December 2007 Volume 2 Number 2 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

Performance Analysis of Decentralized Kalman Filters under Communication Constraints 65
Markus S. Schlosser and Kristian Kroschel, Universitat Karlsruhe, Germany

Performance Evaluation for Automated Threat Detection 77
Robert C. Schrag, Global InfoTek, Inc., USA
Masami Takikawa, Art Technology Group, Inc., USA
Paul Goger, Metron, Inc., USA
James Eilbert, AP Technology, LLC, USA

Target Engageability Improvement through Adaptive Tracking 99
Abder Rezak Benaskeur, Francois Rheaume, Stephane Paradis, Defence R&D Canada-Valcartier, Canada

Misassociation Probability in M2TA and T2TA 113
Javier Areta, Yaakov Bar-Shalom, University of Connecticut, USA
Ron Rothrock, Sparta,Inc., USA

Information for Authors 128

From the Aministrative Editor

Review and Production Process of JAIF Manuscripts...

ISIF

www.isif.org

INTERNATIONAL SOCIETY OF INFORMATION FUSION

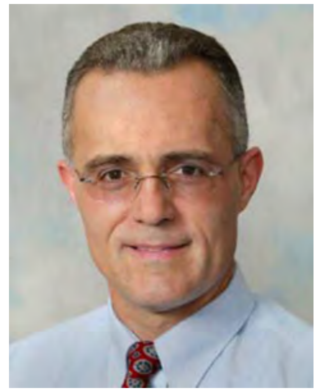
The International Society of Information Fusion (ISIF) seeks to be the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied INFORMATION FUSION technologies. Technical areas of interest include target tracking, detection theory, applications 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 Administrative Editor

December 2007



Review and Production Process of JAIF Manuscripts

The Journal of Advances in Information Fusion (JAIF) has been in publication since the first issue in July 2006. A new issue of the journal is published every six months and posted on the web site of the International Society of Information Fusion (ISIF) at http://www.isif.org. Each issue typically features four to six original articles. Bringing each article to you involves many steps that begin with a new manuscript submission and ends with posting of the final typeset manuscript on the ISIF web site. We are often asked by authors to estimate how long the review and production process takes in JAIF. However, no universal timeline exists for papers to go from manuscript submission to publication, and this is typically true for most journals. Some journals achieve a shorter timeline by reducing the peer review process.

When a new manuscript is submitted to JAIF, it is automatically assigned to the Area Editor for the technical area selected by the corresponding author. The Area Editor assigns an Associate Editor under their area to handle the actual review process. Area Editors can also serve in the role of an Associate Editor. The Associate Editor uses the web-based system to assign three to four reviewers who have the appropriate technical background for evaluating the manuscript. This process often takes more time than one would anticipate because technical experts are busy and not always available to review it in a timely manner. Referees are given forty-five days to complete the review, and typically, the referees take more than forty-five days. Further, in many cases, potential referees do not respond promptly to a request to review a manuscript. As a result of these issues, the Associate Editor has to seek new potential referees for the manuscript further delaying the review process.

The referees' responses usually include detailed comments that are used by the authors to help improve the manuscript and a recommendation on publication of the manuscript. Based on these responses, the Associate Editor makes a decision to accept, reject, or conditionally-accept the manuscript after further revisions. This letter is produced within the JAIF system by editing the appropriate decision letter. Before this letter is sent to the authors, it is reviewed and approved by both the Area Editor and the Editor-in-Chief. The typical manuscript is not accepted after the first cycle of reviews. Conditionally-accepted manuscripts are subjected to one or two (and sometimes three) cycles of additional revisions byte authors before it is accepted for publication. This process takes a few months to more than a year to complete. The reviews of no two manuscripts are identical nor take the same amount of time, and the more problems that the referees find with a manuscript the longer it takes the manuscript to get to publication. While such rounds of revisions might sound painful, the goal is to improve the quality of the paper and make it suitable for a selective archival journal—the stated goal of JAIF from its inception.

When a JAIF manuscript is accepted for publication, the authors are instructed to follow the guidelines found at http://www.isif.org/jaif.htm to prepare all required files for publication. These files are uploaded to the web-based review system or sent electronically via email to the Administrative Editor. It is important for authors get these files in as soon as possible so that the paper will appear in the next the available issue. Further, it is important for authors to submit all required files as these are utilized by the typesetter to produce a professional looking paper. A long delay in getting all required publication files to the Administrative Editor will extend the delay in publishing the associated manuscript.

Manuscript files received by the Administrative Editor are audited for completeness and accuracy, and sent to production after they are found to be acceptable. The typesetter uses LaTex to prepare a draft of the manuscript. The typesetter provides proofs of the manuscript to the Associate Editor-In-Chief, who reviews it for errors (i.e., technical, grammatical, formatting, etc). This review is in addition to the normal peer review performed by the referees. Once this editorial review is complete, the proofs and corrections are sent to the authors for further proofing. The goal of these multiple reviews is to ensure that high quality papers are published in JAIF.

While the delays in publication for the peer review and production processes can at times be frustrating for authors, the standards are set to provide high quality manuscripts to be published in JAIF. This strengthens the reputation of the journal and benefits the authors and research community. Thus, the editorial staff of JAIF is satisfied to publish two issues per year of high quality papers as opposed to more issues with paper of lesser quality. However, as JAIF grows with more manuscript submissions, the number of issues per year will increase.

Thank you again for submitting your manuscripts to JAIF, and the editorial staff encourage you to continue to consider JAIF for future manuscripts and serve our research community as a reviewer for JAIF.

Robert Lynch
Administrative Editor

Performance Analysis of Decentralized Kalman Filters under Communication Constraints

MARKUS S. SCHLOSSER
KRISTIAN KROSCHEL

Distributed fusion architectures are often used in multi-sensor target tracking as they are more robust and more flexible than centralized architectures. Furthermore, they allow for a reduction in the required communication bandwidth with only limited effect on the estimation performance. The trade-off between bandwidth and performance is analyzed in detail for the special case of a decentralized Kalman filter. As a result of this study, a conservative fusion approach for such systems with a reduced communication rate is proposed.

Manuscript received August 20, 2004; revised April 13, 2006, November 4, 2006, and October 8, 2007; released for publication November 8, 2007.

Refereeing of this contribution was handled by Benjamin Slocumb.

This work is part of the Sonderforschungsbereich (SFB) No. 588 *"Humanoide Roboter—Lernende und kooperierende multimodale Roboter"* at the University of Karlsruhe. The SFB is supported by the Deutsche Forschungsgemeinschaft (DFG).

Authors' addresses: M. S. Schlosser, Deutsche Thomson OHG, Karl-Wiechert-Allee 74, 30625 Hannover, Germany, E-mail: (markus.schlosser@thomson.net); K. Kroschel, Institut für Nachrichtentechnik, Universität Karlsruhe, 76128 Karlsruhe, Germany, E-mail: (kroschel@int.uni-karlsruhe.de).

1557-6418/07/$17.00 © 2007 JAIF

1. INTRODUCTION

In target tracking, multi-sensor systems are becoming more and more popular [14]. The advantages especially for physically distributed sensors are obvious: multiple viewing angles, different strong points of different sensors, and a higher robustness due to the inherent redundancy. On the other hand, some kind of fusion is necessary to integrate the data from the different sensors and to extract the desired information about the targets.

Traditionally, centralized fusion architectures have been used as their application is straightforward. All the data from the different sensors is sent to a single location to be fused. In recent years, increasing emphasis has been placed on distributed fusion where several fusion nodes exist in the network, like e.g., the Decentralized Kalman Filter (DKF) [11, 27], which is studied here, but also the covariance method [2], the federated filter [6, 7], a fusion system based on channel filters [23], and, most recently, a unified framework for optimal linear estimation fusion [16–21, 29].

As usual, the approaches based on Kalman filters are thereby mainly restricted to the linear Gaussian case. Furthermore, the unified framework is theoretically very insightful. As detailed in [17], the required generalized covariance matrix can, however, only be calculated accurately for some special cases. In many cases, it needs to be approximated numerically or even manually tuned. Finally, even if the covariance matrix can be determined accurately, this need not necessarily be possible in a recursive way so that no recursive estimator can be designed [16].

In a distributed fusion system, the sensor measurements are processed locally to produce state estimates, which are then transmitted between the fusion nodes. This approach is conceptually more complex as, even for statistically independent measurements, the local state estimates are correlated in time and among each other. In contrast to centralized fusion, there is also the danger of reusing information. Common information has to be detected and discarded in the fusion process. Additionally, the task of data association in tracking multiple targets, which is already difficult and still an active area of research for centralized architectures [4], becomes even more complex in the distributed case where only parts of the data are available at each fusion node. Finally, distributed fusion can even be inherently suboptimal [18]. A sufficient condition for distributed fusion to be optimal, however, is that the measurement noises are uncorrelated, which is often at least approximately given in real world scenarios.

On the other hand, the advantages of such distributed fusion architectures are a higher robustness due to a redundancy of fusion nodes and a lower processing load at each fusion node. It is also easier to integrate or scale existing systems. Therefore, distributed fusion is especially advantageous for large scale systems with many sensors. For such a system, another problem typically consists in only a limited amount of communication bandwidth being available. In this case, distributed fusion opens up the possibility to trade off bandwidth against performance by letting the fusion nodes communicate at reduced rates.

In [8, 12, 22], it was already pointed out that, if the communication rate is reduced, the information conveyed by the different local processors becomes correlated due to propagating the same underlying process noise. However, no quantitative measures for the amount of performance degradation were given. In [10], formulas describing the steady state performance of a DKF for arbitrary communication rates were derived and compared with simulative results for a linear system with two sensors and a nearly constant velocity model for the target dynamics. In our work, a simple formula for the performance degradation in the worst case of ignoring the correlation completely is derived. Its evaluation is straightforward compared with the solution of the asymmetric Lyapunov equation in [10]. As it represents the worst case, this analysis can be used to turn the too optimistic estimates of a system with reduced communication rate into conservative ones. Furthermore, the simulative study is extended to a nearly constant acceleration model, which results in some further insights.

The rest of this paper is organized as follows: The Decentralized Kalman Filter (DKF) is introduced in Section 2. In Section 3, the bandwidth requirements of centralized and distributed fusion architectures are compared with each other. The possibility for a reduction in communication rate is detailed in Section 4, and the resulting performance analyzed in Section 5. Based on the results, a new conservative fusion approach is proposed in Section 6. Finally, Section 7 recapitulates the most important findings.

2. DECENTRALIZED KALMAN FILTER

For a Kalman Filter (KF) to be applicable, the target's dynamics need to be modeled by the following state space equation

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k) \tag{1}$$

where $\mathbf{x}(k)$ is the state vector of the target at time instant k, typically containing the target position, velocity etc. $\mathbf{F}$ is the time-invariant state transition matrix and $\mathbf{w}(k)$ a white noise sequence with covariance matrix $\mathbf{Q}(k)$ representing the process noise. The state transition matrix $\mathbf{F}$ could just as well be time-variant for a KF. As this is not the case for the models studied later, it is omitted here.

Respectively, the linear measurement models are given by

$$\mathbf{y}_i(k) = \mathbf{H}_i\mathbf{x}(k) + \mathbf{v}_i(k) \tag{2}$$

where $\mathbf{y}_i(k)$ is the observation vector of the ith sensor, $i = 1,\ldots,N$. $\mathbf{H}_i$ is the corresponding measurement matrix and $\mathbf{v}_i(k)$ a zero-mean, white noise sequence with covariance matrix $\mathbf{R}_i(k)$ representing the measurement noise. To avoid stability issues [28], we assume for simplicity that every sensor is able to measure the complete position of the target and, thus, that the system is fully observable. Furthermore, only one easily detectable target shall be present in the scene without any clutter so that the task of data association is trivial.

According to these model equations, the *Centralized* KF (CKF) algorithm with multiple inputs in its information form can be described as recursively performing the following two steps to calculate the overall state estimate $\hat{\mathbf{x}}_{\mathrm{CKF}}(k \mid k)$ and error covariance matrix $\mathbf{P}_{\mathrm{CKF}}(k \mid k)$ at time instant k [3]:

1. Prediction

$$\hat{\mathbf{x}}_{\mathrm{CKF}}(k \mid k-1) = \mathbf{F}\hat{\mathbf{x}}_{\mathrm{CKF}}(k-1 \mid k-1) \tag{3}$$

$$\mathbf{P}_{\mathrm{CKF}}(k \mid k-1) = \mathbf{F}\mathbf{P}_{\mathrm{CKF}}(k-1 \mid k-1)\mathbf{F}^T + \mathbf{Q}(k-1). \tag{4}$$

2. Estimate correction

$$\hat{\mathbf{x}}_{\mathrm{CKF}}(k \mid k) = \mathbf{P}_{\mathrm{CKF}}(k \mid k)\left(\mathbf{P}_{\mathrm{CKF}}^{-1}(k \mid k-1)\hat{\mathbf{x}}_{\mathrm{CKF}}(k \mid k-1) + \sum_{i=1}^{N}\mathbf{H}_i^T\mathbf{R}_i^{-1}(k)\mathbf{y}_i(k)\right) \tag{5}$$

$$\mathbf{P}_{\mathrm{CKF}}^{-1}(k \mid k) = \mathbf{P}_{\mathrm{CKF}}^{-1}(k \mid k-1) + \sum_{i=1}^{N}\mathbf{H}_i^T\mathbf{R}_i^{-1}(k)\mathbf{H}_i. \tag{6}$$

This is called the information form as the inverse of the covariance matrix $\mathbf{P}^{-1}$ is a measure for the accuracy of the corresponding state estimate $\hat{\mathbf{x}}$ and thus for the information contained in it. Accordingly, $\mathbf{P}^{-1}(k \mid k-1)$ determines the weight given to $\hat{\mathbf{x}}(k \mid k-1)$ in (5).

In the *Decentralized* KF (DKF), Local Kalman Filters (LKFs) produce estimates $\hat{\mathbf{x}}_i(k \mid k)$ based on the information available from a single sensor i using the standard KF equations, i.e., (3)–(6) with $N = 1$. At a Fusion Center (FC), these estimates are fused together to form the overall state estimate $\hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k)$ [22]:

$$\hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k) = \mathbf{P}_{\mathrm{DKF}}(k \mid k)\left(\mathbf{P}_{\mathrm{DKF}}^{-1}(k \mid k-1)\hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k-1) + \sum_{i=1}^{N}[\mathbf{P}_i^{-1}(k \mid k)\hat{\mathbf{x}}_i(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-1)\hat{\mathbf{x}}_i(k \mid k-1)]\right) \tag{7}$$

$$\mathbf{P}_{\mathrm{DKF}}^{-1}(k \mid k) = \mathbf{P}_{\mathrm{DKF}}^{-1}(k \mid k-1) + \sum_{i=1}^{N}[\mathbf{P}_i^{-1}(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-1)] \tag{8}$$

where $\mathbf{P}_{\mathrm{DKF}}$ and $\mathbf{P}_i$ are the error covariance matrices of the state estimates $\hat{\mathbf{x}}_{\mathrm{DKF}}$ at the FC and $\hat{\mathbf{x}}_i$ at the LKFs, respectively.

The state estimate $\hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k)$ in (7) can easily be shown to be equivalent to $\hat{\mathbf{x}}_{\mathrm{CKF}}(k \mid k)$ in (5): Solving (5) in the single sensor case, i.e., $N = 1$, for the weighted measurement $\mathbf{H}_i^T\mathbf{R}_i^{-1}(k)\mathbf{y}_i(k)$ leads to the equivalence between $\mathbf{H}_i^T\mathbf{R}_i^{-1}(k)\mathbf{y}_i(k)$ in (5) and the gain in information between the predicted local estimates $\hat{\mathbf{x}}_i(k \mid k-1)$ and the corrected ones $\hat{\mathbf{x}}_i(k \mid k)$ in (7)

$$\mathbf{H}_i^T\mathbf{R}_i^{-1}(k)\mathbf{y}_i(k) = \mathbf{P}_i^{-1}(k \mid k)\hat{\mathbf{x}}_i(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-1)\hat{\mathbf{x}}_i(k \mid k-1). \quad (9)$$

Therefore, it is not the information contained in the local estimates itself but the gain in information that counts.

3. BANDWIDTH REQUIREMENTS

Looking at (7) and (8), it is sensible to save bandwidth in a DKF by directly transmitting the vector[1]

$$\Delta\hat{\mathbf{x}}_{\mathrm{weighted},i}(k) := \mathbf{P}_i^{-1}(k \mid k)\hat{\mathbf{x}}_i(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-1)\hat{\mathbf{x}}_i(k \mid k-1) \quad (10)$$

and the matrix

$$\Delta\mathbf{I}_i(k) := \mathbf{P}_i^{-1}(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-1) \quad (11)$$

instead of $\mathbf{P}_i^{-1}(k \mid k)$, $\hat{\mathbf{x}}_i(k \mid k)$, $\mathbf{P}_i^{-1}(k \mid k-1)$ and $\hat{\mathbf{x}}_i(k \mid k-1)$. This not only saves half of the communication bandwidth but also processing power at the FC.

Despite these savings, the bandwidth required by a distributed fusion network may still be higher compared with a centralized architecture, as the information packages are usually larger due to the state vector being of a higher dimension than the measurement vector. On the other hand, mislocalizations due to clutter are filtered out locally and need not be transmitted. Furthermore, many sensors do not provide position measurements of the object directly but merely scan the scene, like e.g., laser-radars. In this case, the bandwidth requirements would increase dramatically if the measurements were not preprocessed locally and, thus, a distributed fusion architecture lends itself naturally.

For a large system with many sensors, it is also likely that the central processor or the communication network are not able to handle the large amount of data transmitted by the sensors. In this case, centralized fusion is no longer applicable except if a corresponding amount of measurements is discarded completely. This is, however, likely to affect the estimation performance severely if no smart communications resource management techniques are applied [24]. In this case, distributed fusion opens up the possibility to distribute the processing load and to save the necessary bandwidth by letting the fusion nodes communicate less frequently.

Additionally, solar and battery powered sensors without wiring have become popular recently as they are easy to install [5, 13]. Due to their limited energy resources, it is important to save transmission energy even if sufficient bandwidth is available. The economical usage of transmission energy will become even more important in the future as the performance of microprocessors increases continuously according to Moore's law so that they need less and less energy for the same task. The energy needed for the transmission of the data is, however, mostly unaffected by these improvements [5].

Finally, it is sensible to adapt the reduction in communication rate to the movement of the tracked target. If the object is stationary or if it moves with a nearly constant velocity, its future state can be predicted more reliably and the communication rate can be reduced further than for a maneuvering target [9].

4. REDUCED COMMUNICATION RATE

As far as the DKF is concerned, reducing the communication rate at which information packages are sent from the LKFs to the FC by a factor m results in all predictions by one step being replaced with predictions by m steps in (7) and (8):

$$\hat{\mathbf{x}}_{\mathrm{DKF}_m}(k \mid k) = \mathbf{P}_{\mathrm{DKF}_m}(k \mid k)\left(\mathbf{P}_{\mathrm{DKF}_m}^{-1}(k \mid k-m)\hat{\mathbf{x}}_{\mathrm{DKF}_m}(k \mid k-m) + \sum_{i=1}^{N}[\mathbf{P}_i^{-1}(k \mid k)\hat{\mathbf{x}}_i(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-m)\hat{\mathbf{x}}_i(k \mid k-m)]\right) \quad (12)$$

$$\mathbf{P}_{\mathrm{DKF}_m}^{-1}(k \mid k) = \mathbf{P}_{\mathrm{DKF}_m}^{-1}(k \mid k-m) + \sum_{i=1}^{N}(\mathbf{P}_i^{-1}(k \mid k) - \mathbf{P}_i^{-1}(k \mid k-m)). \quad (13)$$

Any other communication issues are ignored for simplicity. Like in Section 2, it is still assumed that all sensors run synchronously and that information packages travel over communication links without any delay. Furthermore, the update rate at the LKFs is not affected. The LKFs still run at the sensor observation rate.

Unfortunately, if the predictions by one step are replaced by predictions by m steps, the gain in information is no longer based on one measurement but on m measurements and $m-1$ predictions. As these $m-1$ predictions are subject to the same process noise in all LKFs, the gain in information of the different LKFs is no longer statistically independent.

---

[1] "$x := \ldots$" means "x is defined as $\ldots$".

As this statistical dependence is not taken into account during the fusion process in the FC, the performance of the DKF degrades for such a reduced communication rate [10, 22]. On the other hand, if $\mathbf{w}(k) \equiv 0$, i.e., the target's dynamics can be modeled exactly, the local estimates $\hat{\mathbf{x}}_i(k \mid k)$ are not correlated and, therefore, the communication rate can be reduced at will without any performance degradation.

For less and less frequent communication between the LKFs and the FC, i.e., $m \to \infty$, the information contained in the predicted state estimates becomes less and less reliable. This is represented by the inverses of the corresponding error covariance matrices $\mathbf{P}_{\mathrm{DKF}}^{-1}(k \mid k-m)$ and $\mathbf{P}_i^{-1}(k \mid k-m)$ approaching zero. Thus, no weight is given to these estimates and they can be discarded in (12) and (13) leading to

$$\hat{\mathbf{x}}_{\text{naïve}}(k \mid k) = \mathbf{P}_{\text{naïve}}(k \mid k) \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k) \hat{\mathbf{x}}_i(k \mid k) \tag{14}$$

$$\mathbf{P}_{\text{naïve}}^{-1}(k \mid k) = \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k). \tag{15}$$

This is equivalent to the so-called naïve fusion architecture that assumes the local state estimates $\hat{\mathbf{x}}_i(k \mid k)$ and $\hat{\mathbf{x}}_j(k \mid k)$ to be statistically independent for all $i \neq j$.

As a consequence, $m \to \infty$ would result in a system where infinite intervals lie between two communication cycles so that no data would ever be transmitted. The performance of this hypothetical system can, however, readily be determined using the naïve fusion architecture whose track estimate can be formed for every time instant k, i.e., at the sensor observation rate.

From a different perspective, this somewhat astonishing finding can be explained as follows: As already stated, the local state estimates are statistically dependent due to propagating the same underlying process noise. This statistical dependence is properly corrected for in the DKF of (7) and (8) by the term

$$\begin{aligned} \mathbf{X}_{\mathrm{DKF}}(k \mid k-1) &:= \mathbf{P}_{\mathrm{DKF}}^{-1}(k \mid k-1) \hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k-1) \\ &\quad - \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k-1) \hat{\mathbf{x}}_i(k \mid k-1). \end{aligned} \tag{16}$$

In the DKF$_m$ of (12) and (13), this term is approximated by

$$\begin{aligned} \mathbf{X}_{\mathrm{DKF}_m}(k \mid k-m) &:= \mathbf{P}_{\mathrm{DKF}_m}^{-1}(k \mid k-m) \hat{\mathbf{x}}_{\mathrm{DKF}_m}(k \mid k-m) \\ &\quad - \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k-m) \hat{\mathbf{x}}_i(k \mid k-m) \end{aligned} \tag{17}$$

and the DKF$_{\text{naïve}}$ of (14) and (15) neglects the statistical dependence completely, i.e., $\mathbf{X}_{\mathrm{DKF}_{\text{naïve}}}(k \mid 0) \equiv 0$. Therefore, it can serve as a worst case scenario for such a reduction in communication rate, i.e., the estimate $\hat{\mathbf{x}}_{\text{naïve}}(k \mid k)$ in (14) is always less accurate than $\hat{\mathbf{x}}_{\mathrm{DKF}_m}(k \mid k)$ in (12), for all values of m.

Finally, it should be noted that $\mathbf{P}_{\mathrm{DKF}}$ is a valid estimate of the mean square error as it is an optimal Kalman filter. This is, however, neither true for $\mathbf{P}_m$ nor for $\mathbf{P}_{\text{naïve}}$ (except for $\mathbf{w}(k) \equiv 0$). As the DKF$_m$ and the DKF$_{\text{naïve}}$ ignore the correlation between the local estimates at least partially, they overestimate their own performance, i.e.,

$$\begin{aligned} \mathbf{P}_{\text{naïve}} \le \mathbf{P}_{\mathrm{DKF}_m} \le \mathbf{P}_{\mathrm{DKF}} &= \mathrm{E}\{\tilde{\mathbf{X}}_{\mathrm{DKF}} \tilde{\mathbf{X}}_{\mathrm{DKF}}^T\} \\ &\le \mathrm{E}\{\tilde{\mathbf{X}}_{\mathrm{DKF}_m} \tilde{\mathbf{X}}_{\mathrm{DKF}_m}^T\} \le \mathrm{E}\{\tilde{\mathbf{X}}_{\text{naïve}} \tilde{\mathbf{X}}_{\text{naïve}}^T\} \end{aligned} \tag{18}$$

where $\tilde{\mathbf{x}}_*(k \mid k) = \hat{\mathbf{x}}_*(k \mid k) - \mathbf{x}(k)$ and "$\mathbf{A} \le \mathbf{B}$" means that the difference $\mathbf{B} - \mathbf{A}$ is positive semidefinite. The relationship in (18) is valid for "$(k \mid k)$" as well as "$(k \mid k-1)$".

For complex systems containing several hierarchies or even feedback loops, this inconsistency between estimated and true error covariance matrix can build up and cause stability problems. Furthermore, it may cause problems during the association of the measurements to the tracked objects.

5. PERFORMANCE ANALYSIS

In this section, the performance degradation due to infrequent communication is investigated in detail. First, the performance of the DKF$_{\text{naïve}}$ defined by (14) and (15) is derived theoretically. As detailed in Section 4, it presents an upper bound on the performance degradation for a reduction in communication rate, as it ignores the correlation between the local estimates completely. Its performance is equivalent to a hypothetical system where infinite intervals lie between two communication cycles, i.e., $m \to \infty$. Second, the theoretical performance degradation is compared with simulative results. Finally, a simulative study concerning the performance of the DKF$_m$ of (12) and (13) is conducted for realistic reductions in the communication rate $m \ll \infty$.

5.1. Theoretical Analysis of DKF$_{\text{naïve}}$

The theoretical performance of the DKF$_{\text{naïve}}$ defined by (14) and (15) can be derived as follows. Introducing $\tilde{\mathbf{x}}_*(k \mid k) = \hat{\mathbf{x}}_*(k \mid k) - \mathbf{x}(k)$ in (14) leads to

$$\begin{aligned} \mathbf{P}_{\text{naïve}}^{-1}(k \mid k) \tilde{\mathbf{x}}_{\text{naïve}}(k \mid k) &= \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k) \tilde{\mathbf{x}}_i(k \mid k) \\ &\quad - \left[\mathbf{P}_{\text{naïve}}^{-1}(k \mid k) - \sum_{i=1}^{N} \mathbf{P}_i^{-1}(k \mid k) \right] \mathbf{x}(k). \end{aligned} \tag{19}$$

According to (15), the second term is zero. Therefore, the true error covariance matrix $\mathbf{P}_{\text{true}}(k \mid k) :=$

$E\{\tilde{\mathbf{x}}_{\text{naïve}}(k \mid k)\tilde{\mathbf{x}}^T_{\text{naïve}}(k \mid k)\}$ of the $\text{DKF}_{\text{naïve}}$ satisfies

$$\mathbf{P}^{-1}_{\text{naïve}}(k \mid k)\mathbf{P}_{\text{true}}(k \mid k)\mathbf{P}^{-1}_{\text{naïve}}(k \mid k)$$
$$= \sum_{i=1}^{N}\left(\mathbf{P}_i^{-1}(k \mid k) + \sum_{j=1 j\neq i}^{N}\mathbf{P}_i^{-1}(k \mid k)\boldsymbol{\Sigma}_{i,j}(k \mid k)\mathbf{P}_j^{-1}(k \mid k)\right) \tag{20}$$

where the cross-covariance $\boldsymbol{\Sigma}_{i,j}(k \mid k)$ between two local estimates $\hat{\mathbf{x}}_i(k \mid k)$ and $\hat{\mathbf{x}}_j(k \mid k)$ can be determined using the following equation [1]

$$\boldsymbol{\Sigma}_{i,j}(k \mid k) = (\mathbf{I} - \mathbf{K}_i(k)\mathbf{H}_i)\cdot(\mathbf{F}\boldsymbol{\Sigma}_{i,j}(k-1 \mid k-1)\mathbf{F}^T + \mathbf{Q})\cdot(\mathbf{I} - \mathbf{K}_j(k)\mathbf{H}_j)^T. \tag{21}$$

$\mathbf{K}_i(k)$ and $\mathbf{K}_j(k)$ denote the Kalman gains. Using (15) once more, (20) becomes

$$\mathbf{P}_{\text{true}}(k \mid k) = \mathbf{P}_{\text{naïve}}(k \mid k) + \mathbf{P}_{\text{naïve}}(k \mid k)\cdot\left(\sum_{i=1}^{N}\sum_{j=1 j\neq i}^{N}\mathbf{P}_i^{-1}(k \mid k)\boldsymbol{\Sigma}_{i,j}(k \mid k)\mathbf{P}_j^{-1}(k \mid k)\right)\mathbf{P}_{\text{naïve}}(k \mid k). \tag{22}$$

Therefore, the inconsistency of (18) between the true covariance matrix $\mathbf{P}_{\text{true}} = E\{\tilde{\mathbf{x}}_{\text{naïve}}(k \mid k)\tilde{\mathbf{x}}^T_{\text{naïve}}(k \mid k)\}$ and the estimated covariance matrix $\mathbf{P}_{\text{naïve}}$ consists in the term depending on the cross-covariance $\boldsymbol{\Sigma}_{i,j}$ between the local estimates.

The case of identical sensors, i.e., $\mathbf{P}_i(k \mid k) = \mathbf{P}_{\text{LKF}}(k \mid k)\ \forall i$, also results in $\boldsymbol{\Sigma}_{i,j}(k \mid k) = \boldsymbol{\Sigma}_{j,i}(k \mid k) = \boldsymbol{\Sigma}_{\text{LKFs}}(k \mid k)\ \forall i,j$. Therefore, the true error covariance matrix simplifies to

$$\begin{aligned}\mathbf{P}_{\text{true}}(k \mid k) &= \mathbf{P}_{\text{naïve}}(k \mid k) + N(N-1)\mathbf{P}_{\text{naïve}}(k \mid k)\\ &\quad\cdot\mathbf{P}^{-1}_{\text{LKF}}(k \mid k)\boldsymbol{\Sigma}_{\text{LKFs}}(k \mid k)\mathbf{P}^{-1}_{\text{LKF}}(k \mid k)\mathbf{P}_{\text{naïve}}(k \mid k)\\ &= \mathbf{P}_{\text{naïve}}(k \mid k) + \frac{N-1}{N}\boldsymbol{\Sigma}_{\text{LKFs}}(k \mid k)\end{aligned} \tag{23}$$

as $\mathbf{P}_{\text{naïve}}(k \mid k) = (1/N)\mathbf{P}_{\text{LKF}}(k \mid k)$.

The evaluation of (22) and (23) is straightforward compared with the solution of the asymmetric Lyapunov equation in [10] describing the steady state performance of a DKF for arbitrary communication rates, as the latter consists of several non-trivial systems of linear equations. A further advantage consists in the fact that the term depending on the cross-covariance $\boldsymbol{\Sigma}_{i,j}$ in (22) and (23) represents an explicit explanation for the inconsistency of the naïve fusion architecture.

5.2. Simulative Analysis of $\text{DKF}_{\text{naïve}}$

In addition to validating the theoretical analysis of the $\text{DKF}_{\text{naïve}}$, the following simulative study shall provide some quantitative numbers for its performance degradation and its inconsistency in typical tracking scenarios. In contrast to [10], the study not only covers a

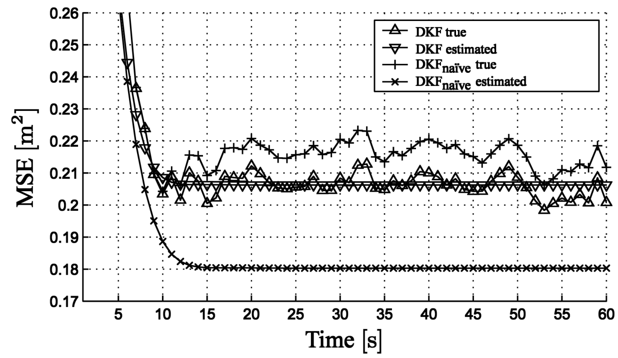


Fig. 1. Comparison between true and estimated MSE of DKF and $\text{DKF}_{\text{naïve}}$ ($q_{\text{NCV}} = 0.01\text{m}^2/(\text{s}^3)$, $T_s = 1$ s, $\sigma_v = 1$ m).

nearly constant velocity but also a nearly constant acceleration model for the target dynamics. The results lead to a generalization of the target maneuvering index [3], which is commonly used as the decisive parameter for describing the performance of a Kalman filter. For simplicity the comparison is restricted to the steady state.

5.2.1. Nearly Constant Velocity

First, $N = 2$ sensors track a target whose one-dimensional dynamics can be described by the following Nearly Constant Velocity (NCV) model [3]

$$\begin{bmatrix} x(k+1) \\ \dot{x}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x(k) \\ \dot{x}(k) \end{bmatrix} + \mathbf{w}_{\text{NCV}}(k) \tag{24}$$

where the process noise is given by

$$\mathbf{w}_{\text{NCV}}(k) = \int_0^{T_s}\begin{bmatrix} T_s - t \\ 1 \end{bmatrix} u_{\text{NCV}}(kT_s + t)dt \tag{25}$$

and $u_{\text{NCV}}(t)$ is zero-mean continuous random noise with power spectral density q_{NCV}, leading to the following process noise covariance matrix

$$\mathbf{Q}_{\text{NCV}} = \begin{bmatrix} \frac{1}{3}T_s^3 & \frac{1}{2}T_s^2 \\ \frac{1}{2}T_s^2 & T_s \end{bmatrix} q_{\text{NCV}}. \tag{26}$$

The measurements $y_i(k)$ are the position in Cartesian coordinates

$$y_i(k) = x(k) + v_i(k), \qquad i = 1,2 \tag{27}$$

with variances $\sigma^2_{v,i}$.

Fig. 1 presents the comparison between the true and estimated Mean Square Error (MSE)

$$\text{MSE} \approx E\{(\hat{x}_*(k \mid k) - x(k))^2\} \tag{28}$$

of the position component in $\hat{\mathbf{x}}_{\text{DKF}}(k \mid k)$ of (7) and $\hat{\mathbf{x}}_{\text{naïve}}(k \mid k)$ of (14) for a typical example, where $q_{\text{NCV}} = 0.01\text{m}^2/(\text{s}^3)$, $T_s = 1$ s and $\sigma_v = 1$ m. The "true" MSEs are thereby averaged on 5000 Monte Carlo runs. In accordance with (18), it can be seen how the DKF slightly outperforms the $\text{DKF}_{\text{naïve}}$ by approximately 4%. Furthermore, the DKF estimates its accuracy correctly

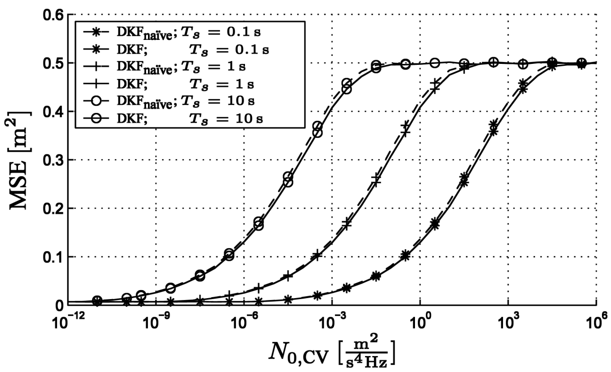


Fig. 2. Comparison between $\mathrm{MSE}_{\text{naïve}}$ and $\mathrm{MSE}_{\mathrm{DKF}}$ as a function of the process noise q_{NCV} ($\sigma_v = 1$ m).

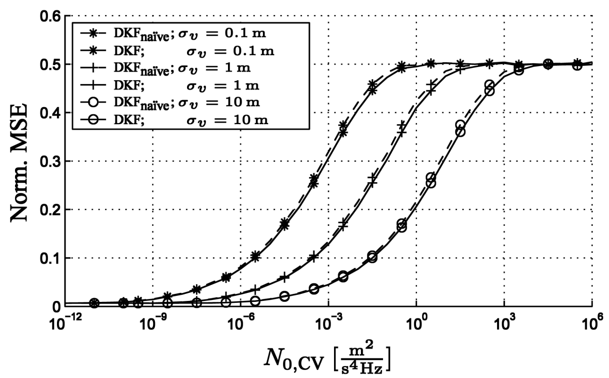


Fig. 3. Comparison between $\mathrm{MSE}_{\text{naïve}}$ and $\mathrm{MSE}_{\mathrm{DKF}}$ normalized by σ_v^2 as a function of the process noise q_{NCV} ($T_s = 1$ s).

whereas the $\mathrm{DKF}_{\text{naïve}}$ produces estimates of its MSE that are about 16% more accurate (smaller) as compared to the DKF_m even though its true MSE is slightly worse.

Fig. 2 displays the true MSEs of DKF and $\mathrm{DKF}_{\text{naïve}}$ averaged over time as a function of the process noise q_{NCV} for three different sampling periods $T_s = 0.1$ s, 1 s, 10 s and a measurement noise $\sigma_v = 1$ m. Fig. 3 shows the corresponding comparison for varying $\sigma_v = 0.1$ m, 1 m, 10 m and $T_s = 1$ s. To allow for a comparison, the MSEs are normalized by σ_v^2 this time. For every point on each line, 1000 Monte Carlo runs with 400 measurements per simulation were performed. To obtain an estimate of the steady state performance, the averages are only based on the last 200 measurements.

For high values of q_{NCV}, the curves in both figures approach the variance of the combined measurements $\sigma_{v,\mathrm{comb}}^2 = \sigma_{v,1}^2\sigma_{v,2}^2/(\sigma_{v,1}^2 + \sigma_{v,2}^2) = 0.5\sigma_v^2$. For low values of q_{NCV}, they should approach zero as KFs are able to estimate the position arbitrarily precisely if no process noise is present and if a sufficient number of observations is available. Due to a finite number of observations, this is not the case here.

In both Figs. 2 and 3, changing the parameter merely results in a shift of the corresponding curves. Their shape is not affected. If T_s is increased by a factor of 10, q_{NCV} needs to be decreased by a factor of 1000 to obtain the same results. On the other hand, if σ_v is

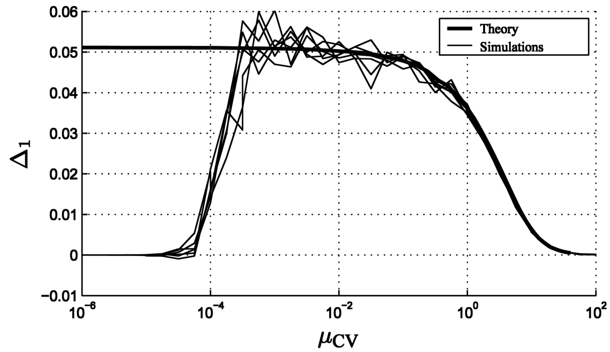


Fig. 4. Relative difference Δ_1 between $\mathrm{MSE}_{\text{naïve}}$ and $\mathrm{MSE}_{\mathrm{DKF}}$ as a function of the target maneuvering index μ_{NCV}.

increased by a factor of 10, q_{NCV} needs to be increased by a factor of 100.

Therefore, using the target maneuvering index [3]

$$\mu_{\mathrm{NCV}} := \sqrt{\frac{q_{\mathrm{NCV}} T_s^3}{\sigma_v^2}} \tag{29}$$

as independent variable leads to an invariance against a variation in the sampling period T_s and the measurement noise σ_v. This can be seen in Fig. 4, which shows the relative difference[2]

$$\Delta_1 = \frac{\mathrm{MSE}_{\text{naïve}} - \mathrm{MSE}_{\mathrm{DKF}}}{\mathrm{MSE}_{\mathrm{DKF}}} \stackrel{!}{\equiv} \frac{P_{\mathrm{true}}^{(1,1)} - P_{\mathrm{DKF}}^{(1,1)}}{P_{\mathrm{DKF}}^{(1,1)}} \tag{30}$$

between the six corresponding dashed and solid curve pairs in Figs. 2 and 3. The identity with the right-hand side follows from the definition of $\mathbf{P}_{\mathrm{true}}$ in (23) and the fact that the DKF estimates its accuracy correctly. This analytical prediction of Δ_1 is represented by the bold line.

For large values of μ_{NCV}, the different curves all approach zero. This can be explained by the predicted states $\hat{\mathbf{x}}_i(k \mid k-1)$ being not only correlated but also unreliable in this case. As large values of μ_{NCV} also imply a large process noise q, the predicted states are given almost no weight compared with the measurements $y_i(k)$ in calculating $\hat{\mathbf{x}}_i(k \mid k)$ in the LKFs. Note that the Kalman filters are of little use in this case.

For small values of μ_{NCV}, (23) predicts Δ_1 to approach 5.1% asymptotically, whereas the simulations show Δ_1 to approach zero. This significant difference can be explained by the analytical curve predicting the steady state behavior, whereas the steady state is never reached during the simulations for such small values of μ_{NCV}. As described by (21) and as depicted in Fig. 5, the cross-covariance $\mathbf{\Sigma}_{i,j}(k \mid k)$ between the local estimates rises only slowly for low values of μ_{NCV}. Therefore, the estimates $\hat{\mathbf{x}}_{\mathrm{DKF}}(k \mid k)$ and $\hat{\mathbf{x}}_{\text{naïve}}(k \mid k)$ are quasi identical and Δ_1 equals zero during the initialization phase.

[2] "$\{!\}$ over $\{\equiv, =, <, >\}$" means "shall be" or "needs to be." Furthermore, the superscript $(1,1)$ denotes the upper-left element of the matrix.

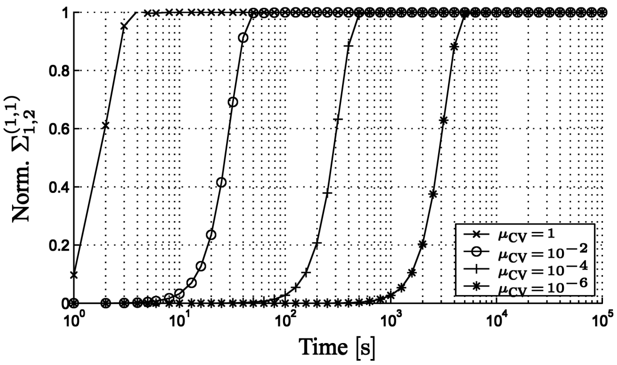


Fig. 5. Time dependency of the normalized cross-covariance between the local position estimates.

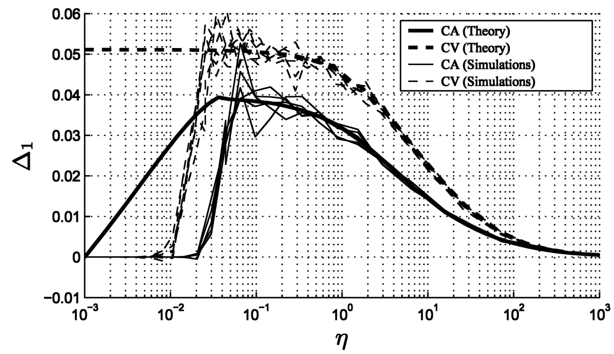


Fig. 7. Relative difference Δ_1 between $\mathrm{MSE}_{\text{naïve}}$ and $\mathrm{MSE}_{\text{DKF}}$ as a function of the weighting ratio η.

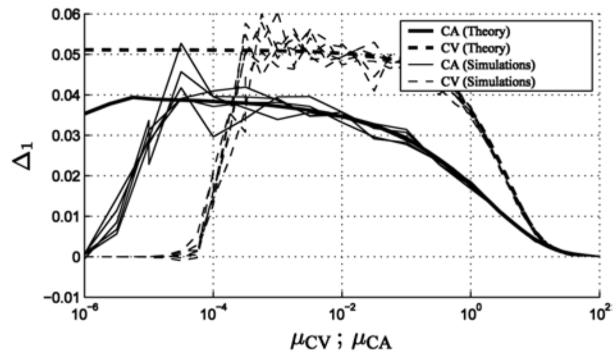


Fig. 6. Relative difference Δ_1 between $\mathrm{MSE}_{\text{naïve}}$ and $\mathrm{MSE}_{\text{DKF}}$ as a function of the target maneuvering index μ.

5.2.2. Nearly Constant Acceleration

In this section, the previous analysis is extended to a Nearly Constant Acceleration (NCA) model being used for the target dynamics, i.e.,

$$\begin{bmatrix} x(k+1) \\ \dot{x}(k+1) \\ \ddot{x}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T_s & \frac{1}{2}T_s^2 \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k) \\ \dot{x}(k) \\ \ddot{x}(k) \end{bmatrix} + \mathbf{w}_{\mathrm{NCA}}(k) \tag{31}$$

where the process noise is given by

$$\mathbf{w}_{\mathrm{NCA}}(k) = \int_0^{T_s} \begin{bmatrix} \frac{1}{2}(T_s - t)^2 \\ T_s - t \\ 1 \end{bmatrix} u_{\mathrm{NCA}}(kT_s + t)dt \tag{32}$$

and $u_{\mathrm{NCA}}(t)$ is a zero-mean continuous random noise signal with power spectral density q_{NCA}, leading to the following process noise covariance matrix

$$\mathbf{Q}_{\mathrm{NCA}} = \begin{bmatrix} \frac{1}{20}T_s^5 & \frac{1}{8}T_s^4 & \frac{1}{6}T_s^3 \\ \frac{1}{8}T_s^4 & \frac{1}{3}T_s^3 & \frac{1}{2}T_s^2 \\ \frac{1}{6}T_s^3 & \frac{1}{2}T_s^2 & T_s \end{bmatrix} q_{\mathrm{NCA}}. \tag{33}$$

Fig. 6 shows the relative difference Δ_1 between the true MSEs of the $\mathrm{DKF}_{\text{naïve}}$ and the DKF for the NCV and the NCA model. For the latter, the target maneuvering index is defined as [3]

$$\mu_{\mathrm{NCA}} := \sqrt{\frac{q_{\mathrm{NCA}} T_s^5}{\sigma_v^2}}. \tag{34}$$

Like in the nearly constant velocity case, the curves are invariant against a variation in T_s and σ_v. On the other hand, the curves for the two different target models clearly do not match.

A comparison between the definitions of the target maneuvering indices in (29) and (34) and the respective process noise covariance matrices in (26) and (33) reveals that μ^2 is proportional to the ratio $Q^{(1,1)}/\sigma_v^2$ between the variance of the predicted position estimate *due to the process noise* and the variance of the measurement. Therefore, μ is an indicator for how much weight is given to the measurement and the predicted estimate during the track update in the LKFs, respectively (see (5) with $N = 1$).

On the other hand, as described in (4), the accuracy of the prediction in the LKFs does not only depend on the covariance of the process noise $\mathbf{Q}$ but also on the state transition matrix $\mathbf{F}$ and the inaccuracy of the last estimate $\mathbf{P}(k-1 \mid k-1)$. The correct ratio

$$\eta := \frac{P_{\mathrm{LKF,\,pred}}^{(1,1)}}{\sigma_v^2} \tag{35}$$

between the weight given to the measurement and the weight given to the predicted position estimate can be taken from the simulations. For the case of a NCV and a NCA model, it can, however, also be calculated for the steady state by numerically solving a system of nonlinear equations (as detailed in the Appendix).

The resulting curves for the relative difference Δ_1 as a function of η are displayed in Fig. 7, where the bold lines again present the analytical predictions. This time the curves for the NCV and the NCA model show the same progression. Only the maximum degradation for the NCA model of around 4% is lower than the 5.1% for the NCV model, as also predicted by the theory. This can be explained by the same weighting ratio η being reached for a smaller process noise q for the NCA model. As the common process noise is responsible

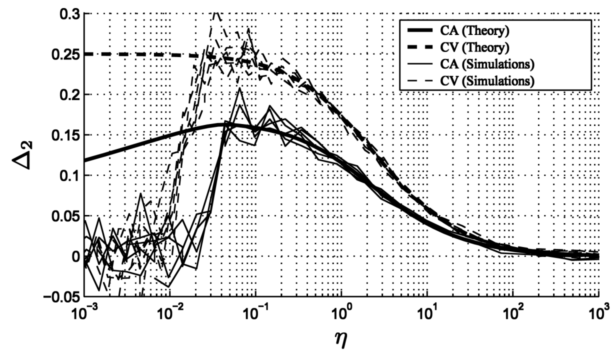


Fig. 8. Relative difference Δ_2 between true and estimated MSE of $\text{DKF}_{\text{naïve}}$ as a function of the weighting ratio η.

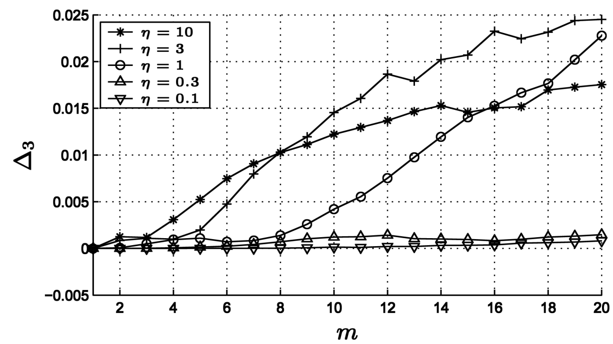


Fig. 9. Relative difference Δ_3 between $\text{MSE}_{\text{DKF}_m}$ and MSE_{DKF} as a function of the the update rate m for the nearly constant velocity model (T_s = 1 s, σ_v = 1 m).

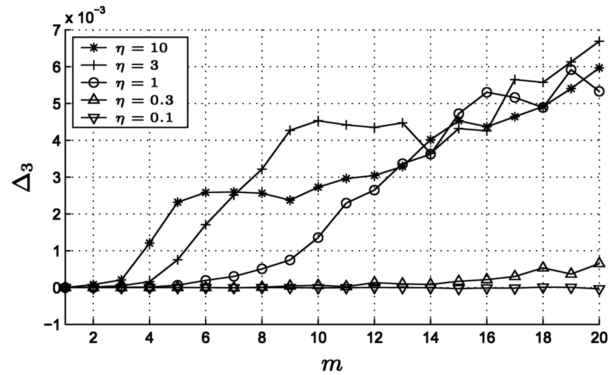


Fig. 10. Relative difference Δ_3 between $\text{MSE}_{\text{DKF}_m}$ and MSE_{DKF} as a function of the the update rate m for the nearly constant acceleration model (T_s = 1 s, σ_v = 1 m).

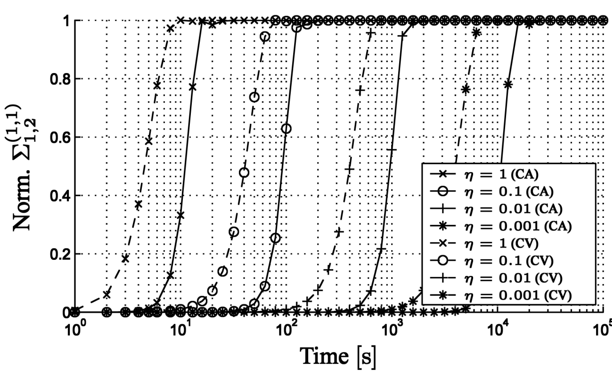


Fig. 11. Time dependency of the normalized cross-covariance between the local position estimates.

for the correlation between the local estimates, the correlation is also lower for the NCA model.

Note that for small values of η even the analytical prediction fails for the NCA model. This is due to the numerical solution of the system of non-linear equations, which is needed to determine the steady state performance of the KFs, reaching its limits for such small values. Simulations with 10000 measurements indicate that Δ_1 stays at its maximum also in the NCA case.

KFs do not only produce state estimates $\hat{\mathbf{x}}$ but also calculate an accuracy of these estimates in form of the error covariance matrix $\mathbf{P}$. As stated earlier, the DKF estimates its accuracy correctly whereas the $\text{DKF}_{\text{naïve}}$, although performing slightly worse, estimates its accuracy even better than that of the DKF.

Fig. 8 shows this difference

$$\Delta_2 = \frac{\text{MSE}_{\text{naïve}} - \text{MSE}_{\text{estimated}}}{\text{MSE}_{\text{estimated}}} \stackrel{!}{=} \frac{P^{(1,1)}_{\text{true}} - P^{(1,1)}_{\text{naïve}}}{P^{(1,1)}_{\text{naïve}}} \tag{36}$$

between the true and estimated MSE of the $\text{DKF}_{\text{naïve}}$ as a function of the weighting ratio η. It can be seen how this difference Δ_2 can again be predicted very precisely by (23). The shape of these curves is very similar to those in Fig. 7 for the difference between the true MSE of the $\text{DKF}_{\text{naïve}}$ and the DKF. The maxima are, however, a lot higher at 16% and 25%, respectively. This significant overestimation of its own performance in the $\text{DKF}_{\text{naïve}}$ can lead to severe stability problems if the estimates are propagated to other fusion nodes in the system or even fed back to the local estimators.

5.3. Performance Analysis of DKF_m

As already stated, the $\text{DKF}_{\text{naïve}}$ presents an upper bound on the performance degradation for a reduction in communication rate. Its performance is equivalent to a hypothetical system where infinite intervals lie between two communication cycles, i.e., $m \to \infty$. In this section, the performance degradation is examined for realistic reductions in communication rate $m \ll \infty$ using the DKF_m of (12) and (13).

To this end, Figs. 9 and 10 show the relative difference

$$\Delta_3 = \frac{\text{MSE}_{\text{DKF}_m} - \text{MSE}_{\text{DKF}}}{\text{MSE}_{\text{DKF}}} \tag{37}$$

between the MSE of the position component in the state estimate $\hat{\mathbf{x}}_{\text{DKF}_m}(k \mid k)$ of (12) and $\hat{\mathbf{x}}_{\text{DKF}}(k \mid k)$ of (7) as a function of the update rate m for the nearly constant velocity (NCV) and nearly constant acceleration (NCA) model, respectively. The sampling period was set to T_s = 1 s and the measurement noises to σ_v = 1 m. This time 10000 Monte Carlo runs were performed on simulations with 400 measurements.

In both figures, it can be seen that the maximum difference of the $\mathrm{DKF}_{\text{naïve}}$ (see Fig. 7) is not reached in the shown interval $1 \leq m \leq 20$ even for large values of η and that Δ_3 remains almost zero for small values of η. The latter is due to the state vector $\mathbf{x}(k)$ changing only slowly in such scenarios. Therefore, the term $\mathbf{X}_{\mathrm{DKF}_m}(k \mid k-m)$ of (17), which corrects for the statistical dependence between $\hat{\mathbf{x}}_1(k \mid k)$ and $\hat{\mathbf{x}}_2(k \mid k)$ in the DKF_m, stays longer an accurate estimate of the true term $\mathbf{X}_{\mathrm{DKF}}(k \mid k-1)$ of (16) in the DKF.

A comparison between the curves for the NCV and the NCA model shows that those for the NCA model are significantly lower. This can be explained by the maximal degradation being lower, as indicated in Fig. 7, and the maximum cross-covariance also being reached more slowly, as indicated in Fig. 11. The reason for this is again that a smaller process noise q is needed for the same weighting ratio η in the NCA case, and the common process noise being responsible for the correlation between the local estimates.

As a result, it can be concluded that, even for the worst case of $\eta \in [1, 10]$ and the NCV model, the communication rate between the LKFs and the FC can be reduced by at least a factor $m = 8$ without introducing an additional error of more than 1%. For sensors producing measurements at high rates, the error due to infrequent communication is typically even smaller as the short sampling period T_s results in a small target maneuvering index μ and, thus, a small weighting ratio η. Exactly in these scenarios, a reduction in communication rate is most likely to be desirable.

6. CONSERVATIVE FUSION APPROACH

Sophisticated approaches for distributed tracking systems exist that are specifically designed to be robust against unmodeled correlations, like e.g., the covariance intersection method [15] or the bounded covariance inflation method [25]. On the other hand, the results of the last section give rise to a conservative, but simple alternative fusion approach for a reduction in communication rate.

Combining the results of Section 5 with (18) it can be seen that $\mathbf{P}_{\mathrm{DKF}_m}(k \mid k)$ is a slightly too optimistic and $\mathrm{E}\{\tilde{\mathbf{X}}_{\text{naïve}}(k \mid k)\tilde{\mathbf{X}}_{\text{naïve}}(k \mid k)^T\}$ a slightly too conservative estimate of the true accuracy $\mathrm{E}\{\tilde{\mathbf{X}}_{\mathrm{DKF}_m}(k \mid k)$ $\tilde{\mathbf{X}}_{\mathrm{DKF}_m}(k \mid k)^T\}$ of the DKF_m in (12) and (13). As $\mathrm{E}\{\tilde{\mathbf{X}}_{\text{naïve}}(k \mid k)\tilde{\mathbf{X}}_{\text{naïve}}(k \mid k)^T\}$ is equivalent to $\mathbf{P}_{\text{true}}(k \mid k)$, it can be determined using (22) and (23). Therefore, the DKF_m of (12) and (13) can be used without alteration for the estimation of $\hat{\mathbf{x}}_{\mathrm{DKF}_m}(k \mid k)$, i.e., $\mathbf{P}_{\mathrm{DKF}_m}(k \mid k)$ is still used during the recursive estimation. To obtain a conservative estimate, $\mathbf{P}_{\mathrm{DKF}_m}(k \mid k)$ is then simply replaced by $\mathbf{P}_{\text{true}}(k \mid k)$.

The computation of $\mathbf{P}_{\text{true}}(k \mid k)$ in (22) and (23), however, requires the knowledge of $\mathbf{\Sigma}_{i,j}(k \mid k)$ in (21) and, thus, of $\mathbf{K}_i(k)$, for all i and k. In the studied case with a time-invariant state-space representation of the tracking system, the $\mathbf{K}_i(k)$ can simply be (pre-)computed at the FC as they do not depend on the actually observed measurements [3]. In many real-world scenarios, this condition is, however, likely to be violated, like e.g., for a maneuvering target or if the measurement accuracy depends on the distance from the sensor to the target. In this case, the FC may determine the gain sequence at least approximately, in particular if additional information is provided by the LKFs. As (21) is not restricted to the $\mathbf{K}_i(k)$ being the Kalman gains, a predictable sequence of suboptimal filter gains may also be used.

Finally, if the calculation of $\mathbf{P}_{\text{true}}(k \mid k)$ for every time step k shall be avoided, an even more conservative approximation consists in determining the maximum difference $\Delta\mathbf{P}_{\max}$ between $\mathbf{P}_{\text{true}}(k \mid k)$ and $\mathbf{P}_{\text{naïve}}(k \mid k)$, and adding this difference to $\mathbf{P}_{\mathrm{DKF}_m}(k \mid k)$, i.e.,

$$\mathbf{P}_{\mathrm{c}}(k \mid k) = \mathbf{P}_{\mathrm{DKF}_m}(k \mid k) + \Delta\mathbf{P}_{\max}. \tag{38}$$

The difference between $\mathbf{P}_{\text{true}}(k \mid k)$ and $\mathbf{P}_{\text{naïve}}(k \mid k)$ is the term depending on the cross-covariance $\mathbf{\Sigma}_{i,j}(k \mid k)$ between the local estimates in (22) and (23). As the cross-covariance builds up slowly, this term is maximum for the steady state, i.e.,

$$\Delta\mathbf{P}_{\max} = \mathbf{P}_{\text{true}}(\infty \mid \infty) - \mathbf{P}_{\text{naïve}}(\infty \mid \infty). \tag{39}$$

As the Kalman gains $\mathbf{K}_i(\infty)$ are constant, (21) turns into a simple system of linear equations. Consequently, if the $\mathbf{K}_i(\infty)$ can be determined (like e.g., for the NCV and the NCA model), this is also true for $\mathbf{\Sigma}_{i,j}(\infty \mid \infty)$ and, thus, $\Delta\mathbf{P}_{\max}$.

7. CONCLUSIONS

If the communication rate between the Local Kalman Filters (LKFs) and the Fusion Center (FC) needs to be reduced to save communication bandwidth, the performance of the fusion process degrades as the information provided by the different sensors becomes correlated due to propagating the same underlying process noise. A simulative study is performed for a simple system with two local sensors, which are both able to measure the position of the target. It is shown that the ratio η between the weight given to the measurement and the weight given to the predicted position estimate during the track update in the LKFs is more useful in describing the performance degradation than the target maneuvering index. Furthermore, it is found that, even for the worst case of $\eta \in [1, 10]$ and the nearly constant velocity model, the communication rate between the LKFs and the FC can be reduced by at least a factor of 8 without introducing an additional error of more than 1%.

Furthermore, a simple formula for the performance degradation in the worst case of ignoring the correlation due to such a reduction in communication rate completely is derived. As its evaluation is straightforward compared with the solution of the asymmetric Lyapunov equation for the general case, it can be used for a conservative fusion architecture where the slightly too optimistic state estimates due to such a reduction in communication rate are replaced by slightly too conservative ones.

APPENDIX

In the steady state, the Kalman filter simplifies to

$$\hat{\mathbf{x}}(k \mid k) = \mathbf{F}\hat{\mathbf{x}}(k-1 \mid k-1) + \mathbf{K}(\mathbf{y}(k) - \mathbf{H}_i\hat{\mathbf{x}}(k \mid k-1)) \tag{40}$$

with a constant Kalman gain $\mathbf{K}$. If, as in our case, only the position of the object is measured in 1D, $\mathbf{K}_i$ becomes a vector

$$\mathbf{K}_{\mathrm{NCV}} =: \begin{pmatrix} \alpha \\ \beta/T_s \end{pmatrix} \qquad \text{and} \qquad \mathbf{K}_{\mathrm{NCA}} =: \begin{pmatrix} \alpha \\ \beta/T_s \\ \gamma/T_s^2 \end{pmatrix} \tag{41}$$

in the nearly constant velocity (NCV) and nearly constant acceleration (NCA) case, respectively. Accordingly, the filters are denoted optimal α-β- and α-β-γ-filter [3]. It can easily be shown [26] that

$$\eta := \frac{P_{\mathrm{pred}}^{(1,1)}}{\sigma_v^2} = \frac{\alpha}{1-\alpha}. \tag{42}$$

The corresponding formulas to calculate α and β for a NCV model excited by discretized continuous-time white process noise can be found in [3]:

$$\alpha = \sqrt{2\beta + \frac{\beta^2}{12}} - \frac{\beta}{2} \tag{43}$$

$$\frac{\beta^2}{1-\alpha} = \frac{q_{\mathrm{NCV}} T_s^3}{\sigma_v^2} =: \mu_{\mathrm{NCV}}^2. \tag{44}$$

The following equations for α, β and γ for a NCA model being excited by a discretized continuous-time white process noise can be obtained using the same approach:

$$\beta^2 = 2\alpha\gamma \tag{45}$$

$$\frac{\gamma^2}{120\alpha^2} - \frac{\gamma}{2\alpha} + \frac{4\gamma}{\alpha\beta} - \frac{2\gamma}{\beta} = 1 \tag{46}$$

$$\frac{\gamma^2}{1-\alpha} = \frac{q_{\mathrm{NCA}} T_s^5}{\sigma_v^2} =: \mu_{\mathrm{NCA}}^2. \tag{47}$$

REFERENCES

[1] Y. Bar-Shalom
On the track-to-track correlation problem.
IEEE Transactions on Automatic Control, **26**, 2 (Apr. 1981), 571–572.

[2] Y. Bar-Shalom and L. Campo
The effect of the common process noise on the two-sensor fused-track covariance.
IEEE Transactions on Aerospace and Electronic Systems, **22**, 6 (Nov. 1986), 803–804.

[3] Y. Bar-Shalom and T. Fortmann
Tracking and Data Association.
New York: Academic Press, 1988.

[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] R. R. Brooks
Data Fusion for a Distributed Ground-Based Sensing System.
In D. L. Hall and J. Llinas (Eds.), *Handbook of Multisensor Data Fusion*, Boca Raton: CRC Press, 2001, 25-1–25-9.

[6] N. A. Carlson
Federated square root filter for decentralized parallel processes.
IEEE Transactions on Aerospace and Electronic Systems, **26**, 3 (May 1990), 517–525.

[7] N. A. Carlson
Federated filter for distributed navigation and tracking applications.
In *Proceedings of the ION 58th Annual Meeting/CIGTF 21st Guidance Test Symposium*, June 2002, 340–353.

[8] K. C. Chang, R. K. Saha and Y. Bar-Shalom
On optimal track-to-track fusion.
IEEE Transactions on Aerospace and Electronic Systems, **33**, 4 (Oct. 1997), 1271–1276.

[9] K. C. Chang, X. Yin and R. K. Saha
Evaluating a linear predictive algorithm for bandwidth conservation.
IEEE Transactions on Aerospace and Electronic Systems, **36**, 4 (Oct. 2000), 1407–1414.

[10] K. C. Chang, T. Zhi and R. K. Saha
Performance evaluation of track fusion with information matrix filter.
IEEE Transactions on Aerospace and Electronic Systems, **38**, 2 (Apr. 2002), 455–466.

[11] C.-Y. Chong
Hierarchical estimation.
In *Proceedings of the MIT/ONR Workshop on C3*, July 1979, 205–220.

[12] O. E. Drummond
On track and tracklet fusion filtering.
In *Proceedings of the SPIE Conference on Signal and Data Processing of Small Targets*, 2002, vol. 4728, 176–194.

[13] P. F. Gorder
Sizing up smart dust.
Journal on Computing in Science and Engineering, **5**, 6 (Nov.–Dec. 2003), 6–9.

[14] D. L. Hall and J. Llinas (Eds.)
Handbook of Multisensor Data Fusion.
Boca Raton, FL: CRC Press, 2001.

[15] S. Julier and J. K. Uhlmann
General Decentralized Data Fusion with Covariance Intersection (CI).
In D. L. Hall and J. Llinas (Eds.), *Handbook of Multisensor Data Fusion*, Boca Raton, FL: CRC Press, 2001, 12-1–12-5.

[16] X. R. Li
Unified optimal linear estimation fusion—Part VII: Dynamic systems.
In *Proceedings of the International Conference on Information Fusion* (Fusion), July 2003, 445–462.

[17] X. R. Li and J. Wang
Optimal linear estimation fusion—Part II: Discussions and examples.
In *Proceedings of the International Conference on Information Fusion* (Fusion), July 2000, MoC2-18–MoC2-25.

[18] X. R. Li and K.-S. Zhang
Optimal linear estimation fusion—Part IV: Optimality and efficiency of distributed fusion.
In *Proceedings of the International Conference on Information Fusion* (Fusion), Aug. 2001, WeB1-19–WeB1-26.

[19] X. R. Li, K.-S. Zhang, J. Zhao and Y.-M. Zhu
Optimal linear estimation fusion—Part V: Relationships.
In *Proceedings of the International Conference on Information Fusion* (Fusion), July 2002, 497–504.

[20] X. R. Li and P. Zhang
Unified optimal linear estimation fusion—Part III: Cross-correlation of local estimation errors.
In *Proceedings of the International Conference on Information Fusion* (Fusion), Aug. 2001, WeB1-11–WeB1-18.

[21] X. R. Li, Y.-M. Zhu, J. Wang and C.-Z. Han
Optimal linear estimation fusion—Part I: Unified fusion rules.
IEEE Transactions on Information Theory, **49**, 9 (Sept. 2003), 2192–2208.

[22] M. E. Liggins, C.-Y. Chong, I. Kadar, M. G. Alford, V. Vannicola and S. Thomopoulos
Distributed fusion architectures and algorithms for target tracking.
Proceedings of the IEEE, **85**, 1 (Jan. 1997), 95–107.

[23] E. Nettleton, H. Durrant-Whyte and S. Sukkarieh
A robust architecture for decentralised data fusion.
In *Proceedings of the International Conference on Advanced Robotics* (ICAR), 2003.

[24] D. Nicholson, C. M. Lloyd, S. J. Julier and J. K. Uhlmann
Scalable distributed data fusion.
In *International Conference on Information Fusion* (Fusion), July 2002, vol. 1, 630–635.

[25] S. Reece and S. Roberts
Robust, low-bandwidth, multi-vehicle mapping.
In *International Conference on Information Fusion* (Fusion), 2005.

[26] M. S. Schlosser
Lineare Schätzer zur multisensoriellen Objektverfolgung in verteilten, nichtlinearen Systemen.
Ph.D. thesis, Institut für Nachrichtentechnik, Universität Karlsruhe, 2006.

[27] J. L. Speyer
Computation and transmission requirements for a decentralized linear-quadratic-gaussian control problem.
IEEE Transactions on Automatic Control, **AC-24**, 2 (Apr. 1979), 266–269.

[28] H. Weiss and J. B. Moore
Improved extended Kalman filter design for passive tracking.
IEEE Transactions on Automatic Control, **AC-25**, 4 (Aug. 1980), 807–811.

[29] K.-S. Zhang, X. R. Li, P. Zhang and H.-F. Li
Unified optimal linear estimation fusion—Part VI: Sensor data compression.
In *Proceedings of the International Conference on Information Fusion* (Fusion), July 2003, 221–228.

Markus Schlosser studied electrical engineering at the University of Karlsruhe, Germany, the Ecole Supérieure d'Ingénieurs en Electronique et Electrotechnique (ESIEE) in Paris, France, and the University of Southampton, United Kingdom. He received the Dipl.-Ing. degree and the Dr.-Ing. degree in electrical engineering from the University of Karlsruhe, Germany, in 2000 and 2005, respectively.

From 2000 to 2005, he was a research assistant in the Communications Group at the University of Karlsruhe. Working on a collaborative project to build a humanoid robot, his research focused on the man-machine interface and in particular the audio-visual attention system as well as audio-visual speaker tracking and identification. In 2005, he joined the Digital Audio Processing Laboratory, Thomson Corporate Research, Hannover, Germany. He is not only engaged in developing advanced audio, but also video processing technologies for production and post-production. His research interests include audio and video processing, data fusion, object tracking, and statistical information theory.

Kristian Kroschel In 1971 he received the Ph.D. degree from the University of Karlsruhe. He studied electrical engineering at the Universities of Karlsruhe and Erlangen-Nürnberg, Germany.

Since 1977 he has been a Professor of Communication Engineering at the Department of Electrical Engineering and Information Technology. From 1987 through 1991 he headed in parallel the group Digital Signal Processing and Diagnosis at the Fraunhofer Institute of Information and Data Processing (IITB) in Karlsruhe. He has held the following positions at the Department of Electrical Engineering and the University of Karlsruhe, respectively: Vice Dean academic year 1982/83, member of the administration board of the university 1984–1992, member of the senate of the university 1986–1990, Student Dean 2000–2002.

On the field of signal processing and communication he has published more than 60 papers and 3 books, partly together with other authors. His special interest is on signal processing of acoustic and video signals applied to man-machine communication. Several times he was a visiting professor, at the Technion Haifa, Israel, the University of California at Santa Barbara, the Bandung Institute of Technology, Indonesia. He is a member of the VDE, the German association of engineers in electrical engineering, electronics, and information technology.

Performance Evaluation for Automated Threat Detection

ROBERT C. SCHRAG
MASAMI TAKIKAWA
PAUL GOGER
JAMES EILBERT

We have developed a performance evaluation laboratory (PE Lab) to assess automated technologies that fuse fragmentary, partial information about individuals' activities to detect modeled terrorist threat individuals, groups, and events whose evidence traces are embedded in a background dominated by evidence from similarly modeled non-threat phenomena. We have developed the PE Lab's main components—a test dataset generator and a hypothesis scorer—to address two key challenges of counter-terrorism threat detection performance evaluation:

- **Acquiring adequate test data to support systematic experimentation; and**
- **Scoring structured hypotheses that reflect modeled threat objects' attribute values and inter-relationships.**

The generator is parameterized so that the threat detection problem's difficulty may be varied along multiple dimensions (*e.g.*, dataset size, signal-to-noise ratio, evidence corruption level). We describe and illustrate, using a case study, our methodology for constraint-based experiment design and non-parametric statistical analysis to identify which among varied dataset characteristics most influence a given technology's performance on a given detection task.

The scorer generalizes metrics (precision, recall, F-value, area under curve) traditional in information retrieval to accommodate partial matching over structured case hypothesis objects with weighted attributes. Threat detection technologies may process time-stamped evidence in either batch, forensic mode (to tender threat event hypotheses retrospectively) or in incremental, warning mode (to tender event hypotheses prospectively—as "alerts"). Alerts present additional scoring issues (*e.g.*, timeliness) for which we discuss practical techniques.

PE Lab technology should be similarly effective for information fusion or situation assessment technologies applied in other domains (besides counter-terrorism), where performance evaluation presents similar challenges.

Manuscript received April 23, 2007; revised October 29, 2008; released for publication February 29, 2008.

Refereeing of this contribution was handled by Pierre Valin.

Authors' addresses: R. Schrag, Global InfoTek, Inc., 1920 Association Dr., Suite 200, Reston, VA 20191, E-mail: (rschrag@globalinfotek.com); M. Takikawa, Art Technology Group, Inc., 673 NW Jackson Ave., Corvallis, OR 97330, E-mail: (takikawa@atg.com); P. Goger, Metron, Inc., 19111 Freedom Dr., Suite 800, Reston, VA 20190, E-mail: (goger@metsci.com); J. Eilbert, AP Technology, LLC, 6 Forrest Central Dr., Titusville, NJ 08560, E-mail: (jleilbert@verizon.com).

R. Schrag and M. Takikawa were employed by Information Extraction and Transport, Inc. (IET) when this work was begun. J. Eilbert was employed by CHI Systems, Inc.

1. INTRODUCTION

Threat detection by sifting high-volume data streams for indicators has been likened to the problem of recognizing a complete "threat" needle by selecting from among many haystack-sized piles of threat and non-threat needle pieces [33]. Under this analogy, problem difficulty may vary depending on factors such as how many stacks there are, how many threat and non-threat needles are distributed among them, and how like are threat and non-threat needles. A key goal in developing a performance evaluation laboratory (PE Lab) is to understand how variation along dimensions like these can affect the performance of a threat detection technology.

As the haystack analogy suggests, many characteristics that contribute to threat detection's difficulty may be modeled simply using convenient abstractions of real-world phenomena. We want to identify well-performing regions of an information fusion approach—*e.g.*, its power to resolve ambiguities arising from partial, potentially corrupted, and temporally overlapping evidence fragments. We deliberately aim to drive the evaluated technology toward explicit representations of and reasoning about structured data and connections between entities and events. Abstraction serves to factor out issues inessential to this, and we model key relationships among threat and non-threat actors, events, and evidence characteristics approximating qualitative real-world relationships and quantitative values. We also factor out user interaction—*e.g.*, evidence visualization and mixed-initiative hypothesis development—so that technology evaluation is in principle entirely automated (although in practice we have not yet required hands-off execution for detection technologies).

We have followed these principles in developing the PE Lab's dataset generator during a multi-year, multi-contractor, multi-agency Government research program, where it has served in several program-wide technology evaluations that have necessitated our development of novel, compatible scoring methods. Many results have already been reported [1] [2] [3] [4] [5] [7] [8] [9] [10] [11] [12] [13] [14] [18] [22] [23] [24] [25] [29] [30] [31] [32] [40] [41] [43] [44].

The PE Lab is schematized in Fig. 1, where square-cornered boxes represent artifacts, round-cornered boxes represent processes, and arrows represent flow of artifacts. The threat detection component—assumed to employ link discovery (LD) technology and also referred to here as an LD component—is rendered 3-dimensionally to indicate its status outside of the PE Lab proper (as the technology under test).

Synthetic dataset generation creates evidence used to challenge LD and (synthetic) ground truth used in scoring LD's hypotheses. LD processes evidence to hypothesize threat phenomena. Generation uses simulation driven by discrete, stochastic event patterns that also are provided to LD. Hypothesis scoring compares technologies' output hypotheses to ground truth.

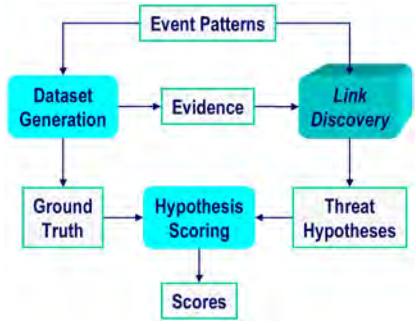


Fig. 1. PE Lab schematic.

In subsequent sections, we describe the following.

- Abstract challenge problem domain (Section 2)
- General hypothesis scoring methods (Section 3)
- Alert scoring methods (Section 4)
- Experiment design to identify performance influences in the problem space (Section 5)
- PE Lab advantages for information fusion system design (Section 6)
- Conclusions (Section 7)

2. COUNTER-TERRORISM THREAT DOMAIN

Synthetic datasets have the advantage for evaluation that (synthetic) ground truth is readily available for scoring. To support unclassified, exploratory counter-terrorism research, we have developed synthetic datasets presenting the same key sources of threat detection difficulty that intelligence analysts have described in real data. In the terminology of the information fusion community [37] [38], real and synthetic datasets present common "referencing" and "registration" problems: *Is the "man in the white shirt" in one report the same "man in a white shirt" described in another report about a different event?* They also present several types of "association" problems: *Are several lower-level events all parts of the same higher-level event, or are people members of the same organization?* Finally, they present "estimation" problems: *Is a group of events that have already been associated really an instance of a particular type of behavior, and if so can upcoming events be predicted based on our model of the behavior?* The issues addressed by PE Lab-based evaluation fall mainly in the Joint Director of Laboratories data fusion model's Level 2, estimation of relationships among entities—or situation assessment [21].

We generate our synthetic datasets over an artificial world that is tunable, mitigating privacy and security classification concerns and supporting systematic experimentation. That our artificial world is also abstract facilitates parameterized overlap between threat and non-threat activities and de-emphasizes knowledge representation and reasoning requirements in comparison to (threat) signal detection requirements, consistent with the funding program's goals. Our synthetic datasets, while thus simplified, present deliberately selected technical challenges.

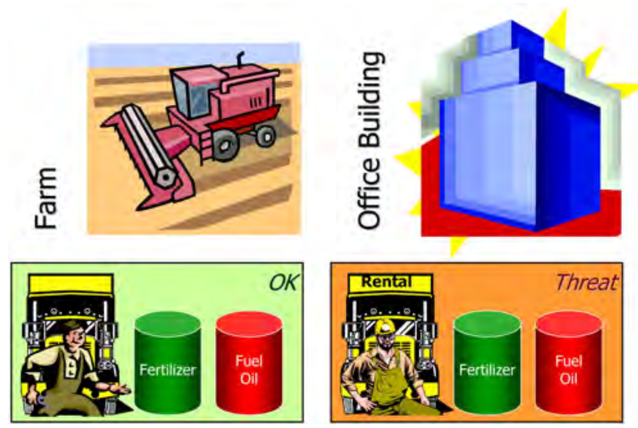


Fig. 2. Real-world motivation for challenge problem.

Fig. 2 exhibits some real-world motivation behind the abstract, artificial world challenge problem domain we have developed. The PE lab's dataset generator uses an artificial world abstraction style inspired by that of Hats [27] [28]. A key difference is that the PE Lab is structured deliberately to emphasize exploratory experimentation, as described in Section 5.

On the left-hand side of Fig. 2, "Farmer Fred" buys fertilizer and fuel oil and transports these *via* truck to his farm. He applies the fertilizer using his tractor which (along with his truck) burns the fuel oil. (Fred is an honest, hard-working man.) On the right-hand side, "Demolition Dan" acquires the same resources but mixes them into a slurry that he transports (*via* rental truck) to the basement of an office building. (Dan is up to no good.)

In the artificial world, capabilities (like farming and demolition) and resources (like fertilizer and fuel oil) are mapped to abstract elements that individuals can possess intrinsically or acquire. Infrastructure elements (like office buildings) are mapped to "targets" that support both legitimate/productive and destructive modes of use or "exploitation." Non-threat and threat individuals (like Fred and Dan) each may belong to any of various groups whose members collaborate in sub-group teams towards different goals. Exploitations play out the general scheme of Fig. 3.

The exploitation scheme on the left-hand side of Fig. 3 includes four main, sequential subevents, each of which unfolds through several levels of task decomposition (illustrated in Fig. 4), bottoming out in transactions with record types indicated on the right-hand side of Fig. 3 and at the bottom of Fig. 4. In a threat exploitation, the final, consummation phase—in which capabilities and resources are (destructively) applied to the target—defines the time by which alerting must occur to be at all effective. Transactions appearing in incrementally presented, time-stamped evidence are the

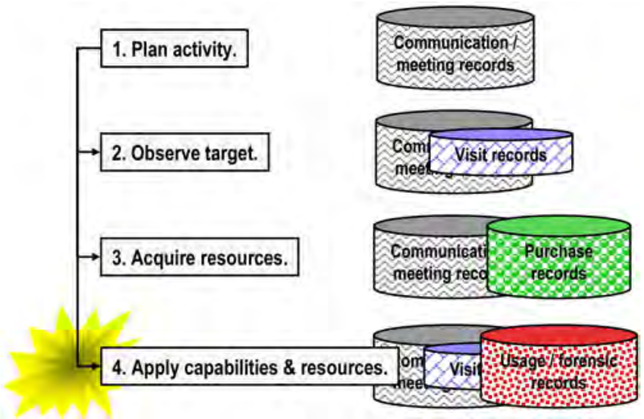


Fig. 3. Generic exploitation scheme.

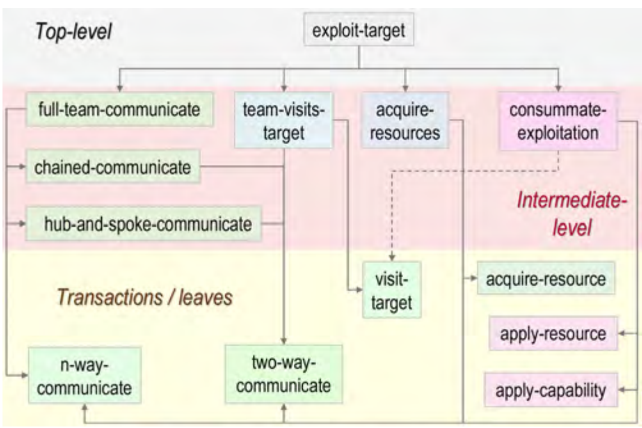


Fig. 4. Invocation relationships among event generation patterns.

sole basis LD has for issuing alerts; intermediate-level events are never materialized in evidence.

In the real world, people typically interact simultaneously in several different social spheres associated with (*e.g.*) work, family, faith, neighborhood, sports/hobbies, civic involvement, shopping, and other relationships. People interact to coordinate times and locations for all of their activities, negotiate inter-activity constraints, and travel as necessary to interact. To make large dataset generation efficient, we have abstracted away such details, modeling all group activities with the same abstraction (the exploitation pattern), allowing individuals to participate in arbitrarily many activities simultaneously, and assuming that all activities take place in a single location (*e.g.*, a metropolitan area).

The challenge to threat detection technology is to identify and report threat *cases*—top-level objects with attributes and values summarizing extant threat phenomena at a level sufficient for scoring. The case types that are LD is tasked to detect include threat actors (groups, individuals, and their aliases) and (ideally, impending) threat events/attacks. To perform this challenge, an automated threat detector is given information about the underlying artificial world that is relatively complete (excepting only a few, novel exploitation modes) and about events and actors that is only partial—per settings of "observability" parameters, as depicted notionally in Fig. 5.

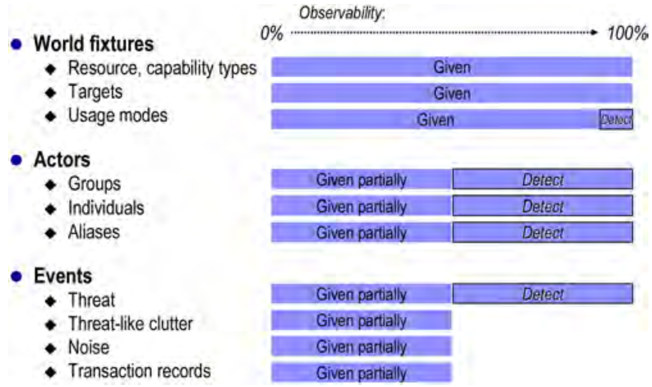


Fig. 5. Threat detection objectives and notional instance observabilities.

We further describe the artificial world problem domain as follows.

- *Individuals* have *assets.*
  —They have permanent *capabilities.*
  —They can acquire consumable *resources* as necessary to exploit a target in one of its modes.
- Both resources and capabilities are abstract enumerations.
- *Exploitation modes* are sets of capabilities and resources.
  —*Vulnerability modes* are exploited by threat actors.
  —*Productivity modes* are exploited by both threat and non-threat actors.
- *Groups* are collections of individuals. Only *threat individuals* belong to *threat groups.* Both threat and *non-threat* individuals can belong to *non-threat groups.* Groups have designated exploitation modes—vulnerability modes for threat groups and productivity modes for both group types. A group can exploit a target that exhibits one of its modes.
- Groups have subgroups—*exploitation teams*—that focus on particular exploitation modes for which a team has qualified members.
- Groups' and teams' member individuals tend to share abstract *social/demographic attributes.*
- *Noise events* masking threat activity occur at several levels. We refer to non-threat exploitations as *clutter. Structured noise* events share intermediate structure with exploitations. *Transaction noise* events are atomic.

In this world, inter-connections abound. Modes overlap with respect to capabilities and resources (as suggested in Fig. 2). Groups overlap with respect to modes, as do targets. Individuals overlap with respect to teams and groups and with respect to capabilities. Exploitations overlap in time with each other and with noise and clutter events. All of these inter-connections contribute to threat detection difficulty.

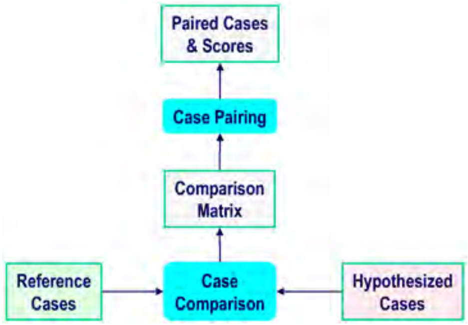


Fig. 6. Generic hypothesis scoring scheme.

3. GENERAL HYPOTHESIS SCORING METHODS

We want to score structured hypotheses that reflect modeled threat objects' attribute values and interrelationships—*e.g.*, a threat event case mentions a threat group case, which includes threat individuals that may be named by their aliases. (Section 3.2 gives full attribute type details.) In this object-oriented context, we need metrics analogous to traditional information fusion's probability of detection and probability of false alarm. For this purpose, we generalize the related recall and precision metrics from traditional information retrieval to accommodate partial matching over structured objects with weighted attributes. Fig. 6 depicts the generic scoring scheme. We require LD to return hypothesis objects that are definite (incorporate neither logical nor probabilistic uncertainty).

In Fig. 6, the reference cases are summaries computed from ground truth, and the hypothesized cases are from LD. Because case objects have significant structure, we want to credit LD when hypothesized cases approximately match reference cases. Match quality is determined by case comparison. When a hypothesized case object's existence has been inferred from lower-level evidence, we can decide which reference case to pair the hypothesized one with only by comparing the hypothesized case with all relevant reference cases—on the basis of their attribute values. We store comparison results for the candidate pairs in a matrix. With inexact matching, it also can be ambiguous which of the one-to-one mappings admitted by the candidate pairs should be selected, so we use an algorithm that optimizes dataset-level scores. Given these pairs, we compute scores for object-oriented metrics based on traditional precision, recall, and F-value metrics.

Subsequent subsections present our scoring methods in more detail, as follows.

- Section 3.1 summarizes the issues of case comparison and case pairing that arise with inexact structured object matching.
- Section 3.2 summarizes the scored object types and attributes in our counter-terrorism domain.
- Section 3.3 presents the algorithmic details of case comparison.

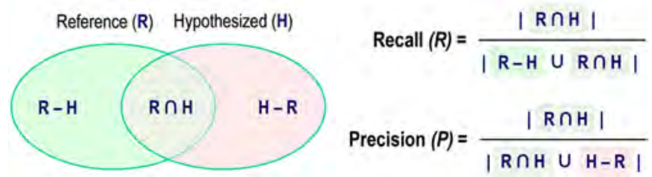


Fig. 7. Traditional precision and recall.

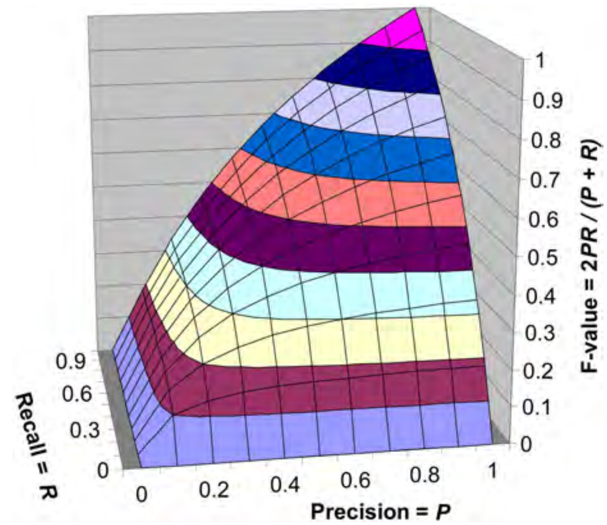


Fig. 8. The F-value surface.

- Section 3.4 describes how we apply the algorithm at the attribute and value level.
- Section 3.5 summarizes additional extant and contemplated hypothesis scoring capabilities.
- Section 3.6 discusses others' work related to our hypothesis scoring approach.

3.1. Case Comparison and Pairing

Case comparison determines the quality of match between any two cases. We characterize this quality by generalizing the traditional precision and recall metrics that presume exact matching between hypothesized and reference items. Fig. 7 illustrates the traditional versions of these metrics.

Traditionally, recall R is the number of valid hypotheses divided by the number of detection targets (the required number of valid hypotheses). Precision P is the number of valid hypotheses divided by the number of all hypotheses.

A single metric to summarize the values of recall and precision is frequently useful, and it is traditional to appeal to F-value = $2PR/(P+R)$—the harmonic mean of precision and recall. (When both precision and recall are zero, we define F-value as zero.)

F-value (shown in Fig. 8 and also known as "F-score" or "F-measure") has the same extremes as a simple average of precision and recall but discounts differences between them (less aggressively than $\min(P,R)$ would discount such differences).

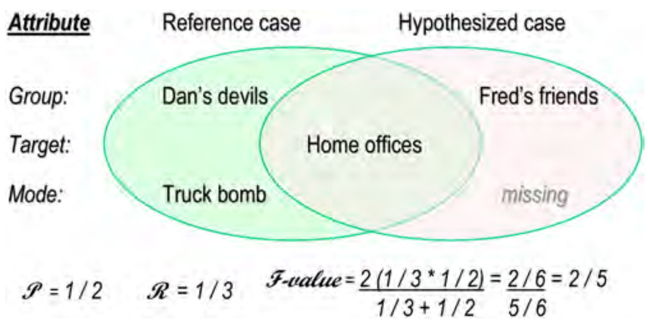


Fig. 9. Object-oriented metrics.

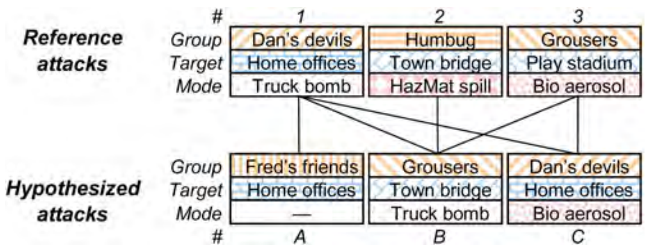


Fig. 10. Case pairing issue.

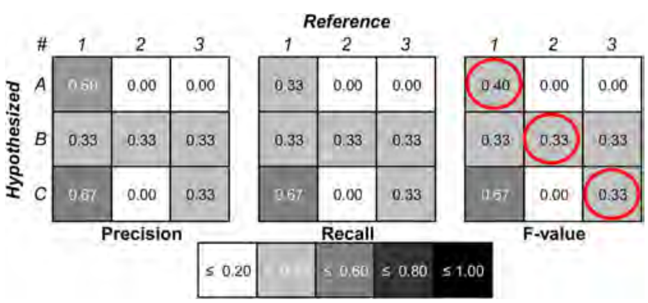


Fig. 11. Case pairing matrix and metrics.

To accommodate inexact matching over structured case objects, we define object-oriented versions of precision, recall, and F-value, as illustrated in Fig. 9. Our complete definitions—in Section 3.3—address object attributes that may be weighted differently, so that attributes contribute to scores non-uniformly.

Of the three attribute values in the reference case of Fig. 9, the hypothesized case agrees only for the Target attribute, so the object-oriented recall score $\mathcal{R}$ is 1/3. Of the two attributes included in the hypothesis, only one agrees with the reference, so the object-oriented precision score P is 1/2. The corresponding object-oriented F-value ($\mathcal{F}$-value) is 2/5, as shown.

Case pairing determines which hypothesized cases to pair with which reference cases—since this may not be obvious, as illustrated in Fig. 10.

In Fig. 10, we have three reference and three hypothesized attack cases. (Reference Case 1 and Hypothesized Case A correspond to the pairing of Fig. 9.) Links appear in the bipartite graph between reference and hypothesized cases wherever these share one or more attributes. Fig. 11 illustrates how we perform one-to-one case pairing using a matrix over all possible pairings.

In Fig. 11, we compute per-pair object-oriented precision, recall, and F-value (as in Fig. 9). Then we use an optimization algorithm to select (red-circled) pairs leading to the greatest average object-oriented F-value. So, we have computed a matching for Fig. 10's bipartite case graph including just the strictly vertical edges.

Case pairing is necessary only for objects whose existence LD has hypothesized based on lower-level evidence, when we require it to invent a unique identifier (UID) in its own namespace. Otherwise LD reports objects' UIDs as they appear in evidence. We forcibly pair any like-UID hypothesized and reference objects, and we omit them from the case pairing matrix.

When the numbers of reference and hypothesized cases do not match, we (effectively) pad the matrix, as necessary to make it square, with null cases. Precision and recall with respect to null cases are defined as zero.

As an optimization algorithm to select a best-scoring one-to-one case pairing, we have often used greedy, heuristic search with a sparse matrix representation that can process thousands of structured hypothesized and reference cases in tens of minutes on conventional hardware. We also have implemented an optimal assignment algorithm [19] that can process hundreds of cases in minutes. The greedy algorithm always selects next the best score in the yet-unselected row or column with the greatest standard deviation among F-values, thus has $O(n^2)$ behavior. In practice, when its F-values differ from those optimally computed, it is only by a few percentage points. The optimal algorithm is $O(n^3)$. For small n (up to about 10,000), both algorithms are dominated by the $O(n^2)$ matrix set-up time. Our current implementation doesn't support non-sparse, square matrices of more than about 5,000 on a side, though, and, as the optimal algorithm does not readily accommodate a sparse matrix representation, we fall back to the greedy algorithm as an alternative. We also are interested in the optimal forward/reverse asymmetric assignment algorithm of Bertsekas and Castañon [6] which is reported to work efficiently with sparse matrices.

To help illustrate how we develop dataset-level metrics, Fig. 12 depicts a somewhat larger, notional case pairing matrix. Again, it holds computed object-oriented F-values for candidate reference-*versus*-hypothesized case pairings.

In Fig. 12, we again have circled entries to optimize average F-value in a one-to-one pairing. Generally, we admit entries to candidacy per a user-specified F-value threshold that must be exceeded for a hypothesis to be deemed adequate for an analyst or other consumer. The larger, red circles apply when all non-zero entries are candidates for pairing (*i.e.*, when the threshold is zero). The smaller, yellow circles apply when the threshold is set at 0.75. Notice that the pairings under the different thresholds are different—each case pairing process considers just the eligible entries. Giving these pairs and object-oriented precision and recall scores, we compute dataset-level precision and recall. Under zero thresholding, the dataset's average object-oriented F-value is

#	1	2	3	4	5	6	7	8	9	10
A	0.23	0.04	0.09	0.96	0.95	0.27	0.60	0.12	0.22	0.69
B	0.30	0.68	0.12	0.26	0.62	0.11	0.37	0.77	0.88	0.57
C	0.95	0.96	0.25	0.12	0.63	0.50	0.53	0.27	0.94	0.89
D	0.85	0.81	1.00	0.47	0.20	0.56	0.37	0.99	0.45	0.80
E	0.62	0.40	0.45	0.65	0.16	0.62	1.00	0.02	0.22	0.06
F	0.53	0.48	0.19	0.04	0.53	0.69	0.57	0.78	0.95	0.12
G	0.63	0.92	0.02	0.50	0.02	0.03	0.89	0.25	0.95	0.73
H	0.61	0.04	0.30	0.12	0.58	0.05	0.76	0.48	0.91	0.33
I	0.21	0.38	0.97	0.46	0.95	0.34	0.92	0.35	0.32	0.17
J	0.16	0.91	0.40	0.48	0.51	0.11	0.61	0.24	0.30	0.59

Fig. 12. Larger, notional case pairing matrix.

#	1	2	3	4	5	6	7	8	9	10
A	0.61	0.52	0.54	0.98	0.98	0.63	0.80	0.50	0.61	0.84
B	0.65	0.84	0.56	0.63	0.81	0.55	0.68	0.89	0.94	0.78
C	0.98	0.98	0.62	0.56	0.82	0.75	0.78	0.63	0.97	0.95
D	0.92	0.91	1.00	0.74	0.60	0.78	0.68	1.00	0.73	0.90
E	0.81	0.70	0.72	0.82	0.58	0.81	1.00	0.51	0.61	0.53
F	0.77	0.75	0.60	0.52	0.76	0.84	0.78	0.89	0.98	0.56
G	0.82	0.96	0.51	0.75	0.51	0.52	0.94	0.63	0.97	0.86
H	0.81	0.52	0.65	0.56	0.79	0.52	0.88	0.74	0.96	0.66
I	0.61	0.69	0.98	0.73	0.98	0.67	0.96	0.67	0.66	0.59
J	0.58	0.96	0.70	0.74	0.76	0.55	0.81	0.62	0.65	0.80

Fig. 13. Notional recall values for Fig. 12's F-values.

0.834, and the traditional metrics aren't particularly informative (always equal one), as all cases (given equal numbers) are paired. Under non-zero thresholding, we always apply traditional metrics given these pairings, so the dataset-level precision, recall, and F-value under the 0.75 threshold all equal 0.8. (A potential expedient would be to develop different higher-thresholded dataset-level scores from a single, zero- or otherwise-lower-thresholded case pairing by dropping any below-threshold pairs.)

To support forthcoming examples, we exhibit, in Fig. 13 and Fig. 14, matrices with notional object-oriented precision and recall values that combine to yield the F-values in Fig. 12. For each cell in the F-value matrix, we have set $\mathcal{R} = (1 + \mathcal{F}\text{-value})/2$ and $\mathcal{P} = \mathcal{R}(\mathcal{F}\text{-value})/(2\mathcal{R} - \mathcal{F}\text{-value})$. We have set the F-value threshold for case pairing at zero and include only the larger, red circles from Fig. 12.

When LD can rank its hypotheses with respect to estimated quality, this ranking supports developing a precision-recall curve and computing the area under the curve (AUC). Any consistently applied variant of precision and recall—*e.g.*, using any consistent F-value threshold—suffices here. Fig. 15 illustrates the AUC for the example values in Fig. 13 and Fig. 14, under two different hypothesis rankings.

At each ith curve point, we compute precision and recall with respect to the full reference case set and the set of LD's 1st- through ith-ranked hypotheses. Fig. 15 notes the hypotheses accounted for by each rectangle contributing to the example's AUC supposing LD ranks its hypotheses in the order (A, B, C, D, E, F, G, H, I, J). Instead of performing full case pairing at each point, we expediently take the case pairings over the full sets of reference and hypothesized cases as authoritative and impose them as we consider each successively presented case to develop the curve.

#	1	2	3	4	5	6	7	8	9	10
A	0.14	0.02	0.05	0.94	0.93	0.17	0.48	0.07	0.13	0.58
B	0.19	0.57	0.07	0.16	0.51	0.06	0.25	0.69	0.83	0.45
C	0.93	0.94	0.15	0.07	0.51	0.37	0.43	0.17	0.91	0.85
D	0.78	0.74	1.00	0.35	0.12	0.44	0.25	0.99	0.33	0.72
E	0.50	0.28	0.32	0.54	0.09	0.50	1.00	0.01	0.14	0.03
F	0.41	0.37	0.12	0.02	0.40	0.58	0.44	0.69	0.93	0.07
G	0.52	0.88	0.01	0.38	0.01	0.02	0.84	0.16	0.93	0.63
H	0.49	0.02	0.20	0.07	0.45	0.02	0.67	0.36	0.88	0.22
I	0.13	0.27	0.95	0.34	0.93	0.22	0.89	0.23	0.21	0.10
J	0.09	0.87	0.28	0.35	0.39	0.06	0.49	0.15	0.19	0.47

Fig. 14. Notional precision values for Fig. 12's F-values.

AUC summarizes LD's performance on a dataset in a way that rewards good rankings. When LD returns lower-quality hypotheses earlier, these drag down incrementally computed scores for later points in the curve as well. When LD returns its hypotheses in the order (E, D, I, C, A, G, F, B, H, J) induced by decreasing F-value scores, the AUC rises to 0.851.

Note that our detection task, where structured threat hypotheses must be developed from lower-level evidence, leads to expectations different from those for the traditional information retrieval task, where every presented item merely must be classified as positive or negative. In the traditional setting, the AUC expected from a strategy of pure guessing is 0.5. In our setting, some reference threats may never be reported, given what-

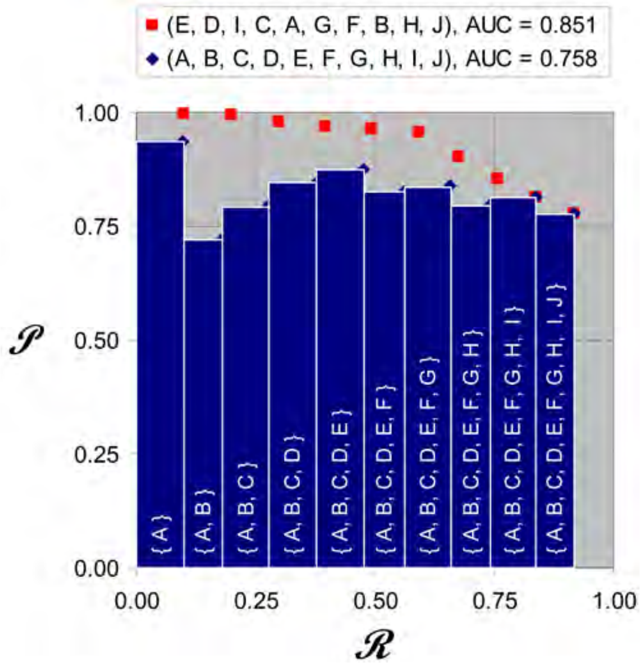


Fig. 15. Precision-recall curve and area for the example values in Fig. 13 and Fig. 14.

ever practical minimum-estimated precision threshold a detector may set. Also, the universe of potential—*e.g.*, syntactically admissible—hypotheses for a given dataset is practically unbounded, rather than limited to presented items as in the traditional case, so that precision scores may be dominated by a practically arbitrary number of false-positive responses (making guessing a practically ineffective strategy).

The size of the information fusion hypothesis space also presents issues for some other metrics commonly used in performance analysis of (binary) classifiers. In particular, the so-called "false-positive rate" used in receiver operator characteristic (ROC) curves and in the calculation of the "detection capability" metric of Gu *et al.* [16][1] assumes a practically enumerated set of "true-negative" responses (*i.e.*, the presented items known in ground truth to be non-threat). True negatives also must be enumerated for the machine learning community's commonly used "accuracy" metric (the number of true—positive and negative—responses divided by the number of all—true and false—responses). Schrag and Takikawa [35] describe analogies between our hypothesis scoring approach and binary classification.

3.2. Scored Object Types and Attributes

Table I presents the types and attributes that are considered during scoring in the counter-terrorism domain, per the artificial world's representation. Along with each attribute is specified its domain, scoring weight (reflecting the challenge problem developer's intuition of an attribute's importance), and reference attribute cardinality (either single or multiple).

TABLE I
Scoring-Relevant Types and Attributes

Scoring-relevant type	Scoring-relevant attribute	Attribute domain	Reference attribute: Weight	Reference attribute: Cardinality
VulnerabilityExploitationCase				
	startingDate	Date (integer)	1	1
	endingDate	Date (integer)	2	1
	minAssetApplicationEndingDate	Date (integer)	2	1
	maxAssetApplicationEndingDate	Date (integer)	2	1
	performedBy	ThreatGroup	3	1
	directingAgent	ThreatIndividualEC	2	1
	deliberateActors	ThreatIndividualEC	1	1+
	targetInExploitation	ExploitationTarget	5	1
	modelnExploitation	VulnerabilityMode	4	1
ThreatGroup				
	exploitsVulnerabilities	VulnerabilityMode	1	1+
	memberAgents	ThreatIndividualEC	1	1+
ThreatIndividualEC				
	hasMember	ThreatIndividual	1	1+
ThreatIndividual				
ExploitationTarget				
VulnerabilityMode				
	modeCapabilities	Capability		1+
	modeResourceTypes	ResourceType		1+
Capability				
ResourceType				

The first three types in Table I are just the scored case types. The remaining types are those that appear as values of scored attributes of cases (*i.e.*, as subcases) or in turn as values of the subcases' attributes.

For each type, each instance also has a unique identifier (UID) by which it may be referred to. An object of class ThreatIndividualEC is used to represent an equivalence class (EC) of threat individual identities, supporting aliases. In attribute values, we interpret any of an EC's member UIDs as denoting the full EC.

Note that an instance of the exploit-target event generation pattern in Fig. 4 is represented in Table I only at the top level (the VulnerabilityExploitationCase threat event type). To simplify scoring, we have engineered this object type to include relevant attributes that it might not explicitly have otherwise—*e.g.*, minAssetApplicationEndingDate, determined by comparing occurrence dates of (lower-level) events associated with the apply-resource and apply-capability patterns.

Note also that event objects have as attribute values objects of other types, some of which also are scored. We rely on the fact that our counter-terrorism domain's supercase type-to-subcase (whole-to-part) type graph (depicted in Fig. 16) is directed and acyclic, as we compute scores for leaf types first and work our way up to root types. (In Fig. 16, only the object types requiring case pairing are shown.)

[1]The authors, working in the domain of computer network intrusion detection, describe an "intrusion detection capability" metric which is in fact applicable in any binary classification setting.

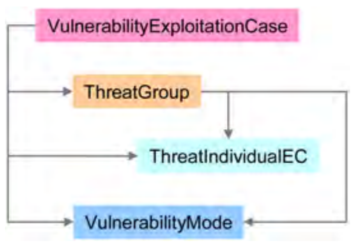


Fig. 16. Counter-terrorism domain supercase type-to-subcase type graph.

3.3. Algorithmic Details of Case Comparison

We compare two like-type cases to determine their object-oriented precision $\mathcal{P}$ and object-oriented recall $\mathcal{R}$, as follows.

We treat a case as a set of assertions regarding the case's attribute values—*e.g.*, (hasMembers Group-4931 Individual-2437). Note that a given case can have multiple, distinct assertions pertaining to a given (multi-valued) attribute. *e.g.*, Group-4931 can have more than one member. Note also that the reference and hypothesized cases can have different numbers of assertions and of attributes, depending on numbers of values per attribute reported by the reference and by the hypothesis.

For each case type, for each defined attribute, a case scoring specification indicates an assertion weight, as summarized in Table I. For a given attribute, the same weights are used for assertions of hypothesized as for those of reference cases.

For a given reference case with the set of assertions $\{r_1, r_2, \ldots, r_m\}$ and corresponding set of weights $\{w_1, w_2, \ldots, w_m\}$, we define the "object-oriented recall basis" $\mathcal{R}_b = \sum_{(i=1\ldots m)} w_i$. (So, each weight is counted once for each assertion in which the attribute appears.) For a given hypothesized case with the set of assertions $\{h_1, h_2, \ldots, h_n\}$ and corresponding set of weights $\{w_1, w_2, \ldots, w_n\}$, we similarly define the "object-oriented precision basis" $\mathcal{P}_b = \sum_{(j=1\ldots n)} w_j$. Note that, for a given comparison of two cases, $\mathcal{R}_b$ and $\mathcal{P}_b$ may differ depending on numbers of values per attribute reported by the reference and by the hypothesis.

We pair reference and hypothesis attribute assertions one-to-one, computing for each pair (r_i, h_i) the following (per the next section's rules for assertion comparison).

- Object-oriented recall $\mathcal{R}(r_i, h_i)$
- Object-oriented precision $\mathcal{P}(h_j, r_j)$

We define the "object-oriented recall contribution" $\mathcal{R}_c$ as the sum over the hypothesized case's assertions of assertion weight w_i pro-rated by the corresponding recall—$\mathcal{R}_c = \sum_{(i=1\ldots n)} \mathcal{R}(r_i, h_i) * w_i$. The "object-oriented precision contribution" $\mathcal{P}_c$ is the sum over the reference case's assertions of assertion weight w_j pro-rated by the corresponding precision—$\mathcal{P}_c = \sum_{(j=1\ldots m)} \mathcal{P}(h_j, r_j) * w_j$.

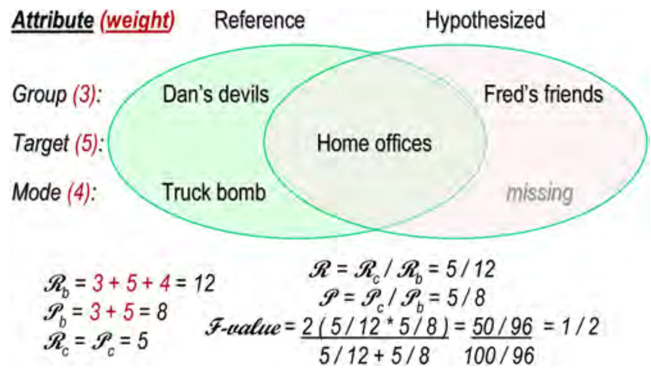


Fig. 17. Object-oriented metrics with non-uniform attribute weights.

For a given pair of reference and hypothesized cases, we define the following.

$$\mathcal{R} = \mathcal{R}_c / \mathcal{R}_b$$

$$\mathcal{P} = \mathcal{P}_c / \mathcal{P}_b$$

$$\mathcal{F}\text{-value} = 2(\mathcal{P} * \mathcal{R}) / (\mathcal{P} + \mathcal{R}).$$

We compute the metrics for a given dataset's cases of a given type as follows. Let N_R be the number of reference cases and N_H the number of hypothesized cases. Let the set $\{p_1, p_2, \ldots, p_o\}$ be the computed pairs, and $\mathcal{R}(p_k), \mathcal{P}(p_k)$ the object-oriented recall and precision (respectively) of the kth pair. Then for the dataset we have the following.

$$\mathcal{R} = \left(\sum_{(k=1\ldots o)} \mathcal{R}(p_k) \right) \Big/ N_R$$

$$\mathcal{P} = \left(\sum_{(k=1\ldots o)} \mathcal{P}(p_k) \right) \Big/ N_H.$$

Fig. 17 adds non-uniform attribute weights to the example of Fig. 9 to illustrate their use in the object-oriented metrics.

The metrics' sensitivities to specified attribute weights depend on the numbers of values for each attribute in compared cases (so, how many times each weight is counted) and—for nested objects—on weights applied in supercases.

3.4. Pairing and Comparison for Attribute Values

We require that paired assertions have the same attribute, so single-valued attributes pair straightforwardly and multi-valued attributes require a one-to-one pairing over their values. In principle, the values may be scalars (of type, *e.g.*, Date) or structured objects (cases or other objects—*e.g.*, Targets) that have UIDs. Per Table I, we have no multi-valued attributes with scalar values in the counter-terrorism domain. Thus, we emphasize here the pairing of multi-valued attribute assertions with nested object values.

We have two alternative general methods for pairing multi-valued attribute assertions.

- Rely on a global one-to-one pairing between all the hypothesized and reference instances of the nested case types for any multi-valued attributes. This has the advantage of consistent pairing across all contexts (because the same hypothesized case is always paired with the same reference case).
- Compute a local one-to-one pairing addressing just the context of a given candidate pair of hypothesized and reference cases and their attribute values. This has the advantage of optimizing per-candidate pair scores—at the expense of global consistency (because different hypothesized cases may be paired with different reference cases in different contexts). Local pairing over nested cases with multi-valued attributes might be prohibitively expensive, but these do not arise in our counter-terrorism domain.

Once pairs have been established over the candidate hypothesized and reference objects' attributes, we can read off each pair's object-oriented precision and recall. We have two alternative general methods of doing this.

- Interpret the computed comparison scores smoothly (accept them at face value)—wherever they fall in the interval $[0,1]$. Smooth comparison reflects the combined matching quality of a given case and all of its nested subcases.
- Interpret the selected pairing crisply, by returning one—for both precision and recall—if the hypothesized object has been paired with a reference object, zero otherwise. Thus, scores fall in the set $\{0,1\}$. Crisp attribute comparison minimizes the impact of inexact matching of nested subcases. The crisp setting matters when the F-value threshold for case pairing is zero. Under non-zero thresholding, comparison is always "crisp" (using traditional metrics, as noted in Section 3.1).

Of the object types with attributes considered during scoring, instances only of ThreatIndividualEC appear as values of multi-valued attributes in other types (VulnerabilityExploitationCase and ThreatGroup). So, in our counter-terrorism domain, only instances of type ThreatIndividualEC may require local case pairing. For these instances, a pair's object-oriented precision and recall depend on how we handle any aliases.

- If both the reference and hypothesized ECs are singleton (perhaps because the dataset does not invoke aliasing), the pair's object-oriented precision and recall are both one if the UIDs match, otherwise both zero.
- Otherwise (at least one of the reference and hypothesized ECs is not singleton), we have two choices.
  —Apply smooth or crisp comparison to the globally or locally computed pairing of reference and hypothesis ThreatIndividualECs.
  —Anti-alias by appeal to the ECs defined in ground truth, then (after discarding any resulting duplicate assertions) score as for singletons. This may be appropriate when LD does not have access to an alias detection capability.

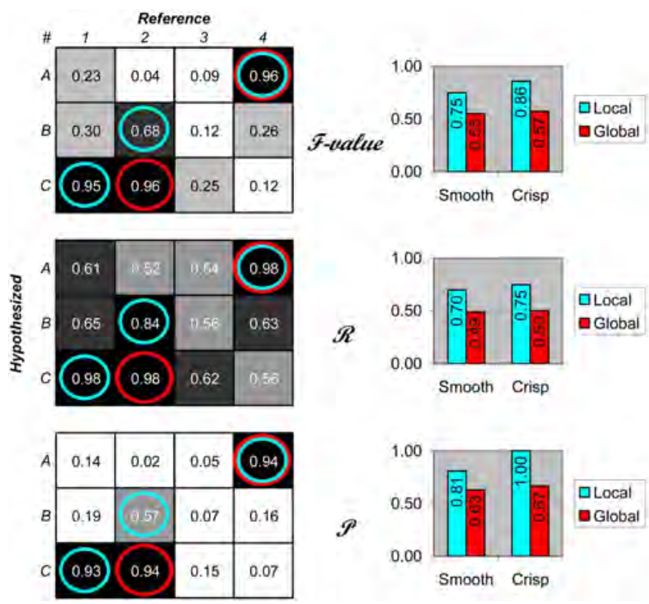


Fig. 18. Alternative pairings (local and global) and interpretations (smooth and crisp) for objects (of type ThreatGroup) with object-valued attributes (of type ThreatIndividualEC).

To illustrate the above concepts, suppose that Fig. 12 compares objects of type ThreatIndividualEC. Suppose that we would like to compare two ThreatGroups and that we have for the present zeroed out the scoring weight for the exploitsVulnerabilities attribute. Then only the memberAgents attribute counts; suppose for the compared groups it has the following values.

- Reference: {EC-1, EC-2, EC-3, EC-4}
- Hypothesized: {EC-A, EC-B, EC-C}

We have reproduced the relevant portions of Fig. 12, Fig. 13, and Fig. 14 in Fig. 18. The global pairing is given by the larger, red circles, the local pairing by the smaller, cyan ones.

Of the types in Table I, the following either have no attributes or have no attributes that are considered during scoring.

- Instances of type Date are represented by integers. The object-oriented precision and recall for a pair of reference and hypothesized Dates are both (identically) defined in terms of their normalized temporal distance, as illustrated in Fig. 19. (The "ratio" parameter is computed with respect to a nominal distance specified for the given attribute.)
- The types Target, Capability, and ResourceType have no scored attributes. If a pair's UIDs are the same, the object-oriented precision and recall are both one, otherwise both zero.
- For the type VulnerabilityMode, we require strict set equivalence. For each of the two multi-valued attributes, if a pair's respective sets of attribute values

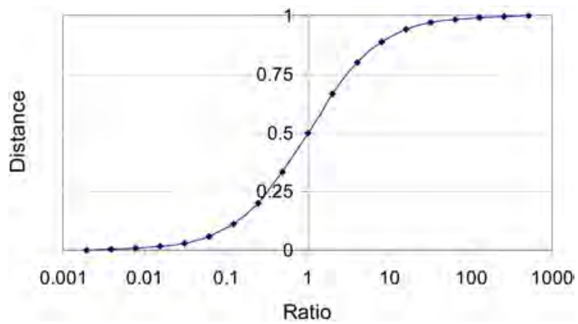


Fig. 19. A normalization function for Date distances.

have the same members, the pair's object-oriented precision and recall are both one, otherwise both zero.

3.5. Additional Hypothesis Scoring Capabilities

The PE Lab also implements many-to-one and many-to-many case pairing methods that—unlike their one-to-one counterpart—do not necessarily penalize LD for submitting multiple competing hypotheses. We have considered (but not yet implemented) support for logical and probabilistic uncertainty. One practical approach may be for LD to submit disjunctive hypotheses with each disjunct's probability noted. Entire disjunctions would then be paired one-to-one with reference cases.

Because we include some (partially described, potentially corrupted) case objects in evidence, we would like a scoring method to factor this information out of reported scores. We score evidence's case objects directly (treating them as hypotheses), to establish a baseline, as described in Section 5.4. To obtain more diagnostic value, we are contemplating a refinement to develop separate hypothesis scores regarding reference case content that is:

- Correct in evidence;
- Clearly incorrect (corrupted) in evidence;
- Ambiguous in evidence; or
- Omitted from evidence.

Our hypothesis scorer's many parameters (only some of which have been mentioned here) support customized experimentation. For example, we have on occasion (as in Section 3.4's illustration) tailored case scoring specifications to zero out the weights of attributes that a given technology does not address. Another parameter lets us count the weight of each multi-valued attribute only once for a given object instance, rather than counting the attribute's weight time it appears in the object. Counting just once is appropriate when attributes' relative imports are roughly independent of their per-instance cardinalities, which agreed with our intuition when we applied the overall PE lab approach in the computer network intrusion detection information fusion domain [34].

We are considering the following alternative treatment of temporal information (*e.g.*, events' endpoints)

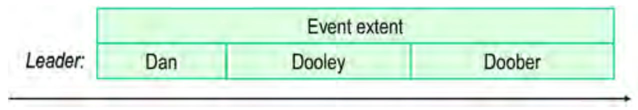


Fig. 20. Ground truth representation of a time-varying attribute.

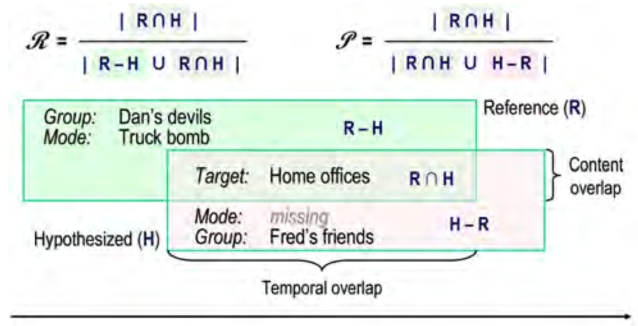


Fig. 21. Temporal object-oriented recall and precision.

to accommodate scored objects with time-varying properties that may be included in future versions of the artificial world or that may arise in other application domains—*e.g.*, groups with time-varying membership (memberAgents) or events during which an exploitation's leader (directingAgent) role may alternate among team members. Instead of treating an event's endpoints as individual scalar attributes (startingDate and endingDate), we would record, for each object attribute with temporal extent, a set of contiguous temporal intervals over which its potentially different values hold. In ground truth, for each single-valued attribute there would be a set of disjoint, adjacent intervals that together cover the object's extent, as in Fig. 20.

Fig. 21 suggests how we would compute temporal object-oriented precision and recall for the reference and hypothesized threat events shown in Fig. 9. (For simplicity, Fig. 21 omits the weights of Fig. 17. The fact that our current threat events have static attributes also simplifies the illustration, in that we require only a single temporal interval per event.)

Note that scores for these temporal versions of object-oriented recall and precision are zero whenever the intervals do not overlap, which means we can omit any non-overlapping pairs from the case pairing matrix without excluding any pairs with non-zero comparison scores.

Under some circumstances, it might make sense to compare temporal objects with attributes summarized at different levels of temporal abstraction; thus two objects' coarse time structures could be seen to match relatively well even when their fine time structures did not.

3.6. Related Work (Hypothesis Scoring)

Our object-oriented metrics may be compared with other metrics of inexact matching, such as the graph edit distance metric used in LAW [44]. Some of the key differences between this approach and ours are:

- Our strong object orientation *versus* their accommodation of arbitrary relationship graphs;
- Our separate tracking of false-positive and false-negative discrepancies *versus* their uniform distance tracking; and
- Their accommodation of ontological distances between typed nodes that we haven't required (because our ontology of case- and scoring-relevant attribute types in Table I includes no subtype relationships).

Our reliance on a directed acyclic graph of subcase types means that (given global pairing of multi-valued attributes) we compare two cases of a given type only once, regardless of how interconnected the different-type hypotheses may be. It also allows us to deprecate the fine differences between nested cases in comparison to the coarser differences at the supercase level. They also take advantage of graphs' hierarchical structure and cache computed subgraph distances locally, but not globally as we do. Ghallager [15] surveys additional graph-based pattern matching metrics.

Working in the computer network intrusion detection domain, Tadda *et al.* [39] have adopted our unweighted object-oriented metrics and our one-to-one case pairing (with recall, rather than F-value, used in case comparison matrices) and have developed some original metrics—inspired by metrics in the field of target tracking—to summarize a resulting case pairing matrix. This is for a single-level case structure where each attribute value is assumed to appear in no more than one hypothesis.

To support scoring in the network intrusion detection PE Lab, Schrag and Takikawa [34] developed object-oriented precision and recall scores at different levels by defining different case types corresponding to different abstractions (information-reducing mappings) of a full-information case type accommodating hypotheses from the technology under test. This requires separate case pairings for the different case types but supports scoring the abstractions using common metrics. Precision for the most-reduced case type, a bag of evidence, corresponds to the "track purity" metric used in target tracking.

Mahoney *et al.* [26] describe scoring for a two-level military situation hypothesis that includes a joint probability distribution (*e.g.*, a Bayesian network) regarding component objects' existence and attribute values. They compute (un-normalized) attribute value distances at the situation component level, accounting for hypothesized probabilities and applying attribute weights. At the situation level, they invoke a distance threshold to qualify potential matches of hypothesized and reference components, then develop all possible sets of one-to-one pairs and estimate the likelihood of the observed distance given that the pair is a correct match and given that the pair does not match. By aggregating the likelihoods computed for a given pair across the sets, they estimate the overall likelihood for each pair. It is not clear whether this work could be generalized easily to address more deeply nested hypotheses.

4. ALERT SCORING METHODS

The scoring methods of Section 3 allow us to compare, for a given scored object type, a static set of hypothesized objects against a static set of reference objects. They thus are appropriate for scoring a batch of event hypotheses tendered retrospectively/forensically. By themselves, though, they are inadequate for scoring *alerts*—event hypotheses tendered prospectively, for warning, when LD incrementally processes time-stamped event records in evidence. Here, we describe methods that additionally account for alert scoring's dynamics—whether the alerts in effect during reference events are good hypotheses and whether they can be used to support effective warning.

Schrag *et al.* [36] describe an approach that was implemented late in the development of PE Lab and suggest an alternative practical approach and an associated (deemed-impractical) idealized approach. All three of these variants rely on specified costs of false-positive and false-negative reporting that are applied uniformly over a portion of the reference event's temporal extent. As explained further in Section 4.2, they also conflate evaluation of the quality of a prospectively tendered event hypothesis and of the decision about whether or not to tender it, in that they presume some mitigating action will be attempted for all tendered alerts. Recognizing the potential value of pure hypothesis quality evaluation and of more sophisticated models for cost and response, we here sketch two alternative, complementary approaches.

- Section 4.1 describes a cost-free approach based on the temporal variant of object-oriented precision and recall described in Section 3.5.
- Section 4.2 describes an alert-free, cost-based approach that excludes precision and recall because it doesn't even require the technology under test to present alerts, rather merely to invoke response actions when it considers this advantageous with reference to a furnished cost model.
- Section 4.3 describes related work.

4.1. Cost-free Alert Scoring

As LD advances incrementally through time-stamped evidence, it must examine only the events that are reported (with respect to simulation time) either at or before its current processing time. LD may submit alerts at any (simulation) time, with any frequency it chooses. Finer incremental processing time intervals may incur greater overall processing time but also may afford more opportunity to detect impending threat events and issue alerts sooner. Coarser intervals may not pick up evidence regarding threat events until after they have been consummated—and so miss the opportunity for alerting. Requiring alerts at fixed simulation-time frequencies would entail similar issues. Requiring alerts at specified simulation times (*e.g.*, near threat event consummations) might engender gaming by LD.

LD may supersede an alert tendered earlier in simulation time with one tendered later, or it may retract an alert without superseding it. We call an alert that has been tendered and that has been neither superseded nor retracted an "active" alert. The temporal scores we've described in Section 3.5 for retrospectively tendered event hypotheses can be construed as applying to the hypotheses active at the end of simulation time. For any given simulation timepoint, we could make a similar comparison of the active alerts—either to the full set of reference events or just to those that have started. If we averaged the resulting scores across timepoints, though, earlier events that were considered in more per-timepoint scores would have disproportionate effect. We propose instead to consider all simulation timepoints simultaneously, admitting an alert as a candidate for pairing with a reference event based on its activity status (*i.e.*, whether it is active or not) at an anchor point in the reference when it might reasonably last be considered useful—the first time one of its capability or resource assets is applied to the target (noted in the attribute minAssetApplicationEndingDate) and after which we may consider the attack's success to be inexorable.

While we have thus shifted the focus from individual simulation timepoints to the reference events themselves, we'd like to go a bit further and reflect how the quality of alerting has evolved over the course of each reference event (*i.e.*, not just at one anchor point). Given unlimited computing power, we might compute the per-anchor point scores using as anchors all earlier points in the reference event (*i.e.*, considering just the alerts active at each point). While this may be feasible when there are few alerts, for more general practicality we must limit our invocation of the expensive case pairing operation. We propose to do so by taking the case pairing that is computed for the reference cases' minAssetApplicationEndingDates as authoritative and by chaining backward from the alerts paired there to any alerts they have superseded.[2] Finally, we can take the average values we have computed for temporal precision and recall (or for temporal F-value) over each reference case.

In the example of Fig. 22, Alert 3 tendered at Simulation Time 3 supersedes Alert 2 tendered at Time 2. Alert 4 (not shown) is tendered past Anchor Point A

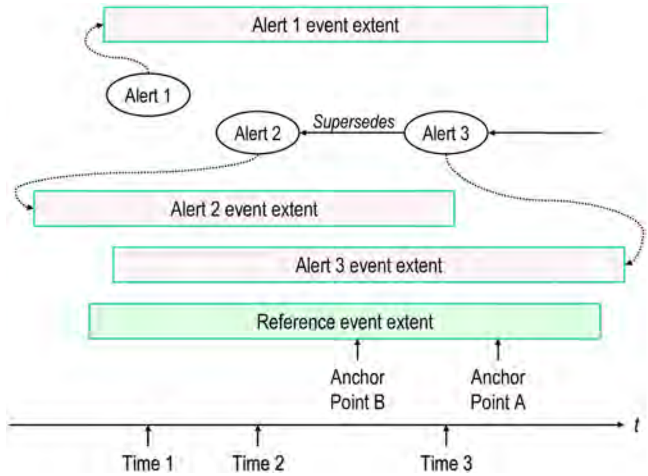


Fig. 22. Alerts in a supersession chain.

corresponding to the minAssetApplicationEndingDate, so is not eligible for comparison to this reference event. Both Alert 1 and Alert 3 are active with respect to Anchor Point A; Alert 2, the better match, is paired, and the alert supersession chain it heads becomes authoritative over the rest of the reference event. Even though Alert 1 is a better match at Anchor Point B than Alert 2, Alert 1 is not considered. The overall score computed for this reference event thus is based on Alert 3 from Time 3 to the event's end, on Alert 2 from Time 2 to Time 3, and on no alert before Time 2.

Under this treatment of supersession chains, LD should supersede one alert with another when it believes they apply to the same event. Otherwise, it should retract the earlier alert and start a new chain. We require supersession chains to be non-branching. Backward branching would introduce ambiguity as we chain backward. Forward branching might make two reference intervals end up at the same alert—if the heads of their respective chains were paired with different reference cases.

4.2. Alert-free Response Action Scoring

In the earlier approaches [36], we specified costs for false-positive and false-negative predictions of an attack on a target, indicating (respectively) the costs of actions taken in response to a false prediction (*e.g.*, escalated protection, evacuation) and of inaction resulting from a non-prediction (destruction of target, loss of life). We applied these costs uniformly over the portion of the reference event's temporal extent preceding the minAssetApplicationEndingDate and discounted them to reward good hypotheses—discounting the false-positive cost by object-oriented precision and the false-negative cost by object-oriented recall. See Fig. 23, whose dark-shaded fraction corresponds to the discount and light-shaded fraction to the assessed cost.

In the real world, the costs of action and of inaction depend on complex interactions. Even in our artificial world, it's hard to tell a consistent story (in terms

[2]The earlier practical and implemented approaches (Schrag *et al.* 2006) chained backward over supersession links established as authoritative by pairing reference cases with retrospectively tendered hypotheses. While this affords more perspective for LD, it might (especially since some reference events become visible in evidence after they are complete) be considered artificial.

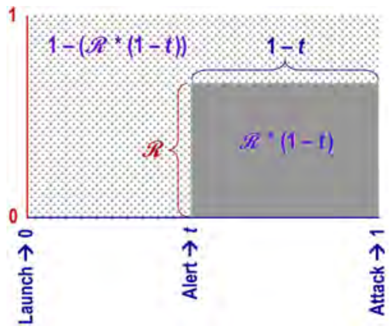


Fig. 23. Discounting uniform false-negative cost by object-oriented recall (in earlier approaches).

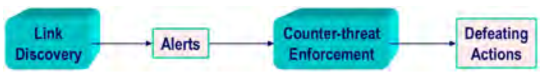


Fig. 24. Processes in automated threat defeat.

of actions and effects) about why false-negative and false-positive costs should be applied symmetrically and uniformly over time.

More fundamentally, our funding program's scope was limited to an LD component tendering threat hypotheses (see Fig. 1), not a "counter-threat enforcement" component issuing actions to interdict threat agents or to defeat their attacks (see Fig. 24). Imposing an action interpretation on LD's prospective event hypotheses effectively required our evaluation participants to conflate enforcement with LD and to make decisions about whether to issue alerts at all.

In hindsight, we recommend that cost-based scores be developed following an approach closer to that taken by Morrison *et al.* [27] in the Hats simulator, where an agent with threat detection capability (like that of LD) acts as a player in a game. The player agent must act on whatever threat event hypotheses it develops to interdict suspected or known threat actors before they inflict damage. Performance is determined by the game's final score that accounts for incurred costs associated with things like surveillance, damage, and false arrest. Because the player agent is included in the Hats simulation, its actions can have downstream effects as rich as the simulation supports. For example, an incarcerated agent will not participate in future attacks.

The simulator currently underlying PE Lab's dataset generator supports no such enforcement actions. One reason we wanted datasets that could be processed entirely off-line was to compare easily performance results for a given dataset across different technologies under test, without the technologies' different sets of enforcement actions resulting in different simulation histories (effectively making the datasets different and the technologies' scores incomparable). The closest we can come to Hats' style of cost-based performance evaluation and still maintain dataset-based comparability is to admit "hypothetical" actions and effects that do not actually manifest in simulation history. Instead, we propose to infer tendered actions' (determined or likely) effects and (determined or expected) costs according to a model that we furnish. Such a hypothetical action approach also might be used to evaluate an enforcement component's performance against real-world historical datasets, which inherently require off-line processing.

TABLE II
Attack Effects and Eosts

Attack Effect	One-time cost
Good individual dies.	High
Target is destroyed.	Very High

TABLE III
Hypothetical Counter-Threat Enforcement Actions, Effects, and Costs

Enforcement Action	Unit-time cost	Latency	Effect
Escalate target security.	Low	Short	CE can detain, evacuate.
Detain individual.	Medium	Nil	*Role-dependent…*
Exploitation team leader:			Exploitation is abandoned.
Half of exploitation team members:			Exploitation is abandoned.
Evacuate target.	High	Medium	Nobody good at target dies.

To illustrate how this approach might be realized in our artificial world, Table II suggests some notional one-time costs to be associated with a successful attack. Table III suggests continuing (per-time unit) costs that a counter-threat enforcement (CE) component could incur in its actions to defeat attacks.

Table III also suggests time constraints. Details follow.

- Upon invoking the action to escalate security at a target, there is a short "security escalation period" (latency). Thereafter, CE can detain (effectively incarcerate) individuals at that target and can evacuate it.
- The action to evacuate a target must be invoked a medium time before the attack's last resource or capability is applied (maxAssetApplicationEndingDate) to avoid the deaths of visitors (everyone whose last visit was to the target, including attackers') when the target is destroyed.
- CE may choose to detain (*e.g.*) suspected threat individuals or any individual it can determine may possess a resource or capability supporting a suspected-impending threat mode. The exploitation team leader or at least half its members must be detained throughout the asset application interval (from minAssetApplicationEndingDate to maxAssetApplicationEndingDate) to defeat the attack. We could (as in Hats) accord greater cost to detaining a non-threat individual.

Fig. 25 illustrates CE's hypothetical actions to defeat a specific attack, in a scenario where CE is able to detain three attackers but doesn't have confidence that it can thwart the attack. It decides to evacuate and is able to do so in time.

Fig. 26 illustrates (not to scale) cumulative costs associated with the enforcement actions and with the attack in Fig. 25. In the described scenario, costs in the

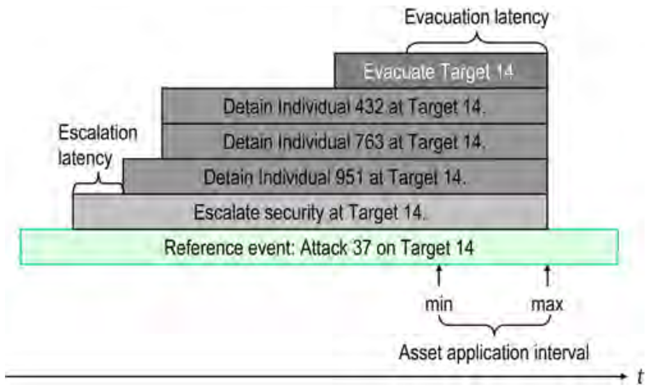


Fig. 25. Actions to defeat a specific attack.

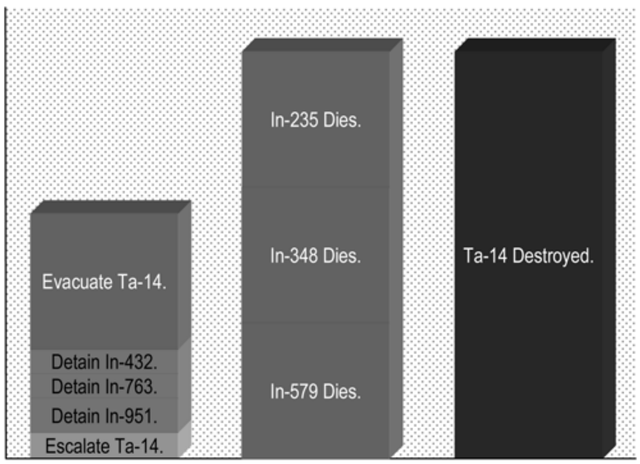


Fig. 26. Costs of hypothetical enforcement (left), enforcement-defeated loss of life (center), and attack damage (right).

center column are averted; those in the left and right columns are incurred.

Precise scalar costs would need to be tuned relative to each other to ensure that datasets would pose reasonable challenges. *e.g.*, we generally would want CE to incur a higher cost for incarcerating everyone it ever sees than it incurs for doing nothing. A limited cost budget (*e.g.*, one that would support only a limited number of concurrent individual detentions or of total target evacuation days) also could help to ensure plausible CE threat-defeating strategies.

To facilitate comparison across datasets, scores may be normalized in each dataset to the cost that would be incurred if CE were to do nothing (and all attacks were successful).

4.3. Related Work (Alert Scoring)

Besides Hats, related work appears to be limited. Others have evaluated event prediction where exact hypothesized-to-reference case matching is appropriate. Weiss and Hirsh [42] specialize precision to discount temporally close false-positive hypotheses. Létourneau, Famili, and Matwin [20] apply a nonmonotonic timeliness function to determine rewards and penalties for true- and false-positive predictions of appropriate aircraft component replacement times. Different metrics are certainly appropriate in different contexts, and we believe the accommodation of inexactly matching hypothesized and reference cases and attendant case pairing entail issues for structured threat alert scoring that others have not addressed. Mahoney *et al.* [26] suggest some overall strategies for comparing hypothesized *versus* reference situation histories and note the requirement for timeliness, without directly addressing threat event prediction.

TABLE IV
Coarse Problem Space Dimensions

Group Connectivity	How many groups an individual belongs to
Noise, Clutter	How much threat masking
Dataset Size	How many observable transactions
Population Size	How many individuals
Pattern Complexity	Minimalistic *vs.* richer threat event modeling
Observability	How likely observations are
Corruption	How corrupted observations are
Aliasing	How frequently aliases are used
Event Confusability	How like are threat and non-threat activities
Target Duty Level	How busy targets are
Individual Duty Level	How busy individuals are

5. EXPERIMENT DESIGN TO IDENTIFY PERFORMANCE INFLUENCES IN THE PROBLEM SPACE

We now describe our methodology for constraint-based experiment design and analysis to identify which among varied dataset characteristics most influence a given technology's performance on a given detection task. We illustrate this methodology with a case study using object-oriented F-value as the performance metric of interest. We describe our experimental approach, summarized as follows, in subsequent subsections.

1. Collapse many (fine) problem space parameters into a few dimensions with discrete (coarse) difficulty settings (Section 5.1).
2. Specify a mix of experimental datasets that maximizes diversity over the difficulty settings (Section 5.2).
3. Exercise participating detection technology configurations over datasets in the mix (Section 5.3).
4. Score technologies' output hypotheses relative to an baseline derived from evidence (Section 5.4).
5. Determine the statistical significance of apparent problem space performance influences by technology and detection objective (Section 5.5).

Section 5.6 describes some related experimental design work.

5.1. Problem Space Discretization

The coarse problem space dimensions are summarized in Table IV.

Each coarse dimension corresponds to one or more fine parameters. For some dimensions, we discretize the fine parameters based on quantitative annual or

TABLE V
Fine Parameter Discretizations by Problem Difficulty

Group Connectivity	None		Easy		Fair		Hard	
Individual status:	Threat	Non-threat	Threat	Non-threat	Threat	Non-threat	Threat	Non-threat
Mean groups per individual	1	1	2	4	4	6	6	10
Dev. groups per individual	0	0	1	2	2	3	3	4

TABLE VI
Fine Parameter Discretizations per Annual Performance Goals

Population Size	Y1	Y2.5	Y3
Number of individuals	~1,000	~10,000	~100,000
Mean threat group membership	20	80	80
Dev. threat group membership	5	20	20
Number of capabilities	50	100	150
Number of resources	50	100	150

Dataset Size	Y1	Y2.5	Y3
Number of observable transactions	N / A	~100,000	~1,000,000

Noise, Clutter	Y1	Y2.5	Y3
Threat-to-clutter event ratio	0.08	0.008	0.0008
Structured event SNR	0.08	0.008	0.0008
Transaction event SNR	0.08	0.008	0.0008
Individual SNR	0.4	0.08	0.008
Group SNR	0.8	0.16	0.016

TABLE VII
Cross-Parameter Constraints

Noise, Clutter	Dataset Size	Population Size	Pattern Complexity	Rationale for prohibited combinations of settings listed in rows
	Y2.5	Y3		People doing too few things during a simulation
Y3	Y2.5			Too few threat events (maybe none)
Y1	Y3		Thin	Too many threat events to score practically
Y3	Y3		Fat	Too few threat events (maybe none)
Y2.5	Y2.5		Fat	
	Y3	Y2.5	Fat	Too many time ticks for incremental threat detection to be viable (currently)
	Y3	Y1	Fat	
	Y2.5	Y1	Fat	

semi-annual performance goals (set by the funding program)—see Table VI. For other dimensions we chose to explore, we discretize into difficulty settings such as Easy, Fair, Hard—see Table V for an example. We apply a stop-light color-coding over the discretized settings, adding light green for very easy settings and dark red for very hard.

5.2. Dataset Mix Specification

Several factors make effective experimentation challenging in this context. The evaluation dataset mix is scoped to occupy a few solid weeks of coordinated program effort. Processing is not always hands-off, with several disparate component developers sometimes manually handling intermediate results within a single technology configuration. A star-shaped experimental design with fixed baseline settings and single-dimension departures might serve individual technology configurations with single detection objectives, but—with each dataset—we must test multiple configurations over multiple objectives. What's easier for one technology/objective combination might be harder for another. At evaluation time, we have somewhat sparse prior performance data from dry-run activities. We need an experiment that effectively tests over multiple baselines simultaneously, so we choose a diversity-maximizing, fractional factorial design.

We take the following steps, discussed below, to maximize diversity.

1. Specify cross-dimension settings constraints that ensure well dataset generation.

2. Perform constraint satisfaction to develop an initial dataset mix.

3. Perturb the initial mix in hill-climbing to optimize the experiment's coverage.

Table VII indicates some prohibited coarse setting combinations and associated rationale. *e.g.*, the Fat setting (corresponding to rich threat event modeling—resulting in more atomic transactions per threat event) results in too few threat events when the signal-to-noise ratio used is too low for the dataset size.

Other combinations of coarse settings over these dimensions have been verified to generate well datasets. The coarse discretizations themselves assure compatibilities at the fine parameter level. For example, the various signal-to-noise (threat-to-non-threat) ratios (SNRs) for a given coarse setting in Table VI are coordinated so that there are enough individuals to satisfy the generator's minimum group size requirement. The discretization process thus factors out such fine, numerical constraints (whose violation would raise run-time exceptions), so that coarse constraint satisfaction over symbolic domains is sufficient for the dataset mix specification/experiment design.

The constraint satisfaction problem is challenging in that we want a number of dataset specifications that draw without replacement from settings pools, fixed for each dimension, until all the pools are exhausted. The pool for Group Connectivity, *e.g.*, includes six instances each for the tokens None, Easy, Fair, and Hard. We have implemented an algorithm to specify a dataset mix respecting both the constraints and the pools. Alternatively, if the exact numbers of tokens drawn from each pool is not critical, we can draw with replacement to generate a random dataset specification and discard this if it does not satisfy constraints (or if it is a duplicate), until we have enough datasets.

With an initial mix in hand, we perform a hill-climbing random walk over the space of well datasets, swapping any two datasets' like settings along a given dimension whenever this decreases the maximum number of like settings shared across all datasets.

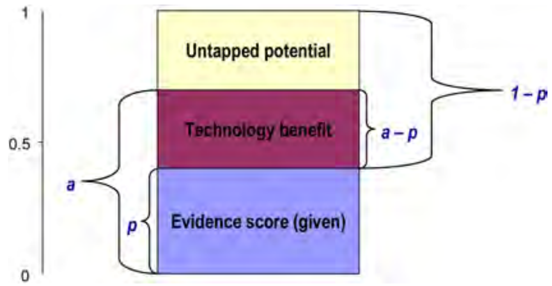


Fig. 27. Relative scoring $r = (a - p)/(1 - p)$.

5.3. Detection Technology Exercise

Technology developers receive the test datasets in database form and are required to return ranked hypotheses in the scorer's input format for each of the detection objectives noted in Fig. 5.

5.4. Adjusting Hypothesis Scores Relative to an Evidence Baseline

To compare dimension influences across different datasets requires comparable scores. As explained below, our default ("absolute") scoring method credits hypothesis content that is patently manifest in datasets to different extents. Comparability requires a ("relative") scoring method that factors this content out.

Evidence provided to LD (as illustrated in Fig. 5) includes partial top-level case descriptions for some instances of the detection object types (threat event, group, individual, and alias association). These descriptions, notionally corresponding to a legacy intelligence database, afford starting places for the detection process. The completeness, consistency, and transparency of these descriptions with respect to ground truth depend on settings for the Observability, Corruption, and Alias dimensions. In absolute scoring, LD gets credit for reporting detection objects whether the same information appears in evidence or not.

In relative scoring, the detection task may be reinterpreted as, "Find unknown and correct misreported threat objects and their attribute values." Let a stand for LD's absolute score, and let p stand for the score for returning exactly all and only the top-level threat case content provided in evidence. We use p as a baseline in computing the relative score $r = (a - p)/(1 - p)$. See Fig. 27.

Note that r can be negative—if LD does not perform as well as the baseline. Note also that the relative score rewards LD for any improvements to top-level threat case content provided in evidence—for supplying missing attribute values or correcting corrupted ones.

5.5. Identifying Performance Influences

Because of the coarse discretizations and constraints, our experiment design must be "unbalanced" (*i.e.*, have unequal numbers of settings within and across dimensions). This requires us to invent novel techniques to identify performance-influencing dataset characteristics, rather than, *e.g.*, applying ANOVA over coefficient means among regression fits.

TABLE VIII
Ranked Settings Significance Testing

Observed	Number	Number	Ideal	Distance
Easy	1	1	Easy	0
Easy	2	2	Easy	0
Hard	1	3	Easy	1
Easy	3	4	Easy	1
Easy	4	5	Easy	1
Easy	5	1	Hard	3
Covert	1	2	Hard	1
Hard	2	3	Hard	1
Hard	3	4	Hard	1
Hard	4	5	Hard	1
Hard	5	6	Hard	1
Hard	6	7	Hard	2
Covert	2	8	Hard	3
Hard	7	9	Hard	3
Covert	3	1	Covert	8
Hard	8	2	Covert	3
Hard	9	3	Covert	2
Significance	0.0081	$= a/(a+b)$	Total	32

Distance of observed ordering

a b

0 2 ... 14 30... ...62 ... 126 ... ... 256 256 256

Relative scores support ranking experimental datasets by LD's performance for a given objective. Under this ranking, we expect the settings for a dataset dimension with significant performance influence to tend to exhibit the expected difficulty order—*e.g.*, "Easy, Fair, Hard" or "Y1, Y2.5, Y3." To determine the significance of the settings order actually observed, we first compute its distance to the expected, or "ideal," order, as illustrated in Table VIII. We first number the tokens for each setting (*e.g.*, "Hard") consecutively as they appear in each of the observed and ideal orders. Then, for each so-numbered token, we compute the distance between its ranks in the two orders. Finally, we sum the rank distances for all the setting tokens—yielding in the example an aggregate distance of 32.[3]

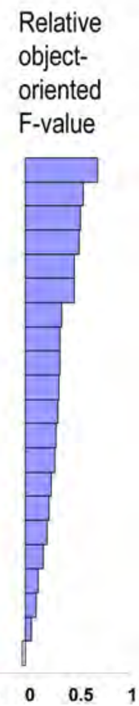
To determine the extent to which the observed order is significant with respect to the ideal—the extent to which the observed could have arisen strictly by chance, with lower values indicating greater significance—we similarly compute distances (represented in the abbreviated vector at the bottom of Table VIII) to the ideal from a sufficient number $N = a + b$ of randomly generated token orders, counting the number of times a the observed order is at least as close and reporting significance as a/N. The significance computation thus accounts both

[3]This example is taken from an experiment earlier than that reflected in Table IX. Here, the Observability dimension is discretized into just three settings: Easy, Hard, Covert.

TABLE IX
Performance Influence Case Study (Group Detection)

Dataset Number	Group Connectivity	Noise, Clutter	Dataset Size	Population Size	Observable 2-way-comms per Individual	Pattern Complexity	Observability	Corruption	Aliasing	Confusability	Target Duty Level	Individual Duty Level
	None, Easy, Fair, Hard	Y3, Y2.5, Y1	Y3, Y2.5	Y1, Y2.5, Y3		Thin, Fat	Perfect, Easy, Fair, Hard, Y2.5, Covert	None, Fair, Hard	None, Easy, Fair, Hard	Easy, Fair, Hard	Easy, Fair, Hard	Easy, Fair, Hard
25	None	Y3	Y3	Y2.5	170.77	Thin	Fair	None	None	Easy	Easy	Easy
27	None	Y3	Y3	Y1	1982.75	Thin	Y2.5	Hard	None	Easy	Easy	Hard
EP	Easy	Y1	Y2.5	Y1	127.80	Thin	Easy	None	None	Easy	Easy	Easy
24	None	Y2.5	Y3	Y3	13.41	Fat	Hard	None	None	Easy	Easy	Easy
20	None	Y2.5	Y2.5	Y2.5	16.35	Thin	Perfect	None	Easy	Easy	Fair	Fair
38	Hard	Y3	Y3	Y1	1818.23	Thin	Perfect	None	None	Easy	Hard	Hard
28	None	Y1	Y2.5	Y2.5	14.47	Fat	Fair	Fair	None	Hard	Hard	Fair
18	Easy	Y2.5	Y2.5	Y1	136.69	Thin	Perfect	None	Fair	Fair	Hard	Easy
30	Easy	Y3	Y3	Y2.5	175.22	Thin	Easy	Hard	Easy	Fair	Hard	Easy
35	Fair	Y3	Y3	Y1	1787.58	Thin	Covert	Hard	None	Hard	Fair	Easy
39	Hard	Y3	Y3	Y1	1890.39	Thin	Fair	Fair	None	Fair	Fair	Fair
40	Easy	Y1	Y2.5	Y2.5	12.22	Fat	Fair	None	None	Easy	Fair	Hard
36	Easy	Y1	Y2.5	Y2.5	12.37	Fat	Hard	Fair	None	Easy	Easy	Fair
26	Fair	Y2.5	Y3	Y3	17.77	Fat	Easy	Hard	None	Hard	Hard	Easy
23	Hard	Y2.5	Y3	Y3	14.65	Fat	Y2.5	Hard	None	Easy	Easy	Fair
22	Hard	Y2.5	Y3	Y3	15.71	Fat	Hard	Fair	None	Easy	Hard	Hard
19	Fair	Y2.5	Y2.5	Y2.5	14.89	Thin	Hard	Fair	None	Fair	Easy	Hard
37	Fair	Y1	Y2.5	Y2.5	12.46	Fat	Covert	Hard	None	Hard	Fair	Fair
29	Hard	Y1	Y2.5	Y2.5	12.67	Fat	Covert	None	None	Hard	Easy	Easy
32	Fair	Y1	Y2.5	Y2.5	13.11	Fat	Easy	Fair	Hard	Fair	Fair	Easy
31	Hard	Y1	Y2.5	Y2.5	10.68	Fat	Hard	Hard	Fair	Fair	Fair	Hard

Significance level: 1, 0.1, 0.01, 0.001, 0.0001



for the closeness of the observed order to the ideal and for variability of settings among the datasets.

By way of a case study, we include Table IX, covering results for a selected technology configuration [2] with the group detection objective, (to provide membership lists for all of the threat groups). Table IX covers an additional dimension (not included in Table IV) relevant to the technology configuration: Observed 2-way-comms per Individual. The 21 datasets processed using the selected technology are sorted by group detection performance (noted lower right).[4] Each dataset dimension column is headed by an idealized settings order. Under the dimension name, significance is plotted on a log scale.

With a scoring option in effect to resolve aliases automatically from ground truth, Group Connectivity is the most significant influence: chance probability = 0.0006. (Without this option, Aliasing is.) We split the dataset mix along this dimension to continue analysis, with results shown below.

[4]The experiment reflected in Table IX included 24 datasets developed from the pool specifications discussed in Section 5.2. The results were developed in the context of a technology integration experiment; a different group detection technology was used to process the three datasets omitted from Table IX.

- Group Connectivity (GC) *at 0.0006 significance*:
  —GC = None: *(No dimension of convincing significance)*
  —GC = Easy: *(No dimension of convincing significance)*
  —GC = Fair or Hard: Observed 2-way-comms per Individual *at 0.0005 significance*

Both Group Connectivity and Observed 2-way-comms per Individual are relevant to group detection intuitively as well as in the group detector's implementation.

5.6. Related Work (Experiment Design)

Hoffman and Jameson [17] present a multi-sensor, multi-target geospatial tracking testbed with a multi-dimensional dataset generation facility to explore the performance boundaries of a particular data fusion system implementation. They identify dimensions of problem complexity (or solution difficulty), generate datasets with parameter values varying along the different dimensions (apparently following the same kind of star-shaped experimental design we discussed in Section 5.2), and determine limits of acceptable performance for the fusion system and subsystems using tra-

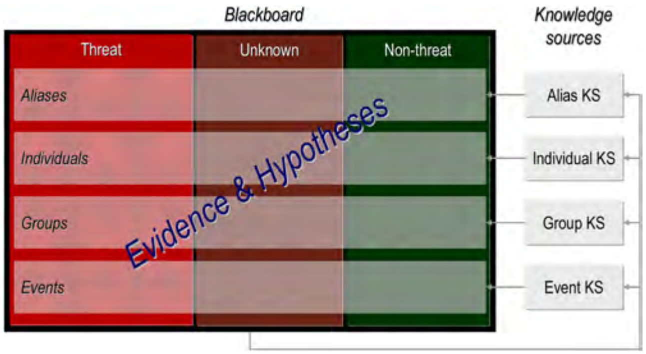


Fig. 28. Blackboard-based component integration.

ditional precision and recall along with other tracking metrics. Our work differs from theirs in several ways.

- Our problem focuses on inference of higher-level activities from discrete transaction evidence.
- Our evaluation applies object-oriented metrics to structured hypotheses.
- Our experimentation employs a fractional factorial design to differentiate performance of multiple solution implementations.
- Our analysis uses a novel rank correlation test to determine the problem dimensions most affecting a given technology's performance.

6. PE LAB ADVANTAGES FOR INFORMATION FUSION SYSTEM DESIGN

The overall PE Lab supports advanced threat detection technology development in several ways.

As reported here, we assess technical progress through program-wide evaluation and identify particular problem characteristics most influential to a technology's performance. Besides assisting individual technologists, this process can identify alternative technologies' relative strengths and elucidate potentially advantageous combinations.

Within a functional architecture (such as the blackboard architecture schematized in Fig. 28), we can employ the PE Lab to validate assumptions about the performance of a downstream component (or blackboard knowledge source—KS) based on that of an upstream one.

Suppose, *e.g.*, that a group detector depends on an alias resolver to deliver sufficiently de-aliased evidence about individuals. If the resolver is not yet performing at a goal level meeting the detector's input specs, we can still ascertain validity of performance claims for the latter by stubbing the former with a direct feed of evidence having per-spec de-aliasing. This can help to pinpoint performance gaps among functional components early in the development process.

In the future, we hope to facilitate such exploratory experimentation *via* a PE Lab-based component test harness and a program-wide commitment to automated (*i.e.*, hands-off) component execution. This has the potential to institutionalize the evaluation/experimentation process as a near-continuous loop in which experiments result in performance feedback to technology developers and developers respond to performance deficits with updated component versions. It also would enhance opportunities for large-scale experimentation.

7. CONCLUSIONS

Our hypothesis scoring methods are applicable in principle to performance evaluation in any domain where technologies return instances of one or more structured object types, given a problem for which an answer key is available. We expect that our alert scoring methods may be applied with benefit in other information fusion domains where hypothesis timeliness is important. Our experimental design methods may benefit other fusion (especially situation assessment) applications during exploratory system design and development.

PE Lab datasets and documentation are currently available to U.S. Government-approved users. Documentation covers concept of operations, event generation pattern language and counter-terrorism domain patterns, dataset generation algorithms, ontology, database schema, case scoring specifications, hypothesis format, and user instructions for the hypothesis scoring and dataset generation software.

ACKNOWLEDGMENTS

We thank our colleagues who have contributed to the PE Lab's overall design, implementation, and application, notably (with apologies to anyone we overlook) Jaffar Adibi, Chad Bisk, Chris Boner, Dan Bostwick, Hans Chalupsky, Thayne Coffman, Hans Dettmar, Seth Greenblatt, Ian Harrison, Dan Hunter, David Jensen, Kathy Laskey, Sherry Marcus, Andrew Moore, Kendra Moore, Peter Olson, David Page, Nick Pioch, Mike Pool, Josh Powers, Eric Rickard, Ben Rode, Jeff Schneider, Ted Senator, Andrei Sobolevskii, Sam Steingold, Afzal Upal, Andre Valente, Didier Vergamini, Bob Washburn, and Ed Wright. We additionally thank Randy Dieterle, Brandon Goldfedder, Tod Levitt, and anonymous reviewers for their constructive comments.

REFERENCES

[1] J. Adibi, T. Barrett, S. Bhatt, H. Chalupsky, J. Chame and M. Hall
Processing-in-memory technology for knowledge discovery algorithms.
In *Proceedings of the Second International Workshop on Data Management on New Hardware* (DaMoN 2006), June 2006, http://www.isi.edu/~hans/publications/DAMON06.pdf.

[2] J. Adibi and H. Chalupsky
Scalable group detection via a mutual information model.
In *Proceedings of the First International Conference on Intelligence Analysis* (IA-2005), https://analysis.mitre.org/proceedings/Final_Papers_Files/355_Camera_Ready_Paper.pdf.

[3] J. Adibi, H. Chalupsky, E. Melz and A. Valente
The KOJAK group finder: Connecting the dots via integrated knowledge-based and statistical reasoning.
In *Proceedings of the Sixteenth Innovative Applications of Artificial Intelligence Conference* (IAAI-04), 2004, http://www.isi.edu/∼hans/publications/IAAI04.pdf.

[4] J. Adibi, P. R. Cohen and C. T. Morrison
Measuring confidence intervals in link discovery: A bootstrap approach.
In *Workshop on Link Analysis and Group Detection* (LinkKDD2004), held at the Tenth Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining, Aug. 2004, http://crue.isi.edu/files/papers/adibi_2006_1141857733.pdf.

[5] D. W. Aha, J. W. Murdock and L. A. Breslow
Managing terrorist activity hypotheses.
Technical Report AIC-03-188, Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence, Intelligent Decision Aids Group, Washington, D.C., 2003.

[6] D. P. Bertsekas and D. A. Castañon
A forward/reverse auction algorithm for asymmetric assignment problems.
Technical Report LIDS-P-2159, Massachusetts Institute of Technology, 1993, http://citeseer.ist.psu.edu/bertsekas93 forwardreverse.html.

[7] C. Boner
Automated detection of terrorist activities through link discovery within massive databases.
In Proceedings of the *American Association for Artificial Intelligence (AAAI) Spring Symposium on A.I. Technologies for Homeland Security*, Palo Alto, CA, Mar. 2005.

[8] C. M. Boner
Novel, complementary technologies for detecting threat activities within massive amounts of transactional data.
In *Proceedings of the First International Conference on Intelligence Analysis* (IA-2005), https://analysis.mitre.org/proceedings/Final_Papers_Files/ 318_Camera_Ready_Paper.pdf.

[9] J. Coble, D. Cook and L. Holder
Structure discovery in sequentially-connected data streams.
International Journal on Artificial Intelligence Tools, **15**, 6 (Dec. 2006), http://www.eecs.wsu.edu/∼holder/papers/CobleIJAIT06.pdf.

[10] J. Coble, R. Rathi, D. Cook and L. Holder
Iterative structure discovery in graph-based data.
International Journal on Artificial Intelligence Tools, **14**, 1–2 (Feb.–Mar. 2005), http://www.eecs.wsu.edu/∼holder/papers/CobleIJAIT05.pdf.

[11] D. Davenport and C. R. Hill
Fast abductive reasoning over ontologies.
In *American Association for Artificial Intelligence (AAAI) Fall Symposium on Capturing and Using Patterns for Evidence Detection*, 2006, http://demo.cs.brandeis.edu/papers/author.html#viswanathan.

[12] J. Davis, V. Santos Costa, I. M. Ong, D. Page and I. Dutra
Using Bayesian classifiers to combine rules.
In *3rd Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Workshop on Multi-Relational Data Mining (MRDM 2004)*, 2004, http://www.cs.wisc.edu/∼dpage/nbayes.pdf.

[13] J. Davis, I. Dutra, D. Page and V. Santos Costa
Establishing identity equivalence in multi-relational domains.
In *Proceedings of the First International Conference on Intelligence Analysis* (IA-2005), 2005, https://analysis.mitre.org/proceedings/Final_Papers_Files/ 378_Camera_Ready_Paper.pdf.

[14] J. Eilbert, J. Hicinbothom and A. Karnavat
A cognitive architecture focusing on situation estimation: merging perceptual and cognitive processes.
In *Behavior Representation in Modeling and Simulation* (BRIMS), 2005.

[15] B. Gallagher
Matching structure and semantics: A survey on graph-based pattern matching.
In *American Association for Artificial Intelligence (AAAI) Fall Symposium on Capturing and Using Patterns for Evidence Detection*, 2006.

[16] G. Gu, P. Fogla, D. Dagon, W. Lee and B. Škoríc
Measuring intrusion detection capability: An information-theoretic approach.
In *Proceedings of ACM Symposium on InformAtion, Computer and Communications Security* (ASIACCS'06), 2006, http://www-static.cc.gatech.edu/∼guofei/paper/Gu_asiaccs06_cid.pdf.

[17] M. Hoffman and S. Jameson
Complexity and performance assessment for data fusion systems.
Lockheed Martin Advanced Technology Laboratories, 1998, http://www.atl.external.lmco.com/overview/papers/920.pdf.

[18] L. Holder, D. Cook, J. Coble and M. Mukherjee
Graph-based relational learning with application to security.
Fundamenta Informaticae special issue on Mining Graphs, Trees and Sequences, **66**, 1–2 (Mar. 2005), 83–101, http://www.eecs.wsu.edu/∼holder/papers/HolderFI05.pdf.

[19] R. Jonker and A. Volgenant
A shortest augmenting path algorithm for dense and sparse linear assignment problems.
Computing. **38**, 4 (Dec. 1987), 325–340.

[20] S. Létourneau, F. Famili and S. Matwin
Data mining for prediction of aircraft component replacement.
In *IEEE Intelligent Systems* special issue on Data Mining, (Dec. 1999), 59–66, http://www.site.uottawa.ca/∼stan/papers/1999/componentFailureIEEE.pdf.

[21] J. Llinas, C. Bowman, G. L. Rogova, A. Steinberg, E. Waltz and F. White
Revisiting the JDL data fusion model II.
In *Proceedings of the Seventh International Conference on Information Fusion* (Fusion 2004), June 28–July 1, 2004, Stockholm, Sweden, 1218–1230, http://www.fusion2004.foi.se/papers/IF04-1218.pdf.

[22] S. A. Macskassy and F. Provost
A simple relational classifier.
In *2nd Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Workshop on Multi-Relational Data Mining (MRDM 2003)*, http://www-ai.ijs.si/SasoDzeroski/MRDM2003/proceedings/macskassy.pdf.

[23] S. A. Macskassy and F. Provost
Suspicion scoring based on guilt-by-association, collective inference, and focused data access.
In *Proceedings of the International Conference on Intelligence Analysis*, 2005, https://analysis.mitre.org/proceedings/Final_Papers_Files/ 273_Camera_Ready_Paper.pdf.

[24] S. A. Macskassy, and F. Provost
Suspicion scoring based on guilt-by-association, collective inference, and focused data access.
In *Proceedings of the North American Association for Computational Social and Organization Sciences (NAACSOS) Conference*, 2005, http://www.research.rutgers.edu/∼sofmac/paper/naacsos2005-app/ macskassy-naacsos2005-app.pdf.

[25] S. A. Macskassy and F. Provost
A brief survey of machine learning methods for classification in networked data and an application to suspicion scoring.
In *Workshop on Statistical Network Learning* (poster) at *23rd International Conference on Machine Learning*, 2006, http://www.research.rutgers.edu/∼sofmac/paper/sna2006/abstract.html.

[26] S. Mahoney, K. Laskey, E. Wright and K. Ng
Measuring performance for situation assessment.
In *Proceedings of the Military Sensing Symposia (MSS) National Symposium on Sensor and Data Fusion*, San Antonio, TX, 2000, http://ite.gmu.edu/∼klaskey/papers/Mahoney_et_al_Fusion2000.pdf.

[27] C. Morrison, P. Cohen, G. King, J. Moody and A. Hannon
Simulating terrorist threat in the Hats simulator.
In *Proceedings of the First International Conference on Intelligence Analysis*, MITRE Corp., 2005, https://analysis.mitre.org/proceedings/Final_Papers_Files/ 265_Camera_Ready_Paper.pdf.

[28] C. T. Morrison and P. R. Cohen
The Hats information fusion challenge problem.
In *Proceedings of the 9th International Conference on Information Fusion* (Fusion 2006), special session on Distributed Inference and Decision-Making in Multisensor Systems, 2006, http://crue.isi.edu/files/papers/clayton_2006_1151624442.pdf.

[29] M. Mukherjee and L. Holder
Graph-based data mining for social network analysis.
In *Workshop on Link Analysis and Group Detection* (LinkKDD2004), held at the *Tenth Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining*, Aug. 2004, http://www.eecs.wsu.edu/∼holder/papers/HolderFI05.pdf.

[30] Patrick Pantel
Alias detection in malicious environments.
In *American Association for Artificial Intelligence (AAAI) Fall Symposium on Capturing and Using Patterns for Evidence Detection*, 2006, http://www.patrickpantel.com/cgi-bin/Web/Tools/ getfile.pl?type=paper&id=2006/aaai-fs06.pdf.

[31] N. J. Pioch, D. Hunter, J. V. White, A. Kao, D. Bostwick and E. K. Jones
Multi-hypothesis abductive reasoning for link discovery.
In *Workshop on Link Analysis and Group Detection* (LinkKDD2004), held at the *Tenth Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining*, Aug. 2004, www.cs.cmu.edu/∼dunja/LinkKDD2004/Nicholas-Pioch-LinkKDD-2004.pdf.

[32] J. Potts, L. Holder, D. Cook and J. Coble
Learning concepts from intelligence data embedded in a supervised graph.
In *Proceedings of the International Conference on Intelligence Analysis*, May 2005, http://www.eecs.wsu.edu/∼holder/papers/PottsIA05.pdf.

[33] T. Senator
Evidence Extraction and Link Discovery program.
In *DARPATech 2002 Symposium*, 2002, http://www.darpa.mil/DARPATech2002/presentations/iao_pdf/speeches/SENATOR.pdf.

[34] R. C. Schrag and M. Takikawa
Cyber situation assessment evaluation participant's guide.
Information Extraction and Transport, Inc., July 2006.

[35] R. C. Schrag and M. Takikawa
Scoring hypotheses from threat detection technologies: Analogies to machine learning evaluation.
In *American Association for Artificial Intelligence (AAAI) Workshop on Evaluation Methods for Machine Learning II*, 2007, http://www.globalinfotek.com/schrag/Schrag-Takikawa-AAAI-2007-WS04-color.pdf.

[36] R. C. Schrag, M. Takikawa, P. Goger and J. Eilbert
Scoring alerts from threat detection technologies.
In *American Association for Artificial Intelligence (AAAI) Fall Symposium on Capturing and Using Patterns for Evidence Detection*, 2006, http://www.globalinfotek.com/schrag/Schrag-et-al-AAAI-Fall-2006.pdf.

[37] A. N. Steinberg
Unification across data fusion levels 1–3.
In *Proceedings, Seventh International Conference on Information Fusion*, Stockholm, Sweden, 2004.

[38] A. N. Steinberg
Open networks: Generalized multi-sensor characterization.
In *Proceedings of the Ninth International Conference on Information Fusion*, Florence, Italy, 2006.

[39] G. Tadda, J. Salerno, D. Boulware, M. Hinman and S. Gorton
Realizing situation awareness in a cyber environment.
In Belur V. Dasarathy (Ed.), *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006*, *Proceedings of SPIE* vol. 6242, 2006, 624204-(1–8).

[40] J. Thomere, I. Harrison, J. Lowrance, A. Rodriguez, E. Ruspini E and M. Wolverton
Helping intelligence analysts detect threats in overflowing, changing and incomplete information.
In *Proceedings of the 2004 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, July 2004, 39–45, http://www.ai.sri.com/pubs/files/1042.pdf.

[41] J. E. Tierno
Performance analysis and prediction for data mining systems.
In *American Association for Artificial Intelligence* (AAAI), Spring Symposium A.I. Technologies for Homeland Security, Palo Alto, CA, Mar. 2005.

[42] G. Weiss and H. Hirsh
Learning to predict rare events in event sequences.
In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 1998, 359–363, http://www.cs.rutgers.edu/∼gweiss/papers/ts-workshop.pdf.

[43] J. V. White and C. G. Fournelle
Threat detection for improved link discovery.
In *Proceedings of the International Conference on Intelligence Analysis*, May 2005, https://analysis.mitre.org/proceedings/Final_Papers_Files/ 333_Camera_Ready_Paper.pdf.

[44] M. Wolverton and J. Thomere
The role of higher-order constructs in the inexact matching of semantic graphs.
In D. Mladenic, N. Milic-Frayling and M. Grobelnik (Eds.), *Link Analysis: Papers from the 2005 American Association for Artificial Intelligence (AAAI) Workshop*, technical report WS-05-07, 1–7, http://www.ai.sri.com/pubs/files/1146.pdf.

Robert Carl Schrag was born at Griffiss Air Force Base in Rome, NY on December 3, 1957, the son of Vernon Dale Schrag and the former Marilyn Louise German. After high school in Endicott, NY, he attended the State University of New York at Binghamton (now Binghamton University) and the Massachusetts Institute of Technology and received a B.A. in natural sciences from The Johns Hopkins University in 1979.

In 1980, Schrag joined a new artificial intelligence research group at Rome Air Development Center (RADC—now part of the Air Force Research Lab) at Griffiss. While at RADC, Schrag monitored contracts and performed in-house research in knowledge-based systems and logic programming. He received an M.S. in computer and information sciences from Syracuse University in 1983 and in 1984 joined VOIS, Inc.—a small company researching speech recognition and psychoacoustics in Endicott. In 1985, he returned to logic-based artificial intelligence, joining the Honeywell Systems and Research Center in Minneapolis, MN.

Schrag led his first sponsored research effort at Honeywell, developing temporal reasoning technology in collaboration with Tom Dean of Brown University, with funding from the Defense Advanced Research Projects Agency (DARPA) and RADC. Having enjoyed the collegiality of this experience, he decided to return to school in 1992. At the University of Texas at Austin, he studied computer science, beginning under Ben Kuipers and finishing under Dan Miranker. He also spent the summer of 1993 at AT&T Bell Telephone Laboratories in Murray Hill, NJ, working with James Crawford. His 1996 Ph.D. dissertation, entitled "Search in SAT/CSP: Phase Transitions, Abstraction, and Compilation," focused on propositional satisfiability (SAT) and the finite-domain constraint satisfaction problem (CSP).

In 1996, Schrag joined Information Extraction and Transport, Inc. (IET), of Arlington, VA, where he developed and conducted multi-participant evaluations for six different artificial intelligence research programs funded by DARPA (High Performance Knowledge Bases, Rapid Knowledge Formation, Evidence Extraction and Link Discovery) and other U.S. Government agencies. At IET, he also led an effort to build a multiple-context facility for DARPA's UltraLog military logistics agent-based planning and scheduling environment.

In 2007, Schrag joined Global InfoTek, Inc., of Reston, VA, where he is a principal scientist working on applications of technology to U.S. national security problems.

Robert Carl Schrag was born again (in Jesus Christ) on December 3, 1986. In 1990, he covenanted marriage with the former Robin Sue Maxwell of Montevideo, MN. At this writing, they have two children and reside in Fairfax County, VA.

Masami Takikawa received his B.A. and M.A. in behavioral science from Hokkaido University in 1986 and 1988, respectively, and his M.S. and Ph.D. in computer science from Oregon State University in 1992 and 1998, respectively.

He is currently Architect of On-Demand Personalization at Art Technology Group (ATG), Inc., working on the application of relational real-time Bayesian inference technology to e-commerce. From 1997 to 2007, Dr. Takikawa was principal engineer at Information Extraction & Transport (IET), Inc., specializing in the development and application of various AI technologies to data fusion, situation assessment, and link discovery for government and industry.

Paul Goger graduated from William and Mary in 2001 with an undergraduate degree in mathematics and computer science. In his last year, he completed his senior thesis with Professor Michael Trosset in stochastic optimization. After college, Paul was hired by Metron, Inc., Reston, VA.

His early areas of work involved data fusion applied to the problem of anti-submarine warfare. For the last several years he has been working in the realm of counterterrorism.

James L. Eilbert earned a Ph.D. in biomathematics from North Carolina State University, Raleigh, NC (1980), and a M.S. in applied mathematics from Courant Institute, New York University, New York, NY (1975).

He is currently CEO of AP Technology, LLC in Titusville, NJ. He has twenty-five years experience in research and development focusing on cognitive modeling and computer vision. He has been principal investigator on projects that developed techniques for information fusion and situation assessment, cognitive modeling of intelligence analysts, modeling vision with a focus on the interaction of perception and cognition, modeling context and context switching, and automatic target recognition. He has applied statistical techniques and neural networks to the analysis of medical, robotic, and satellite imagery. He is currently working on biologically plausible models of attention and emotion, and their role in situation estimation and activity recognition.

Target Engageability Improvement through Adaptive Tracking

ABDER REZAK BENASKEUR
FRANÇOIS RHÉAUME
STÉPHANE PARADIS

This paper addresses the joint problem of target engageability assessment and engageability improvement in naval Anti-Air Warfare operations. An integrated approach that aims to minimize the detect-to-engage sequence is proposed. It uses an estimation of the search-to-lock-on time of the fire control radar to evaluate the engageability of targets. The latter is then improved through the control of tracking operations. Weapons assignment process and the resulting engagement plan are adjusted based on the results of both the assessment and the improvement of the engageability. A quantitative evaluation of the proposed approach was performed using a simulation and performance evaluation environment developed at Defence Research and Development Canada–Valcartier. Although simple sensors and weapons models used in the presented work, encouraging results were obtained with scenarios involving generic supersonic Anti-Ship Missiles. In such scenarios, the proposed adaptive tracking strategy was able to provide timely engagements compared to a conventional engagement strategy.

Manuscript received March 28, 2007; revised October 15 and March 26, 2008; released for publication May 13, 2008.

Authors' adresses: Defence R&D Canada–Valcartier, 2459 Blvd. Pie XI North, Quebec (Qc), G3J 1X5, Canada.

1. INTRODUCTION

Reaction time of current and future naval warships is eroded since they are expected to operate in a large variety of situations with constantly increasing complexity. To cope with diverse air and surface threats, the warships, either operating in a single ship configuration or within a task group, will require their combat power resources to be efficiently managed. The coordination and tight integration during the deployment of these resources will also be required. Decision support aids can help in overcoming the inherent complexity of the naval Command and Control (C^2) process and the underlying combat resource management problem [15].

This paper addresses two problems related to target engageability in naval warfare operations: engageability assessment and engageability improvement. Engageability is defined as the feasibility of engagement actions against designated targets. Engageability assessment is concerned with the evaluation of the feasibility of engagement actions based on the involved combat resources, the environmental condition, and the geometry of the engagement. The problem of engageability improvement goes beyond the assessment and aims at making non-feasible engagements feasible by changing the engagement geometry and dynamics.

The focus of this paper is on on Anti-Air Warfare (AAW) [7], and more specifically the problem of *combat power management* to counter Anti-Ship Missiles (ASM). This problem is very constrained by the availability of the combat resources, the most important ones being hardkill and softkill weapons. Furthermore, most of the weapons rely on supporting resources for their deployment. An example of such supporting resources is given by the Fire Control Radars (FCR) [22] that offer a limited number of concurrent channels. Since typical AAW hardkill weapons require FCR support, the engageability of targets using hardkill is very dependent upon the availability of FCR. This is not the case for softkill, which can be fired without the need of FCR.

In this work, the availability of FCRs is used as a key parameter in the assessment of the target engageability using hardkill weapons. When required and possible, the engageability is improved by adapting the object[1] tracking functionality [1, 4] using an on-line estimation of the FCR search-to-lock-on time. This represents a new approach, which is partly inspired from the work of [10, 11] on covariance control. If the target is not and cannot be made engageable, softkill engagements are advocated. As detailed in the sequel and under given conditions, both engageability assessment and improvement exploit the dependency of the FCR search and lock-on duration on the error covariance of the track of the target to be engaged; the track being provided by the surveillance system. Scenarios involving

[1]Here we make a clear distinction between objects and targets, since not all objects will become targets (from the engagement perspective).

generic supersonic Anti-Ship Missiles (ASMs) are used to demonstrate the proposed approach. The evaluation is performed within a simulation and evaluation environment developed at Defence Research and Development Canada–Valcartier. This environment is a combination of a set of tools, including the Simulation Environment for the Analysis of the Tactical Situation (SEATS) test-bed [16], Ship Air Defense Model (SADM) simulator [21] and Concept Analysis and Simulation Environment for Automatic Target Tracking and Identification (CASE-ATTI) test-bed [20].

This paper is original in that it is one among the very few to address the engageability assessment problem, and the first to propose an engageability improvement approach, and to integrate it with object tracking and weapons assignment functionalities.

The paper is organized as follows. Section 2 presents the naval Command and Control (C^2) problem. Section 3 discussed the target engageability concept and the role of fire control operation in the detect-to-engage sequence. A method for assessing the target engageability based on FCR search-to-lock-on time is presented in Section 4. The result of engageability assessment is exploited in Section 5 to proposed a target engageability improvement solution. The simulation results and their discussion are given in Section 6.

2. COMMAND AND CONTROL PROBLEM

Military Command and Control (C^2) is a very complex problem and often this complexity rises from the multitude, the heterogeneity and the inter-relationships of the systems and resources involved. This is in general the case when simultaneous engagements, involving heterogeneous sensor and/or weapon systems, can take place. Decision support aids can help in overcoming the inherent complexity of simultaneous engagements.

Naval tactical C^2, which defines the context of this work, can be decomposed into a set of generally accepted functions that must be executed within some reasonable delays to ensure mission success. A very high-level description of those functions, related to battlespace management, is given below (Fig. 1). Note that the presented C^2 model is proposed here for our specific target engagement application. Waltz and Llinas [23] present a more generic description and review of the more the general Command, Control and Communications (C^3) problem.

2.1. Surveillance

Surveillance includes *object detection*, *object tracking*, and *object identification*. Object detection is very dependent upon the sensors performance. Object tracking uses the sensor data to estimate the current kinematical properties of the object, and predict their future positions. Object identification (and classification) assesses the identity and the class of objects. This also results in the resolution of true objects from decoys.

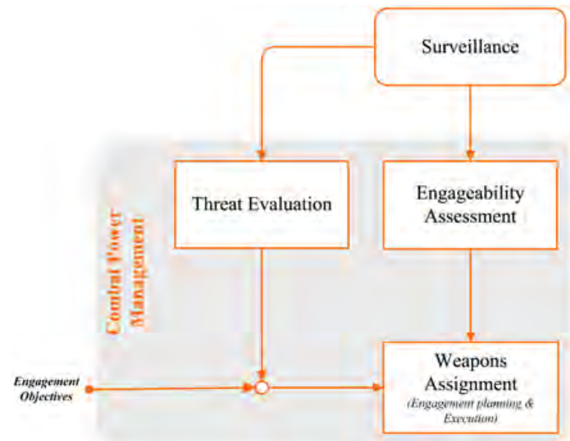


Fig. 1. Global view of C^2 process.

2.2. Combat Power Management

To defend itself, a warship relies on a set of tactical resources, which we will refer to as Combat Power (CP). These consist mainly of weapons, sensors, navigation, and communication systems. For a typical frigate, such as the Canadian Halifax Class, the Anti-Air Warfare (AAW) weapons include hardkill and softkill. Hardkill weapons are directed to intercept its target and actively destroy it through direct impact or explosive detonation in the proximity of the target. Hardkill weapons for a typical frigate include Surface to Air Missiles (SAM), an intermediate range Gun, and a Close-In Weapons System (CIWS). Softkill weapons use techniques to deceive or disorient the target to cause it to destroy itself, or at least lose its lock on its intended target (*i.e.*, ownship or the high value unit). The AAW softkill weapons for a typical frigate include decoys (Radio Frequency/Infrared) and jamming systems (on-board/off-board).

Combat Power Management[2] (CPM) functionalities include, as depicted on Fig. 1, *threat evaluation*, *engageability assessment* and *weapons assignment*, which are described below.

2.2.1. Threat Evaluation

Threat evaluation establishes the intent and the capability of the non-friendly entities within a certain Volume Of Interest (VOI) and for a specific reference point. It refers to the ongoing process of determining if an entity intends to inflict evil, injury, or damage to the defending forces and/or their interests, along with the ranking of such entities according to the level of threat they pose. In this work, threat value computation is based on the Closest Point of Approach (CPA).

[2]Also referred to as Threat Evaluation and Weapons Assignment (TEWA).

2.2.2. Engageability Assessment

Engageability assessment [9, 12, 8] concerns the evaluation of own force's engagement options feasibility against the non-friendly entities within the VOI. This process is intended to help the weapons assignment process by eliminating candidate solutions that violate one or more hard constraints. The latter will therefore not be feasible. Several aspects can be taken into consideration during this process, such as Rules Of Engagement (ROE), blind zones, ammunition availability, etc.

2.2.3. Weapons Assignment

Weapons assignment makes decisions on how to deal with the identified threats (and that become targets now). This process can be subdivided into several subproblems that include mainly a *response planning* and *response execution and monitoring*. Response planning ensures that one or more weapons are assigned to engage each target, including the assignment of supporting resources (as sensors, communications, etc.). This is about assignment of both resources (a pure allocation problem) and start and end times to activities (a pure scheduling problem). We talk about joint resource allocation and scheduling problems, that generates a ranked engagement list of the targets for the response execution module. Response execution and monitoring is the process by which the planned response is executed in real-time. This also includes the execution monitoring functionality. Since the responses are executed in a dynamic environment, subject to uncertainty, changing goals, and changing conditions, the actual execution contexts will be different from the projected ones.[3] Monitoring is required to help detect, identify and handle contingencies caused by uncertainty and changing nature of the environment.

3. FIRE CONTROL AND ENGAGEABILITY PROBLEM

To provide response to an ASM attack, human operators in charge of AAW go through a standard sequence of operations referred to as *detect-to-engage sequence*. This temporal sequence starts with the object detection by the surveillance system and ends with the object (now a target) engagement. Fig. 2 illustrates the main operations within this sequence. These include the object detection and tracking by the surveillance system, FCR cueing, acquisition by FCR (that relates to the FCR search-to-lock-on) and engagement.

The duration of the detect-to-engage sequence is crucial to the ship survival. Short detect-to-engage sequence provides room for the re-engagement of the same target (in the case of a target miss assessment) or the engagement of one or more different targets (in the case of a target kill/seduction assessment).

[3]The ones that motivated the construction of the original response.

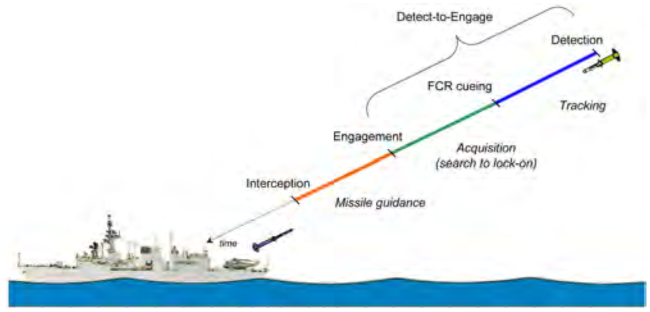


Fig. 2. Detect-to-engage sequence.

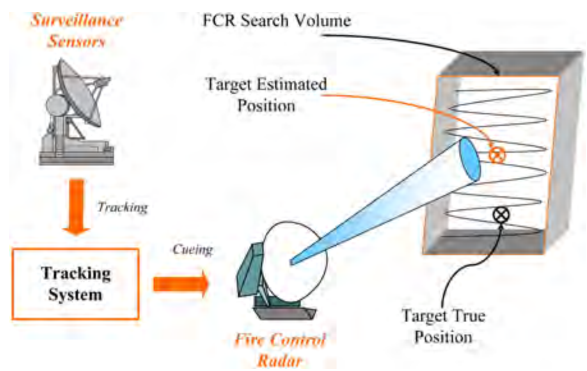


Fig. 3. Fire control cueing.

The duration of the whole detect-to-engage sequence depends on the individual durations of the composing functions, over which the decision making (human, automation or both) has some control. This is especially the case for the FCR search-to-lock-on. This control will be used, as explained in the sequel, to evaluate and optimize the duration of the detect-to-engage sequence in order to improve the engageability of targets.

3.1. FCR Cueing

The FCR system effectively offers two concurrent fire channels for the hardkill weapons that provide high accuracy track data for target engagements. Fig. 3 illustrates how the surveillance sensors and the related tracking system cue the FCR system to help it acquire the target and provide a hardkill firing solution [17]. It is assumed here that FCR cueing includes the designation phase in which the FCR is directed to the estimated location of the target.

Upon detection by the 2D surveillance radars, the contact information is provided to the tracking system that maintains a more accurate estimation of the object position and infers its identity and classification. In the sequel, only the object position will be considered. It is given by the state estimate $\hat{\mathbf{x}} = [\hat{x}, \hat{y}, \hat{\dot{x}}, \hat{\dot{y}}]^T$ and the related error covariance matrix $\mathbf{P}$ that represents some measure of the tactical picture accuracy. Note that in [19], accuracy is defined in terms of the Root Mean Square Error (RMSE) of a track. Here we will assume a consistent tracking and optimal filter [3], where the target behavior follows the motion model used by the tracking filter, so that the RMSE and the error covariance will converge towards the same value.

Given its high-risk consequences, the engagement phase requires more precise information than the surveillance operations. This is why the (3D) FCR must take over the less accurate (and 2D) surveillance radars. To provide such accurate information, the FCR will have to acquire and then track the designated target by itself. Therefore, once a decision is made to engage a given target, the corresponding positional information is used to cue the FCR, *i.e.*, to delimit its search region (Fig. 3) for its search and lock-on phase. This phase starts once the FCR begins its scan[4] and ends when the FCR locks on the target.

3.2. FCR Search-to-Lock-on Time

Following a specific pattern, the FCR will scan the specific region of the Volume Of Interest (VOI) until it detects and locks on the target for which a track is then maintained. The target course and speed contained in this FCR track is then used to compute a Predicted Intercept Point (PIP) inside the weapon engagement envelope. The goal is to provide guidance (for the missile) or the pointing (for the gun) information toward the engaged target.

During this target acquisition, or search-to-lock-on, phase, the FCR has a search time that depends on several factors, such as: the ownship weapons properties, Command and Control System (CCS) performance, the operator skill/training, the engaged target characteristics, etc. Nevertheless, the search-to-lock-on duration should be limited to avoid wasting the valuable and scarce reaction time.

3.3. Track Accuracy, Search-to-Lock-on, and Engageability

The accuracy of the information cued to the FCR determines the volume it must scan. The time it will take to re-acquire the target, that is the duration of the search-to-lock-on operation, depends in a non-linear manner of the volume to be scanned and the detection probability of the FCR. This duration is subtracted from the total reaction time available to the decision-maker and/or combat power management capability. Note that the ship survival is very depending upon this reaction time.

A poor track accuracy causes the FCR to search in a large volume, so that it will take more time to acquire the target. This may lead to grave consequences on ownship safety. Therefore, the engageability of the targets is a function of the accuracy of their tracks as provided by the surveillance system. Hence, controlling the track accuracy, on the surveillance side, offers a means to improve the targets engageability and increases the chance of ownship to achieve its engagement objectives.

[4]Upon cueing from the surveillance system.

4. ENGAGEABILITY ASSESSMENT

Engageability assessment defines the process of evaluating the engageability of a specific target, *viz.*, evaluating the ability to successfully execute a specific engagement action against a specific target. Success here is related to the ability to undertake the action, given the tactical situation, and not to the outcome of the undertaken action. Actions refer to defensive strategies where one or more weapons are assigned to the target. Engageability is assessed over a multi-dimensional space, which includes time, space, frequency spectrum, etc.

In this work, the focus is on time. The engageability is defined as the feasibility, in terms of scheduling or time-lining, of a specific engagement as defined by the duration of the detect-to-engage process. The evaluation also considers target state and characteristics as well as characteristics of the defensive weapons and of their related resources. This evaluation aims at reducing the combat power management problem complexity and save the weapons assignment planning time by discarding inconsistent candidate solutions. Thus, a feasible alternative must verify a set of constraints and will be eliminated if it violates any one. For example, an alternative is retained if, for each considered hardkill engagement,

1) the requested FCR is available;
2) the target to be engaged is within the range of the selected FCR;
3) the interception will occur within the weapon envelope; and
4) the target is not in the blind zones of FCR and weapons.

In the remaining, only the availability of the FCR is considered for the engageability assessment, where the assessment process considers time constraints over the predicted timeline of the interception.

4.1. FCR-Based Engageability Assessment

The proposed engageability assessment computation is based on estimations of both the search-to-lock-on time (t_s) and of the detect-to-engage duration (t_{de}). The engageability of a specific target depends on its kinematics as well as on the properties of the weapons ($\mathcal{W}$) and the characteristics of the FCR.

There is a minimum admissible range of interception $r_i^{\min}$ that depends on both the weapon and the FCR. Here it is assumed that $r_i^{\min}$ corresponds to the weapon's minimum effective range $r_w^{\min}$. Accordingly, the predicted target intercept range r_i should always be such that

$$r_i \geq r_i^{\min} = r_w^{\min} \tag{1}$$

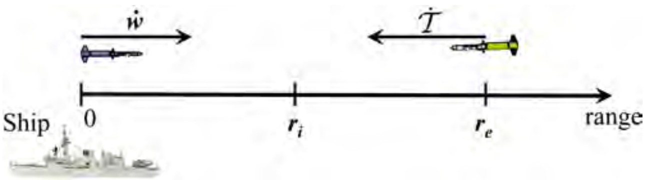


Fig. 4. Target interception range r_i.

where r_i is determined based on the target velocity $\dot{\mathcal{T}}$, the target range r_e at the beginning of the engagement, and the weapon velocity $\dot{w}$ (see Fig. 4)

$$r_i = \left[\frac{\dot{w}}{\dot{w} + \dot{\mathcal{T}}}\right] r_e. \tag{2}$$

The assessment of engageability is performed as by equation (1), with focus on specific steps in the engagement. First, given $r_w^{\min}$ there is a minimum target range $r_e^{\min}$ at which the weapon must be fired to make the interception happen within its effective range. If the weapon $\mathcal{W}$ is launched while the target $\mathcal{T}$ has already passed $r_e^{\min}$, it will be too late, *i.e*, $r_i \notin [w^-, w^+]$. Therefore, the weapon $\mathcal{W}$ must be fired while the target $\mathcal{T}$ is beyond $r_e^{\min}$. To make this interception possible, the target $\mathcal{T}$ must be acquired by the FCR at a range r_l such that

$$r_l \geq r_e \geq r_e^{\min}. \tag{3}$$

From Equation 2, it is clear that the intercept range r_i is dependent on r_e (target range at the end of the detect-to-engage sequence). The sequence includes detection, surveillance sensor tracking, FCR cueing, FCR search-to-lock-on, FCR tracking, and finally weapon launch. The duration t_{de} of the whole sequence is function of the different phases, as follows

$$t_{\text{de}} = t_{\text{tr}} + t_s + \Gamma(t_{\text{det}}, t_{\text{cue}}, t_{\text{tr}f}, t_{\text{wi}}) \tag{4}$$

where the durations of detection (t_{det}), FCR cueing (t_{cue}), FCR tracking ($t_{\text{tr}f}$), and weapon launch initialization (t_{wi}) are assumed non-controllable (for this work) and gathered in a single function Γ. t_{tr} and t_s designate respectively the durations of tracking (with the surveillance sensors) and search-to-lock-on of the FCR. A limit $t_{\text{de}}^{\max}$ on the detect-to-engage time t_{de} is set using the limit $r_e^{\min}$ defined by the minimum range beyond which the target must be engaged

$$t_{\text{de}} \leq t_{\text{de}}^{\max} = \frac{r_d - r_e^{\min}}{\dot{\mathcal{T}}} \tag{5}$$

where r_d is the range at which the threat is detected by the surveillance sensors, and $r_e^{\min}$ is given by

$$r_e^{\min} = r_w^{\min}\left[1 + \frac{\dot{\mathcal{T}}}{\dot{w}}\right]. \tag{6}$$

Consequently, the detect-to-engage time t_{de} has an influence over the predicted intercept range r_i. Thus, any constraint on r_i can be reformulated as a constraint on t_{de}. Furthermore, the duration t_{de} is mainly determined by the length t_s of the search and lock-on phase of the FCR and by the duration t_{tr} of the tracking phase (by the surveillance system). Therefore, the constraint on t_{de} can be re-expressed as a constraint on t_s

$$t_{\text{de}} = t_{\text{tr}} + t_s \leq t_{\text{de}}^{\max} \Rightarrow t_s \leq t_s^{\max} \tag{7}$$

where

$$t_s^{\max} = t_{\text{de}}^{\max} - t_{\text{tr}} \tag{8}$$

$$= \frac{r_d - r_e^{\min}}{\dot{\mathcal{T}}} - t_{\text{tr}} \tag{9}$$

$$= \frac{r_d - r_w^{\min}}{\dot{\mathcal{T}}} - \frac{r_w^{\min}}{\dot{w}} - t_{\text{tr}}. \tag{10}$$

Moreover, since the duration of the search and lock-on phase of the FCR depends on the uncertainty related to the track of the target ($\mathbf{P}$), the established constraints on t_s can be re-expressed as constraints on $\mathbf{P}$. Note also the all of the constraints described above can be re-expressed in terms of time instead of range. The next section will show how the search-to-lock-on time and the detect-to-engage time can be estimated.

4.2. Estimation of the Sequences Duration

The core idea of the presented work is to generate engagement strategies that exploit contextual information. This information is given, in this work, by an estimate of the duration of the engagement sequence, focusing on the search-to-lock-on time of the FCR. These estimated values, shown in Fig. 2, will be used both for assessing and improving target engageability. As mentioned above, the estimated search-to-lock-on time ($\hat{t}_s$) of the FCR is a key parameter of the engagement sequence. It is evaluated based on both the characteristics of the target and of the FCR.

Error covariance of the track handed-over to the FCR also influences the estimation in the determination of the search volume and conditional detection probabilities. Assuming Gaussian noise for both the target dynamics and the measurement process, let $\hat{\mathbf{x}}$ and $\mathbf{P}$ represent the target state estimate and its error covariance matrix respectively, $\mathbf{Q}$ the process noise covariance matrix for the discrete time interval h and $\mathbf{R}$ the measurement error covariance matrix. Considering regular measurement updates at an update interval h, the track accuracy represented by $\mathbf{P}$ can be expressed in terms of the tracking time t_{tr}:

$$\mathbf{P} = \mathcal{F}(t_{\text{tr}}) \tag{11}$$

$\mathcal{F}$ is evaluated recursively by applying the Kalman covariance update equations given by

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{W}_{k+1}\mathbf{H}_{k+1}\mathbf{P}_{k+1|k} \tag{12}$$

with

$$\mathbf{W}_{k+1} = \mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^T[\mathbf{H}_{k+1}\mathbf{P}_{k+1|k}\mathbf{H}_{k+1}^T + \mathbf{R}]^{-1} \tag{13}$$

and

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^T + \mathbf{Q} \tag{14}$$

where **H** and **F** are the measurement and state transition matrices respectively. Thus, for a tracking duration of $n \times h$ seconds, **P** is obtained recursively by evaluating equation (15) n times. The estimated search-to-lock-on time can be expressed in terms of **P**

$$\hat{t}_s = \mathcal{G}(\mathbf{P}) = \mathcal{G}(\mathcal{F}(t_{\text{tr}})) \tag{15}$$

where $\mathcal{G}$ is the estimation function. As more to **P**, $\mathcal{G}$ depends on other variables that include the characteristics of the FCR. There is no explicit analytical form for the function $\mathcal{G}$, which instead is computed recursively. More details on the computation of $\mathcal{G}$ are given in [18] for a standard fixed swath search pattern of the FCR. In short, the estimation function defined in [18] considers

1) a FCR model that comprises beam shape, direction displacement speeds and search pattern;
2) a search area that is defined by delimiting an amount of the localization probabilities given by **P**;
3) conditional detection probabilities for the FCR;
4) a multi-scan time estimation related to a cumulative detection probability;

Let $p_a(t)$ be the density function associated with the probability that the FCR detects and acquire the target at time t. Then the search-to-lock-on time estimation function $\mathcal{G}$ is defined as

$$\hat{t}_s = \mathcal{G}(\mathbf{P}) = \int_0^{+\infty} p_a(t) t \, dt \tag{16}$$

where t_s can be seen as a random variable with mean $\hat{t}_s$. p_a depends on the target localization probability p_L and on the conditional detection probability $p_{D|L}$ that depends on the properties/performance of the sensor used. This is where the track accuracy represented by **P** has an influence since p_L is defined according to the Gaussian target distribution subsumed by the error covariance matrix **P** [18, 17]. Note that the fact that p_L is Gaussian does not imply that t_s has a Gaussian distribution. Also, in this work, $\hat{t}_s$ will be considered as an exact estimation of the search-to-lock-on, as a primary study, although it is acknowledged that future work should consider the probabilistic nature of t_s. $p_{D|L}$ depends on the properties/performance of the FCR.

Moreover, the discretized form of (16) is

$$\hat{t}_s = \sum_k p_a(t_k) t_k \tag{17}$$

where $p_a(t_k)$ represents the probability mass function. Note that $p_a(t_k)$ does not have an explicit analytical form. It is computed recursively [18].

Finally, substituting (15) into (7), the estimated detect-to-engage time is

$$\hat{t}_{\text{de}} = t_{\text{tr}} + \sum_k p_a(t_k) t_k \tag{18}$$

$$\hat{t}_{\text{de}} = t_{\text{tr}} + \mathcal{G}(\mathcal{F}(t_{\text{tr}})) \tag{19}$$

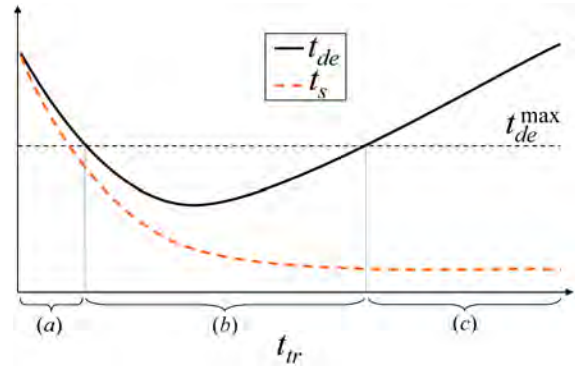


Fig. 5. Detect-to-engage (t_{de}) and search to lock-on (t_s) durations in terms of tracking time t_{tr}. Measurement updates are assumed to occur regularly so that their number is proportional to the tracking time t_{tr}. (a) & (c) Target not engageable. (b) Target engageable.

which shows that the detect-to-engage time t_{de} (or its estimate $\hat{t}_{\text{de}}$) can be expressed as a function of t_{tr}.

Before going further about the estimation of the search-to-lock and time, it must be acknowledged that the presented estimation functions are restrained to specific tracking conditions to produce useful results for the development of target engagement strategies. For instance, the presented strategies in the next paragraphs necessitate having a monotonically decreasing search-to-lock-on time function in terms of the tracking duration. This implies conditions on the tracking process that mainly involve the process noise, the measurement noise and the measurement update rate for the tracking filter. Also, it is also assumed that measurements are received regularly over time such that as the tracking process goes on more measurements are received and the track uncertainty gets reduced. As more, to obtain a monotonically decreasing track error covariance, a relatively high measurement update rate is needed and the ratio of the process noise over measurement noise must be low enough. This should allow having a monotonically decreasing search-to-lock-on time function and a corresponding detect-to-engage function that is characterized with a single minimum as in Fig. 5.

The functions for $\hat{t}_s$ and $\hat{t}_{\text{de}}$ in terms of the tracking duration t_{tr} that we are needing to apply adaptive tracking strategies are illustrated on Fig. 5, where we have $t_{\text{de}} = t_s + t_{\text{tr}}$.

In comparison, Fig. 6 shows the same functions obtained experimentally with $h = 0.4$ and with surveillance sensor measurement noise standard deviations $\sigma_\beta = 0.035$ rad for bearing and $\sigma_r = 1$ m for range.[5] The process noise follows the pulse model with power spectral density (standard deviation) of 1 m^2/s^3 [4]. The tracking parameters were adjusted to provide the monotonically decreasing search-to-lock-on time and the minima of the detect-to-engage function.

[5]The tracking system converts the sensor measurements from Polar to Cartesian coordinates using the conventional coordinate transformation [4, 2].

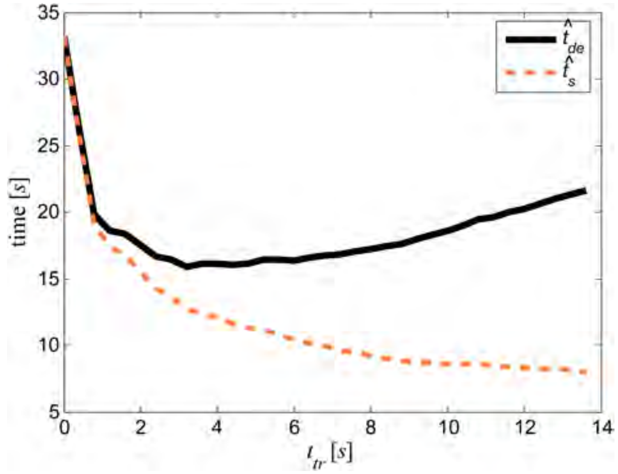


Fig. 6. Estimated detect-to-engage ($\hat{t}_{\mathrm{de}}$) and search to lock-on ($\hat{t}_s$) durations in terms of tracking time t_{tr} for a target initially situated at 75 km from the ship and with speed of 700 m/s in the direction of the ship.

Hence, controlling duration t_{tr} before cueing the FCR in order to satisfy the conditions expressed in (1) and (5) can ensure that the target is engageable. As explained below, three situations may occur (Fig. 5) based on the constraint $t_{\mathrm{de}}^{\max}$:

a) A too short tracking time causes a too low track accuracy, and therefore a too long search-to-lock-on time. The target is not engageable since the predicted target intercept range will be below its minimum limit defined in (1).

b) A good compromise between tracking time and search-to-lock-on time. The target is engageable since the predicted target intercept range will be above its minimum limit defined in (1).

c) A too long tracking time before cueing the FCR. The FCR takes a short time to lock on the target (due to the high accuracy of the track). Nevertheless, the gain in search-to-lock-on time cannot compensate for the long time spent in tracking. As for (a), the target is not engageable.

The proposed target engageability improvement solution, presented in the next section, will help maintaining situation (b) for different engagement scenarios.

5. TARGET ENGAGEABILITY IMPROVEMENT

As stated in the previous section, the engageability assessment aims at supporting the weapons assignment planning process. Instead of performing it in open-loop manner (Fig. 1), we propose a closed-loop approach that combines the engageability assessment and engageability improvement, as shown in Fig. 7.

The concept of engageability improvement goes beyond the assessment concept by changing the engagement geometry and dynamics, to make non-feasible engagements feasible. In this work, it is shown that the engageability can be improved through the minimization of the detect-to-engage time in given in equation (19). The proposed target engageability improvement approach is based on the control of the FCR cueing time. This is achieved through feedback to the object tracking function, such as illustrated in Fig. 7. Both the engageability assessment and the engageability improvement functions use an estimation of the FCR's search-to-lock-on time, and interact with the data fusion (*i.e.*, object tracking), threat evaluation and weapons assignment processes.

The engageability improvement function controls the cueing time of the FCR by setting a tracking duration t_{tr} for each target based on the desired search-to-lock-on time and the underlying track accuracy (Fig. 8). Practically, the tracking duration is determined itera-

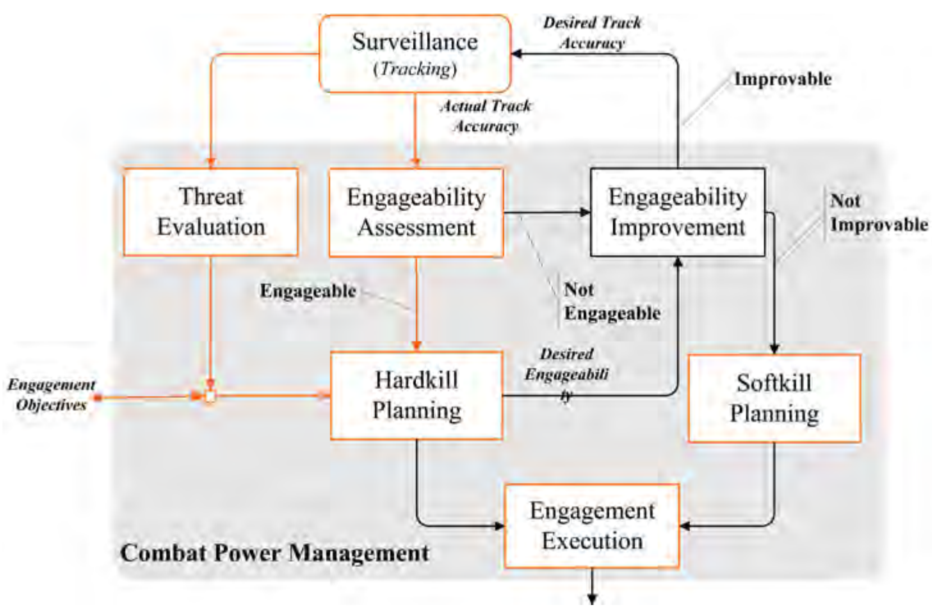


Fig. 7. Target engageability improvement in the C^2 process.

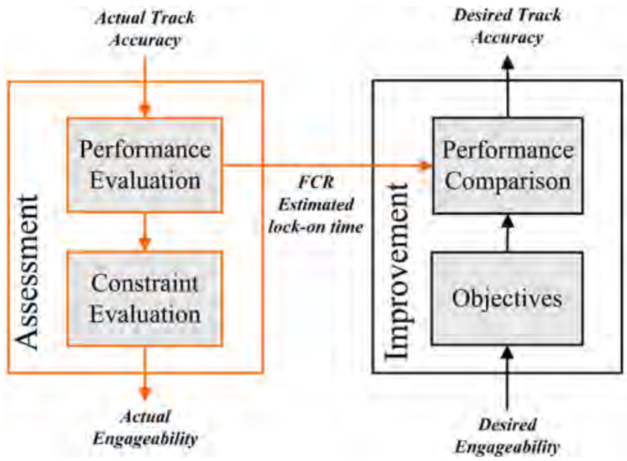


Fig. 8. Engageability assessment and engageability improvement interaction.

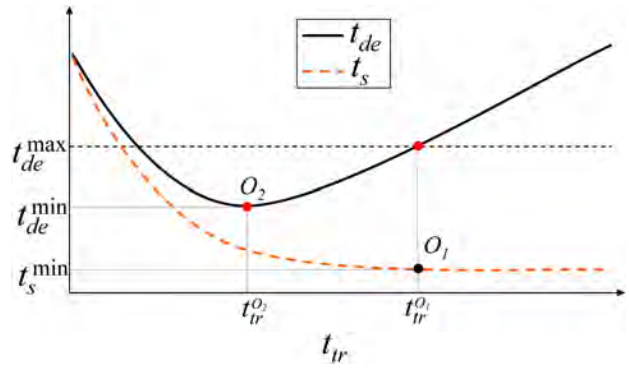


Fig. 9. Objectives of the detect-to-engage sequence.

tively at the instant where the estimated search-to-lock-on time (and the underlying error covariance of the track) reaches an objective threshold shown in Fig. 8. It is chosen such that some operational objectives on the detect-to-engage and search-to-lock-on durations are achieved. These objectives depend mainly on the number of planned engagements and their configurations. In the followings, two operational objectives (labeled respectively O_1 and O_2 on Fig. 9) are considered.

O_1—aims at intercepting the target as close as possible to the weapon's minimum intercept range $r_i^{\min}$. This translates in long tracking time, minimum[6] search-to-lock-on time ($t_s = t_s^{\min}$) and high probability of interception. On the other hand, this corresponds to the maximum detect-to-engage duration ($t_{\mathrm{de}} = t_{\mathrm{de}}^{\max}$), as illustrated on Fig. 9).

O_2—aims at intercepting the target such as to minimize the total detect-to-engage duration ($t_{\mathrm{de}} = t_{\mathrm{de}}^{\min}$). This consists in finding the tracking duration and cueing time that guarantee the achievement of this objective.

The proposed engageability improvement approach considers the initial hardkill engagement plan for each target. Then, based on the engageability assessment, the detect-to-engage objectives are adjusted to improve engageability. When the target is not engageable, and there is no room for improvement, hardkill engagements are dropped and softkill is recommended instead for the concerned targets. When the engagement objectives are met and the engageability of each target is satisfactory, the hardkill plan is made available for execution and the FCR is cued following the computed engagement schedule. The execution also includes the softkill strategies for the targets that were not considered engageable.

Fig. 10 illustrates an engagement sequence example of two targets, the engageability the second target is improved through the minimization of the detect-to-engage duration for the first target. It is shown that both targets can be intercepted in time with the appropriate selection of the tracking durations $t_{\mathrm{tr}1}$ and $t_{\mathrm{tr}2}$.

[6]Which makes FCR more available for other engagements and also minimizes the signature (*i.e.*, detectability) of ownship.

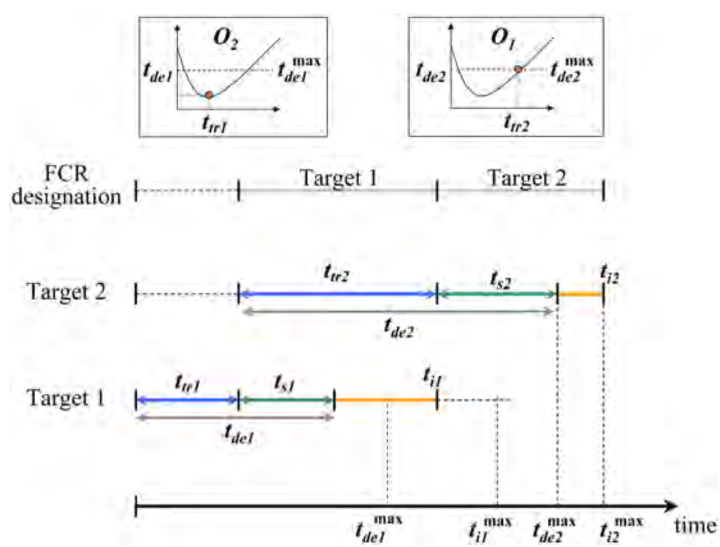


Fig. 10. Engagement sequence of two targets, with detect-to-engage time minimization on the first target (t_{dk} is detection time and t_{ik} is interception time for target k).

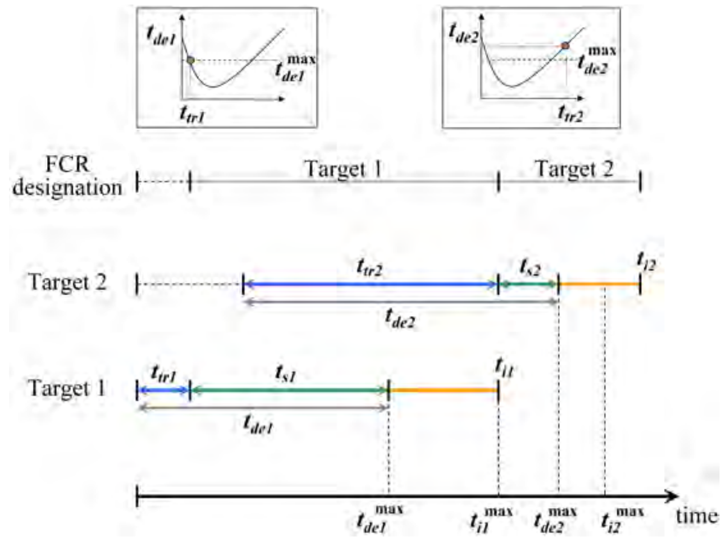


Fig. 11. Engagement sequence of two targets, without detect-to-engage time minimization on the first target (t_{dk} is detection time and t_{ik} is interception time for target k).

On the other hand, Fig. 11 shows the engagement sequence of the same two targets where no improvement of engageability was used. Because the tracking durations $t_{\mathrm{tr}1}$ and $t_{\mathrm{tr}2}$ are not set correctly, one of the targets will not be engageable.

6. SIMULATION AND RESULTS

A quantitative evaluation of the proposed target engageability improvement approach was performed using a combination of the SEATS test-bed [16], SADM simulator [21] and CASE-ATTI test-bed [20]. The target engageability improvement approach is based on minimization of the detect-to-engage time. The demonstration uses the search-to-lock-on time estimator and the FCR model presented in [17].

Two scenarios, featuring a warship that is attacked respectively by one or two supersonic ASMs, are presented. The scenarios were defined such that the duration of the detect-to-engage sequence is critical to the ship survival. The simulated scenarios, including weapons and targets characteristics, are kept simple to avoid incorporating any military CLASSIFIED information. Nonetheless, the simulation remains rich enough to illustrate the benefits of the proposed approach as a first study.

The ship is assumed equipped with SAMs as primary hardkill weapons. The SAM minimum intercept range is assumed to be $r_w^{\min} = 1000$ m. Any interception below this range is considered to be highly unlikely successful. In this case, the defending ship would be hit by the ASM. Also, it is assumed that the ship has only one FCR available.[7] Fig. 12 and Fig. 15 illustrate the two used scenarios as scripted in the simulation environment, using STAGE™.

For each of the two scenarios presented in the next sections, the performance of the defending ship is evaluated using two different defensive strategies. The first one is a conventional engagement method that does not rely on the estimation of the search-to-lock-on time. With this strategy, the FCR is cued as soon as the ASM is detected and a confirmed track is established. This corresponds to the typical tactic used by most navies in the world. The second strategy exploits engageability assessment and improvement concepts. Although more sophisticated hardkill and softkill coordination strategies exist [14, 13, 5, 6], softkill combat resources are used, in this work, as second resorts in cases where hardkill engagements are deemed not feasible. Finally, it must noted that the actual search-to-lock-on time t_s and its estimated value $\hat{t}_s$ are considered identical in this simulation. Hence the results should be treated as average results rather than single instances out a probability distribution. However, a future work that will consider the probabilistic aspect of the search-to-lock-on time will offer a natural extension to this work.

6.1. Single Target Scenario

The first scenario (Fig. 12) considers a closing single supersonic ASM with a zero CPA relative to the ownship.

This scenario provides the ship with conditions for re-engagement should a miss occur. More precisely, it is assumed that the target is missed at its first engagement. The miss is due to the SAM performance and a second engagement is then required to intercept the target. In that case, a second SAM could be launched shortly after the miss assessment. The following will show how the minimization of the detect-to-engage sequence can provide the opportunity of a second engagement

[7]Note that Canadian Frigates of Class Halifax have two FCRs.

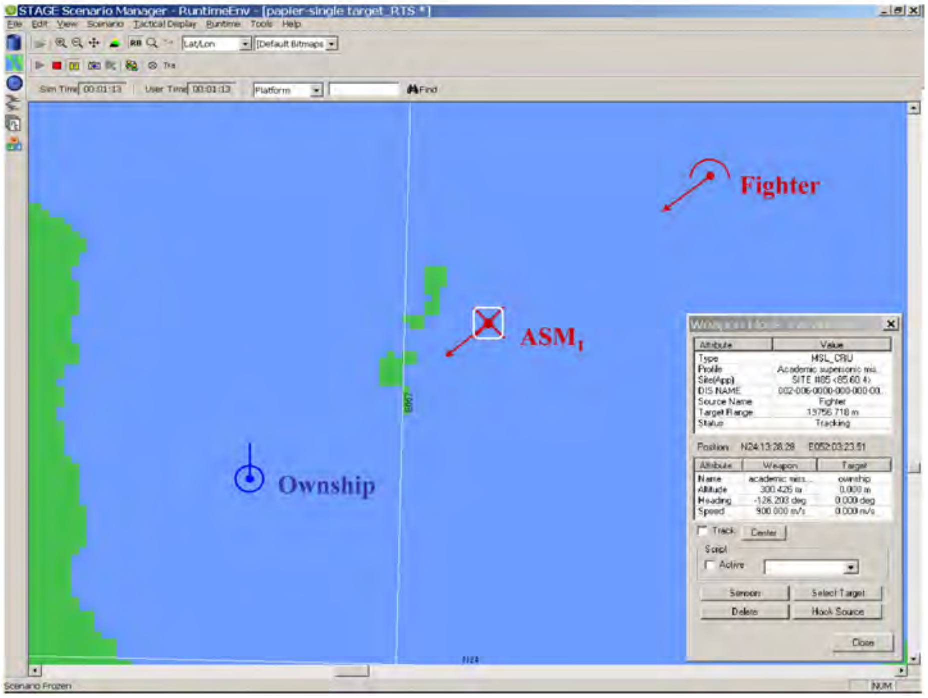


Fig. 12. Single target scenario in the STAGE (within SEATS test-bed).

compared to the conventional engagement method and under specific target conditions.

Suppose a scenario that starts at $t = 0.0$ s with the detection of the ASM by the surveillance system at the initial range of 26000 m. The ASM has and initial altitude of 300 m and a speed of 900 m/s. Under these conditions, the threat time-to-go (or time on flight) is about 29 s. The tracking parameters are given in Table I. Note that the parameters were set to study how the estimation of the search-to-lock-on time and related cueing strategies could improve the engageability of targets.

6.1.1. Conventional Engagement

With the conventional engagement strategy, the FCR is cued by the surveillance system as soon as a confirmed track is obtained, that is 2.9 s after the first detection. The FCR locks on the target at 21.3 s. One second later, a first SAM is fired. It misses the target at time $t_m = 26.5$ s, as illustrated in Fig. 13. According to the minimum intercept range ($r_w^{\min}$) of the SAM, the maximum intercept time is:

$$t_i^{\max} = \frac{r_d - r_w^{\min}}{\dot{T}} = \frac{26000 - 1000}{900} = 27.78 \text{ s} \qquad (20)$$

so that the target must be intercepted before $t = 27.78$ s. A second SAM could be fired not until $t = 27.5$ s, so it that would be too late to intercept the threat. The warship is hit by the target at 28.9 s, unless a softkill is used as a backup strategy.

TABLE I
Tracking Parameters for the Single-Target and Two-Target Scenarios

Track update period (h)	0.4 s
Search and surveillance radar accuracy in bearing (σ_β)	0.035 rad
Search and surveillance radar accuracy in range (σ_r)	1 m
Process noise power spectral density	1 m^2/s^3

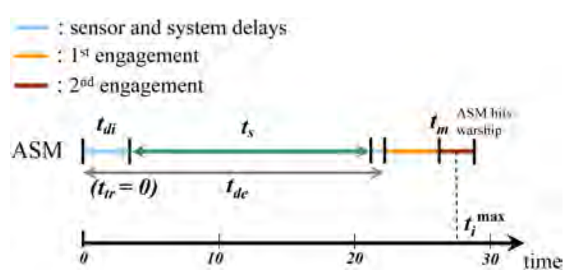


Fig. 13. Engagement sequence without detect-to-engage time minimization (single target scenario).

6.1.2. Detect-to-Engage Time Minimization

Using the minimization of the detect-to-engage time, the FCR is not cued as soon as a confirmed track is obtained. Instead, it is cued once the minimum value of the detect-to-engage duration is reached (*i.e.*, at $t_{\text{tr}} = 9.9$ s). A shown in Fig. 14, this causes the FCR to lock on the target at 15.4 s, offering a gain of 5.9 s even if cueing occured 7.0 s later compared to the conventional engagement case. A first missile is then fired at 16.4 s and misses the target at 24.4 s, which leaves enough time for a re-engagement with a predicted interception

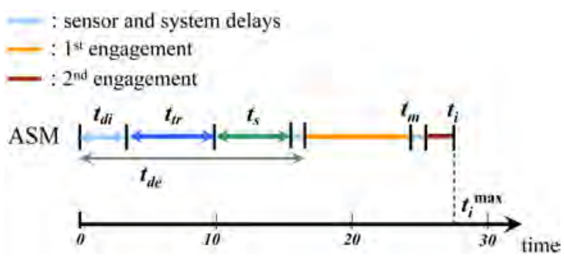


Fig. 14. Engagement sequence with minimization of the detect-to-engage time (single target scenario).

time below t_i^{max}. A second missile is fired at 25.4 s and hits the target at 27.7 s, at 1112 m from the ownship. Table II summarizes the results.

TABLE II
Results of the Conventional Engagement and the Engageability Improvement Method for the Single Target Scenario

Function	Time ([s])	ASM range ([m])
Without Engageability Improvement		
Cueing	2.9	23390
Acquisition (lock-on)	21.3	6830
1st engagement	22.3	5930
1st miss	26.5	2114
2nd engagement	(27.5)	(1217)
Interception	none, the ASM hits the warship at 28.9 s	
With Engageability Improvement		
Cueing	9.9	17090
Acquisition (lock-on)	15.4	12140
1st engagement	16.4	11240
1st miss	24.4	4014
2nd engagement	25.4	3114
Interception	27.7	1112

6.2. Two-Target Scenario

This second scenario (Fig. 15) illustrates the engagement of two closing supersonic ASMs, again with zero CPA relative to the defending ship.

ASM$_1$ has an initial range of 19000 m, an altitude of 300 m and its speed is 900 m/s. **ASM**$_2$ has an initial range of 23000 m, an altitude of 300 m and its speed is 900 m/s as well. **ASM**$_1$ pops up at 0.0 s, while **ASM**$_2$ pops up at 2.0 s. It is assumed that the ownship is aware that an attack by more than one ASM is potentially high.

6.2.1. Conventional Engagement

Without minimization of the detect-to-engage time of its, **ASM**$_1$ is intercepted just before it reaches the minimum intercept range (1000 m) of the SAM. This leaves no time for engaging **ASM**$_2$, which is detected by the surveillance radar at 2.0 s. At that time, the FCR is busy on **ASM**$_1$. The assignment of the FCR to **ASM**$_2$ takes place at 20.9 s and the acquisition occurs at 23.9 s. Another SAM could be fired at 24.9 s, but it would be too late to prevent **ASM**$_2$ from hitting the ship (at 27.56 s).

This is shown in Fig. 16, where the cueing of the FCR as soon as **ASM**$_1$ is detected ($t_{tr1} = 0$) has resulted in a long search-to-lock-on time ($t_{s1} = 13.9$ s). The evaluation of the engageability before the engagement takes place will allow the recommendation of a softkill strategy.

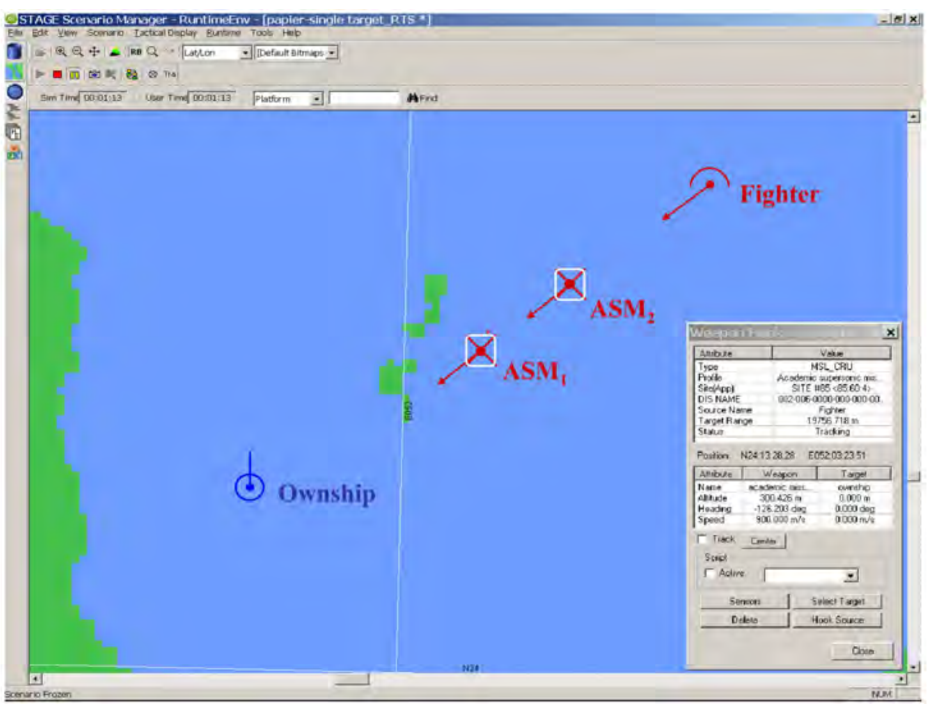


Fig. 15. Two-target scenario in the STAGE (within SEATS test-bed).

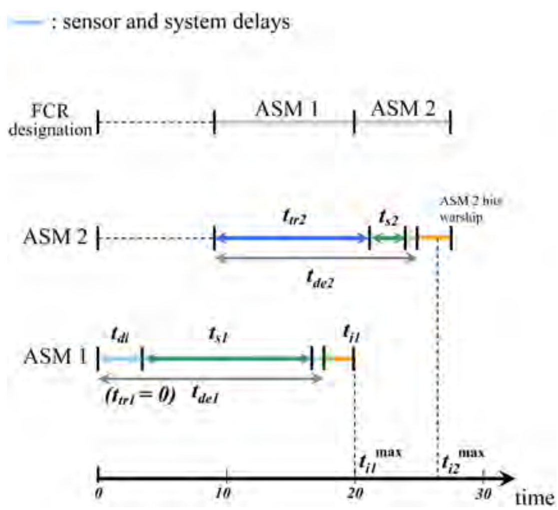


Fig. 16. Engagement sequence without detect-to-engage time minimization on the first target (two-target scenario).

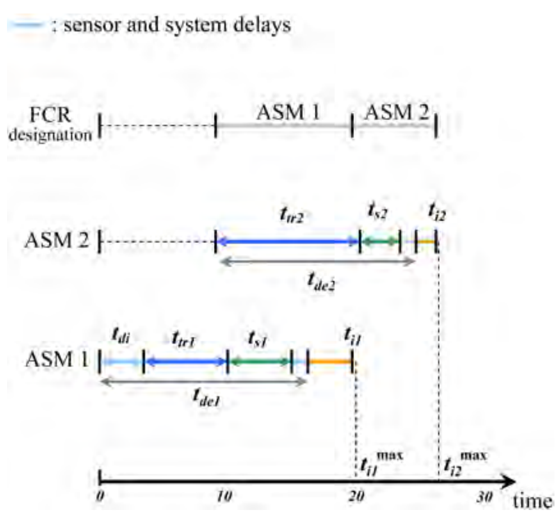


Fig. 17. Engagement sequence with detect-to-engage time minimization (two-target scenario).

6.2.2. Detect-to-Engage Time Minimization

With minimization of the detect-to-engage time, the FCR is assigned to the first detected target (**ASM**$_1$) at 9.9 s, that is 7.0 s after track confirmation. The FCR locks on the threat at 14.9 s (instead of 16.8 s with the conventional engagement method). A SAM is then fired at 15.9 s and hits **ASM**$_1$ at 19.25 s. Prior to that, the second target was detected by the surveillance radar at 2.0 s. The FCR was busy until made available once **ASM**$_1$ is assessed killed. The FCR is then assigned to **ASM**$_2$ at 20.25 s. It locks on it at 23.25 s (instead of 23.9 s with the conventional engagement method). A SAM is fired at 24.25 s against **ASM**$_2$, which is intercepted at 26.38 s. Fig. 17 shows the complete engagement sequence and Table III summarizes the results.

TABLE III
Conventional Engagement and the Detect-to-Engage Time Minimization Method for the Two-Target Scenario

Target	Function	Time ([s])	ASM range ([m])
Conventional Engagement			
ASM$_1$	Cueing and designation	2.9	16390
	Acquisition (lock-on)	16.8	3880
	Engagement	17.8	2980
	Interception	19.9	1064
ASM$_2$	Cueing and designation	20.9	5990
	Acquisition	23.9	3290
	Engagement	24.9	2390
	Interception	none, **ASM**$_2$ hits the warship at 27.56 s	
Detect-to-Engage Time Minimization			
ASM$_1$	Cueing and designation	9.9	10090
	Acquisition (lock-on)	14.9	5590
	Engagement	15.9	4690
	Interception	19.25	1675
ASM$_2$	Cueing and designation	20.25	6575
	Acquisition (lock-on)	23.25	3875
	Engagement	24.25	2975
	Interception	26.38	1063

6.3. Discussion

The two presented scenarios showed that the FCR cueing time t_{tr} can have a significant impact on the engageability of the targets and can be used as a control variable to influence the engagement sequence. The presented results are based on several assumptions regarding the engagement configuration, the shipboard resources, as well as the behavior of the different estimation algorithms used. For instance, for less critical situations (*e.g.*, single subsonic missile attack), cueing the FCR a few seconds later or a few seconds sooner may not impact much the outcome of the engagement. Also, the tracking parameters showed in Table I assumes that the error covariance to be monotonically decreasing in terms of the tracking time t_{tr}. A minimum detect-to-engage time strategy would be irrelevant in the cases where the error covariance does not decrease with time. Moreover, the results are very dependent on the model the FCR and the corresponding search-to-lock-on time, which can change considerably with the track error covariance [17, 18].

7. CONCLUSION

This paper considered the problem of target engagement in the naval Anti-Air Warfare operations. The naval Command and Control process was briefly described and an approach to improve the engageability of the targets was proposed. The proposed approach combines object tracking, threat evaluation, and weapons assignment in a closed-loop and integrated manner, and uses an estimation of the search-to-lock-on time of the FCR to control the tracking and cueing operations. Two scenarios, involving supersonic targets with the inherent short reaction time, were used to show the benefit of the proposed approach on the overall detect-to-engage sequence and the ownship survival. For the two presented illustrative scenarios, the conventional engagement method was unable to cope with the short reaction time constraints, and failed to defeat the threats using its hardkill resources. The engageability improvement strategy based on minimization of the detect-to-engage time provided a better way to exploit the same available and scarce reaction time.

REFERENCES

[1] Y. Bar-Shalom (Ed.)
Multitarget-Multisensor Tracking: Applications and Advances.
Artech House, 1990.

[2] Y. Bar-Shalom and X.-R. Li
Multitarget-Multisensor Tracking: Principles and Techniques.
Storrs, CT, 1995.

[3] Y. Bar-Shalom, X. R. Li and T. Kirubarajan
Estimation with Applications to Tracking and Navigation.
New York: Wiley, 2001.

[4] S. Blackman and R. Popoli
Design and Analysis of Modern Tracking Systems.
Norwod, MA: Artech House, 1999.

[5] D. Blodgett, S. Paquet, P. Plamondon, B. Chaïb-draa and P. Kropf
Coordinating plans for agents performing AAW hardkill and softkill for frigates.
In *Proceedings of The 2001 AAAI Fall Symposium Series*, North Falmouth, MA, July 2001, 1–8.

[6] D. Blodgett, P. Plamondon, Cha¨b-draa, P. Kropf and É. Bossé
A method to optimize ship maneuvers for the coordination of hardkill and softkill weapons within a frigate.
In *7th International Command and Control Research and Technology Symposium* (7th ICCRTS), Quebec, Sept. 2002.

[7] C. Brown, P. Fagan, A. Hepplewhite, B. Irving, D. Lane and E. Squire
Real-time decision support for the anti-air warfare commander.
In *6th International Command and Control Research and Technology Symposium* (6th ICCRTS), Annapolis, MD, 2001.

[8] B. A. Chalmers
Supporting threat response management in a tactical naval environment.
In *8th ICCRTS*, June 2003.

[9] H. Irandoust and A. Benaskeur
Argumentation-based decision support in naval tewa context.
In *6th Workshop on Computational Models of Natural Argument*, Riva del Garda, Italy, Aug. 2006.

[10] M. Kalandros and L. Y. Pao
Controlling target state estimate in centralized multisensor systems.
In *Proceedings of the 1998 American Control Conference*, Philadelphia, PA, June 1998, 2749–2753.

[11] M. Kalandros and L. Y. Pao
Sensor management for tracking interacting targets.
In *Proceeding of the Workshop on Estimation, Tracking, Fusion and Control: A Tribute to Yaakov Bar-Shalom*, Monterey, CA, May 2001.

[12] H. T. Kauderer
Air-directed surface-to-air missile study methodology.
Johns Hopkins APL Tech. Dig., **21**, 2 (2000), 244–250.

[13] T. Liang
Getting the act together.
In *International Defence Review*, Netherland, 1995.

[14] T. Liang and K. D. Liem
Integrated naval air defence: Co-ordinating hardkill and softkill weapons.
In *International Defense Review 6*, 1992, 567–569.

[15] S. Paradis, A. Benaskeur, M. Oxenham and P. Cutler
Threat evaluation and weapons allocation in network-centric warfare.
In *Proceedings of Fusion 2005*, Philadelphia, PA, July 2005.

[16] S. Paradis and J. Roy
An architecture and a facility for the integration of all levels of data fusion.
In *Proceedings of the Third International Conference on Information Fusion* (Fusion 2000), Paris, July 2000.

[17] F. Rheaume and A. Benaskeur
Impact of the tactical picture quality on the fire control radar search-lock-on time.
Technical Report TR 2003-284, Defence R&D Canada–Valcartier, 2003.

[18] F. Rheaume and A. R. Benaskeur
Fire control-based adaptation in data fusion applications.
In *Proceedings of 10th International Conference on Information Fusion* (Fusion 2007), Quebec, July 2007.

[19] R. L. Rothrock and O. E. Drummond
Performance metrics for multiple-sensor multiple-target tracking.
In O. E. Drummond, Ed., *Proceedings of SPIE, Signal and Data Processing of Small Targets*, vol. 4048, presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, July 2000, 521–531.

[20] J. Roy, É. Bossé, and D. Dion
CASE_ATTI: An algorithm-level testbed for multi-sensor data fusion.
Technical Report DREV-R-9411, Defence R&D Canada–Valcartier, May 1995.

[21] SADM
Ship Air Defense Model user manual, BAE Systems Australia, http://www.sadm.biz.

[22] U.S. Navy Advancement Fire Controlman Course Manuals
Fire Controlman, Volume 2—Fire-Control Radar Fundamentals.
Naval Education and Training Porfessional Development and Technology Center, NAVEDTRA 14099, Oct. 2000.

[23] E. Waltz and J. Llinas
Multi-sensor Data Fusion.
Boston: Artech House, 1990.

Dr. Abderrezak Benaskeur holds a Ph.D. degree in control theory (2000) from Université Laval, Quebec City, a M.Sc. in computer engineering (1993) from Université Paris-XI/Orsay, and a B.Sc. degree (1992) in control theory from the University of Constantine.

He is a defence scientist at Defence R&D Canada–Valcartier since 2000. His research activities cover: military command and control, distributed combat power management in naval operations, planning and scheduling, control systems, and adaptive data fusion.

François Rhéaume obtained his B.Sc. degree in computer engineering (2001) and M.Sc. degree in electrical engineering (2002) from Université Laval.

He is a defence scientist at Defence R&D Canada–Valcartier since 2002. His research interests include adaptive data fusion, pattern recognition, military command and control and distributed combat power (sensor and weapons) management in naval operations.

Stéphane Paradis has been a defence scientist at DRDC–Valcartier since 1989. He is now heading the Intelligence and Information section at DRDC–Valcartier. His area of expertise is in decision support enabling technologies, which includes visualization and human-computer interaction techniques, for situation and threat assessment applied to maritime command and control operations. He has a strong expertise in information fusion, information and knowledge management, AI (specifically blackboard technologies, automated inference reasoning engines, neural networks) and modeling and simulation.

Misassociation Probability in M2TA and T2TA

JAVIER ARETA
YAAKOV BAR-SHALOM
RON ROTHROCK

This paper presents procedures to calculate the probability that the measurement or the track originating from an extraneous target will be (mis)associated with a target of interest for the cases of Nearest Neighbor and Global association. For the measurement-to-track (M2T) case, it is shown that these misassociation probabilities depend, under certain assumptions, on a particular—covariance weighted—norm of the difference between the targets' predicted measurements. For the Nearest Neighbor M2T association, the exact solution, obtained for the case of equal track covariances, is based on a noncentral chi-square distribution. An approximate solution is also presented for the case of unequal track prediction covariances. For the Global M2T association case an approximation is presented for the case of "similar" track covariances. In the general case of unequal track covariances where this approximation fails, a more complicated but exact method based on the inversion of the characteristic function is presented. The track-to-track (T2T) association case involves correlated random variables for which the exact probability density function is very hard to obtain. Moment matching approximations are used that provide very accurate results. The theoretical results, confirmed by Monte Carlo simulations, quantify the benefit of Global vs. Nearest Neighbor M2T association. These results are applied to problems of single sensor as well as centralized fusion architecture multiple sensor tracking.

Manuscript received October 3, 2007; revised May 7, 2008; released for publication June 16, 2008.

Refereeing of this contribution was handled by Yvo Boers.

Authors' addresses: J. Areta and Y. Bar-Shalom, Univ. of Connecticut, ECE Dept., Storrs, CT 06269-2157, E-mail: (areta@ee.uconn.edu), (ybs@ee.uconn.edu); R. Rothrock, Sparta, Arlington, VA 22203, E-mail: (Ron_Rothrock@rosslyn.sparta.com).

1557-6418/07/$17.00 © 2007 JAIF

1. INTRODUCTION

This paper deals with the closed form misassociation probability formula for measurement-to-track association (M2T) and track-to-track association (T2T). The emphasis of this work is in closely spaced targets, which is much more prevalent in the real world than association of clutter to target tracks. Thus clutter is not considered in the sequel. In the first sections we develop the procedure to calculate the probability that the measurement associated by a likelihood based assignment algorithm to a target of interest originates from another (extraneous) target as a function of the state estimates and covariances of the tracks. Both a Nearest Neighbor[1] (NN) as well as Global[2] (G) assignment are considered. An approximate procedure is developed for the T2T association, as a closed form of the probability density function is very hard to find, due to the existing correlation between the track estimates. These closed form expressions should be useful when the knowledge of the performance of a system is to be quantified, for example, in the selection of a radar given it accuracy and the expected scenarios it could encounter. Also as in [6], it could be used to predict the number of measurements needed to achieve a certain performance. The model used for the targets is deterministic—they are located at a certain separation distance in the measurement space, expressed in terms of the track state estimates mapped into the measurement space. The association problem[3] was investigated in [10] for a different model, namely, the targets were assumed randomly distributed (i.i.d. uniform in a hyperball of a sufficiently large radius). Extensive work on the association of tracks from two sources, using kinematic, feature and classification information was done in [14, 7]. In [9] a more complex T2TA problem accounting for registration errors and mismatch in the number of tracks is considered. To obtain meaningful results, the track model considered is simplistic, assuming isotropic errors of the same variance. The model considered here allows performance evaluation of association algorithms under more realistic conditions, namely, with arbitrary measurement prediction covariances and the results are expressed in terms of the target separation distance.

Section 2 formulates the M2T association problem. The calculation of the misassociation probability for a Nearest Neighbor association is described in Section 3

[1]Strictly speaking, this is local NN because it considers the association of only one measurement to a track at a time (see [1] Section 3.2), and the measurement/track with smaller association cost is assigned to it. The other measurement is associated to the remaining track.

[2]This considers simultaneously all the measurements and tracks so the assignment is chosen as the association pair with overall smaller cost, and, unlike the local NN, has a unique solution [11].

[3]Two-dimensional (2-D, also known as "single frame"), i.e., between two lists—the list of measurements from the latest scan/frame and the list of tracks. Association is called sometimes "correlation"; since correlation has a well defined meaning in probability/statistics, we will not use it for association.

for the situation where the innovation covariances of the two tracks are equal as well as a generalization for unequal innovation covariances case. Section 4 introduces the Global association criterion, and the probabilities of misassociation for the two innovation covariance cases are obtained. The case of T2T association is considered in Section 5. Simulation results presented in Section 6 compare the theoretical calculations with Monte Carlo runs. Conclusions are presented in Section 7.

2. FORMULATION OF THE M2T ASSOCIATION PROBLEM

The predicted measurements (at the current time, not indicated for simplicity) for the two targets are denoted as $\hat{z}_i$, with associated covariances S_i, $i = 1,2$. These covariances are detailed in the sequel.

The pdf of the measurement prediction from the target of interest, designated as 1, is

$$p(z_1) = \mathcal{N}(z_1; \hat{z}_1, S_1) \tag{1}$$

while the pdf of the measurement prediction from the extraneous target, designated as 2, is

$$p(z_2) = \mathcal{N}(z_2; \hat{z}_2, S_2). \tag{2}$$

It is assumed that the assignment algorithm, using the likelihood function (or likelihood ratio) as a criterion, will associate to target t the measurement whose likelihood of having originated from target t is the largest. The likelihood of measurement z_i having originated from target t is given by the pdf of a measurement originating from target (track) t—the predicted measurement pdf—evaluated at z_i, namely,

$$\begin{aligned}\Lambda_{it} &= P_{D_t}\mathcal{N}(z_i; \hat{z}_t, S_{it})\\ &\stackrel{\Delta}{=} P_{D_t}|2\pi S_{it}|^{-1/2}\exp[-\tfrac{1}{2}(z_i - \hat{z}_t)' S_{it}^{-1}(z_i - \hat{z}_t)]\end{aligned} \tag{3}$$

where P_{D_t} is the detection probability of target t, and

$$S_{it} = H_t P_t H_t' + R_i \tag{4}$$

where H_t is the measurement matrix for track t and R_i is the measurement noise covariance for z_i. Since this noise covariance can be a function of the SNR, it has the index of the measurement. The likelihood ratio of originating from this track vs. from (random) clutter is this likelihood function divided by a constant, which is the spatial density of the clutter, assumed Poisson distributed [4].

Thus the index of the measurement that will be associated with track t is [1]

$$\begin{aligned}i^*(t) &= \arg\max_i [P_{D_t}\mathcal{N}(z_i; \hat{z}_t, S_{it})]\\ &= \arg\min_i [(z_i - \hat{z}_t)' S_{it}^{-1}(z_i - \hat{z}_t) + \ln|2\pi S_{it}|].\end{aligned} \tag{5}$$

Note that the target detection probability does not appear in the final expression above because all the likelihoods of association with track t have the same multiplier. In the case that the innovation covariance matrices are equal[4]

$$S_{it} = S_t \qquad \forall \quad i \tag{6}$$

then the log of the determinant of the covariance matrix in (5) is the same for all i and

$$i^*(t) = \arg\min_i [(z_i - \hat{z}_t)' S_t^{-1}(z_i - \hat{z}_t)]. \tag{7}$$

Consequently, under assumption (6), for track 1 the associated measurement report (AMR) will be the one whose normalized (Mahalanobis) distance squared to $\hat{z}_1$, given by

$$D(z, \hat{z}_1) = (z - \hat{z}_1)' S_1^{-1}(z - \hat{z}_1) \tag{8}$$

is the smallest. This amounts to a "local Nearest Neighbor" (designated as NN) assignment. Therefore, the misassociation event ($\mathrm{MA}_{21}^{\mathrm{NN}}$) that the measurement from target 2 is assigned to track 1 (which represents target 1) occurs if

$$\{\mathrm{MA}_{21}^{\mathrm{NN}}\} \stackrel{\Delta}{=} \{D(z_2, \hat{z}_1) < D(z_1, \hat{z}_1)\}. \tag{9}$$

The analysis of misassociation in the case of a global assignment (G) will be presented later.

3. NEAREST NEIGHBOR M2T MISASSOCIATION

3.1. Equal Innovation Covariances

This section evaluates the probability of misassociation of the NN assignment technique which considers tracks independently under a simplifying assumption. Assuming that

$$z_1 \sim \mathcal{N}(\hat{z}_1, S_1) \tag{10}$$

the pdf of $D(z_1, \hat{z}_1)$ is chi-square with n_z (dimension of z) degrees of freedom (d.o.f.), to be denoted as

$$p_{D(z_1,\hat{z}_1)}(x) = \chi^2_{n_z}(x). \tag{11}$$

To obtain the pdf of the "competition," $D(z_2, \hat{z}_1)$, it is rewritten as

$$\begin{aligned}D(z_2, \hat{z}_1) &= (z_2 - \hat{z}_1)' S_1^{-1}(z_2 - \hat{z}_1)\\ &= (z_2 - \hat{z}_2 + \hat{z}_2 - \hat{z}_1)' S_1^{-1}(z_2 - \hat{z}_2 + \hat{z}_2 - \hat{z}_1).\end{aligned} \tag{12}$$

Note that the above contains, in addition to the deterministic quantity $\hat{z}_2 - \hat{z}_1$, the difference $z_2 - \hat{z}_2$. The latter is random with covariance S_2, but the quadratic form in (12) contains the matrix S_1.

As shown below, the pdf of (12) is noncentral chi-square if the matrix in the quadratic form is the covariance of $z_2 - \hat{z}_2$. Consequently, it will be first assumed

[4]This holds approximately when a single sensor tracks two close targets.

that[5]

$$S_1 = S_2 = S. \tag{13}$$

Using the Cholesky decomposition of S^{-1}

$$S^{-1} = (S^{-1/2})' S^{-1/2} \tag{14}$$

one can rewrite (12) as

$$D(z_2,\hat{z}_1) = [S^{-1/2}(z_2 - \hat{z}_2) + S^{-1/2}(\hat{z}_2 - \hat{z}_1)]' \times [S^{-1/2}(z_2 - \hat{z}_2) + S^{-1/2}(\hat{z}_2 - \hat{z}_1)]. \tag{15}$$

Denoting the n_z-vector

$$\xi_{21} \stackrel{\Delta}{=} [S^{-1/2}(z_2 - \hat{z}_2) + S^{-1/2}(\hat{z}_2 - \hat{z}_1)] \tag{16}$$

the distance (15) is its norm squared, i.e.,

$$D(z_2,\hat{z}_1) = \xi_{21}'\xi_{21} = \sum_{i=1}^{n_z} (\xi_{21}(i))^2. \tag{17}$$

Since

$$\mathrm{cov}[\xi_{21}] = S^{-1/2} S (S^{-1/2})' = I \tag{18}$$

the components $\xi_{21}(i)$ of ξ_{21} are independent Gaussian random variables with nonzero means and unity variance. Thus

$$\xi_{21}(i) \sim \mathcal{N}(\bar{\xi}_{21}(i), 1), \qquad i = 1,\ldots,n_z. \tag{19}$$

where $\bar{\xi}_{21}(i)$ it the i-th element of $[S^{-1/2}(\hat{z}_2 - \hat{z}_1)]$, $i = 1,\ldots,n_z$.

Consequently [15], the pdf of (17) is noncentral chi-square with n_z d.o.f. and non-centrality parameter

$$\begin{aligned}\lambda &= \sum_{1}^{n_z} ([S^{-1/2}(\hat{z}_2 - \hat{z}_1)]_i)^2 \\ &= [S^{-1/2}(\hat{z}_2 - \hat{z}_1)]'[S^{-1/2}(\hat{z}_2 - \hat{z}_1)] \\ &= (\hat{z}_2 - \hat{z}_1)' S^{-1} (\hat{z}_2 - \hat{z}_1).\end{aligned} \tag{20}$$

This pdf is denoted as

$$p_{D(z_2,\hat{z}_1)}(x) = \chi^2_{n_z,\lambda}(x). \tag{21}$$

The cumulative distribution function (cdf) corresponding to the above will be denoted as $X^2_{n_z,\lambda}(x)$ and a routine for its evaluation (to be needed below) is available from [5].

The probability of the misassociation event (9) is then given by

$$\begin{aligned}\mathsf{P}_{\mathrm{MA}_{21}^{\mathrm{NN}}} &= \mathsf{P}\{D(z_2,\hat{z}_1) < D(z_1,\hat{z}_1)\} \\ &= \int_0^\infty \mathsf{P}\{D(z_2,\hat{z}_1) < x\} p_{D(z_1,\hat{z}_1)}(x)dx \\ &= \int_0^\infty X^2_{n_z,\lambda}(x)\chi^2_{n_z}(x)dx.\end{aligned} \tag{22}$$

[5]For simplicity, the single indexing of covariances as in (1)–(2) is used in the sequel.

3.2. Unequal Innovation Covariances

This section evaluates the probability of misassociation of the NN assignment technique which considers tracks independently in the general case. While assumption (13) is somewhat limiting, it is not unreasonable to assume that two targets in the same neighborhood have the same state estimation covariance. If (13) is not satisfied, then (16) has to be replaced by

$$\zeta_{21} \stackrel{\Delta}{=} [S_1^{-1/2}(z_2 - \hat{z}_2) + S_1^{-1/2}(\hat{z}_2 - \hat{z}_1)] \tag{23}$$

and

$$D(z_2,\hat{z}_1) = \zeta_{21}'\zeta_{21} = \sum_{i=1}^{n_z} (\zeta_{21}(i))^2. \tag{24}$$

The covariance of (23) is

$$\mathrm{cov}[\zeta_{21}] = S_1^{-1/2} S_2 S_1^{-1/2} \neq I \tag{25}$$

i.e., its components are not independent anymore and (24) is not chi-square distributed. Consequently, one cannot use anymore (22) to evaluate the probability of the misassociation event (9).

In this case the exact distribution of (24) is needed. However, this is not known because the covariance of $z_2 - \hat{z}_2$ is S_2 but the norm is w.r.t. $S_1 \neq S_2$. A moment matching technique will be used to approximate its distribution.

Considering only the zero-mean part of (23), its norm squared is

$$D_0(z_2,\hat{z}_1) = [S_1^{-1/2}(z_2 - \hat{z}_2)]'[S_1^{-1/2}(z_2 - \hat{z}_2)] \tag{26}$$

and its mean is

$$\begin{aligned}E[D_0(z_2,\hat{z}_1)] &= E[[S_1^{-1/2}(z_2 - \hat{z}_2)]'[S_1^{-1/2}(z_2 - \hat{z}_2)]] \\ &= \mathrm{tr}[S_1^{-1}S_2].\end{aligned} \tag{27}$$

Under the equal covariance assumption, the above would have been equal to n_z (the trace of the $n_z \times n_z$ identity matrix, as in (18)).

Based on the above observation, we will scale $D(z_2,\hat{z}_1)$ to match its mean to what it would have been in the equal covariance case (i.e., n_z) as follows

$$D^*(z_2,\hat{z}_1) \stackrel{\Delta}{=} \alpha_{21} D(z_2,\hat{z}_1) \tag{28}$$

where

$$\alpha_{21} \stackrel{\Delta}{=} \frac{n_z}{\mathrm{tr}[S_1^{-1}S_2]} \tag{29}$$

and will approximate the distribution of $D^*(z_2,\hat{z}_1)$ as noncentral chi-square with n_z d.o.f. and noncentrality parameter

$$\begin{aligned}\lambda^* &= \alpha_{21} \sum_{i=1}^{n_z} ([S_1^{-1/2}(\hat{z}_2 - \hat{z}_1)]_i)^2 \\ &= \alpha_{21}(\hat{z}_2 - \hat{z}_1)' S_1^{-1} (\hat{z}_2 - \hat{z}_1).\end{aligned} \tag{30}$$

The probability of the misassociation event (9) is then given by

$$\begin{aligned}
\mathsf{P}_{\mathrm{MA}_{21}^{\mathrm{NN}}} &= \mathsf{P}\{D(z_2,\hat{z}_1) < D(z_1,\hat{z}_1)\} \\
&= \mathsf{P}\left\{\frac{D^*(z_2,\hat{z}_1)}{\alpha_{21}} < D(z_1,\hat{z}_1)\right\} \\
&= \int_0^\infty \mathsf{P}\{D^*(z_2,\hat{z}_1) < \alpha_{21}x\} p_{D(z_1,\hat{z}_1)}(x)dx \\
&= \int_0^\infty X^2_{n_z,\lambda^*}(\alpha_{21}x)\chi^2_{n_z}(x)dx. \qquad (31)
\end{aligned}$$

Since the above is an approximation, its quality will be evaluated via Monte Carlo runs in Section 6.

4. M2T MISASSOCIATION IN A GLOBAL ASSIGNMENT

If a Global assignment [2] is used, then a misassociation (swap of measurements for two targets) occurs if

$$D_{21} + D_{12} < D_{11} + D_{22} \qquad (32)$$

since, as in (5), the covariance determinants cancel. This assumes the same noise covariances for the two measurements and the same detection probabilities for the two targets. The evaluation of the probability of this event is done next. Note that the distance term $D(z_i,\hat{z}_j)$ is now noted as D_{ij}, for simplicity. No "gating" [1] (i.e., infinite gating threshold) is assumed because it would require truncated pdfs, which would complicate the analysis and would make little difference in the results because the gates are, typically, above 99%.

The inequality (32) is rewritten so that the random variables in it (z_1, z_2) are each on one side only, namely,

$$D_{21} - D_{22} < D_{11} - D_{12}. \qquad (33)$$

The l.h.s. of the above is

$$\begin{aligned}
D_{21} - D_{22} &= (z_2 - \hat{z}_1)'S_1^{-1}(z_2 - \hat{z}_1) - (z_2 - \hat{z}_2)'S_2^{-1}(z_2 - \hat{z}_2) \\
&= z_2'(S_1^{-1} - S_2^{-1})z_2 - z_2'S_1^{-1}\hat{z}_1 - \hat{z}_1'S_1^{-1}z_2 \\
&\quad + z_2'S_2^{-1}\hat{z}_2 + \hat{z}_2'S_2^{-1}z_2 + \hat{z}_1'S_1^{-1}\hat{z}_1 - \hat{z}_2'S_2^{-1}\hat{z}_2. \qquad (34)
\end{aligned}$$

As before, this equation should be analyzed for the case of equal covariances, where cancelation of the first term occurs, and for different covariances.

4.1. Global Assignment Misassociation with Equal Innovation Covariances

When both covariances are equal,[6] the quadratic term in (34) vanishes, thus the distribution of $D_{21} - D_{22}$ is Gaussian since

$$\begin{aligned}
D_{21} - D_{22} &= -2(\hat{z}_1 - \hat{z}_2)'S^{-1}z_2 + \hat{z}_1'S^{-1}\hat{z}_1 - \hat{z}_2'S^{-1}\hat{z}_2 \\
&\stackrel{\Delta}{=} c'z_2 + b \stackrel{\Delta}{=} \Delta_{21}. \qquad (35)
\end{aligned}$$

Similarly, the term $D_{11} - D_{12}$ can be written as

$$\begin{aligned}
D_{11} - D_{12} &= -2(\hat{z}_1 - \hat{z}_2)'S^{-1}z_1 + \hat{z}_1'S^{-1}\hat{z}_1 - \hat{z}_2'S^{-1}\hat{z}_2 \\
&\stackrel{\Delta}{=} c'z_1 + b \stackrel{\Delta}{=} \Delta_{12}. \qquad (36)
\end{aligned}$$

Note that these two Gaussian random variables (RV) are independent because each depends on only one of the measurements. Thus, as they are Gaussian and independent, their difference $\Delta_{21} - \Delta_{12}$ is also Gaussian

$$\Delta_{21} - \Delta_{12} \sim \mathcal{N}(2(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2), 8(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)) \qquad (37)$$

so the probability of the misassociation event $\mathrm{MA}^{\mathrm{G}}_{21,12} = \{\Delta_{21} < \Delta_{12}\}$ can be calculated in terms of the cumulative density function Φ of a standard Normal random variable as

$$\begin{aligned}
\mathsf{P}_{\mathrm{MA}^{\mathrm{G}}_{21,12}} &= \mathsf{P}\{\Delta_{21} - \Delta_{12} < 0\} \\
&= \mathsf{P}\left\{\frac{\Delta_{21} - \Delta_{12} - 2(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)}{[8(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)]^{1/2}} \right. \\
&\qquad \left. < \frac{-2(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)}{[8(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)]^{1/2}}\right\} \\
&= \Phi\left(-\left[\frac{(\hat{z}_1 - \hat{z}_2)'S^{-1}(\hat{z}_1 - \hat{z}_2)}{2}\right]^{1/2}\right). \qquad (38)
\end{aligned}$$

4.2. Global Assignment Misassociation with Unequal Innovation Covariances

In the case that the covariances are different, the quadratic term in (34) does not vanish. Its contribution is higher when the covariance matrices are very different. Thus two approaches to approximate the distribution of the difference will be investigated. The first one approximates the distribution as a noncentral chi-square, using the first two moments, and is expected to provide better results when the quadratic term dominates and has positive eigenvalues. The other approach is to fit a Gaussian with matched mean and variance, and is expected to work better when the contribution of the quadratic term is small.

4.2.1. Moment matching approaches

For the first approach, denote

$$A_{21} \stackrel{\Delta}{=} S_1^{-1} - S_2^{-1} \qquad (39)$$

$$b_{21} \stackrel{\Delta}{=} S_1^{-1}\hat{z}_1 - S_2^{-1}\hat{z}_2 \qquad (40)$$

$$c_{21} \stackrel{\Delta}{=} \hat{z}_1'S_1^{-1}\hat{z}_1 - \hat{z}_2'S_2^{-1}\hat{z}_2 \qquad (41)$$

[6] A similar approach has been taken in [10] assuming, however, random location of the targets according to a spatial Poisson process. In our case the probability of error is a function of a normalized distance between the targets.

one has, by completing the quadratic form

$$\begin{aligned}D_{21} - D_{22} &= z_2' A_{21} z_2 - z_2' b_{21} - b_{21}' z_2 + c_{21} \\ &= (z_2 - A_{21}^{-1} b_{21})' A_{21} (z_2 - A_{21}^{-1} b_{21}) \\ &\quad - b_{21}' A_{21}^{-1} b_{21} + c_{21}.\end{aligned} \tag{42}$$

Using the following notation

$$d_{21} \triangleq -b_{21}' A_{21}^{-1} b_{21} + c_{21} \tag{43}$$

$$G_{21} \triangleq (z_2 - A_{21}^{-1} b_{21})' A_{21} (z_2 - A_{21}^{-1} b_{21}) \tag{44}$$

expression (34) becomes

$$D_{21} - D_{22} = G_{21} + d_{21}. \tag{45}$$

Note that (44) can be rewritten as

$$G_{21} \triangleq (z_2 - \hat{z}_2 + \hat{z}_2 - A_{21}^{-1} b_{21})' A_{21} (z_2 - \hat{z}_2 + \hat{z}_2 - A_{21}^{-1} b_{21}) \tag{46}$$

which would be exactly noncentral chi-square distributed if the matrix in the quadratic form would have been the covariance of z_2. As in (28), define

$$G_{21}^* \triangleq \beta_{21} G_{21} \tag{47}$$

where

$$\beta_{21} \triangleq \frac{n_z}{\mathrm{tr}(A_{21} S_2)}. \tag{48}$$

Then, G_{21}^* is approximately (by moment matching) noncentral chi-square distributed with n_z d.o.f. and noncentrality parameter

$$\lambda_{21}^* \triangleq \beta_{21} (\hat{z}_2 - A_{21}^{-1} b_{21})' A_{21} (\hat{z}_2 - A_{21}^{-1} b_{21}). \tag{49}$$

This is written as

$$G_{21}^* \sim \chi^2_{n_z, \lambda_{21}^*}. \tag{50}$$

A similar definition yields G_{12}^*, which is the negative of the r.h.s. of (33). The pdf of G_{12}^* is the same as in (50) with the indices 1 and 2 switched. Furthermore, G_{21}^* and G_{12}^* are independent.

The misassociation event for a Global assignment between tracks 1 and 2 is thus

$$\mathrm{MA}^{\mathrm{G}}_{21,12} = \left\{ \frac{G_{21}^*}{\beta_{21}} + d_{21} < -\left[\frac{G_{12}^*}{\beta_{12}} + d_{12} \right] \right\}. \tag{51}$$

The probability of the above is then obtained as

$$\mathsf{P}_{\mathrm{MA}^{\mathrm{G}}_{21,12}} = \int_0^\infty X^2_{n_z, \lambda_{21}^*} \left[-\beta_{21} \left(\frac{x}{\beta_{12}} + d_{12} + d_{21} \right) \right] \chi^2_{n_z, \lambda_{12}^*}(x) dx. \tag{52}$$

Note from (43) that $d_{12} = -d_{21}$ and thus (52) becomes

$$\mathsf{P}_{\mathrm{MA}^{\mathrm{G}}_{21,12}} = \int_0^\infty X^2_{n_z, \lambda_{21}^*} \left[-\frac{\beta_{21}}{\beta_{12}} x \right] \chi^2_{n_z, \lambda_{12}^*}(x) dx. \tag{53}$$

For the second approach, the mean and variance of the differences[7] $\Delta_2 \triangleq D_{21} - D_{22}$ and $\Delta_1 \triangleq D_{11} - D_{12}$ are required to approximate their distributions by a Gaussian pdfs. From (42) we have that $D_{21} - D_{22}$ is a quadratic expression in z_2, and similarly for $D_{11} - D_{12}$, thus

$$\Delta_2 = D_{21} - D_{22} = z_2' A_{21} z_2 - 2 b_{21} z_2 + c_{21} \tag{54}$$

$$\Delta_1 = D_{11} - D_{12} = z_1' A_{21} z_1 - 2 b_{21} z_1 + c_{21}. \tag{55}$$

Note that the above two RVs are independent. Using the results in the appendix showing the variance of a quadratic form, one has (approximately)

$$\Delta_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{56}$$

where

$$\mu_i = \mathrm{tr}(A_{21} S_i) + \hat{z}_i' A_{21} \hat{z}_i - 2 b_{21}' \hat{z}_i + c_{21} \tag{57}$$

$$\sigma_i^2 = 2\mathrm{tr}(A_{21} S_i A_{21} S_i) + 4[A_{21} \hat{z}_i - b_{21}]' S_i [A_{21} \hat{z}_i - b_{21}]. \tag{58}$$

As in Section 4.1, these two Gaussian random variables are independent, so the misassociation probability can be calculated in terms of the cumulative density function Φ of a standard Normal random variable as

$$\begin{aligned}\mathsf{P}_{\mathrm{MA}^{\mathrm{G}}_{21,12}} &= \mathsf{P}\{\Delta_2 - \Delta_1 < 0\} \\ &= \mathsf{P}\left\{ \frac{\Delta_2 - \Delta_1 - \mu_2 + \mu_1}{(\sigma_1^2 + \sigma_2^2)^{1/2}} < \frac{\mu_1 - \mu_2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} \right\} \\ &= \Phi\left(\frac{\mu_1 - \mu_2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} \right).\end{aligned} \tag{59}$$

4.2.2. Global assignment misassociation with unequal innovation covariances—exact solution

The approximation methods of the previous subsection will be shown to work well when the covariances are not very different, that is, when the matrix $A_{21} = S_1^{-1} - S_2^{-1}$ has small eigenvalues compared to the values of the element in b_{21}. If this is not the case, something that happens when S_1 greatly differs from S_2, the distribution of the quadratic form is not easy to approximate, and a method to obtain the true distribution is required.

In the following subsection a method to numerically obtain the cdf of a noncentral quadratic function will be delineated, following [8]. Analytical expressions have been derived via a series representation, similarly to [12], for the case of real Gaussian random variables, but the convergence of such series is not assured, and hence only the numerical integration method is presented.

[7]Different notations are used than in Section 4.1 because their expressions are different.

The quadratic function of Gaussian random variables (42) is replicated here without subscripts for clarity

$$Q(Z) = Z'AZ + Z'b + b'Z + c$$
$$= (Z + A^{-1}b)'A(Z + A^{-1}b) - b'A^{-1}b + c \tag{60}$$

where the N-vector Z is Gaussian

$$Z \sim \mathcal{N}(\mu, \Sigma). \tag{61}$$

Neglecting the constant, (60) can be expressed as a weighted sum of noncentral chi-square random variables by writing Z in terms of $X \sim \mathcal{N}(0, I)$

$$Z = \Sigma^{1/2}X + \mu \tag{62}$$

as

$$Q(Z) = [X + \Sigma^{-1/2}(\mu + A^{-1}b)]'[\Sigma^{1/2}]'$$
$$\times A\Sigma^{1/2}[X + \Sigma^{-1/2}(\mu + A^{-1}b)]$$
$$= \sum_{k=1}^{N} \lambda_k \chi^2_{1,\rho_k^2} \tag{63}$$

where the matrix $[\Sigma^{1/2}]'A\Sigma^{1/2}$ has distinct eigenvalues λ_k and eigenvector matrix T, and

$$\rho \stackrel{\Delta}{=} [\rho_1 \ldots \rho_N]' = T\Sigma^{-1/2}(\mu + A^{-1}b). \tag{64}$$

The characteristic function of $Q(Z)$ is

$$\phi(t) = \prod_{k=1}^{N}(1 - 2i\lambda_k t)^{-1/2} \exp\left(i\sum_{r=1}^{N} \frac{\rho_r^2 \lambda_r t}{1 - 2i\lambda_r t}\right). \tag{65}$$

This function can be inverted as in [8], yielding

$$P(Q < q) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u)}{u\kappa(u)} du \tag{66}$$

where

$$\theta(u) = \frac{1}{2}\sum_{k=1}^{N}\left[\tan^{-1}(\lambda_k u) + \frac{\rho_k^2 \lambda_k u}{1 + \lambda_k^2 u^2}\right] - \frac{qu}{2} \tag{67}$$

$$\kappa(u) = \prod_{k=1}^{N}(1 + \lambda_k^2 u^2)^{1/4} \exp\left(\frac{1}{2}\sum_{r=1}^{N}\frac{\rho_r^2 \lambda_r^2 u^2}{1 + \lambda_r^2 u^2}\right). \tag{68}$$

Also the probability density function is obtained as

$$f(q) = \frac{1}{\pi}\int_0^\infty \frac{\cos\theta(u)}{\kappa(u)} du. \tag{69}$$

These integrals can be truncated to some finite upper limit since $\kappa(u)$ is an increasing function in u, and the integral is approximated using Simpson's rule, following the suggestions in [8].

Having both the pdf and cdf of the quadratic functions, the exact probability of misassociation is obtained as in the previous sections. While this characteristic function based procedure is exact, unlike the moment matching procedure from the previous subsection, it is more costly. However, the moment matching procedure is accurate enough in certain circumstances, to be specified in the next section.

5. FORMULATION OF THE T2T ASSOCIATION PROBLEM

In this section we consider the case of two sensors m, n, generating track estimates from two targets i, j. The track estimate of target i generated by sensor m after receiving the measurement at time k is noted as $\hat{x}_i^m(k \mid k)$. Some authors [6, 13] have tried to obtain the T2T misassociation probability, but without considering the correlation of the estimation errors (due to the common process noise [1]) and have also considered only a single track, thus no global association results have been reported.

In the case of Gaussian measurements, the log likelihood ratio for the common origin association consists of two terms. One is the normalized distance squared (NDS) and the other is a ratio of the determinant of the innovation covariance matrix and the density μ_{ex} of extraneous tracks [4]. In the case of tracks $\hat{x}_j^n(k \mid k)$ and $\hat{x}_i^m(k \mid k)$ the NDS takes the form

$$D_{ij}^{mn}(k) = (\hat{x}_i^m(k \mid k) - \hat{x}_j^n(k \mid k))'[T_{ij}^{mn}(k)]^{-1}(\hat{x}_i^m(k \mid k) - \hat{x}_j^n(k \mid k)) \tag{70}$$

where $T_{ij}^{mn}(k)$ is the track difference covariance, which will be given later.

The likelihood based association cost for tracks $\hat{x}_j^n$ and $\hat{x}_i^m$ is the negative log likelihood ratio (NLLR)

$$C_{ij}^{mn} = D_{ij}^{mn} + \ln(|2\pi T_{ij}^{mn}|^{1/2}/\mu_{ex}). \tag{71}$$

The superscripts will be dropped when possible as we are checking for association between tracks originated from sensors m and n only and not across time. Also the time index k will be dropped, for brevity, thus for example $T_{ij}^{mn}(k)$ becomes T_{ij}.

5.1. Nearest Neighbor T2T Association Criterion

Using the NLLR cost as a modified distance definition, the Nearest Neighbor (NN) misassociation event ($\mathrm{MA}_{ij}^{\mathrm{NN}}$) that the estimate of target i obtained by sensor m is assigned to track j from sensor n instead of being assigned to track i from sensor n, is defined by

$$\{\mathrm{MA}_{21}^{\mathrm{NN}}\} \stackrel{\Delta}{=} \{C_{ij} < C_{ii}\}. \tag{72}$$

Consider each of these cost terms separately, and note that each estimate can be expressed as its true value plus the error term $\hat{x}_i^m(k \mid k) = x_i(k) + \tilde{x}_i^m(k \mid k)$, where $x_i(k)$ is the true state of target i (regardless of the sensor).

For the term C_{ii}, i.e., the cost of associating the tracks corresponding to target i obtained at the two sensors, the covariance matrix

$$\begin{aligned} T_{ii} &= E\{(\tilde{x}_i^m - \tilde{x}_i^n)'(\tilde{x}_i^m - \tilde{x}_i^n)\} \\ &= E\{(\tilde{x}_i^m)'\tilde{x}_i^m\} + E\{(\tilde{x}_i^n)'\tilde{x}_i^n\} \\ &\quad - E\{(\tilde{x}_i^m)'\tilde{x}_i^n\} - E\{(\tilde{x}_i^n)'\tilde{x}_i^m\} \\ &= P_i^m + P_i^n - P_i^{mn} - P_i^{nm} \end{aligned} \tag{73}$$

is required.

The autocovariance terms are obtained from the estimation algorithm (a Kalman filter in our case) and the crosscovariance is present due to the common process noise. This crosscovariance terms can be obtained in an iterative way [1] by the Lyapunov type equation

$$\begin{aligned} P_i^{mn}(k) &= E\{(\tilde{x}_i^m)'\tilde{x}_i^n\} \\ &= [I - W_i^m H_i^m][F P_i^{mn}(k-1)F' + Q][I - W_i^n H_i^n]'. \end{aligned} \tag{74}$$

The distance term in the cost can then be written as

$$\begin{aligned} D_{ii} &= (x_i + \tilde{x}_i^m - x_i - \tilde{x}_i^n)'[T_{ii}(k)]^{-1}(x_i + \tilde{x}_i^m - x_i - \tilde{x}_i^n) \\ &= (\tilde{x}_i^m - \tilde{x}_i^n)'[T_{ii}(k)]^{-1}(\tilde{x}_i^m - \tilde{x}_i^n). \end{aligned} \tag{75}$$

This random variable has χ^2 distribution, but its dependence on the other distance term D_{ij} (through $\tilde{x}_i^m$) precludes the usage of the results obtained for the M2T association.

The term C_{ij} is the cost of associating track i obtained from sensor m to the track j obtained by sensor n. In this case the covariance matrix T_{ij} is simply $P_i^m + P_j^n$, as the track errors are not correlated. Then, the distance D_{ij} can be written as

$$\begin{aligned} D_{ij} &= (x_j + \tilde{x}_j^m - x_i - \tilde{x}_i^n)'[T_{ij}(k)]^{-1}(x_j + \tilde{x}_j^m - x_i - \tilde{x}_i^n) \\ &= (\tilde{x}_i^m - \tilde{x}_j^n - c)'[T_{ij}(k)]^{-1}(\tilde{x}_i^m - \tilde{x}_j^n - c) \end{aligned} \tag{76}$$

where $c = x_i - x_j$ is the separation between tracks.

The exact probability of misassociation can be obtained as

$$\begin{aligned} \mathsf{P}_{\mathrm{MA}_{ij}^{\mathrm{NN}}} &= \mathsf{P}\{C_{ij} < C_{ii}\} \\ &= \mathsf{P}\{D_{ij} - D_{ii} + \gamma_{ij} < 0\} \\ &= \int_{\mathcal{A}} [(\tilde{x}_i^m - \tilde{x}_j^n - c)'[T_{ij}(k)]^{-1}(\tilde{x}_i^m - \tilde{x}_j^n - c) \\ &\qquad - (\tilde{x}_i^m - \tilde{x}_i^n)'[T_{ii}(k)]^{-1}(\tilde{x}_i^m - \tilde{x}_i^n) + \gamma_{ij}] \\ &\qquad \times p(\tilde{x}_i^m, \tilde{x}_i^n, \tilde{x}_j^m) d\tilde{x}_i^m d\tilde{x}_i^n d\tilde{x}_j^m \end{aligned} \tag{77}$$

where

$$\mathcal{A} = \{\{\tilde{x}_i^n \tilde{x}_i^m \tilde{x}_j^n\} : (D_{ij} - D_{ii}) < 0\} \tag{78}$$

and

$$\begin{aligned} \gamma_{ij} &= \ln(|2\pi T_{ij}^{mn}|^{1/2}/\mu_{ex}) - \ln(|2\pi T_{ij}^{mn}|^{1/2}/\mu_{ex}) \\ &= \ln(|T_{ij}^{mn}|^{1/2}/|T_{ii}^{mn}|^{1/2}). \end{aligned} \tag{79}$$

The integration region $\mathcal{A}$ is very difficult to find, and this does preclude the usage of (77) for the calculation of the misassociation probability. Another approach is to obtain the pdf of the cost difference, but the fact that two of the estimates are correlated makes the exact pdf calculation very complex. These are the reasons that lead to the moment matching approach technique described next.

As the distance formulas involve quadratic terms, closed form first and second order moments can be obtained, which depend only on the correlation matrices involved. So, if we define

$$\eta_1 = D_{ij} - D_{ii} + \gamma_{ij} \tag{80}$$

the first and second moments of η_1, μ_{η_1} and $\sigma_{\eta_1}^2$, are obtained using (133) and (134) in the appendix. These moments are used to match both a Gaussian distribution as well as a shifted chi-square distribution, to obtain approximate misassociation probabilities.

The Gaussian approximation $\xi_1 \approx \eta_1$ follows by defining

$$\xi_1 \sim \mathcal{N}(\mu_{\eta_1}, \sigma_{\eta_1}^2). \tag{81}$$

Then the approximate misassociation probability is given by

$$\mathsf{P}^{\mathrm{G}}_{\mathrm{MA}_{ij}^{\mathrm{NN}}} = \Phi\left(\frac{\mu_{\eta_1}}{\sigma_{\eta_1}}\right) \tag{82}$$

where $\Phi(\cdot)$ is the normal standard cumulative distribution function.

The chi-square approximation $\zeta_1 \approx \eta_1$ is based on the definition of the random variable ζ_1 in terms of a shifted chi-square random variable w with k degrees of freedom

$$\zeta_1 = \varrho_1 + w \tag{83}$$

where

$$\varrho_1 = \mu_{\eta_1} - \sigma_{\eta_1}^2/2 \tag{84}$$

$$k = \sigma_{\eta_1}^2/2 \tag{85}$$

and $X_k^2(\cdot)$ is the cumulative distribution function for a chi-square random variable with k degrees of freedom. Then the approximate misassociation probability is given by

$$\mathsf{P}^{\chi^2}_{\mathrm{MA}_{ij}^{\mathrm{NN}}} = X_k^2(\varrho_1). \tag{86}$$

5.2. Global T2T Association Criterion

Similarly to the NN case, the misassociation event $\mathrm{MA}_{ij}^{\mathrm{G}}$ that the estimate of target i and j obtained by sensor m are respectively assigned using a global approach to tracks j and i from sensor n instead of being assigned to tracks i and j from sensor n, is defined by

$$\{\mathrm{MA}^{\mathrm{G}}\} \stackrel{\Delta}{=} \{C_{ij} + C_{ji} < C_{ii} + C_{jj}\}. \tag{87}$$

The exact probability of misassociation is again very difficult to obtain, thus a moment matching approach to a Gaussian and a shifted chi-square random variable will be used to obtain two approximations. Defining

$$\eta_2 = D_{ij} + D_{ji} - D_{ii} - D_{jj} + \gamma \tag{88}$$

where

$$\gamma = \ln(|T_{ij}^{mn}|^{1/2}|T_{ji}^{mn}|^{1/2}/(|T_{ii}^{mn}|^{1/2}|T_{jj}^{mn}|^{1/2})). \tag{89}$$

The first and second moments of η_2, μ_{η_2} and $\sigma_{\eta_2}^2$, are obtained using (140) and (141) in the appendix.

The Gaussian approximation $\xi_2 \approx \eta_2$ is, as before,

$$\xi_2 \sim \mathcal{N}(\mu_{\eta_2}, \sigma_{\eta_2}^2) \tag{90}$$

and this yields the approximate misassociation probability as

$$\mathsf{P}_{\mathrm{MA^G}}^{\mathcal{N}} = \Phi\left(\frac{\mu_{\eta_2}}{\sigma_{\eta_2}}\right). \tag{91}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The chi-square approximation $\zeta_2 \approx \eta_2$ is also as before based on the definition of the random variable ζ_2 in terms of a shifted chi-square random variable ς with k degrees of freedom

$$\zeta_2 = \varrho_2 + \varsigma \tag{92}$$

where

$$\varrho_2 = \mu_{\eta_2} - \sigma_{\eta_2}^2/2 \tag{93}$$

$$k = \sigma_{\eta_2}^2/2. \tag{94}$$

Denoting $X_k^2(\cdot)$ to the cumulative distribution function of a chi-square random variable with k degrees of freedom, the approximate misassociation probability is given by

$$\mathsf{P}_{\mathrm{MA^G}}^{\chi^2} = X_k^2(\varrho_2). \tag{95}$$

6. SIMULATION RESULTS

A number of cases with unequal covariances are considered to compare the techniques for the Nearest Neighbor of Section 3 to Monte Carlo results. As a limiting case, the equal covariance situation is also considered.

Two targets, moving in 3 dimensions, are considered. Their motion is modeled by a NCV (nearly constant velocity) model [3] in each Cartesian coordinate with Gaussian zero-mean white process noise with PSD (power spectral density) $\tilde{q}$, uncorrelated across the coordinates. Position measurements z_s are obtained with probability of detection one at sampling intervals of T in spherical ("s") coordinates (range, azimuth and elevation) with additive Gaussian zero-mean white noise with covariance

$$R_s = \mathrm{diag}[\sigma_r^2,\ \sigma_a^2,\ \sigma_e^2]. \tag{96}$$

These measurements are converted into Cartesian coordinates ("C") in the standard manner, resulting in z_C, with covariance matrix (at a particular position x_p relative to the radar), denoted as $R_C(x_p)$. The tracking filter is then linear [3]. No clutter or false measurements are considered in this work. In an actual tracking algorithm the covariance of the converted measurements is evaluated at the predicted position or at the measurement itself, whichever is more accurate. Here the converted measurement covariances (for the two targets) will be evaluated at the predicted locations of the corresponding measurements, $\hat{z}_{C_t}$, $t = 1,2$, which will be also a parameter of the evaluation to be carried out. These predicted measurements will quantify the separation between the targets.

To simulate a case where the measurement covariances are unequal, it is assumed that target t was observed n_t times, $t = 1,2$. These will yield the innovation covariances S_t, $t = 1,2$ (in Cartesian coordinates, denoted only with the target subscript for simplicity). With the values of $\hat{z}_{C_t}$, and S_t, $t = 1,2$, a random number generator will be used to generate the measurements

$$z_t \sim \mathcal{N}(\hat{z}_{C_t}, S_t), \qquad t = 1,2. \tag{97}$$

Let $z_t(j)$ denote the measurements in Monte Carlo run j, $j = 1,\ldots,N$. Using these, denote the indicator variable of the misassociation event (9) as

$$\chi(j) = \begin{cases} 1 & \text{if} \quad D(z_2,\hat{z}_1) < D(z_1,\hat{z}_1) \\ 0 & \text{otherwise} \end{cases} \tag{98}$$

where the distances $D(z,\hat{z})$ are defined in (12). The theoretical probability that the above indicator will be unity (i.e., a misassociation occurs) is given by (31), to be denoted now as P.

Thus the test whether the theoretical probability matches the outcomes (98) will be based on the statistic

$$\hat{\mathsf{P}} = \frac{1}{N}\sum_{j=1}^{N}\chi(j) \tag{99}$$

and, in order for P to be acceptable,[8] one has to have (see, e.g., [3], Sect. 2.6.4)

$$|\hat{\mathsf{P}} - \mathsf{P}| < 2\sqrt{\frac{\mathsf{P}(1-\mathsf{P})}{N}} \tag{100}$$

or,

$$|\hat{\mathsf{P}} - \mathsf{P}| < 2\sqrt{\frac{\hat{\mathsf{P}}(1-\hat{\mathsf{P}})}{N}} \tag{101}$$

based on the 95% probability region.

[8]Both $\hat{\mathsf{P}}$ and P will be subscripted later by NN for the Nearest Neighbor assignment misassociation and by G for the Global assignment.

TABLE I
The M2T Misassociation Probabilities for Various Covariances and Separations (Scenario 1) for the Nearest Neighbor Method

n_1	n_2	$S_1/10^4$			$S_2/10^4$			c	δ	$\mathsf{P}_{\mathrm{MA}^{\mathrm{NN}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{NN}}}$
		1.202	−1.182	−0.011	1.202	−1.182	−0.012	.03	0.306	.490	.495
30	30	−1.182	1.202	−0.013	−1.182	1.202	−0.011	.1	1.019	.400	.400
		−0.011	−0.013	2.385	−0.012	−0.011	2.385	.3	3.058	.063	.065
		1.202	−1.182	−0.011	1.372	−1.352	−0.039	.03	0.305	.450	.455
30	10	−1.182	1.202	−0.013	−1.352	1.374	−0.012	.1	1.017	.372	.369
		−0.011	−0.013	2.385	−0.039	−0.012	2.815	.3	3.051	.065	.058
		1.202	−1.182	−0.011	1.810	−1.794	−0.085	.03	0.267	.322	.328
30	5	−1.182	1.202	−0.013	−1.794	1.845	−0.062	.1	0.892	.275	.273
		−0.011	−0.013	2.385	−0.085	−0.062	4.164	.3	2.675	.069	.066

Note: The minor differences between the covariances of the two targets for $n_1 = n_2 = 30$ are due to the fact that the conversions from spherical to Cartesian coordinates of the measurement noise covariances (which amount to linearizations) were done at slightly different points, (109) and (110), respectively.

The following values for the parameters of the problem were considered:

$$\tilde{q} = 5\ \mathrm{m}^2/\mathrm{s}^3 \tag{102}$$

$$T = 1\ \mathrm{s} \tag{103}$$

$$\sigma_r = 10\ \mathrm{m} \tag{104}$$

$$\sigma_a = \sigma_e = 1\ \mathrm{mrad} \tag{105}$$

$$x_p = [10^5\ \ 10^5\ \ 10^3]'\ \mathrm{m} \tag{106}$$

$$n_1 = 30 \tag{107}$$

$$n_2 = 5; 10; 30 \tag{108}$$

$$\hat{z}_1 = x_p \tag{109}$$

$$\hat{z}_2 = x_p + c[10^2\ \ 10^2\ \ 10^2]'\ \mathrm{m}. \tag{110}$$

In the above c is a coefficient that yields several separations.

In Scenario 1 the same radar located at the origin of the coordinate system is tracking both targets. In Scenario 2, target 2 has been tracked (for n_2 samples) by another radar located at $[2 \cdot 10^5, 0, 0]$. In this case the ellipsoids corresponding to the two covariance matrices are approximately perpendicular.

A "normalized separation distance" between the two targets, denoted as δ, is evaluated, following (16), according to

$$\delta^2 \triangleq (\hat{z}_2 - \hat{z}_1)'[S_1^{-1/2}]'S_2^{-1/2}(\hat{z}_2 - \hat{z}_1). \tag{111}$$

This gives a measure of the closeness of the two targets for the case of different innovation covariances.

6.1. M2T Misassociation with Nearest Neighbor Assignment

Table I shows for the values of n_i listed above, the resulting covariances S_i, and for the three separations, defined by $c = 0.03; 0.1; 0.3$, the resulting normalized separation distances, the theoretical $\mathsf{P}_{\mathrm{MA}^{\mathrm{NN}}}$ (based on the non-central χ^2 distribution, presented in Section 4) as well as the average $\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{NN}}}$ from 1000 Monte Carlo runs. In all cases the differences between these probabilities is well within the limits given in (100) for $N = 1000$. Therefore the theoretical misassociation probabilities $\mathsf{P}_{\mathrm{MA}^{\mathrm{NN}}}$, as computed by the method presented in Section 4, are remarkably accurate.

Note that for decreasing n_2 the $\mathsf{P}_{\mathrm{MA}^{\mathrm{NN}}}$ for equivalent separation does also decrease. This seems counterintuitive, as larger number of measurements correspond to having more information available at the tracker, and thus a better (smaller) probability of misassociation is expected. The reason for this phenomenon is in the relative size of the innovation covariance matrices. When these are similar the chance of confusing the measurements is high as the measurements from both targets are within the same distance to any of tracks. Instead, if one of the ellipsoids is larger than the other, the distance of some of the measurements corresponding to this track (the ones occurring outside the smaller ellipsoid) to the center of the smaller covariance track will be larger, thus making it harder to misassociate them.

6.2. M2T Misassociation with Global Assignment

Table II shows the comparison of Nearest Neighbor association results with the approximate Global association misassignment probability evaluation algorithms from 4.2 (Gaussian fit and χ^2 fit via moment matching) with the results for Monte Carlo runs.

The χ^2 fit for the Global assignment approach does not work when the covariance matrices are very similar, as the noncentrality parameter depends on the inverse of the difference of these matrices. On the other hand, the Gaussian fit method does give accurate results over all the range of parameters, showing that for this Scenario the linear term dominates over the quadratic.

For the case of Scenario 2, the covariance ellipses S_1 and S_2 are quite different as a result of the location of the sensors during the initial estimation periods. In this case, prior to the current time when the association is considered, sensor 1 has been tracking target 1 only

TABLE II
The M2T Misassociation Probabilities for Various Covariances and Separations using Both the Global and Nearest Neighbor Approaches (Scenario 1)

n_1	n_2	c	$\mathsf{P}_{\mathrm{MA}^{\mathrm{NN}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{NN}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{G}}}$	$\mathsf{P}^{\mathcal{N}}_{\mathrm{MA}^{\mathrm{G}}}$	$\mathsf{P}^{\chi^2}_{\mathrm{MA}^{\mathrm{G}}}$
		.03	.490	.495	.410	.416	N/A
30	30	.1	.400	.400	.235	.235	N/A
		.3	.063	.065	.016	.015	N/A
		.03	.450	.455	.401	.399	.126
30	10	.1	.372	.369	.230	.232	N/A
		.3	.065	.058	.016	.015	N/A
		.01	.322	.328	.316	.312	.312
30	5	.1	.275	.273	.224	.235	.231
		.3	.069	.066	.029	.039	.032

TABLE IV
The T2T Misassociation Probabilities for Different Track Accuracies and Separations using Both the Global and Nearest Neighbor Approaches

n_1	n_2	c	$\mathsf{P}^{\mathcal{N}}_{\mathrm{MA}^{\mathrm{NN}}}$	$\mathsf{P}^{\chi^2}_{\mathrm{MA}^{\mathrm{NN}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{NN}}}$	$\mathsf{P}^{\mathcal{N}}_{\mathrm{MA}^{\mathrm{G}}}$	$\mathsf{P}^{\chi^2}_{\mathrm{MA}^{\mathrm{G}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{G}}}$
		.01	.484	.553	.480	.464	.476	.464
10	10	.2	.236	.251	.225	.123	.112	.100
		.4	.043	.026	.028	.073	.011	.015
		.01	.345	.396	.321	.258	.269	.240
30	30	.2	.252	.274	.233	.162	.161	.149
		.4	.116	.100	.096	.055	.035	.034
		.01	.466	.542	.457	.344	.394	.339
10	50	.2	.344	.394	.339	.236	.244	.227
		.4	.165	.161	.152	.066	.049	.052

TABLE III
The M2T Misassociation Probabilities for Various Covariances and Separations using the Global and Nearest Neighbor Approaches (Scenario 2)

n_1	n_2	$S_1/10^4$			$S_2/10^4$			c	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{NN}}}$	$\hat{\mathsf{P}}_{\mathrm{MA}^{\mathrm{G}}}$	$\mathsf{P}^{chf}_{\mathrm{MA}^{\mathrm{G}}}$
		1.202	−1.182	−0.011	1.2027	−0.8072	−0.0126	.03	.169	.136	.134
30	30	−1.182	1.202	−0.013	−0.8072	1.2028	−0.0088	.1	.164	.130	.129
		−0.011	−0.013	2.385	−0.0126	−0.0088	2.3852	.3	.131	.103	.095
		1.202	−1.182	−0.011	1.3724	−0.6369	−0.0395	.03	.118	.101	.100
30	10	−1.182	1.202	−0.013	−0.6369	1.3740	−0.0317	.1	.116	.097	.096
		−0.011	−0.013	2.385	−0.0395	−0.0317	2.8157	.3	.108	.086	.081
		1.202	−1.182	−0.011	1.8106	−0.1957	−0.0853	.01	.070	.061	.062
30	5	−1.182	1.202	−0.013	−0.1957	1.8452	−0.0820	.1	.069	.060	.061
		−0.011	−0.013	2.385	−0.0853	−0.0820	4.1647	.3	.068	.059	.056

and sensor 2 has been tracking target 2 only because of occlusion conditions. Then, after the initial estimation periods, both targets are visible for sensor 1 and a centralized fusion architecture [1] is assumed. The two measurements of sensor 1 are to be associated with the two tracks—one from sensor 1, the other from sensor 2.

Because of the different past "history" of these tracks, the difference matrix A_{21} is no longer "close" to zero, and in general it can have positive and negative eigenvalues,[9] so the approximate algorithms for the Global assignment from Section 4.2 give inaccurate values (for the Nearest Neighbor the evaluation algorithm from Section 4 works well). Since the distribution of such quadratic form is difficult to find, the characteristic function based method of Section 4.2 is needed. The results are shown in Table III.

6.3. T2T Misassociation Probabilities

Consider again Scenario 2, where two sensors (local trackers) generate track estimates from two targets, but now the estimation is done simultaneously by the sensors, and the results transmitted to a fusion center. The probability of misassocition is again parameterized by the separation between the targets. In this case we are not interested in the measurement-to-track association (which is assumed to be done perfectly at each sensor) but in the track-to-track association performed at the fusion center. We consider that the local tracks are based on different numbers of measurements, n_1 for the first sensor and n_2 for the second, modeling different times of target acquisition. The probability of detection is considered to be 1.

Table IV shows the results obtained by the two approximation methods for the cases of NN and global association criteria. Also the misassociation probability estimated from 1000 Monte Carlo runs is shown to validate the results obtained.

It can be seen that for large target separation the probability of misassociation goes to zero, as expected, and that in this case the chi-square approximation is very accurate, and much better than the normal approximation. For smaller separation, when the misassociation probability is close to one half, the Gaussian approximation is better, although it may mismatch the real value by up to a 10%. Overall both approximations provide larger probabilities than the true one, so the lowest value estimate should be used to guarantee an error below 10%.

[9]This scenario was devised to see if the evaluation technique works for indefinite difference matrices.

The advantage of using global assignment vs. nearest neighbor is clear, unless the targets are so far apart that it is obvious how to do the association, or so close that no matter which method is used, the misassociation probability is around 0.5.

7. CONCLUSIONS

For the M2T association problem, the probability that the measurement from an extraneous target will be (mis)associated with a target of interest by the (local) Nearest Neighbor association was evaluated exactly for the case of equal track prediction covariances and approximately for the case of unequal covariances. It was shown that this misassociation probability depends on a particular—covariance weighted—norm of the difference between the targets' predicted measurements designated as the "separation" of the tracks. Numerical simulations confirm the accuracy of the solutions presented for the misassociation probabilities.

For the Global association, in the case of very dissimilar track covariances the approximation methods do not work, and a characteristic function based method, which is more expensive computationally but exact, was presented with excellent results. The probability formulas derived as well as the Monte Carlo runs show the benefit of the Global (G) vs. Nearest Neighbor (NN) associations, especially in the case of similar track covariances. Future work will involve considering the multiframe or multidimensional association (MDA) case.

The T2T association problem is harder, and only approximate results are presented, which nonetheless provide accurate results for both global and NN association criteria. The estimated probabilities are never off by more than a 10% of the true value. It has been shown that a chi-square matching of the statistic gives the best results when the separation is large, and for the case of smaller separations a Gaussian matching provides better results.

APPENDIX A. MOMENT MATCHING OF QUADRATIC FUNCTIONS: UNCORRELATED CASE

Consider the random variable defined by

$$v = u'Au + b'u + c \tag{112}$$

where A is any real $n\times n$ matrix, b is a real $n\times 1$ vector, c is a real scalar and u is a Gaussian random vector

$$u \sim \mathcal{N}(\mu,\Sigma). \tag{113}$$

Define the zero mean version of u as $\tilde{u} = u-\mu$, an rewrite

$$\begin{aligned} v &= (\tilde{u}+\mu)'A(\tilde{u}+\mu) + b'(\tilde{u}+\mu) + c \\ &= \tilde{u}'A\tilde{u} + \tilde{u}'A\mu + \mu'A\tilde{u} + \mu'A\mu + b'\tilde{u} + b'\mu + c \\ &= \tilde{u}'A\tilde{u} + \tilde{b}'\tilde{u} + \tilde{c} \end{aligned} \tag{114}$$

where $\tilde{b} = A'\mu + A\mu + b$ and $\tilde{c} = \mu A'\mu + b'\mu + c$ The mean value of v is

$$\begin{aligned} E\{v\} = \bar{v} &= E\{\tilde{u}'A\tilde{u}\} + \tilde{b}'E\{\tilde{u}\} + \tilde{c} \\ &= \mathrm{tr}(A\Sigma) + \tilde{c} \end{aligned} \tag{115}$$

where the expected value of a quadratic form is taken from [3] Section 1.4.15 The variance of v is

$$\begin{aligned} \mathrm{Var}\{v\} &= E\{(v-\bar{v})^2\} = E\{(\tilde{u}'A\tilde{u} + \tilde{b}'\tilde{u} - \mathrm{tr}(A\Sigma))^2\} \\ &= E\{\tilde{u}'A\tilde{u}\tilde{u}'A\tilde{u}\} + \tilde{b}'E\{\tilde{u}\tilde{u}'\}\tilde{b} + \mathrm{tr}(A\Sigma)^2 + 2E\{\tilde{u}'A\tilde{u}'\tilde{u}'\}\tilde{b} \\ &\quad - 2E\{\tilde{u}'A\tilde{u}\}\mathrm{tr}(A\Sigma) - 2\tilde{b}'E\{\tilde{u}\}\mathrm{tr}(A\Sigma) \\ &= \mathrm{tr}(A\Sigma)^2 + 2\mathrm{tr}(A\Sigma A\Sigma) + \tilde{b}'\Sigma\tilde{b} + \mathrm{tr}(A\Sigma)^2 - 2\mathrm{tr}(A\Sigma)^2 \\ &= 2\mathrm{tr}(A\Sigma A\Sigma) + [(A+A')\mu + b]'\Sigma[(A+A')\mu + b] \end{aligned} \tag{116}$$

where, as before, a compact expression for the fourth moment of u is used, and the terms containing odd powers of u are zero.

APPENDIX B. MOMENT MATCHING OF QUADRATIC FUNCTIONS: CORRELATED CASE

Consider four random vectors $\hat{x}_i^m, \hat{x}_i^n, \hat{x}_j^m$ and $\hat{x}_j^n$, Gaussian distributed

$$\begin{bmatrix} \hat{x}_i^m \\ \hat{x}_i^n \\ \hat{x}_j^m \\ \hat{x}_j^n \end{bmatrix} = \begin{bmatrix} \tilde{x}_i^m + x_i \\ \tilde{x}_i^n + x_i \\ \tilde{x}_j^m + x_j \\ \tilde{x}_j^n + x_j \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} x_i \\ x_i \\ x_j \\ x_j \end{bmatrix}, \begin{bmatrix} P_i^m & P_{ii}^{mn} & 0 & 0 \\ P_{ii}^{nm} & P_i^n & 0 & 0 \\ 0 & 0 & P_j^m & P_{jj}^{mn} \\ 0 & 0 & P_{jj}^{nm} & P_j^n \end{bmatrix} \right)$$

the four possible distance terms are

$$D_{ij} = (\tilde{x}_i^m - \tilde{x}_j^n - s)'[P_i^m + P_j^n]^{-1}(\tilde{x}_i^m - \tilde{x}_j^n - s) \tag{117}$$

$$D_{ji} = (\tilde{x}_j^m - \tilde{x}_i^n + s)'[P_j^m + P_i^n]^{-1}(\tilde{x}_j^m - \tilde{x}_i^n + s) \tag{118}$$

$$D_{ii} = (\tilde{x}_i^m - \tilde{x}_i^n)'[P_i^m + P_i^n - P_{ii}^{mn} - P_{ii}^{nm}]^{-1}(\tilde{x}_i^m - \tilde{x}_i^n) \tag{119}$$

$$D_{jj} = (\tilde{x}_j^m - \tilde{x}_j^n)'[P_j^m + P_j^n - P_{jj}^{mn} - P_{jj}^{nm}]^{-1}(\tilde{x}_j^m - \tilde{x}_j^n) \tag{120}$$

where $s = x_j - x_i$.

For convenience, define

$$M = [P_i^m + P_j^n]^{-1};\ N = [P_i^m + P_i^n - P_{ii}^{mn} - P_{ii}^{nm}]^{-1} \tag{121}$$

$$K = [P_j^m + P_i^n]^{-1};\ Q = [P_j^m + P_j^n - P_{jj}^{mn} - P_{jj}^{nm}]^{-1}. \tag{122}$$

To obtain the first and second moments for the difference of distances, the following tools are needed.

From [3], the moments for quadratic and quartic zero mean Gaussian random vectors $x \sim \mathcal{N}(0, R_x)$ are given by

$$E\{x'W_1x\} = \mathrm{tr}(W_1R_x) \tag{123}$$

$$E\{x'W_1xx'W_2x\} = \mathrm{tr}(W_1R_x)\mathrm{tr}(W_2R_x) + 2\mathrm{tr}(W_1R_xW_2R_x). \tag{124}$$

If two zero mean Gaussian random vectors x and y are correlated through R_{xy}, the vector y can be written as

$$y = Rx + Tw \tag{125}$$

where $w \sim \mathcal{N}(0, I)$ and

$$R = R'_{xy}Rx^{-1} \tag{126}$$

$$T = (Ry - R'_{xy}Rx^{-1}R_{xy})^{1/2}. \tag{127}$$

In the above $\Xi^{1/2}$ denotes the Cholesky factor of Ξ, so that $\Xi^{1/2}(\Xi^{1/2})' = \Xi$.

For the NN association case we are interested in the moments of the difference $d_1 = D_{ij} - D_{ii}$. In this case the variables $\tilde{x}_i^m$ and $\tilde{x}_i^n$ are correlated, and can be related as

$$\tilde{x}_i^n = A\tilde{x}_i^m + Bw \tag{128}$$

where $w \sim \mathcal{N}(0, I)$ and

$$A = (P_{ii}^{mn})'(P_i^m)^{-1} \tag{129}$$

$$B = (P_i^n - (P_{ii}^{mn})'(P_i^m)^{-1}P_{ii}^{mn})^{1/2}. \tag{130}$$

Then we have

$$\mu_{d_1} = E\{(\tilde{x}_i^m - \tilde{x}_j^n - s)'M(\tilde{x}_i^m - \tilde{x}_j^n - s) - (\tilde{x}_i^m - \tilde{x}_i^n)'N(\tilde{x}_i^m - \tilde{x}_i^n)\} \tag{131}$$

$$\begin{aligned} &= E\{(\tilde{x}_i^m)'(M-N)\tilde{x}_i^m - 2(\tilde{x}_i^m)'M(\tilde{x}_j^n + s) + (\tilde{x}_j^n)'M\tilde{x}_j^n + s'Ms \\ &\quad - 2(\tilde{x}_j^n)'Ms + (\tilde{x}_i^m)'N\tilde{x}_i^n - (\tilde{x}_i^n)'N\tilde{x}_i^n\} \end{aligned} \tag{132}$$

$$= \mathrm{tr}((M-N)P_i^m) + \mathrm{tr}(MP_j^n) + \mathrm{tr}(2NAP_i^m) - \mathrm{tr}(NP_i^n). \tag{133}$$

Define $\tilde{d}_1 = d_1 - s'Ms$, so that $\mathrm{cov}(d_1) = \mathrm{cov}(\tilde{d}_1)$ as s is a constant. Then after some algebraic operations

$$\begin{aligned} E\{\tilde{d}_1^2\} &= E\{[(\tilde{x}_i^m - \tilde{x}_j^n - s)'M(\tilde{x}_i^m - \tilde{x}_j^n - s) \\ &\qquad - (\tilde{x}_i^m - \tilde{x}_i^n)'N(\tilde{x}_i^m - \tilde{x}_i^n) - s'Ms]^2\} \\ &= \mathrm{tr}((M-N)P_i^m)^2 + 2\mathrm{tr}((M-N)P_i^m(M-N)P_i^m) \\ &\quad + \mathrm{tr}(MP_j^n)^2 + 2\mathrm{tr}(MP_j^nMP_j^n) \\ &\quad - \mathrm{tr}(NP_i^n)^2 + 2\mathrm{tr}(NP_i^nNP_i^n) + 4\mathrm{tr}(Mss'MP_i^m) \\ &\quad + 4\mathrm{tr}(NAP_i^m)\mathrm{tr}(AN'P_i^m) + 8\mathrm{tr}(NAP_i^mAN'P_i^m) \\ &\quad + 4\mathrm{tr}(NBB'NP_i^m) + \mathrm{tr}((M-N)P_i^m)\mathrm{tr}(MP_j^n) \\ &\quad + 2\mathrm{tr}((M-N)P_i^m)\mathrm{tr}(NAP_i^m) + 4\mathrm{tr}((M-N)P_i^mNAP_i^m) \\ &\quad - \mathrm{tr}((M-N)P_i^m)\mathrm{tr}(ANAP_i^m) - \mathrm{tr}((M-N)P_i^mANAP_i^m) \\ &\quad - \mathrm{tr}((M-N)P_i^m)\mathrm{tr}((B'NB)m + 2\mathrm{tr}(MP_j^n)\mathrm{tr}((NAP_i^m) \\ &\quad - \mathrm{tr}(MP_j^n)\mathrm{tr}(NP_i^n) - 2\mathrm{tr}(NAP_i^m)\mathrm{tr}(A'NAP_i^m) \\ &\quad - 4\mathrm{tr}(NAP_i^mA'NAP_i^m) - 2\mathrm{tr}(NAP_i^m)\mathrm{tr}(B'NB) \\ &\quad - 4\mathrm{tr}(NBB'NAP_i^m). \end{aligned} \tag{134}$$

For the global association case we are interested in the moments of

$$d_2 = D_{ij} + D_{ji} - D_{ii} - D_{jj}. \tag{135}$$

In this case the vectors $\tilde{x}_i^m$ and $\tilde{x}_i^n$ are correlated as before and (128) still holds. Also $\tilde{x}_j^m$ and $\tilde{x}_j^n$ are correlated and can be related as

$$\tilde{x}_j^n = C\tilde{x}_j^m + Dx \tag{136}$$

where $x \sim \mathcal{N}(0, I)$ and

$$C = (P_{jj}^{mn})'(P_j^m)^{-1} \tag{137}$$

$$D = \mathrm{chol}(P_j^n - (P_{jj}^{mn})'(P_j^m)^{-1}P_{jj}^{mn}). \tag{138}$$

Then, the moments of interest are

$$\begin{aligned} \mu_{d_2} &= E\{(\tilde{x}_i^m - \tilde{x}_j^n - s)'M(\tilde{x}_i^m - \tilde{x}_j^n - s) \\ &\quad + (\tilde{x}_j^m - \tilde{x}_i^n - s)'K(\tilde{x}_j^m - \tilde{x}_i^n - s) \\ &\quad - (\tilde{x}_i^m - \tilde{x}_i^n)'N(\tilde{x}_i^m - \tilde{x}_i^n) - (\tilde{x}_j^m - \tilde{x}_j^n)'Q(\tilde{x}_j^m - \tilde{x}_j^n)\} \end{aligned} \tag{139}$$

$$\begin{aligned} &= \mathrm{tr}((M-N)P_i^m) + \mathrm{tr}((M-Q)P_j^n) \\ &\quad + \mathrm{tr}((K-N)P_i^n) + \mathrm{tr}((K-Q)P_j^m) \\ &\quad + 2\mathrm{tr}(NAP_i^m) + 2\mathrm{tr}(QCP_j^n). \end{aligned} \tag{140}$$

Define $\tilde{d}_2 = d_2 - s'(M+k)s$, so that $\mathrm{cov}(d_2) = \mathrm{cov}(\tilde{d}_2)$ as s is a constant. Then after some algebraic operations

$$\begin{aligned} E\{\tilde{d}_2^2\} &= \mathrm{tr}((M-N)P_i^m)^2 + 2\mathrm{tr}((M-N)P_i^m(M-N)P_i^m) \\ &\quad + \mathrm{tr}((M-Q)P_j^n)^2 + 2\mathrm{tr}((M-Q)P_j^n(M-Q)P_j^n) \\ &\quad + \mathrm{tr}((N-C)P_i^n)^2 + 2\mathrm{tr}((N-C)P_i^n(N-C)P_i^n) \\ &\quad + \mathrm{tr}((K-Q)P_j^m)^2 + 2\mathrm{tr}((K-Q)P_j^m(K-Q)P_j^m) \\ &\quad + \mathrm{tr}((-2M)P_j^n(-2M)'P_i^m) + \mathrm{tr}((-2M)ss'(-2M)'P_i^m) \\ &\quad + \mathrm{tr}((2M)ss'(2M)'P_j^n) + \mathrm{tr}((-2K)P_i^n(-2K)P_j^m) \\ &\quad + \mathrm{tr}((-2K)ss'(-2K)P_j^m) + \mathrm{tr}((-2K)ss'(-2K)P_i^n) \\ &\quad + \mathrm{tr}((2N)AP_i^m)\mathrm{tr}(A'(2N)AP_i^m) \\ &\quad + 2\mathrm{tr}((2N)AP_i^mA'(2N)AP_i^m) + \mathrm{tr}((2N)BB'(2N)'P_i^m) \\ &\quad + \mathrm{tr}(RCP_j^n)\mathrm{tr}(C'R'P_j^n) + 2\mathrm{tr}(RCP_j^nC'R'P_j^n) \\ &\quad + \mathrm{tr}(RDD'R'P_j^n) + \mathrm{tr}((M-N)P_i^m)\mathrm{tr}((M-Q)P_j^n) \\ &\quad + \mathrm{tr}((M-N)P_i^m)\mathrm{tr}(A'(N-C)AP_i^m) \\ &\quad + 2\mathrm{tr}((M-N)P_i^mA'(N-C)AP_i^m) \\ &\quad + \mathrm{tr}((M-N)P_i^m)\mathrm{tr}(B'(N-C)B) \\ &\quad + \mathrm{tr}((M-N)P_i^m)\mathrm{tr}((K-Q)P_j^m) \end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{tr}((M-N)P_i^m)\mathrm{tr}((2N)AP_i^m)\\
&+ 2\mathrm{tr}((M-N)P_i^m(2N)AP_i^m)\\
&+ \mathrm{tr}((M-N)P_i^m)\mathrm{tr}(R'CP_j^n)\\
&+ \mathrm{tr}((M-Q)P_j^n)\mathrm{tr}((N-C)P_i^n)\\
&+ \mathrm{tr}((M-Q)P_j^n)\mathrm{tr}(C'(K-Q)CP_j^n)\\
&+ 2\mathrm{tr}((M-Q)P_j^nC'(K-Q)CP_j^n)\\
&+ \mathrm{tr}((M-Q)P_j^n)\mathrm{tr}(D'(K-Q)D)\\
&+ \mathrm{tr}((M-Q)P_j^n)\mathrm{tr}((2N)AP_i^m)\\
&+ \mathrm{tr}((M-Q)P_j^n)\mathrm{tr}(R'CP_j^n)\\
&+ 2\mathrm{tr}((M-Q)P_j^nR'CP_j^n)\\
&+ \mathrm{tr}((N-C)P_i^n)\mathrm{tr}((K-Q)P_j^m)\\
&+ \mathrm{tr}(A'(N-C)AP_i^m)\mathrm{tr}((2N)AP_i^m)\\
&+ 2\mathrm{tr}(A'(N-C)AP_i^m(2N)AP_i^m)\\
&+ \mathrm{tr}(B'(N-C)B)\mathrm{tr}((2N)AP_i^m)\\
&+ 2\mathrm{tr}(A'(N-C)BB'(2N)'P_i^m)\\
&+ \mathrm{tr}((N-C)P_i^n)\mathrm{tr}(R'CP_j^n)\\
&+ \mathrm{tr}((K-Q)P_j^m)\mathrm{tr}((2N)AP_i^m)\\
&+ \mathrm{tr}(C'(K-Q)CP_j^n)\mathrm{tr}(R'CP_j^n)\\
&+ 2\mathrm{tr}(C'(K-Q)CP_j^nR'CP_j^n)\\
&+ \mathrm{tr}(D'(K-Q)D)\mathrm{tr}(R'CP_j^n)\\
&+ 2\mathrm{tr}(D'(K-Q)CP_j^nR'D)\\
&+ \mathrm{tr}((-2M)P_j^nC'(-2K)AP_i^m)\\
&+ \mathrm{tr}((-2M)ss'(-2K)'AP_i^m)\\
&- \mathrm{tr}((2M)ss'(-2K)CP_j^n)\\
&+ \mathrm{tr}((2N)AP_i^m)\mathrm{tr}(C'RP_j^n). \qquad (141)
\end{aligned}
$$

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.
J. Wiley and Sons, 2001.

[4] Y. Bar-Shalom, S. S. Blackman and R. J. Fitzgerald
The dimensionless score function for measurement to track association.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 1 (Jan. 2007), 392–400.

[5] DCDFLIB (Double precision Cumulative Distribution Function LIBrary)
http://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware.aspx?Software_Id=21.

[6] A. Farina and R. Miglioli
Association of active and passive tracks for airborne sensors.
Signal Process. **69** ,3 (1998), 209–217.

[7] J. Ferry
XMAP: Track-to-track association with metric, feature, and target-type data.
In *Proceedings of the Conference on Information Fusion*, July 2006.

[8] J. P. Imhof
Computing the distribution of quadratic forms in normal variables.
Biometrika, **48** (1961), 419–426.

[9] M. Levedahl
An analysis of object designation performance using GNN and GNP correlation.
In *Proceedings SPIE Conference on Tracking Small Targets*, Orlando, FL, 2004, 441–451.

[10] S. Mori, K. C. Chang and C. Y. Chong
Performance analysis of optimal data association with application to multiple target tracking.
In Y. Bar-Shalom (Ed.), *Multitarget-Multisensor Tracking: Applications and Advances*, vol. II, Artech House, 1992 (reprinted by YBS Publishing, 1998).

[11] A. Poore and N. Rijavec
Lagrangian relaxation algorithm for multidimensional assignment problems arising from multitarget tracking.
SIAM Journal on Optimization **3**, 3 (Aug. 1993), 544–563.

[12] D. Raphaeli
Distribution of noncentral indefinite quadratic forms in complex normal variables.
IEEE Transactions on Information Theory, **42**, 3 (1996), 1002–1007.

[13] R. Saha
Analytical evaluation of an ESM/radar track association algorithm.
In *Proceedings SPIE Conference on Tracking Small Targets*, Orlando, FL, Apr. 1992, 338–347.

[14] L. Stone, M. Williams and T. Tran
Track-to-track association and bias removal.
In *Proceedings SPIE Conference on Tracking Small Targets*, Oct. 2001, 315–329. http://en.wikipedia.org/wiki/Noncentral_chi-square_distribution.

Javier Areta received the electrical engineer degree in 2001 from the University of La Plata, Argentina (UNLP). He is currently pursuing his Ph.D. in electrical and computer engineering at the University of Storrs, Connecticut. His main interests lie in the areas of tracking, data fusion and signal processing.

Yaakov Bar-Shalom (S'63—M'66—SM'80—F'84) 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.

Ronald Rothrock was born on May 16, 1948. He received his B.S. degree in mechanical and aerospace engineering from Illinois Institute of Technology in 1970. He served six years in the U.S. Navy as an aviator, and earned his M.S. in operations research from Stanford University in 1977.

He worked for BDM International, Inc., (and its eventual purchaser, TRW) in McLean, VA, from 1977 to 1998, concentrating in the areas of defense communications and radar sensor networking and tracking. During 1998 he briefly worked for CSC and then joined SPARTA, Inc., in Arlington, VA, for which he continues to work as a chief engineer and director of a group that develops air vehicle, ballistic missile, and maritime surface vessel tracking, data conditioning, data association, classification, sensor registration, intelligent message pruning for peer-to-peer communications networks, and related algorithms, deliverable software, and simulations.

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

BOARD OF DIRECTORS*

2005–2007	**2006–2008**	**2006–2009**
Per Svensson	David Hall	Stefano Coraluppi
Darko Musicki	Ivan Kadar	Jean Dezert
Eloi Bossé	Elisa Shahbazian	Wolfgang Koch

*Board of Directors are elected by the members of ISIF for a three year term.

PAST PRESIDENTS

Darko Musicki, 2008
Erik Blasch, 2007
Pierre Valin, 2006
W. Dale Blair, 2005
Chee Chong, 2004
Xiao-Rong Li, 2003
Yaakov Bar-Shalom, 2002
Parmod Varshney, 2001
Yaakov Bar-Shalom, 2000
Jim Llinas, 1999
Jim Llinas, 1998

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 2 December 2007 ISSN 1557-6418