ID, and Biometrics. When a passenger books a flight, they may be asked a series of personal questions that will lead to an assessment of their risk. When this is completed, the airline representative will then assess the risk based on

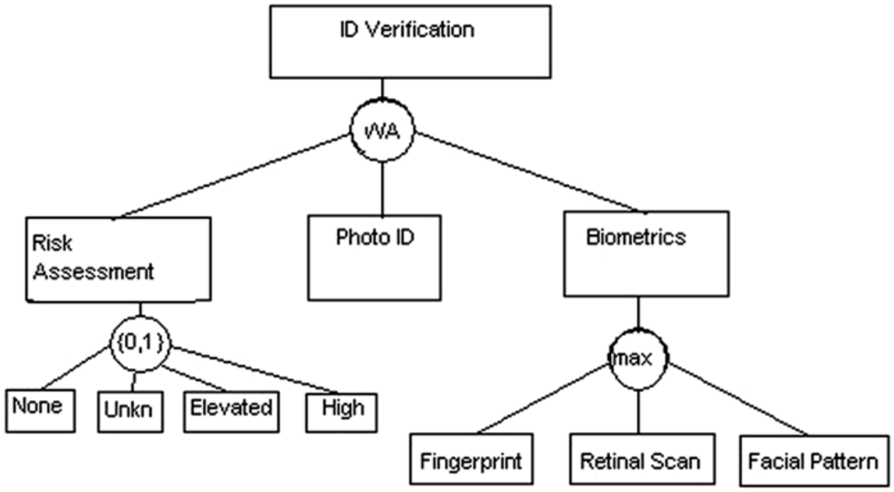


Fig. 2.1.2.1. The feature tree underlying the ID verification node.

the answers to the questions. To illustrate that any function may be used in this level of INFERD, we will introduce a binary function such that each of the four levels of risk assessment will be given a value in $\{0,1\}$ where $0 =$ the level of risk was not given to the passenger and $1 =$ the level of risk was given to the passenger. We will then take a weighted average of the binary values against the weights $[0.00, 0.33, 0.67, 1.00]$ for no risk, unknown risk, elevated risk, and high risk respectively. Next, when a passenger claims their boarding passes at the airport, they are asked to show their photo ID. If that ID matches all known information about the passenger, we give a 0 value to that node; conversely, if there is a discrepancy we will assign a value of 1. Finally, there is a system being worked on and nearly in place in most major airports called CAPPS II (Computer Assisted Passenger Prescreening System). CAPPS II takes biometric information about the passenger and attempts to verify their identity. The system will test fingerprints, retinal scans, and facial patterns of passengers. In our model, we will assign a 0 value if there is no problem identifying the passenger positively. However, if there is an issue with these, we will assign the value 1. Under the biometrics node we will take the maximum using the GOWA function by setting $W = [1,0,0,\ldots,0]$.

Let us assume that the passenger being screened when purchasing their tickets was given a risk assessment of "unknown." When they arrived at the airport, their photo ID matched up. However, when CAPPS II was used there was an identification discrepancy with the retinal scan and the facial pattern (the fingerprint appeared to be correct). The node for Risk Assessment would be given a value $0*0 + 1*0.33 + 0*0.67 + 0*1 = 0.33$. The photo ID node would have a value of 0. The biometrics node has a value of $\max\{0,1,1\} = 1$. Hence, we take the average to obtain the credibility of the Template Node ($j = 1$), ID Verification. $c_1 = (0.33 + 0 + 1)/3 = 0.443$. We will use this value going forward.

2.1.3. Level 2 Fusion—Template Credibility Calculation

Level 2 fusion or Situation Refinement is currently the highest level being implemented in the INFERD engine. The input or Fused Vertex Level to L2 are the credibility values of the Template Nodes, the model is the given template and the output or Fused Graph Level is an overall credibility value for that template. It is these credibility values coupled with the ranking of the templates that provides the system analyst with a situational estimation of their system's current environmental status.

Credibility Values exist for each node in the Template Graph (Feature Nodes and Template Nodes) and the Template Graph itself. While the methods of calculation of these values vary, the meanings of the values remain consistent. A credibility value is simply a likelihood of occurrence that INFERD produces. For Feature Nodes, this value is in $\{0,1\}$ because either the observable captured by that node was input to the system or it was not. For Template Nodes which can represent events, objects, or abstract concepts the value is in $[0,1]$ because this is a much more fuzzy process. The same argument is made for Template Graph credibility calculation as well.

The INFERD engine has imbedded into it, by the system user, templates specific to the given system being studied suggesting the way it works within its environment. Each Template Node may have an underlying Feature Tree that gives INFERD its credibility via L0 and L1 fusion described earlier. The functions applied to the children in the Feature Tree are chosen by the user, the following figure shows maximum and weighted average. The Template Nodes are then linked to each other as deemed reasonable to make up the Template Graph. See Figure 2.1.3.1 for an illustration.

These Template Nodes connect to one another to form a Template Graph. There are three types of nodes that can exist in the graph as defined by their links to other nodes. Extrinsic Nodes are those that have no

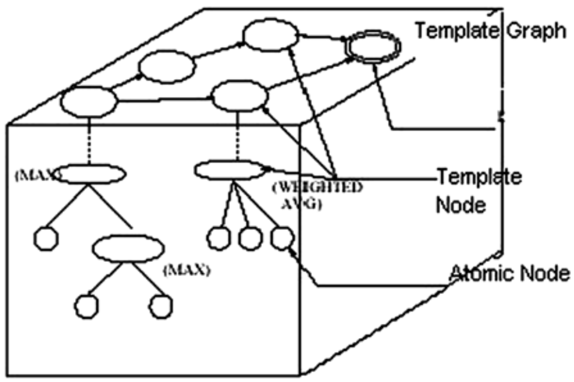


Fig. 2.1.3.1. Illustration of L1 and L2 fusion within INFERD.

precursor Template Nodes. Their credibility value or likelihood of occurrence is based solely on the Feature Nodes contained within its Feature Tree. Intrinsic Nodes are those that have one or more precursor nodes, and also are not reported on at any level by any substructure. These nodes can only be possible if triggered by another node connecting to it in the Template Graph; there is no underlying Feature Tree. Bi-trinsic Nodes are those that are reported on at some level by L0 fusion and also have precursor nodes. The credibility value of nodes of this class can leverage data from its Feature Trees along with its precursor nodes.

For example, let us say we have a cyber network alert system [35] being monitored by INFERD; we may have Attack Templates imbedded into the engine. Each of the Template Nodes would be some sub-situation that could imply a possible attack on ones network. Hence, underlying these Template Nodes would be the Feature Tree including the steps possibly leading up to this part of an attack happening. Note that there can, and most likely will be, many more than one single Template Graph being analyzed by INFERD at any given time.

Continuing our Airport Security example, we design a Template Graph containing seven nodes that each has an important contribution to understanding the credibility of a passenger's safety. INFERD stores many Template Graphs and analyzes them at the same time. Hence, in this example, individual passengers would have their own Template Graph. However, in many other domains there may not be a consistency among Template Graphs; there could be many with different factors. Figure 2.1.3.2 illustrates our Template Graph with node numberings in parentheses.

In our example, we will consider ID Verification among other actions taken by the passenger, most of which are understood in context. When in an airport, one is not allowed under law to speak of "terrorism," "bombs," "guns," etc. Hence, we include node 7 as "forbidden" words. Underlying each of these Template Nodes, there may be a Feature Tree giving a credibility factor denoted by c_j where $j = 1,\ldots,7$. Notice how some Template Nodes also have influences from other Template Nodes in the Template Graph. From above, we have $c_1 = 0.443$; let $c = [0.443, 0.55, 1, 0.01, 0.4, 0.1, 0.15]$. We will work with this Template Graph in the next sections.

Now we describe two approaches that can be used to determine the credibility factor of the Template Graph in the INFERD engine. The first method described will be a parametric approach with the advantages and disadvantages discussed above. The second approach will be the Entropy approach used to combat many of the drawbacks of the parameterized approach.

2.1.3.1. The Parameterized Approach for Credibility Factor

Our first L2 algorithm uses concepts in Statistical Mechanics. During the late 1800s, M. L. Boltzmann and J. W. Gibbs studied in the field of thermodynamics and pioneered what we now know as statistical mechanics. While thermodynamics (in the classical sense) deals with a single system called a macrostate, statistical mechanics studies the sub-components of this system, called microstates. Statistical mechanics is the application of probability theory to the field of mechanics for large populations of particles with respect to their motion subject to forces. The greatest benefit of such a methodology from a physics point of view is that statistical mechanics contains the ability to make macroscopic predictions based on microscopic properties. This ability lends itself directly to a Data Fusion system since raw data enters the system as microscopic properties and the desired result of Situation Awareness is a macroscopic prediction based on the raw data.

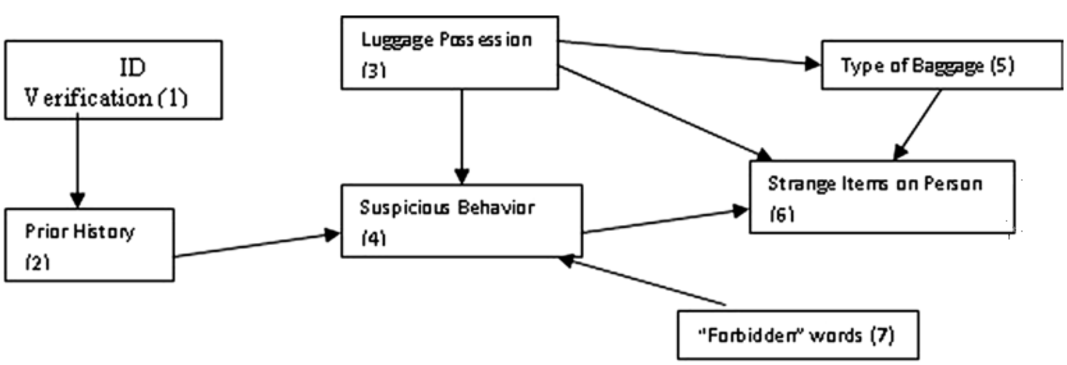


Fig. 2.1.3.2. Example of Template Graph.

One of the very first applications of statistical mechanics to optimization was in the field of Simulated Annealing. Simulated annealing is a generalization of a Monte Carlo method for examining the equations of state and frozen states of *n*-body systems [27]. The concept is based on the manner in which liquids freeze or metals recrystalize in the process of annealing. In an annealing process a melted material, initially at high temperature and disordered, is slowly cooled so that the system at any time is approximately at thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered and approaches a frozen ground state. The original Metropolis scheme was that an initial state of a thermodynamic system was chosen at energy E and temperature T, then by holding T constant the initial configuration is perturbed and the change in energy dE is computed. If the change in energy is negative the new configuration is accepted. If the change in energy is positive it is accepted with a probability given by the Boltzmann factor $e^{-dE/T}$. This process is repeated a sufficient number of times to give good sampling statistics for the current temperature, and then the temperature is decremented and the entire process repeated until a frozen state is achieved at $T = 0$. By analogy the generalization of this Monte Carlo approach to combinatorial problems is straight forward [21]. The current state of the thermodynamic system is analogous to the current solution to the combinatorial problem—the energy equation for the thermodynamic system is analogous to the objective function, and the ground state is analogous to the global minimum. Hence, this notion of simulated annealing can be used in optimization problems that are NP-hard as a brilliant heuristic approach. The basic components of simulated annealing are in statistical mechanics, thus showing a strong tie between the fields of statistical mechanics and optimization. We therefore recommend its use for our purposes in data fusion and as a heuristic for situation state estimation. Claude Shannon found deep links between information theory and thermodynamics. Following the same reasoning a possible link can be drawn between the probability of occurrence of the activity of track of hacker behavior in a noisy environment and the heating and cooling of a metal to a steady state. Thus we investigate this approach as applied to the problem of computer network security.

One of the more important results discovered by Gibbs and Boltzmann describes the probability of a microstate being within a certain energy state. We denote the energy state as E_s; under the assumption of the system at hand being independent of other systems, we can write the probability as follows:

$$P(E_s) = \frac{e^{-E_s/T}}{Z(T)}$$

where T denotes the temperature of the system and $Z(T)$ is a partition function which normalizes probabilities across all states such that $\sum_{s\in S} P(E_s) = 1$.

This concept has been used in various applications in Information Theory [31], Optimization [21], and Decision Theory [13]. The application of the Gibbs-Boltzmann Equation in INFERD begins with defining the Template Graph as the system's macrostate. It's sub-components (Template Nodes) represent the microstates. Let $G(N,A)$ be the macrostate (Template Graph) where N is the set of Template Nodes and A is the set of arcs connecting the nodes. Each node $j \in N$ has a probability of belonging to one of four possible energy states:

$$E_j^H = \text{High}$$

$$E_j^M = \text{Medium}$$

$$E_j^L = \text{Low}$$

$$E_j^0 = \text{Insignificant(Zero)}.$$

Given the discrete nature of the energy states, we must introduce thresholds to determine to which energy state each Template Node j belongs. Hence, we are introducing a parameter that may be set by the INFERD user as they see fit within their system. Let TH^H, TH^M, and TH^L denote the threshold values between the high, medium and low energies respectively. Note that $TH^L < TH^M < TH^H$, and $0 \leq TH^i \leq 1$ for all i in $\{L,M,H\}$. Let c_j denote the credibility of node j coming from the Feature Tree in L1 fusion described above. We can determine the credibility (or probability of a Template Node occurring), P_j, using equation (2.1.3.1.1):

$$P_j = \begin{cases} P(E_j^H) = \dfrac{e^{-\alpha^{-2}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}, & c_j \in [TH^H, 1] \\[2ex] P(E_j^M) = \dfrac{e^{-\alpha^{-1}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}, & c_j \in [TH^M, TH^H) \\[2ex] P(E_j^L) = \dfrac{e^{-\alpha^{-0}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}, & c_j \in [TH^L, TH^M) \\[2ex] P(E_j^0) = \dfrac{e^{-\alpha^{1}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}, & c_j \in [0, TH^L) \end{cases} . \tag{2.1.3.1.1}$$

These thresholds and the constant α ($\alpha \geq 1$), will allow for a parametric approach in determining the energy level of each Template Node j. A higher value of α results in more emphasis being placed on the higher energy states and vice versa. The assigning of value α can be attributed to many different reasons specific to each user; however we suggest that if the user believes their L0 sensors (information detection sensors) are highly reliable, they may opt for a higher α value. In contrast, if the user has less confidence in their sensors detecting incoming information, they may wish to use an α value closer to 1.

We define our partition function as $Z(T) = |N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}$ such that the sum of the probabilities over all energy states is equal to 1, hence meeting the basic axioms of probability:

$$\sum_{s\in S} P(E_s) = \sum_{j=1}^{|N|}(P(E_j^H) + P(E_j^M) + P(E_j^L) + P(E_j^0)) = |N|(P(E_j^H) + P(E_j^M) + P(E_j^L) + P(E_j^0))$$
$$= |N|\left(\frac{e^{-\alpha^{-2}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}} + \frac{e^{-\alpha^{-1}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}} + \frac{e^{-\alpha^{0}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}} + \frac{e^{-\alpha^{1}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right) = 1.$$

Now that we have the probabilities of each Template Node in $G(N,A)$ being in a certain energy state, we use them to obtain an overall credibility factor (CF) for the entire Template Graph. The probability of the state of a node j that has other nodes directed to it (N_j) will be affected by its neighboring nodes as long as the last occurrence of the two events depicting the node (at times r_j and r_h) are within a desirable time frame as set by the user, denoted t_{jh}. It is necessary to define a new set of probabilities for each Template Node that not only take into account its own state probability, but also the current states of the Template Nodes directed to it. Equation (2.1.3.1.2) defines Q_j as these desired probabilities:

$$Q_j = \lambda_j^0 P_j + \sum_{h\in N_j, |r_j - r_h| \le t_{jh}} \lambda_j^h Q_h \qquad \forall \quad j \in N. \tag{2.1.3.1.2}$$

$G(N,A)$ cannot contain any directed cycles. In particular, there will always be at least one sequence for obtaining the revised probabilities, such that no Q_j that depends on another is calculated without the proper adjustment.

Now that we have obtained probability values for each node of the Template Graph considering the topology of $G(N,A)$, we can introduce the overall Credibility Factor (CF) as seen in equation (2.1.3.1.3):

$$CF = \frac{\sum_{j=1}^{|N|} Q_j}{\left(\dfrac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)}. \tag{2.1.3.1.3}$$

The denominator of equation (2.1.3.1.3) simply normalizes the overall Credibility Factor so that when the probabilities of all of the Template Nodes are equal and in the high-energy state, then:

$$CF = \frac{\sum_{j=1}^{|N|} Q_j}{\left(\dfrac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)} = \frac{\sum_{j=1}^{|N|} \dfrac{e^{-\alpha^{-2}}}{|N|\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}}{\left(\dfrac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)} = \frac{\left(\dfrac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)}{\left(\dfrac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)} = 1.$$

The parameters λ are the constraints used to obtain a weighted sum of the state probabilities, such that:

$$\lambda_j^0 + \sum_{h\in N_j} \lambda_j^h = 1, \qquad \forall \quad j \in N, \quad \text{with}$$
$$\lambda_j^0 \ge 0, \qquad \lambda_j^h \ge 0 \qquad \forall \quad h \in N_j, \quad \text{and} \quad \forall \quad j \in N.$$

These λ values represent the importance of connecting nodes as desired by the user. For instance, in some application domains, it can be such that the influence of the Feature Trees with respect to Template Nodes be weighted heavily, and the correlation influence of connected nodes be only marginally considered. In this case, λ_j^0 can be set close to 1, and the λ_j^h values closer to 0. Note that each individual λ_j^h value does not necessarily have to be equal; one can place different weights on each node directed at the node in question.

It is important to note that in order for a consistent calculation of the Q_j values to be possible,

Let us continue our Airport Security example referring back to Figure 2.1.3.2 showing our Template Graph. Recall $c = [0.443, 0.55, 1, 0.01, 0.4, 0.1, 0.15]$. We will set our parameter $\alpha = 2$ and our set of thresholds as $TH^i = [0.25, 0.5, 0.75]$ for low, medium and high respectively. Then using equation (2.1.3.1.1), we can see the probability of being in high, medium, low, and insignificant energy states are $[0.059, 0.046, 0.028, 0.010]$ respectively. Now we can find the Q_j values given (2.1.3.1.2) and the parameters as follows:

$$\lambda_j^0 = \begin{cases} 1, & \text{if} \quad N_j = \emptyset \\ 0.5, & \text{o.w.} \end{cases}$$

$$\lambda_j^h = \frac{0.5}{|N_j|} \qquad \forall \quad h \in \{N_j : N_j \ne \emptyset\}, \quad \text{and} \quad \forall \quad j \in N.$$

We can see from the values given by the c vector, nodes 4, 6, and 7 have insignificant energy levels; nodes 1 and 5 have low energy; node 2 has medium energy; and node 3 has a high energy level. Hence, we use the P_j values calculated above along with the P_j values for each node of the Template Graph with the appropriate weighting to determine the Q_j value for each node j in N. We get $Q = [0.028, 0.037, 0.059, 0.024, 0.044, 0.021, 0.01]$. Then we compute the credibility factor as:

$$CF = \frac{\sum_{j=1}^{|N|} Q_j}{\left(\frac{e^{-\alpha^{-2}}}{\sum_{i=0}^{3} e^{-\alpha^{-(i-1)}}}\right)}$$

$$= \frac{0.028 + 0.037 + 0.059 + 0.024 + 0.044 + 0.021 + 0.01}{0.412}$$

$$= \frac{0.223}{0.412} = .5413 = 54.13\%.$$

Hence, in our example, under the statistical mechanics with parameterization methodology, we obtain a Credibility Factor of 54.13%, suggesting that this given passenger is about 54% probable to be a danger to others in the airport or on the aircraft. It is left to the user's discretion as to what is a large enough credibility in their given system in order to react accordingly.

2.1.3.2. The Entropy Approach for Credibility Factor

To liberate our system from the parametric and rule-based deficiencies listed in Table 1.2.1, we have implemented a novel approach of using Entropy, or a measure of randomness, to calculate the credibility values for our Template Graphs. By determining the inherent level of randomness in a template, and relating it to the maximum amount of randomness possible, we can derive meaning about how likely (credible) that particular template graph is taking place.

The theory of statistical mechanics is governed primarily through the second law of thermodynamics, better known as entropy. Entropy was first used as a measure within the study of thermodynamics, but has since been shown to be valuable in many other areas including psychodynamics, thermoeconomics and information theory. Information theory is useful in many disciplines but is most basically defined as a means to measure the amount of data that can be stored in a communication type medium. Claude Shannon, in 1948, composed a famous work [31] wherein he began to understand the transmission of information through a noisy channel. His fundamental results include the "source coding theorem" which states that the average number of bits of information required to represent the result of an uncertain event is given by entropy. Shannon's "noisy channel coding theorem" suggests that *reliable* communication is possible over noisy channels provided that the rate of communication is below a certain threshold. INFERD is a fusion system where a large amount of information is input, some of which is valuable and some of which is not. This non-valuable information can be considered noise in Information Theoretic terms. Since entropy measures amounts of valuable information throughout a channeling system, it seems appropriate to use such a measure for Situational Assessment within INFERD.

The study of entropy has evolved greatly throughout the years. It has been shown that there are many types of entropy that can be used in many domains. Tsallis [36, 37] introduces a generalized entropy function based on a parameter q.

$$H_q = k\frac{1 - \sum_{i=1}^{W} p_i^q}{q-1}, \qquad \left(\sum_{i=1}^{W} p_i = 1,\ k > 0\right).$$

The question arises as to what the value of q should be in any given domain. In [37], Tsallis discusses three optimization methods that can be used to find the optimal q. In [36], he discusses how in many optimization algorithms and information theory domains, $q \to 1$. He later suggests that while considering a Gaussian distribution, $q \to 1$ thus is the case for many natural phenomenon. Hence, we use the above Tsallis General Entropy Function with $q \to 1$. This gives us Shannon's Entropy Function as seen below.

Claude Shannon studied the discovery of statistical knowledge about a source by use of proper encoding of the information and defined entropy in cooperation with Boltzmann's famous H-Theorem as shown in equation (2.1.3.2.1), where H is entropy, p_i is the probability of being in state i and K is a constant (Boltzmann's constant in thermodynamics) [31].

$$H = -K\sum_{i=1}^{n} p_i \log_2 p_i. \qquad (2.1.3.2.1)$$

Shannon's application of entropy to information theory allows one to find the total amount of randomness embedded in a state-system process. In doing so, there must be an existing alphabet with known probabilities of symbols. Consider the example where we have an alphabet consisting of four symbols with the following probabilities (1/2 1/4 1/8 1/8). Using equation (2.1.3.2.1) we get $H = (1/2)\log_2 2 + (1/4)\log_2 4 + (1/8)\log_2 8 + (1/8)\log_2 8 = 1.75$ bits/symbol. Next consider the case where the probabilities of each symbol are at equality (1/4 1/4 1/4 1/4). Using equation (2.1.3.2.1) we get $H = 2.0$ bits/symbol. Next consider the case where we have the following probabilities (0 0 0 1). Equation (2.1.3.2.1) gives $H = 0.0$ bits/symbol. Note that as the probabilities of each symbol move to equality, the entropy moves to a maximum. This corresponds intuitively with the idea of randomness in a system—as each symbol in the alphabet becomes equally likely to occur, the symbols in the words constructed from that alphabet become less predictable. Also note that as the number of symbols in the alphabet increases, so does the randomness. This follows intuitively as well—if there are more symbols to choose from, predictability becomes more difficult.

Our system does not use alphabets and alphanumeric symbols as discussed in Shannon's paper, but our application in INFERD is in line with the requirements of entropy as defined by Shannon. Here, the "system" is our Template Graph, and the "symbols" are the Template Nodes within the Template Graph. We measure the entropy of the Template Graph by growing or shrinking the Total State Space defined below according to the credibilities of the Template Nodes and keeping the probabilities of each state within each sub-space at equality. By altering the size of the Total State Space to determine entropy as opposed to altering probabilities of states within the space, we develop a monotonically increasing H function with respect to the credibilities of the Template Nodes (c_j).

Before we describe the entropy method for calculating the Credibility Factor we must take into account the topology of the Template Graph. This is a similar procedure to the "Q-function" used in the Statistical Mechanics Methodology in Section 2.1.3.1. In fact, the only parameters used in the Entropy Approach are the same λ values as defined in the previous section. We will use the following equation to determine the new c_j* (the c_j values that take the directions of the Template Graph edges to nodes into account) values to be used in the entropy calculation.

$$c_j* = \lambda_j^0 c_j + \sum_{h\in N_j, |r_j - r_h| \le t_{jh}} \lambda_j^h c_h * \qquad \forall \quad j \in N$$

$$\lambda_j^0 + \sum_{h\in N_j} \lambda_j^h = 1, \qquad \forall \quad j \in N, \quad \text{with}$$

$$\lambda_j^0 \ge 0, \qquad \lambda_j^h \ge 0 \qquad \forall \quad h \in N_j, \quad \text{and} \quad \forall \quad j \in N.$$

We have a Template Graph $G(N,A)$ with a node set N and an arc set A, where the jth node has a normalized credibility factor value of c_j. We seek a normalized scalar aggregation function that approaches zero when all node credibilities tend to zero and approaches unity when all node credibilities tend to unity, and does not require us to take account of the arc set A (which would require an extensive parameterization of the aggregation function.)

Since the only data we intend to use in the aggregation function are the normalized credibility factors c_j, which can be interpreted as individual probabilities of their corresponding Template Nodes being "true," we are motivated to consider the Shannon entropy function as a convenient starting point for building our aggregation function. Shannon entropy is very simple to calculate under the assumption that a system has K equiprobable states, and is given by $H = \log K$ in this case. (The base of the logarithm is immaterial, as changing bases only induces a constant factor multiplying H.)

Thus we consider a system having equiprobable states, where the overall number of states is a decreasing function of the variable $x = \sum_{j=1}^{|N|} c_j$, i.e., the more certain we are of the truth of our composite set of Template Nodes (such that $x \to |N|$), the lower the number of states and hence the smaller the value of H; conversely, as the truth probabilities approach zero ($x \to 0$), the larger the number of states and the larger the value of H.

A simple function for the number of states K that satisfies these properties is

$$K = |N| - \sum_{j=1}^{|N|} c_j + 1.$$

In the two extreme cases, we have

$$H_{\min} = H(x = |N|) = \log K|_{c_j \equiv 1 \ \forall_j} = \log(1) = 0$$

$$H_{max} = H(x = 0) = \log K|_{c_j \equiv 0 \ \forall_j} = \log(|N| + 1).$$

For all values $0 < x < |N|$, we have $\log(|N| + 1) > H(x) > 0$.

Our desired credibility factor $CF(x)$ for the Template Graph should range monotonically between zero and unity as x ranges between its maximum and minimum values, respectively. The simplest function satisfying these properties is similar to the work presented by Pierce and John [28]:

$$CF(x) = \frac{H_{\max} - H(x)}{H_{\max} - H_{\min}} = \frac{\log(|N| + 1) - \log\left(|N| - \sum_{j=1}^{|N|} c_j + 1\right)}{\log(|N| + 1)}.$$

Now let us continue our ongoing example and consider the Template Graph with the same values given before: $c = [0.443, 0.55, 1, 0.01, 0.4, 0.1, 0.15]$. Here we illustrate the entropy approach via example. First, we must account for which nodes are pointed at which (the topology of the Template Graph). We will define our parameter λ just as is done in the prior example in Statistical Mechanics. Using the same routine, we obtain $c_j* = [0.433, 0.492, 1, 0.288, 0.7, 0.285, 0.15]$. We can now use the above equation to find the credibility factor (CF).

$$CF(x) = \frac{H_{\max} - H(x)}{H_{\max} - H_{\min}} = \frac{\log(|N| + 1) - \log\left(|N| - \sum_{j=1}^{|N|} c_j + 1\right)}{\log(|N| + 1)}$$
$$= \frac{\log(8) - \log(7 - (0.433 + 0.492 + 1 + 0.288 + 0.7 + 0.285 + 0.15) + 1)}{\log(8)} = 0.261.$$

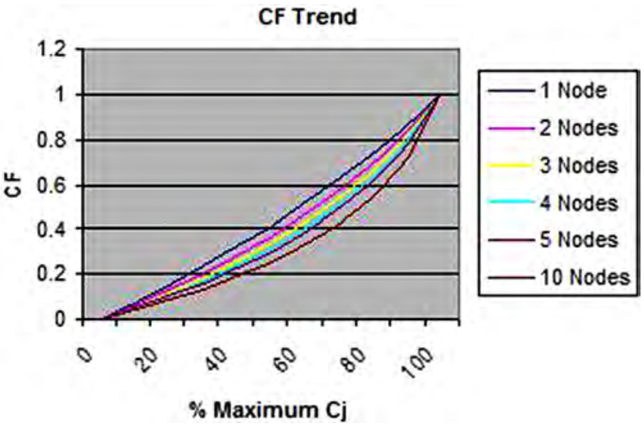


Fig. 2.1.3.2.2. *CF* Trends as node count increases.

Hence we say there is a 26.1% chance that this particular passenger is a danger to those around him.

As mentioned before, CF is the credibility value of the Template Graph and represents the likelihood that the given scenario is taking place. This value is used to rank the Template Graphs and is a simple indicator to the analyst helping them in their decision process of which situations to look into further.

A question can be raised to why a ten node Template Graph is not ranked as credible as a 1 node Template Graph when the sum of the credibilities of the nodes contained within them are at the same percentage level with respect to the maximum $\sum_{j\in N} c_j*$ (refer to Figure 2.1.3.2.2). Recall that as this value increases and decreases we determine the entropy for the graph by decreasing and increasing the size of Ω, respectively. Each state in this state space represents a piece of knowledge defining the scenario that has not been detected in the stream. Templates Graphs with more nodes have more of these states when at the same $\sum_{j\in N} c_j*$ level, which makes intuitive sense because we must detect many more occurrences in the system to be of definite certainty that it has taken place.

2.1.3.3. Other Credibility Factors

The above stated credibility factor calculations determine the reliability of information. However, we believe that although this is a very valuable measure, it is not all inclusive in terms of aiding the system user to make a complete decision. There should be more measures allowing a user to be more well informed of the current situation.

We provide two examples of possible measures that can be defined and embedded into a future version of INFERD. Let's consider the cyber domain as an example. One helpful measure could be "Depth." Cyber attacks are usually accomplished in a progression toward an end goal. This progression is understood, and thus a measure can be defined in order to determine how far into an attack a hacker may be at a certain time. This helps explain the current situation (L2 fusion) as well as understand possible immediate ramifications of a continued attack (L3 fusion). Another helpful measure could be "Breadth" of an attack. Breadth would help understand the entire scope of an attack, thus providing the user with information regarding how many possibilities an attacker would have in the near future.

These credibility measures among others can be very helpful to a user in terms of both situational awareness and impact assessment and will be further explored in future versions of INFERD.

3. CONCLUSIONS AND FURTHER WORK

In this paper we have described our INFERD system in a general sense as it can be applied to various domains depending on the needs of the consumer. We offer a system that is flexible in that a user can adjust the functions used at L1 and L2 fusion as well as input their own scenarios as Template Graphs in order to meet their needs. We describe the JDL definitions used for information fusion and show how INFERD incorporates those steps into its analysis of the system at hand. We offer two opposing viewpoints at the second level of fusion (L2) along with the advantages and disadvantages of each. The Entropy approach we discuss is a new and improved approach with direct ties into information theory as pioneered by Shannon [31].

To initially test INFERD and its fusion capabilities related to the cyber warfare domain, AFRL tasked Skaion Corporation with the job of generating a number of synthetic cyber attack traffic data sets labeled "Blind Tests." These data sets are actual packet and IDS alert information generated from attacks that were run on a virtual computer network with common data set components such as noise injected in. In this first test, INFERD was able to handle 86.4 million alerts over a 24 hour period. These data processing rates are highly above even large computer networks allowing us to claim real-time performance. Future tests will be performed against ground truth information to assess the "quality" of the generated hypotheses and the sensitivity of the generated hypotheses as a function of the parameters of the algorithm.

Advancements in the fusion process itself have been considered and proposed as research for a future phase of the project. Being able to determine credibilities in a given system is just a first step in the process of being able to successfully use that information to perform a desired task with ones system. In the future, there will be work done to make INFERD a self-acting, as well as a self-learning, system. An upcoming stage of our research will be to determine how to make INFERD a self-acting engine for various applications, hence creating a self-governing machine.

Currently, the user of the system must enter the Template Graphs to be analyzed by INFERD. For many application domains, it may be necessary to generate thousands or tens of thousands of these templates in order to appropriately analyze the system. Hence, it would be highly useful to create some sort of automated Template Generation technique. The next stage of our research will be to find a method to generate desirable templates to be inserted into INFERD for analyzing.

ACKNOWLEDGMENTS

This work has been performed under joint funding from the Air Force Research Lab in Rome, NY (AFRL-Rome) and the Advanced Research & Development Activity (ARDA).

REFERENCES

[1] D. Ballard and L. Owsley
Artificial intelligence in the helicopter cockpit of the future.
In *IEEE/AIAA 10th Digital Avionics Systems Conference*, 1991, 125–130.

[2] D. Ballard and L. Rippy
A knowledge-based decision aid for enhanced situational awareness.
In *IEEE/AIAA 13th Digital Avionics Systems Conference*, 1994, 340–347.

[3] S. B. Banks and C. S. Lizza
Pilot's associate: A cooperative, knowledge-based system application.
IEEE Intelligent Systems and Their Applications, **6**, 3 (1991), 18–29.

[4] E. Bengoetxea, P. Larranaga, I. Bloch and A. Perchant
Estimation of distribution algorithms: A new evolutionary computation approach for graph matching problems.
In *EMMCVPR '01: Proceedings of the Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, London, UK, 2001, 454–468.

[5] E. Bengoetxea, P. Larranaga, I. Bloch, A. Perchant and C. Boeres
Inexact graph matching using learning and simulation of Bayesian networks.
In Proceedings of CaNew workshop, ECAI, 2000.

[6] C. Bishop
Neural Networks for Pattern Recognition.
New York: Oxford University Press, 1995.

[7] P. Bradley, U. Fayyad and O. Mangasarian
Mathematical programming for data mining: Formulations and challenges.
INFORMS Journal on Computing, 1999.

[8] A. L. Buczak and R. E. Uhrig
Hybrid fuzzy-genetic technique for multisensor fusion.
Information Sciences, **93**, 3–4 (1996), 265–281.

[9] L. Chin-Teng and C. S. G. Lee
Reinforcement structure/parameter learning for neural-network-based fuzzy logic control systems.
IEEE Transactions on Fuzzy Systems, **2**, 1 (1994), 46–63.

[10] D. Conte, P. Foggia, C. Sansone and M. Vento
Thirty years of graph matching in pattern recognition.
International Journal of Pattern Recognition & Artificial Intelligence, **18**, 3 (May 2004), 265–298.

[11] F. Cuppens and A. Miege
Alert correlation in a cooperative Intrusion Detection Framework, In *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, 2002.

[12] B. Das
Representing uncertainties using Bayesian networks.
(DSTO-TR0918), 1999.

[13] N. Denis and E. Jones
Spatio-temporal pattern detection using dynamic Bayesian networks.
In *42nd IEEE Conference on Decision and Control*, **5** (2003), 4533–4538.

[14] M. H. DeGroot
Optimal Statistical Decisions.
New York: McGraw-Hill, 1970.

[15] J. Friedman
On bias, variance, 0/1-loss, and the curse of dimensionality.
Data Mining and Knowledge Discovery, 1997.

[16] D. L. Hall and J. Llinas
An introduction to multisensor data fusion.
Proceedings of the IEEE, **85**, 1 (1997), 6.

[17] D. Hall and J. Llinas
Handbook of Multisensor Data Fusion.
CRC Press, 2001.

[18] D. Hao, J. Haifeng, H. Zhiyao and L. Haiqing
Data fusion algorithm based on fuzzy logic.par In *Fifth World Congress on Intelligent Control and Automation*, **4** (2004), 3101–3103.

[19] D. Heckerman
Bayesian Networks for Data Mining, Data Mining and Knowledge Discovery.
1997.

[20] N. K. Kasabov
Hybrid fuzzy connectionist rule-based systems and the role of fuzzy fules extraction.
In *International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*, **1** (1995), 49–56.

[21] S. Kirkpatric, C. D. Gelatt, and M. P. Vecchi
Optimization by simulated annealing.
Science, **220**, 4598 (May 13, 1983), new series, 671–680.

[22] A. Lockett
Graph-matching intrusion detection system (GMIDS) whitepaper.
21st Century Technologies, 2003.

[23] A. Mahajan, W. Kaihong and P. K. Ray
Multisensor integration and fusion model that uses a fuzzy inference system.
IEEE/ASME Transactions on Mechatronics, **6**, 2 (2001), 188–196.

[24] I. V. Maslov and I. Gertner
Multi-sensor fusion: An evolutionary algorithm approach.
Information Fusion, in press, corrected proof, 2005.

[25] S. Mathew, C. Shah and S. Upadhyaya
An alert fusion framework for situation awareness of coordinated multistage attacks.
Technical Report 2004-18, CSE Department, SUNY at Buffalo, 2004.

[26] S. Mathew, C. Shah, S. Upadhyaya, M. Sudit, V. Garach, A. Stotz and M. Holender
A thermodynamic approach to intrusion alert fusion.
(submitted to *IEEE*).

[27] N. Metropolis, A. Rosenbluth, R. Rosenbluth, A. Teller and E. Teller
Equation of state calculations by fast computing machines.
Journal of Chemical Physics, **21**, 6 (1953), 1087–1092.

[28] J. Pierce and R. John
An Introduction to Information Theory—Symbols, Signals, and Noise.
New York: Dover Publishers, 1980.

[29] C. Qu and Y. He
A method of threat assessment using multiple attribute decision making.
In *6th International Conference on Signal Processing*, vol. 2, 2002, 1091–1095.

[30] C. Shannon
A mathematical theory of communication.
The Bell System Technical Journal, **27** (1948), 379–423.

[31] J. Salerno, M. Hinman and D. Boulware
Building a framework for situation awareness.
In *Proceedings of the Seventh International Conference on Information Fusion*, FUSION 2004, **1** (2004), 219–226.

[32] J. L. Schlabach, C. C. Hayes and D. E. Goldberg
A genetic algorithm for generating and analyzing battlefield courses of action.
Evolutionary Computation, **7**, 1 (1998), 45–68.

[33] A. N. Steinberg, C. L. Bowman and F. E. White
Revisions to the JDL data fusion model.
In *Proceedings of the SPIE Sensor Fusion: Architectures, Algorithms, and Applications III*, Orlando, FL, **3719** (1999), 430–441.

[34] S. C. Stubberud and K. A. Kramer
Data association for multiple sensor types using fuzzy logic.
In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, **3** (2005), 2154–2159.

[35] M. Sudit, A. Stotz and M. Holender
Situational awareness of a coordinated cyber attack.
SPIE Defense & Security Symposium, Orlando, FL, Mar. 2005.

[36] C. Tsallis
Nonextensive statistics: Theoretical, experimental and computational evidences and connections.
Brazilian Journal of Physics, **29**, 1 (Mar. 1999).

[37] C. Tsallis
Entropic nonextensivity: A possible measure of complexity.
Chaos Solutions and Fractals, **13** (2002), 371–391.

[38] A. Valdes and K. Skinner
Probabilistic alert correlation.
In *Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection*, 2001.

[39] C. Wallace and J. Patrick
Coding Decision Trees.
Monash University, Melbourne, Australia, 1991.

[40] S. Wang and N. P. Archer
A neural network based fuzzy set model for organizational decision making.
IEEE Transactions on Systems, Man and Cybernetics, Part C, **28**, 2 (1998), 194–203.

[41] J. Wang and M. Bender
Connectionist decision support systems for multiple criteria decision making.
In *IEEE International Conference on Systems, Man and Cybernetics*, 1991, 1955–1960.

[42] C. Jeung Won, J. Woo and C. Dong Lae
Situation/threat assessment fusion system (staffs).
In *Proceedings of the Fifth International Conference on Information Fusion*, **2** (2002), 1374–1380.

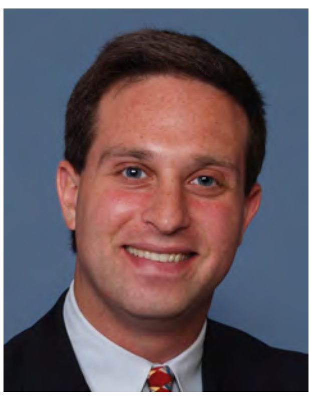
[43] R. R. Yager
On Ordered weighted averaging aggregation operators in multi-criteria decision making.
IEEE Transactions on Systems, Man and Cybernetics, **18** (1988), 183–190.

[44] R. R. Yager
Hierarchical aggregation functions generated from belief structures.
IEEE Transactions on Fuzzy Systems, **8**, 5 (Oct. 2000), 481–490.

[45] R. R. Yager
Generalized OWA aggregation operators.
Fuzzy Optimization and Decision Making, **2** (2004), 93–107.

Moises Sudit obtained his Bachelor of Science in industrial and systems engineering from Georgia Institute of Technology, his Master of Science in operations research from Stanford University and his Doctorate in operations research from Purdue University.

His primary research interests are in the theory and applications of discrete optimization. More specifically, he has been concerned in the design and analysis of methods to solve problems in the areas of integer programming and combinatorial optimization. One primary goal of this research has been the development of efficient exact and approximate (heuristic) procedures to solve large-scale engineering and management problems. As managing director of the Center for Multisource Information Fusion, Dr. Sudit has merged the interests of operations research with information fusion. He has an appointment as research professor in the School of Engineering and Applied Sciences at the University at Buffalo. Dr. Sudit is a NRC Fellow through the Information Directorate at the Air Force Research Laboratory and has received a number of scholarly and teaching awards. He has a number of publications in distinguished journals and has been the principal investigator in numerous research projects.



Michael N. Holender received his M.S. degree in industrial and systems engineering at the State University of New York at Buffalo (UB) with a focus in operations research in 2005. He is currently working towards his Ph.D. in the same field of study and will complete it in 2008. He received his B.S. in mathematics and statistics at Miami University, Oxford, OH, in 2002.

He began his career as an actuarial analyst for Univera Healthcare in Amherst, NY performing underwriting functions and supplying rates for employer groups. He streamlined many reporting and data gathering processes allowing future users to cut their time working on projects drastically while also improving accuracy. These interests lead him to graduate school whereupon he began his studies in optimization and process efficiency. He held both teaching and research assistantships throughout his career at UB. He has also taken on summer internships both with CUBRC (Calspan and University at Buffalo Research Center) and the United States Air Force Research Labs in Rome, NY. He currently works as a research assistant through CUBRC on various projects involving data fusion and conceptual spaces as they apply mathematical programming and optimization techniques. He tutors students of all ages in many different fields of mathematics and statistics. He also consults local businesses by creating easy-to-use software systems to increase process efficiency. Michael will continue his studies while acting as principal instructor of an introductory probability course for junior and senior undergraduate engineers at UB.

During his schooling, Michael has served at local president of Omega Rho. He is also a member of Pi Mu Epsilon Mathematics honorary fraternity. His research interests include data fusion, conceptual spaces, stochastic processes and business efficiency.

Adam D. (David) Stotz received a B.S. in computer engineering from the State University of New York at Buffalo in 2003 and is currently there working toward a Ph.D. in operations research with a primary scholarly research focus on distributed situation assessment.

He is currently employed at CUBRC, Inc. in Buffalo, NY where he has served as a research scientist for four years in the information fusion business line. He participates in the technical oversight of numerous projects within the CUBRC and UB comanaged Center for Multisource Information Fusion (CMIF) led by information fusion pioneering researchers Dr. Jim Llinas and Dr. Moises Sudit. His research activities have focused on high level (L2, L3, and L4) fusion in various application domains including cyber security, maritime domain awareness, and threat based routing among others. He has publications in various journals and has presented at numerous conferences on these topics. Adam was appointed as a National Research Council Fellow at AFRL-Rome Information Directorate in 2005 and received an Ethical Hacking certification from the EC-Council in 2006.

John T. (Terry) Rickard (S'67—M'75—SM'01) received the B.S. and M.S. degrees in electrical engineering from Florida Institute of Technology, Melbourne, Florida, in 1969 and 1971, respectively, and a Ph.D. degree in engineering physics from the University of California at San Diego, La Jolla, California, in 1975. He also received Series 7 and 63 General Securities Licenses in 1995 and a Series 24 General Securities Principal License in 1995.

He has 34 years of experience in technology and financial organizations, all of it in management and technology development positions. He began his career working in digital design and testing with Harris Corporation in 1969. He worked part-time as a graduate student for what was then the Naval Electronics Laboratory Center (now SPAWAR Systems Center) in San Diego, California from 1973 to 1975. In 1975, he cofounded ORINCON Corporation, a San Diego-based company specializing in the design and development of state-of-the-art data and information processing solutions for government and commercial customers. He ended his first career with ORINCON in 1994 as senior vice president and technical director. From 1994 to 2001, he served as President and later Chief Scientific Officer of OptiMark Technologies, Inc. He is a coinventor of the OptiMark transaction matching system and was instrumental in the company's development from a start-up enterprise to an operating entity. Rejoining ORINCON in 2001 as senior vice president, his focus has been on broadening the company's technology base, particularly in machine intelligence. When ORINCON was acquired by Lockheed Martin in 2003, he was appointed to the position of senior principal research scientist. In 2005, he was elected a Senior Fellow of Lockheed Martin, for whom he now works from his home in Larkspur, Colorado. His technical expertise includes signal processing, optimization, neural networks, fuzzy and expert systems, and graphical knowledge representation and inference for machine intelligence. His additional expertise includes financial engineering disciplines such as transaction systems, market structures, financial analytics, data mining, derivatives pricing, risk analysis, and trading strategies. His current research interests are in computational intelligence, conceptual spaces, information fusion, content based information retrieval, and nanotechnology.

Dr. Rickard has served on the boards of directors of three companies and has authored numerous technical publications that have appeared in refereed technical journals, books, and conference proceedings. In addition, he has authored several patents, and has several pending patent applications. He currently serves as the Vice Chairman of the IEEE Computational Intelligence Society, Denver Chapter, and is a board member of the non-profit Golden Triangle Research Institute and a Technical Advisory Board member of a nanotechnology hedge fund. In 2006, he received the Author of the Year Award from Lockheed Martin Integrated Systems and Solutions.

Ronald R. Yager received his undergraduate degree from the City College of New York and his Ph.D. from the Polytechnic University of New York.

He has worked in the area of fuzzy sets and related disciplines of computational intelligence for over twenty-five years. He has published over 500 papers and fifteen books. He is considered one the worlds leading experts in fuzzy sets technology. He was the recipient of the IEEE Computational Intelligence Society Pioneer award in Fuzzy Systems. Dr. Yager is a fellow of the IEEE, the New York Academy of Sciences and the Fuzzy Systems Association. He was given an award by the Polish Academy of Sciences for his contributions. He served at the National Science Foundation as program director in the Information Sciences program. He was a NASA/Stanford visiting fellow and a research associate at the University of California, Berkeley. He has been a lecturer at NATO Advanced Study Institutes. Currently, he is Director of the Machine Intelligence Institute and Professor of Information and Decision Technologies at Iona College. He is editor and chief of the International Journal of Intelligent Systems. He serves on the editorial board of a number of journals including the *IEEE Transactions on Fuzzy Systems, Neural Networks, Data Mining and Knowledge Discovery*, *IEEE Intelligent Systems, Fuzzy Sets and Systems*, the *Journal of Approximate Reasoning and the International Journal of General Systems*. In addition to his pioneering work in the area of fuzzy logic he has made fundamental contributions in decision making under uncertainty and the fusion of information.

Multistatic Sensor Placement: A Tracking Approach

O. ERDINC

P. WILLETT
University of Connecticut

S. CORALUPPI
NATO Undersea Research Centre

Active sonar tracking using measurements from multistatic sensors has shown promise: there are benefits in terms of robustness, complementarity (covariance-ellipse intersection) and of course simply due to the increased probability of detection that naturally accrues from a well-designed data fusion system. It is not always clear what the placement of the sources and receivers that gives the best fused measurement covariance for any target—or at least for any target that is of interest—might be. In this paper, we investigate the problem as one of global optimization, in which the objective is to maximize the information provided to the tracker.

We assume that the number of sensors is given, so that the optimization is done in a continuous space. The strong variability of target strength as a function of aspect is integral to the cost function we optimize. Doppler information is not discarded when constant frequency (Doppler-sensitive) waveforms are available. The optimal placements that result are consistent with our intuition, suggesting that our placement strategy may provide a useful tool in more complex scenarios where intuition is challenged.

Manuscript received July 3, 2007; revised November 11, 2007; released for publication December 4, 2007.

Refereeing of this contribution was handled by Robert Lynch.

This research was supported by the Office of Naval Research under contracts N0014-04-1-0133 and N0014-07-1-0055. This paper is a significantly expanded version of [7], itself a development on [6].

Authors' adresses: O. Erdinc and P. Willett, Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT, Email: {ozgur,willett}@engr.uconn.edu.; S. Coraluppi, Reconnaissance, Surveillance, and Networks Department, NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy, Email: coraluppi@nurc.nato.int.

1557-6418/07/$17.00

1. INTRODUCTION

A. Background

Multistatic sonar networks have the potential to improve anti-submarine warfare (ASW) detection and tracking performance against small, quiet targets in harsh reverberation-limited littoral operating areas. This improved performance comes from increased area coverage, expanded geometric diversity (greater coverage footprint), increased target hold, robustness to sensor loss and jamming, improved localization through cross-fixing (complementarity of uncertainty); and of course through simple gains in probability of detection via data fusion [5].

Moreover, multistatic systems are flexible. It is possible to use different waveforms at different sources, and the ping times can be chosen with greater freedom. In most scenarios, how to choose these parameters to exploit the capabilities of the multistatic sonar system is not immediately obvious—flexibility brings complexity. In practice it is common that parameters such as sensors' locations, waveforms and ping times are chosen heuristically and perhaps not optimally.

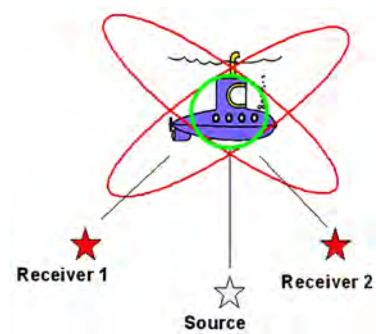


Fig. 1. Cartoon illustrating one of the benefits of using multistatic sonar: complementarity of the localization uncertainties.

In this paper, we investigate the advantages that an optimized sensor placement might offer, and we propose a methodology to determine the optimal placement strategy. Tracking in a complex and time-varying ocean environment is challenging. Multi-path effects, salinity/temperature gradients and geographical constraints may result in highly cluttered and/or low SNR sonar signals. Hence, finding the "best" placement strategy is going to be a considerable help to the tracker. In this work, it is assumed that quickly-deployable short range sensors are used and based on the predicted tracking performance, a sensor re-deployment scheme is proposed.

This study began with [6], which introduced many of the features from this paper (a similar criterion, aspect dependence, blanking zone). In [6] a certain intuitive regularity in optimized sensor layout was noted, and the incremental benefit of complementarity between sensors' perspectives (intersecting covariance ellipses) was seen to diminish after the second source/receiver pair. The results were considerably more stable in [7], presumably due to the use of the *minimax* optimization that we shall discuss shortly. The current paper is more comprehensive, and additionally considers the case of Doppler-sensitive waveforms. We find the work of Hernandez and Horridge [10], [11], who use the posterior Cramer-Rao lower bound (PCRLB), highly relevant. The PCRLB is dynamic and allows for both missed detections and false alarms by incorporating the information reduction factor (IRF) [12]. A question answered in [10] is: Given a present target being tracked and its associated uncertainty, where ought a new sensor be "dropped" to minimize the future uncertainty? In this paper we have a different concern: How should a field of sensors be configured to protect against an intelligent threat? The PCRLB is perhaps a better indication of tracking performance than the metric we shall introduce; but as noted in [10] it is more complex, conservative, and requires a description of the target dynamics and initial uncertainty.

In the next section, we explain our modeling assumptions. Then, we outline the proposed optimization technique, while Sections 4 to 7 report representative results. In the final section, we summarize our contribution.

B. Deployable Experimental Multistatic Sonar (DEMUS) System

We relate our analysis to DEMUS [14], an experimental system designed and used for investigation of the potential of multistatic sonar systems. The DEMUS system is composed of three deployable receiver arrays and one deployable acoustic source. Each system is battery powered, moored to the sea bottom, and communicates with the ship via radio and satellite links sited on a surface buoy. The receiver array records 64 channels of acoustic data (7 arms of 9 staves, plus one in the center). Each vertical stave sums the output of 3 hydrophones. The array's aperture may be scaled by extending the system's arms. The source array is made up of a vertically suspended set of 8 free flooded acoustic rings, capable of transmitting at high power.

Bistatic sonar can have many configurations, and the characteristics of DEMUS, our notional platform, include its relatively large beamwidth (approximately six degrees)—although with sufficient SNR the angular resolution can be made much better via interpolation—and the isotropic nature of its angular resolution. That is, although many sonar arrays have different performance depending on their orientation (e.g., broadside versus endfire in a linear array), DEMUS does not: this removes a parameter from our optimization process and allows us to concentrate on placement alone. In other words, the optimization is done concerning the locations of the sensors, their orientations do not matter.

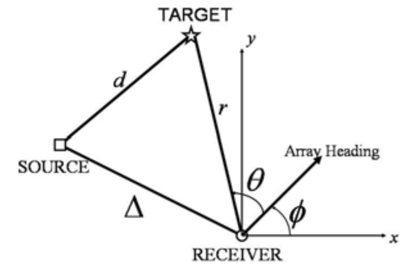


Fig. 2. Bistatic source/receiver/target geometry for a single source/receiver pair.

2. MODELING

A. Measurement Model and Localization Accuracies

In active sonar, the measurements are the return of the transmitted acoustic signal from the target of interest and the time of arrival. Hydrophones (receivers) determine the angle of arrival that gives the localization of the target together with the traveling time of the signal. Further, the transmitted frequencies could be translated from those of the received signal due to the relative motion of the target to the source and/or receiver. This Doppler shift provides information on the relative speed of the targets. In this paper we assume two different cases: first, with Doppler information not available, specifically where a wide-band linear frequency modulated (LFM) sonar signal is used; and second, a constant frequency sonar signal (CW) is used and Doppler information incorporated in the localization analysis.

The measurement model for a multistatic system, with stationary source and receiver, is

$$\begin{bmatrix} r \\ \theta \\ \dot{r} \end{bmatrix} = \begin{bmatrix} r_{tr} \\ \tan^{-1}\left(\dfrac{y-y_r}{x-x_r}\right) \\ \dfrac{\dot{x}(x-x_s)+\dot{y}(y-y_s)}{2r_{ts}} + \dfrac{\dot{x}(x-x_r)+\dot{y}(y-y_r)}{2r_{tr}} \end{bmatrix} + w, \qquad w \sim \mathcal{N}(0,\Sigma) \tag{1}$$

where the target state consists of its position and velocity, $X_t = [x \;\; y \;\; \dot{x} \;\; \dot{y}]'$ and r_{tr} is the range between the target and the receiver. For LFM signals, the range rate measurement $\dot{r}$ is insignificant and hence ignored.

The localization accuracy, i.e. the covariance matrix of the target state estimation after a single observation set $[r \;\; \theta \;\; \dot{r}]'$ is received, is a function of source, receiver, and target states and the selected sonar waveform,

$$R(X_s, X_r, X_t, \omega) = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & 0 & 0 \\ \sigma_{xy} & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x}}^2 & \sigma_{\dot{x}\dot{y}} \\ 0 & 0 & \sigma_{\dot{x}\dot{y}} & \sigma_{\dot{y}}^2 \end{bmatrix} \tag{2}$$

where

- X_s: source's state (location) in Cartesian plane, $s = 1, 2, \ldots, N_s$, N_s is the number of sources.

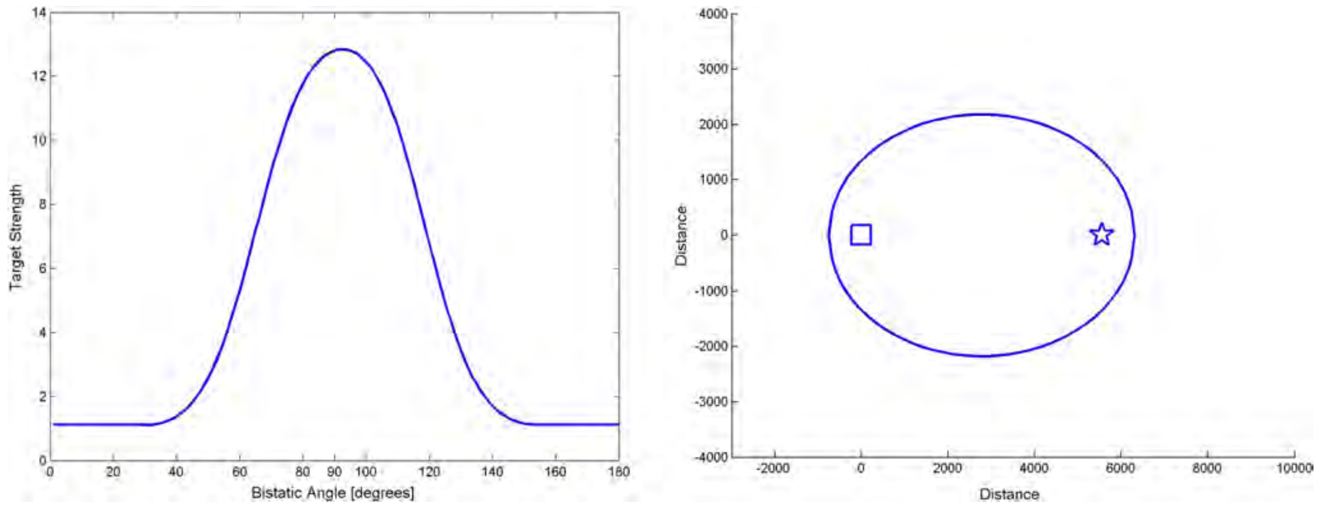


Fig. 3. Left: Nominal target strength [in dB] versus bistatic angle. Specular reflections from the broadside of the target are expected to be considerably stronger than those from oblique or endfire angles. Right: The blanking zone ellipse, with source indicated by the box and receiver by the star. Target detection is not possible within the blanking zone.

- X_r: receiver's state (location) in Cartesian plane, $r = 1,2,\ldots,N_r$, N_r is the number of receivers.
- X_t: target's state (location and velocity) in Cartesian plane.
- ω: selected waveform, $\omega \in \{\text{LFM, CW}\}$.

In recent work [3], [4], localization errors for bistatic and quasi-monostatic contact localization accuracy were derived as a function of the source-target-receiver geometry and assumed error statistics for source and receiver locations, sound speed, time, bearing, and array heading measurements. This study illustrated that the impact of measurement errors on localization accuracy depends highly on the source-target-receiver geometry. Due to space limitations we omit the lengthy equations showing the relations between the errors mentioned above and the components of covariance matrices, and refer the interested reader to the related publications [3] and [4]. Coraluppi has described the measurement errors and, more important, the measurement error covariances, as a function of the fundamental system errors in angle, observation time, array orientation, speed of sound and source receiver locations. Sensitivity analysis of these errors can be found in a consequent publication [9].

B. Target Detection Modeling

In addition to the localization analysis a second element that we require in the optimization metric to be discussed in Section 3-A is a model for target detection capability. Significant work exists on elaborate target strength and signal-to-noise (SNR) modeling; we have chosen to work with simple models that capture the key geometric dependencies relevant to CW and LFM transmitted waveforms.

1) *Aspect Dependence*: In many studies, targets are assumed to be a *point*: the sonar cross-section is independent of the angle of illumination; or, in the case of bistatic systems, independent of the relative angle between source and receiver. Many practical targets do not have this behavior at all. A "specular" return from a target whose broadside is normal to the bistatic angle is apt to be much larger than from a target that presents some other visage. Much is known about some targets. However, to keep our work generic, yet still to capture a flavor of the aspect-dependence that we seek, we have applied a simple target strength (TS) model to represent the aspect angle dependence of the signal return. If the target heading happens to be parallel to the line between the source and the receiver, the target strength is highest. This effect degrades as the angle varies away from the "best" angle; i.e., 90 degrees. Fig. 3 shows the target strength versus bistatic angle. If the expected target heading is known (for example, the surveillance volume is a narrow region that any target must traverse), the sensor placement ought to exploit this information.

2) *Direct Blast*: An important concern in bistatic systems is the direct blast: the signal that arrives at the receiver via the direct path. Propagation speed dictates that no target's return can arrive at the receiver prior to the direct blast; but, more important, since the direct blast is considerably louder than any target reflection, as a practical matter no reception is possible until the direct blast passes over. The direct blast can be very useful in calibration and registration, and consequently is perhaps a great strength of the multistatic architecture. However, there is an unavoidable "blanking zone" (the inside of an ellipse having source and receiver as foci and whose target-locus-receiver distance is the transmitted pulse length times the speed of propagation) as illustrated in Fig. 3, in which the system is blind.

3) *Signal-to-Noise Ratio (SNR) Modeling*: It is assumed that the multistatic system will be capable of transmitting and processing both LFM and CW. This capability is desired in a multistatic system since LFM and CW waveforms are "complementary." When the

target strength is maximum, target heading is parallel to the line between the source and the receiver, the target is in Doppler blind zone, i.e. range rate remains the same no matter what the target speed is. In this situation LFM waveform (or any Doppler-insensitive waveform) would be the right choice to use. On the contrary, when the target heading is perpendicular to bistatic orientation, then the target strength is minimum, whereas the range-rate is highest. Hence, a Doppler-sensitive waveform such as CW would provide high SNR.

In [8], SNR is calculated as a function of source/receiver/target locations and the selected waveform by a model that employs a simplified reverberation-limited sonar equation and a Q-function, which quantifies Doppler performance of sonar waveforms in rejecting reverberation. The model allows for both CW and LFM waveforms, and is sensitive to a number of waveform properties including center frequency, bandwidth, etc. In our work, we use this model and the reverberation-limited active sonar equation becomes (see [8] for details),

$$\mathrm{SNR} = \mathrm{TS} - \mathrm{BSS} - \mathrm{AREA} - Q(\Delta_f) \tag{3}$$

where TS is the target strength (as in Fig. 3-left), BSS is the bottom scattering strength, a parameter that depends on ocean seafloor type and composition (for our purposes, this is constant over the surveillance region), AREA is the area of the ensonified patch (i.e., resolution cell) that is a function of beam-width and the range from receiver to the patch, and the last term, $Q(\Delta_f)$, is the (negative-valued) Q-function, which reduces the amount of reverberation energy as a function of the target's Doppler shift. $Q(\Delta_f)$ is the term that quantifies Doppler-sensitive constant-frequency (CW) waveform's advantage over the FM waveform.

4) *Target Detection Probability and Detection-Localization Coupling*: The target detection probability for source i, receiver j, and waveform w, assuming Swerling I model (i.e., Rayleigh distributed target amplitude), is given by [13]

$$P_d^{\omega}(i,j) = e^{-DT/(1+\mathrm{SNR})} \tag{4}$$

where DT stands for the detection threshold. We set DT at 10 dB. The measurement error assumptions that drive the state estimation covariance (see equation 2) calculations include bearing, timing and frequency shift errors; these are related to the observed SNR as follows:

$$\sigma_\theta = \frac{\lambda}{\sqrt{\mathrm{SNR}}} \tag{5}$$

$$\sigma_\tau = \frac{\gamma}{\sqrt{\mathrm{SNR}}} \tag{6}$$

$$\sigma_f = \frac{\zeta}{\sqrt{\mathrm{SNR}}} \tag{7}$$

where λ, γ, ζ are some constants. This implies that amplitude-weighted interpolation between beams and between matched-filter bins is performed [2]. This coupling is used in our optimization work; that is, for each source-target-receiver geometry, we determine the measurement error standard deviations to be employed.

3. SENSOR PLACEMENT OPTIMIZATION

In this section, we describe the details of the proposed optimization algorithm. We propose an objective function that utilize the state estimation covariance matrix, R, and the probability of detection, P_d. The bearing, timing and the frequency errors that are used in the calculation of R, depend on the SNR value. Hence, R is a function of target orientation, relative Doppler, and source, receiver, target locations; so is P_d.

A. Objective Function

In finding an optimal sensor placement, our main objective is to improve target tracking performance. Hence, we use the "information" flow to the tracker as the basis of the optimization surface. The Fisher information matrix can be seen as a quantification of information in the measurement about the target's state. In Section 2-A, we have shown that the target localization uncertainty R can be derived as a function of source/receiver/target locations and the selected waveform (CW or LFM). The Fisher information matrix is defined as the inverse of the covariance of the estimate:

$$I(X_s, X_r, X_t, \omega) = R(X_s, X_r, X_t, \omega)^{-1}. \tag{8}$$

For optimization purposes, we need a scalar quantity for each source, receiver and target configuration for a given waveform ω (CW or LFM). We use the "information gain"

$$I_{\mathrm{fused}}(X_t) = \sum_{\forall\omega} \sum_{\forall(s,r)\in Y} P_d^{\omega}(s,r) I(X_s, X_r, X_t, \omega). \tag{9}$$

$I_{\mathrm{fused}}(X_t)$ is a function of target location given a particular geometry Y—the locations of sources and receivers, see (14). In other words, the second sum in the equation (9), implies that all source/receiver pairs' locations in the given geometry are considered. Note that equation (9) is based on the simplifying approximation that sensor measurement errors are uncorrelated from one contact to another, and indeed can be related to the PCRLB [10] for the case of a target without process noise and in the absence of false alarms. This is true for contact timing and bearing errors, but is not the case for source and receiver positioning errors, array heading errors, and speed of sound errors. Thus, the expression, while simple and useful for our purposes, has some degree of optimism: the true information gain is upper bounded by this expression.

Direct blast blanking means that for certain source-target-receiver geometries the detection probability that follows from our signal-excess modeling must be replaced by zero. Rather than doing so, and for numerical stability in the optimization process, we choose instead to discount the information gain with a barrier-type function. That is, as the target moves into the direct blast region, it is still detected but with a rapidly increasing localization uncertainty:

$$I(X_s, X_r, X_t, \omega) = e^{-\kappa d} \cdot I(X_s, X_r, X_t, \omega) \qquad (10)$$

where d is the shortest distance between the target and the border of the blanking zone ellipse.

We choose the determinant to be the scalar measure of the quality of information available to the tracker at each waypoint. Moreover, we consider a set of linear target trajectories T, each consisting of several waypoints, as illustrated in Fig. 4. The number of waypoints along each trajectory differs based on the speed of the target and the sampling interval; the latter is chosen so as to have several waypoints for the fastest-moving trajectories of interest. We use the (optimistic) simplifying approximation that information gained along a trajectory is the summation of the information across waypoints. Thus, as the scalar measure for each trajectory $T_i \in T$, we use the summation of determinants of the fused information matrix over all waypoints $w_{ij} \in T_i$:

$$M(T_i) = \sum_j \det(I_{\text{fused}}(w_{ij})). \qquad (11)$$

The objective function may be defined in either an *average* or *worst-case* sense. The former approach seems more applicable to problems where surveillance assets are covert, and is defined as:

$$J = \sum_i M(T_i) \qquad (12)$$

where i is the trajectory index. Alternatively, the objective function J can be defined as the *worst-case* (i.e. smallest) information gain achieved across all trajectories:

$$J = \min_i M(T_i). \qquad (13)$$

Maximization of the latter objective function is in fact the well-known *minimax* criterion: minimization of the maximum possible loss. In an overt network, a threat submarine would try to choose a path so that it would not be detected. Hence, operationally, the minimax criterion makes more sense since it makes sure that there are no "holes" in the surveillance region. We choose it as our objective in the optimization.

Note that this objective incorporates (and maximizes) both localization accuracy and the detection opportunities over the whole trajectory of the target. In other words, it aims to improve the tracking accuracy at all instances of target penetration. Hence, it can be seen that it relates to other operationally meaningful objectives, such as maintaining (not losing) a track, or increasing the target detectability.

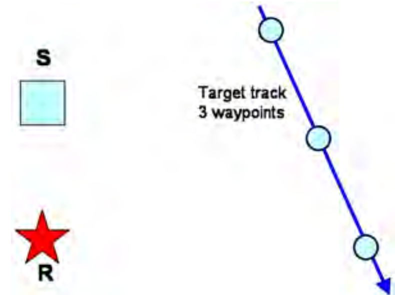


Fig. 4. In optimization, target trajectories are used instead of target grids. A target heading south-east is shown, with 3 waypoints along its trajectory.

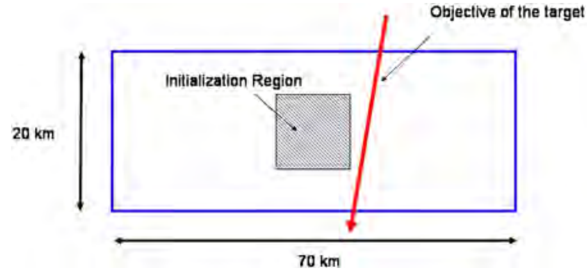


Fig. 5. Barrier Scenario: the best placement should prevent any target from passing this 20 km-by-70 km barrier without being tracked. The sensors are initially randomly placed inside the shaded region.

B. Our Scenario

Here we consider the barrier-scenario; the target submarine aims to pass a barrier 20 km long and 70 km wide (see Fig. 5). The multistatic sonar system has to be placed optimally so that no target can pass this region without being tracked. We consider some possible target trajectories, where the target heading and the speed differ (see Fig. 7; solid lines show 15 different trajectories). The objective is to maximize the provided information in the worst target trajectory. Operationally, the surveillance area should have no "holes." The tracker's performance is expected to meet some requirements even in the worst cases.

C. Likelihood Surfaces

With N available sensors, the parameter space for the optimization algorithm is $2N$, reflecting the need to locate each sensor in both latitude and longitude. We have no particular prescient knowledge on how this $2N$ dimensional surface would look, and hence it is not easy to decide on the most appropriate optimization algorithm. However, we can take 2-dimensional snapshots from this surface. One example is given in Fig. 6. The dark colored area around $(8000, -10000)$ is the best placement for the 3rd receiver at the moment that the snapshot is taken. In this figure, this snapshot reveals a smooth surface, although one that is not necessarily concave. We choose the steepest ascent algorithm, mainly

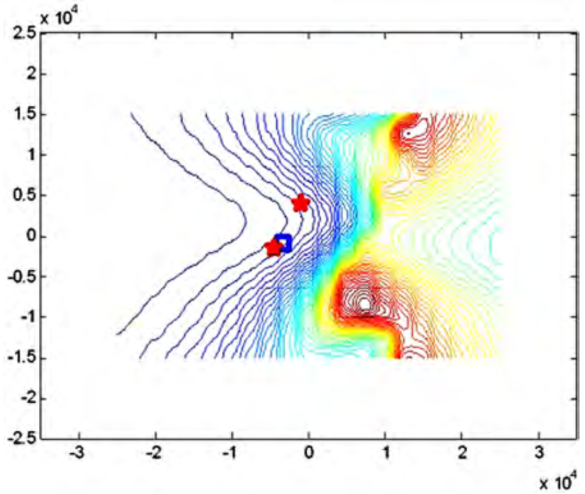


Fig. 6. 2-dimensional snapshot of 8-D surface: 1 source and 2 receiver positions are held, the 3rd receiver is free. The best positions are dark red areas at right.

since it is easy to implement, and it is intuitively easy to monitor its behavior.

D. Steepest Ascent Algorithm

The steepest ascent algorithm is a gradient-based unconstrained optimization technique. Y is a stacked vector of dimension $2N$,

$$Y = [X_{s_1}\ X_{s_2} \cdots X_{s_{N_s}}\ X_{r_1}\ X_{r_2} \cdots X_{r_{N_r}}]' \tag{14}$$

where total number of sensors is $N = N_s + N_r$. In each iteration k, Y^k moves in the direction of the gradient of the objective function, until convergence to a (local) maximum occurs. We have

$$Y^{k+1} = Y^k + \alpha^k \nabla f(Y^k) \tag{15}$$

where α_k is the step size used at iteration k. As previously described, we use objective function

$$f(Y^k) = J = \min_i \sum_j \det(I_{\text{fused}}(w_{ij})) \tag{16}$$

where $i = 1,2,\ldots,$ *number of waypoints*. Due to the complicated nature of the objective function, $f(Y^k)$, it is hard to obtain the gradient analytically. Hence a numerical gradient evaluation scheme is used. We approximate the gradient at the direction i by the central difference formula [1],

$$\frac{\partial f(Y^k)}{\partial Y^i} \approx \frac{1}{h}(f(Y^k + he_i) - f(Y^k - he_i)) \tag{17}$$

where h is fixed for each gradient direction and the e_i is the unit vector in the direction of Cartesian basis vector i. The value of h should be chosen as small as possible, otherwise, the coarsely-discretized objective function may result in erroneous gradient estimates. On the other hand, smaller h may cause numerical problems near the local maximum [1].

Step-size selection is a critical step for fast convergence.[1] The step-size needs to be large enough to reach the local minimum soon, and small enough to prevent oscillation (or large errors) when near the critical point. We apply a successive step-size reduction strategy, the so-called "Armijo rule" [1]. The Armijo rule picks its step-sizes to satisfy the inequality

$$f(Y^k) - f(Y^k + \beta^m s \nabla f(Y^k)) \geq -\sigma \beta^m s \|\nabla f(Y^k)\|^2 \tag{18}$$

where $0 < \beta < 1$ (chosen as $\beta = 0.7$), $0 < \sigma < 1$ (chosen as $\sigma = 0.1$), $s < 1$ and $m = 0,1,2$. The step-size is $\alpha_k = \beta^m s$. The Armijo rule first tests step-size s (i.e., $m = 0$) and then keeps increasing m until the inequality is satisfied. The parameter s (chosen as $s = 0.1$) and σ assure that there is a substantial increase in the objective function for the stepsize α_k. Convergence is declared when m reaches 20, this implying that the algorithm tests a point very close to the current one and there is still no improvement in terms of the objective. We choose s so that the first test point in the gradient direction would be 5 km away from the current location Y^k. Overall, the optimization algorithm works as follows:

1) Randomly initialize the sensors positions.[2]
2) Evaluate $f(Y^k)$ for current position vector Y^k.
3) Evaluate gradient by central difference formula ($2N * 2 = 4N$ $f(Y^k)$ evaluations.)
4) Test step sizes, α^k, according to Armijo rule. (At most m function evaluations.)
5) Update sensor positions using equation (15).
6) Go to step 2.

4. PLACEMENT STRATEGIES WITH LFM WAVEFORMS

We refer to a source-receiver pair as a detection node. In this section, we will consider 2-node and 3-node systems. Besides the main question of how to place these assets, we also aim to address which one of, for instance, the 2-node systems perform better? Is it better to deploy two sources with one receiver, or is the system with two-receivers and a source good enough? The barrier is the region (-35 km, -10 km) to (35 km, 10 km); there are 15 hypothetical target trajectories considered along this barrier. For instance, trajectory 1 represents a target with heading 200 degrees (from North) and 10 kts speed. Along this trajectory, there are 5 waypoints.

[1]We omit the discussions regarding the convergence rate of the steepest ascent algorithm. For a detailed analysis on the subject, see [1].

[2]Steepest ascent is not a stochastic method. Random initialization is hence important. It ensures that there is no bias in the convergence results.

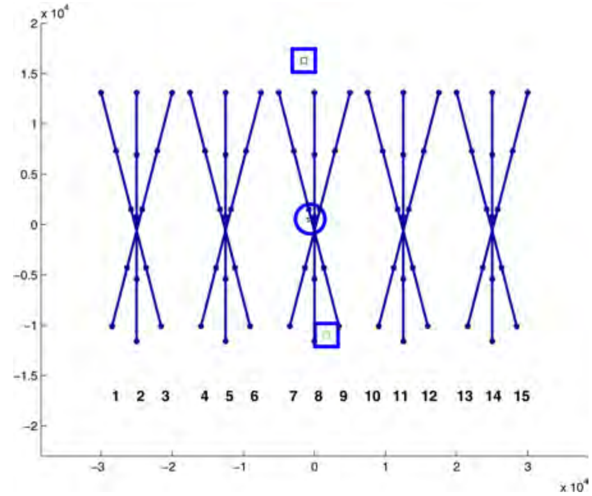


Fig. 7. 2 Sources (the squares), 1 Receiver (the circle) case: The lines are trajectories and the dots represent waypoints of each trajectory. Targets head south. The optimal placement forms a line in the North-South direction. See Table I for scores of trajectories.

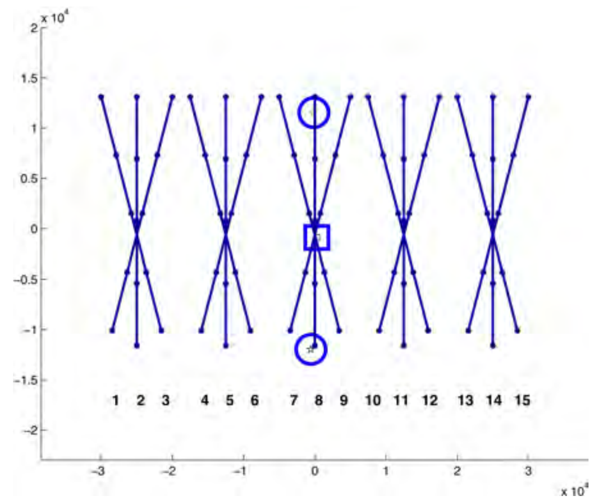


Fig. 8. 1 Source, 2 Receivers case: The optimal placement is very similar to the one in Fig. 7. See Table I for scores of trajectories.

A. Two-Node Cases

We consider two 2-node systems: one source and two receivers, and two sources and one receiver. The optimal placement turns out to be that the sources (blue squares in the Figs. 7 and 8) and receivers (circled star in the Figs. 7 and 8) form a line in the North-South direction. This is intuitive since it allows sensors to see the target from broadside. As explained in Section 2-B1, the target strength is at its maximum if the bistatic angle is close to 90 degrees (i.e., broadside), meaning that the SNR is high. Moreover, the receivers are located so that for any given target location, the orientation of the uncertainty ellipses becomes complementary (see Fig. 1). These lead both to high P_d and to good localization accuracy, hence resulting in good fused information. The objective of maximizing the worst-case information gain leads to a "balanced" deployment solution. During the optimization process, the worst trajectory jumps between the west-most group to the east-most group. Hence, the convergence geometry ends up being in the middle of the barrier.

TABLE I
Scores of Trajectories of Figs. 7 and 8
(worst cases are shown bold)

Trajectories	2S-1R	1S-2R
1	68.17	68.56
2	70.24	72.46
3	71.60	73.83
4	42.50	42.96
5	43.95	46.27
6	37.86	42.83
7	56.43	71.19
8	55.75	58.01
9	40.73	48.30
10	39.15	43.34
11	42.76	43.88
12	44.23	41.65
13	**71.62**	**74.42**
14	71.32	73.68
15	71.13	70.53

Placement scores[3] are obtained by using

$$\text{score} = \sqrt{\frac{\sqrt{1/J}}{\pi}}. \tag{19}$$

Note that if a trajectory consists of a single waypoint (target grid), and assuming the R is round, i.e., it is a circular uncertainty around the target position, the score has a physical meaning: it is the radius of the $1-\sigma$ covariance circle.

Scores corresponding to the trajectories shown in Figs. 7 and 8 are given in Table I. Balanced deployments are evident from the scores (compare the scores of 1, 2, 3 with 13, 14, 15). Another intuitive outcome is apparent from the scores of trajectories 7, 8 and 9: For the single source case, they are higher (worse). When the target penetrates into the blanking zone of the first source/receiver pair, the range between the target and the receiver of the second S/R pair is significantly higher than in the case of 2 sources and 1 receiver. A higher range results in higher localization error, mainly due to the bearing error. So for an LFM waveform, it is better to deploy two sources and a receiver, given the we have only three assets.

B. Multiple-Node Case

We now consider the case that there are 3 receivers with a single source. Fig. 9 shows the outcome of the optimization algorithm. The optimal placement suggests to use a regular geometry that one might find "heuristic." However this solution is not unique. The differ-

[3]Note that the scalar metric used in optimization is the determinant. Score is introduced for easy comparison.

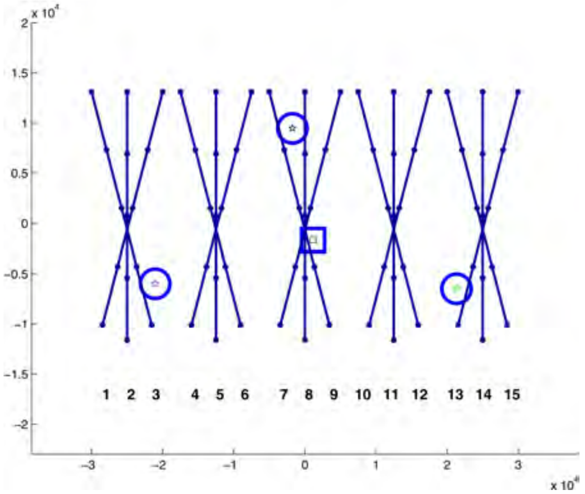


Fig. 9. 3 Receivers, one source case, the optimal placement: Two of the receivers (circles) are spread out to "monitor" the borders, and the source (square) is in the middle with the third receiver is in its north. They form a regular triangular geometry.

ent (random) initial placements result in different "optimal" placements, but they are equivalently good (see the scores from Table II).

It is intuitive that given a solution one can create its symmetric version and achieve the exact same scores in reversed order. Two such solutions are shown in Fig. 10. The initial placements were chosen randomly, but interestingly, the converged geometries are almost mirrored copies of each other. The scores indicate the reversed-ordering effect, and these solutions are as good as the regular-looking one from Fig. 9. Another nice feature is that the range of the scores is not wide, meaning that any path in the barrier would be monitored similarly. This is due to use of the minimax criterion.[4]

[4]Our observation is that another candidate criterion, maximizing the average measure, $\sum_i M(T_i)$, does not necessarily have this behavior. It can achieve an improvement in the average score by only improving some of the trajectories. Such a result would be operationally undesirable.

It is important to note that the steepest ascent algorithm is such that there is no guarantee that convergence is to the global optimum except when the objective function is concave; and from Fig. 6, we cannot assume a concave surface. Hence, our "optimal" solutions are not outcomes of each and every execution of the proposed algorithm. In fact, they are chosen so that they give the best score(s) among many runs of the optimization program. We follow this procedure:

- Execute the optimization process several times (on the order of 10).
- List all convergence scores.
- Determine distinct outcomes with scores close to the best overall score. (The optimal placements result in similar scores, and these scores are much lower than those in the rest of the list. Clustering is done by observation.)
- Ignore suboptimal placements.

It is important to note that all of the optimal placements yield "equivalent" solutions. In other words, they are very similar if one considers rotations and mirror reflections. The solutions given in Fig. 10 are an example for two such placements.

5. PLACEMENT STRATEGIES WITH CW WAVEFORMS

As opposed to an LFM waveform, a pulse at a constant frequency has coarser time resolution, and consequently, position-only estimation using only CW waveform yields a comparably larger uncertainty. On the other hand, a CW waveform provides Doppler information that is a function of the relative velocity of the target. Hence, while a fast target (e.g. 10 kts) yields few waypoints across the barrier, it is more likely to be detected when a CW waveform is used.

Since the information matrix has velocity uncertainty, the score loses its physical meaning when a CW

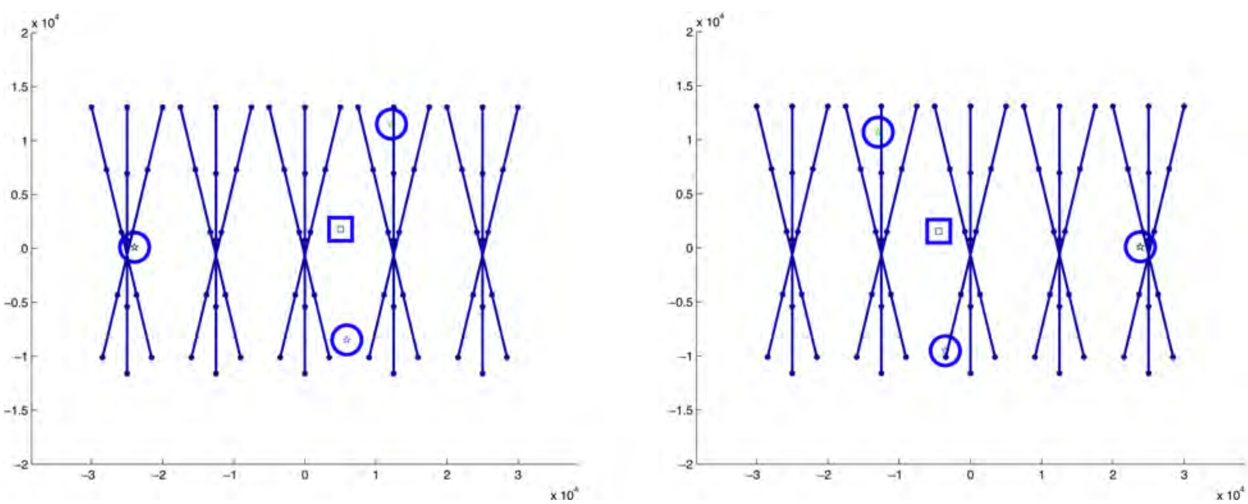


Fig. 10. Equivalent solutions: convergence results for two different runs. Interestingly, the geometries are "mirrors" of each other.

TABLE II
LFM Waveform is used
Scores of Trajectories for Figs. 9–10
(worst scores of each case are shown in bold)

Trajectory No.	1S-3R Fig. 9	1S-3R Fig. 10-left	1S-3R Fig. 10-right
1	52.14	48.51	47.90
2	**52.29**	41.66	52.18
3	47.15	**51.81**	51.08
4	38.91	42.41	30.10
5	36.93	41.49	33.29
6	35.97	43.20	33.82
7	46.56	29.69	34.65
8	51.09	31.91	30.73
9	47.04	29.47	35.12
10	35.60	32.07	42.70
11	35.63	30.51	40.14
12	41.67	29.59	40.53
13	45.51	49.65	**52.21**
14	50.49	50.43	42.29
15	51.86	45.64	47.81

waveform is available. Nonetheless, we use scores in our comparison tables, since they are easier to compare than the values of determinants of the information matrices.

A. Two-Node Cases

We look at the same two systems as before. The optimal placements are given in Figs. 11 and 12. This time the sensors are in the west-east orientation. This is again intuitive since the penetrating target provides high range rate (Doppler) measurement, and hence the information provided to tracker is higher. The complementarity of the waveforms is consistent with the complementarity of the optimal solutions. Another important observation is that the scores from the faster target are much better (lower) than the other one. This indicates that the Doppler information is so dominant that even though the slow target has many more waypoints and hence many more chances to be detected, it is harder to detect it if only a CW waveform is used.

B. Multiple Node Cases

Convergence geometry for the 1 source—3 receivers configuration is given in Fig. 13 in the fast-target case. It is similar to the regular triangle geometry. The eminent structure of the optimal geometry is to put two of the receivers close to both sides of the barrier, and the remaining receiver is placed in the center so that it forms a line with the source in north-south orientation. (The results are similar in the slow-target case.)

More important, this placement does not contradict the one from LFM case. It seems that both high Doppler detection and high target strength detection is possible if 3 source/receiver pairs are available.

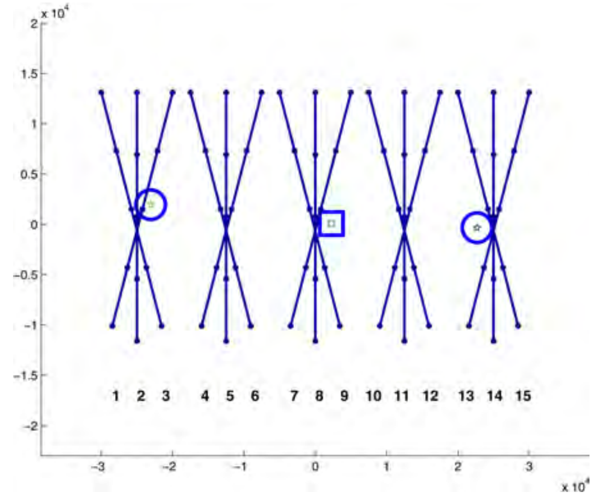


Fig. 11. CW waveform is used: Orientation of the sensors are complementary to the one of LFM waveform. See Table III for the scores. Target speed is high: 10 knots.

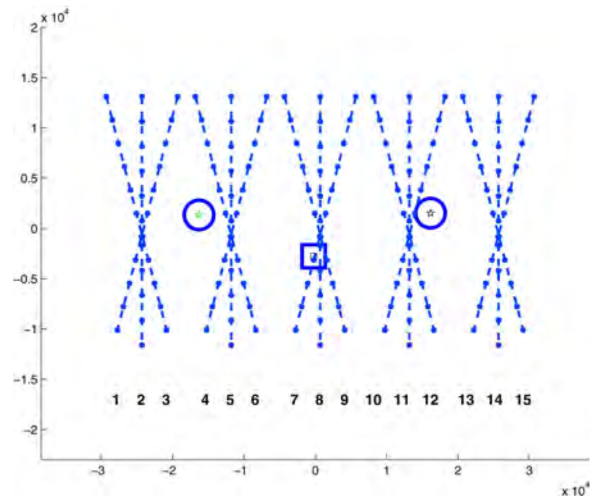


Fig. 12. CW waveform is used: Target speed is 4 knots.

Note that scores for slow target trajectories are worse than faster ones. Indeed, target SNR is considerably higher when target speed is high, resulting in much higher probability of detection, and also better localization. This effect is so dominant that even the fact that slow targets have twice as many waypoints is little help to the tracker.

6. PLACEMENT STRATEGIES WITH BOTH CW AND LFM WAVEFORMS

In the previous two sections, we have analyzed the proposed methodology and reported that the results are consistent with intuition. In this section, we assume the multistatic system is capable of using both waveforms and the target speed is unknown. Hence we consider two extreme cases to analyze the worst-case scenario: a target with 4 knots speed which is hard to detect with a CW waveform and a fast target moving with 10 knots trying to pass the barrier as quickly as possible. Each

TABLE III
Scores of Trajectories of Figures
(worst cases are shown bold)
Scores when CW Waveform is used for the Systems in the Figs. 11 and 12

Trajectory No.	Fast 10 kts	Slow 4 kts
1	4.73	91.63
2	11.72	95.17
3	9.23	84.09
4	7.12	33.39
5	6.27	44.36
6	7.62	23.98
7	11.09	96.03
8	10.19	84.77
9	**11.73**	**96.56**
10	8.07	23.57
11	6.18	44.98
12	8.51	33.77
13	8.18	85.76
14	10.97	96.52
15	8.62	93.18

Note: The slow target gives higher (worse) scores, unlike when an LFM waveform is used.

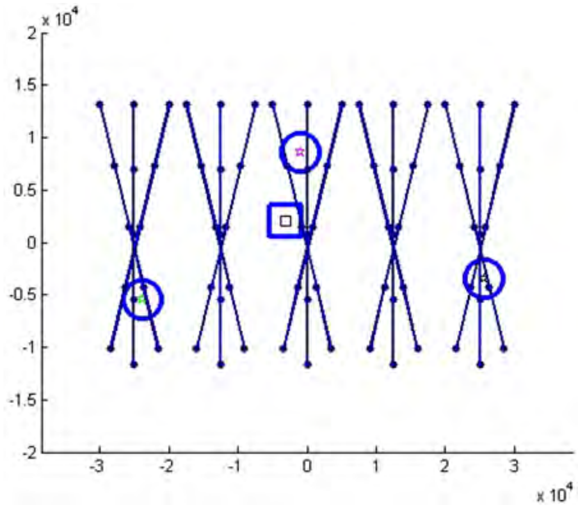


Fig. 13. CW waveform is used: Two of the receivers reach to the sides of the barrier and the source is in the middle with the third receiver.

of the 15 trajectories are duplicated so one trajectory corresponds to a slow target and the other corresponds to a fast one.

For the single source, three receivers case the optimal placement is shown in Fig. 14. The outcome is consistent with the earlier findings so that two of the receivers are placed far out and the third receiver stays in the middle of the barrier close to the source.

7. SENSITIVITY ANALYSIS OF SENSOR PLACEMENT RESULTS

In the previous sections, we have reported "optimal" placement strategies based on a sparsely sampled linear

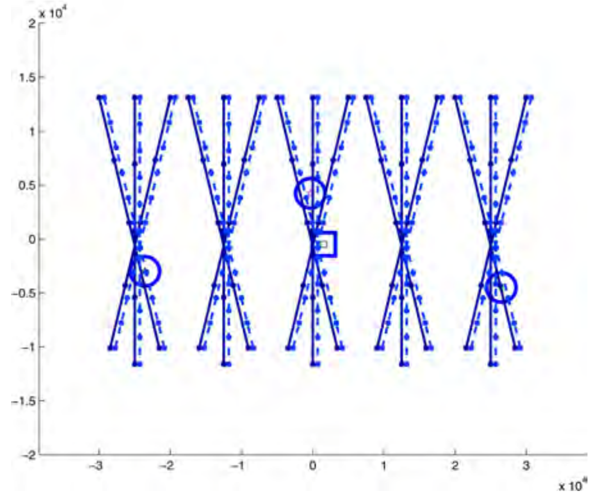


Fig. 14. Both CW and LFM waveforms are available. Trajectories are duplicated corresponding to different target speeds, 4 kts and 10 kts. The geometry is consistent with the earlier placements.

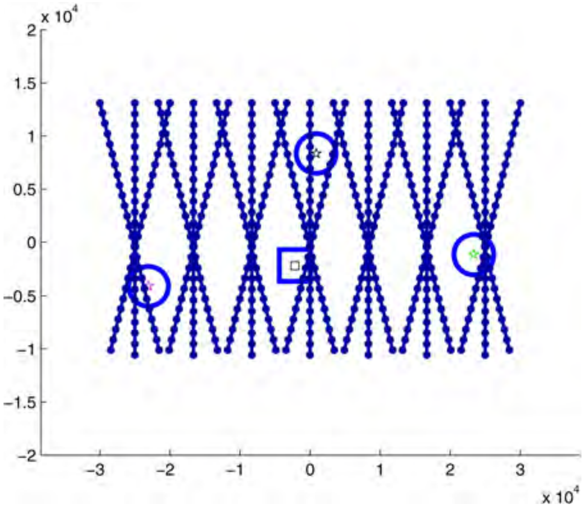


Fig. 15. The same initial placement as for Fig. 9, but with many more trajectories. The optimal placement is the same.

target trajectories. Here we investigate whether this choice has a dramatic impact on our results. The first analysis we consider is to use the result from Fig. 9 and run the optimization again with more trajectories, each of which has many more waypoints. The initial placement is the one in the optimization run resulting in Fig. 9. As seen in Fig. 15, the outcome is almost identical to the former result.

For the second analysis the trajectories are perturbed a random amount as seen in Fig. 16, and the optimization algorithms re-run. The initial placement for optimization is chosen as the optimal placement from Fig. 9. If this placement is optimal, it is expected that perturbation of the target trajectories would have little impact. The result in Fig. 16 confirms this.

In an overt network, there is no reason for a threat target to follow a straight line. Hence, in the last part of sensitivity analysis we consider piece-wise linear tra-

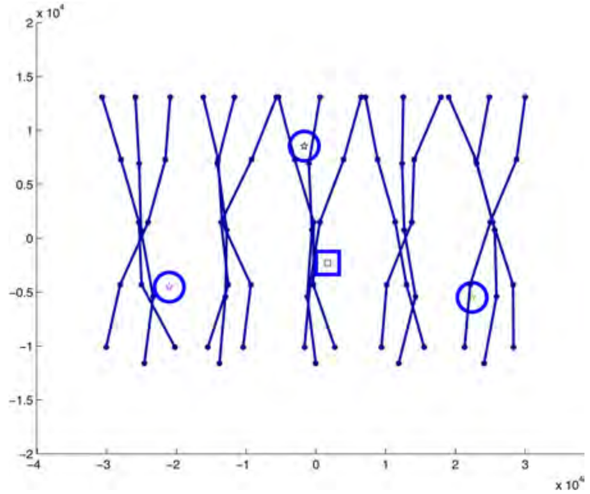


Fig. 16. The optimal placement with perturbed trajectories. The initial point is the optimal placement from Fig. 9.

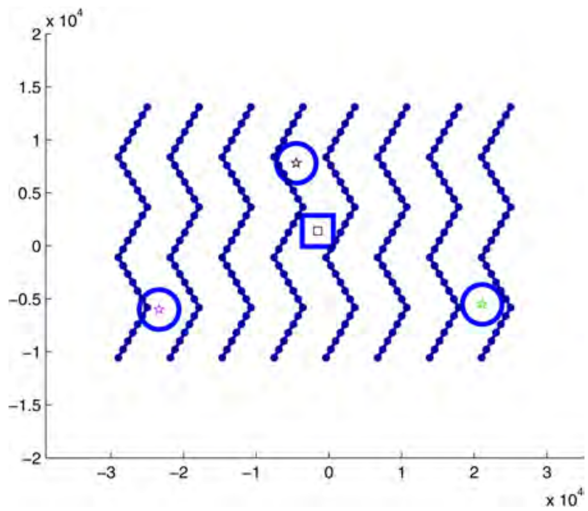


Fig. 17. Targets follow a zig-zag path. It appears that the optimal placement is robust.

jectories: such a trajectory is of interest since the target strength is changing along the trajectory. The result is given in Fig. 17. It appears that the optimal placement from Fig. 9 is robust for different configurations of target trajectories.

8. SUMMARY AND CONCLUSIONS

We propose an optimization technique for the optimal sensor placement for multistatic sonar systems. We study optimal placements in the LFM-only case, the CW-only case, and the combined LFM-CW case, and show that the optimal placements are consistent with our intuition, thus validating our placement methodology and its use as a placement aid in more complex scenarios where intuition is challenged.

An important aspect of the algorithm is that we employ a "minimax" criterion which results in a balanced surveillance performance. This makes sure that there is no path across the barrier for a target yet it remains "unseen."

Some aspects of modeling are important:

- Targets are not "point" targets: we employ an aspect angle dependent target strength model.
- Target Doppler is included in the localization analysis whenever CW waveforms are used.
- It is assumed that targets follow some realistic trajectories; Hence, availability of two complementary waveforms, CW and LFM, is incorporated in the metric.
- The modeling reflects the "Blanking Zone" due to direct blast signal reception.
- Signal Excess is calculated by a model where:
  —A simplified reverberation-limited sonar equation is used;
  —The Q-function is considered, which quantifies the Doppler performance of sonar waveforms in rejecting reverberation.

A scalar metric blends all of the above into a trajectory score, where "information gain" is computed at each waypoint of the trajectory. A steepest ascent algorithm is used for optimization, together with an intelligent step-size selection scheme (Armijo rule), and numerical gradient evaluation techniques.

It is desired to show that the "optimal" placements do, in fact, improve tracking performance. Thus, in future work, we plan to compare actual tracking performance based on optimal sensor placements with performance based on sub-optimal placements. This study would provide further validation that our information-based optimization objective captures the salient dataset characteristics that are required for high-quality tracker outputs and an effective surveillance system.

REFERENCES

[1] D. Bertsekas
Nonlinear Programming.
Athena Scientific, 1999.

[2] W. Burdic
Underwater Acoustic System Analysis.
Prentice Hall, 1984.

[3] S. Coraluppi
Multistatic sonar localization.
IEEE Journal on Oceanic Engineering, **31**, 4 (Oct. 2006).

[4] S. Coraluppi, C. Carthel, D. Hughes, A. Baldacci and M. Micheli
Multi-waveform active sonar tracking.
Proceedings of International Waveform Diversity and Design Conference, Pisa, Italy, June 2007.

[5] H. Cox
Fundamentals of Bistatic Active Sonar in Underwater Acoustic Data Processing.
Kluwer Academic Publishers, 1989.

[6] O. Erdinc, J. Areta, S. Coraluppi and P. Willett
Multistatic sonar sensor placement—A tracking perspective.
Proceedings of 2005 SPIE Conference on Signal and Data Processing of Small Targets, San Diego CA, Aug. 2005.

[7] O. Erdinc, P. Willett and S. Coraluppi
Multistatic sensor placement: A tracking approach.
Proceedings of International Conference on Information Fusion, Florence, Italy, July 2006.

[8] D. Grimmett
Multi-sensor placement to exploit complementary properties of diverse sonar waveforms.
Proceedings of International Conference on Information Fusion, Florence, Italy, July 2006.

[9] D. Grimmett and S. Coraluppi
Sensitivity analysis for multistatic LFAS localization accuracy.
NURC Technical Report SR-386, 2004.

[10] M. L. Hernandez and P. R. Horridge
Advances in the management of multi-sensor systmes with associated applications.
IEE, Target Tracking 2004: Algorithms and Allocations, May 23–24, 2004.

[11] P. R. Horridge and M. L. Hernandez
Multistatic radar resource management.
Proceedings of SPIE Signal and Data Processing of Small Targets, San Diego, CA, Aug. 2003.

[12] R. Niu, P. Willett and Y. Bar-Shalom
Matrix CRLB scaling due to measurements of uncertain origin.
IEEE Transactions on Signal Processing, **49**, 7 (2001), 1325–1335.

[13] C. Rago, P. Willett and Y. Bar-Shalom
Detection-tracking performance with combined waveforms.
IEEE Transactions on Aerospace and Electronic Systems, **34**, 2 (Apr. 1998), 612–624.

[14] M. van Velzen and L. Leutelt
Evaluation of radial receive array performance.
NURC Technical Report SR-400, Oct. 2004.

Ozgur Erdinc received the B.Sc. degree from Istanbul Technical University, Turkey, in 1997, and M.Sc. degree from University of Connecticut, Storrs, CT in 2004.

He is currently a research assistant/Ph.D. student in the Electrical and Computer Engineering department at the University of Connecticut. His primary research interests are in the areas of target tracking, signal processing, detection and estimation.

Peter Willett received his BASc (Engineering Science) from the University of Toronto in 1982, and his PhD degree from Princeton University in 1986. He has been a faculty member at the University of Connecticut ever since, and since 1998 has been a Professor. He was awarded IEEE Fellow status effective 2003. His primary areas of research have been statistical signal processing, detection, machine learning, data fusion and tracking. He also has interests in and has published in the areas of change/abnormality detection, optical pattern recognition, communications and industrial/security condition monitoring. He is editor-in-chief for IEEE Transactions on Aerospace and Electronic Systems, and until recently was associate editor for three active journals—IEEE Transactions on Aerospace and Electronic Systems (for Data Fusion and Target Tracking) and IEEE Transactions on Systems, Man, and Cybernetics, parts A and B. He is also associate editor for the IEEE AES Magazine, associate editor for ISIF's electronic Journal of Advances in Information Fusion, is a member of the editorial board of IEEE's Signal Processing Magazine and was first editor of the AES Magazine's periodic Tutorial issues. He has been a member of the IEEE AESS Board of Governors since 2003. He was General Co-Chair (with Stefano Coraluppi) for the 2006 ISIF/IEEE Fusion Conference in Florence, Italy and for the 2008 ISIF/IEEE Fusion Conference in Cologne, Germany, Program Co-Chair (with Eugene Santos) for the 2003 IEEE SMCC inWashington DC, and Program Co-Chair (with Pramod Varshney) for the 1999 Fusion Conference in Sunnyvale.

Stefano Coraluppi received the B.S. degree in Electrical Engineering and Mathematics from Carnegie Mellon University (Pittsburgh PA) in 1990, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland (College Park MD) in 1992 and 1997, specializing in automatic control. From 1997 to 2002 he was a Senior Research Engineer at ALPHATECH Inc. (Burlington MA), where he worked on multi-sensor data fusion and target tracking for ground surveillance. In 2002 he joined the NATO Undersea Research Centre (La Spezia, Italy) as a Senior Scientist, and works on multistatic active sonar fusion and tracking for maritime surveillance. He served as General Co-chair of the *9th International Conference on Information Fusion* in Florence, Italy, in July 2006. He is an Associate Editor for Target Tracking and Multisensor Systems for the *IEEE Transactions on Aerospace and Electronic Systems*, and is a member of the Board of Directors of the *International Society of Information Fusion*.

Game Theoretic Approach to Threat Prediction and Situation Awareness

GENSHE CHEN

DAN SHEN

CHIMAN KWAN
Intelligent Automation, Inc.

JOSE B. CRUZ, JR.
The Ohio State University

MARTIN KRUGER
Office of Naval Research

ERIK BLASCH
Air Force Research Laboratory

The strategy of data fusion has been applied in threat prediction and situation awareness. The terminology has been standardized by the Joint Directors of Laboratories (JDL) in the form of a so-called "JDL Data-Fusion Model." Higher levels of the model call for prediction of future development and awareness of the development of a situation. It is known that the Bayesian Network is an insightful approach to determine optimal strategies against an asymmetric adversarial opponent. However, it lacks the essential adversarial decision processes perspective. In this paper, a data-fusion approach for asymmetric-threat detection and prediction based on advanced knowledge infrastructure and stochastic (Markov) game theory is proposed. Asymmetric and adaptive threats are detected and grouped by intelligent agent and Hierarchical Entity Aggregation in level-two fusion and their intents are predicted by a decentralized Markov (stochastic) game model with deception in level-three fusion. We have evaluated the feasibility of the advanced data fusion algorithm and its effectiveness through extensive simulations.

Manuscript received August 14, 2006; revised May 3, 2007 and November 15, 2007; released for publication January 16, 2008.

Refereeing of this contribution was handled by Peter Willett.

Authors' adresses: Genshe Chen, Dan Shen, Chiman Kwan, Intelligent Automation, Inc., 15400 Calhoun Dr., Suite 400, Rockville, MD 20855, E-mail: {gchen, dshen, ckwan}@i-a-i.com; Jose B. Cruz, Jr., Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, E-mail: (cruz.22@osu.edu); Martin Kruger, Office of Naval Research, 875 North Randolph Street, Suite 1425, Arlington, VA 22203, E-mail: (Martin_Kruger@onr.navy.mil); Erik Blasch, Air Force Research Laboratory, 2241 Avionics Cir., WPAFB, OH 45433, E-mail: (erik.blasch@wpafb.af.mil).

1. INTRODUCTION

Data fusion has been largely applied to symmetric military warfare in which long-term strategic target development processes have developed the signatures or deductive model-based templates describing the component targets of the fielded adversary forces [14], [27]. Asymmetric adversaries, usually utilizing Camouflages, Concealment, and Deceptions (CC&D), and "unilateral destruction" are quite unpredictable in their behavior, tactics, weapons, and the choice of targets. Information and patterns of behavior that could provide advanced warning of hostile intent are often hidden in a vast background of harmless civilian activity. Automated processing techniques are needed to augment tactical intelligence-analysis capabilities by automatically identifying the militarily-relevant features of all available data of different modalities (e.g., signals intelligence, human intelligence, imagery intelligence, etc.) and recognizing patterns that are out of the ordinary [25] and/or indicate probable hostile intent [18].

As asymmetric warfare becomes more prevalent and introduces new security challenges, there is a critical need for strategies for providing actionable information to military decision makers so that the adversaries' most likely future courses of actions (COAs) can be predicted. By successfully assessing possible future threats from the adversaries, decision makers can make more effective targeting decisions, plan friendly COAs, mitigate the impact of unexpected adversary actions, and direct sensing systems to observe more efficiently adversary behaviors. Information fusion is an efficient method for providing this information by combining diverse data from multiple sources. Many studies have dealt with the information sources directly, which is the first level of fusion (object assessment) and some have aggregated information for level-two fusion—situation assessment (SA) [22]. Information fusion for threat and situation analysis is outlined in [13] with reference to utility value. Others have included SA from cyber-IF domains [20] with elements of SA ontology developments [16]. However, to combat the present and future asymmetric threats to national and international security, information fusion developments must progress beyond current level-one fusion paradigms.

In this research, we developed a data-fusion framework for asymmetric-threat detection and prediction in an urban-warfare setting based on advanced knowledge infrastructure and Markov (stochastic) game theory. It consists of four closely coupled activities: 1) Level-one fusion automates the processing and integration of information from disparate sources to produce an integrated object state. 2) Level-two fusion automates the estimation and groups the cooperative objects which perform common tasks. The main tasks of level-two fusion are estimation and prediction of relations among entities, to include force structure and cross force relations, communications and perceptual influences, phys-

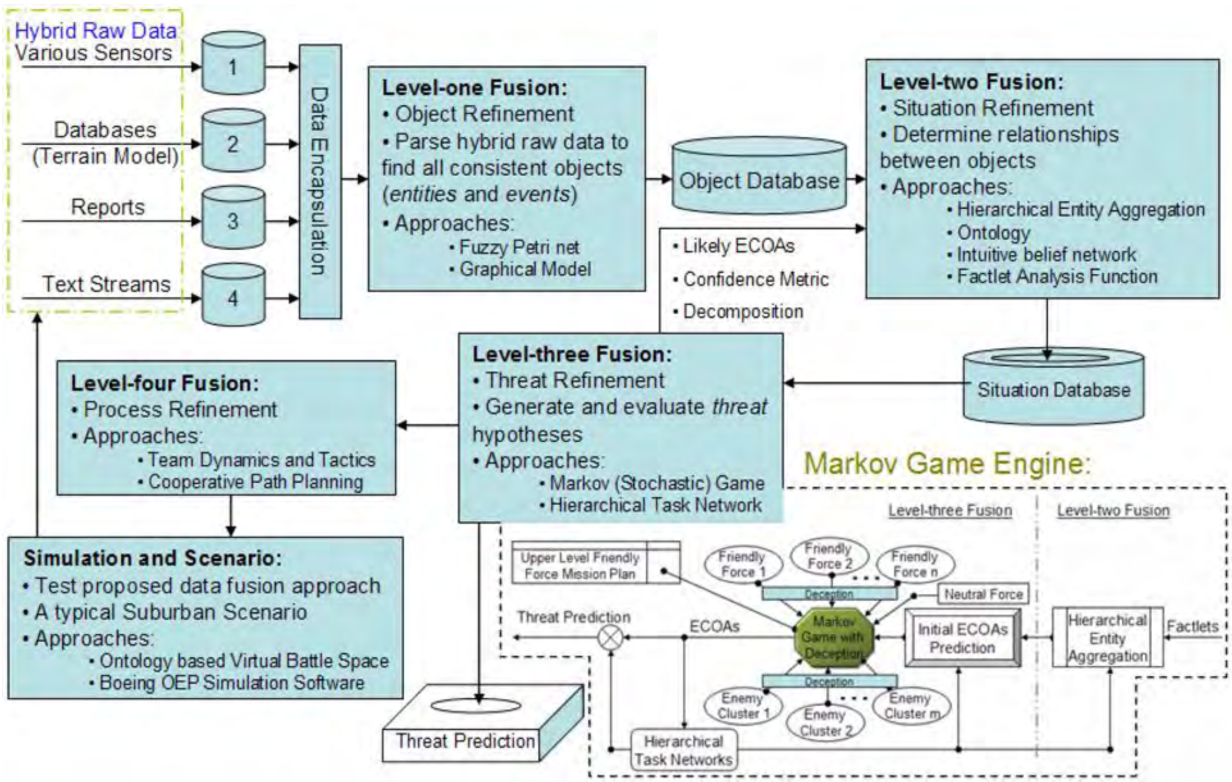


Fig. 1. The overall architecture. (The substructure of the Markov Game engine is also clearly shown in Fig. 2.)

ical context, etc. 3) Level-three fusion automates, infers and predicts the intentions and COAs of asymmetric threats. 4) Level-four fusion uses these COAs to task available sensor assets to optimally minimize cost of operations and decision response time. In particular, asymmetric and adaptive threats are detected and grouped by intelligent agent and Hierarchical Entity Aggregation in level-two fusion and their intents are predicted by a decentralized Markov (stochastic) game model with deception in level-three fusion. Game theory is not a new concept in military and cyber defense decision support. Existing game theoretic approaches [1] [2] [21] for threat detection and decision support are based on static matrix games and simple extensive games, which are usually solved by game trees. However, these matrix game models lack the sophistication to study multi-players with relatively large actions spaces, and large planning horizons. Recently, Brynielsson and Arnborg propose a game theoretic data fusion approach [30] via combining higher level command and control (C2) and Bayesian Network (BN) to solve multiple-decision-makers problems.

We have implemented Hierarchical Entity Aggregation and ontology-based Factlet Analysis Function to detect asymmetric treats at level-two fusion. Factlets are statements or evidence about the situation in the battlespace and they form the main input to the level-two fusion. We have implemented an adversary Markov game [23] model with three players: Red force (enemies), Blue force (friendly forces), and White force (neutral objects) at level-three fusion. Inherent information imperfection is considered and implemented in two methods: 1) the decentralized decision making scheme; and 2) deception with bounded rationality. We have modified our game theoretic sensor management algorithm at level-four fusion.

A software prototype has been developed with a display module based on the *Mixed-Initiative Control of Autonomous Unmanned Units under Uncertainty* (MICA) OEP [28] to integrate levels 1, 2, 3, and 4 data fusion and to demonstrate the performance of our proposed algorithms.

The paper is organized as follows. In Section 2, we will summarize the technical approach, which includes overall architecture, hierarchical entity aggregation at level-two fusion, and Markov game approach at level-three fusion. Section 3 describes the experimental results. Section 4 concludes the paper.

2. THREAT PREDICTION AND SITUATION ANALYSIS

2.1. Overall Structure

The overall architecture of our game theoretic data fusion is shown in Fig. 1. The level-one fusion builds the tracks of enemy targets from the reported data formatted by Data Encapsulation, which is the mechanism whereby the original data are kept hidden from the user and the user can only perform a restricted set of operations on the data. Level-one fusion also writes the Red target track table, which contains time, location, target type beliefs, and other information about each target. The tracks are based on data from the Blue Unmanned Air vehicle (UAV) and Airborne Warning and Control System (AWACS) sensors. Field reports from forward observers and signal intelligence contributes to event data. Level-one fusion establishes and maintains tracks for all ground vehicles, makes track-to-track associations, eliminates duplicates, and also initiates, maintains and drops tracks. The Blue tables of tracks of friendly armament resources contain similar information.

The level-two fusion (situation assessment–SA) performs spatial and temporal processing on tracks produced by level-one multi-sensor, multi-target track fusion, supplemented with intelligence information from both structured data sources such as databases and unstructured data sources such as ontology-based documents. At this level, Hierarchical Entity Aggregation, ontology and Factlet Analysis Function are used to cluster Red entities into groups by position, find the group centers-of-mass, build target group tables, and determine certain critical events and behaviors over time, which it formats into frame structures to pass to the level-three fusion process.

At level-three (threat assessment–TA) fusion, we investigated and demonstrated the effectiveness of Markov game theory. An adversarial Markov game framework is proposed for threat refinement to drive existing and newly formulated models of threat behavior with factlets derived from situation refinement to support the determination of possible enemy course of actions (ECOAs). An artificial intelligence planning concept, Hierarchical Task Network, is exploited to decompose the estimated ECOAs. The decompositions are fed back into and used in level-two fusion to identify and group the enemy entities that pose threats.

At level-four fusion (process refinement), the main tasks are to perform resource allocation and to provide feedback information for fusions at level 1, 2, and 3 to adjust the parameters. We use the method developed by the authors in a Navy funded on-going Phase II project named "Adaptive Cooperative Path and Mission Planning for Multiple Aerial Platforms."

We have conducted the implementation and analysis of several data fusion approaches at every JDL-model level, including conscious effort on the display technology to the user (as proposed in the Data Fusion Information Group (DFIG) [6]). We drive existing and newly formulated algorithms to support the determination of possible enemy COA. Asymmetric threats will be identified efficiently by Hierarchical Entity Aggregation at level-two fusion and assigned special payoff functions in our Markov Game framework at level-three fusion so that the intents of these irrational threats or entities will be efficiently predicted.

Due to page limitations, here we focus only on level-two and level-three data fusion and details can be found in the following subsections. A related paper summarizing our results with respect to level-one fusion algorithm will appear elsewhere.

2.2. Level-Two Fusion—Situation Refinement

The objectives of level-two fusion SA include estimation as to the measurements and observations that are available and establishing relationships between entities, events and the environment. An ontology-based battle-space modeling technique provides feasibility to the representation and organization of the environmental observations in a machine-readable manner. It also facilitates prediction of the potential relationships among the entities.

The Factlet Analysis Functions execute across the extent of the Virtual Battlespace as well as estimate across the objects present and within each analysis perspective, to generate both measured and inferred items of evidence, the "factlets." These Functions are concerned with establishing the "relationships" between objects in the Virtual Battlespace. For example, the Motion Analysis Function considers the movement patterns of groups (established by the Aggregate Analysis Function) of military objects such as armored personnel carriers. The Motion Analysis Function may conclude that the current movement pattern indicates a probing behavior on the part of the adversary, rather than a full scale attack. This prediction becomes a factlet.

In our data-fusion framework, Hierarchical Entity Aggregation [12] [1] [15] (HEA) is exploited to identify and group the entities that pose threats so that level-three TA fusion can be performed efficiently because of the following two major reasons. HEA reduces the ECOA hypothesis space for level-three fusion by reducing the number of potential "threats" to consider. In our approach, applying a Markov (stochastic) game theoretic algorithm to predict ECOA becomes more feasible. The other is that HEA can efficiently identify the asymmetric threats. Entity Aggregation plays an important role in subsequent fusion processing in the way that it provides aggregates that have the same tactical goal. For example, the capabilities and resources of a single terrorist are vastly different from the capabilities and resources of a team of terrorists. As a result, HEA will produce different results when considering a single terrorist or a team of terrorists as a threatening entity. To improve the performance of asymmetric adversary identification, we propose a feedback structure based on a Hierarchical Task Network (HTN) so that the revised asymmetric tactics and strategy can be decomposed and fed back to the HEA.

These identified asymmetric units with the associated aggregations will be handled and refined by

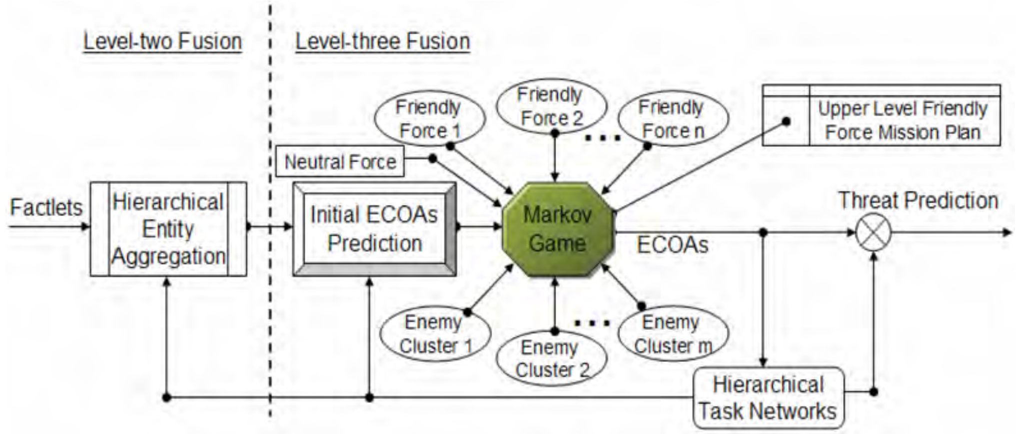


Fig. 2. Structure of level-three fusion (threat refinement).

our proposed Markov games in level-three TA data fusion.

2.3. Level 3 Data Fusion—Threat Refinement

2.3.1. A Decentralized Stochastic Game Theoretical Model

As shown in Fig. 2, a decentralized Markov game is used to model the evolution of ECOAs originated from an initial prediction based on Hierarchical Entity Aggregation.

A Markov (stochastic) game [23] is given by (i) a finite set of players N, (ii) a finite set of states S, (iii) for every player $i \in N$, a finite set of available actions D^i (we denote the overall decision space $D = \times_{i\in N} D^i$, where $\times$ is the multiplication operation), (iv) a transition rule $q: S \times D \to \Delta(S)$, (where $\Delta(S)$ is the space of all probability distributions over S), and (v) a payoff function $r: S \times D \to R^N$. For our threat prediction problem, we obtain the following discrete time Markov game:

Players (Decision Makers)—Although in our distributed (decentralized) Markov game model, each group (cluster, team) makes decisions, there are three main players: enemy, friendly force, and neutral players. All clusters of enemy (friendly force, or neutral) can be considered as a single player since they have a common objective.

State Space—All the possible COAs for enemy and friendly force consist of the state space. An element $s \in S$ is thus a sample of enemy and friendly force COAs composed of a set of triplets (resource, action verb, and objective). As an example, an enemy COA might be: the Red team 1 (resource) attacks (action verb) the Blue team 2 (objective). Similarly, for the friendly force COAs, resource is a friendly asset and objective is an adversary entity. Therefore, we can denote the state and state space as

$$s = (s^B, s^R, s^W)$$

$$S = S^B \times S^R \times S^W$$

where $s^B \in S^B$, $s^R \in S^R$, and $s^W \in S^W$ are the COAs of Blue (friendly) force, Red (enemy) force, and White (neutral) force, respectively.

$s^B = \{(r_i^B, a_i^B, o_i^B) \mid r_i^B \in R^B, a_i^B \in A^B, o_i^B \in O^B\}$ where R^B, A^B and O^B are the set of the resource, action, and objective of Blue force, respectively.

Similarly, the states for red force and white force are denoted as:

$$s^R = \{(r_i^R, a_i^R, o_i^R) \mid r_i^R \in R^R, a_i^R \in A^R, o_i^R \in O^R\}$$

$$s^W = \{(r_i^W, a_i^W, o_i^W) \mid r_i^W \in R^W, a_i^W \in A^W, o_i^W \in O^W\}$$

REMARK 1 It is well known that civilians often play an active role in wars. That is, they are not just passively static but might purposefully take actions to help one side in a battle to minimize their losses or achieve some political purpose. Unfortunately, existing game theoretic models usually do not consider this situation, although collateral damage has been considered in a paper on a two-player game model by Dr. Cruz et al. [10]. In this research, a three-player attrition-type discrete time dynamic game model is formulated with two opposing forces and one civilian player that might be either neutral or slightly biased. In our current implementation, the White units only care about their possible losses. For example, when a dangerous location is detected, normal White units will find a COA to keep themselves as far as possible from the harmful location. In the case where Red force poses as White for deceptive purpose, our algorithm will deem the Red force as White until abnormal activities or deceptions are detected.

Decision—At every time step, each Blue group chooses a list of targets with associated actions and confidences (note that: the probability distribution over the list of targets, i.e., the sum of the confidences should be equal to 1) based on its local battle field information, such as the unit type and positions of possible targets, from level-two data fusion. Let D_i^B denote the decision space of the ith Blue team. Each element d_i^B of D_i^B is defined as

$$d_i^B = \left\{(a_i^B, t_i^B, p_i^B) \mid a_i^B \in A^B, t_i^B \in O^B, 0 < p_i^B \leq 1, \sum p_i^B = 1\right\} \tag{1}$$

where p_i^B is the probability of the action-target couple (a_i^B, t_i^B), which is defined as the action a_i^B to target t_i^B. Therefore, the decision space of Blue $A^1 = \times_{i \in R^B} D_i^B$. (Compared with the standard definition of Markov game model reviewed in the beginning part of Section 2.3.1, D_i^B is the action set of ith member of Blue team, which is deemed as a single player. So, generally, the meaning of A^1 is same as that of D^1 in the standard definition.) As an example, for the Blue small weapon UAV 1 in Blue team 1, its action might be $d_1^B = \{$(attack, Red fighter 1, 0.3), (fly to, Red fighter 2, 0.5), (avoid, Red fighter 3, 0.2)$\}$.

Similarly, each Red cluster (obtained from the level-two data fusion) determines a probability distribution over all possible action-target combinations. Let D_i^R denote the decision space of the ith Red cluster. Each element d_i^R of D_i^R is defined as

$$d_i^R = \left\{(a_i^R, t_i^R, p_i^R) \mid a_i^R \in A^R, t_i^R \in O^R, 0 < p_i^R \leq 1, \sum p_i^R = 1\right\} \tag{2}$$

where p_i^R is the probability of action a_i^R to target t_i^R. Therefore, the decision space of Red force $A^2 = \times_{i \in R^R} D_i^R$. A possible action for Red platform 1 (Red fighter 1) is $d_1^R = \{$(attack, small weapon UAV 1, 0.6), (move to, Blue solider 2, 0.2), (avoid, Blue solider 1, 0.2)$\}$.

REMARK 2 Decision and action verbs are different concepts. A decision is a set of triplets with associated probabilities while an action verb is just a component of the triplet composed of resource, action verb and objective. All actions are included in A^1 for player 1 (Blue force) and A^2 for player 2 (Red force). All action verbs are enumerated in A^B for player 1 (Blue force) and A^R for player 2 (Red force).

The decisions of White objects are relatively simple. The main action type is movement. Let D_i^W denote the decision space of the ith White unit. Each element d_i^B of D_i^B is defined as

$$d_i^W = \left\{(a_i^W, t_i^W, p_i^W) \mid a_i^W \in A^W, t_i^W \in O^W, 0 < p_i^W \leq 1, \sum p_i^W = 1\right\} \tag{3}$$

where p_i^W is the probability of action a_i^W to target t_i^W.

Transition Rule—Due to the uncertainty properties of military environments, we assume that the states of the Markov game have inertia so that the decision makers have more chance in the pursuit of the objective from previous actions. We define an inertia factor vector for each player. Without loss of generality, we take the Blue force as an example, $\eta^B = (\eta_1^B, \eta_2^B, \ldots, \eta_{m_B}^B)^T$, where m_B is the number of the teams or clusters of Blue force, and $0 \leq \eta_j^B \leq 1$, $1 \leq j \leq m_B$. So, for the jth team of the Blue player, there is a probability of η_j^B to keep the current action-target couple and a probability of $(1-\eta_j^B)$ to use a new action composed of action-target couples.

There are two steps to calculate the probability distribution over the state space S, where s_k, s_{k+1} are states at time step k and $k+1$ respectively, and a_k^B, a_k^R and a_k^W are the decisions of Blue force, Red force, and White force, respectively, at time step k.

Step 1 With the consideration of a inertia factor vector η^B, we combine the current state with decisions of both players to obtain fused probability distributions over all possible action-target couples for the Red and Blue forces. To do this, we first decompose the current state into the action-target couples for each team of each player (Red force, Blue force, or White force). Let $\Psi_j^B(s_k)$ denote the resulting action-target couple related to the jth team of the Blue player. For example, if there is one triplet of (Blue team 1, attack, Red fighter 2) in the current state s_k, then the action-target couple for Blue team 1 (the first team of Blue force) is $\Psi_1^B(s_k)$ = (attack, Red fighter 2). Secondly, for each specified team, say the jth cluster of Blue player 2 (Blue force), we fuse its action-target couples via modifying the probability of each possible action-target couple based on the following formula

$$\bar{p}^B((a_j^B, t_j^B) \mid s_k) = \begin{cases} p_j^B(1-\eta_j^B), & \text{if } (a_j^B, t_j^B, p_j^B) \in d_j^B \\ & \text{and } (a_j^B, t_j^B) \notin \{\Psi_j^B(s_k)\} \\ p_j^B(1-\eta_j^B) + \eta_j^B, & \text{if } (a_j^B, t_j^B, p_j^B) \in d_j^B \\ & \text{and } (a_j^B, t_j^B) \in \{\Psi_j^B(s_k)\} \\ \eta_j^B, & \text{if } (a_j^B, t_j^B, p_j^B) \notin d_j^B \\ & \text{and } (a_j^B, t_j^B) \in \{\Psi_j^B(s_k)\} \\ 0, & \text{if } (a_j^B, t_j^B, p_j^B) \notin d_j^B \\ & \text{and } (a_j^B, t_j^B) \notin \{\Psi_j^B(s_k)\} \end{cases}. \tag{4}$$

There are four cases in Eq (4): 1) The action-target couple (a_j^B, t_j^B) only occurs in the current action of the jth cluster of the Blue player and is not in the current state s_k, which can be mathematically represented by $(a_j^B, t_j^B, p_j^B) \in d_j^B$ and $(a_j^B, t_j^B) \notin \{\Psi_j^B(s_k)\}$. Then we know the probability of (a_j^B, t_j^B) in current state s_k is 0 and probability of (a_j^B, t_j^B) in current action is p_j^B. So, according to the definition of inertia, the fused probability of the action-target couple (a_j^B, t_j^B) is $p_j^B(1-\eta_j^B) + 0(\eta_j^B) =$

$p_j^B(1-\eta_j^B)$. 2) The action-target couple (a_j^B, t_j^B) happens both in the current action of the jth cluster of the Blue player and in the current state s_k. Then we know the probability of (a_j^B, t_j^B) in the current state s_k is 1 and probability of (a_j^B, t_j^B) in the current action is p_j^B. So, according to the definition of inertia, the fused probability of the action-target couple (a_j^B, t_j^B) is $p_j^B(1-\eta_j^B) + 1(\eta_j^B) = p_j^B(1-\eta_j^B) + \eta_j^B$. 3) The action-target couple (a_j^B, t_j^B) only occurs in the current state s_k, and then we know the probability of (a_j^B, t_j^B) in current state s_k is 1 and probability of (a_j^B, t_j^B) in the current action is 0. So, according to the definition of inertia, the fused probability of the action-target couple (a_j^B, t_j^B) is $0(1-\eta_j^B) + 1(\eta_j^B) = \eta_j^B$. 4) The action-target couple (a_j^B, t_j^B) occurs neither in the current state s_k nor in the current action of the jth cluster of the Blue player, and then we know the probability of (a_j^B, t_j^B) in the current state s_k is 0 and probability of (a_j^B, t_j^B) in the current action is 0. So, according to the definition of inertia, the fused probability of the action-target couple (a_j^B, t_j^B) is $0(1-\eta_j^B) + 0(\eta_j^B) = 0$.

Similarly, the new probability distribution for the jth team of the Red player (Red force) is

$$\bar{p}^R((a_j^R, t_j^R) \mid s_k) = \begin{cases} p_j^R(1-\eta_j^R), & \text{if} \quad (a_j^R, t_j^R, p_j^R) \in d_j^R \\ & \text{and} \quad (a_j^R, t_j^R) \notin \Psi_j^R(s_k) \\ p_j^R(1-\eta_j^R) + \eta_j^R, & \text{if} \quad (a_j^R, t_j^R, p_j^R) \in d_j^R \\ & \text{and} \quad (a_j^R, t_j^R) \in \Psi_j^R(s_k) \\ \eta_j^R, & \text{if} \quad (a_j^R, t_j^R, p_j^R) \notin d_j^R \\ & \text{and} \quad (a_j^R, t_j^R) \in \Psi_j^R(s_k) \\ 0, & \text{if} \quad (a_j^R, t_j^R, p_j^R) \notin d_j^R \\ & \text{and} \quad p(a_j^R, t_j^R) \notin \Psi_j^R(s_k) \end{cases}. \tag{5}$$

The new probability distribution for jth team of White player (White force) is

$$\bar{p}^W((a_j^W, t_j^W) \mid s_k) = \begin{cases} p_j^W(1-\eta_j^W), & \text{if} \quad (a_j^W, t_j^W, p_j^W) \in d_j^W \\ & \text{and} \quad (a_j^W, t_j^W) \notin \{\Psi_j^W(s_k)\} \\ p_j^W(1-\eta_j^W) + \eta_j^W, & \text{if} \quad (a_j^W, t_j^W, p_j^W) \in d_j^W \\ & \text{and} \quad (a_j^W, t_j^W) \in \{\Psi_j^W(s_k)\} \\ \eta_j^W, & \text{if} \quad (a_j^W, t_j^W, p_j^W) \notin d_j^W \\ & \text{and} \quad (a_j^W, t_j^W) \in \{\Psi_j^W(s_k)\} \\ 0, & \text{if} \quad (a_j^W, t_j^W, p_j^W) \notin d_j^W \\ & \text{and} \quad (a_j^W, t_j^W) \notin \{\Psi_j^W(s_k)\} \end{cases}. \tag{6}$$

Step 2 We determine the probability distribution over the all possible outcomes of state s_{k+1},

$$q(s_{k+1} \mid s_k, a_k^B, a_k^R, a_k^W) = \prod_{j=1}^{m_B} \bar{p}^B((a_j^B, t_j^B) \mid s_k) \prod_{j=1}^{m_R} \bar{p}^R((a_j^R, t_j^R) \mid s_k) \times \prod_{j=1}^{m_W} \bar{p}^W((a_j^W, t_j^W) \mid s_k) \tag{7}$$

when

$$s_{k+1} = \bigcup_{j=1}^{m_B} \{(r_j^B, a_j^B, t_j^B)\} \bigcup_{j=1}^{m_R} \{(r_j^R, a_j^R, t_j^R)\} \bigcup_{j=1}^{m_W} \{(r_j^W, a_j^W, t_j^W)\},$$

otherwise, $q(s_{k+1} \mid s_k, a_k^B, a_k^R, a_k^W) = 0$. Where m_B is the number of the teams or clusters of the Blue player (Blue force), m_R is the number of the teams or groups of the Red player (Red force) and m_W is the number of the units of the White player (White force). $\{(r_i^B, a_i^B, t_i^B)\}$ is the set of all possible (with positive probability) triplets for the ith team of the Blue player. Therefore $\bigcup_{i=1}^{m_B} \{(r_i^B, a_i^B, t_i^B)\}$ contains all the possible (with positive probability) triplets for the Blue force. From step 1, we know that the fused probability of each specified (a_j^B, t_j^B) is $\bar{p}^B((a_j^B, t_j^B) \mid s_k)$ defined in equation (1). With the assumption that all teams of Blue force are independent, we obtain the overall probability of Blue force, $\prod_{j=1}^{m_B} \bar{p}^B((a_j^B, t_j^B) \mid s_k)$. Similarly, $\prod_{j=1}^{m_R} \bar{p}^R((a_j^R, t_j^R) \mid s_k)$ and $\prod_{j=1}^{m_R} \bar{p}^W((a_j^W, t_j^W) \mid s_k)$ are the overall probabilities of the Red and White force, respectively. So the probability distribution over the all possible outcomes of state s_{k+1} (composed of all possible sub-states of Blue, Red, and White force) can be calculated via equation (7).

Payoff Functions—In our proposed decentralized Markov game model, there are two levels of payoff function for each player (Blue, Red or White).

The lower (local) level payoff functions are used by each team or cluster to determine the team actions based on the local information. For the jth team of Blue force, the payoff function is defined as $f_j^B(\tilde{s}_j^B, d_j^B, W_k^B)$, where $\tilde{s}_j^B \subseteq s$ is the local information (note that in a distributed and partial observable framework, local information for each player means the battle or state information is available to the player.) obtained by the team, and W_k^B, the weights for all possible action-target couples of Blue force, is announced to all Blue teams and determined according the top level payoff functions by the supervisor of Blue force.

$$f_j^B(\tilde{s}_j^B, d_j^B, W_k^B) = \sum_{(a_i^B, t_i^B, p_i^B) \in d_j^B} w^B(j, a_i^B, t_i^B, W_k^B) p_i^B g^B(j, a_i^B, t_i^B, \tilde{s}_j^B) \tag{8}$$

where, $w^B(j,a_i^B,t_i^B,W_k^B)$ will calculate the weigh for any specified action-target couple for the jth team of Blue force from the W_k^B, p_i^B is the probability of the action-target couple (a_i^B,t_i^B), and $g^B(j,a_i^B,t_i^B,\tilde{s}_j^B)$ will determine the gain from the action-target couple (a_i^B,t_i^B) for the jth team of Blue force according to the positions and features, such as platform values and defense/offense capability, of the Blue and Red platforms. Similarly, we obtain the lower level payoff functions for the jth team of Red or enemy force,

$$f_j^R(\tilde{s}_j^R,d_j^R,W_k^R) = \sum_{(a_i^R,t_i^R,p_i^R)\in d_j^R} w^R(j,a_i^R,t_i^R,W_k^R)p_i^R g^R(j,a_i^R,t_i^R,\tilde{s}_j^R) \tag{9}$$

$$f_j^W(\tilde{s}_j^W,d_j^W,W_k^W) = \sum_{(a_i^W,t_i^W,p_i^W)\in d_j^W} w^W(j,a_i^W,t_i^W,W_k^W)p_i^W g^W(j,a_i^W,t_i^W,\tilde{s}^W). \tag{10}$$

REMARK 3 For some asymmetric threats, such as suicide bombers, the payoff functions may only consider the loss of the Blue side. For some camouflage and concealment entities, their objectives are to hide themselves and move close to the Blue units. Other deception units will do some irrational and additional movements to hide their true goals.

REMARK 4 People usually think of a military conflict situation as a zero-sum game—a game with a winner and a loser. In zero-sum game theory, the players have opposite objectives. If one player maximizes an objective function, the other automatically minimizes it. This is equivalent to a player maximizing an objective function and the other player maximizing the negative of the same function. Since the sum of the objective functions is zero, the game is called a zero-sum game. But when there are significant differences between the cultures of the Red and Blue forces and significant differences in the valuations of their assets and their opponent's assets, the zero-sum game approach in general is not representative. For example, a Blue objective might be to preserve as much of the Blue assets and to destroy as much of the Red assets as possible. However, recent experience with terrorist type battles suggests that the Red force may not be as concerned as the Blue force with preserving its own assets. The objectives in such a situation are not opposite of each other and a nonzero-sum approach would be much more appropriate.

The top (global) level payoff functions are used to evaluate the overall performance of each player.

$$J^B = \sum_k \left[\sum_{j=1}^{m_B} f_j^B(\tilde{s}_j^B,d_j^B,W_k^B)\right] \tag{11}$$

$$J^R = \sum_k \left[\sum_{j=1}^{m_R} f_j^R(\tilde{s}_j^R,d_j^R,W_k^R)\right] \tag{12}$$

$$J^W = \sum_k \left[\sum_{j=1}^{m_W} f_j^W(\tilde{s}_j^W,d_j^W,W_k^W)\right] \tag{13}$$

where k is the time index. In our approach, the calculation of the lower level payoffs are distributed and sent back to commander/supervisor via communication networks.

REMARK 5 Since the gain functions $g^B(j,a_i^B,t_i^B,\tilde{s}_j^B)$ for Blue force, $g^R(j,a_i^R,t_i^R,\tilde{s}_j^R)$ for Red force and $g^W(j,a_i^W,t_i^W,\tilde{s}_j^W)$ for White force are different functions, asymmetric force and cost utilities can be straightforwardly represented in our model. In addition, after an irregular adversary is detected, a different type of gain function will be assigned dynamically.

REMARK 6 In our Markov game model, the states used in the control strategies are the estimates of the future systems states. These estimates will evaluate or update following the Markov decision process in the Markov game framework, in which the interactions are considered. At each time k, the process will be repeated based on the observed current system states.

Strategies—In this project, we have tried several well known types of strategies. Here we only give a brief description about three of them:

Pure Nash Strategies with a finite horizon. In game theory, the Nash equilibrium (named after John Nash [17] who proposed it) is a kind of optimal collective strategy in a game involving two or more players, where no player has anything to gain by changing only his or her own strategy. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep their's unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium. In our approach, we use a game search tree to find the solution.

Mixed Nash Strategies. A mixed strategy is used in game theory to describe a strategy comprised of possible actions and an associated probability, which corresponds to how frequently the action is chosen. Mixed strategy Nash equilibria are equilibria where at least one player is playing a mixed strategy. Nash proved that that every finite game has Nash equilibria but not all have a pure strategy Nash equilibrium.

Correlated Equilibria [26]. Unlike Nash equilibria, which are the concept of equilibria formulated in independent strategies, correlated equilibria were developed from correlated strategies in non-cooperative games. The correlated equilibrium of a Markov game describes a solution for playing a dynamic game in which players are able to communicate but are self-interested. Based on the signals, which are generated by the correlated devices and announced to each decision maker, players choose their actions according to the received private signals. There are two types of correlation devices: autonomous and stationary devices. An autonomous correlation device is a pair $\mathcal{D} = (((M_n^i)_{i\in N}, d_n)_{n\in N})$, where (i) M_n^i is a finite set of signals for player i at time step n, and (ii) $d_n : M(n) \to \Delta(M_n)$, $M_n = \times_{i\in N} M_n^i$ and $M(n) = M_1 \times M_2 \times \cdots \times M_{n-1}$. A stationary correlation device is a pair $\mathcal{D} = (((M^i)_{i\in N}, d))$, where $d \in \Delta(M)$ and $M = \times_{i\in N} M^i$. Actually, a stationary correlation device is a special case of an autonomous correlation device, where M_n^i is independent of n and d_n is a constant function that is independent of n.

Given a correlation device $\mathcal{D}$, we define an extended game $G(\mathcal{D})$. The game $G(\mathcal{D})$ is played exactly as the original game, but at the beginning of each stage n, a signal combination $m_n = (m_n^i)_{i\in N}$ is drawn according to the probability function $d_n(m_1, m_2, \ldots, m_{n-1})$ and each player i is informed of m_n^i. Then each decision maker must base his choice of actions on the received signal. Any deviator will be punished via his min-max value. The punishment only occurs if a player disobeys the recommendation of the device. Every Markov game with an autonomous correlated device admits a correlated equilibrium [26].

REMARK 7 In our proposed approach, the solution to the Markov game model is obtained via a K time-step look-ahead approach, in which we only optimize the solution in the K time-step horizon. We set K as 5 during the simulations of the Section 3—Experiments. Actually, this suboptimal technique is used successfully for calculations in games such as chess, backgammon, and monopoly.

2.3.2. Hierarchical Task Network

Once the ECOA hypotheses have been generated, they must be evaluated. However, since the generated hypotheses are not directly observable, they are not suitable for correctness testing. As with any hypothesis test, observables must be identified. These observables act as indicators to refute or support ECOA hypotheses. A Hierarchical Task Network (HTN) planner [11] is employed to decompose ECOA hypotheses into observable task sequences.

A construct known as the Hierarchical Task Network (HTN) provides a representation of tasks at various levels of specificity. The HTN not only mimics the variation in specificity found in military echelons, it also allows a computational construct for analyzing ECOAs. In our game theoretic approach to level-three fusion (threat assessment), the HTN is employed to provide a method for decomposing high-level ECOAs into more specific tasks. The HTN representation is the basis of most modern planning algorithms. It is based on the concept that humans plan by decomposing tasks into smaller ones until a sequence of tractable tasks are found that satisfy the objective [7]. These are tasks that the fusion processes attempt to infer or observe directly and are assumed to be tractable.

3. EXPERIMENTS

In the simulation part, we build a virtual battle-space and a typical urban scenario based on the ontology concept, which is an explicit, formal, machine-readable semantic model that defines the classes (or concepts) and their possible inter-relations specific to some specified domain. To simulate our data fusion approach, we implemented and tested our battle-space, scenario and algorithms on our prototype software with developed and funded cooperative path planning and mission planning algorithms [8], [9], [24].

3.1. Scenario Description

We used a scenario shown in Fig. 3 to demonstrate the performance of our proposed threat prediction and situation awareness algorithm. In the shown urban environment, the Blue force's missions are to capture two bridges and to do security patrol on the main roads connecting the two bridges. The Blue ground force consists of 3 teams of three soldiers each with sniper rifles. The Red force includes 3 armed fighters and 3 asymmetric adversaries hiding in and acting like the White objects (the civilians and vehicles). We assume there is an asymmetry in total forces between Blue side and Red side. Blue has more soldiers than Red. Moreover, the objectives of Blue side and Red side are asymmetric: the objectives of Red side are to kill Blue forces without considering the loss of themselves and the consideration of collateral damage. The main challenge for both sides is to understand the situation from the fused sensor data and predict the intent of the opponent under the "believed" war situation.

REMARK 8 In this scenario, the kill probability (of each weapon type) and the target value of each unit (Blue, Red, and White force) are pre-specified.

3.2. Implementation

To demonstrate our approach, we developed simulation software (Fig. 4) as a controller module for the MICA (Mixed Initiative Control of Automa-Teams) Open Experimental Platform (OEP) [28].

For the scenario (Fig. 3), the possible actions for blue side are "Blue Team 1 move to Bridge 1," "Blue Team 2 Attack Red Fighter 2," or "Blue Team 3 Halt." In general, R^B = {Blue Team 1, Blue Team 2, Blue Team 3}, A^B = {Move to, Attack, Halt}, and O^B = {Red Fighter 1, Red Fighter 2, Red Fighter 3, Bridge 1, Bridge 2, Dummy, Detected Asymmetric Threats}. The possible actions for red force are "Fighter 1 attack Blue team 1," "Asymmetric threat acts as a civilian." In general, R^R = {Fighter 1, Fighter 2, Fighter 3, Asymmetric threat}, A^R = {Move to, Attack, Act as a civilian, Halt}, and O^R = {Blue team 1, Blue team 2, Blue team

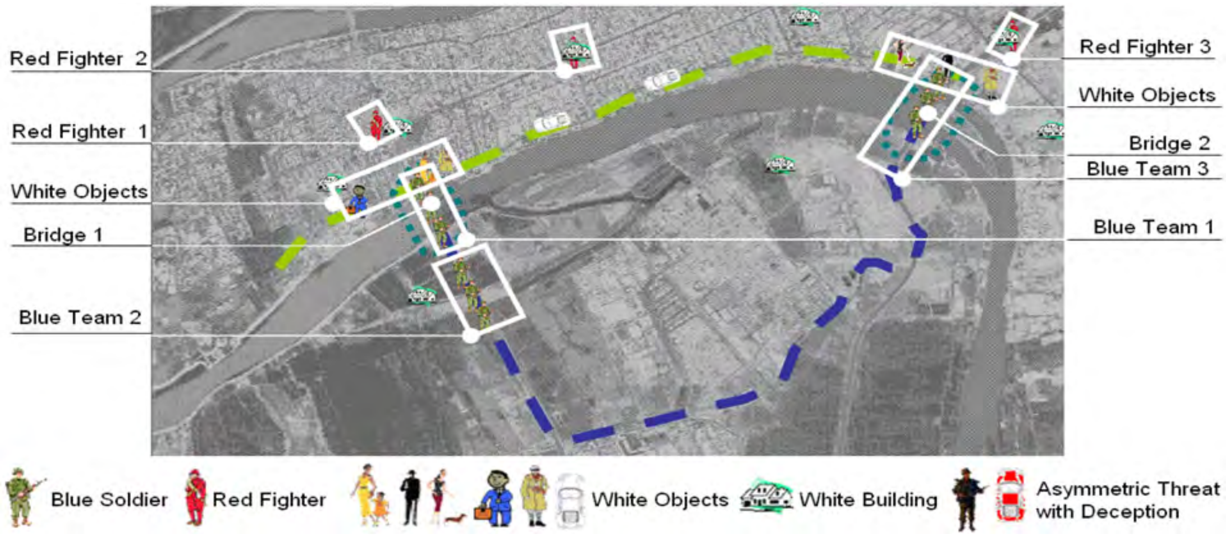


Fig. 3. A simulated scenario—urban warfare for combating guerrilla forces.

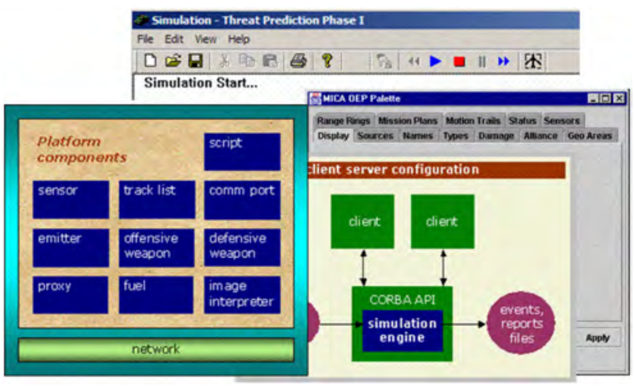


Fig. 4. Simulation Software—a controller module for MICA OEP virtual battlespace.

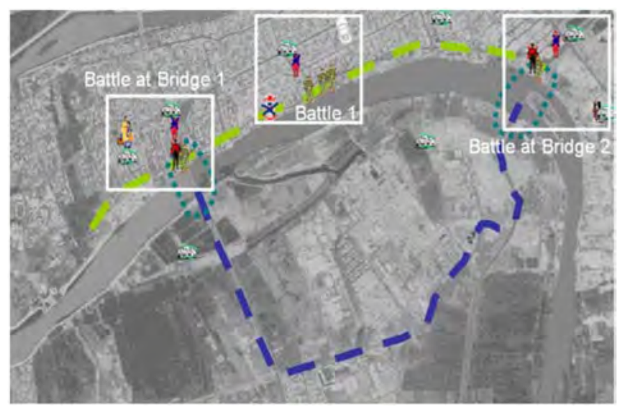


Fig. 5. Result of a simulation run.

3, Bridge 1, Bridge 2, Dummy}. The actions for White side include "Civilian 1 move to safe place," "Civilian 2 move to Bridge 1," and "Civilian 3 move to dangerous place." Mathematically, $R^W = \{$Civilian 1, Civilian 2,...$\}$, $A^W = \{$Move to, Halt$\}$, and $O^R = \{$Bridge 1, Bridge 2, Dummy, Safe place, Dangerous place$\}$.

In this simulation we set all inertia values to 0.1 and we also assume that the there is no measurement error for the Blue, Red, and White forces.

The objective of the Blue force is to save Bridge 1, Bridge 2, Blue teams, and Civilians; and eliminate Red Fighter 1, Red Fighter 2, Red Fighter 3 and possible asymmetric threats. The goal of the Red side is to Destroy Bridge and Kill Blue teams (we assume that Red force has to kill Blue teams nearby before destroying Bridge 1 or 2). The White force's goal is to protect civilians. Each side will estimate the information of damage status (probability and expectation value) and calculate its cost function based on the unit values: Bridge (100), Blue team (50), Red Fighter (20), Asymmetric threat (50), Civilian (0 for "don't care about collateral damage" or 10 for "care about collateral damage"). We set the kill probability to 0.5.

To solve the Markov game problem, we have conducted a numerical procedure to compute the strategies with a K-step look-ahead horizon. We first convert the Markov game to several MDPs (one MDP for each player with every possible combination of K-step strategies of the other players) and several one-step static matrix games (one game for each player at every current system state). Then existing algorithms (MDP MATLAB toolbox and Gambit [29]) will be exploited to solve the MDPs and matrix games.

3.3. Experiments

For the scenario, in a specific simulation run (Markov game approach with correlated equilibrium) as shown in Fig. 5, Blue team 1 and Blue team 3 were assigned to secure Bridge 1 and Bridge 2, respectively, almost for the whole simulation period of 30 minutes. Blue team with 3 Blue soldiers was doing security patrol on the two major roads connecting two bridges and

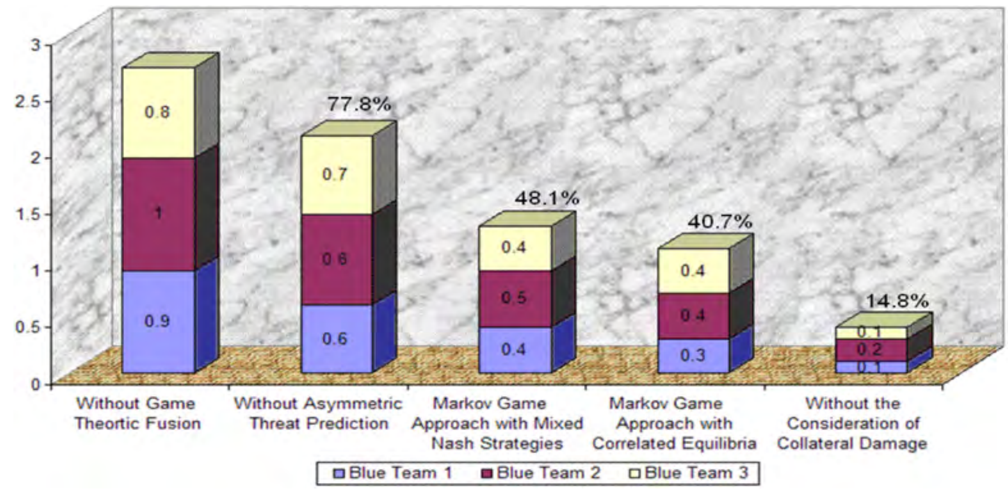


Fig. 6. Damage comparison of various options.

some important areas. On the other hand, Red fighters and asymmetric adversaries are trying their best to kill Blue forces. The first battle happened when Red Fighter 2 tried to attack Blue Team 2 with the help of an asymmetric White vehicle with deception (hiding in White vehicles). During this period, one asymmetric adversary vehicle, which posed civic activities at first and carried out abnormal activities during the battle, is detected and killed. Without the help of the Red vehicle, Red fighter 2 was killed by Blue team 2. Almost at the same time, the asymmetric adversaries near Bridge 1 and Bridge 2 were attacking the Blue team 1 and 2. At this stage, two civilians were detected and killed as asymmetric adversaries. Without the help from the asymmetric adversaries with deception, Red fighter 1 and 3 were killed by Blue team 1 and 3 at Bridges 1 and 2, respectively. In this specific run, there is no loss of Blue soldiers since our algorithm predicted the intents of the Red side correctly and promptly.

In addition to the explained run, we performed many experiments. We compared the results using the various options, such as without game theoretic fusion (without level-two or level-three fusion, and a Bayesian Network approach), without asymmetric-threat prediction (with level-two fusion but the payoff function of game model at level-three fusion doesn't change dynamically), game approach with mixed Nash strategy, game approach with correlated equilibria, and the game approach without collateral damage consideration in the cost function of Blue side. Since the simulation is stochastic, the results consist of the mean of 10 runs for each case, which are shown in Fig. 6 (Only the damage information for the Blue side is shown). From the damage comparison results, we can see that our Markov game approach with correlated equilibrium and deception consideration for threat detection and situation awareness is better than the other methods except the game approach without collateral damage consideration.

4. CONCLUSIONS

Game theoretic tools have a potential for threat prediction that takes uncertainties in Red plans and deception possibilities into consideration. In this paper, we have evaluated the feasibility of the Markov game theoretic data fusion algorithm. The effectiveness has been demonstrated through extensive simulations. The scalability and stability analysis of our game theoretic approach is one direction of future research.

ACKNOWLEDGMENTS

This research was supported by the U.S. Navy under contract number N00014-05-M-0205. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Navy.

REFERENCES

[1] A. Agah, S. K. Das and K. Basu
A non-cooperative game approach for intrusion detection in sensor networks.
Vehicular Technology Conference, VTC2004-Fall, 2004, 2902–2906.

[2] T. Alpcan and T. Basar
A game theoretic application to decision and analysis in Network Intrusion Detection.
In *Proceedings of 42nd IEEE Conference on Decision and Control*, Maui, HI, 2003, 2595–2600.

[3] D. Barbará, Y. Li and J. Couto
COOLCAT: An entropy-based algorithm for categorical clustering.
In *Proceedings of the 7th International Conference on Information and Knowledge Management*, McLean, VA, Nov. 04–09, 2002.

[4] T. Basar and G. J. Olsder
Dynamic Noncooperative Game Theory.
SIAM Series in Classics in Applied Mathematics, Philadelphia, PA, (2nd ed.), Jan. 1999.

[5] BBN Technologies
Homepage: http://openmap.bbn.com/.

[6] E. Blasch and S. Plano
DFIG level 5 supporting situation assessment reasoning.
In *Proceeding of the 8th International Conference on Information Fusion*, Philadelphia, PA, July 25–29, 2005.

[7] M. G. Ceruti and M. Bowman
Expertise, pattern management, and decision making: Challenges in human informatics.
In *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics* (SCI 2001) and *the 7th International Conference on Information Systems Analysis and Synthesis* (ISAS 2001), Orlando, FL, July 22–25, 2001.

[8] G. Chen, C. Kwan, D. Shen, J. B. Cruz and A. Vannevel
Adaptive cooperative path planning for multiple platforms.
In *Proceedings of IEEE International Conference on Robotics and Biomimetics* (ROBIO 2005), Hong Kong and Macau, June 29–July 3, 2005.

[9] G. Chen, D. Shen, J. B. Cruz, C. Kwan, S. Cox, S. P. Riddle and C. Matthews
A novel cooperative path planning for multiple aerospace vehicles.
In *Proceedings of AIAA infotech@aerospace*, Arlington, VA, Sept. 26–29, 2005.

[10] J. B. Cruz, Jr., G. Chen, D. Garagic, X. Tan, D. Li, D. Shen, M. Wei and X. Wang
Team dynamics and tactics for mission planning.
In *Proceedings of IEEE Conference on Decision and Control*, Maui, HI, Dec. 9–12, 2003.

[11] K. Erol, D. Nau and J. Hendler
HTN planning: Complexity and expressivity.
In *Proceedings of The 12th National Conference on Artificial Intelligence*, Seattle, WA, July 31–Aug. 4, 1994.

[12] W. J. Farrell, III
Hierarchical entity aggregation for situation assessment: Identifying asymmetric aggregates.
Military Sensing Symposia, National Symposium on Sensor and Data Fusion, Laurel, MD, June 2004.

[13] D. Kettani and J. Roy
A qualitative spatial model for information fusion and situation analysis.
In *Proceedings of the 3rd International Conference on Information Fusion*, Paris, France, July 10–13, 2000.

[14] J. Llinas, C. L. Bowman, G. Rogova, A. N. Steinberg, E. Waltz and F. E. White
Revisiting the JDL data fusion model II.
In *Proceedings of the 7th International Conference on Information Fusion*, Stockholm, Sweden, June 28–July 1, 2004.

[15] C. Looney and L. Liang
Cognitive situation and threat assessments of ground battlespaces.
In *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Queensland, Australia, Aug. 8–11, 2003.

[16] C. J. Matheus, K. P. Baclawski and M. M. Kokar
Derivation of ontological relations using formal methods in a situation awareness scenario.
In *Proceedings of SPIE—Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003*, vol. 5099, Apr. 2003, 298–309.

[17] J. Nash
Noncooperative games.
Annals of Mathematics, **54** (1951), 286–295.

[18] A. Pawlowski, S. Gigili and Frank Vetesi
Situation and threat refinement approach for combating the asymmetric threat.
Military Sensing Symposia, National Symposium on Sensor and Data Fusion, San Diego, CA, Aug. 13–15, 2002.

[19] J. J Salerno, E. Blasch, M. Hinman and D. M. Boulware
Evaluating algorithmic techniques in supporting situation awareness.
In *Proceeding of the 8th International Conference on Information Fusion*, Philadelphia, PA, July 25–29, 2005.

[20] J. J. Salerno, M. Hinman and D. Boulware
Building a framework for situation awareness.
In *Proceedings of the 7th International Conference on Information Fusion*, Stockholm, Sweden, June 28–July 1, 2004.

[21] K. Sallhammar, S. J. Knapskog and B. E. Helvik
Using stochastic game theory to compute the expected behavior of attackers.
In *Proceedings of 2005 Symposium on Applications and the Internet Workshops*, 2005.

[22] J. Schubert
Evidential force aggregation.
In *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Queensland, Australia, Aug. 8–11, 2003.

[23] L. S. Shapley
Stochastic games.
In *Proceedings of the National Academy of Sciences of the United States of America*, vol. 39, Oct. 1953, 1095–1100.

[24] D. Shen, J. B. Cruz, G. Chen, C. Kwan, S. Cox, S. P. Riddle and C. Matthews
A game theoretic approach to mission planning for multiple aerial platforms.
In *Proceedings of AIAA infotech@aerospace*, Arlington, VA, Sept. 26–29, 2005.

[25] S. Singh, W. Donat, H. Tu, K. Pattipati and P. Willett
Anomaly detection via feature-aided tracking and hidden Markov models.
IEEE Aerospace Conference, Big Sky, MT, Mar. 3–10, 2007.

[26] E. Solan and N. Vieille
Correlated equilibrium in stochastic games.
Games and Economic Behavior, **38**, 2 (Feb. 2002), 362–399.

[27] N. Steinberg, C. L. Bowman and F. E. White
Revisions to JDL data fusion model.
In *Proceedings of the SPIE Conference on Sensor Fusion: Architectures, Algorithms, and Applications*, vol. 3719, Sunnyvale, CA, July 1999.

[28] User Guide for the Open Experimental Platform (OEP), version 1.3, Boeing, Mar. 2003.

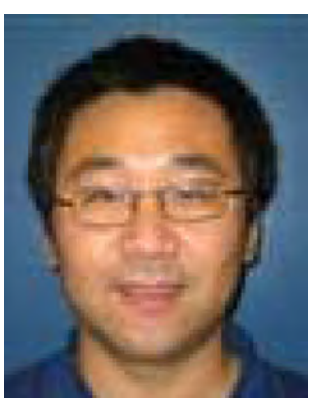
[29] R. D. McKelvey, A. M. McLennan and T. L. Turocy
Gambit: Software Tools for Game Theory.
Version 0.2007.01.30, http://gambit.sourceforge.net.

[30] J. Brynielsson and S. Arnborg
An information fusion game component.
Journal of Advances in Information Fusion, **1**, 2 (Dec. 2006), 108–121.

Genshe Chen received the B.S. and M.S. in electrical engineering, Ph.D. in aerospace engineering, in 1989, 1991 and 1994 respectively, all from Northwestern Polytechnical University, Xian, P. R. China.

Currently Dr. Chen is the Vice President and CTO of DCM Research Resources LLC, Germantown, MD, where he directs the research and development activities for Government Services and Commercial Solutions. Prior to founding DCM Research Resources, he was the program manager in Networks, Systems and Control at Intelligent Automation, Inc., leading research and development efforts in target tracking, information fusion and cooperative control. He was a Postdoctoral Research Associate in the Department of Electrical and Computer Engineering of the Ohio State University from 2002 to 2004. He worked at the Institute of Flight Guidance and Control of the Technical University of Braunshweig (Germany) as an Alexander von Humboldt research fellow and at the Flight Division of the National Aerospace Laboratory of Japan as a STA fellow from 1997 to 2001. He did postdoctoral work at the Beijing University of Aeronautics and Astronautics and Wright State University from 1994 to 1997.

He has served as the Principal Investigator/Technical lead for about 40 projects, including more than 25 U.S. government projects such as maneuvering target detection and tracking, cooperative control for teamed unmanned aerial vehicles, a stochastic differential pursuit-evasion game with multiple players, multi-missile interception, asymmetric threat detection and prediction, space situation awareness, cyber defense, and space-time adaptive processing, etc. His technical expertise also includes game theoretic estimation and control, threat prediction and information fusion, guidance and control of manned and unmanned vehicles, GPS/INS/image integrated navigation system, computational intelligence and data mining, hybrid system theory and Markov chain, signal processing and computer vision, pattern recognition, biometrics, Bayesian networks and influence diagrams, social network analysis, simulation and training, and GIS. Dr. Chen has about 100 professional publications.



Dan Shen received the B.S. degree in Automation from Tsinghua University, Beijing, China, in 1998, the M.S. and Ph.D. degree in electrical engineering from the Ohio State University (OSU), Columbus, in 2003, 2006, respectively. Currently, he is a research scientist at Intelligent Automation, Inc., Rockville, MD. From 1998 to 2000, he was with Softbrain Software Co., Ltd., Beijing, China, as a Software Engineer. From September 2005 to March 2006, he was an intern at Intelligent Automation, Inc. His research interests include game theory and its applications, optimal control, and adaptive control.

Chiman Kwan's primary research areas include robust and adaptive control methods, digital signal and image processing, neural networks, flight control and simulation, and fuzzy logic control. Dr. Kwan received his Ph.D. in May 1993 and already has had 39 journal papers published in archival journals. He has had about 90 additional papers published in major conference proceedings. He is currently the Vice President of Research & Development at IAI, leading research in signal/image processing and control. Before joining IAI, he used to work for SSC (Superconducting Super Collider Lab.) from April 1991 to February 1994, where he was heavily involved in the modeling, simulation, and design of modern digital controllers and signal processing algorithms for the beam control system. He received an invention award for his work at SSC. After the demise of SSC, he joined the Automation and Robotics Research Institute in Fort Worth where he applied intelligent control methods such as neural networks and fuzzy logic to the control of power systems, robots, and motors.

Jose B. Cruz, Jr. received his B.S. degree in electrical engineering (summa cum laude) from the University of the Philippines (UP) in 1953, the S.M. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge in 1956, and the Ph.D. degree in electrical engineering from the University of Illinois, Urbana-Champaign, in 1959. He is currently a Distinguished Professor of Engineering and Professor of Electrical and Computer Engineering at the Ohio State University (OSU), Columbus. Previously, he served as Dean of the College of Engineering at OSU from 1992 to 1997, Professor of electrical and computer engineering at the University of California, Irvine (UCI), from 1986 to 1992, and at the University of Illinois from 1965 to 1986. He was a Visiting Professor at MIT and Harvard University, Cambridge, in 1973 and Visiting Associate Professor at the University of California, Berkeley, from 1964 to 1965. He served as Instructor at UP in 1953–1954, and Research Assistant at MIT from 1954 to 1956. He is the author or coauthor of six books, 21 chapters in research books, and numerous articles in research journals and refereed conference proceedings.

Dr. Cruz was elected as a member of the National Academy of Engineering (NAE) in 1980. In 2003, he was elected a Corresponding Member of the National Academy of Science and Technology (Philippines). He is also a Fellow of the American Association for the Advancement of Science (AAAS), elected 1989, a Fellow of the American Society for Engineering Education (ASEE), elected in 2004, and a Fellow of International Federation of Automatic Control (IFAC), appointed 2007. He received the Curtis W. McGraw Research Award of ASEE in 1972 and the Halliburton Engineering Education Leadership Award in 1981. He is a Distinguished Member of the IEEE Control Systems Society and received the IEEE Centennial Medal in 1984, the IEEE Richard M. Emberson Award in 1989, the ASEE Centennial Medal in 1993, and the Richard E. Bellman Control Heritage Award, American Automatic Control Council (AACC), 1994. In addition to membership in NAE, ASEE, and AAAS, he is a Member of the Philippine American Academy for Science and Engineering (Founding member, 1980, President 1982, and Chairman of the Board, 1998–2000), Philippine Engineers and Scientists Organization (PESO), National Society of Professional Engineers, Sigma Xi, Phi Kappa Phi, and Eta Kappa Nu. He served as a Member of the Board of Examiners for Professional Engineers for the State of Illinois, from 1984 to 1986. He served on various professional society boards and editorial boards, and he served as an officer of professional societies, including IEEE, where he was President of the Control Systems Society in 1979, Editor of the IEEE Transactions on Automatic Control, a Member of the Board of Directors from 1980 to 1985, Vice President for Technical Activities in 1982 and 1983, and Vice President for Publication Activities in 1984 and 1985. Currently, he serves as Chair (2004–2005) of the Engineering Section of the American Association for the Advancement of Science (AAAS).

Martin Kruger is currently serving as the Intelligence, Surveillance and Reconnaissance Thrust Area Manager for the Expeditionary Warfare Maneuver Warfare & Combating Terrorism Science and Technology Department at the Office of Naval Research. In that capacity, he is responsible for maturing and transitioning applicable technology. Research interests include sensing, data fusion & visualization, resource management and information dissemination. The overall objective of the program is to increase the efficiency and effectiveness of the translation of intelligence requirements to actionable intelligence relevant to the Global War on Terror.

Before coming to ONR, Mr. Kruger served as a research and development manager for the Future Theater Air and Missile Defense program office at the Naval Sea Systems Command. He has also worked for the Marine Corps Warfighting Laboratory and for the Naval Surface Warfare Center Indian Head Division. Mr. Kruger started his career as a Naval Officer, serving as an instructor at the Naval Nuclear Propulsion School.

After leaving active duty, Captain Martin Kruger has continued serving the Navy as a drilling reservist. Reserve assignments have included four command tours, one each at a shipyard, a SUPSHIP, a NAVSEA field activity and a Weapon Station. He is currently serving as a Chief Ordnance Inspector.

Martin Kruger holds Bachelor in engineering in Chemical Engineering, a Master of Science in Industrial Chemistry and a Masters in Business Administration. He is also a graduate of the Naval War College and is Level 3 Certified in Program Management.

Erik Blasch received his B.S. in mechanical engineering from MIT and Masters in mechanical and industrial engineering from Georgia Tech and MBA, MSEE, from Wright State University and a Ph.D. from WSU in EE. Dr. Blasch also attended Univ of Wisconsin for an MD/PHD in Mech. Eng. until being called to Active Duty in the United States Air Force. Currently, he is a Fusion Evaluation Tech Lead for the Air Force Research Laboratory, Adjunct Professor at WSU, and a reserve Major with the Air Force Office of Scientific Research.

Dr. Blasch was a founding member of the International Society of Information Fusion (ISIF) and the 2007 ISIF President. Dr. Blasch has many military and civilian career awards; but engineering highlights include team member of the winning '91 American Tour del Sol solar car competition, '94 AIAA mobile robotics contest, and the '92 AUVs competition where they were first in the world to automatically control a helicopter. Since that time, Dr. Blasch has focused on Automatic Target Recognition, Targeting Tracking, and Information Fusion research compiling 200+ scientific papers and book chapters. He is active in IEEE and SPIE including regional activities, conference boards, journal reviews and scholarship committees.

Track-to-Track Association Using Attributes

YAAKOV BAR-SHALOM
University of Connecticut

HUIMIN CHEN
University of New Orleans

The problem of track-to-track association—a prerequisite for the fusion of tracks—has been considered in the literature for tracks described by kinematic states and, more recently, has been generalized to include additional (continuous valued) feature and (discrete valued) attribute variables which pertain to those tracks. These approaches allow the search for the maximum likelihood (ML) or maximum a posteriori (MAP) association. However, while for kinematic variables there is a "gating" procedure based on a Gaussian distribution—which corresponds to a Neyman-Pearson test of "common origin" (actually, "same kinematic state") with selectable power—there is no simple counterpart of this for attributes. The sufficient statistic for the optimal association test (in the Neyman-Pearson sense) based on discrete-valued target classification information observables (attributes) is derived and its relationship with the class probability vector is discussed. Based on this, "attribute gates" are presented, which allow a Neyman-Pearson test for "same class" with the desired power.

Manuscript received December 8, 2006; revised August 1, 2007; released for publication January 30, 2008.

Refereeing of this contribution was handled by Benjamin Slocumb.

Authors' adresses: Y. Bar-Shalom, University of Connecticut, Dept. of Electrical & Computer Engineering, Storrs, CT 06269-2157, E-mail: (ybs@ee.uconn.edu); H. Chen, University of New Orleans, Dept. of Electrical Engineering, New Orleans, LA 70148, E-mail: (hchen2@uno.edu).

1557-6418/07/$17.00

1. INTRODUCTION

The problem of track-to-track association (T2TA)—a prerequisite for the fusion of tracks—has been considered initially in the literature for tracks described by kinematic states [1]. More recently, it has been generalized to include additional (continuous valued) feature and (discrete valued) attribute variables which pertain to those tracks.[1] These approaches allow the search for the maximum likelihood (ML) or maximum a posteriori (MAP) association.

It turns out that, under the Gaussian assumption on the estimation errors—which applies to the kinematic states and, possibly, the features—a simple sufficient statistic exists for the track association hypothesis testing and, consequently, it is easy to find the threshold for the desired Neyman-Pearson test of "common origin" (actually, "same kinematic state") with a selectable power. However, this does not apply for attributes, which are discrete valued. In this paper the sufficient statistic for the optimal association test in the Neyman-Pearson sense is derived for discrete-valued attribute/classification information and its relationship with the class probability vector is discussed.

Feature-aided T2TA was presented in [19, 20, 8]. A comprehensive procedure for incorporation of attributes and their possible dependence on the features was presented in [23, 14, 15] and shown to be amenable to obtain the MAP association of tracks from two sensors using linear programming. A multiple model approach for feature aided tracking (FAT) was presented in [22]. Classification-aided tracking with measurement-to-track association via multidimensional assignment (MDA) was discussed in [4]. However, while these approaches provide the ML or MAP association, they do not provide the means to set up a statistical hypothesis test with a desired power.

Target features and attributes/classification outputs are in general useful for track-to-track association especially when targets are closely spaced and the association based on the kinematic states only is unreliable. In some cases one deals with sensors that provide target attributes but the associated kinematic information is highly inaccurate. In such a case it is of interest to provide association decisions based on the attributes/classification information alone. This is the major motivation for the present work.

The rest of the paper is organized as follows. Section 2 discusses briefly the use of (continuous valued) kinematic and feature variables for track-to-track for association. The modeling and use of (discrete valued) attribute/classification information for track-to-track association is presented in Section 3. The modeling of the

[1] Examples of features are radar cross-section and target length. Examples of attributes are number of engines of an aircraft and type of emitter/waveform. Target classes can be, e.g., fighter vs. bomber or specific aircraft type. A detailed discussion of target features, attributes and classification can be found in [9, 10, 11, 12, 13].

classifier is discussed and the classification sufficient statistic is derived under the assumption of a constant "confusion matrix." The sufficient statistic calculation from the class probability vector is presented. To obtain the Neyman-Pearson test for "common class," the statistical characterization of the classification information sufficient statistic is developed in Section 4. Track-to-track association using all the information is discussed in Section 5. This section also presents a suboptimal test statistic for association based on classification information, as well as the (optimal in the Neyman-Pearson sense) likelihood ratio test and discusses the methodology for the performance evaluation. Numerical examples of the use of the likelihood function and likelihood ratio test are given in Section 6. Section 7 presents conclusions.

2. TRACK-TO-TRACK ASSOCIATION USING KINEMATIC AND FEATURE STATES

The problem of track-to-track association (T2TA) has been considered in the literature only for tracks described by kinematic states. If tracks also include continuous valued nonkinematic features which are estimated together with the kinematic states, they should be considered as part of the state in the process of track-to-track association. Once a decision on common origins of the tracks (actually, "same state") is made, the state estimates of those tracks deemed to correspond to the same target (have the same true state) can be fused to yield a more accurate target state estimate.

Assuming that the estimates of these features are obtained in a manner similar to the kinematic variables, their errors can be taken as zero mean Gaussian random variables. In this case, the association likelihood function based on the *augmented state*, consisting of the kinematic and feature components, can be expressed in the standard form [1]. This holds for continuous valued observations of both continuous as well as discrete valued features (the latter was discussed and illustrated in [13]). A special case is the target radar cross-section (RCS), which is positive and a Gaussian model is not appropriate. While a Swerling fluctuation model can be easily used for measurement to track association [1], for track-to-track association, there is no known sufficient statistic in this case for the hypothesis test, like the difference of state estimates when they have Gaussian errors. While a possible approach could be to use the difference of the RCS estimates (or their logarithms) with an (approximate) Gaussian assumption, the true RCS for different sensors is probably not the same due to the different aspect angles, which would make its use questionable. The situation of discrete valued observations, which is related to attributes/classifications, is discussed in Section 3.

For a pair of tracks, one can accept the "same state" hypothesis H_1: $\mathcal{C}^{ij}$ for track i from one sensor and track j from another sensor if the normalized distance between their (augmented) local state estimates $\hat{x}^i$ and $\hat{x}^j$ (at a common time, not indicated in the above notation for simplicity), which is chi-square distributed, is "not too large." Specifically, the squared norm

$$D^{ij} = (\hat{x}^i - \hat{x}^j)'[T^{ij}]^{-1}(\hat{x}^i - \hat{x}^j) \quad (1)$$

has to be within the $1-\alpha$ probability region of the chi-square distribution with n_x degrees of freedom for acceptance of the same state hypothesis, i.e.,

$$D^{ij} \le \chi^2(1-\alpha) \quad (2)$$

where the notation from [3] has been used. In (1)

$$T^{ij} = P^i + P^j - P^{ij} - P^{ji} \quad (3)$$

is the covariance of the difference between the local state estimation errors, which includes the local estimation error covariances P^i, P^j and the crosscovariance term P^{ij} due to the common process noise [1]. Note that for the feature part of the (augmented) state, which can be assumed in general time invariant without process noise, the crosscovariance of feature estimation errors between two local trackers is zero. For a practical way to obtain the crosscovariances for kinematic state components, see [7]. The acceptance region for H_1, defined by (2), is called *kinematic gate* [1].

According to the above, $\mathcal{C}^{ij}$ is rejected *if there is too much evidence against it*—the difference between the estimates is too large (relative to their accuracies, quantified by the covariance matrices) to accept that they are from the same true state. The hypothesis H_1: $\mathcal{C}^{ij}$ is called in the literature "common origin," but it is more accurate to call it "same true kinematic state," as (4) indicates.

Due to the correlation in time of the track state estimation errors [3], the test (1) is based on the track estimates at a single (common) point in time. Typically, since sensors (and the corresponding local estimators) are not synchronized, one of the state estimates will be a (short interval) prediction.

The test statistic (1), while commonly used in the literature [1, 6] without proof of its validity, was proven for the first time in [18]. This proof is briefly outlined below, because it will be the basis of a similar approach for the case of track association with classification information to be presented in the next section.

The likelihood function of the *same kinematic state* hypothesis $\mathcal{C}^{ij}$ is the pdf of the track sufficient statistics (the kinematic state estimates) conditioned on $\mathcal{C}^{ij}$, namely,

$$\Lambda_{\mathrm{kin}}(\mathcal{C}^{ij}) \triangleq p(\hat{x}^i, \hat{x}^j \mid \mathcal{C}^{ij}) = \int_{\mathcal{V}} p(\hat{x}^i, \hat{x}^j \mid x) p(x) dx \quad (4)$$

where x is the true common state of the two tracks and $\mathcal{V}$ is the region in which x takes values. The r.h.s. of (4) follows from the total probability theorem.

Assuming the joint pdf of the local state estimates to be

$$p(\hat{x}^i, \hat{x}^j \mid x) = \mathcal{N}\left(\begin{bmatrix} \hat{x}^i \\ \hat{x}^j \end{bmatrix}; \begin{bmatrix} x \\ x \end{bmatrix}, \begin{bmatrix} P^i & P^{ij} \\ P^{ji} & P^j \end{bmatrix}\right) \tag{5}$$

and using a *diffuse (noninformative) prior* [3] within a "sufficiently large" region $\mathcal{V}$, with volume V,

$$p(x) = V^{-1} \tag{6}$$

the likelihood function (4) becomes [18]

$$\Lambda_{\text{kin}}(\mathcal{C}^{ij}) = V^{-1}\mathcal{N}(\hat{x}^i - \hat{x}^j; 0, T^{ij}). \tag{7}$$

This shows the validity of using the test (2), which is based on the likelihood function (7) and states the following:

"Reject the same state (and thus the common origin) hypothesis if the normalized distance is in the α-tail of its distribution."

Note that the likelihood function (7) is a Gaussian pdf with the simple sufficient statistic (1). The Gaussian pdf makes it very easy to arrive at the test (2), which excludes its tail. This is the main reason to assume the estimation error of a continuous valued feature variable to be Gaussian.

REMARK The sufficient statistic (1) follows from the likelihood function of the "common origin" hypothesis. Assuming for the alternative hypothesis H_0: $\mathcal{C}^{i\neq j}$ ("different origin") a uniform diffuse distribution [3] i.e., a constant

$$\Lambda_{\text{kin}}(\mathcal{C}^{i\neq j}) = c \tag{8}$$

yields the likelihood ratio for the Neyman-Pearson test as (7) divided by the above constant c, which is irrelevant—the likelihood function and likelihood ratio tests are effectively the same (and have the same ROC curve). Consequently, (1), which can be seen to be the negative log-likelihood ratio, is optimal in the Neyman-Pearson sense and the power of the test (2) is $1 - \alpha$.

The difference of the estimates, as in (7), will not be an exact sufficient statistic for the situation of continuous valued observations on a discrete valued feature. Nevertheless, as shown in Section 5.2, one can use such an approximate sufficient statistic effectively.

The generalization of (7) to an arbitrary number of tracks can be found in [5]. This allows the association of tracks from an arbitrary number of sensors based on their local estimates, covariances and crosscovariances.

3. TRACK-TO-TRACK ASSOCIATION USING DISCRETE ATTRIBUTE/CLASSIFICATION INFORMATION

This section deals with the modelling and use of (discrete valued) attribute/classification information for track-to-track association. The modelling of the classifier is discussed and the classification sufficient statistic is derived under the assumption of a constant "confusion matrix."

Consider the case where a track contains observations of discrete valued attributes from which one can infer the target's class. The following model will be assumed for the target classes. It is assumed that there are N_c classes of targets. Let the possible classes be

$$\kappa \in K = \{1, \ldots, N_c\}. \tag{9}$$

The target class is assumed to be time invariant.

If two tracks belong to different classes, they obviously cannot have a common origin. The converse, however, is not true: if two tracks belong to the same class, they do not necessarily originate from the same target, unless the class is a unique identity. Thus, what can be accomplished with class information is testing whether two tracks belong to the *same class*, rather than originating from the same target. This is similar to the test based on kinematic state estimates where the test is, rigorously speaking, "same kinematic state" rather than "same origin."

3.1. Modelling of the Classifier Output

Let ζ denote the output of the classifier. The classifier's output is an attribute (an element of a discrete set), which is related to the presumed class to which the target under consideration belongs, as discussed below. The output set can have, in general, a larger number of elements than the set of target classes.[2] Then

$$\zeta \in K_a = \{1, \ldots, N_a\} \supseteq K \tag{10}$$

and it is assumed that one has

$$c_{nm} = P\{\zeta = m \mid \kappa = n\}, \qquad n = 1, \ldots, N_c, \quad m = 1, \ldots, N_a \tag{11}$$

which are the elements of the "confusion matrix" (see, e.g., [17])

$$C = [c_{nm}]. \tag{12}$$

Note that c_{nm} is the **likelihood** (probability of the observable conditioned on the truth of interest, see, e.g., [3]) of the true class being n when the classifier output (the observable) is $\zeta = m$. Thus the class likelihood function for classifier output m is the mth column of the confusion matrix C. One can conceivably have the likelihood functions depend on additional variables, like target kinematic state (e.g., aspect angle, distance to target), lighting, etc.

In the sequel it is assumed that all elements in the confusion matrix are constant and the same across dif-

[2] For example, one can have an "undetermined" class or a "class n_1 or n_2." Following [13], we use the term attribute for the observable from which a probabilistic inference can be made on the target class.

ferent classifiers. These restrictions can be removed by using a time argument and/or a classifier index, in which case, for classifier i at time k one would have the likelihoods $c^i_{nm}(k)$. However, in this case there is no sufficient statistic as in Section 3.3. Furthermore, it will be assumed that the classifier outputs are, conditioned on the truth, independent across time (it has "white" errors) and independent of the kinematic and feature variables. This is the counterpart of the white measurement noise for the kinematic measurements.[3]

3.2. Update of the Classification Probabilities

Denote by μ^0_n the prior probability of class n (prior to the observation under consideration). The posterior (or updated) probability of a target being in class n, given that the classifier's output is m, is

$$\mu_n = P\{\kappa = n \mid \zeta = m\} = \frac{c_{nm}\mu^0_n}{\sum_{l=1}^{N_c} c_{lm}\mu^0_l}. \qquad (13)$$

The corresponding class probability vector of the target under consideration can be written as

$$\mu = \frac{c_m \otimes \mu^0}{c'_m \mu^0} \qquad (14)$$

where c_m is the mth column of C, μ^0 is the prior probability vector and $\otimes$ is the Schur-Hadamard product (term by term) [16].

Similarly, for a track—*a sequence of associated measurements that includes classification information*—the updated class probability vector at time k, with classifier output m, is given by

$$\mu(k) \stackrel{\Delta}{=} \mathrm{col}[P\{\kappa = n \mid \zeta(k) = m, \zeta^{k-1}\}] = \frac{c_m \otimes \mu(k-1)}{c'_m \mu(k-1)} \qquad (15)$$

where ζ^{k-1} denotes the cumulative classification information at time $k-1$ and

$$\mu(0) = \mu^0 \qquad (16)$$

is the prior before getting any classification information.

It is worth mentioning that it is not via the probabilities (14) that the observations are used in the update (15), unless they are based on a uniform (i.e., noninformative) prior [3]. The update (15) requires the latest observation to enter via the *likelihood function* c_m. If one has only probabilities as in (14) with nonuniform priors, one can use the update/fusion procedure from Sec. 8.5.2 of [1], which avoids the "double counting" of the prior information.

[3]Just like the measurements are correlated because they observe (nearly) the same state, the classifier outputs will be correlated because the observe the same true variable, the class of the same target. However, the errors of the classifiers, like the errors of the sensor providing the kinematic measurements, are assumed to be white.

3.3. The Sufficient Statistic for Classification

Denote the output of the classifier at time t as $m(t)$, $t = 1,\ldots,k$. The recursion (15) can be rewritten as follows

$$\mu(k) = \frac{1}{\alpha} c_{m(k)} \otimes c_{m(k-1)} \otimes \cdots \otimes c_{m(1)} \otimes \mu^0 \qquad (17)$$

where α is the normalizing constant.

Note that, since the target class was assumed to be time invariant, the actual times at which the classifier outputs are generated are not relevant. Furthermore, because of the commutativity of the Schur-Hadamard product in (17), it can rewritten as follows

$$\mu(k) = \frac{1}{\alpha} c_1^{[\nu_1]} \otimes c_2^{[\nu_2]} \otimes \cdots \otimes c_{N_a}^{[\nu_{N_a}]} \otimes \mu^0 \qquad (18)$$

where ν_m is the number of times the output of the classifier was m and $c_m^{[\nu_m]}$ is c_m raised to the power ν_m with the Schur-Hadamard product.

Since

$$c_m^{[\nu_m]} = [(c_{1m})^{\nu_m}\ (c_{2m})^{\nu_m} \cdots (c_{N_c m})^{\nu_m}]', \qquad m = 1,\ldots,N_a \qquad (19)$$

the sufficient statistic for the classifier output can be seen to be the number of times each output class was generated, i.e., it is the vector

$$\nu = [\nu_1,\ldots,\nu_{N_a}]'. \qquad (20)$$

The existence of the above sufficient statistic hinges on the assumption that the confusion matrix is constant. Otherwise it appears that there is no such sufficient statistic.

3.4. Calculation of the Sufficient Statistic from the Classification Probabilities

In practice it is more likely that the information provided by the classifier will be the class probability vector (18) rather than the sufficient statistic (20). In this case we need to recover the vector ν from the vector μ. Note that both ν and μ consist of N_a elements but the elements of μ sum up to unity after normalization by α, which is not known. The elements of ν sum up to N, the number of times the classifier provided an output i.e.,

$$\sum_{m=1}^{N_a} \nu_m = N. \qquad (21)$$

It is assumed that N and μ^0 are known. This allows the substitution

$$\nu_1 = N - \sum_{m=2}^{N_a} \nu_m \qquad (22)$$

in (18), which can be written for component n as (the time index is omitted)

$$\mu_n = \frac{\mu_n^0}{\alpha}\prod_{m=1}^{N_a}(c_{nm})^{\nu_m} = \frac{\mu_n^0(c_{n1})^N}{\alpha}\prod_{m=2}^{N_a}\left(\frac{c_{nm}}{c_{n1}}\right)^{\nu_m}, \qquad n = 1,\ldots,N_c. \tag{23}$$

The above is a set of N_c equations in the unknowns consisting of ν_m, $m = 2,\ldots,N_a$, and α.

Taking the log of equations (23), one obtains a set of N_c linear equations in the N_a unknowns ν_m, $m = 2,\ldots,N_a$, and $\log\alpha$. If $N_a = N_c$, this set has a unique solution, allowing us to obtain ν_m, $m = 2,\ldots,N_a$ (note that α, while part of the solution, is of no interest); ν_1 follows from (22). This provides the complete solution in this case for the classification sufficient statistics ν_m, $m = 1,\ldots,N_a$. As it will be shown later, this vector ν (and not μ) will be needed for the test whether two tracks belong to the same class. If $N_a > N_c$, then in general one cannot find the classification sufficient statistics ν_m, $m = 1,\ldots,N_a$ uniquely by solving the above equations because there are not enough equations. In this case, if one has only the classification probabilities, then the likelihood function of the same class hypothesis cannot be fully specified. If $N_c > N_a$, then one can use a subset of N_a equations from (23) to obtain the likelihoods.

4. THE SAME CLASS LIKELIHOOD FUNCTION FROM CLASSIFICATION INFORMATION

This section presents the probability mass function (pmf) of the classifier's sufficient statistic, which is the basis of the Neyman-Pearson test.

The pmf, denoted as $P[\cdot]$, of the cumulative (local) classifier information using the sufficient statistic for the (local) track i is, for a total number of N^i classifier outputs, if the true class is n, given by the multinomial distribution [21]

$$P[\nu^i \mid \kappa^i = n] = P[\nu_1^i,\ldots,\nu_{N_a}^i \mid \kappa^i = n] = N^i!\prod_{m=1}^{N_a}\frac{c_{nm}^{\nu_m^i}}{\nu_m^i!} \tag{24}$$

where the total number of classifier outputs is

$$\sum_{m=1}^{N_a}\nu_m^i = N^i. \tag{25}$$

For simplicity, we assume that the local classifiers have the same confusion matrix C which is known at the fusion center. Furthermore, the outputs of the two classifiers are assumed, conditioned on the truth, independent.[4] Then the likelihood of the "same class" hypothesis H_1: $\mathcal{C}^{ij}$ can be written as follows

$$\begin{aligned}\Lambda_{\text{class}}(\mathcal{C}^{ij}) &\triangleq P[\nu^i,\nu^j \mid \mathcal{C}^{ij}] = \sum_{n=1}^{N_c}P[\nu^i,\nu^j \mid \kappa^i = \kappa^j = n]\mu_n^0\\ &= \sum_{n=1}^{N_c}P[\nu^i \mid \kappa^i = n]P[\nu^j \mid \kappa^j = n]\mu_n^0\\ &= \sum_{n=1}^{N_c}N^i!N^j!\left[\prod_{m=1}^{N_a}\frac{c_{nm}^{\nu_m^i+\nu_m^j}}{\nu_m^i!\nu_m^j!}\right]\mu_n^0\end{aligned} \tag{26}$$

where κ^i and κ^j are the true classes of tracks i and j, respectively. Note that while in (4) the total probability theorem was used with the diffuse (or improper) prior (6) to yield (7), in (26) the proper (and not necessarily uniform) prior μ^0 was used (since κ takes values in a finite set). In (26) it is assumed that the classification errors are independent across time and sensors.

Using the same approach as in the case of the continuous valued states, we propose, based on the class information, to reject the common origin hypothesis *if there is too much evidence against it*: if the likelihood (26) is in the tail of its distribution.

We are faced with two problems here:

1) There is no simple sufficient statistic similar to the normalized distance in the continuous/Gaussian case. The difference of the two vectors ν^i and ν^j is not the exact sufficient statistic for the hypothesis test. Nevertheless, a suboptimal method based on this is explored in the next section and evaluated later.

2) While there is an expression for the likelihood function pmf, to find its "tail," an exhaustive evaluation of all its point masses is needed: these point masses have to be ordered and the "tail" identified.

An alternative approach would be to use a Monte Carlo method to determine whether a particular pair ν^i, ν^j is in the tail of the distribution. To evaluate a 5% tail probability with $2\sigma = 1\%$, one needs 2000 runs (of $N^i + N^j$ classifications), which is not too excessive.

In the examples presented later, the exhaustive evaluation of the likelihood function pmf is carried out, since it is not too expensive computationally. The number of points at which (26) has to be evaluated is obtained below.

The number of points (different outcomes) for the pmf (24) is

$$M(N_a,N^i) = (N_a)^{N^i} - \sum_{\substack{k_1,\ldots,k_{N_a}\in\{0,1,\ldots,N^i\}\\ k_1+\cdots+k_{N_a}=N^i}}\left(\frac{(N^i)!}{k_1!\cdots k_{N_a}!}-1\right). \tag{27}$$

The above result follows from Eq. (1.18) in Ch. 2 of [21]. Namely, it is the number of elements in the expansion of a multinomial of N_a elements raised to

[4] This is the counterpart of assuming independent measurement noises at different sensors.

the power N^i: it is $(N_a)^{N^i}$ less the number of outcomes which are equivalent because the order does not matter. The latter is given, for a certain outcome, by the corresponding multinomial coefficient less one because for each outcome we remove only the duplicates.

The number of points at which (26) has to be evaluated is then

$$M(N_a, N^i, N^j) = M(N_a, N^i)M(N_a, N^j). \tag{28}$$

5. TRACK-TO-TRACK ASSOCIATION TESTING USING ALL THE INFORMATION

This section proposes a procedure to test for common origin by decoupling the continuous variables from the discrete ones. Then the details of a simpler suboptimal test as well as the optimal (likelihood ratio) test using the discrete attribute variables are presented.

5.1. Test using the Likelihood Function of H_1

It seems reasonable to accept that tracks i and j are from the same target if

1) Their continuous valued state (kinematic and feature) estimates are "close enough" to accept that their kinenatic/feature stats are the same—they satisfy (2), and

2) Their classification does not present strong evidence that they cannot be the same, i.e., $\Lambda_{\text{class}}(\mathcal{C}^{ij})$, given in (26), falls into the $1-\alpha$ probability concentration region under H_1. In this case the "same class" hypothesis H_1 is accepted.

This sequence of tests assumes implicitly that the classification errors are independent of the kinematic and feature variables. While one can write a joint likelihood if there is a dependence [23], a joint test does not seem to be available at this point due to the mixed nature (continuous-discrete) of the likelihood function and the ensuing lack of a joint sufficient statistic.

5.2. Test using the Difference of Local Classification Sufficient Statistics

To reduce the complexity of the likelihood function given in (26) and thus simplify the computation of the probability region (the "attribute gate"), we also consider using the difference of the local classification sufficient statistics in the hypothesis test. The results are particularly useful for the "same class" test involving only two targets.

As a preliminary, consider the difference

$$z = x - y \tag{29}$$

of two discrete-valued random variables

$$x \in \{0,\ldots,n_x\} \qquad y \in \{0,\ldots,n_y\} \tag{30}$$

which yields

$$z \in \{-n_y,\ldots,n_x\}. \tag{31}$$

Then

$$P\{z = n\} = \sum_{k=n}^{n_x} P\{x = k\}P\{y = k - n\}. \tag{32}$$

Let the difference of the local classification sufficient statistics given in (20), now superscripted by the sensor index, be

$$\delta^{ij} = [\delta_1^{ij},\ldots,\delta_{N_a}^{ij}] = [\nu_1^i - \nu_1^j,\ldots,\nu_{N_a}^i - \nu_{N_a}^j] \tag{33}$$

which yields

$$\delta_k^{ij} \in \{-N^j,\ldots,N^i\}, \qquad k = 1,\ldots,N_a. \tag{34}$$

The pmf[5] of the above vector is, based on (32), given by

$$\begin{aligned} P[\delta^{ij} \mid \kappa = n] = \sum_{k_1=\delta_1^{ij}}^{N^i} \cdots \sum_{k_{N_a}=\delta_{N_a}^{ij}}^{N^i} & P\{\nu_1^i = k_1,\ldots,\nu_{N_a}^i = k_{N_a} \mid \kappa = n\} \\ & \times P\{\nu_1^j = k_1 - \delta_1^{ij},\ldots,\nu_{N_a}^i = k_{N_a} - \delta_{N_a}^{ij} \mid \kappa = n\} \end{aligned} \tag{35}$$

where the probabilities in the N_a-fold summation above are given in (24).

5.3. Test using the Likelihood Ratio of H_1 vs. H_0

There is an alternative approach to use the classification information: rely on the likelihood ratio of the two hypotheses, rather than only the likelihood function of H_1. This is done as follows.

The hypothesis H_0: $\mathcal{C}^{i\neq j}$ that the two targets are different, i.e., *belong to different classes*, is composite and can be written with the total probability theorem as follows

$$\begin{aligned} \Lambda_{\text{class}}(\mathcal{C}^{i\neq j}) &\triangleq P[\nu^i,\nu^j \mid \mathcal{C}^{i\neq j}] \\ &= \sum_{n=1}^{N_c} \sum_{l=1,l\neq n}^{N_c} P[\nu^i,\nu^j \mid \kappa^i = n \neq \kappa^j = l]\mu_n^0\mu_l^0 \\ &= \sum_{n=1}^{N_c} \sum_{l=1,l\neq n}^{N_c} N^i!N^j! \left[\prod_{m=1}^{N_a} \frac{c_{nm}^{\nu_m^i} c_{lm}^{\nu_m^j}}{\nu_m^i!\nu_m^j!}\right] \mu_n^0\mu_l^0. \end{aligned} \tag{36}$$

The test statistic (based on the classification information only) is then the ratio of (26) to (36) i.e.,

$$\lambda_{\text{class}}(\mathcal{C}^{ij} : \mathcal{C}^{i\neq j}) = \frac{\Lambda_{\text{class}}(\mathcal{C}^{ij})}{\Lambda_{\text{class}}(\mathcal{C}^{i\neq j})} \tag{37}$$

[5] Denoted as $P[\cdot]$, while $P\{\cdot\}$ denotes the probability of an event.

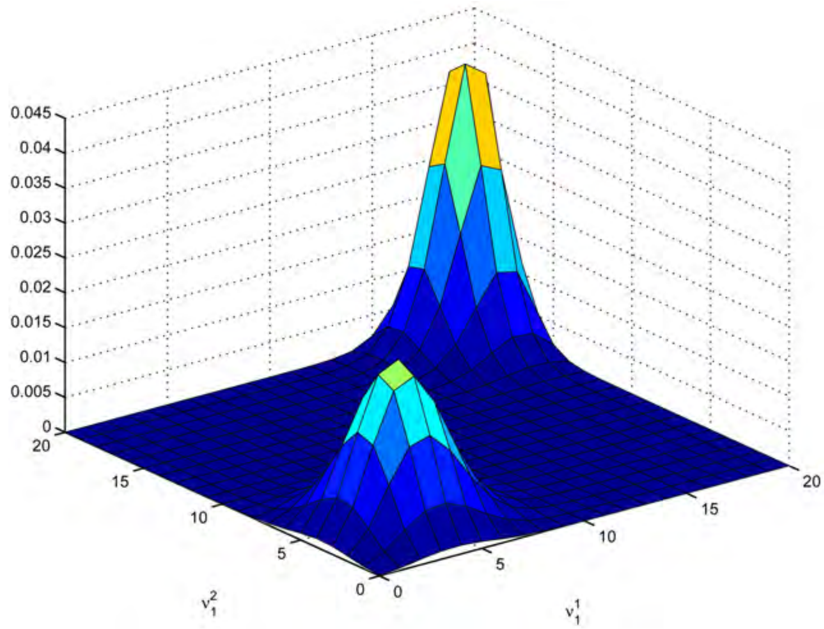


Fig. 1. The probability mass function of the local classifier sufficient statistics (ν_1^1, ν_1^2) under the same class hypothesis (likelihood function of H_1).

and the acceptance region for H_1 is obtained by finding the $1-\alpha$ probability concentration region for this ratio under H_1.

For the kinematic and feature variables, typically, the alternate hypothesis is usually modeled by a diffuse pdf [3], i.e., it amounts to a constant that rescales the likelihood function into the likelihood ratio. Thus they are effectively the same test. Consequently, we shall focus on the tests based on classification variables.

5.4. Performance Evaluation of the Tests

To find the probability of false alarm (acceptance of H_1 when H_0 is true) for the likelihood function based test, one has to evaluate numerically the probability mass of (26) under H_0 in the acceptance region for H_1.

For the likelihood ratio based test, one has to evaluate the probability mass of the likelihood ratio test statistic (37) under H_0 in the acceptance region for H_1. These are illustrated in the next section.

6. NUMERICAL EXAMPLES

Assume that each target belongs to one of $N_c = 2$ classes and each classifier output provides target attribute (taken here as its class, i.e., $N_a = 2$) with the accuracy given by the following time invariant confusion matrix

$$C = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}. \tag{38}$$

Based on the sufficient statistics from two local classifiers, we want to test whether the two targets belong to the same class (hypothesis H_1) vs. different classes (hypothesis H_0). Note that the common origin hypothesis will be rejected if according to the hypothesis test it is very implausible. Assume that the target has equal prior probability of belonging to each of the two classes and the total number of outputs for each local classifier is $N^i = 20$, the pmf of the sufficient statistic under the "same class" hypothesis is shown in Figure 1 for different ν_1^i, $i = 1,2$.[6]

We can see the two peaks at $(18,18)$ and $(4,4)$ corresponding to the two possible classes of the truth. In general the pmf will have N_a (number of attributes) peaks due to the uncertainty about the truth. The decision region based on the *likelihood function* (26) to allow 5% missed detection of H_1 is shown in Figure 2. This follows from the $\alpha = 5\%$ tail probability mass of the pmf plotted in Figure 1. Unlike in the Gaussian case (for continuous valued states), the "same class" test (which is for discrete states), in general, yields an irregular decision region. This is the *attribute gate*, defined in the ν_1^i, $i = 1,2$ space. In this example we have only two possible combinations for the evaluation of (36) when the two targets belong to two different classes, so one can evaluate the likelihood function of H_0 relatively easily.

The pmf of $\nu_1^2 - \nu_1^1$ is shown (as a bar plot) in the top part of Figure 3. It is symmetric around zero, as expected. The (moment matched) normal probability plot (cdf—cumulative distribution function) in the bottom part of Figure 3 indicates that the difference of the local classification statistics (whose cdf is the shown staircase function) can be well approximated by a Gaussian distribution. By numerical calculation we found that the decision region (attribute gate for the above difference)

[6] In view of (25) and the fact that there are only two classes, ν_1^i is a sufficient statistic for classifier i.

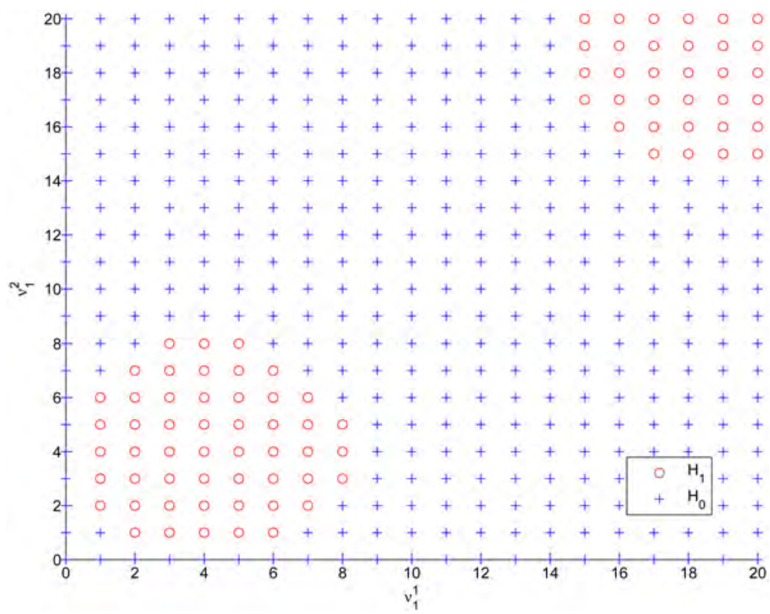


Fig. 2. Decision regions (attribute gate in the (ν_1^1, ν_1^2) space) based on the likelihood function of H_1 with 5% probability of incorrectly rejecting it.

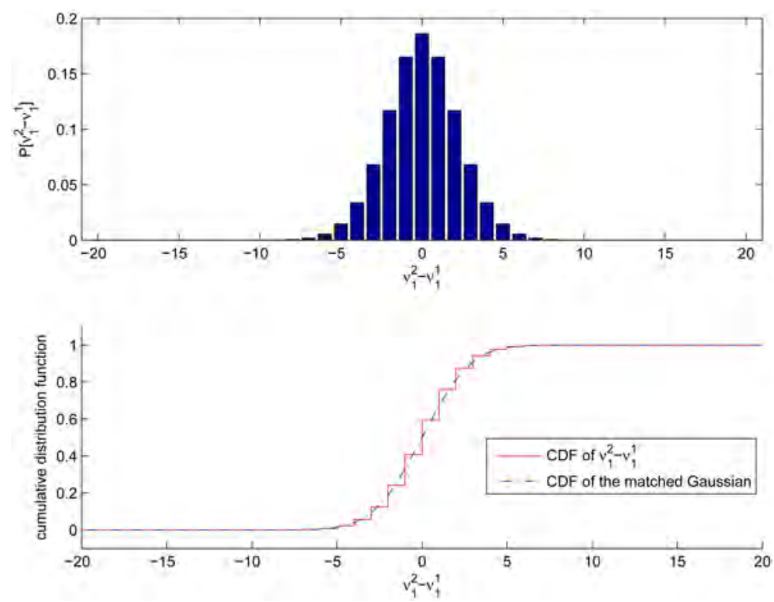


Fig. 3. The probability mass function of the approximate local classifier sufficient statistic $(\nu_1^2 - \nu_1^1)$ and the corresponding moment-matched normal probability distribution under the "same class" hypothesis (likelihood function of H_1). The 95% attribute gate for this difference is $[-4,4]$.

to allow 5% missed detection of H_1 is $[-4,4]$. This is much simpler than the class gate given in Figure 2.

The log-likelihood ratio surface of the same class (H_1) vs. the different classes hypothesis (H_0)—the ratio (28)—is shown in Figure 4. The prior for the classes is taken here as uniform. The decision region based on the *likelihood ratio* (28) to allow $\alpha = 5\%$ missed detection of H_1 is shown in Figure 5.

Comparison of the various tests

By numerically evaluating the pmf of the likelihood ratio statistic (37) under H_0 in the acceptance region for H_1, we find the false alarm probability (acceptance of H_1 when H_0 is true) as 8.1×10^{-8}. Compared with the decision region shown in Figure 2, which yields a false alarm probability of 1.9×10^{-7}, the likelihood ratio test has only a marginal performance gain in this case. Interestingly, the decision region based on the difference of the local classification statistics yields a false alarm probability of 9.1×10^{-5}, which is higher than both of the above, but still quite small.

7. SUMMARY AND CONCLUSIONS

The likelihood function and likelihood ratio based tests for track-to-track association using classification

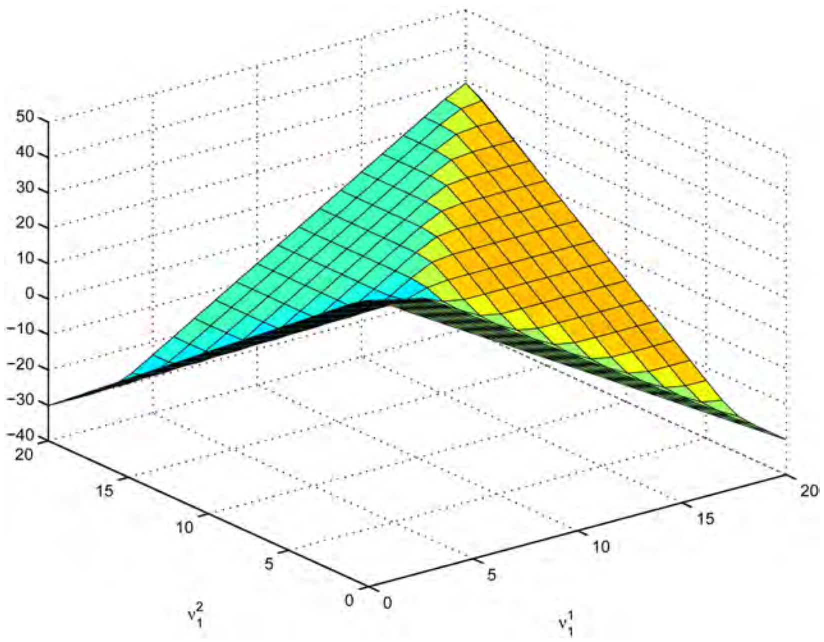


Fig. 4. The surface of the log-likelihood ratio between the same class (H_1) and different classes (H_0) vs. the local classifier sufficient statistics (ν_1^1, ν_1^2). Note the two peaks, one for (ν_1^1, ν_1^2) both large, one for both small.

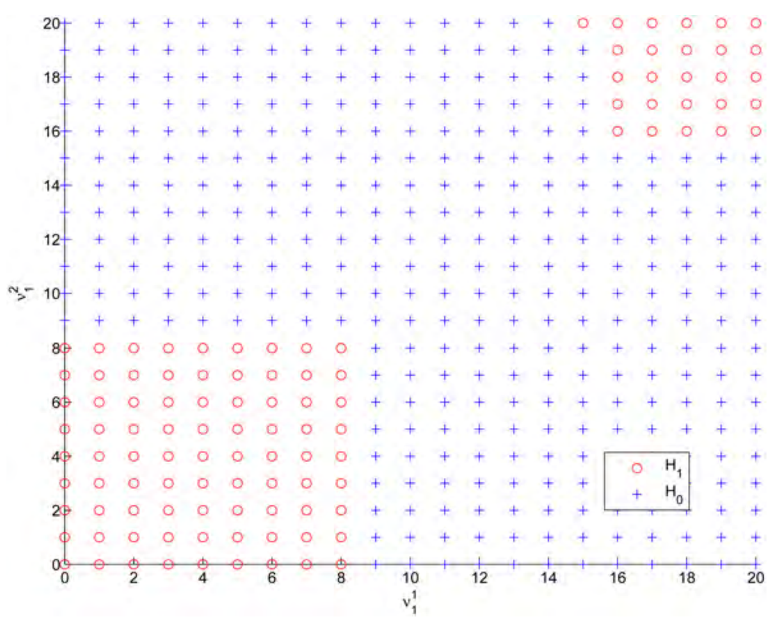


Fig. 5. Decision region for the likelihood ratio test with 5% probability of incorrectly rejecting H_1. This is the 95% attribute gate in the (ν_1^1, ν_1^2) space.

information have been derived and the means for evaluating their performance have been presented and illustrated. The sufficient statistic for the optimal association test in the Neyman-Pearson sense was obtained and its relationship with the class probability vector was discussed. The likelihood ratio test does not appear to have significant advantage over the test based on the likelihood function. The simplest test, based on the difference of the local sufficient statistics—similar to the test for continuous valued state estimates—yields a false alarm probability which is higher than for both of the above tests, but still quite small. The generalization to different and possibly time varying confusion matrices is a topic for future investigation.

Acknowledgments

Stimulating discussions with Drs. John Edwards, Oliver E. Drummond and J. Kent Haspert from the SIAP TCAIET task force are gratefully acknowledged. Thanks to an anonymous reviewer who pointed out that Eq. (27) was incorrectly written in the original version.

REFERENCES

[1] Y. Bar-Shalom and X. R. Li
Multitarget-Multisensor Tracking: Principles and Techniques.
YBS Publishing, 1995.

[2] Y. Bar-Shalom and W. D. Blair (Eds.)
Multitarget-Multisensor Tracking: Applications and Advances, vol. III.
Artech House, 2000.

[3] Y. Bar-Shalom, X. R. Li and T. Kirubarajan
Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction.
Wiley, 2001.

[4] Y. Bar-Shalom, T. Kirubarajan and C. Gokberk
Tracking with classification-aided multiframe data association.
IEEE Transactions on Aerospace and Electronic Systems, **41**, 3 (July 2005), 868–878.

[5] Y. Bar-Shalom and H. Chen
Multisensor track-to-track association for tracks with dependent errors.
Journal of Advances in Information Fusion, **1**, 1 (July 2006), 3–15. Also in *Proceedings of 43rd IEEE Conference on Decision and Control*, Nassau, Bahamas, Dec. 2004.

[6] S. Blackman and R. Popoli
Design and Analysis of Modern Tracking Systems.
Artech House, 1999.

[7] H. Chen, T. Kirubarajan and Y. Bar-Shalom
Performance limits of track-to-track fusion vs. centralized estimation.
IEEE Transactions on Aerospace and Electronic Systems, **39**, 2 (Apr. 2003), 386–400.

[8] C.-Y. Chong and S. Mori
Metrics for feature-aided track association.
In *Proceedings of 9th International Conference on Information Fusion*, Florence Italy, July 2006.

[9] O. E. Drummond
On features and attributes in multisensor multitarget tracking.
In *Proceedings of 2nd International Conference on Information Fusion*, Sunnyvale, CA, July 1999.

[10] O. E. Drummond
Integration of features and attributes into target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 4048, Orlando, FL, Apr. 2000.

[11] O. E. Drummond
Feature, attribute and classification aided target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 4473, San Diego, CA, Aug. 2001.

[12] O. E. Drummond
On categorical feature aided target tracking.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5204, San Diego, CA, Aug. 2003.

[13] O. E. Drummond
Tracking and classification with attribute data from legacy sensors.
In *Proceedings of ONR/GTRI Workshop on Tracking*, Key West, FL, June 2004.

[14] J. Ferry and L. D. Stone
Track-to-track association with missing features.
In *Proceedings of MSS Sensor and Data Fusion*, Monterey, CA, May 2005.

[15] J. Ferry
XMAP: Track to track association with metric, feature, and target type data.
In *Proceedings of 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[16] P. R. Halmos
Finite-Dimensional Vector Spaces.
Van Nostrand, 1958.

[17] J. K. Haspert, J. F. Heagy and R. J. Sullivan
General approach to template based target recognition.
IEEE Systems Magazine, **22**, 9, Part 2, (Sept. 2007), 53–67.

[18] X. D. Lin, Y. Bar-Shalom and T. Kirubarajan
Multisensor bias estimation using local tracks without a priori association.
In *Proceedings of SPIE Conference on Signal and Data Processing of Small Targets*, vol. 5204, paper #70, San Diego, CA, Aug. 2003.

[19] S. Mori and C.-Y. Chong
Effects of unpaired objects and sensor biases on track-to-track association: Problems and solutions.
In *Proceedings of MSS National Symposium on Sensor and Data Fusion*, vol. 1, 2000, 137–151.

[20] S. Mori and C.-Y. Chong
Evaluation of data association hypotheses: Non-Poisson IID cases.
In *Proceedings of 7th International Conference on Information Fusion*, Stockholm, Sweden, July 2004.

[21] E. Parzen
Modern Probability and Its Applications.
Wiley, 1960.

[22] B. J. Slocumb and M. E. Klusman
A multiple model SNR/RCS likelihood ratio score for radar-based feature-aided tracking.
In *Proceedings of SPIE, Conference on Signal and Data Processing of Small Targets*, vol. 5913, San Diego, CA, Aug. 2005.

[23] L. D. Stone, M. L. Williams and T. M. Tran
Track to track association and bias removal.
In *Proceedings of SPIE Conference Signal and Data Processing of Small Targets*, vol. 4728, Orlando, FL, Apr. 2002.

Yaakov Bar-Shalom was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1970, all in electrical engineering.

From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, CA. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and Marianne E. Klewin Professor in Engineering. He is also director of the ESP Lab (Estimation and Signal Processing) at the University of Connecticut. His research interests are in estimation theory and stochastic adaptive control and he has published over 360 papers and book chapters in these areas. In view of the causality principle between the given name of a person (in this case, "(he) will track," in the modern version of the original language of the Bible) and the profession of this person, his interests have focused on tracking.

He coauthored the monograph *Tracking and Data Association* (Academic Press, 1988), the graduate text *Estimation with Applications to Tracking and Navigation* (Wiley, 2001), the text *Multitarget-Multisensor Tracking: Principles and Techniques* (YBS Publishing, 1995), and edited the books *Multitarget-Multisensor Tracking: Applications and Advances* (Artech House, Vol. I 1990; Vol. II 1992, Vol. III 2000). He has been elected Fellow of IEEE for "contributions to the theory of stochastic systems and of multitarget tracking." He has been consulting to numerous companies, and originated the series of Multitarget Tracking and Multisensor Data Fusion short courses offered at Government Laboratories, private companies, and overseas.

During 1976 and 1977 he served as associate editor of the *IEEE Transactions on Automatic Control* and from 1978 to 1981 as associate editor of *Automatica*. He was program chairman of the 1982 American Control Conference, general chairman of the 1985 ACC, and cochairman of the 1989 IEEE International Conference on Control and Applications. During 1983–1987 he served as chairman of the Conference Activities Board of the IEEE Control Systems Society and during 1987–1989 was a member of the Board of Governors of the IEEE CSS. Currently he is a member of the Board of Directors of the International Society of Information Fusion and served as its Y2K and Y2K2 President. In 1987 he received the IEEE CSS distinguished Member Award. Since 1995 he is a distinguished lecturer of the IEEE AESS. He is corecipient of the M. Barry Carlton Awards for the best paper in the *IEEE Transactions on Aerospace and Electronic Systems* in 1995 and 2000, and received the 1998 University of Connecticut AAUP Excellence Award for Research, the 2002 J. Mignona Data Fusion Award from the DoD JDL Data Fusion Group, and the 2008 IEEE D. J. Picard Medal for Radar Technologies and Applications.

Huimin Chen received the B.E. and M.E. degrees from Department of Automation, Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, in 2002, all in electrical engineering.

He was a post doctorate research associate at the Physics and Astronomy Department, University of California, Los Angeles, and a visiting researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University in 2002 where his research focus was on weak signal detection for single electron spin microscopy. He joined the Department of Electrical Engineering, University of New Orleans in January 2003 as an assistant professor. His research interests are in general areas of signal processing, estimation theory, and information theory with applications to target detection and target tracking.

Information for Authors

The Journal of Advances in Information Fusion (JAIF) is a semi-annual, peer-reviewed archival journal that publishes papers concerned with the various aspects of information fusion. The boundaries of acceptable subject matter have been intentionally left flexible so that the journal can follow the research activities to better meet the needs of the research community and members of the International Society of Information Fusion (ISIF).

Restrictions on Publication: The International Society of Information Fusion (ISIF) publishes in JAIF only original material that has not been either published or submitted for publication in an archival journal. However, prior publication of an abbreviated and/or preliminary form of the material in conference proceedings or digests shall not preclude publication in JAIF when notice is made at the time of submission and the previous publications are properly referenced in the manuscript. Notice of prior publication of an abbreviated version of the manuscript must be given at the time of submission to the transactions. Submission of the manuscript elsewhere is precluded unless it is refused for publication or withdrawn by the author. Concurrent submission to other publications and this journal is viewed as a serious breach of ethics and, if detected, the manuscript under consideration will be rejection.

Sanctions for Plagiarism: If a manuscript is substantially similar to any previous journal submission by any of the authors, its history and some explanation must be provided. If a manuscript was previously reviewed and rejected by a different journal, the authors shall provide to the area editor copies of all correspondence involving the earlier submission. In addition, the authors must discuss the reasons for resubmission and be prepared to deliver further material if more is requested. Submission of a previously rejected manuscript is acceptable as long as the motive for submission to JAIF is explained and the authors have addressed the feedback of the earlier reviewers. The authors of a manuscript that is found to have been plagiarized from others or contain a crossover of more than 25% with another journal manuscript by any of the authors will incur sanctions by the ISIF. The following sanction will be applied for any of the above-noted infractions: (1) immediate rejection of the manuscript in question; (2) immediate withdrawal of all other submitted manuscripts by any of the authors to any of the ISIF publications (journals, conferences, workshops); (3) prohibition against all of the authors for any new submissions to ISIF publications, either individually, in combination with the authors of the plagiarizing manuscript as well as in combination with new coauthors. The prohibition shall continue for at least two years from notice of suspension.

Types of Contributions: Contributions may be in the form of regular papers or correspondence. The distinction between regular papers and correspondence is not one of quality, but of nature. Regular papers are to be a well-rounded treatment of a problem area, while a correspondence item makes one or two points concisely. In regular papers, the title, abstract, and introduction should be sufficiently informative to illuminate the essence of the manuscript to the broadest possible audience and to place the contributions in context with related work. The body of the manuscript should be understandable without undue effort by its intended audience. Correspondence should be less discursive but equally lucid. Authors should be aware that a well-written correspondence are usually published more rapidly than a regular paper.

Submission of Manuscript for Review: Manuscripts are submitted electronically via the internet at http://jaif.msubmit.net. For peer review, manuscripts should be formatted in a single column and double spaced.

General Information for JAIF Authors and Submission of Final Manuscript: General information for JAIF authors is available at http://www.isif.org/jaif.htm, and specific document preparation information for final manuscript for publication is given below and can be found at http://www.isif.org/JAIF Manuscript Preparation1.doc.

Copyright: It is the policy of the ISIF to own the copyright to the technical contributions it publishes; authors are required to sign an ISIF copyright transfer form before publication. This form appears in JAIF from time to time and is available on the web site listed above. If a copyright form is not submitted with the manuscript, it will be requested upon contribution acceptance. Publication will not take place without a completed copyright form.

Page Charges: Since ISIF has elected to support the publication of JAIF with the revenues of conferences and membership dues, page charges are not mandated for publication in JAIF. Page charges for journals of professional societies are traditional, and at some time in the future, page charges may be introduced to cover the expenses associated with the publication of JAIF. Payment of page charges will not be a prerequisite for publication.

Discloseability: The ISIF must, of necessity, assume that material submitted for publication is properly available for general dissemination to the audiences for which JAIF serves. It is the responsibility of the author, not ISIF, to determine whether disclosure of materials requires prior consent of other parties, and, if so, obtain it.

Preparation of Manuscript for Publication: After a manuscript is recommended for publication by the editorial staff, the corresponding author prepares the manuscript according to the guidelines below and submits the final version of the manuscript as a pdf file and the supporting manuscript files. Before preparing their manuscript for publication, authors should check for the more update version of the guidelines at http://jaif.msubmit.net.

All manuscripts must be submitted electronically in pdf format with the supporting files in one of the following formats: Plain TeX, AMS-TeX, LaTeX, or MS-Word with the MathType extension for equations and in-text mathematical material. The pdf file must be created with graphics resolution set to 300 dpi or greater. The author must approve the pdf file because it will be used as the basis for rectifying any inconsistencies in the supporting files. The supporting files must be prepared according to the following guidelines.

- Text files should include title, authors, affiliations, addresses, a footnote giving the dates for submission and revision and name of the editor that handed the review, and an optional footnote acknowledging a sponsor for the research. The information must be included in the text files in this order.
- Abstract must be included. The abstract must include the key words for the manuscript and include no more than 300 words for a regular paper and 150 words for a correspondence.
- Authors should number main section headings as 1, 2, 3, etc., and subsections as 1.1, 1.2, or 2.1. All headings should be typed with title format (i.e., first letter caps and lower case)—not in ALL CAPS. The typesetter will convert to all caps in final formatting as required.
- Authors should format references very carefully according to the examples given below. Style for authors' names are initials followed by last name and it must be followed precisely. References must be listed alphabetically by (i) last name of first author, (ii) initials of first author, (iii) last name of second author, (iv) initials of second author, etc. For manuscripts with common authors and order of authors, the date of publication should be used to select order with the earlier publications being list first. The names of publications must be spelled completely. In the references, the author should use only approved abbreviations that include: No., Vol., p., pp., AIAA, IEEE, IEE, and SPIE.
- Authors who use one of the TeX variants must provide a list of their macros and the files for the macros. Authors who use LaTeX must include the bbl and aux files.
- All figures must be submitted electronically as HIGH resolution (300 dpi or better) in Color graphics or GRAYSCALE (where shading is involved) or BLACK AND WHITE (if simple line art) graphics files (tif, eps, jpg, bmp). Each figure must be supplied as a separate graphics file. Graphics (or captions) should NOT be embedded in the text files. The figures must be included in the pdf file of the full article and the pdf file must be created with graphics resolution set to 300 dpi or greater.
- A separate file including all figure captions must be included.
- Each table must be submitted in a separate file. Tables (or captions) should NOT be included in text files and should be in form of DATA—rather than graphics—files.
- A separate file including all table captions must be included.
- A separate text of the biography of each author must be submitted. The text file should be less than 500 words.
- Separate graphics files of each author's photo should be provided as a grayscale graphics file or a color graphics file.

Examples of the references are alphabetized correctly and listed below.

BOOK:

[1] R. E. Blahut
Theory and Practice of Error Control Codes. Reading, MA: Addison-Wesley, 1983.

PROCEEDINGS ARTICLE:

[2] T. Fichna, M. Gartner, F. Gliem, and F. Rombeck
Fault-tolerance of spacebome semiconductor mass memories.
In Twenty-Eighth Annual International Symposium on Fault-Tolerant Computing, Digest of Papers, 1998, 408–413.

BOOK:

[3] P. K. Lala
Fault Tolerant and Fault Testable Hardware Design. Englewood Cliffs, NJ: Prentice-Hall, 1985.

WEB SITE:

[4] National Semiconductors Inc.
Homepage: http://www.national.com.

PROCEEDINGS ARTICLE:

[5] C. Paar and M. Rosner
Comparison of arithmetic architectures for reed-solomon decoders in reconfigurable hardware.
In *Proceedings of the Symposium on Field-Programmable Custom Computing Machines*, Apr. 1997, 219–225.

JOURNAL ARTICLE:

[6] N. R. Saxena and E. J. McCluskey
Parallel signature analysis design with bounds on aliasing.
IEEE Transactions on Computers, **46**, 4 (Apr. 1997), 425–438.

[7] C. I. Underwood and M. K. Oldfield
Observations on the reliability of cots-device-based solid state data recorders operating in low-Earth orbit.
IEEE Transactions on Nuclear Science, **47**, 4 (June 2000), 647–653.

WEB SITE:

[8] Xilinx Inc.
Homepage: http://www.xilinx.com.

INTERNATIONAL SOCIETY OF INFORMATION FUSION

ISIF Website: http://www.isif.org

SOCIETY VISION

The International Society of Information Fusion (ISIF) seeks to be the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied information fusion technologies.

SOCIETY MISSION

Advocate

To advance the profession of fusion technologies, propose approaches for solving real-world problems, recognize emerging technologies, and foster the transfer of information.

Serve

To serve its members and engineering, business, and scientific communities by providing high-quality information, educational products, and services.

Communicate

To create international communication forums and hold international conferences in countries that provide for interaction of members of fusion communities with each other, with those in other disciplines, and with those in industry and academia.

Educate

To promote undergraduate and graduate education related to information fusion technologies at universities around the world. Sponsor educational courses and tutorials at conferences.

Integrate

Integrate ideas from various approaches for information fusion, and look for common threads and themes—look for overall principles, rather than a multitude of point solutions. Serve as the central focus for coordinating the activities of world-wide information fusion related societies or organizations. Serve as a professional liaison to industry, academia, and government.

Disseminate

To propagate the ideas for integrated approaches to information fusion so that others can build on them in both industry and academia.

Call for Papers

The Journal of Advances in Information Fusion (JAIF) seeks original contributions in the technical areas of research related to information fusion. Authors of papers in one of the technical areas listed on the inside cover of JAIF are encouraged to submit their papers for peer review at http://jaif.msubmit.net.

Call for Reviewers

The success of JAIF and its value to the research community is strongly dependant on the quality of its peer review process. Researchers in the technical areas related to information fusion are encouraged to register as a reviewer for JAIF at http://jaif.msubmit.net. Potential reviewers should notify via email the appropriate editors of their offer to serve as a reviewer.

Additional Copies

If you need additional copies, please contact Tammy Blair at 404-407-7726 or tammy.blair@gtri.gatech.edu for pricing information.

Journal of Advances in Information Fusion Volume 2 Number 1 June 2007 ISSN 1557-6418