RC24404 (W0711-038) November 7, 2007
Computer Science

IBM Research Report

A Framework for Quality of Information Analysis for Detection-Oriented Sensor Network Deployments

Sadaf Zahedi, Mani B. Srivastava
Electrical Engineering Department
University of California, Los Angeles
Los Angeles, CA 90095-1594

Chatschik Bisdikian
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

A Framework for Quality of Information Analysis for Detection-oriented Sensor Network Deployments

Sadaf Zahedi *and* Mani B. Srivastava
Univ. of California, Los Angeles
Electrical Engineering Department
Los Angeles, CA 90095-1594, USA
{*szahedi,mbs*}*@ee.ucla.edu*

Chatschik Bisdikian
IBM Corp.
T. J. Watson Research Center
Hawthorne, NY 10514, USA
bisdik@us.ibm.com

Abstract

With their increasing capabilities, research efforts in sensor network have been spreading on a variety of aspects. In this paper, we present a system-oriented, layered approach for evaluating their application-related performance. Focusing on sensor-enabled detection systems, an application planners point of view of the system is considered and a hypothesis-testing-based analysis framework for evaluating the quality of information (QoI) supported by a sensor network deployment is explored. The QoI properties of centralized, distributed and hybrid decision topologies are investigated and trade-offs explored at the sensor, cluster, and system-level. In the process, the computationally powerful concept of a QoI equivalent sensor is presented and applied in the aforementioned explorations. Finite size networks are considered and limiting behavior and dominance properties are also investigated. Finally, extensions of the analysis framework to faulty sensor and the impact of calibration are also investigated.[1]

1. Introduction

With advances in computing and communication technologies, low(er) cost, intelligent sensor-enabled systems (from single-sensor to highly distributed multi-sensor systems) find their way in a multitude of application environments in areas as diverse as the habitant monitoring, forest monitoring, utility grid monitoring, environmental control, machinery control, intelligence gathering and enemy activity surveillance, and so on.

Commensurate to their increasing capabilities, recent years have seen increased research efforts in several areas related to sensor systems including ad hoc deployment and operation of sensor networks, energy-aware architectures and protocols, coverage and localization, efficient query dissemination, sensor calibration and error-management, data cleaning, and so on. These broad research efforts notwithstanding, we find increased need for additional research directions that focus on (or, at least, are highly influenced by) the need to develop generic computational aids that would allow (sensor) application planners and the sensor system designers that support them to study design alternatives in advance of deploying sensor networks.

Specifically, our work is motivated by the following usage scenario. An application *planner* plans the deployment of a sensor-enabled application. While the planner sets the requirements for the application, he/she employs a sensor system *designer* to design and deploy a sensor system that would (hopefully) meet the information needs of the application planned. In a fashion analogous to QoS (quality of service) and SLAs (service level agreements) in networking, the planner specifies the information needs of his/her application via the *quality of information* (QoI) derived from the sensor system, underscoring his/her application-centric interest/view of the problem (i.e., capturing the effectiveness of the end-result), rather than a sensor-network-centric view of the problem.

QoI has typically tied to the study of structured, in some sense, information like data residing in databases and considers such information aspects like data consistency, completeness, currency, etc. [1, 2, 3]. However, the QoI topic is very broad and still quite open when applied to sensor networks. While the development of a definition for QoI is still a work in progress, we use QoI to capture the effectiveness or capability of the sensed data and the information derived from them to "paint" those aspects of

[1]Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

the real world, that are of importance accurately enough for an application to perform its task at a required level of effectiveness.

How QoI is expressed is very closely related to the application that will use the information. Given the wide application space for sensor networks, for our usage scenario we have elected to consider the family of *event detection* applications. This application family is at the core of a wide range of sensor-enabled decision making (or, action taking) tasks. Specifically, we consider events that produce "wave-like" signals, the *event signature*, like acoustic, seismic, electromagnetic, etc., typically encountered during monitoring for surveillance and intelligence gathering purposes like detecting presence of enemy weaponry, hostile activities (e.g., gunfire, explosions), monitoring remote territories, and so on. For this family of applications, how accurately the real world can be painted will depend on how effectively the sensor-based detection system detects the event. Thus, we elect as QoI attributes of interest the *detection probability* P_d of correctly detecting the occurrence of the event and the *false alarm rate* P_f, i.e., the probability of declaring the occurrence when it did not occur. Later in the paper, we will use classical hypothesis testing for detecting signals in noisy environments [4, 5] as the key analysis technique for these QoI attributes. Note that we point to explosion and gunfire detection type of applications as representative of transient events, and underscore that in our research we do not want to be limited only to the more traditional persistent events.

In order for the designer in our usage scenario to design and deploy a detection system that achieves the required QoI levels set-forth by the application planner, it would be advantageous that he has at his/her disposal a computational aid that can use to easily test different design configurations and evaluate the QoI and sensor-system deployment trade-offs. Furthermore, for this computational aid to be useful, it must be reusable in many situations and, thus, it would also be advantageous that it is not limited to a specific set of system and modeling assumptions, e.g., a specific sensor node topology, or signal propagation models but allows different system models and design objectives to be considered. In pursuing these objectives, we introduced in [6] the first instance of a QoI-inspired analysis framework that can serve as the foundation for creating the reusable computational aid. The framework serves as a solutions methodology (a *meta-solution*, if you like) that captures the key aspects of solving detection problems that can be reused as the designer tests for candidate deployments under various assumptions.

In [6], we highlighted the framework and considered at a high-level a few study cases. In this paper, we both broaden and deepen our presentation in these topics. Specifically, in this paper, we provide analytical results for QoI for networks with finite number of sensors with arbitrary topologies, along with computationally attractive evaluation expressions, we also study their limiting behavior and derive fundamental dominance relationships that aid in establishing bounding approximations to the system performance. We study the relative performance of various detection architectures and make use of equivalent sensors to simplify this analysis. We also study the sensitivity of the performance results with respect to the positional accuracy of sensors relative to the event location. Finally, we present sensor model generalizations that accommodate potentially faulty operations, e.g., uncalibrated sensors.

While, we are not aware of a prior work on a framework along the aforementioned objectives, considerable amount of work has been done in modeling and analyzing the detection systems considering various parameters and models for signal-to-noise ratio (SNR), channel fading, spatial correlations, and so on [7, 8, 9, 10, 11]. In [12] an optimum fusion rule for a multi-sensor system is derived, when the probabilities of detection and false alarm of each sensor is known to the fusion center. In [9] a decision fusion rule is proposed based on a "counting" policy, where the analysis is based on a Poisson sensor distribution model and consider the limiting performance of the system as the number of sensors increases. In [10] a inhomogeneous situation has been considered as a result of the spatial distribution of the nodes and channel fading and a system-level analysis is provided just for large number of sensors. In [11] the decision fusion algorithm is also studied and system level performances for large networks. Additionally, an approximation for the system decision threshold is derived that provides performance guarantees when the false alarm probabilities are identical for all sensors. In [13] and [14] a hybrid (neither centralized, nor distributed) energy-driven detection scheme is proposed based on the binary observations of the sensors and study the trade-off between detection accuracy and energy consumption. The above studies consider fixed and persistent events.

Any of these analysis can be incorporated within the framework to provide performance results for the the specific models studied in them. The organization of the paper is as follows: In section 2, we introduce the reference model for the system under consideration and the general toolkit framework. In section 3, we present the core analysis approach based on hypothesis testing. In section 3, we discussed the application of the core QoI analysis in detection sensor-systems. In sections 5 and 6, we studied two steps that can be incorporated as a part of post-processing and preprocessing layers of our framework, and finally we conclude in section 7 with some concluding remarks.

2. The reference detection system model and QoI analysis framework

The analysis framework is based on a fairly general detection system model, which we introduce next.

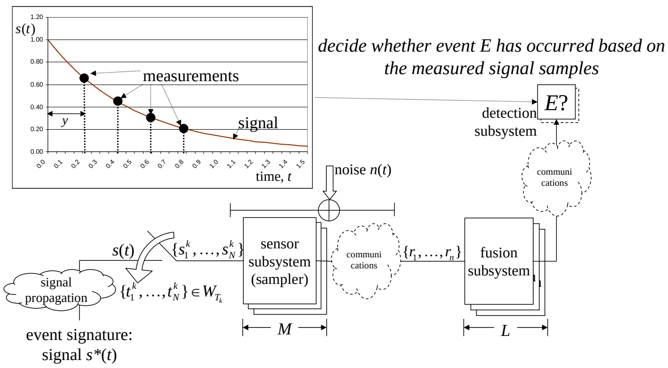

Figure 1. Functional architecture of the reference detection system.

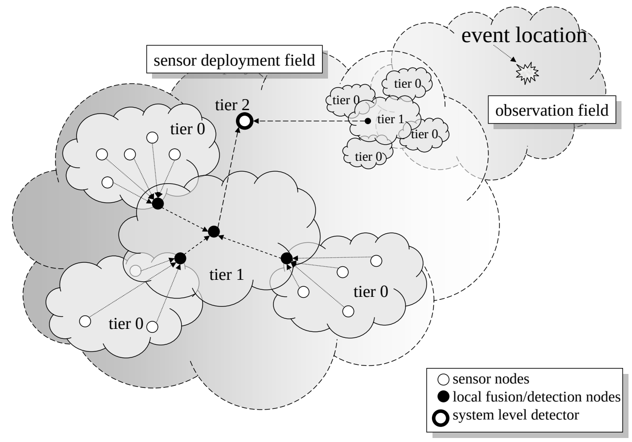

Figure 2. Multi-tiered detection topology.

2.1. The reference detection system

Figure 1 shows the reference architecture of our sensor-based detector system. It partitions the sensing operation into three functional subsystems: (a) the *sensor subsystem* or *sampler*; (b) the *fusion subsystem*; and (c) the *detection subsystem*. Each of these subsystems operates at a different level of the event detection process. The sensor subsystem, comprising a collection of $M \geq 1$ sensor nodes, samples the physical world (in search for an event signature) and pass these samples to the fusion subsystem. The fusion subsystem, comprising a collection of $L \geq 1$ fusion centers, operates on the samples it receives (which could be corrupted by noise) to produce a "summary" description of the samples. Finally, this summary is used by the detection subsystem to decide whether an event of interest has occurred or not.

Based on the system topology considered, there could be multiple detection subsystems, some of them performing local detection (concerned with what a subset of sensors "says" about the event occurrence) and *one* associated with the system-level detection (concerned with what the overall system says about the event occurrence). The three subsystems may be collocated or separated as could be the case of a networked sensing system. At current time, we focus only on the information, e.g., samples, to be transferred between these subsystems and not how the transfer process may affect them.

According to the reference model, the sensor subsystem records observations $\{r_1, r_2, \ldots\}$ from environmental samples that it takes. These records comprise samples $\{s_1, s_2, \ldots\}$ of the event signature (if the event occurred, otherwise the s_i's are null) and noise. When the event has occurred, the sequence $\{s_1^k, s_2^k, \ldots\}$ (we will use the sensor index k only when necessary) can be thought of as samples of the signal $s^k(t)$ that is the *projection* of the original event signature $s^*(t)$ projected at the location of the k-th sensor, $k \in \{1, \ldots, M\}$. As discussed in [6], the projection metaphor is key in developing our framework, to be discussed shortly.

The general system model in figure 1 is applicable to a variety of detection architectures/topologies including single- and multiple-sensor networks, centralized, distributed, single-tiered and multi-tiered hybrid topologies (where detection decisions made at one tier combine their decisions to support decision making at a higher tier). Figure 2 shows a fairly general sensor detection topology that covers the aforementioned alternatives. Note that for simplicity reasons, the fusion and detection subsystems for the detection system in figure 2 are assumed to be co-located, however, we do not restrict our system to this case only. Finally, the figure eludes to the rather general *area of engagement* for the sensor deployment that we consider. The area of engagement comprises a *deployment field* where sensor can be deployed and an *observation field* (or, incident field) where the events of interest occur. These two fields may have an arbitrary geography, comprised contiguous or non-contiguous regions, overlap or not, etc. This area of engagement represents a generalization of the sensor coverage models typically considered in the literature where sensors are located in the midst of the area that they sense.

2.2. The QoI analysis framework

Prior research in the area has studied sensor-network based detection for very specific system models, e.g., for a given signal signature (quite often a persistent one with constant amplitude), a given deployment architecture, e.g., either centralized and/or distributed, possibly a given propagation/attenuation model for the event signature, and so on. While such analysis provides significant insight to the performance of a potential deployment, such an insight would closely reflect and be intimately tied to the analysis assumptions made. However, the specific analysis assumptions may bear limited (if any) resemblance to the application realities and requirements (e.g., transient vs. persistent signals, non-overlapping deployment and observation fields, non-homogenous deployment and observation fields, different signal propagation models, and so on). Therefore, we have opted to consider a "higher-level" analysis methodology, or *analysis framework*, that will allows us to compose specific solutions from any number of existing or new analysis techniques.

Our framework is based on identifying key analysis procedures within general context of our reference systems. Based on the aforementioned projection metaphor, we noted that anything that influences the event signature propagation from an event location to a sensor can be separated from the subsequent detection (QoI) analysis. Specifically, since decision making is really made based on the recorded observations, if the signature projections were somehow known, one could have proceeded with the detection QoI analysis unbeknownst to how they were formed, which in general is the result of the original event signature, of course, and the geometry of the system and the geography it is deployed in.

With the above observations in hand, we proposed in [6] an analysis framework comprised three major processing layers: (a) input pre-processing; (b) detection QoI analysis (the core analysis engine); and (c) output post-processing. The input pre-processing is related to anything that may influence the original event signature until it is recorded by the sensor (i.e., it generates the sequence of projection samples) like the deployment and observation topologies (which determine the relative geography between sensors and events), the signal propagation and attenuation models (which determine how the original signal projects itself at the sensor locations), the sampling policies (which determine which sensors contribute which samples to the detection process), the noise models (which determine the distortion process of the signal), the measurement error models (which model the errors in the reported value from the sensor platform faults) and so on. The QoI core analysis engine calculates the QoI attributes for the derived set of projection samples based on a hypothesis testing formulation, to be highlighted later in the section. Finally, given the application requirements, post-processing of the QoI analysis results may be necessary, for example, to calculate averages over an observation region, or to calculate optimal position of sensors, performance sensitivity analysis (which we give an example of later in the paper), and so on. During post-processing, services of the core analysis engine may be requested again for the QoI analysis of the system for a different set of system parameters.

The analysis framework is summarized in figure 3, which shows an instance of the framework on the right and the role of its users, the planner and the designer, on the left. Specifically, the planner specified the desired application and its constraints, e.g., the area of engagement, the application domain, which may reflect the class of event sig-

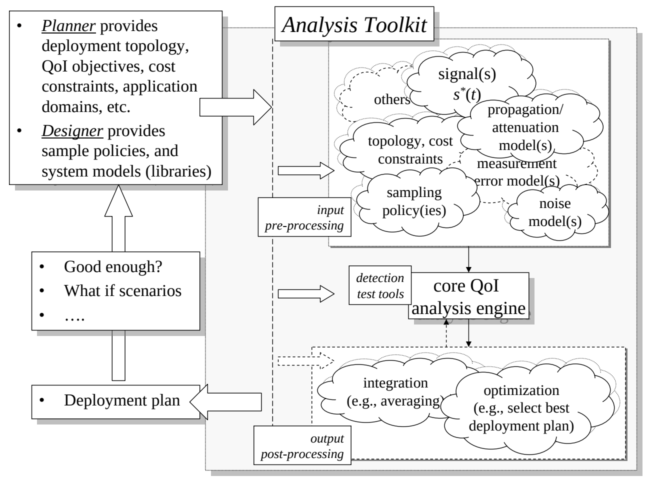

Figure 3. The QoI analysis framework and toolkit.

natures to be detected, the desired QoI levels, budget constraints, and so on. Consequently, the designer supplements these in his/her analysis tool with system level models, e.g., propagation models, noise models, communication and interference models, sampling policies, and so on, that are appropriate for the the deployment in the area of engagement. These libraries may already be present in an analysis toolkit built around the framework or be custom built for the specific deployment, and thus becoming part of the library for future use. The framework in figure 3 was drawn with this system-oriented toolkit view of it. The clouds shown in the processing layers of the framework represent analysis libraries that the designer can choose from to evaluate the QoI performance of alternative deployment plans. Finally, when a deployment plan with the corresponding QoI levels have been evaluated, the designer in consultation with the planner may decide as to whether the expected performance of the system is satisfactory enough, or additional analysis and trade-off studies for "what if" situations are required. Note that in addition, to planning a future deployment, the designer can use the toolkit to evaluate the QoI performance of an existing system and study whether any fine-tuning of its operational parameters is required.

Each of the processing layers can be studied on its own right separately to the extend necessary, e.g., utilize portions of existing studies like the ones referenced in this paper. Under the framework umbrella, these separate studies can be combined to provide answers to a variety of deployment situations in a reusable fashion. Note that prior detection system studies that focus a specific set of system assumptions represent monolithic (non-layered) top-down interpretations of the framework which increases analysis complexity (the entire problem needs to be studied in one-shot) and reduces application adaptability.

We close this section with a definition of an equivalent sensor, introduced in [6], that we use for simplifying the analysis of certain sensor fusion topologies.

Definition 1. *The equivalent sensor of a multi-sensor system is a single-sensor sensing system that achieves the same QoI level as the multi-sensor system using the same observation set.*

Equivalent sensors are not real sensors but rather mathematical artifacts to aid analysis. When an equivalent sensor, e.g., detector sensor, can be constructed, analysis tools (including both analytical and simulation ones) could be built that serve the user of the sensor-enabled applications much better than tools that expose the complexity of the entire sensor network to them.

3. The core QoI analysis engine: Likelihood ratio tests & hypothesis testing

We build or core QoI analysis engine for our framework around the hypothesis testing analysis techniques [5, 4] Event detection analysis is based on tests to decide whether the hypothesis of the occurrence of an event is supported by the observations made or not. Next, we highlight the key results from these techniques and later on we apply them for the QoI analysis of specific detection architectures. We consider binary testing, but extensions to multiple-hypothesis testing are also possible.

In order to test the hypothesis that an event of interest occurred (hypothesis H_1), or not (the *null* hypothesis H_0) the detector accumulates observations $r_1,\dots,r_N$ which satisfy under

$$\text{hypothesis } H_1: \; r_i = s_i + n_i, \text{ or, under} \tag{1a}$$

$$\text{hypothesis } H_0: \; r_i = \quad n_i, \tag{1b}$$

where, under hypothesis H_1, s_i represents the value of the signal at the i-th sampling instance, while, under both hypotheses, n_i represents an additive noise component that is added to the i-th sample, and r_i represents the i-th measurement that is contributed to the fusion subsystem. The test is represented by the *likelihood ratio test* (LRT) comparing a ratio of conditional probabilities to a threshold:

$$\Lambda(\mathbf{r}_N) = \frac{P(\mathbf{r}_N|H_1)}{P(\mathbf{r}_N|H_0)} \underset{\text{select } H_0}{\overset{\text{select } H_1}{\gtrless}} \eta, \tag{2}$$

where $P(\mathbf{r}_N|H_z)$ represents the probability density function for the observation vector $\mathbf{r}_N = [r_1,\dots,r_N]^T$ conditioned on hypothesis H_z, $z \in \{0,1\}$; for notational brevity, in the sequel, we will skip the size index N unless necessary. For a zero mean, additive and stationary Gaussian noise process with covariance matrix $\mathbf{C} = E\left(\mathbf{n}^T\mathbf{n}\right)$, where $\mathbf{n} = [n_1,\dots,n_N]$, the LRT in (2) becomes:

$$l \triangleq \mathbf{r}^T\mathbf{C}^{-1}\mathbf{s} \underset{H_0}{\overset{H_1}{\gtrless}} \eta^* \triangleq \ln(\eta) + 0.5\,\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}. \tag{3}$$

The expression $\psi^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$ represents the *signal-to-noise ratio* (SNR) for this system. The covariance matrix $\mathbf{C}$ of the noise process in (3) can capture correlations across both the spatial (i.e., across the sensors) and temporal (along a single sensor) dimensions. Actually, the entries in the covariance matrix are strongly related to the way that samples from the various sensors contribute to the decision making, which we refer to as the sampling policy. Note that if the noise samples exhibit only spatial or only temporal correlation we may reorder the sample contributions at the fusion subsystem in such a way that $\mathbf{C}$ becomes block diagonal, and hence $\mathbf{C}^{-1}$ is block diagonal too with each block in the latter being the inverse of each block in the former matrix.

With regard to the reference system in figure 1, l, which is a *sufficient statistic* for this test, represents the summary operation performed by the fusion subsystem. On the other hand the comparison between l and η^*, for deciding in favor of the one or the other hypothesis, is performed by the the detection subsystem. The Bayesian hypothesis testing, where the a priori probabilities for the two hypotheses, P_0 and P_1 are assumed known, minimizes the cost of making a decision and [5]

$$\eta = \frac{P_0}{P_1} \cdot \frac{C_{10} - C_{00}}{C_{01} - C_{11}}, \tag{4}$$

where C_{ij} represents the cost of deciding in favor of H_i when H_j is true. The Neyman-Pearson test is an alternative LRT which does not utilize a priori probabilities and its objective is to maximize the probability of detection maintaining a predefined false alarm rate.

4. Detection sensor-system topologies

In this section, we apply the core QoI analysis from the last section to a collection of specific but very important sensor systems: (a) single-sensor detector; (b) centralized multi-sensor detector; (c) fully distributed multi-sensor detector; and (d) hybrid multi-sensor detector. While results presented have independent merit, like the dominance relationships in section 4.3, the following presentation should also be viewed as a series of steps followed within the core QoI engine to accommodate even more advanced deployment scenarios.

4.1. The single-sensor detection system

The single-sensor detection systems and its QoI performance serves as the building block for the QoI analysis of the multi-sensor considered later in the paper. The probability of detection P_d and false alarm rate P_f are given by [5, 15]:

$$P_d = \Pr(l \geq \eta^\star|H_1) = 1 - \Phi\left(\frac{\ln(\eta)}{\psi} - \frac{\psi}{2}\right), \text{ and} \tag{5a}$$

$$P_f = \Pr(l \geq \eta^\star|H_0) = 1 - \Phi\left(\frac{\ln(\eta)}{\psi} + \frac{\psi}{2}\right), \tag{5b}$$

respectively, where $\Phi(\cdot)$ is the cumulative distribution function of the normalized Gaussian random variable $\mathscr{N}(0,1)$. For uncorrelated noise samples, the SNR ψ^2 is given by $\psi^2 = \left(\sum_{i=1}^{N} s_i^2\right)/\sigma^2$.

Variations and different interpretations of the formulae in (5) will be used subsequently in the analysis of multi-sensor detection systems.

4.2. L=1: Centralized multi-sensor detection system

The centralized multi-sensor system with only one fusion subsystem ($L=1$) and one detection subsystem serves a performance benchmark for any detection topology and we highlight here a key result from [6]. Specifically, the expressions in (5) still holds true with $\psi^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$, see discussion following (3). For a diagonal covariance matrix of the form $\mathbf{C} = \mathrm{diag}\left(\sigma_1^2\mathbf{I}_{N_1}, \sigma_2^2\mathbf{I}_{N_2}, \ldots, \sigma_M^2\mathbf{I}_{N_M}\right)$, the system-wide SNR is given by

$$\psi_{sys}^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = \sum_{k=1}^{M}\left\{\frac{1}{\sigma_k^2}\sum_{i=1}^{N_k}(s_i^k)^2\right\} = \sum_{k=1}^{M}\psi_k^2, \quad (6)$$

where ψ_k^2 is the SNR attributed to samples from sensor k. In other words, the system-wide SNR is decomposable to the SNRs at the individual sensor level.

It follows from the above and definition 1 that:

Corollary 1. *A centralized multi-sensor detection system with a Gaussian noise process possesses an equivalent sensor with SNR* $\psi^2 = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}$. *If in addition the noise process is an independent process as well, the SNR is decomposable as in* (6).

We will make use of the uniformly equivalent sensor when later in this section we consider the performance of 2-tier (hybrid) architectures like the one shown in figure 2 with only tiers 0 and 1 involved.

4.3. L=M: Fully distributed multi-sensor detection system

At the other end the broad spectrum of detection architectures lies the fully distributed ($L=M$) detector. In the fully distributed (or, simply, distributed) case, each sensor is bound to a separate, dedicated fusion and detection subsystems. In the distributed detector (and more generally in any non-centralized detector), decision is made in two steps. In the first step, sensors make *sensor-* or *local-level decision* and then the local decisions are fused to obtain the *system-level* ones. Clearly the QoI performance at the sensor level follows from the study of the single-sensor detector in section 4.1, where the SNR ψ_k (instead of ψ) in (5) is calculated separately for each sensor k.

Fusing the local decision to derive the system-level decision requires the use of a *detection policy*. A variety of policies can and have been considered and typically will involve some form of a counting strategy, e.g., decide that the event has occurred if at least, say, Q out of the M sensors indicate that it has. A more general weighted sum approach has been studied in [12] where each sensor's decision (+1 in favor of H_1 and -1 in favor of H_0) is weighted by the probability of miss and false alarms. As previously mentioned, a library from these policies can easily be adopted and incorporated with the library of models and analysis techniques that a toolkit build around our framework can have. The rather elaborate expression that results from the weighted policy in [12] is not very amenable to analytical evaluation with regard to the QoI analysis. Instead, next we will study the QoI performance of a counting-based detection policy.

Definition 2. *Let a* Q-count policy *be the detection policy according to which a decision in favor of* H_1 *is made if at least* Q *out of the* M *sensors decide in favor of the event occurrence. We will refer to this as the* $\{Q,M\}$ *detection system.*

Let P_k^d and P_k^f represent the probability of detection (under hypothesis H_1) and false alarm (under hypothesis H_0) for sensor k, respectively. Due to the similarity of several of the expressions involving these two probabilities, we may will also write generically P_k^z, where $z \in \{d, f\}$ for these probabilities. The system-wide probability $P_z(Q;M)$ is equal to the probability that of all collections $\mathscr{S}_q^M$ of sets of sensors $\mathbf{x}_q$ with exactly q sensors, there is at least one set $\mathbf{x}_q$ with $q \geq Q$ sensors all of which declare in favor of the event occurrence, i.e., (where $y=1$ when $z=d$, and 0 when $z=f$):

$$\begin{aligned} P_z(Q;M) &= Pr(q \geq Q | H_y) \\ &= \sum_{q=Q}^{M}\left\{\sum_{\mathbf{x}_q \in \mathscr{S}_q^M}\left[\Big(\prod_{\substack{\text{sensor} \\ m \in \mathbf{x}_q}} P_m^z\Big)\Big(\prod_{\substack{\text{sensor} \\ n \notin \mathbf{x}_q}}(1-P_n^z)\Big)\right]\right\}. \end{aligned} \quad (7)$$

Admittedly as the number of sensors M increases, the computation burden of (7) becomes daunting. However, assuming knowledge of the QoI performance of the $\{Q-1, M-1\}$ and $\{Q, M-1\}$ detection systems and tallying the cases that will bring it up to the $\{Q,M\}$ system, the following simple recursion can be derived:

$$P_z(Q;M) = (1-P_M^z)P_z(Q;M-1) + P_M^z P_z(Q-1;M-1). \quad (8)$$

The recursions have the following trivial boundary conditions, which follow directly from (7):

$$P_z(1;M) = 1 - \prod_{k=1}^{M}\left(1-P_z^k\right), \text{ and } P_z(M;M) = \prod_{k=1}^{M}(P_z^k). \quad (9)$$

In recursion (8), the M-sensor system must be a direct *augmentation* of the $M-1$-sensor system, i.e., constructed by simply adding the M-th sensor to the already existing $M-1$

sensors of the "smaller" network without moving any of them.

In the special, albeit unlikely case, where the probabilities P_z^k are independent of the sensor k, i.e., $P_k^z = p_z$ for all $k \in \{1,\ldots,M\}$, (7) reduces to the well known tail of a binomially distributed random variable:

$$P_z^{eq}(Q;M) = \sum_{q=Q}^{M} \left[\binom{M}{q} (p_z)^q (1-p_z)^{M-q} \right]. \tag{10}$$

Such *equiprobable* detection systems have been considered in the past, especially for Neyman-Pearson LRTs where the false alarm rate (i.e., the p_f in the above notation) is set at its maximum allowable value [9, 11]; [10] analyzes a system under the simplifying assumption that both the probability of detection and false alarm are equal for all sensors. For such systems, it follows from the central limit theorem (CLT) applied to the i.i.d. Bernoulli random variables that "add-up" to the aforementioned binomially distributed random variable, that as M increases:

$$P_z^{eq}(Q;M) \xrightarrow[M\to\infty]{} 1 - \Phi\Big(\frac{Q - Mp_z}{\sqrt{Mp_z(1-p_z)}}\Big), \tag{11}$$

where the convergence is in the sense that the difference of the two terms reduces to 0 with increasing M. However, the above limiting behavior is not unique to the equiprobable system only. It follows from the the generalized CLT [16] applied to the "well-behaved" 0-1 Bernoulli random variables that

$$P_z(Q;M) \xrightarrow[M\to\infty]{} 1 - \Phi\Big(\frac{Q - \sum_{i=1}^M P_i^z}{\sqrt{\sum_{i=1}^M P_i^z(1-P_i^z)}}\Big). \tag{12}$$

Next we establish an important dominance property between the "equiprobable" system above and the general distributed detection system considered here. Let μ_x and σ_x^2 represent the mean and variance of a random variable x.

Lemma 4.1. *Let I_k, $k \in \{1,\ldots,M\}$ be a collection of M independent indicator (i.e., Bernoulli) random variables and $p_k = \Pr(I_k = 1)$. Let also $\overline{I}_k$, $k \in \{1,\ldots,M\}$ be a collection of corresponding "average" i.i.d. indicators such that $\overline{p} = \Pr(\overline{I}_1) = \ldots = \Pr(\overline{I}_M = 1)$ where $\overline{p} = \big(\sum_{i=1}^M p_i\big)/M$. Then, the counter functions $C = \sum_{i=1}^M I_i$ and $\overline{C} = \sum_{i=1}^M \overline{I}_i$ satisfy*

$$\mu_C = \mu_{\overline{C}} \quad \text{and} \quad \sigma_C^2 \le \sigma_{\overline{C}}^2, \tag{13}$$

with equality holding if-and-only-if $p_1 = \ldots = p_M = \overline{p}$.

Proof. The equality of the means follows directly from the definition of C and $\overline{C}$; each of these means is equal to $M\overline{p} = \sum_{i=1}^M p_i$. From the independence of the I_k's and $\overline{I}_k$'s, the variances for the two counter functions are given by:

$$\sigma_C^2 = \sum_{i=1}^M p_i(1-p_i) = \sum_{i=1}^M p_i - \sum_{i=1}^M (p_i)^2, \text{ and} \tag{14a}$$

$$\sigma_{\overline{C}}^2 = M\overline{p}(1-\overline{p}) = \sum_{i=1}^M p_i - \frac{\big(\sum_{i=1}^M p_i\big)^2}{M}. \tag{14b}$$

Therefore, $\sigma_C^2 \le \sigma_{\overline{C}}^2 \Leftrightarrow \big(\sum_{i=1}^M p_i\big)^2 \le M\sum_{i=1}^M (p_i)^2$, which holds true from Chebyshev's inequality with equality holding true iff $p_1 = \ldots p_M$. □

Let us consider a distributed detection system (the original system) and its corresponding equiprobable system that "happens" to satisfy $P_z^{eq} = \big(\sum_{k=1}^M P_z^k\big)/M$. Then, given that decision making in these systems depends on the number of sensors that decide that the event has occurred, it follows directly from the above lemma that:

Corollary 2. *The decision process of the equiprobable detection system exhibits higher variability than that of its corresponding original system.*

Theorem 4.1. *Given an original distributed system employing Q-count detection policy, the system-wide QoI attribute probabilities $P_z(Q;M)$ and $P_z^{eq}(Q;M)$, $z \in \{d,f\}$, satisfy:*

$$\text{as } M \to \infty, \quad P_z(Q;M) \le P_z^{eq}(Q;M). \tag{15}$$

Proof. For a sufficiently large M, the two probabilities can come and *remain* arbitrarily close to their Gaussian limiting values in (11) and (12), in which case they will also remain distinctively apart from each other. Since $\Phi(\cdot)$ is an increasing function of its argument, it follows from Lemma 4.1 and the definition of the equiprobable system that

$$\Phi\Big(\frac{Q - \sum_{i=1}^M P_i^z}{\sqrt{\sum_{i=1}^M P_i^z(1-P_i^z)}}\Big) \ge \Phi\Big(\frac{Q - MP_z^{eq}}{\sqrt{MP_z^{eq}(1-P_z^{eq})}}\Big). \tag{16}$$

The theorem now follows from (11) and (12) again. □

Corollary 3. *As M increases, the performance the QoI attribute for the $\{Q^{eq}, M\}$ equiprobable system is a close approximation to that of the $\{Q, M\}$ original system, when*

$$Q = MP_z^{eq} + \frac{\sigma}{\sigma^{eq}}\big(Q^{eq} - MP_z^{eq}\big) = \sum_{i=1}^M P_i^z + \frac{\sigma}{\sigma^{eq}}\big(Q^{eq} - \sum_{i=1}^M P_i^z\big), \tag{17}$$

where $\sigma = \sqrt{\sum_{i=1}^M P_i^z(1-P_i^z)}$ and $\sigma^{eq} = \sqrt{MP_z^{eq}(1-P_z^{eq})}$.

4.4. $1 < L < M$: Hybrid multi-sensor detection systems

For completeness, we briefly discuss the hybrid detection system where the number of fusion subsystems are somewhere between 1 and M. Assuming a two-tier system, this means that *clusters* of sensors contribute their samples to tier-0 fusion subsystems, i.e., operate in a centralized detection manner analogous to the centralized system in section 4.2. Also, there may be additional sensors that make decisions locally in a manner analogous to the single-sensor system in section 4.1. Decisions from the clusters then fuse with the later local decisions at a tier-1 fusion subsystem in a manner analogous to the second decision step discussed for the distributed detection system in section 4.3. Thus, hybrid detection systems combine elements from all previous architectures presented and their QoI performance analysis follows from a combination of the previous analyses. Note by substituting entire clusters of sensors with their equivalent sensors, this case reduces to the study of a distributed sensor system with L sensors.

The assignment of sensors to clusters is an open research issue but we expect that both performance objectives and geography constraints (e.g., proximity constraints) may influence the assignment. In such assignments, there is an additional trade-off to be considered between cost of data communication and achievable QoI, as it "costs" more to transmit data to a central location to achieve higher QoI.

Next the QoI performance of the various topologies through some numerical examples.

4.5. Sensor topology comparisons

In this subsection, we will compare the QoI performance of two hybrid systems with a fully distributed system and also with its corresponding centralized system as discussed in sections 4.2, 4.3 and 4.4. For this comparisons we assume the observation of an event occurring at a specific location, i.e., the observation field is a single-point set and the sensors are deployed at specific locations relative to the event location. This is indeed a simple case, but since we are interested in comparing different systems against this case, it provides a clear basis for the comparison. More elaborate deployment and observation field are under study.

Let the physical topology of the system be represented by the distance vector $\mathbf{d} = [d_1, \ldots, d_M]^T$, where d_k is the distance of the path k that a signal takes from the event location to sensor k. Over path k, let $a_k(t)$ be the *attenuation* the signal experiences, and let v be the propagation speed for the signal. Assuming that an event occurs at time $t = 0$ and possesses the (transient) event signature $s^*(t)$, the signal signature seen by sensor k (the k-th event signature projection), $1 \leq k \leq M$, would be (excluding any noise components):

$$s_k(t) = a_k(t)s^*(t - \tau_k)u(t - \tau_k), \quad (18)$$

where $u(t)$ is the unit step function. The *time shift* τ_k is due to the propagation delay to sensor k and equals $\tau_k = d_k/v$. This pre-processing relates to the physical topology (geometry) of the sensor network and has been used in this numerical results. Note that the expression in (18) represents an example of a pre-processing operation that projects the original event signature from the location of the event occurrence to the location of the sensor. For our comparisons, we assume the attenuation function to be

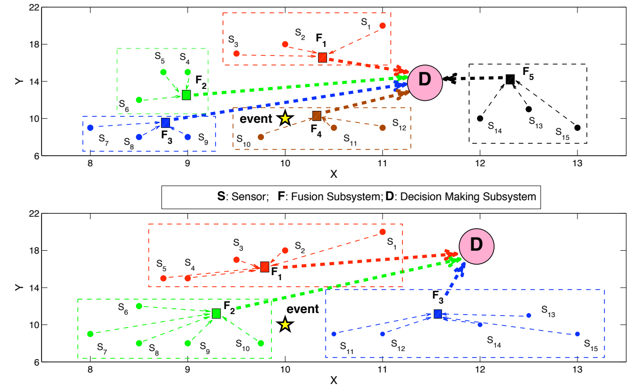

Figure 4. Topology of the network and two different detection architecture

$a_k(t) = a_k = 1/(1 + d_k^2)$. As discussed in [17], attenuation function of acoustic signals may change at different distances from the source of the signal and the best attenuation models can be obtained via experimental data on specific sensor deployments.

We start with the multi-sensor system $(M = 15)$ in figure 4. We have assumed that event signature is the transient $s^*(t) = 1 - t^2$ (with a lifespan of one time unit) and that each sensor contributes $N = 20$ samples. We consider varying number L of fusion subsystems where $L = [1, 3, 5, 15]$; the two end cases represent the centralized and distributed cases respectively, the rest are examples of hybrid topologies.

The upper part of the figure (4) shows the case that $L = 5$, with each clusters having 3 sensors contributing samples to the (tier-0) fusion subsystem. The fusion subsystems, then, send their local decisions to a single decision subsystem for the system-wide decision making. The lower part of the figure (4) shows the case where we have 3 clusters of sensors. The clustering in both cases are based on the relative distance between the neighboring sensors and the event location.

When $L \neq 1$, the decision subsystems uses a Q-count policy to make its system-wide decision based from the lower tier-decisions. Figure 5 summarizes this results in this case. Assuming Additive White Gaussian (AWG) noise with the same variance level σ for each sensor, we study the QoI performance for various σ and a priori probabilities P_0 and P_1; we use $\eta = P_0/P_1$ assuming $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 0$ in (4). The cases $\eta = 1$, i.e., $P_0 = P_1$ which will serve as our reference point, $\eta = 1.3$ (> 1) and $\eta = 0.7$ (< 1) as well. We capture QoI through a single metric the *probability of error* P_e which aggregates the QoI attributes: $P_e = P_1(\eta P_f + 1 - P_d)$;(for notational brevity, we have dropped the Q and M from the notation. As expected, centralized architecture achieves the best (i.e. smallest) P_e of all architectures as it makes best use of the *all* available information.

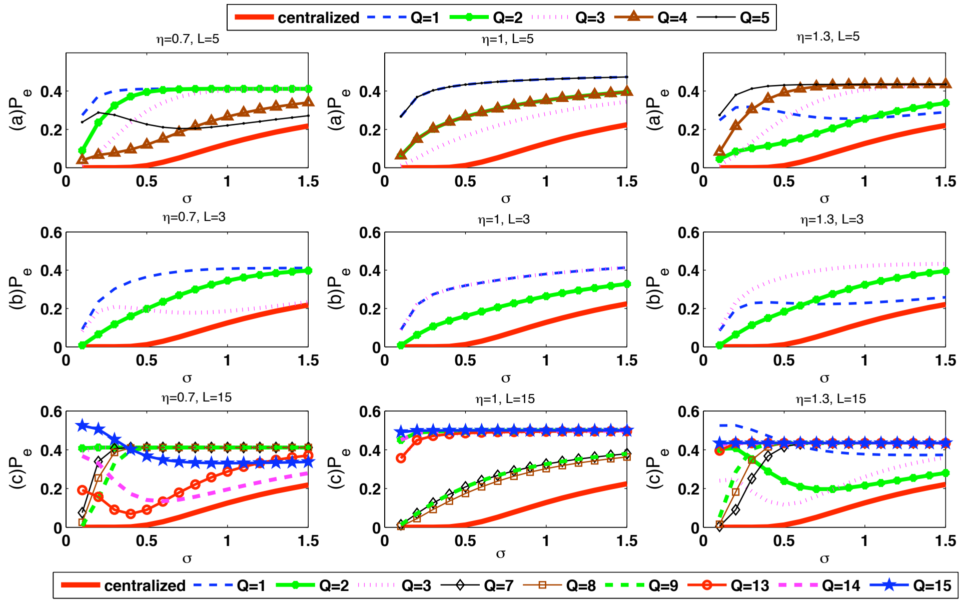

Figure 5. Hybrid approach((a)M=15,L=5 and $\eta = [.7, 1, 1.3]$,(b)M=15,L=3 and $\eta = [.7, 1, 1.3]$,(c)M=15,L=15 and $\eta = [.7, 1, 1.3]$

For $\eta = 1$ when comparing the QoI performance of the different detection policies for a specific clustering and detection architecture,the Q_{opt}s for the cases of $L = 5$, $L = 3$ and $L = 15$ are $Q = 3$, $Q = 2$ and $Q = 8$ which is compatible with the discussion in section 4.3. Note that P_f and P_d, not shown, decrease with increasing Q as the detection policy decides in favor of H_0 more liberally; this trend holds true for all cases of η studied.

The case where $\eta = 1.3$ is shown in figure 5 on third column. First of all, with $\eta > 1$ (i.e., $P_0 > P_1$), the condition for deciding in favor of H_1 becomes harder to achieve when compared with the $\eta = 1$ case, see (3). Therefore, P_d and P_f decrease in magnitude when compared with these probabilities when $\eta = 1$ (for the same σ) for both architectures. We again see the centralized one achieving the best performance. With respect to the distributed architecture. It is also notable that the best detection policy for all three cases is attained at a Q value that is smaller than the best Q of the similar case with $\eta = 1$. The reduction of the optimal Q threshold in the case where $\eta > 1$ when compared with that for $\eta = 1$ follows from the fact that with decreasing P_1 it becomes harder for any sensor to declare in favor of H_1. Therefore, having even a small number of sensors declaring in favor of H_1 is reason enough to decide in favor of H_1 system-wide.

Finally, the case where $\eta = 0.7$ is shown in figure 5. Arguing as before, the behavior of the QoI metrics in this case relative to those of the $\eta = 1$ case is in reverse order to the behavior experienced when $\eta = 1.3$. With respect to P_e, while the centralized architecture still performs the best.

Regarding the best Q at least when $\eta = 1$, it follows from (5) that $P_k^d \geq P_k^f$ for any k and, furthermore, when $\eta = 1$, $P_k^d = 1 - P_k^f$; see [18] for an extensive discussion regarding these probabilities in relation to η. The error probability in this case is $P_e = P_1(P_f + 1 - P_d)$. Based on the relationship between P_k^d and P_k^f for the various sensors and (7), it follows that P_e exhibits a symmetrical behavior with Q centered at $M/2$ and, hence, P_e possesses an extremum at (or the floor or ceiling of) this value. By calculating the sign of the difference $\Delta P_e = P_e(Q;M) - P_e(Q+1,M)$ close to $M/2$ it can be shown that this extremum is actually a minimum for P_e. For other ηs, we conjecture that in general the best detection policy for the distributed architectures is achieved at a threshold Q that decreases away from $M/2$ (and possibly toward 1) as η increases above 1, and increases away from $M/2$ (and possibly toward M) as η decreases below 1.

Next we study an evaluation step that may occur in the post processing layer of our framework. It relates to location sensitivity analysis that may performed after the QoI results have been obtained via the core QoI analysis.

5. Position sensitivity analysis

It should be clear that the QoI performance will depend on the distance between the sensor and event, therefore it would be desirable to study the sensitivity of the performance on location errors, e.g., what if sensors were not placed exactly where they were supposed to, or have an erroneous understanding of their exact location.

If the received signal follows the model in (18), then

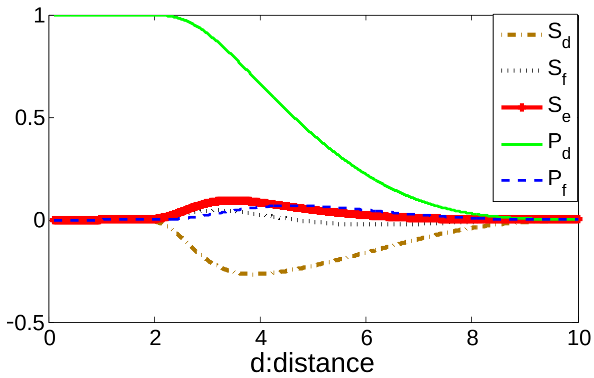

Figure 6. Sensitivity of QoI metrics to distance measurement, for single sensor

sensitivity in the probability of detection S_d for a single sensor system can be derived as follows:

$$S_d = \frac{\partial P_d}{\partial d} = -\frac{1}{\sqrt{2\pi\sigma^2 E}} \exp\left(\frac{-(\frac{ln\eta}{\psi} - \frac{\psi}{2})^2}{2}\right) * \left(\frac{-\ln\eta}{\psi^2} - \frac{1}{2}\right)\left(\frac{E}{a}\frac{\partial a}{\partial d} - E'\frac{\partial \tau}{\partial d}\right), \tag{19}$$

where $E = a^2(d)\sum_{i=1}^{i=N} S^2(t_i - \tau(d))$ and $E' = \partial E/\partial t$. Similarly, the sensitivity of the false alarm is:

$$S_d = \frac{\partial P_d}{\partial d} = -\frac{1}{\sqrt{2\pi\sigma^2 E}} \exp\left(\frac{-(\frac{ln\eta}{\psi} + \frac{\psi}{2})^2}{2}\right) * \left(\frac{-\ln\eta}{\psi^2} + \frac{1}{2}\right)\left(\frac{E}{a}\frac{\partial a}{\partial d} - E'\frac{\partial \tau}{\partial d}\right), \tag{20}$$

Finally $S_e = P_0\partial P_f/\partial f + P_1\partial P_d/\partial d$. It follows from (19) and (20) that when $\eta = 1$ then $S_e = S_f$.

Equations (19), (20) show how sensitivity of the QoI metrics relates to the selected propagation model and also the received signal power (E) at sensors. Figure 6, shows the sensitivity of the QoI metrics of a single sensor system where source generates the signal of $s^*(t) = 10(1-t^2)$ (with a lifespan of one time unit) and the signal attenuates with $a = 1/(1+d^2)$, and the noise level is $\sigma^2 = 1$; sensor takes $N = 20$ samples, and $\eta = 3$. As figure 6 shows, the QoI metrics face different level of sensitivity by changing the relative distance. As expected, as the sensor is located further from the event the location sensitivity decreases.

6. Faulty sensors and measurements

In this section, we introduce a new modeling contribution to our analysis framework. Specifically, in addition to the additive environmental noise, the n_i samples in (1), we consider sensor systems with faulty, biased, uncalibrated sensors. Enriching our framework with sensor fault models and analysis techniques will allow us (or better, the framework users like the system designer) to accommodate the calibration errors, while these errors are one of the major obstacles to practical use of the sensor networks [19].

Starting with a single sensor system, we first consider a linearly biased measurement model [20], where the sensor response exhibits an unknown gain a and offset b, so that instead of sampling s_i as in (1), the sensor reports $as_i + b$, plus noise, instead. Incorporating this fault model as a preprocessing step in our framework, the hypothesis testing formulation in (1) becomes:

$$\text{hypothesis } H_1 : r_i = as_i + b + n_i, \tag{21}$$

$$\text{hypothesis } H_0 : r_i = b + n_i. \tag{22}$$

If we were to consider a multi-sensor system the gains and offsets for the different sensors may vary, but we don't pursue this case further in this paper. Since the gain and offset parameters are unknown, the hypothesis test of (21) will be a composite one, where we not only decide in favor of one of the hypotheses but estimate the unknown parameters a and b as well.

To model the problem under each hypothesis, we can use linear model [4] as following:

$$\mathbf{r}_{|H_i} = \mathbf{H}_i\theta_i + \mathbf{n}, \text{ where } i \in \{0,1\} \tag{23}$$

Where $\mathbf{r} = [r_1, ..., r_N]^T$ is column vector of all observations and the noise vector $\mathbf{n} = [n_1, ..., n_N]^T$ has a $\mathcal{N}(\mathbf{0}, \mathbf{C})$ distribution and is independent of θ_i. $\mathbf{H}_1$ comprises of two columns, the first containing the samples s_i and the second being the unity vector and $\mathbf{H}_0$ comprises of two columns, the first being the zero vector and the second being the unity vector. The unknown parameter vectors are $\theta_1 = [a,b]^T$,and $\theta_0 = [0,b]^T$. If we assume parameters are deterministic, then the linear model becomes classical linear model, however no Uniformly Most Powerful (UMP) test for classical linear model exists[4]. Moreover, in our specific application, specially for a situation of the transient events, we have to be able to make the decision based on the limited (relatively few) number of available samples in timely manner. Since there is no UMP test available for classical linear model, the precision of its estimate with relatively small number of samples in not guaranteed. However in real experiments and based on available training data, we may be able to assume a gaussian distribution for parameter vectors $\theta_i \sim \mathcal{N}(\mu_{\theta_i}, \mathbf{C}_{\theta_i})$. With this prior distribution assumption for parameters, the model (23) will be considered as bayesian linear model, then the posterior distribution $p(\theta_i|\mathbf{r})$ is also Gaussian with mean and covariance matrix given by (see [4, Th. 10.3]):

$$E(\theta_i|\mathbf{r}) = \mu_{\theta_i} + \mathbf{C}_{\theta_i}\mathbf{H}_i^T(\mathbf{H_i}C_{\theta_i}\mathbf{H}_i^T + \mathbf{C})^{-1}(\mathbf{r} - \mathbf{H}_i\mu_{\theta_i})\text{and,} \tag{24}$$

$$C(\theta_i|\mathbf{r}) = \mathbf{C}_{\theta_i} - \mathbf{C}_{\theta_i}\mathbf{H}_i^T(\mathbf{H_i}C_{\theta_i}\mathbf{H}_i^T + \mathbf{C})^{-1}(\mathbf{H}_i\mathbf{C}_{\theta_i}), \tag{25}$$

respectively.

Then the Minimum Mean Square Error (MMSE) Estimate of the θ_i will be $\hat{\theta}_i = E(\theta_i|R)$ where $i \in \{0,1\}$. Having these estimates for the unknown parameters, we can substitute them in the (21) and reuse the likelihood ratio test with estimated parameters. Then the decision test can be reduces to the following:

$$\hat{l} \triangleq \mathbf{r}^T \mathbf{C}^{-1}(\mathbf{s_1}(\hat{\theta}_\mathbf{1}) - \mathbf{s_0}(\hat{\theta}_\mathbf{0})) \underset{H_0}{\overset{H_1}{\gtrless}} \eta^* \triangleq \ln(\eta) + \frac{1}{2}(\mathbf{s_1}(\hat{\theta}_\mathbf{1})^T \mathbf{C}^{-1}\mathbf{s_1}(\hat{\theta}_\mathbf{1}) - \mathbf{s_0}(\hat{\theta}_\mathbf{0})^T \mathbf{C}^{-1}\mathbf{s_0}(\hat{\theta}_\mathbf{0})). \quad (26)$$

Where $s_i(\hat{\theta}_i) = H_i\hat{\theta}_\mathbf{i}$. To compute the QoI metrics, assuming $\Delta = s_1(\hat{\theta}_1) - s_0(\hat{\theta}_0)$, we have $\mu_{\hat{l}|H_i} = s_i(\hat{\theta}_i)^T C^{-1}\Delta$, and $\sigma^2_{\hat{l}|H_i} = \Delta^T C^{-1}\Delta$, where $i \in \{0,1\}$. Then QoI metrics can be computed with the following formulation:

$$P_z = \Pr(\hat{l} \geq \eta^\star|H_i) = 1 - \Phi\left(\frac{\eta^* - \mu_{\hat{l}|H_i}}{\sigma_{\hat{l}|H_i}}\right), \quad (27)$$

where $i = 1$ when $z = d$, and 0 when $z = f$. If we are not aware of calibration error, then under both hypotheses $\breve{\theta} = [1;0]$ will be considered; which results in $s_1(\breve{\theta}) = s$, and $s_0(\breve{\theta}) = 0$, but it is notable that observations come from the sensor samples with gain a and offset b, so $\mu_{\hat{l}|H_1} = (as + b\vec{\mathbf{1}})^T C^{-1}s$ and $\mu_{\hat{l}|H_0} = b\vec{\mathbf{1}}^T C^{-1}s$ and $\sigma^2_{\hat{l}|H_1} = \sigma^2_{\hat{l}|H_0} = \sigma^2 = s^T C^{-1}s$, then using the similar approach as (27), we can compute the probability of detection and false alarm for this case.

Figure 7 compares the probability of error (P_e) of three specific cases as noise level (σ) changes; in the first case, the complete knowledge of the sensors gain and offset is provided, in this case equation (27) with θ equal to the exact values of a and b. The second case is the situation that calibrated error model is considered, but the deterministic knowledge of the parameters is not available, and the estimated parameters is required, and the third case is the situation that we have not considered any measurement error for the sensor-collected data, and we use the gain of one and offset of zero in sensor measurements. In this study case it is assumed that, we have a single-sensor system which contributes $N = 20$ samples in detection process, and is located in distance of $d = 3$, from a source with signal signature of $s^*(t) = 1 - t^2$. It is also assumed that in reality sensor has a gain of $a = 1.4$ and offset of $b = 0.4$, and we have prior knowledge of $\theta \sim \mathcal{N}([1, 0.2], I_2)$ for parametres. As figure 7 shows, if we don't have any knowledge about sensor's gain and offset, we will face large probability of error in decision making, for this particular case since actual gain is larger than one and we have positive offset, for smaller noise levels(σ) we have large probability of false alarm which results in large probability of error as well. As it is expected and shown in figure 7 for a case of perfect knowledge of the calibration factors, probability

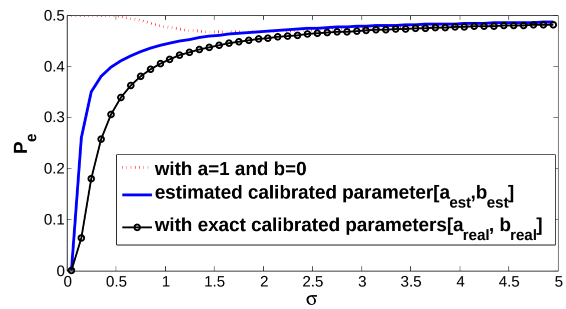

Figure 7. P_e analysis of the un-calibrated sensors

of error will become less, and this result also shows that using the estimated version of the calibration factors will help us to get close to the exact model. It is notable that when the noise level increases, it affects the estimator performance; results in figure 7 also confirms this situation. It is also notable that using larger number of samples from observations, can improve the estimator performance. Finally this discussion and the provided approach and results, show that to achieve a better QoI, it is required to consider faults and measurement errors in the system.

7. Discussion and Concluding Remarks

Research output on sensor networks has been increasingly covered many different aspects. Operation, configuration, management, information retrieval and dissemination, hardware and software applications, middleware services, programming paradigms, are just few of the areas that considerable research capital has been invested and output has been generated the past few years. Within this sea of groundbreaking research, the research area that has attracted our interest did not, to the best of our knowledge, exist, at least not in any form of distinctively identifiable manner. We have decided to study sensor networks from the vantage point of applications, that have interest to use them not only to receive data from them, but also have knowledge of their quality.

For our setting, quality of information (QoI) represents the level of confidence that sensor-data-dependent applications may place on derived information from the sensor network(s) that support the application. Knowledge of the QoI allows mission/application planners to enhance their system with contingency plans in anticipation of "errors" that the network of sensors and their fusion modules accurately capturing and interpreting the real world.

Our research anchors around the mission or application planner who would like to deploy a sensor-enabled application, and is interested to know the level (or at least the ranges) of QoI which will be receiving; therefore the application can be trained or designed to accommodate uncertainties in the received data flows. Within this context, we anticipate that an evaluation toolkit will be very effective aid in calculating QoI. The toolkit will comprise a collection of analysis methodologies that a system designer (supporting the application planner) can choose for the evaluation. The design philosophy behind such a toolkit is what we are currently researching.

Considering the rather broad application space of detection of signal-producing events, we have proposed and studied a layered analysis framework that allows the composition of performance analysis solutions from a number of constituent analysis techniques. The layered approach separates the solution approach into tree steps: a preprocessing step that focuses on what affects signals from their source to their destination; a core analysis of QoI; and, finally, a post processing layer that operates on the results of the QoI analysis to build the final desired solution. While we have presented mostly the QoI analysis layer considering various decision topologies, we presented examples of both preprocessing (like signal attenuation and sensor deployment) and postprocessing (like the position sensitivity analysis).

We have based our QoI analysis on hypothesis testing, and derived QoI performance for a number of topologies. We made use of equivalent sensors to reduce the analysis of hybrid system to those of fully distributed systems, and derive both sensor level and system-level QoI performance metrics. In the latter case, we study both finite and "infinite" sensor systems and derived a continuum of computationally attractive solutions spanning both small and large systems. We derived dominant relationships, that allow to calculate performance metrics by constructing and analyzing simpler systems.

As we move forward with research we will be pursuing enriching our framework with additional capabilities studying big problems at small atomic steps. While in [6] we discusse how we may accommodate the lack of knowledge of when a signal have started , by bounding the performance of our system, we recognize that additional work is needed in this area. Along with it we see the need to study sampling policies and how they may influence the decision. Our focus now is purely on the information processing, however if we were to consider the time it takes to propagate sensor samples and local decision to higher-tier fusion and decision subsystems, the impact of communication will need to be studied.

References

[1] S. Ehikioya, "A characterization of information quality using fuzzy logic," in *18th Int'l Conf. of the North American Fuzzy Information Processing Society,(NAFIPS 1999)*, June 10-12, 1999.

[2] T. Dasu and T. Johnson, *Exploratory Data Miningand Data Cleaning*. Wiley Interscience, 2003.

[3] "The MIT total data quality management (TDQM) program," http://web.mit.edu/tdqm/www/index.shtml.

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*. Prentice Hall, 1998, vol. 1.

[5] H. L. van Trees, *Detection, Estimation, and Modulation Theory, part I*. John Wiley & Sons, 1968.

[6] S. Zahedi and C. Bisdikian, "A framework for qoi-inspired analysis for sensor network deployment planning," in *2nd Int'l Workshop on Performance Control in Wireless Sensor Networks (PWSN)*, Austin, TX, USA, Oct. 23, 2007.

[7] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors:part i-fundamentals," in *Proceedings of the IEEE*, vol. 85, no. 1, Jan. 1997, pp. Page(s):54 – 63.

[8] R. Blum, S. Kassamand, and H. Poor, "Distributed detection with multiple sensors:part ii-advanced topics," in *Proceedings of the IEEE*, vol. 85, no. 1, Jan. 1997, pp. Page(s):64–79.

[9] R. Niu and P. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," in *EURASIP Journal on Wireless Communications and Networking*, vol. 5, no. 4, 2005, pp. Page(s):462 –472.

[10] S. K. Jayaweera and K. Al-Tarazi, "Large system decision fusion performance in inhomogeneous sensor networks," in *Proc. of the IEEE Int'l Conf. on Video and Signal Based Surveillance*, 2006.

[11] N. Katenka, E. Levina, and G. Michailidis, "Local vote decision fusion for target detection in wireless sensor networks," in *Joint Research Conference on Statistics in Quality Industry and Technology*, Knoxville, TN, USA, June 7-9, 2006.

[12] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 98–101, 1986.

[13] L.Yu, L.Yuang, G. Qu, and A. Ephremides, "Energy-deriven detection scheme with with gauranteed accuracy," in *in Proc. IPSN'06*, Nashville,TN, USA, April 9-21, 2006.

[14] A. E. L. Yu, *Detection, Energy, and Robustness in Wireless Sensor Networks*. John Wiley & Sons, 2006.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*. Prentice Hall, 1998, vol. 2.

[16] X.Guyon, *Random Fields on a network: Modeling, Statistictics, and Applications*. Springer-Verlag, 1995.

[17] T. Pham, "Distributed source localization and tracking algorithms for ad-hoc acourstic sensor networks," Ph.D. dissertation, University of Maryland, College Park, 2006.

[18] C. Bisdikian, "Quality of information trade-offs in the detection of transient phenomena," in *SPIE Unattended Ground, Sea, and Air Sensor Technologies And Applications IX Conference*, Orlando, FL, USA, April 9-13, 2007.

[19] P. Buonadonna, D. Gay, J. Hellerstein, W.Hong, and S. Madden, "Task:sensor network in a box," *Proc. Second European Workshop on Wireless Sensor Networks*, 2005.

[20] L. Balzano and R. Nowak, "Blind calibration of sensor networks," *Proc. Information Processing in Sensor Networks (IPSN)*, 2007.