Effects of Computing and Communications on State Fusion Over Long-Haul Sensor Networks

Nageswara S. V. Rao
Oak Ridge National Laboratory
Oak Ridge, TN 37831
Email: raons@ornl.gov

Katharine Brigham and B. V. K. Vijaya Kumar
Carnegie Mellon University
Pittsburgh, PA 15213
Email: {kbrigham,kumar}@ece.cmu.edu

Qiang Liu and Xin Wang
State University of New York at Stony Brook
Stony Brook, NY 11794
Email: {qiangliu,xwang}@ece.sunysb.edu

***Abstract*—We consider a sensor network where the state estimates are sent over long-haul connections from sensors to a fusion center to be correlated and fused to generate a global state estimate. The network loss and latency determine the state estimates received within a time-window at the fusion center, which in turn determine the quality $Q_{CC}(\tau)$ of the global state estimate computed within the allocated time τ. We derive a probabilistic performance bound on $Q_{CC}(\tau)$ as a function of the distribution of state estimates as well as the computation and network parameters, which shows a separation between the contributions of communications and computing parts. We then show that measurements based on field tests can be used to derive confidence bounds for $Q_{CC}(\tau)$. We complement the analytical results with component simulations and a system-level simulation of tracking a ballistic object to illustrate the qualitative effects of computing, including allocated computation time, and communications, including loss rates and transport protocols.**

***Index Terms*—Long-haul sensor networks, state fusion, communication protocols, intractable computation.**

I. Introduction

Sensor networks have been deployed in a variety of applications including environmental monitoring, target tracking, and health care [1], [2]. In particular, information fusion methods have been utilized to combine data from networked sensors to support detection and tracking tasks [3], [4]. Several of these works consider a "small" communications network so that the effects of losses and latency can be mitigated by an appropriate choice of data processing algorithms [5], [6]. In sharp contrast, sensors in *long-haul* sensor networks may be distributed across the globe and/or in space, in applications such as monitoring greenhouse gas emissions using airborne and ground sensors [7], processing global cyber events using cyber sensors distributed over the Internet [8], space exploration using a network of telescopes [9], and target detection and tracking for air and missile defense [10]. The response time requirements of of such long-haul sensor networks could be quite varied, from a few seconds in detecting cyber attacks on critical infrastructures to years in detecting global trends in greenhouse gas emissions. We focus on a particular class of long-haul sensor networks that are tasked to detect and track events and/or targets within a timescale of seconds.

We consider long-haul sensor networks, wherein the state estimates are sent to a fusion center over long-haul connections, such as submarine and terrestrial fiber connections or satellite links that could be tens of thousands of miles long. The sensor data received at the fusion center within a time-window (of few seconds) are correlated and fused to generate global state estimates of targets within the "field of view" of one or more sensors. The correlation algorithm determines the groups of state estimates, such that each group is hypothesized to correspond to a single target. The fusion algorithm combines the state estimates of each group into a single global estimate corresponding to the underlying target. These algorithms, typically nonlinear, are used to solve computationally intractable problems [11], [12], and hence the quality of their output for a fixed time allocation τ could vary significantly. Their execution time depends not only on the number of the input state estimates but also on their inter-relationships; for example, well-clustered state estimates are easier to correlate than uniformly dispersed ones. Furthermore, the communication round-trip time (RTT) is of the same order as the algorithm time-window, and random communication losses and latency variations will be reflected in the different sensor estimates arriving within the time-window at the fusion center. As a result, both the computing and communications parameters contribute to the quality of the fused state estimates. The performance measure of the global state estimate generated by the network is given by $Q_{CC}(\hat{X}_1, \ldots, \hat{X}_N, \tau) \in [0, 1]$, which typically represents the normalized state estimation error. The goal is to keep this performance measure as small as possible, and in particular to keep it lower than a specified value δ with a certain level of confidence.

A complete and precise evaluation of the performance measure of the global state estimate would require an in-depth knowledge of the underlying distributions of state estimate errors, correlation and fusion algorithms, network communications parameters, and their mutual dependencies. Such detailed knowledge is often not available; nevertheless, the network is deployed in such a way that field system tests can be conducted using targets of known states. Although significant expense and time are often entailed for such field tests that could limit the total number of runs, the correlation and fusion algorithms can be executed off-line using the field measurements and different time allocations at significantly lower costs. We limit our treatment to estimating approximations to the performance bounds of the global state estimate based on a limited number of such tests.

Our focus in this work includes: (a) analytical methods to assess the quality of fused state estimates with different communication and computation scenarios and (b) estimation

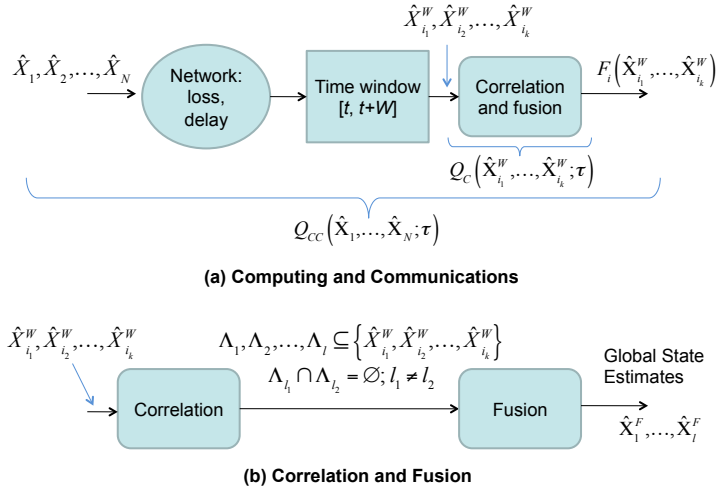


Fig. 1. Decomposition of computing and communications parts

methods to utilize measurements from field system tests to support such assessments. Our contributions are two-fold:

(a) We derive an expression for the expected value $\bar{Q}_{CC}(\tau)$ of the global estimate as a function of the fixed computing time τ and the network loss probability and latency. This analysis shows a degree of separation between the contributions of network and computation parameters to $\bar{Q}_{CC}(\tau)$, so that they can be analyzed independently.

(b) We demonstrate that measurements of $Q_{CC}(\tau)$ from field system tests can be used to derive robust estimates of the probability with which the quality of the fused estimate is bounded by δ for an allocated time τ, namely $\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta\right\}$.

We complement these analytical results by a series of simulations with different levels of abstractions to illustrate the qualitative effects of (i) computing, including types of fusers and allocated computation time, and (ii) communications, including loss rates and transport protocols.

This paper is organized as follows: In Section II we derive a performance bound for $Q_{CC}(\hat{X}_1,\ldots,\hat{X}_N,\tau)$ as a function of τ and parameters of the network and processing algorithms. We describe performance bounds on estimates of Q_{CC} using measurements from field tests in Section III. Component and system-level simulation results are presented in Section IV, and the conclusions are presented in Section V.

II. Computing and Communications Decomposition

The computation and communication processes involved in the generation of global state estimates are illustrated in Fig. 1. A total of N sensors are used to generate the state estimates $\hat{X}_1,\hat{X}_2,\ldots,\hat{X}_N$, which are then sent to the fusion center. However, due to the long-haul connections, these packets are fairly susceptible to losses and significant delays so that only a subset of the state estimates, $\hat{X}^W_{i_1},\ldots\hat{X}^W_{i_k}$, arrive within a time-window $[t,t+W]$ at the fusion center. The received state estimates are then input into the correlation and fusion algorithms. The output of the correlation algorithm consists of groups of correlated state estimates $\Lambda_1,\Lambda_2,\ldots,\Lambda_l\subseteq\{\hat{X}^W_{i_1},\ldots\hat{X}^W_{i_k}\}$; $\Lambda_{l_1}\cap\Lambda_{l_2}=\varnothing$, for $l_1\neq l_2$, such that each Λ_i is hypothesized to correspond to a single target i. Each set of correlated state estimates is then fused to generate the global state estimates.

With the estimates $\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}$ received at the fusion center, the computation time of the correlation and fusion algorithms is $T_C\left(\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}\right)$. If allocated time τ is smaller than this quantity, the algorithms may not run to completion, thereby affecting the quality of the output. Let $Q_C\left(\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k};\tau\right)$ denote the performance measure of the correlation and fusion output normalized to interval $[0,1]$, where 0 represents no error and 1 represents the maximum error. By combining computing and communications, the quality of the estimate generated by the network in response to the state estimates $\hat{X}_1,\hat{X}_2,\ldots,\hat{X}_N$ is $Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)$. This is a random quantity with two types of contributors: (a) distributions of state estimates, including those of the target, measurement process and state estimation algorithms, and (b) network loss and latency distributions that determine the state estimates that arrive within the time-window $[t,t+W]$ at the fusion center. The expectation of the performance measure is given by

$$\bar{Q}_{CC}(N,\tau)=\int_{\hat{X}_1,\ldots,\hat{X}_N}Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)dP_{\hat{X}_1,\ldots,\hat{X}_N}=$$
$$\int_{\hat{X}_1,\ldots,\hat{X}_N}\int_{k\geq 0:\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}}Q_C\left(\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k};\tau\right)dP_{\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}}$$
$$\cdot\,\mathbf{P}\left\{k\geq 0:t_{\hat{X}^W_{i_1}},\ldots,t_{\hat{X}^W_{i_k}}\in[t,t+W]\right\}dP_{\hat{X}_1,\ldots,\hat{X}_N}, \tag{1}$$

where $t_{\hat{X}^W_{i_j}}$, $j=1,2,\ldots,k$, denotes the time at which the state estimate $\hat{X}^W_{i_j}$ arrives at the fusion center. This expression demonstrates the contributions of randomness due to (a) state estimates reflected in the measure $Q_C\left(\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k};\tau\right)$ and (b) communications network parameters reflected in the probability $\mathbf{P}\left\{t_{\hat{X}^W_{i_1}},\ldots,t_{\hat{X}^W_{i_k}}\in[t,t+W]\right\}$.

We are interested in the probability of ensuring the quality measure of the final estimate to be below δ, when time τ is allocated to the correlation and fusion algorithms; that is, $\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta\right\}$. We decompose this quantity by conditioning on the state estimates as follows

$$\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta\right\}= \tag{2}$$
$$\int\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta|\hat{X}_1,\ldots,\hat{X}_N\right\}dP_{\hat{X}_1,\ldots,\hat{X}_N}.$$

A lower bound on $\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta|\hat{X}_1,\ldots,\hat{X}_N\right\}$ is derived as follows

$$\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1,\ldots,\hat{X}_N;\tau\right)<\delta|\hat{X}_1,\ldots,\hat{X}_N\right\}=$$
$$\int_{k\geq 0:\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}}\mathbf{P}\left\{Q_C\left(\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k};\tau\right)<\delta\right\}dP_{\hat{X}^W_{i_1},\ldots,\hat{X}^W_{i_k}}$$
$$\cdot\,\mathbf{P}\left\{k\geq 0:t_{\hat{X}^W_{i_1}},\ldots,t_{\hat{X}^W_{i_k}}\in[t,t+W]\right\}$$

$$\geq \sum_{k=0}^{N} \left(1 - \frac{E_k\left[Q_C(\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W; \tau)\right]}{\delta} \right) \cdot \mathbf{P}\left\{ k \geq 0 : t_{\hat{X}_{i_1}^W}, \ldots, t_{\hat{X}_{i_k}^W} \in [t, t+W] \right\} \quad (3)$$

where we have utilized Chebyshev's inequality; namely, for $X > 0$, we have $\mathbf{P}\{X < \delta\} \geq 1 - \frac{E[X]}{\delta}$, and the expectation $E_k[.]$ is taken over k received state estimates. The above expression shows the separation between the computation and communications parts, which can be analyzed independently. The first term depends on the distribution of the state estimates and the parameters of the correlation and fusion algorithms, in particular, on the random event $\left\{ T_C\left(\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W\right) < \tau \right\}$ and the resulting $Q_C\left(\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W; \tau\right)$. The second term does not depend on the state estimates and can be analyzed using the properties of network communications, such as latency and loss rate of the connection and time-out parameters of the transport protocols employed.

A. Computation Time

The computation time $T_C\left(\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W\right)$ is composed of the execution time of both correlation and fusion algorithms and could vary significantly depending on the underlying algorithms. In general, the correlation problem is computationally intractable [11] in that its worst-case execution time of any algorithm could be exponential in k. The computation time of linear fusers is $O(k)$, whereas more complicated fusers may have significantly longer execution time [12]. Consequently, the quality of the output typically degrades with decreasing τ, the time allocated for executing these algorithms. For example, the multiple-hypothesis correlation algorithm is implemented as a tree search [11], which is then curtailed to limit its execution time to τ. Such search termination leads to a sub-optimal correlation quality as reflected in the fuser output. In the special case where it is known a priori that there is only a single target, the correlation is much simpler since it only needs to identify the state estimates that correspond to the target as opposed to spurious states such as those of debris or noise.

When a fixed time τ is allocated for correlation and fusion, the expected quality with k state estimates is given by

$$\bar{Q}_C(k, \tau) = \int_{\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W} Q_C\left(\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W, \tau\right) dP_{\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W}. \quad (4)$$

Then, the overall correlation and fusion quality is

$$\bar{Q}_C(\tau) = \sum_{k=0}^{N} P_k \bar{Q}_C(k, \tau), \quad (5)$$

where P_k is the probability of receiving exactly k updates within $[t, t+W]$ at the fusion center. In general, the quantities $\bar{Q}_C(k, \tau)$ and $\bar{Q}_C(\tau)$ cannot be exactly computed since they depend on the distribution of states $\hat{X}_{i_1}^W, \ldots, \hat{X}_{i_k}^W$, which subsumes those of state estimates and communications network, as well as the properties of correlation and fusion algorithms. We consider that the sensor network has been deployed so that system tests can be conducted and the correlation and fusion algorithms can be executed. In tests, targets with known states are provided to the system and the corresponding outputs are measured, including the state estimates at sensors and fusion center, and outputs of correlation and fusion algorithms. In Section III, we describe a method that utilizes data from field tests to estimate confidence bounds on $\bar{Q}_C(k, \tau)$ and $\bar{Q}_C(\tau)$.

B. Communication Time

We consider a simple message loss model for TCP communications where each message is lost with probability p independently of other messages. At a high-level abstraction, TCP sends a message – the state estimate in our case – and waits for a time-out period T_{TO} to receive an acknowledgment, after which the message is re-sent if no acknowledgment is received. Typically, T_{TO} is several times the RTT of the connection, and over long-haul connections it could be of the order of seconds, the same as that of the time-window of correlation and fusion algorithms. Let T_{EE}^T denote the time at which a message is received at the receiver using TCP, and T_L the latency of the connection. The probability that a message will be delivered after exactly i losses is $p^i(1-p)$, which corresponds to $T_{EE}^T = iT_{TO} + T_L$. Then, the expected time at which the message is received using TCP is

$$E\left[T_{EE}^T\right] = T_L + \sum_{i=0}^{\infty} ip^i(1-p)T_{TO} = T_L + \frac{p}{1-p}T_{TO}, \quad (6)$$

and the second moment is given by

$$E\left[\left(T_{EE}^T\right)^2\right] = T_L^2 + 2T_{TO}T_L\frac{p}{1-p} + T_{TO}^2\frac{p(1+p)}{(1-p)^3}. \quad (7)$$

From these expressions, the probability that a message will be successfully delivered within the time-window $[T_L, T_L + W]$ is given by

$$\begin{aligned} \mathbf{P}\left\{T_{EE}^T - T_L < W\right\} &= 1 - \mathbf{P}\left\{T_{EE}^T - T_L > W\right\} \\ &\geq 1 - \frac{E\left[(T_{EE}^T - T_L)^2\right]}{W^2} \\ &= 1 - \frac{T_{TO}^2 p(1+p)}{W^2(1-p)^3}, \end{aligned} \quad (8)$$

where we have applied the Chebyshev's inequality of the second order: $\mathbf{P}\{X > \delta\} \leq \frac{E[X^2]}{\delta^2}$. For a given connection loss probability p, this probability bound could be lower than desired. In particular, for connections of tens of thousands of miles, T_{TO} is of the order of seconds, and even small loss probabilities will lead to messages missing the time-window by a few seconds at the fusion center. At the cost of utilizing additional bandwidth, the message delivery probability can be improved by replicating the state estimates and sending them over different TCP streams. Such streams could be sent over different paths if available and staggered across different TCP ports, and we call such protocol the Duplicated Staggered TCP (DSTCP).

We send three copies of each state estimate over the connection, which reduces the loss rate to p^3. Let T_{EE}^{DI} denote the time at which the state estimate is received using such DSTCP method. Then we have

$$E\left[T_{EE}^{DI}\right] = T_L + \frac{p^3}{1-p^3}T_{TO}, \quad (9)$$

which is smaller than $E\left[T_{EE}^{T}\right]$ for $p > 0$. Proceeding as above, we have the message delivery probability given by

$$\mathbf{P}\left\{T_{EE}^{DI} - T_L < W\right\} \geq 1 - \frac{T_{TO}^2 p^3(1+p^3)}{W^2(1-p^3)^3}, \qquad (10)$$

which is higher than in the case of TCP.

We now consider a more general scenario where message loss is correlated. Let p_1 denote the probability of losing a duplicate state estimate given the first one is lost. For loss caused by buffer overflows at the routers and hosts, p_1 could be much higher than p. Then for DSTCP that sends three copies of the state estimate, let T_{EE}^{DC} denote the time at which the message is received at the fusion center. Similar to the previous case, we have

$$E\left[T_{EE}^{DC}\right] = T_L + \frac{pp_1^2}{1-pp_1^2}T_{TO}, \qquad (11)$$

which is smaller than $E\left[T_{EE}^{T}\right]$ for $p > 0$, $p_1 > 0$ but larger than $E\left[T_{EE}^{DI}\right]$ when $p_1 > p$. Similarly, we have the message delivery probability given by

$$\mathbf{P}\left\{T_{EE}^{DC} - T_L < W\right\} \geq 1 - \frac{T_{TO}^2 pp_1^2(1+pp_1^2)}{W^2(1-pp_1^2)^3}, \qquad (12)$$

which is higher than that in the case of TCP. In general, the number of replicated state estimates needed to meet a specified probability of message delivery can be estimated with p, T_{TO} and T_L, by generalizing the above derivations. On long connections that are not bandwidth-constrained, DSTCP provides a practical solution to improve the quality of the fused state estimate. Such a method can be implemented at the application level on the end hosts.

III. Field Test Measurements

In field tests, the state estimates corresponding to known targets are sent over the network to the fusion center, and the algorithms are executed with a fixed execution time τ. The collected measurements are used to measure the performance of the fused state corresponding to (i) $\bar{Q}_C(\tau)$ for fixed execution time τ, and (ii) $\bar{Q}_C(k,\tau)$ for k received state estimates at the fusion center for fixed τ. The system executes l system tests wherein the numbers of state estimates reaching the fusion center within time-window $[t, t+W]$ are given as $k_1, k_2, \ldots, k_l$. The corresponding measured performance estimates are $q_{k_1,\tau}, q_{k_2,\tau}, \ldots, q_{k_l,\tau}$. We define the empirical mean

$$\hat{Q}(\tau) = \frac{1}{l}\sum_{i=1}^{l} q_{k_i,\tau}, \qquad (13)$$

and use it as an estimate for $\bar{Q}_C(\tau)$. The closeness of this estimate is given by the Hoeffding's inequality [13]

$$\mathbf{P}\left\{\left|\bar{Q}_C(\tau) - \hat{Q}(\tau)\right| > \epsilon\right\} < 2e^{-2\epsilon^2 l}, \qquad (14)$$

which improves with the number of measurements l.

We now address the issue of extrapolating the above result for a specified number k of the state estimates used at the fusion center. Let $\hat{r}(k)$ represent the regression estimate of $\bar{Q}_C(k,\tau)$ for fixed τ. It is assumed that $\hat{r}(k)$ is an non-increasing function of k in that the overall quality of the fused state improves (at least does not degrade) with more state updates arriving within time-window $[t, t+W]$ at the fusion center. Thus $\hat{r}_C(.)$ is chosen from a set of bounded monotone functions of a single operand, denoted by $\mathcal{M}_1$. Let $\hat{r}_C$ minimize the empirical error

$$\min_{r_C \in \mathcal{M}_1} \frac{1}{l}\sum_{i=1}^{l}\left(r_C(k_i) - q_{k_i,\tau}\right)^2. \qquad (15)$$

The prediction capability of $\hat{r}_C$ is given by Vapnik's finite sample theory [13] as follows:

$$\mathbf{P}\left\{\max_k \left|\bar{Q}_C(k,\tau) - \hat{r}_C(k)\right| > \epsilon\right\} \leq 8ele^{-\epsilon l/4}. \qquad (16)$$

which also improves with the number of measurements l but slower than Hoeffding's bound above.

IV. Simulation Examples

A. Component Simulations

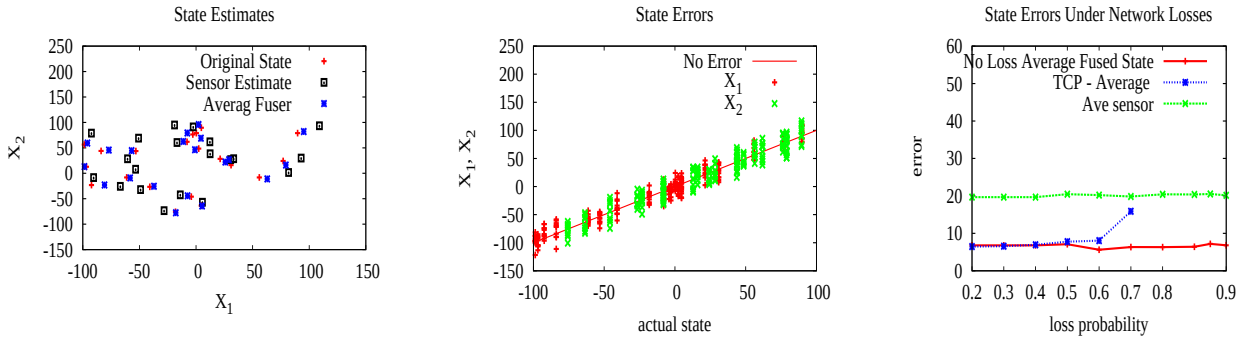
We now present simplified simulations of a single 3D target to illustrate the analytical results. Here, the states are generated uniformly within $[-A, A]^3$ area as in Fig. 2(a), which shows two of the three coordinates. The communication losses are simulated by using TCP message delivery rates[1] computed based on connection loss probabilities. In this example, RTT is 1 second corresponding to about 10,000-mile connection, and the time-window is 10 seconds.

(a) *Average Fuser - Unbiased independent errors:* The sensor errors are zero-mean and statistically independent in the range $[-B, B]$, as shown in Fig. 2(b). For this case, the fuser averages the state estimates that arrive within the time-window, and provides a substantial improvement in the state error as shown in Fig. 2(c); the Euclidean-distance error of the fused estimate is around 8 compared to that of the average sensor error at around 20. Thus, the fusion of sensor estimates is a good choice in this case if there are no network losses. As the communications loss probability is increased, TCP losses lead to the degradation of state estimate as shown in Fig. 2(c). When the loss rate exceeded 0.7 no messages were received within the window at the fusion center in 20 instances we simulated.

(b) *Nonlinear Fuser - Biased errors:* To illustrate the effects of the fuser, we consider biased sensor errors as shown in Fig. 3(a), where positive state values have a positive bias and negative state values a negative bias. In this case, the average fuser is no longer as effective, but a nonlinear fuser that applies a correction based on the sign of the coordinate and then computes an average, leads to a much improved state estimate as shown in Fig. 3 (b). This fuser is more complex and performs better than its linear counterpart; nevertheless, its performance also degrades with the connection loss probability in a qualitatively similar manner.

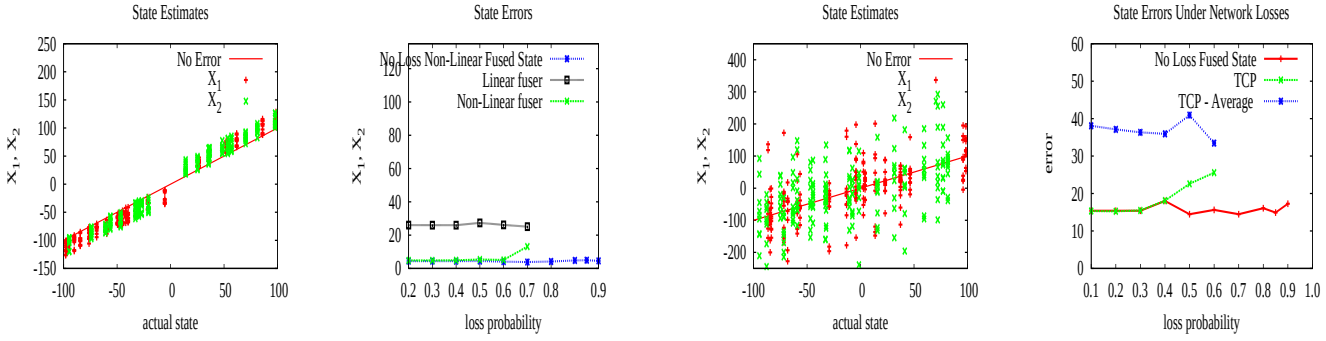
(c) *Linear Fuser - Unbiased errors with difference variances:* Next, we consider a case where sensor errors have zero bias

[1]Due to the loss recovery mechanism of TCP, the message delivery rate is higher than one minus the connection loss probability and increases with window-size W.



(a) original and estimated states (b) state error distributions (c) state errors

Fig. 2. Average fuser under unbiased independent errors and network losses



(a) biased sensor errors (b) state errors - nonlinear fuser (c) sensor errors with different variances (d) state errors - linear fuser

Fig. 3. Linear and nonlinear fusers under network losses

but have different variances as shown in Fig. 3(c). In this case, a linear fuser with coefficients inversely proportional to the sensor error covariances performs better than the average fuser, as illustrated in Fig. 3(d). And the effects of communication losses are quite similar to previous cases.

As illustrated in these cases, the fusion method determines the quality of the estimate and the execution time. Nevertheless, in each case the fuser performance degrades in a qualitatively similar way as the loss probability is increased under TCP. But DSTCP achieves a better message delivery rate than TCP, and improves the fused state estimate as will be illustrated next.

Results of 20 field test point observations for a loss rate of 40% are shown in Fig. 4. The number of state updates received using DSTCP is higher than that of TCP, which consequently leads to a lower average state error as shown in Fig. 4(a). Using the Hoeffding's inequality presented in previous section, we computed the confidence levels for 95%–99% intervals as shown in Fig. 4(b). For a 95% interval the confidence is close to 0.99, but it is only around 0.4 for a 98.5% interval. By varying the connection loss probability from 0.1 to 0.8, we collected the state errors as a function of the number of state estimates. The k values corresponding to different loss rates are shown in Fig. 4(c) along with the corresponding regression fits. The confidence estimates for the regression estimation are shown for 50%-90% intervals in Fig. 4(d), which are less tight than those of point estimates given by Hoeffding's formula since the underlying regression estimation problem is more complex than that of estimating the scalar quantity $\bar{Q}_C(\tau)$.

B. System-Level Simulation

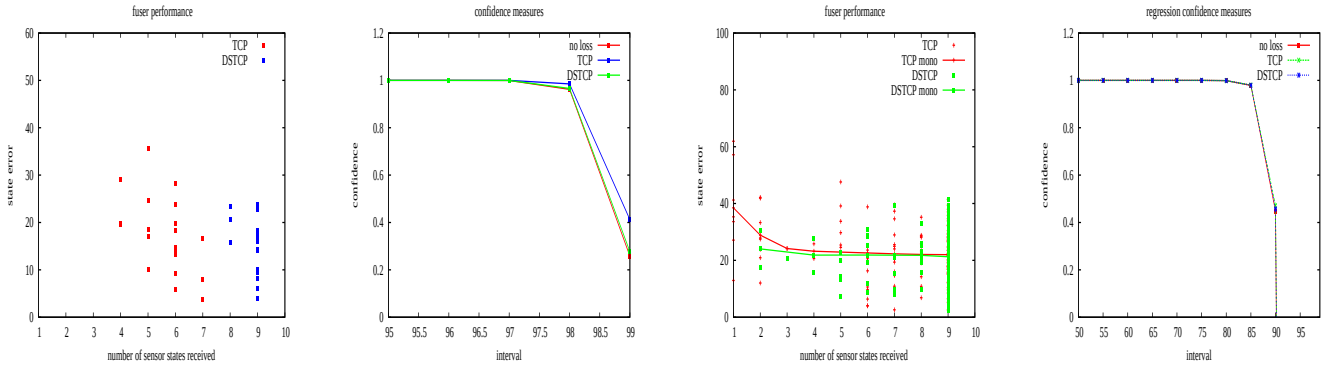
We now consider system-level simulations using a ballistic target model with known distributions and properties of the state estimate errors, correlation and fusion algorithms, and network communication parameters. The states of the coasting ballistic targets are simulated using the following state-space model [14]:

$$\dot{\mathbf{x}} \triangleq \begin{bmatrix} \dot{\mathbf{p}} \\ \dot{\mathbf{v}} \end{bmatrix} = \mathbf{f}\left(\begin{bmatrix} \mathbf{p} \\ \mathbf{v} \end{bmatrix}\right) \triangleq \begin{bmatrix} \mathbf{v} \\ \mathbf{a}_G(\mathbf{p}) \end{bmatrix}. \tag{17}$$

The target state vector $\mathbf{x} = \begin{bmatrix} \mathbf{p}^T & \mathbf{v}^T \end{bmatrix}^T$, where $\mathbf{p} = \begin{bmatrix} x & y & z \end{bmatrix}^T$ and $\mathbf{v} \triangleq \dot{\mathbf{p}} = \begin{bmatrix} \dot{x} & \dot{y} & \dot{z} \end{bmatrix}^T$ are the target position and velocity vectors, respectively. $\mathbf{a}_G(\mathbf{p})$ is the gravitational acceleration under the spherical Earth model [14]:

$$\mathbf{a}_G(\mathbf{p}) = -\frac{\mu}{p^2}\mathbf{u}_p = -\frac{\mu}{p^3}\mathbf{p}, \tag{18}$$

where $\mathbf{p}$ is the vector from the Earth's center to the target, $p \triangleq \|\mathbf{p}\|$ is its length, $\mathbf{u}_p \triangleq \mathbf{p}/p$ is the unit vector in the



(a) state error (b) Hoeffding's confidence estimates (c) regressions (d) regression confidence estimates

Fig. 4. Confidence measures based on field test simulations

direction of $\mathbf{p}$, and $\mu = 3.986012 \times 10^5$ km^3/s^2 is the Earth's gravitational constant. The algorithm used for state propagation can be found in [15].

A state estimate $\hat{X}_i$ from sensor i is generated by adding random Gaussian noise to the true states:

$$\hat{X}_i = X + n_i, \tag{19}$$

where X is the true target state for a target, and $n_i \sim \mathcal{N}(\underline{0}, \Sigma)$. Σ is a diagonal matrix

$$\Sigma = diag\left(\begin{bmatrix}\sigma_x^2 & \sigma_y^2 & \sigma_z^2 & \sigma_{\dot{x}}^2 & \sigma_{\dot{y}}^2 & \sigma_{\dot{z}}^2\end{bmatrix}\right), \tag{20}$$

where σ_x^2, σ_y^2, and σ_z^2 are the position error variances, and $\sigma_{\dot{x}}^2$, $\sigma_{\dot{y}}^2$, and $\sigma_{\dot{z}}^2$ are the velocity error variances.

The states for two different targets are generated using the initial states shown in Table I. Two sensors are assumed to be tracking each target at all times (for a total of 4 sensors in these simulations) although the correlator does not always correctly assign the state estimates from each sensor to the correct target. The state estimate error parameters are set as follows for both sensors: $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$ and $\sigma_{\dot{x}}^2 = \sigma_{\dot{y}}^2 = \sigma_{\dot{z}}^2 = 10^{-4}$.

TABLE I
INITIAL TARGET STATES

Target	Position x, y, z (km)	Velocity $\dot{x}$, $\dot{y}$, $\dot{z}$ (km/s)
1 [16]	113.75, 3950, 5150	0.94, 3.33, -6.0125
2	114.25, 3945, 5140	0.92, 3.2, -6.0

1) Communications and Computing: The communication losses are again simulated by using TCP and DSTCP message delivery rates as described in the last subsection. For correlation/data association, we consider a simple model: the state estimate of Target i is assigned to Target j by the correlator with probability a_{ij}:

$$a_{ij} = \begin{cases} 1-(M-1)p, & i = j \\ p, & i \neq j \end{cases}, \quad i,j = 1,...,M, \quad M \geq 2, \tag{21}$$

where M is the number of targets, and p is the probability that we assign a state estimate to any incorrect target. For the case where $M = 1$, data association is not required since there is only one target that the state estimate can belong to (i.e., for $M = 1$, $p = 0$). In these simulations we will only examine the more interesting case where $M \geq 2$.

We model p here to be the following function of the number of targets, M, and the fixed execution time, τ:

$$p = \frac{1}{M(1+\beta\tau)}, \quad p \leq \frac{1}{M}, \tag{22}$$

where β is a scaling parameter. This function has the following properties: (a) as the number of targets M increases, the total probability of misassignment, $(M-1)p$, increases; (b) as τ increases, p decreases; (c) if $\tau = 0$, there is an equal probability of assigning a state estimate to any target; and (d) as $\tau \to \infty$, $p \to 0$. However, of note is that p is unlikely to be zero given an unlimited execution time for the correlator; more likely there may be some lower bound p_{min} on the probability p, which may be added to p such that if $\tau \to \infty, p \to p_{min}$.

Once the received state estimates have been assigned to a target, a linear fuser is then employed to fuse the state estimates assigned to each target, which is given by

$$\hat{X}_j^F = \left(\sum_{i=1}^{L} P_{j_i}^{-1}\right)^{-1} \sum_{i=1}^{L} P_{j_i}^{-1} \hat{X}_{j_i}^W, \tag{23}$$

where $\hat{X}_j^F$ is the fused estimate of target j, $\hat{X}_{j_i}^W$ is a received state estimate that has been assigned to target j and P_{j_i} is its corresponding state error covariance matrix, and L is the total number of estimates that have been assigned to this target.

2) Performance Bounds: The mean-squared error (MSE) is used as the quality measure $Q_{CC}\left(\hat{X}_1,...,\hat{X}_N;\tau\right)$ for the fused result to analyze the lower bound derived in Eq. (3). The expected performance measure based on receiving k state estimates, $\bar{Q}_C(k;\tau)$ in Eq. (4), is computed for Target j using the following conditional expectation, conditioned on the state estimates that are assigned to target j:

$$f = E\left[\left(X_j - \hat{X}_j^F\right)^2 \mid \hat{X}_{j_1}^W,...,\hat{X}_{j_L}^W \in \Lambda_j\right], \tag{24}$$

where Λ_j is the group of correlated state estimates $\hat{X}_{j_1}^W,...,\hat{X}_{j_L}^W$ that corresponds to target j, and $\hat{X}_j^F$ is as defined in Eq.

(23). By taking the expected value of Eq. (24), we obtain the following expression for $\bar{Q}_C(k;\tau)$ for target j:

$$\bar{Q}_C^{(j)}(k,\tau) = \sum_{\hat{X}_{j_1}^W,\ldots,\hat{X}_{j_L}^W \in \Lambda_j} f \cdot \mathbf{P}\left\{\hat{X}_{j_1}^W, \ldots, \hat{X}_{j_L}^W \in \Lambda_j\right\}. \quad (25)$$

This summation is taken over all possible sets of received state estimates that are assigned to target j.

The probability that k state estimates arrive within the given time window is given by

$$\mathbf{P}\left\{t_{\hat{X}_{i_1}^W}, \ldots, t_{\hat{X}_{i_k}^W} \in [t, t+W]\right\} = \binom{N}{k} p_l^k (1-p_l)^{N-k}. \quad (26)$$

where p_l is the message delivery probability, and N is the total number of state estimates from the sensors. The lower bound on the quality of the overall fused estimate is then given by

$$\mathbf{P}\left\{Q_{CC}\left(\hat{X}_1, \ldots, \hat{X}_N; \tau\right) < \delta | \hat{X}_1, \ldots, \hat{X}_N\right\} \geq \sum_{k=0}^{N} \left(\frac{1-\bar{Q}_C^{(j)}(k,\tau)}{\delta}\right) \binom{N}{k} p_l^k (1-p_l)^{N-k} \quad (27)$$

3) Simulation Results: A total of 5,000 monte-carlo simulations were run for a 60-second trajectory with loss rates of 0 and 0.5 and computing time of $\tau = 0.1$ and 1 (which correspond to target misassignment probabilities of 0.33 and 0.08, respectively, from Eq. (22)). Message delivery probabilities were computed for the TCP and DSTCP protocols using Eqs. (8) and (10). An equation for the theoretical lower bound for the position MSE[2] in one coordinate for these simulations was derived using Eq. (27), and these bounds as well as the estimated probability $\hat{P}\left\{Q_{CC}(\hat{X}_1, ..., \hat{X}_N; \tau < \delta\right\}$ from the simulations are shown in Fig. 5. The effects of communications can be seen for fixed τ, and the effects of computing are shown for fixed loss rates. The theoretical bound in Eq. (27) is the minimum guaranteed probability for the final error to be below δ, and its estimate $\hat{P}\{.\}$ based on these simulations is consistently better, thereby confirming that the guarantee is indeed met. The plots in Fig. 5 confirm that when the DSTCP protocol is used, the performance of the fused estimate improves in all cases (in both the tightness of the bounds and the quality measures). Overall, the bounds are fairly tight for longer computing time and high message delivery probabilities, but worsen as the loss rate increases and the computation time is shortened, although the computing time appears to have a stronger impact. This suggests that the effects of computing on the overall quality may not be as strong as the bound indicates.

Without the knowledge of the underlying distributions, however, $\bar{Q}_C(k,\tau)$ cannot be directly computed; but the regression $\hat{r}_C(k)$ can be used as its approximation. Using the same 60-second trajectory with $\tau = 0.1$ and a loss rate of 0.5, overall quality measurements were collected over 100 simulations to yield 6,000 test measurements (i.e., $l = 6,000$ in Eq. (16)). Similarly to Fig. 4(c), Fig. 6 shows a plot of the final MSEs as a function of the number of estimates received in one simulation, along with its regression fit. Fig. 7 shows a plot of the regression fit and the corresponding theoretical MSE.

[2]The MSEs are normalized so that a value of 1 represents the highest error.

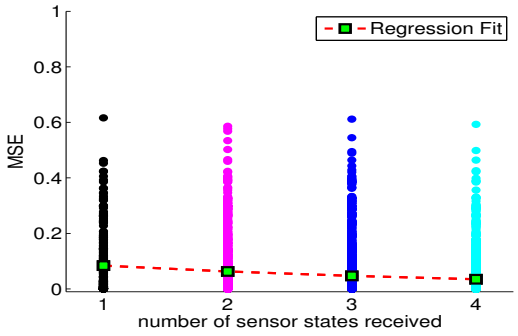


Fig. 6. State errors as a function of the number of estimates received and its corresponding regression fit.

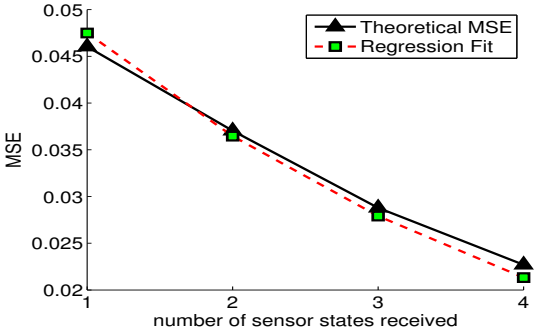


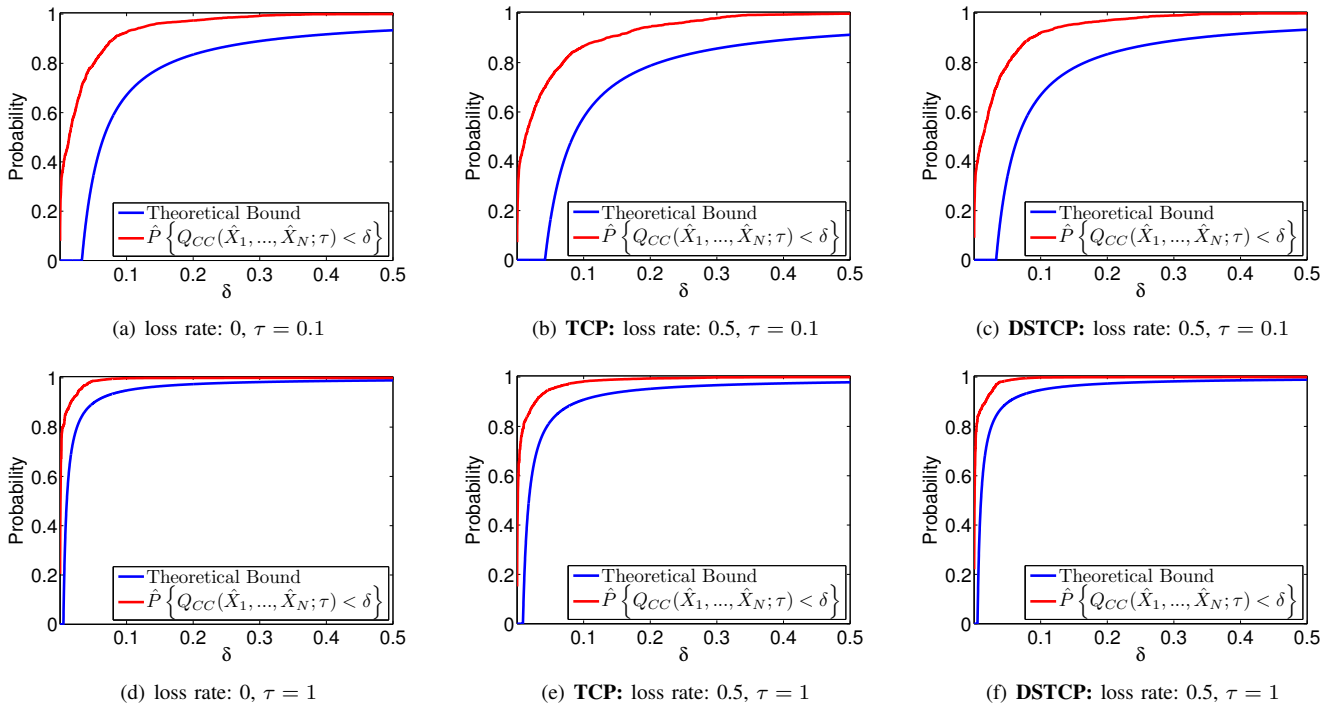
Fig. 7. Regression fit and Theoretical MSE

Putting it all together, Fig. 8 shows a plot of the probability that the final state estimate will be below some δ, the theoretical lower bound on this probability using $\bar{Q}(k,\tau)$, and the estimated lower bounds generated using the regression function $\hat{r}_C(k)$ which was computed 1,000 times over different simulations. This plot also shows the upper and lower bounds of the confidence interval (computed using Eq. (16) for a confidence level of 0.9), demonstrating that a majority of the bounds computed using the regression estimates fall within this interval, with 8.1% of the estimates outside of the confidence interval in this particular example.

V. Conclusions

We presented an analytical model to assess the performance of state fusion over a long-haul sensor network wherein the state estimates are correlated and fused at the fusion center to generate global state estimates. In particular, we analyzed the performance measure $Q_{CC}(\tau)$ of the global state estimate computed within the allocated time τ under varying degrees of network loss and latency. Its performance bound enabled us to separate the contributions of computing and communications, and we illustrated their individual qualitative effects on the fused state estimates using simulations. We also showed that the measurements of $Q_{CC}(\tau)$ based on field tests can be used to derive its confidence level.

We consider this work to be an initial step in assessing the effects of computing and communications on the state estimation over long-haul networks, as demonstrated in the simulations. While the simulations are somewhat simplified, they illustrate the potential of this method to provide qualitative insights into the system performance. More detailed analysis



(a) loss rate: 0, $\tau = 0.1$ (b) **TCP:** loss rate: 0.5, $\tau = 0.1$ (c) **DSTCP:** loss rate: 0.5, $\tau = 0.1$

(d) loss rate: 0, $\tau = 1$ (e) **TCP:** loss rate: 0.5, $\tau = 1$ (f) **DSTCP:** loss rate: 0.5, $\tau = 1$

Fig. 5. Plots of the probability that the overall MSE is less than δ for different values of τ and loss rates using the TCP and DSTCP protocol. The lower bound for this probability given in Eq. (3).

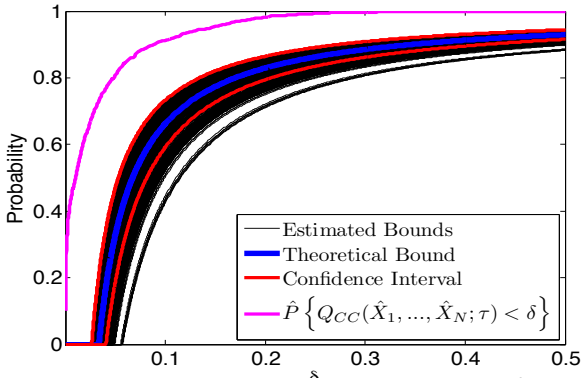


Fig. 8. Plot of the probability that the overall MSE is less than δ for $\tau = 0.1$ and a loss rate of 0.5, and the theoretical and estimated lower bounds generated using the regression estimates. Also shown is the corresponding confidence interval bounds for $\hat{r}_C(k)$ for a confidence level of 0.9.

of the correlation and fusion algorithms would be of future interest. Also, more extensive system-level simulations and measurements from deployed systems would provide better insights into the practical system performance.

Acknowledgments

This work is funded by the Mathematics of Complex, Distributed, Interconnected Systems Program, Office of Advanced Computing Research, U.S. Department of Energy, and SensorNet Project of Office of Naval Research, and is performed at Oak Ridge National Laboratory managed by UT-Battelle, LLC for U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

[1] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*. John Wiley and Sons, Pub., 2010.

[2] D. T. H. Lai, M. Palanoswami, and R. Begg, Eds., *Healthcare Sensor Networks: Challenges Towards Practical Implementation*. CRC Press, 2011.

[3] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishers, 2011.

[4] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.

[5] Y. Xiao, H. Chen, and F. H. Li, Eds., *Handbook on Sensor Networks*. World Scientific, 2010.

[6] F. Zhao and L. Guibas, *Wireless Sensor Networks*. Elsevier, 2004.

[7] "Earth systems research laboratory, national oceanic and atmospheric administration," //www.esrl.noaa.gov/gmd/ccgg/index.html.

[8] "Network telescope research," 2011, www.caida.org/research/security/telescope/.

[9] U. of Southampton, "Global network of new-generation telescopes will track astrophysical events as they happen, *ScienceDaily*," Jan. 2011.

[10] W. Boord and J. B. Hoffman, *Air and Missile Defense Systems Engineering*. CRC Press, 2012.

[11] A. B. Poore, "Multidimensional assignment problems in arising in multitarget and multisensor tracking," in *Nonlinear Assignment Problems: Algorithms and Applications*, P. M. Pardalos and L. S. Pitsoulis, Eds. Kluwer Academic Publishers, 2000, pp. 13–38.

[12] N. S. V. Rao, "Measurement-based statistical fusion methods for distributed sensor networks," in *Distributed Sensor Networks*, S. S. Iyengar and R. R. Brooks, Eds. Chapman and Hall/CRC Publishers, 2005.

[13] V. N. Vapnik, *Statistical Learning Theory*. John-Wiley and Sons, New York, 1998.

[14] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part ii: Motion models of ballistic and space targets," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 1, pp. 96–119, Jan. 2010.

[15] M. Yeddanapudi, Y. Bar-Shalom, K. R. Pattipati, and S. Deb, "Ballistic missile track initiation from satellite observations," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, no. 3, pp. 1054–1071, Jul. 1995.

[16] T. H. Kerr, "Streamlining measurement iteration for ekf target tracking," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 408–421, Mar. 1991.