Distributed Detection in Sensor Networks with Limited Range Sensors [*]

Erhan Baki Ermis, Venkatesh Saligrama
Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215
Email: {ermis, srv}@bu.edu

September 10, 2013

Abstract

We consider a multi-target detection problem over a sensor network (SNET) with limited range sensors and communication constraints. This problem complements the widely considered decentralized detection problem where all sensors observe the same target. While the necessity for global collaboration is clear in the decentralized detection problem, the benefits of collaboration with limited range sensors is unclear and has not been widely explored. In this paper we develop a distributed detection approach based on recent development of the false discovery rate (FDR). We show that collaboration among sensors can not only be beneficial, but also necessary, even in the limit of non-overlapping sensing regions for each sensor. Our scheme characterizes the optimal achievable performance, i.e., maximizes the detection rate subject to false discovery rate and communication constraints. We further demonstrate scalability to large SNETs by showing that the upper bound on the communication complexity scales linearly with the number of actual targets and is independent of the number of sensors. Finally, we deal with situations where the sensing model may be uncertain and establish robustness of our techniques to such uncertainties.

1   Introduction

The design and deployment of sensor networks (SNET) for decision making pose fundamental challenges due to energy constraints and uncertain environments. It is well known that energy limits the ability to collaborate by limiting the communication range, capacity, and system lifetime. On the other hand, to overcome uncertainty, which arises from noisy measurement of distributed phenomena, and to meet the global constraints of the network, some limited form of collaboration among sensor nodes is necessary.

1

The general question of dealing with distributed data in the context of detection has been an active topic of research (see [2, 3, 6–8, 12, 15, 17, 19, 20, 25–28] and references therein).

In this paper we focus on one such problem where we seek to minimize the miss rate subject to false alarm and communication cost constraints. Preliminary work along these lines have been presented in some of our earlier papers [10, 11, 29]. Specifically, we develop solutions for detection of distributed events, sources, or abnormalities that are localized. By localization we mean that the event has inherent locality and multiplicity. For instance multiple targets at several locations coupled with the fact that each sensor has only a limited detection/sensing range and can detect targets only within this range. In addition, we assume communication constraints limiting the number of bits that can be transmitted within the network.

This problem complements conventional decentralized detection and estimation problems, where noisy information about a single event is measured by the entire network. The global phenomenon can be one of several different discrete possibilities or a continuum in estimation problems. Researchers have investigated several architectures ranging from fusion centric to ad-hoc consensus based approach [2, 3, 8, 15, 17, 19, 20, 25–28]. Our problem, namely, localized detection, arises naturally whenever each sensor has a limited sensing range and there are a multiplicity of targets as illustrated in Figure 1. Limited sensing range arises as a consequence of natural attenuation of signal energy with distance. Multiplicity issue can arise either as a consequence of multiple targets or as a consequence of spatial variability of the underlying field resulting in negligible correlation between two well-separated spatial locations. Our focus in this paper is on the first aspect, namely, multiplicity of targets and issues associated with the second aspect is subject of future research.

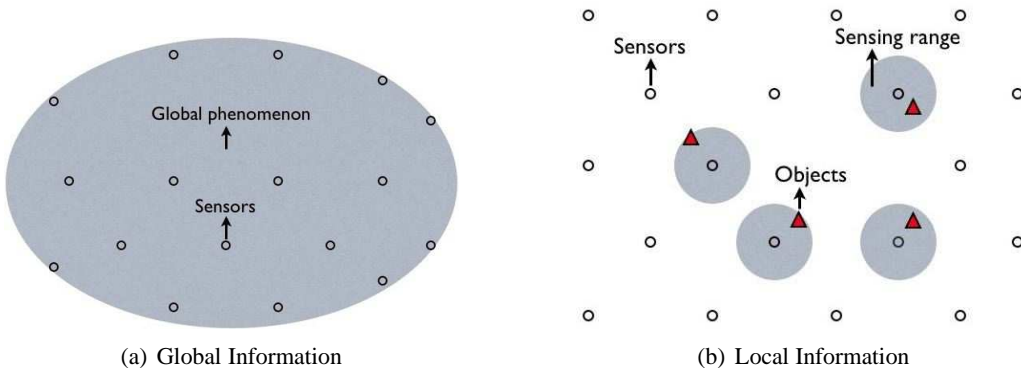(a) Global Information   (b) Local Information

Figure 1: Decentralized Detection vs. Localized Detection: In decentralized detection the sensors observe a single global phenomenon, whereas in localized detection the sensors observe multiple local phenomena.

In contrast to the conventional decentralized detection problem where sensor data is correlated across all the sensor nodes, the issue here is that the sensing regions of each sensor partially intersects with a few sensors' sensing regions. In this light one could contemplate that local decisions are "optimal" at least in the limiting case when sensing regions have no overlap. This strategy by its nature does not require any collaboration. Nevertheless, it turns out that controlling false alarm probability in this manner invariably

results in poor detection performance. This problem is related to the so called multiple comparison testing problems in statistical literature [16]. In this setting a data set is given with each data point coming from one of two exponential distributions and the objective is to associate the data point with the correct distribution.

It turns out that at a fundamental level our challenge does not primarily arise from decentralized processing but rather in using false alarm probability as a performance metric [1, 4, 5]. We further quantify it for general distributions in Section 3.1 using information theoretic techniques and show that indeed global false alarm and miss probability together are strictly bounded away from zero.

The multiple comparison problem has led to various formulations and a number of associated approaches [4, 21]. A promising direction introduced recently is the relaxation of false alarm probability to the false discovery rate (FDR) [4, 14, 24]. In contrast to false alarm probability, FDR controls the expected ratio of false positives to the declared positives. It turns out that this relaxation leads to significant improvement in detection power even when the probability distributions under significant(non-null) hypothesis is unknown. This concept has been extended to the Bayesian setting [14] when all of the distributions are known. The FDR concept has also been applied to a variety of problems ranging from estimation theory [1] to practical applications [13].

Motivated by these results we adopt FDR as a performance measure for our SNET problem. We characterize the fundamental trade offs between global performance (false alarms and miss rate) and communication cost in a non-bayesian setting. We develop a framework to minimize the miss rate subject to communication cost and worst-case misclassification constraints.

Our technique involves an optimal local domain transformation at each sensor node. The main insight in developing this technique is that *false discovery rate constraint is satisfied for all transformations that are measure invariant with respect to distribution of observations under* H_0. The approach we develop is easily decentralized and is robust to approximate knowledge of observation statistics. This is particularly useful since the sensing range is typically unbounded and has a power-law decay behavior. Moreover, this decentralized algorithm has very elegant scalability properties. A very interesting implication of this work is that the detection performance of a wireless SNET is comparable to that of a wired network of sensors in the following sense: *corresponding to an achievable (centralized) false positive threshold the communication complexity grows in proportion to the actual number of events, sources or abnormalities while achieving the same centralized performance*.

The main contributions of our paper are fourfold:

(a) Development of distributed detection theory in the context of limited sensing scenarios.

(b) Extensions of optimal false discovery rate theory to situations when the distributions under null and significant hypothesis are partially known.

(c) Extensions of optimal false discovery rate theory to situations when the distribution under null hypothesis is partially known and the mixture statistics can be approximately learnt.

(d) A general definition of p values that allows control of FDR when the distribution of observations under H_0 is a member of a family of distributions.

3

The last two aspects are particularly useful when the target observation models are unknown but a mixture model of null & target statistics can be partially learnt from data.

The organization of the paper is as follows: in Section 2 we present an overview of the problem setup. In Section 2.1 we describe ideal and non-ideal sensing models and define the space of decision rules. In ideal sensing the targets are detectable if the sensors are within a fixed range of the target and no signal is received by a sensor outside this range. Non-ideal sensing relaxes this assumption to account for unbounded range, which arises naturally from power law decay of signals. In Section 3 we present the general formulation of the problem, discuss probability of false alarm formulation and why it is not suitable in this setup. Finally we describe the false discovery rate, discuss its scalability property, and present a formulation of the problem from the FDR perspective. In Section 4 we explain the suboptimal nature of the FDR approach and develop the domain transformed FDR (DTFDR) scheme. Here we compare FDR and DTFDR, describe a robust distributed DTFDR algorithm, and present optimality results. In Section 5 we present the robustness analysis of the DTFDR algorithm, and explain how non-ideal sensing model can be addressed in the DTFDR framework through a robustness approach. In Section 6 we present simulations and discuss some interesting results. Finally, in Section 7 we present our concluding remarks.

2   Setup

We consider a non-Bayesian setting where an unknown number of targets are distributed on a sensor field of m sensors. We assume no prior information on the number of targets, and their potential locations. Targets are observed by a SNET in which the sensor nodes are distributed uniformly. We wish to identify, via distributed strategies, the set of sensors that have an object in their sensing range.

We can consider two equivalent models for our problem.

Target Centric Model: In this scheme, the targets generate a signal field over the sensor network and the sensors sample the field at their locations. For example, this scheme is applicable when several seismic sensors are deployed over a field. In this scheme the significant hypothesis (H_1) for a sensor is the event that the sensor is within the effective region of a target, and the null hypothesis(H_0) is the event that the sensor is outside the effective region of all targets.

Sensor Centric Model: Alternatively, we could consider a sensor centric scheme where each sensor has the ability to detect targets in a given range. The vertices denote targets and sensors and the edges denotes sensors in the range of the target. There are two types of edges that are possible corresponding to ideal and non-ideal sensing models. In the ideal sensing model the sensing range is bounded and the edges of the graph depict sensors in the "finite range" of the target. In the non-ideal sensing model we can also have "weak links" that depicts weak dependence of sensor observations on distant targets. This model is realistic and accounts for power-law decay of signals with distance. We deal with non-ideal sensing situation by adopting a robustness perspective, i.e., both the null and significant hypothesis are viewed as composite hypothesis. For instance, the observation statistics for sensors far away from the targets is modeled as belonging to a family of distributions.

An important point is that the graph is unknown a priori since the locations of targets are unknown. Also

note that our main objective here is only to identify sensors that are within the target range and does not deal with the important problem of associating sensor measurements that correspond to the same target.

The observation vector is denoted by $Y = (Y_s : s \in \mathcal{S})$, where throughout the paper $\mathcal{S}$ represents the set of sensors that form the SNET, and Y_s represents the collection of measurements taken by sensor $s \in \mathcal{S}$. The realization of observation vector Y is denoted by $y = (y_s : s \in \mathcal{S})$. For definiteness we focus on the case when Y has a continuous distribution. The conditional probability distribution function of the observation vector Y_s at sensor s under each hypothesis $H^s = H_i$, $i = 0, 1$ is denoted by $G_{is}(\cdot)$, where H^s denotes the hypothesis at sensor s. We assume a general structure on the problem in the sense that G_{1s} belongs to a class of distributions $\mathcal{G}_1$, and G_{0s} belongs to a class of distributions $\mathcal{G}_0$. Let $\mathcal{S}_0 = \{s \in \mathcal{S} : H^s = H_0\}$ with cardinality m_0 and $\mathcal{S}_1 = \{s \in \mathcal{S} : H^s = H_1\}$ with cardinality m_1.

2.1 Mathematical Modeling

We describe mathematical models for ideal and non-ideal sensing. The ideal sensing model accounts for situation where targets can be sensed only if a sensor is within a fixed range of target.

Ideal Sensing Model: In general the joint density of observations factorizes to conditionals as in equation 1 below.

$$g(y_1, y_2, \ldots, y_m \mid H_{i1}, H_{i2}, \ldots H_{im}) = \prod_{s \in \mathcal{S}_0} g(y_s \mid H^s) \prod_{s \in \mathcal{S}_1} g(y_s \mid \{H^s\}_{s \in \mathcal{S}_1}) \tag{1}$$

In addition note that if there is at most a single sensor that can be close to a target, the above equation can be further factorized as follows:

$$g(y_1, y_2, \ldots, y_m \mid H_{i1}, H_{i2}, \ldots H_{im}) = \prod_{s \in \mathcal{S}_0} g(y_s \mid H^s) \prod_{s \in \mathcal{S}_1} g(y_s \mid H^s) \tag{2}$$

In general we will deal with the first model for the rest of the paper while dealing with ideal sensing.

Non-Ideal Sensing Model: Here we assume that the observed signal from each target decays as a function of distance between the target and the observing sensor. Therefore the target signal is no longer constant within the region around the target, and is no longer zero outside this region. As a consequence, sensors outside the effective region also observe a signal from each target. In this model the above factorization does not hold. However, we assume that the joint density lies within a ball of the factorizations with respect to some metric. We will elaborate further on the choice of this metric in the upcoming sections. Therefore we assume that the following holds for the non-ideal sensing model:

$$g(y_1, y_2, \ldots, y_m \mid H_{i1}, H_{i2}, \ldots H_{im}) \in \mathcal{B}_\Delta \{ \prod_{s \in \mathcal{S}_0} g(y_s \mid H^s) \prod_{s \in \mathcal{S}_1} g(y_s \mid \{H^s\}_{s \in \mathcal{S}_1}) \} \tag{3}$$

$$\tag{4}$$

5

where, $B_\Delta(g)$ is a family of distributions of radius Δ (in an appropriate topology/metric) centered around $g(\cdot)$.

To deal with the non-ideal sensing model, we relax the assumption that $\mathcal{G}_0$ and $\mathcal{G}_1$ are singletons. We deal with this case from a robustness perspective, i.e., as a perturbation of the ideal sensing model, in the last part of the paper.

We next formalize distributed decision strategies to complete the problem setup.

Distributed Decision Rules: A distributed decision rule is a set of individual decision rules where each sensor makes a decision in a finite alphabet set $\mathcal{A}$ based on its cumulative local observation and available information from other sensors, i.e.

$$u_s(t) : \{Y_s \times I_s(t)\} \mapsto \mathcal{A}$$

where $u_s(t)$ is the decision of sensor s, $I_s(t)$ is the causally available information to sensor s from other sensors until time t. Note that $I_s(\cdot)$ can be constant as a function of time, which implies no information is shared between sensors. The collection of decisions of all sensors forms the decision rule at time t for the SNET problem:

$$u(t) = \{u_s(t)\}_{s \in \mathcal{S}}$$

Each sensor may then choose to communicate this local decision (or any other information) to other sensors in the SNET. Later we discuss schemes for efficient communication of these decisions. Ultimately, the SNET reaches a decision when all the sensors have decided between null and significant hypothesis. The objective is to devise a distributed decision rule such that it has large detection power, leads to a small measure of false alarms, and has a minimal communication cost to the SNET.

We focus on a subclass of strategies, which ultimately proves to be scale-wise optimal. In this setup, at any intermediate time each sensor is either undecided or concluded that there is a target in its vicinity. Those sensors that conclude that there is a target communicate their one-bit decisions to the SNET. These decisions in combination with local data is utilized to make decisions by the undecided sensors. To be concrete, each sensor makes a local decision, i.e., $u_s(\cdot) \in \{\phi, H_0, H_1\}$. The reason u_s is one bit, although it seems counterintuitive, is that ϕ stands for undecided, and thus the actual decision is one bit. Sensors with $u_s(t) = H_1$ communicate their decisions to the entire network. It turns out that what is required is the number of sensors at time t that decide H_1 and not really the identity of these sensors. This fact is advantageous from the point of view that there are a large collection of algorithms [9, 18] that can perform such counting operations efficiently.

These set of strategies for each sensor can be viewed as feedback strategy involving the SNET, where information from sensors that have so far made their decisions is fed back to the SNET. This is shown schematically in Figure 2.

We point out that in contrast to the above strategy, a *non-adaptive strategy* is one where each sensor makes its decision by ignoring the available information from the SNET, i.e. $u_s(t) : \{Y_s\} \mapsto \mathcal{A}$. Although this approach is very simple to implement, it is usually blind to the global constraints of the network.
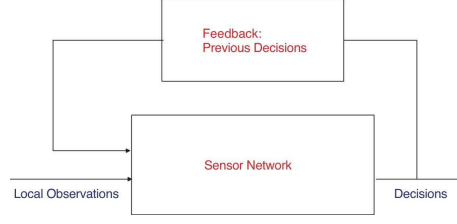
Figure 2: Feedback Structure for Adaptive Decision Making.

3 Problem Formulation

The most general form of our problem is as follows: maximize worst-case detection power subject to communication and worst case false alarm constraints.

In order to further specify the problem we first define the local and global false alarm and miss events. Broadly, local false alarm event is the event that a sensor with no target in its sensing radius deciding otherwise, i.e. $LFA_s = \{u_s = H_1 \ \& \ H^s = H_0\}$ and the local miss event is the event that a sensor with a target in its sensing radius deciding otherwise, i.e. $LM_s = \{u_s = H_0 \ \& \ H^s = H_1\}$. Based on these definitions, the global false alarm event is the event that there is at least one local false alarm event, i.e., $GFA = \{\cup_{s \in \mathcal{S}} LFA_s \neq \phi\}$ and the global miss event is the event that there is at least one local miss event, i.e. $GM = \{\cup_{s \in \mathcal{S}} LM_s \neq \phi\}$.

In view of these events, consider the table below, where m is the number of samples (or sensor nodes) known in advance. Here the observable random variable R is the total number of sensors that decide presence of a target within their sensing region, and the unobservable random variable V is the total number of sensors that commit a local false alarm. The global false alarm event is the event $\{V \geq 1\}$.

3.1 Issues with Error Probability Formulation:

In detection problems, a very common quantity of interest is the probability of false alarm. Conventionally, one could formalize the problem as minimizing global probability of miss under global false alarm probability and communication constraints. Ignoring the communication constraints, a well known approach to this formulation is the Bonferroni procedure [23] and it leads to very conservative results. The basic issue is that as the number of observations grow larger, constraining the global probability of false alarm requires constraining the local probability of false alarm at much smaller levels. This is due to the multiplicity of false alarms as we explain via an example below. Such stringent constraints in local probability of false alarm in turn diminishes the detection power. The difficulties associated with this formulation for multiple comparison test problems have been widely recognized and we will consider a very different approach. Consider the following example for the multiplicity issue:

Example 1: Suppose $m = 10^4$ sensors scattered on a field. We desire a global false alarm probability, $P_F^g \leq 0.2$. Suppose local false alarm probability, P_F^l is bounded at 0.2 at each sensor. Then as many as 20%

of the sensors are expected to report false alarms with high probability. This implies that P_F^g will be close to one since with high probability at least one sensor will commit a false alarm. Therefore, to reduce P_F^g, P_F^l has to be bounded at much smaller levels, which in turn reduces the detection power greatly.

The false alarm and miss probabilities are associated with the random variables V and T respectively in the above table. Example 1 is based on a local thresholding strategy. Nevertheless, as we see show below, the error probability formulation is fundamentally flawed. We show via information-theoretic arguments that asymptotically the error probability is bounded away from zero by $\mathcal{H}(H^s \mid Y_s)$, the conditional entropy of H^s.

Theorem 3.1 *Consider the following optimization problems—worst-case and Bayesian with uniform prior distribution over all hypothesis—with performance costs γ_w, γ_b respectively. The minimization is taken over all centralized decision strategies, $u(Y^N)$, for N observations from the sensors under the ideal sensing model. Then,*

$$\gamma_w = \min_{u} \max_{H^s} \left(Pr\{V \geq 1 \mid \{H^s : s \in \mathcal{S}\}\} + Pr\{T \geq 1 \mid \{H^s : s \in \mathcal{S}\}\} \right) \geq \gamma_b$$

where

$$\gamma_b = \min_{u} Pr(V \geq 1) + Pr(T \geq 1) \geq \mathcal{H}(H^s \mid Y_s) - \frac{1}{N}$$

where $\mathcal{H}(\cdot \mid \cdot)$ is the conditional entropy. It follows that there exists no decision strategy for which both false alarm and miss probability can simultaneously be smaller than $\mathcal{H}(H^s \mid Y_s)/2$.

Proof: First note that from Lagrangian duality it follows that, $\gamma_w \geq \gamma_b$. Consequently, we are left to establish a bound for the Bayesian problem. Now we observe that the error event,

$$E = \{u(Y^N) \neq \{H^s : s \in \mathcal{S}\}\} = \{V \geq 1\} \cup \{T \geq 1\}$$

Therefore, from Fano's inequality it follows that for any strategy ϕ:

$$\Pr(V \geq 1) + \Pr(T \geq 1) \geq \Pr(E) \geq \frac{1}{N}\mathcal{H}(\{H^s : s \in \mathcal{S}\} \mid Y^N) - \frac{1}{N} = \mathcal{H}(H^s \mid Y_s) - \frac{1}{N}$$

The last equality follows from the independence assumptions. ∎

Remark: H^s is a binary random variable and so its entropy (or conditional entropy) is always smaller than one. Nevertheless, depending on measurement noise at each sensor $\mathcal{H}(H^s \mid Y_s)$ could be arbitrarily close to one.

3.2 False Discovery Rate Formulation

In the FDR [4] formulation, instead of controlling the probability of false alarm, we control the worst case expected ratio of the number of sensors that commit local false alarm (V), to the total number of sensors that declare presence of object within their sensing radius (R), i.e. $FDR = \max_{\{H^s\}_{s \in \mathcal{S}}} E\{V/R \mid \{H^s\}_{s \in \mathcal{S}}\}$.

For simplicity of notation, from here on we will write $FDR = E\{V/R\}$ and $\Pr\{\cdot \mid \{H^s\}_{s \in \mathcal{S}}\} = \Pr\{\cdot\}$ whenever it is clear from the context.

$$
\begin{aligned}
FDR &= E(Q) = E\{V/(V+S)\} = E\{V/R\} \\
&= E\{V/R|R > 0\}\Pr\{R > 0\} + E\{V/R|R = 0\}\Pr\{R = 0\} \\
&= E\{V/R|R > 0\}\Pr\{R > 0\}
\end{aligned}
$$

The last equality follows from the convention that $E\{V/R|R = 0\} = 0$. This is intuitively pleasing, because if no sensors declare presence of target within their sensing radius then no false discoveries are committed.

Inherently FDR is a function of two variables: the cardinality of $\mathcal{S}_1$ (m_1), and the decision rule (u); i.e. $FDR = \Xi(m_1, u)$. Now define $\Xi^*(u) = \max_{m_1} \Xi(m_1, u)$. Since the number of targets are unknown, $\Xi^*(u)$ is a proper definition of the worst-case FDR. For a given decision rule, $Q = V/R$ is a function of all the observations, therefore this expectation is in fact with respect to the joint density function of observations. However, for the ideal sensing model the joint density function can be simplified. The explicit form of this expectation is given below, and other expectations have analogous forms.

$$
\begin{aligned}
E(Q) &= \int Q(y_1, y_2, \ldots, y_m \mid H^1, H^2, \ldots, H^m) \, dp(y_1, \ldots, y_m \mid H^1, H^2, \ldots, H^m) \\
&= \int Q(y_1, y_2, \ldots, y_m \mid H^1, H^2, \ldots, H^m) \prod_{s \in \mathcal{S}_0} dp(y_s \mid H^s) \prod_{s \in \mathcal{S}_1} dp(y_s \mid \{H^s\}_{s \in \mathcal{S}_1})
\end{aligned}
$$

It is easy to establish that the probability of global false alarm is bounded from below [4] by FDR, i.e., $P_F^g = \Pr\{V \geq 1\} \geq E(Q) = FDR$ with equality when $m = m_0$. Thus, FDR is a weaker notion in terms of global false alarm probability. However this relaxation of the false alarm constraint can improve the detection power significantly, therefore FDR is a desirable approach in many problems.

Expected number of misses is a function of m_1, u, and $\mathcal{G}_1$ in which F_{1s} belongs; i.e. $E(T) = \Phi(m_1, \mathcal{G}_1, u)$. Again, we define $\Phi^*(\mathcal{G}_1, u) = \max_{m_1} \Phi(m_1, \mathcal{G}_1, u)$, which captures the worst-case scenario for the expected number of misses. Further define $\Phi^{**}(u) = \max_{G_{1s} \in \mathcal{G}_1} \Phi^*(\mathcal{G}_1, u)$, which takes into account the family of distributions under positive hypotheses.

Introducing a communication variable for each sensor at each time, we are ready to formalize the problem in its most general form:

$$
\begin{aligned}
&\min_u \; \Phi^{**}(u) \;\; subject \; to: \\
&\Xi^*(u) \leq \gamma, \\
&\sum_{s,t} C(u_s(t)) \leq \alpha
\end{aligned}
$$

where, $C(u_s(t))$ is the communication cost for sensor s at time t. The sensor can choose whether or not to communicate. The cost is zero if no bits are communicated and $\log |\mathcal{A}|$ bits in general when the decision,

$u_s(t)$ is communicated. In our case $\log|\mathcal{A}| = 1$.

Remark: One could also consider the same problem with false non-discovery rate (FNR) introduced by [14], i.e., $\Phi(m_1, \mathcal{G}_1, u) = E(T/(U+T)) = FNR$ and develop similar results. This is the natural counterpart of FDR, which represents the number of $H_1 s$ among the samples declared as H_0, and it can be interpreted as the miss rate. Although we make remarks about this view at several different parts of our paper, we keep our focus on the case where $\Phi(m_1, \mathcal{G}_1, u) = E(T)$ for simplicity of exposition.

For the ideal sensing model $\mathcal{G}_1$ is a singleton, and therefore $\Phi^{**}(u) = \Phi^*(\mathcal{G}_1, u)$. In the following sections we will develop the solution for the ideal sensing model. Later we will extend these results to a family of distributions based on a robustness approach to solve the non-ideal sensing model problem.

Controlling the FDR: When the false discovery rate was introduced, a decision making algorithm was proposed along with it. This algorithm takes the value γ as an input, and in return guarantees a false discovery rate below this input. The FDR procedure is described below and presented in Figure 3.

FDR procedure:

1. Calculate the p values for all the observations
2. Order the p values in ascending order
3. Find the largest index, i_{max}, such that $p_i \leq \frac{i}{m}\gamma$
4. Declare p_j significant for $0 \leq j \leq i_{max}$

The following definition of p value of a random variable X is used in this paper:

$$P(X) = \int_X^\infty g_0(t)dt = 1 - G_0(X) \tag{5}$$

where g_0 is the probability density function (pdf) of the observations under H_0. It is obvious that p value of a random variable is also a random variable.

Let $Y_{0s} \sim G_{0s}$ and $Y_{1s} \sim G_{1s}$. Define $P_{0s} = P(Y_{0s})$, and similarly $P_{1s} = P(Y_{1s})$ and let F_{0s} and F_{1s} be their corresponding distribution functions, i.e. $P_{0s} \sim F_{0s}$ and $P_{0s} \sim F_{0s}$. The family $\mathcal{G}_0$ is transformed to a new family $\mathcal{F}_0$ and $\mathcal{G}_1$ is transformed to a new family $\mathcal{F}_1$ by this transformation.

Lemma 3.2 *The random variable P_0 is distributed uniformly in (0,1) for all continuous distributions of X_0.*

Proof:

$$Pr\{P_0(X) > P_0(x)\} = Pr\{X < x\} = G_0(x)$$
$$1 - Pr\{P_0(X) \leq P_0(x)\} = G_0(x)$$
$$Pr\{P_0(X) \leq P_0(x)\} = 1 - G_0(x) = P_0(x)$$

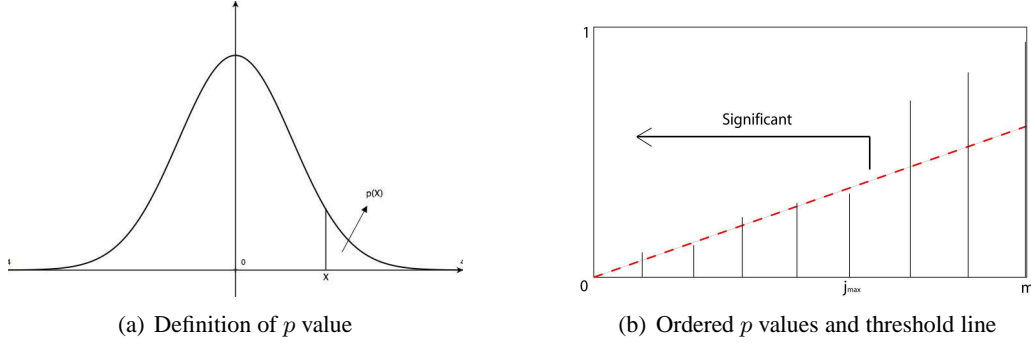(a) Definition of p value   (b) Ordered p values and threshold line

Figure 3: FDR algorithm

Note that for a given value x, $P_0(x)$ is not random. Now put $P_0(x) = p$. Then $Pr\{P_0(X) \le p\} = p$ and the result follows. ∎

The following theorem and its proof are also presented in [5].

Theorem 3.3 *For independent test statistics under null hypothesis, and for any configuration of positive hypotheses, the above procedure controls the FDR at level γ.*

Proof: $\gamma_i = i\gamma/m$. Let P_{0i} be the m_0 p values. Denote with $C_i(k)$ the event that if p_i is declared H_1, exactly $k-1$ other p values are declared H_1. Then;

$$
\begin{aligned}
E(V/R) &= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \mathrm{Pr}\{P_{0i} \le \gamma_k, C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \mathrm{Pr}\{P_{0i} \le \gamma_k\} \mathrm{Pr}\{C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \frac{\gamma k}{m} \mathrm{Pr}\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} \sum_{k=1:m} \mathrm{Pr}\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} = \frac{\gamma m_0}{m} \le \gamma
\end{aligned}
$$

The second equality follows because P_{0i} is independent of all other p values. See [5] for further details. ∎

Below we establish the simple but very important scaling property of the FDR procedure. It is due to this property that we can limit the communication budget with an upper bound that depends on the number of sensors that have a target within their sensing range, and not the total number of sensors in the SNET.

Theorem 3.4 (Scaling) *Let $m_1 = m - m_0$ be the number of sensors with positive hypothesis. The expected ratio of m_1 to the number of sensors that are declared to be significant (R) is lower bounded by $1 - \gamma$; i.e. $E\{m_1/R\} \ge 1 - \gamma$.*

11

Proof: (Scaling) We know that the FDR procedure guarantees $E\{V/R\} = E\{V/(V + S)\} \leq \gamma$. Evidently $S \leq m_1$, meaning the number of correct detections cannot exceed the number of sensors with positive hypotheses. Then:

$$E\{V/(V + S)\} = 1 - E\{S/(V + S)\} \Rightarrow E\{S/(V + S)\} \geq 1 - \gamma$$

$$1 - \gamma \leq E\{S/(V + S)\} \leq E\{m_1/(V + S)\} = E\{m_1/R\}$$

which concludes the proof of this theorem. ∎

This scaling property is important from the SNET perspective in the sense that the algorithm scales with m_1, and not the size of the SNET.

Although FDR procedure is a good candidate solution to the problem formulated above, it suffers from two significant drawbacks that make it unsuitable in our applications:

- First, the FDR procedure does not take into account the knowledge of the probability distributions that generate the samples under H_1. The focus primarily is to reduce false positives and there is no control over the miss rate. Therefore, depending on the problem the detection power can be improved.

- Second, the FDR strategy does not lend itself easily to decentralized implementation when F_1 is not concave.

In the next section we develop methods to overcome these issues.

4   Domain Transformed FDR Procedure

In many cases the adaptive solution that has been described in Section 3.2 is sufficient. However, FDR procedure performs best when the realizations of P_1 are clustered near zero, and that may not necessarily be the case as seen in the following example.

Example 2: Consider two Gaussian random variables with $g_0(x) \sim N(0, 1)$ and $g_1(x) \sim N(-4, 1)$, and consider the FDR constraint $\gamma = 0.05$. The goal is to detect as many samples of P_1 as possible from a mixture of samples subject to FDR constraint, γ. In this case most of the realizations of the random variable P_1 are close to 1 rather than 0, and FDR procedure will not declare them as significant.

In this situation the FDR procedure suffers from severe increase in miss rate. The issue here is that the algorithm searches for significant observations that are less than or equal to the small FDR threshold, while all the significant observations are near 1. To overcome this problem, consider the following transformation on the random variables P_0 and P_1:

$$\hat{P}_i = 1 - P_i, i = 0, 1 \tag{6}$$

Since P_0 is uniformly distributed in (0,1), it is obvious that $\hat{P}_0$ is also uniformly distributed in (0,1). Observe, however, that most of the realizations of $\hat{P}_1$ are close to 0. The transformation shifts all the significant p values to near 0 without changing the distribution of the insignificant observations, generating a more

suitable data set for the algorithm. Therefore, when the FDR procedure is performed on this new set of p values, more of the observations coming from H_1 will be declared as significant, thus the detection power is increased. Furthermore the FDR constraint γ is still satisfied since the transform preserves the $U(0,1)$ distribution of the p values for the observations coming from H_0.

Example 3: Another example of a case when realizations of P_1 are away from zero is angle observations in clutter. Conditioned on the predicted target location the observation from the target has a distribution around some angle θ, whereas the clutter is distributed uniformly between $-\pi$ and π, which pushes P_1 away from zero to around some θ'.

Such cases where realizations of P_1 are away from zero can arise when the null hypothesis is a mixture of various hypotheses. Then the variation of observations under null hypothesis can be much larger than the variation of observations under positive hypothesis, which can lead to the failure of FDR procedure. In this section, we develop a method to overcome the difficulties encountered by the FDR procedure. Conveniently, as a result of this method, $E(T)$ is also minimized. Before we go any further, we restate an important insight on FDR procedure, which follows from Lemma 3.2 and Theorem 3.3 :

Proposition 4.1 *False discovery rate constraint is satisfied for FDR procedures applied along with transformations that are measure invariant with respect to the distribution of the P_0.*

Proof: Applying a measure preserving transform with respect to the distribution of P_0 leaves the random variable P_0 uniformly distributed in (0,1). Since the distribution under null hypothesis is preserved, following the proof of theorem 3.3 from [4] gives the result stated in the proposition. ■

Below, we show that when the distribution of observations under positive hypothesis is available, one can design more powerful methods than the one described above, while still remaining within the FDR constraint γ. Making use of the assumption that the distributions of the observations are known under null and positive hypotheses, we introduce a transformation in the p domain. The transformation is simple in nature, and is a reorientation of the p domain. Despite its simplicity, it has three very important properties to note:

1. It preserves uniform distribution of p values under null hypothesis,
2. It maps an arbitrary pdf of p values to a monotonically decreasing one,
3. Asymptotically the transformed p values are samples of a convex function.

The transformation is dependent on the distribution of the observations under H_1, but not the realizations themselves. Therefore it can be defined apriori. Due to the first property, the FDR constraint is satisfied after the transformation, and by the second property the detection power of FDR procedure is maximized. The third property plays a crucial role in the implementation of FDR procedure in a distributed manner.

13

4.1 Transformation of p Domain

Let μ be any probability measure that is absolutely continuous with respect to Lebesgue measure, and let ϕ be the pdf of a continuous random variable in a subset of (0,1). Define the tuple:

$$(\alpha_\mu(y), \beta_\mu(y)) = (E_U[I_{\{\phi(x) \geq y\}}(x)], E_\mu[I_{\{\phi(x) \geq y\}}(x)])$$

where U stands for the uniform measure. Now, define the new measure in the following way:

$$\hat{\mu}(0, \alpha_\mu(y)) = \beta_\mu(y)$$

If $\alpha_\mu(y)$ has a jump at $y = y_0$ from a to b, then set

$$\hat{\mu}(0, z) = \frac{\beta_\mu(y_0^-) - \beta_\mu(y_0^+)}{b - a}(z - a) + \beta_\mu(y_0^+)$$

for $z \in (a, b)$, which corresponds to a conditionally uniform distribution in (a,b). We now define $\hat{F}_1(\alpha_\mu(y)) = \beta_\mu(y)$.

Various terms involved in this definition can be understood through Figure 4, for which $\phi = d\mu/dx$.
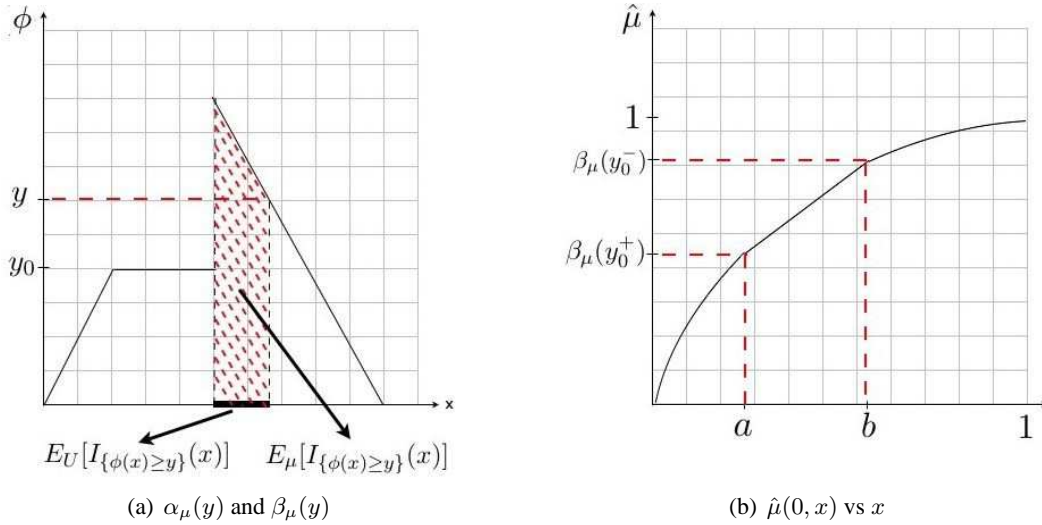

(a) $\alpha_\mu(y)$ and $\beta_\mu(y)$

(b) $\hat{\mu}(0, x)$ vs x

Figure 4: Illustration of transformation elements

Below we show that $\hat{F}_1(\cdot)$ is a distribution function.

Proposition 4.2 $\hat{F}_1(\alpha_\mu(y)) = \hat{\mu}(0, \alpha_\mu(y)) = \beta_\mu(y)$ *is a distribution function.*

Proof: To show that $\hat{F}_1(\cdot)$ is a distribution function, we need to show that **(1)** $\hat{F}_1(\cdot)$ is monotone increasing, **(2)** $\hat{F}_1(\cdot)$ is right continuous, **(3)** $\lim_{x \to -\infty} \hat{F}_1(x) = 0$ and $\lim_{x \to +\infty} \hat{F}_1(x) = 1$.

(1) $\hat{F}_1(\alpha_\mu(y))$ is monotone increasing in $\alpha_\mu(y)$. This is because $\alpha_\mu(y)$ increases as y decreases, and as y decreases $\beta_\mu(y) = E_\mu[I_{\{\phi(x) \geq y\}}(x)]$ increases.

(2) Next, $\hat{F}_1(\alpha_\mu(y))$ is a right-continuous function. To show this, consider some α', and appropriate y' and β'. $\alpha_\mu(y) \downarrow \alpha'$ as $y \uparrow y'$. But, $y \uparrow y' \Rightarrow \beta_\mu(y) \downarrow \beta'$, and since $E_\mu[I_{\{\phi(x) \geq y\}}(x)]$ is right continuous so is $\hat{F}_1(\alpha_\mu(y)) = \beta_\mu(y)$.

(3) Finally, we show that $\lim_{\alpha_\mu(y) \to 0} \hat{F}_1(\alpha_\mu(y)) = 0$ and $\lim_{\alpha_\mu(y) \to 1} \hat{F}_1(\alpha_\mu(y)) = 1$. The reason that we only consider 0 and 1 as the limit points is because $\alpha_\mu(y) \in [0, 1]$ by definition.

Note that since ϕ is a pdf of a continuous random variable, there exists a y' such that $\{x : \phi(x) \geq y'\}$ is empty. Therefore, as $y \uparrow y'$, $\alpha_\mu(y) \downarrow \alpha_\mu(y') = 0$, and $\beta_\mu(y) \downarrow \beta_\mu(y') = 0$. This shows the first part. Next, consider the case when $y \downarrow 0$. In this case $\alpha_\mu(y) \uparrow \alpha_\mu(0) = 1$ and $\beta_\mu(y) \uparrow \beta_\mu(0) = 1$. This establishes that $\hat{F}_1(\alpha_\mu(y))$ is a distribution function. ■

We now show that $\hat{\mu}$ is absolutely continuous with respect to Lebesgue measure, thus admits a pdf.

Proposition 4.3 $\hat{\mu}$ *is absolutely continuous with respect to Lebesgue measure.*

Proof: Over any zero measure set A with respect to Lebesgue measure, $\alpha_\mu(y) = 0$ and $\beta_\mu(y) = 0$ and therefore $\hat{\mu}(A) = 0$. ■

Now let $f_1(\cdot)$ be the pdf of P_1, which exists since P_1 is a continuous random variable. Define the transformation, $\mathcal{T}$ as follows:

1. Let $y_{max} = \sup_x\{f_1(x)\}$;
2. Define the tuple $(\alpha_\mu(y), \beta_\mu(y)) = (E_U[I_{\{f_1(x) \geq y\}}(x)], E_\mu[I_{\{f_1(x) \geq y\}}(x)])$ for $y \in (0, y_{max})$ where μ is the measure of P_1; i.e. $\mu = \int f_1$. Then the new measure: $\hat{\mu}(0, \alpha_\mu(y)) = \beta_\mu(y)$ $\forall y \in (0, y_{max})$
3. Let $\hat{f}_1(\cdot)$ be the corresponding density of $\hat{\mu}$. Generate the transformed random variable $\hat{P} = \mathcal{T}[P]$ as follows: For $P \in (0, 1)$ find $Y = f_1(P)$; then find the set $S = \{x : \hat{f}_1(x) = Y\}$ and choose $\hat{P}$ randomly from S.

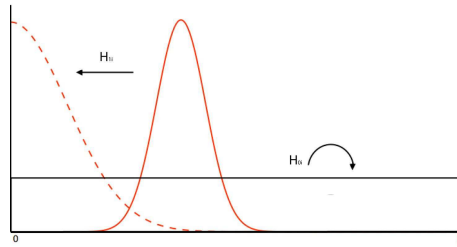Figure 5 illustrates the nature of this transformation.

Figure 5: Illustration of domain transform: a monotonically decreasing density is obtained.

The following procedure will be referred as the Domain Transformed FDR (DTFDR) procedure throughout the paper:

15

1. Apply the transformation $\mathcal{T}$ to realizations of the random variables P_0 and P_1

2. Follow the FDR procedure

The propositions below directly follow:

Proposition 4.4 *$\mathcal{T}$ is a measure invariant transformation for uniform distribution in (0,1).*

Proof: For uniform distribution, $\alpha_\mu(y)$ has a jump from 0 to 1 at $y_0 = 1$. Furthermore, $\beta_\mu(y_0^+) = 0$ and $\beta_\mu(y_0^-) = 1$. Therefore, for $z \in (0,1)$, following the definition of $\hat{\mu}$ we have $\hat{\mu}(0,z) = z$, and the uniform distribution is preserved. ∎

Proposition 4.5 *The DTFDR procedure controls the false discovery rate at the same level as the FDR procedure.*

Proof: The proposition is a direct consequence of Proposition 4.1 and Lemma 3.2 and Proposition 4.4. Observe that by Lemma 3.2 P_0 has uniform measure, and by Proposition 4.4 so does $\mathcal{T}[P_0]$. Then, by Proposition 4.1, the false discovery rate will not change if samples of H_0 are obtained from $\mathcal{T}[P_0]$ instead of P_0. ∎

Proposition 4.6 *The measure transformation $\mathcal{T}$ converts an arbitrary continuous density of P_1 to a monotonically decreasing density over $(0,1)$; i.e. $\hat{f}_1(\hat{p})$ is monotonically decreasing in $\hat{p}$.*

Proof: Note that $\beta_\mu(y)$ is concave as a function of $\alpha_\mu(y)$. To show this consider $y_1 > y_2 > \ldots > y_n$ and a sequence of sets $A_1 \subset A_2 \subset \ldots \subset A_n$ such that $A_i = \{x : f_1(x) \geq y_i\}$. Note that $\mathcal{L}(A_1) \leq \mathcal{L}(A_2) \leq \ldots \leq \mathcal{L}(A_n)$ where $\mathcal{L}$ denotes the Lebesgue measure and $\sup_{x \in A_{i+1} - A_i} f_1(x) \leq \inf_{x \in A_i} f_1(x)$. Since $\sup_{x \in A_{i+1} - A_i} f_1(x) \leq \inf_{x \in A_i} f_1(x)$, $d\beta_\mu/d\alpha_\mu$ is monotonically decreasing in these sets. Therefore, the new measure $\hat{F}_1(\alpha_\mu) = \hat{\mu}(0, \alpha_\mu)$ is concave, and the proposition follows. ∎

Proposition 4.6 leads us to the following useful result: Assume that we are given a distribution that is a mixture of f_1 and $f_0 = U(0,1)$. Then, no matter what the mixing parameter is, we can perform the domain transformation to obtain a mixture that is monotonically decreasing. We present this result as a lemma below. The proof follows as an immediate consequence of Proposition 4.6.

Lemma 4.7 *Let $f = \epsilon f_1 + (1 - \epsilon)f_0$ be a mixture density with $f_0 = U(0,1)$. The transformation $\mathcal{T}_f$ defined using f generates a monotonically decreasing density $\hat{f}_1$ and retains $\hat{f}_0 = U(0,1)$.*

This is particularly useful when we know the mixture density and that f_0 is uniform. Then, this lemma implies that we need not know the mixing parameter ϵ or f_1 in order to be able to define a transformation that converts the f_1 density to a monotonically decreasing one.

Proposition 4.8 *For a continuously distributed P_1, the mixture cumulative distribution function (CDF) of $\hat{P}_0 = \mathcal{T}[P_0]$ and $\hat{P}_1 = \mathcal{T}[P_1]$ is concave.*

Proof: The observations from null hypotheses are distributed uniformly, i.e. $P_0 \sim U(0,1)$. Proposition 4.4 implies that the transformation preserves this distribution, i.e. $\hat{P}_0 = \mathcal{T}[P_0] \sim U(0,1)$. Furthermore, due to Proposition 4.6 the pdf of $\hat{P}_1 = \mathcal{T}[P_1]$ is monotonically decreasing. Thus the mixture density is monotonically decreasing and the mixture CDF is concave. ∎

4.2 Optimality of DTFDR

Note that for any random variable P, $\mathcal{T}[P]$ is the stochastically smallest random variable attainable while preserving the uniform distribution. Therefore, we have established so far that the transformation in the p domain described in Section 4.1 clusters the significant observations near zero with high probability.

Here we define threshold strategies, and we show that with the domain transformation a threshold strategy is optimal.

Definition 4.9 *Assume a clustering problem of a set of observations $X = \{x_1, x_2, \ldots, x_m\}$ into two subsets, X_1 and X_2 such that $X_1 \cap X_2 = \phi$ and $X_1 \cup X_2 = X$. A threshold strategy is one that computes a threshold $t(x_1, x_2, \ldots, x_m)$, and maps $X_1 = \{x : x \in X, x \leq t(x_1, x_2, \ldots, x_m)\}$.*

Given the above definition, we now establish the optimality of threshold strategies for multiple comparison problems after the domain transformation.

Let $f_0(\cdot)$ and $f_1(\cdot)$ be pdfs of observations under H_0 and H_1 respectively. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ be given, where P_i are independent random variables having pdfs f_0 or f_1 with unknown prior probabilities $\Pr\{H^i = H_0\}$ and $\Pr\{H^i = H_0\}$ respectively. Now consider the partitioning problem as described in the above definition. Let S be a decision rule that chooses $\mathcal{P}_S \subset \mathcal{P}$ and labels H_1. Define $P^* = \max_i\{P_i \in \mathcal{P}_S\}$ and $P_* = \min_i\{P_i \in \mathcal{P}_S^c\}$ where $\mathcal{P}_S^c = \mathcal{P} - \mathcal{P}_S$. Now define a new decision rule S' as follows: If $\mathcal{P}_s \neq \phi$ and $P_* \leq p < P^*$ for some p, then S' chooses $\mathcal{P}_{S'} = (\mathcal{P}_S - \{P^*\}) \cup \{P_*\}$ and labels H_1. In all other cases S' chooses $\mathcal{P}_{S'} = \mathcal{P}_s$. With this setup, the following theorem establishes that with the introduction of domain transformation, the optimal strategy for a multiple comparison problems is a thresholding strategy.

Theorem 4.10 (Thresholding Strategies) *Let relevant quantities be defined as above. Assume that $\Pr\{H^i = H_0\} = \Pr\{H^j = H_0\}$, $\forall i, j : 1, \ldots, m$. If $f_0(\cdot) = U(0,1)$ and $f_1(\cdot)$ is a monotonically decreasing pdf, then the false discovery rate of the strategy S' is less than or equal to that of the strategy S, i.e., $FDR_S \geq FDR_{S'}$.*

Proof: See appendix. ∎

Corollary 4.11 *If $\Pr\{H^i = H_0\} = \Pr\{H^j = H_0\}$, $\forall i, j : 1, \ldots, m$ and $f_0(\cdot) = U(0,1)$ and $f_1(\cdot)$ is a monotonically decreasing pdf, a thresholding strategy is optimal for the partitioning problem.*

Proof: The proof follows by induction from the proof of Theorem 4.10. ∎

Theorem 4.12 (Optimality) *DTFDR procedure is minimax optimal, i.e., minimizes the expected number of misses, $E(T)$ subject to an FDR constraint, γ.*

Proof: (Optimality) Theorem 4.10 implies that the optimal strategy is a thresholding strategy for the partitioning problem when $f_0(\cdot) = U(0,1)$ and $f_1(\cdot)$ is monotonically decreasing. Note that for any problem, after the domain transformation, indeed $f_0(\cdot) = U(0,1)$ and $f_1(\cdot)$ is monotonically decreasing.

Since the transformation converts the pdf of P_1s to a monotonically decreasing one, as the indices of the p values increase, the likelihood of each p value being a sample of P_1 monotonically decreases. Moreover,

the transformation generates the stochastically smallest random variable $\mathcal{T}[P]$. Thus, regardless of the point of termination, the algorithm terminates with the largest possible expected ratio of number of H_1s to number of H_0s up to the point of termination.

Now, note that in the worst case the FDR constraint is tight, i.e. $FDR = \gamma$, which in attained when all p values are from H_0s, i.e., $m_0 = m$. Therefore, DTFDR is minimax optimal. ∎

4.3   Performance Comparison of DTFDR and FDR

Before proceeding any further, the term *stochastically larger* [22] must be introduced: We say that the random variable X is stochastically larger than the random variable Y, denoted $X \geq_{st} Y$, when $F_X(a) \leq F_Y(a)$ for all a.

Lemma 4.13 *Let $X_1..X_n \in (0,1)$ be n independent random variables with common density function f_X and let $Y_1..Y_n \in (0,1)$ be n independent random variables with common density function f_Y. Also, let $X_{(i)}$ and $Y_{(i)}$ denote the ith smallest of $X_1..X_n$ and $Y_1..Y_n$ respectively. If $F_X(t) \geq F_Y(t) \; \forall t \in (0,1)$, then $Y_{(i)} \geq_{st} X_{(i)}$.*

Proof: See appendix. ∎

The important implication of this lemma is captured in the following corollary. Although we have proved optimality of DTFDR, we present the corollary and its proof for completeness.

Corollary 4.14 *For any given set of p values with known distributions and any integer k, the probability of declaring the first k p values as significant is larger under the DTFDR procedure than the FDR procedure.*

Proof: Let $\hat{P}_1 = \mathcal{T}[P_1]$. By construction of the transformation, the density of $\hat{P}_1$, $\hat{f}_1$, dominates the density of P_1, f_1. In other words, $\hat{P}_1 \leq_{st} P_1$. Therefore, the results of the lemma 4.13 apply to random variables P_1 and $\hat{P}_1$.

First, assume that the observations contain only samples from H_0. Since the random variable $\mathcal{T}[P_0]$ is stochastically equivalent to P_0, the probability of declaring k of them significant is equal for all k with both procedures. Let p' be the kth p value. Next, when the samples of H_1 are added one by one, probability of the index of p' increasing to $k+1$ is larger with addition of samples from $\mathcal{T}[P_1]$ in comparison to addition of samples from P_1. This is because P_1 is stochastically larger than $\mathcal{T}[P_1]$. But, since $p' \leq k\gamma/m$ implies $p' \leq (k+1)\gamma/m$ the DTFDR procedure increases the probability of a p value being declared as significant. Furthermore, this argument is valid for all $k \leq m$, since $\mathcal{T}$ converts an arbitrary continuous density of P_1 to a monotonically decreasing one, which concludes the proof of the corollary. ∎

Figure 6 demonstrates the detection power of DTFDR procedure in comparison to that of the FDR procedure. The DTFDR procedure is uniformly stronger than the FDR procedure in terms of detection power.

4.4   Distributed DTFDR Algorithm

The DTFDR procedure described above is a centralized procedure, in which the ordering of p values is necessary. However, in a SNET context a decentralized algorithm is desired. Since ordering of p values is

(a) Original and transformed pdf of P_1 (not normalized)

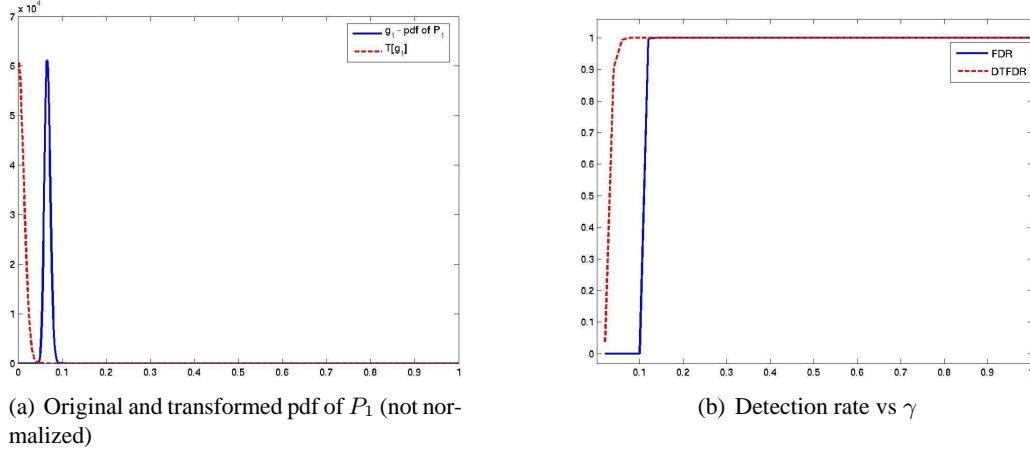(b) Detection rate vs γ

Figure 6: Comparison of detection performance for FDR and DTFDR procedures.

costly in terms of communications, we use feedback to accomplish the linearly increasing thresholding. The decentralized procedure is described below.

At iteration t each sensor keeps a threshold variable $l(i_t) = i_t\gamma/m$ and a counter $count_t = \#\{j : p_j \leq l(i_t)\}$. Furthermore, a minimum number of tests, say k, is predetermined. Initialize $i_1 = 1$ and $count_0 = 0$. Then, at iteration t:

1. Sensor j decides H_{1j} if $p_j \leq l(i_t)$ and H_{0j} otherwise. If decided H_{1j}, announces to the network if it has not done so at iterations $1 \ldots t - 1$;

2. Assume R_t sensors decide H_1 and declare to the network. Update variables:

   (a) If $R_t > 0$, set: $count_t = count_{t-1} + R_t$, $i_{t+1} = count_t + 1$, mark iteration t_{max}, go to step 1;

   (b) If $R_t = 0$ and $i_t < k$ set: $count_t = count_{t-1}$, $i_{t+1} = count_t + 1$, go to step 1;

   (c) If $R_t = 0$ and $i_t > k$ $halt$;

3. Label sensors that declare H_1 until iteration t_{max} as significant.

The feedback in the network makes it possible to retain the dynamic, linearly increasing structure of the centralized thresholding strategy. Furthermore, only the sensors that have decided H_1 declare their decisions to the network with a one bit message and excessive communications due to a large number of false alarms are avoided.

4.5   Scalability of Distributed DTFDR

Note that the distributed algorithm is a first-crossing procedure. Once the threshold line crosses below the ordered p values, the algorithm terminates. When there is a large number of observations, sampling noise is reduced and the ordered p values form a convex function. In this case, the first-crossing and the last-crossing

procedures are the same, and they terminate at the same point. However, when there is not enough samples to diminish the sampling noise, then the ordered p values do not form a convex function, and therefore the first-crossing and the last-crossing procedures have different termination points. We wish to capture the last-crossing procedure via the above algorithm without incurring excessive communication cost. Below we show that the above procedure can be made to achieve the last-crossing performance with high probability when there is sampling noise.

Let the largest p value that is below its corresponding threshold be p'. Due to the nature of the distributed algorithm, it is important that all of the p values that are smaller than p' are also below their corresponding thresholds, otherwise the algorithm terminates prematurely. Here we to show that the above distributed strategy is scalable and that the probability of terminating prematurely decays exponentially fast with k. In other words, the distributed algorithm can be made to perform as well as the centralized algorithm without incurring a communication cost that scales with m, the number of sensors. For this part, let $\theta_0 = m_0/m$, $\theta_1 = m_1/m$, be the ratio of observations from each hypothesis. We use the definition of $l(i) = i\gamma/m$, $i = 1 \ldots m$ as the threshold line for the centralized FDR procedure. We assume that the pdf of P_1 is monotonically decreasing due to the domain transformation.

Lemma 4.15 *Let p_k be the kth smallest p value. If $E(p_{\lceil \frac{k}{1-\epsilon} \rceil}) \leq l_k$, then $Pr\{p_k > l_k\}$ decays exponentially fast with k.*

Proof: Let $N_k = \#\{j : p_j \leq l_k\} = \sum_{j=1}^{m} I_{\{p_j \leq l_k\}}$. By the switching lemma (see for example [1]) the following relationship holds for any k: $\{E(p_{\lceil \frac{k}{1-\epsilon} \rceil}) \leq l_k\} \Leftrightarrow \{E(N_k) \geq \lceil \frac{k}{1-\epsilon} \rceil\}$. Therefore, $E(p_{\lceil \frac{k}{1-\epsilon} \rceil}) \leq l_k \Rightarrow E(N_k) \geq \frac{k}{1-\epsilon}$ and $k \leq E(N_k)(1-\epsilon)$.

$$
\begin{aligned}
Pr\{p_k > l_k\} &= Pr\{N_k < k\} & (7)\\
&\leq Pr\{N_k < E(N_k)(1-\epsilon)\} & (8)\\
&\leq exp\{-\frac{\epsilon^2 E(N_k)}{2}\} & (9)\\
&\leq exp\{-\frac{\epsilon^2 k}{2(1-\epsilon)}\} & (10)
\end{aligned}
$$

Inequality 9 follows from the Chernoff bound, and inequality 10 follows from the application of switching relation along with the assumption of the theorem. ∎

Theorem 4.16 *Consider the first crossing procedure with k (constant) number of tests, as described in Section 4.4. For that distributed implementation:*
a) $FDR \leq \gamma$ *and*
b) *The expected number of bits required to detect the objects is upper bounded by* $\max\{k, E(\frac{m_1}{1-\gamma})\}$

Proof: **a)** We show this part by showing that the distributed algorithm is in fact equivalent to the centralized algorithm, and that presetting k tests affects only the communication cost. If there exists a $p_i \leq i\gamma/m$, $i \geq$

k, then the effect of k preset tests is washed out. This is because the centralized algorithm would also declare all p values less than p_i significant. Therefore $FDR \leq \gamma$ in this case. If there is no such p_i, $i \geq k$, then the algorithm chooses the largest p value $p_j \leq j\gamma/m$, $j < k$, and declares all smaller p values significant. Therefore the distributed algorithm is equivalent to the centralized FDR procedure, hence $FDR \leq \gamma$.
b) Without the k preset tests the upper bound was given by Theorem 3.4, and the result is immediate from there. ∎

The theorem states that after a certain number of p values, say k, are tested against their corresponding thresholds, one can decide whether or not to continue the distributed algorithm with an exponentially small probability of error. This justifies presetting a k number of tests at the beginning of the algorithm. As expected, to achieve this robustness property one must endure k bits of communication cost. Note, however, that k can be fixed a priori and does not depend on the size of the SNET as long as the number of sensors is larger than k.

5 Robustness Analysis and Non-Ideal Sensing Model

The theoretical development until this point of the paper is based on ideal sensing models where $\mathcal{G}_1$ and $\mathcal{G}_0$, and hence $\mathcal{F}_1$ and $\mathcal{F}_0$, were singletons. It remains to show that this method is applicable to problems with non-ideal sensing models where they are not singletons but families of distributions.

To motivate the robustness development, consider a simple setup: a sensor field with targets be $T = \{t_1, t_2, \ldots, t_k\}$ with uniform signal power θ. Define $d(s, t)$ to be the Euclidian distance between sensor s and target t. Assume that the target signal θ decays rapidly with distance. The null hypothesis for a sensor is that it is outside the effective region of all targets, and the alternative is that it is inside the effective region of a target. Now, let J be a set of targets that are observed by sensor s under the alternative hypothesis. Then, with n_s and ν_s being noise at sensor s for null and alternative hypotheses respectively, the observation model at sensor s for the non-ideal case is as follows:

$$H_0 \quad : \quad X_s = \theta'_s + n_s, \text{ where } \theta'_s = \sum_{j \notin J} \frac{\theta}{d(s, t_j)}$$

$$H_1 \quad : \quad X_s = \theta_s + \nu_s, \text{ where } \theta_s = \sum_{j \in J} \frac{\theta}{d(s, t_j)}$$

There is an uncertainty in the observation model due to θ_s and θ'_s terms, because J and $d(s, t_j)$ are unknown.

For families $\mathcal{F}_0$ such that $\mathcal{F}_0 = \{F : |F(x) - F_0(x)| \leq \epsilon x\}$ for nominal $F_0(x) = x$, the proof of FDR theorem gives us an immediate non-asymptotic robustness result, which we state by the following theorem.

Theorem 5.1 *For families $\mathcal{F}_0$ such that $\mathcal{F}_0 = \{F : |F(x) - F_0(x)| \leq \epsilon x\}$ for nominal $F_0(x) = x$, the false discovery rate is bounded by $\gamma(1 + \epsilon)$, i.e. $FDR \leq \gamma(1 + \epsilon)$.*

Proof: Following the proof of Theorem 3.3, let $\gamma_i = i\gamma/m$. Then we have:

$$
\begin{aligned}
E(V/R) &= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \Pr\{P_{0i} \leq \gamma_k, C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \Pr\{P_{0i} \leq \gamma_k\} \Pr\{C_i(k)\} \\
&\leq \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \frac{\gamma k}{m} (1+\epsilon) \Pr\{C_i(k)\} \\
&= (1+\epsilon) \sum_{i=1:m_0} \frac{\gamma}{m} \sum_{k=1:m} \Pr\{C_i(k)\} \\
&= (1+\epsilon) \sum_{i=1:m_0} \frac{\gamma}{m} = (1+\epsilon) \frac{\gamma m_0}{m} \leq (1+\epsilon)\gamma
\end{aligned}
$$

∎

Although this result gives us insight about the robustness property of DTFDR, the assumptions are too strong. We next develop a general result that is valid under weaker assumptions.

Let $\mathcal{G}_0$ be a family of distributions in the observation space. To define a transformation from the observation space to p space we choose a member of $\mathcal{G}_0$, say G_{0i}, and define the transformation as $P(X) = \int_X^\infty g_0^i(x)dx$. If there exists a stochastically largest member of $\mathcal{G}_0$, then we choose that member to define the transformation. This guarantees that all the members of the new family $\mathcal{F}_0$ in the p domain have the following property: $F_{0i}(x) \leq x$. Then, following the proof of Theore 3.3, it is clear that $FDR \leq \gamma$. As an example, consider a family of Gaussian distributions with the same variance and different means. The one with the largest mean is the stochastically largest one. We now formalize this result.

Lemma 5.2 *Let X be a continuous random variable with distribution $G_X \in \mathcal{G}_0$. Let $P(X) = \int_X^\infty g_0(x)dx$ for some $G_0 \in$ go. If $G_0(x) \leq G_X(x) \ \forall x$, then $F_P(p) \leq p$ where F_P is the distribution function of P.*

Proof:

$$
\begin{aligned}
\Pr\{P(X) \leq P(x)\} &= \Pr\{X > x\} \\
&= 1 - \Pr\{X \leq x\} \\
&= 1 - G_X(x) \\
&\leq 1 - G_0(x) \\
&= \int_x^\infty g_0(x')dx' = P(x)
\end{aligned}
$$

Putting $P(x) = p$ we get $\Pr\{P(X) \leq p\} \leq p$, which completes the proof. ∎

Observe that not for every family $\mathcal{G}_0$ there will be a member which is stochastically largest. It is therefore important to be able to develop a method that can address families where there is no stochastically largest

member. Consider for example a family of two Gaussians with a common zero mean, but different variances. In this case non of the distributions is stochastically larger than the other. However, we can find regions in which one dominates the other, and define the transformation accordingly. We formalize this idea in the following lemma.

Lemma 5.3 *Let $\mathcal{G}_0$ be a countable family of distributions and $A_i = \{x : G_{0i}(x) \leq G_{0j}(x) \ \forall \ G_{0j} \in \mathcal{G}_0\}$. Also let X be a continuous random variable with distribution function $G_X \in \mathcal{G}_0$. If the transformation is defined as $P(X) = \sum_i I_{\{X \in A_i\}}(1 - G_{0i}(X))$, then $F_P(p) \leq p$ where F_P is the distribution function of P.*

Proof:

$$
\begin{aligned}
\Pr\{P(X) \leq P(x)\} &= \sum_i \Pr\{P(X) \leq P(x) \mid X \in A_i\}\Pr\{X \in A_i\} \\
&= \sum_i \Pr\{X > x \mid X \in A_i\}\Pr\{X \in A_i\} \\
&= \Pr\{X > x\} \\
&= 1 - G_X(x) \\
&\leq 1 - \sum_i I_{\{x \in A_i\}}G_{0i}(x) = P(x)
\end{aligned}
$$

Again, putting $P(x) = p$ we get $\Pr\{P(X) \leq p\} \leq p$, which completes the proof. ∎

The construction of the transformation in Lemma 5.3 is more general than that in Lemma 5.2. However, there is an immediate generalization to uncountable families, and we present that generalization in the following lemma.

Lemma 5.4 *Let $\mathcal{G}_0$ be a family of distributions and let X be a continuous random variable with distribution function $G_X \in \mathcal{G}_0$. Define a new function as $G^*(x) = \inf_{G \in \mathcal{G}_0} G(x)$. Now define the transformation as $P(X) = 1 - G^*(X)$. Then $F_P(p) \leq p$ where F_P is the distribution function of P.*

Proof: We first need to show that G^* is non-decreasing. Then the rest of the proof is similar to that of Lemma 5.2.

If G^* is not non-decreasing, there exists some $x < y$ and ϵ such that $G^*(x) > G^*(y) + \epsilon$. Now, since G^* is the point-wise infimum, there exists some $G \in \mathcal{G}_0$ such that $G(y) < G^*(y) + \epsilon$. But G is a distribution function, and hence non-decreasing. That implies $G(x) \leq G(y)$. Putting the inequalities together we get $G(x) \leq G(y) < G^*(y) + \epsilon < G^*(x)$. However, this contradicts that G^* is the point-wise infimum at x, and therefore G^* must be non-decreasing. Now using that G^* is non-decreasing, we have:

$$
\begin{aligned}
\Pr\{P(X) \leq P(x)\} &= \Pr\{G^*(X) \geq G^*(x)\} \\
&= \Pr\{X \geq x\} \\
&= 1 - G_X(x) \\
&\leq 1 - G^*(x) = P(x)
\end{aligned}
$$

23

Again, putting $P(x) = p$ we get $\Pr\{P(X) \le p\} \le p$, which completes the proof. ∎

Remark: In fact it can be shown that G^* is a distribution function by showing that G^* is upper semi-continuous in addition to the above property. In appropriate circumstances resampling can be done with respect to G^*. However for our purposes it suffices that it is a non-decreasing function, therefore we omit the proof that G^* is a distribution function.

Using the construction of Lemma 5.4 we develop with the following theorem that for any family of continuous distributions $\mathcal{G}_0$, the FDR is bounded from above by γ.

Theorem 5.5 *Let X be a continuous random variable with distribution $G_X \in \mathcal{G}_0$ for a family of distributions $\mathcal{G}_0$. If the p transformation be defined as in Lemma 5.4, then the false discovery rate is bounded by γ.*

Proof: Let $\gamma_k = \gamma k/m$. Observing that the distribution function $F_{0i}(x) \le x$ for all null distributions, and following the proof of Theorem 3.3 we have:

$$
\begin{aligned}
E(V/R) &= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k}\Pr\{P_{0i} \le \gamma_k, C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k}\Pr\{P_{0i} \le \gamma_k\}\Pr\{C_i(k)\} \\
&\le \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k}\frac{\gamma k}{m}\Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} \sum_{k=1:m} \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} = \frac{\gamma m_0}{m} \le \gamma
\end{aligned}
$$

The inequality follows from $\Pr\{P_{0i} \le \gamma_k\} \le \gamma_k$ which is a result of Lemma 5.3. ∎

The core idea behind Theorem 5.5 is the construction of a p value definition such that the distribution of $P(X)$ satisfies $F_P(p) \le p$. We have shown how we can generate such distributions. However, when we apply the domain transformation to F_P, it is not necessarily the case that $\hat{F}_P(p) \le p$. Therefore, we must devise a robust solution that can address problems where $\hat{F}_P(p) \le p$ may not hold true.

For this part we choose the Kolmogorov distance as the relevant metric and consider families $\mathcal{F}_0$ of size ϵ such that $F_0(x) = x$ is a member of $\mathcal{F}_0$. Formally, let $\mathcal{F}_0$ be a family of distributions such that $\mid F_{0i}(x) - F_0(x) \mid \le \epsilon$ for all $F_{0i} \in \mathcal{F}_0$. Observe that this construction allows the case $\mid F_{0i}(0) - F_0(0) \mid = \epsilon$, which introduces a point mass difference at zero and includes the worst case scenario. The point mass can occur when the support of the member of the families do no overlap as in the following example with uniform distributions: Assume that the observations are obtained from the density g_0 and the p values are obtained via g_0'. In that case, all the observations that belong to the interval (a, b) have p values that equal to zero, and the probability of getting a zero p value is $\epsilon = (b - a)/b$.

Based on this family of distributions, we modify the FDR procedure to compensate for the uncertainty in the family, and establish that the new procedure guarantees false discovery rate control at γ. The modification adds ϵ to the p values and sets the minimum number of tests to be performed to $k \geq \epsilon m$. This addition of ϵ to the p values allows for the control of FDR. We present this result in the following theorem, the proof of which can be found in the appendix.

Theorem 5.6 *Let P_{0i} be a random variable with distribution function $F_{0i} \in \mathcal{F}_0$, where $\mathcal{F}_0$ is a family of distributions such that $\mid F_{0j}(x) - x \mid \leq \epsilon$ for all $F_{0j} \in \mathcal{F}_0$. If $P'_{0i} = P_{0i} + \epsilon$, then false discovery rate is controlled at level γ with the FDR procedure for a set of p values defined in the same fashion to P'_{0i}.*

Proof: See appendix. ∎

Note that until this point we developed our results based on the FDR procedure and have not considered the domain transformation. We now introduce the domain transformation and transform the family $\mathcal{F}_0$ to a new family $\hat{\mathcal{F}}_0$. In the following theorem we present conditions on $\mathcal{F}_0$ such that the domain transformation generates a new family $\hat{\mathcal{F}}_0$ that is of size ϵ in terms of Kolmogorov distance and includes uniform distribution.

Theorem 5.7 *Let μ_0 be the measure associated with $F_0 \in \mathcal{F}_0$ and μ_1 be the measure associated with of $F_1 \in \mathcal{F}_0$. Define $\hat{F}_0$ and $\hat{F}_1$ to be the respective distributions after the transformation. If $d_{tv}\{\mu_0, \mu_1\} \leq \epsilon$ then $\sup_x \mid \hat{F}_0(x) - \hat{F}_1(x) \mid \leq \epsilon$ where for any measurable space Ω, $d_{tv}(\mu, \nu) = \sup_{A \subset \Omega} \mid \mu(A) - \nu(A) \mid$.*

Proof: Let $A_x \subset (0,1)$ be the set that gets mapped to the set $(0, x)$ by the transformation. Since $d_{tv}\{\mu_0, \mu_1\} \leq \epsilon$, by definition of total variation distance we have $\mid \mu_0(A_x) - \mu_1(A_x) \mid \leq \epsilon$. Noting that $\mu_i(A_x) = \hat{\mu}_i(0, x)$, $i = 0, 1$ we have $\mid \hat{\mu}_0(0, x) - \hat{\mu}_1(0, x) \mid \leq \epsilon$ and the result follows. ∎

The implication of Theorem 5.7 is as follows: if the family $\mathcal{F}_0$ is of size ϵ in terms of total variation distance and contains the uniform distribution, the domain transform preserves the size ϵ in terms of Kolmogorov distance. Then we can apply the modified FDR procedure and ensure that false discovery rate is bounded by γ as the result of Theorem 5.6 indicates.

We now extend the development of Theorem 5.7 to $\mathcal{F}_1$, the family of distributions under H_1. The proof of this extension is more involved, and is presented in the appendix.

Theorem 5.8 *Let μ_1 be the measure associated with $F_1 \in \mathcal{F}_1$ and μ_2 be the measure associated with of $F_2 \in \mathcal{F}_1$. Define $\hat{F}_1$ and $\hat{F}_2$ to be the respective distributions after the transformation. If $d_{tv}\{\mu_1, \mu_2\} \leq \epsilon$ then $\sup_x |\hat{F}_1(x) - \hat{F}_2(x)| \leq \epsilon$ where for any measurable space Ω, $d_{tv}(\mu, \nu) = \sup_{A \subset \Omega} |\mu(A) - \nu(A)|$.*

Proof: See appendix. ∎

The implication of Theorem 5.8 is that when the domain transformation is applied to an $\mathcal{F}_1$ that has no concave member, and is such that all members of $\mathcal{F}_1$ are within an ϵ ball in terms of total variation distance, the new family of functions $\hat{\mathcal{F}}_1$ carries a concave member, and is such that all members are within an ϵ ball in terms of the Kolmogorov distance.

Based on the result of Theorem 5.8, we now develop the robustness property with respect to FNR under an asymptotic regime, where $FNR = E(T/(m - R))$. This is the natural counterpart of FDR, which

represents the number of H_1s among the samples declared as H_0, and it can be interpreted as the miss rate. Let F be the concave distribution of observations under positive hypotheses such that $F'(0) > \beta$ where $\beta = (\frac{1}{\gamma} - \theta_0)/(1 - \theta_0)$. [14] show that if c is the solution to $F(x) = \beta x$, asymptotically the following holds true:

$$FNR = \frac{\theta_1(1 - F(c))}{\theta_1(1 - F(c)) + \theta_0(1 - c)}$$

.

The following theorem establishes the the robustness property with respect to FNR.

Theorem 5.9 *Let F_N be the concave nominal distribution of observations under positive hypotheses such that $F'_N(0) > \beta$ and F_A be the actual distribution of observations under positive hypotheses such that $F'_A(0) > \beta$. Let c be the solution to $F(x) = \beta x$ and u the solution to $F(x) - \epsilon = \beta x$. If $|F_N(x) - F_A(x)| \le \epsilon(x)$, then for $\xi = (F_N(c) - F_N(u))/(c - u)$,*

$$FNR_{actual} \le FNR_{nominal} + \frac{1}{(1 - \gamma)^2} \frac{\epsilon}{[1 - F'_N(\xi)/\beta]}$$

where $\epsilon = \sup_x \{\epsilon(x)\}$.

Proof: See appendix. ∎

Note that since $F_n(x)$ and $F_n(x) - \epsilon$ are both concave, $F'_N(u) < \beta$ and $F'_N(c) < \beta$. Therefore, as $\epsilon \downarrow 0$ the bound gets tighter.

The results we have developed above lead us to the following conclusion: As long as the family of distributions that are considered under the non-ideal sensing model are small in terms of the total variation distance, the distributed detection problem can be addressed within the DTFDR framework in the asymptotic regime with graceful degradation of performance with respect to the size of the families.

Similar results can be developed using Prokhorov metric where for any metric space Ω, $d_P(\mu, \nu) = \inf\{\epsilon > 0 : \mu(B) \le \nu(B^\epsilon)$ for all Borel sets $B\}$ where $B^\epsilon = \{x : \inf_{y \in B} d(x, y) \le \epsilon\}$. For this extension the Levy metric takes the place of Kolmogorov metric. Furthermore, this extension allows us to consider singular distributions as well as continuous ones for the positive hypotheses. We present this extension in the following theorem and conclude this section.

Theorem 5.10 *Let μ_1 be the measure associated with $F_1 \in \mathcal{F}_1$ and let μ_2 be the measure associated with any $F_2 \in \mathcal{F}_1$. If $d_P\{\mu_1, \mu_2\} \le \epsilon$ then $\sup_x \{\hat{F}_1(x) - \hat{F}_2(x + 2\epsilon)\} \le \epsilon$ where $\hat{F}_1(\cdot)$ and $\hat{F}_2(\cdot)$ are the distribution functions obtained by the transformation with respect to F_1.*

Proof: See appendix. ∎

6 Simulations

Below we present a detection simulation in which we use FDR and DTFDR procedures. Earlier we developed the robustness property with respect to the perturbation of the model under positive hypotheses. Here

we demonstrate the case where we assume perturbations to the model under the null hypotheses. We present the simulation results with a similar setup to the one we presented earlier.

The sensor field for the simulations is a grid of size 100x100, where each pixel is assumed to have a sensor, and the sensors observe the signal within their pixel. Let the set of targets be $T = \{t_1, t_2, \ldots, t_k\}$ with uniform signal power θ for simplicity. Define $d(s, t)$ be the Euclidian distance between sensor s and target t. We assume that the target signal is a constant θ within a certain effective region of the target, and decays rapidly with distance outside this region. The null hypotheses for a sensor is that it is outside the effective region of all targets, and the alternative is that it is inside the effective region of a target. Then, with n_s and ν_s being noise at sensor s for null and alternative hypotheses respectively, the observation model at sensor s for the non-ideal case is as follows:

$$
\begin{aligned}
H_0 &: \quad X_s = \theta'_s + n_s \\
H_1 &: \quad X_s = \theta_s + \nu_s
\end{aligned}
$$

In the ideal sensing model, $\theta'_s = 0$ and $\theta_s = \theta$. In the non-ideal sensing model, $\theta'_s \in [0, 0.1]$, $\theta_s \in [\theta - 0.1, \theta]$. The uncertainty arises from the model we described earlier in Section 5. The results demonstrate the robustness of DTFDR procedure to such non-idealities. For the simulation, the FDR threshold was set to $\gamma = .15$, $\theta = 2.8$, the effective radius of the target $r_{eff} = 2.5$ pixels. $n_s \sim N(0, 1)$ and $\nu_s \sim N(0, 0.05)$. There were 10 targets on the field. The communication constraint α was varied and the results are presented for illustrative cases in Figure 7 for the non-ideal sensing model.

In the ideal sensing model, for $\alpha \leq 150$ implementation of the FDR procedure was unable to detect the sensors with H_1 hypotheses, whereas the DTFDR procedure was able to do so. As the communication constraint was loosened, the performance of DTFDR procedure increased accordingly, yet keeping the false alarms at low levels. Although the FDR procedure also detected some sensors with H_1 hypotheses, observe that the FDR algorithm suffers from more false alarms.

In the non-ideal sensing model, note that the FDR procedure fails to detect with any amount of communication budget. This is because the ordered p values are always above their corresponding levels. However, although the exact distribution is not known under positive hypotheses, the domain transformation is performed successfully, and this allows for successful detection of significant sensors.

7 Conclusion

In this paper we developed tools for detection of localized events, sources, or abnormalities within SNETs. Unlike decentralized detection, the focus here was on problems involving localized information, where, only a small number of sensors in the vicinity of the phenomena are in the field of observation. For localized phenomena the main difficulty arises from the coupling of: **a)** limited sensing range of sensors which provide the sensors with uncorrelated information; **b)** limited energy, which constrains communication among sensor nodes; **c)** uncertainty in the number of targets and possible locations; **d)** multiplicity of false alarms. Although not evident at first sight, these fundamental difficulties call for collaboration in the SNET in order to meet global constraints.

We characterized fundamental trade offs between global performance (false alarms and miss rate) and communication cost. We developed a novel framework, Domain Transformed FDR, to minimize the expected number of misses subject to worst-case misclassification and communication constraints. This involved a transformation that is measure preserving with respect to the null distribution, and application of the FDR procedure. We developed and proposed a robust distributed algorithm to implement DTFDR, and presented optimality and scaling results. We showed that thanks to the robustness property, problems with non-ideal sensing model can be addressed in this framework. We also presented the elegant scaling result of the FDR algorithm, which has the striking implication that this strategy can be utilized efficiently no matter what the SNET size is.

A Proof of Theorem 4.10

Let $m = m_0 + m_1$ and let $\omega \in (0,1)^m$. Let $B = \{\omega : P_* \leq p, P^* > p\}$ have a nonzero measure. Outside B, S itself is a thresholding strategy and $S = S'$. Therefore we only need to consider the set B. Let $i^*(\omega)$ and $i_*(\omega)$ be the indices of P^* and P_* in the set $\mathcal{P}$. Now, consider a $B' \subset B$ in which $i^*(\omega) = i^*$ and $i_*(\omega) = i_*$ are fixed.

Define $D_S(i) = I(P_i \in \mathcal{P}_S)$, $D_{S'}(i) = I(P_i \in \mathcal{P}_{S'})$, and $A(i) = I(H^i = H_0)$. Then,

$$
\begin{aligned}
FDR_S - FDR_{S'} &= E(\frac{V_S}{R_S}|B') - E(\frac{V_{S'}}{R_{S'}}|B') \\
&= E(\frac{\sum_{i=1:m} D_S(i)A(i)}{\sum_{i=1:m} D_S(i)}|B') - E(\frac{\sum_{i=1:m} D_{S'}(i)A(i)}{\sum_{i=1:m} D_{S'}(i)}|B') \\
&= E(\frac{\sum_{i=1:m} D_S(i)A(i)}{\sum_{i=1:m} D_S(i)} - \frac{\sum_{i=1:m} D_{S'}(i)A(i)}{\sum_{i=1:m} D_{S'}(i)}|B')
\end{aligned}
$$

Now, note that the cardinality of P_S and $P_{S'}$ are the same, and $D_S(i) \neq D_{S'}(i)$ only for i^* and i_*. Therefore we can rewrite the above difference as follows:

$$
\begin{aligned}
FDR_S - FDR_{S'} &= E(\frac{\sum_{i=1:m,i\neq\{i^*,i_*\}} D_S(i)A(i) + D_S(i^*)A(i^*) + D_S(i_*)A(i_*)}{\sum_{i=1:m} D_S(i)}|B') \\
&\quad - E(\frac{\sum_{i=1:m,i\neq\{i^*,i_*\}} D_{S'}(i)A(i) - D_{S'}(i^*)A(i^*) - D_{S'}(i_*)A(i_*)}{\sum_{i=1:m} D_S(i)}|B') \\
&= E(\frac{D_S(i^*)A(i^*) - D_{S'}(i_*)A(i_*)}{\sum_{i=1:m} D_S(i)}|B') \\
&= E(\frac{A(i^*) - A(i_*)}{\sum_{i=1:m} D_S(i)}|B')
\end{aligned}
$$

$$FDR_S - FDR_{S'} \geq \frac{E(A(i^*) - A(i_*))|B')}{m}$$

$$= \frac{E(I(H^{i^*} = H_0) - I(H^{i_*} = H_0)|B')}{m}$$

$$= \frac{\Pr\{H^{i^*} = H_0|B'\} - \Pr\{H^{i_*} = H_0|B'\}}{m}$$

$$= \frac{1}{m}\left[\frac{\Pr\{H^{i^*} = H_0, B'\}}{\Pr\{B'\}} - \frac{\Pr\{H^{i_*} = H_0, B'\}}{\Pr\{B'\}}\right]$$

$$= \frac{1}{m}\left[\frac{\Pr\{B'|H^{i^*} = H_0\}\Pr\{H^{i^*} = H_0\}}{\Pr\{B'\}} - \frac{\Pr\{B'|H^{i_*} = H_0\}\Pr\{H^{i_*} = H_0\}}{\Pr\{B'\}}\right]$$

$$= \frac{1}{m}\left[\frac{\Pr\{P^* > p|H^{i^*} = H_0\}\Pr\{P_* \leq p\}\Pr\{H^{i^*} = H_0\}}{\Pr\{P^* > p\}\Pr\{P_* \leq p\}}\right.$$

$$\left. - \frac{\Pr\{P_* \leq p|H^{i_*} = H_0\}\Pr\{P^* > p\}\Pr\{H^{i_*} = H_0\}}{\Pr\{P^* > p\}\Pr\{P_* \leq p\}}\right]$$

Here observe that $\Pr\{H^{i^*} = H_0\} = \Pr\{H^{i_*} = H_0\}$ by the hypothesis of the theorem, and $\Pr\{P_* \leq p\} = 1 - \Pr\{P^* > p\}$ due to independence assumption. Therefore,

$$FDR_S - FDR_{S'} \geq \frac{\Pr\{H^{i^*} = H_0\}}{m}\left[\frac{(1-p)(1 - \Pr\{P^* > p\})}{\Pr\{P^* > p\}\Pr\{P_* \leq p\}} - \frac{p\Pr\{P^* > p\}}{\Pr\{P^* > p\}\Pr\{P_* \leq p\}}\right]$$

$$= \frac{\Pr\{H^{i^*} = H_0\}}{m\Pr\{P^* > p\}\Pr\{P_* \leq p\}}[1 - \Pr\{P^* > p\} - p + p\Pr\{P^* > p\} - p\Pr\{P^* > p\}]$$

$$= \frac{\Pr\{H^{i^*} = H_0\}}{m\Pr\{P^* > p\}\Pr\{P_* \leq p\}}[F_{P^*}(p) - p]$$

$$\geq 0$$

The last inequality follows from the fact that $f_0 = U(0,1)$ and f_1 is monotonically decreasing. ∎

B   Proof of Lemma 4.13

$$f_{X_{(i)}}(t) = \frac{n!}{(i-1)!(n-i)!}(F_X(t))^{i-1}(1 - F_X(t))^{n-i}f_X(t)$$

$$F_{X_{(i)}}(t) = \frac{n!}{(i-1)!(n-i)!}\int_0^t (F_X(x))^{i-1}(1 - F_X(x))^{n-i}f_X(x)dx$$

$$= \frac{n!}{(i-1)!(n-i)!}\int_0^{F_X(t)} u^{i-1}(1 - u)^{n-i}du$$

29

By the same approach it is easy to see that

$$
\begin{aligned}
F_{Y_{(i)}}(t) &= \frac{n!}{(i-1)!(n-i)!} \int_0^t (F_Y(y))^{i-1}(1 - F_Y(y))^{n-i} f_Y(y) dy \\
&= \frac{n!}{(i-1)!(n-i)!} \int_0^{F_Y(t)} u^{i-1}(1-u)^{n-i} du
\end{aligned}
$$

By hypothesis of the lemma, $F_X(t) \geq F_Y(t)$ $\forall t \in (0,1)$, and since $0 \leq u \leq 1$, it follows that $F_{X_{(i)}}(t) \geq F_{Y_{(i)}}(t)$ $\forall t \in (0,1)$, which concludes the proof of the lemma. ∎

C   Proof of Theorem 5.6

Proof: Again let $\gamma_k = \gamma k/m$. Then:

$$
\begin{aligned}
E(V/R) &= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \Pr\{P'_{0i} \leq \gamma_k, C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \Pr\{P'_{0i} \leq \gamma_k\} \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \Pr\{P_{0i} \leq \gamma_k - \epsilon\} \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} F_{0i}(\gamma_k - \epsilon) \Pr\{C_i(k)\} \\
&\leq \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} [(\gamma_k - \epsilon) + \epsilon] \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \sum_{k=1:m} \frac{1}{k} \frac{\gamma k}{m} \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} \sum_{k=1:m} \Pr\{C_i(k)\} \\
&= \sum_{i=1:m_0} \frac{\gamma}{m} = \frac{\gamma m_0}{m} \leq \gamma
\end{aligned}
$$

D   Proof of Theorem 5.8

For the particular case construct the new measures using $\phi(\cdot) = f_1(\cdot)$:

$$
(\alpha_{\mu_1}(y), \beta_{\mu_1}(y)) = (E_U[I_{\{f_1(x) \geq y\}}(x)], E_{\mu_1}[I_{\{f_1(x) \geq y\}}(x)])
$$

$$(\alpha_{\mu_2}(y), \beta_{\mu_2}(y)) = (E_U[I_{\{f_1(x) \geq y\}}(x)], E_{\mu_2}[I_{\{f_1(x) \geq y\}}(x)])$$

Note that $0 \leq \alpha_{\mu_1}(\cdot) = \alpha_{\mu_2}(\cdot) \leq 1$. Then,

$$\hat{F}_1(\alpha_{\mu_1}(y)) = \hat{\mu}_1(0, \alpha_{\mu_1}(y)) = \beta_{\mu_1}(y)$$

$$\hat{F}_1(\alpha_{\mu_1}(y)) = \hat{\mu}_2(0, \alpha_{\mu_1}(y)) = \beta_{\mu_2}(y)$$

By definition of total variance distance, $\sup_B |\mu_1(B) - \mu_2(B)| \leq \epsilon$, which implies

$$\sup_y |E_{\mu_1}[I_{\{f_1(x) \geq y\}}(x)] - E_{\mu_2}[I_{\{f_1(x) \geq y\}}(x)]| \leq \epsilon \Rightarrow |\beta_{\mu_1}(y) - \beta_{\mu_2}(y)| \leq \epsilon \ \forall y$$

Finally, $\sup_x |\hat{F}_1(x) - \hat{F}_2(x)| = \sup_x |\beta_{\mu_1}(y) - \beta_{\mu_2}(y)| \leq \epsilon$. ∎

E   Proof of Theorem 5.9

Define the lower bound of $F_A(x)$ as $F_L(x) = F_N(x) - \epsilon$, such that $F_L(x) \leq F_A(x) \leq F_N(x)$. Let c be the point that solves $F_N(x) = \beta x$ and u be the point that solves $F_N(x) - \epsilon = \beta x$, where $\beta = (\frac{1}{\gamma} - \theta_0)/(1 - \theta_0)$. Defining $\theta = \theta_1/\theta_0$, we can rewrite FNR as a function of decision point x as follows: $FNR = \frac{\theta \frac{1-F(x)}{1-x}}{1 + \theta \frac{1-F(x)}{1-x}}$. Then, we first start with the following ratio:

$$
\begin{aligned}
\frac{FNR_{actual} - FNR_{nominal}}{FNR_{nominal}} &= (\frac{\theta \frac{1-F_L(u)}{1-u}}{1 + \theta \frac{1-F_L(u)}{1-u}} - \frac{\theta \frac{1-F_N(c)}{1-c}}{1 + \theta \frac{1-F_N(c)}{1-c}}) \times \frac{1 + \theta \frac{1-F_N(c)}{1-c}}{\theta \frac{1-F_N(c)}{1-c}} \\
&= \frac{\theta(\frac{1-F_L(u)}{1-u} - \frac{1-F_N(c)}{1-c})}{(1 + \theta \frac{1-F_N(c)}{1-c})(1 + \theta \frac{1-F_L(u)}{1-u})} \times \frac{1 + \theta \frac{1-F_N(c)}{1-c}}{\theta \frac{1-F_N(c)}{1-c}} \\
&= \frac{(\frac{1-F_L(u)}{1-u} - \frac{1-F_N(c)}{1-c})}{(1 + \theta \frac{1-F_L(u)}{1-u})} \times \frac{1}{\frac{1-F_N(c)}{1-c}} \\
&\leq \frac{(\frac{1-F_L(u)}{1-u} - \frac{1-F_N(c)}{1-c})}{\frac{1-F_N(c)}{1-c}} \\
&= \frac{F_N(c) - F_L(u) + u - c}{(1-u)(1-F_N(c))} \\
&= \frac{F_N(c) - F_L(u) + F_L(u)/\beta - F_N(c)/\beta}{(1-u)(1-F_N(c))} \\
&= \frac{(1 - 1/\beta)(F_N(c) - F_L(u))}{(1-u)(1-F_N(c))} \\
&\leq \frac{(1 - 1/\beta)(F_N(c) - F_L(u))}{(1-c)(1-F_N(c))}
\end{aligned}
$$

Note that

$$F_N(c) - F_L(u) = \beta(c - u)$$

and therefore for some $u \leq \xi \leq c$

$$
\begin{aligned}
F_N(c) - F_L(u) &= F_N(c) - F_N(u) + F_N(u) - F_L(u) \\
&= F_N'(\xi)(c - u) + F_N(u) - F_L(u) \\
&= F_N'(\xi)\frac{F_N(c) - F_L(u)}{\beta} + F_N(u) - F_L(u)
\end{aligned}
$$

Then collecting $F_N(c) - F_L(u)$ terms together

$$(F_N(c) - F_L(u))(1 - F_N'(\xi)/\beta) = F_N(u) - F_L(u)$$

and

$$F_N(c) - F_L(u) = \frac{F_N(u) - F_L(u)}{1 - F_N'(\xi)/\beta} \leq \frac{\epsilon}{1 - (F_N'(\xi)/\beta)}$$

Then we have

$$
\begin{aligned}
\frac{FNR_{actual} - FNR_{nominal}}{FNR_{nominal}} &\leq \frac{(1 - 1/\beta)}{[1 - (F_N'(\xi)/\beta)](1 - c)(1 - F_N(c))}\epsilon \\
&= \frac{1}{[1 - (F_N'(\xi)/\beta)]\frac{1-F_N(c)}{1-c}(1-c)^2}\epsilon \\
&\leq \frac{1}{[1 - (F_N'(\xi)/\beta)]FNR_{nominal}(1-c)^2}\epsilon \\
&\leq \frac{\epsilon}{[1 - (F_N'(\xi)/\beta)]FNR_{nominal}(1-\gamma)^2}
\end{aligned}
$$

where the second inequality follows from $c \leq 1/\beta$. This is because c is the solution to $F(x) = \beta x$, and $F(x) \leq 1$. Therefore $(1 - 1/\beta)/(1 - c) \leq 1$ and the inequality follows. The fourth inequality can easily be seen from the expression of FNR. Now, multiplying both sides by $FNR_{nominal}$ and adding $FNR_{nominal}$ to both sides we get the result. ∎

F   Proof of Theorem 5.10

For the particular case construct the new measures using $\phi(\cdot) = f_1(\cdot)$. For some y let $A_y = \{x : f_1(x) \geq y\}$. Then

$$\hat{\mu}_1(0, \alpha_{\mu_1}(y)) = E_{\mu_1}[I_{\{f_1(x) \geq y\}}(x)] = E_{\mu_1}[I_{\{A_y\}}(x)]$$

Now, for some y' such that $E_U[I_{\{A_y^\epsilon\}}(x)] = E_U[I_{\{f_1(x) \geq y'\}}(x)]$,

$$\hat{\mu}_2(0, \alpha_{\mu_2}(y')) = E_{\mu_2}[I_{\{f_1(x) \geq y'\}}(x)] = \sup_{B:E_U[I_{\{B\}}(x)]=E_U[I_{\{A_y^\epsilon\}}(x)]} \{E_{\mu_2}[I_{\{B\}}(x)]\}$$

Then,

$$
\begin{aligned}
\hat{\mu}_2(0, \alpha_{\mu_2}(y')) + \epsilon - \hat{\mu}_1(0, \alpha_{\mu_1}(y)) &= \sup_{B:E_U[I_{\{B\}}(x)]=E_U[I_{\{A_y^\epsilon\}}(x)]} \{E_{\mu_2}[I_{\{B\}}(x)]\} + \epsilon - E_{\mu_1}[I_{\{A_y\}}(x)] \\
&\geq E_{\mu_2}[I_{\{A_y^\epsilon\}}(x)] + \epsilon - E_{\mu_1}[I_{\{A_y\}}(x)] \\
&\geq 0
\end{aligned}
$$

where the last inequality follows from the fact that $\mu_1(A) \leq \mu_2(A^\epsilon) + \epsilon \ \forall A$. Now, note that $\alpha_{\mu_2}(y') \leq \alpha_{\mu_2}(y) + 2\epsilon$ and thus $\hat{\mu}_2(0, \alpha_{\mu_2}(y) + 2\epsilon) + \epsilon - \hat{\mu}_1(0, \alpha_{\mu_1}(y)) \geq 0$. Since the argument holds for an arbitrary y $\sup_y\{\hat{\mu}_2(0, \alpha_{\mu_2}(y) + 2\epsilon) + \epsilon - \hat{\mu}_1(0, \alpha_{\mu_1}(y))\} \geq 0$ and the proof follows. ∎

References

[1] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Technical Report, Statistics Department, Stanford University*, 2000.

[2] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron. Distributed bayesian hypothesis testing in sensor networks. In *American Control Conference*, Boston, MA, July 2004.

[3] S. Appadwedula, V. V. Veeravalli, and D. L. Jones. Robust and locally-optimum decentralized detection with censoring sensors. In *5th Int. Conf. on Information Fusion*, 2002.

[4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 1995.

[5] Y. Benjamini and D. Yekuteli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29, 2001.

[6] T. Berger, Z. Zhang, and H. Vishwanathan. The ceo problem. *IEEE Transactions on Information Theory*, May 1996.

[7] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, 27, 1982.

[8] J.F. Chamberland and V. Veeravalli. Asymptotic results for decentralized detection in power constrained wireless sensor networks. *IEEE Special Issue on Wireless Sensor Networks*, pages 1007–1015, August 2004.

[9] D.Mosk-Aoyama and D.Shah. Computing separable functions via gossip. *ACM Symposium on Principles of Distributed Computing*, July 2006.

[10] E. Ermis and V. Saligrama. Adaptive sampling strategies for detection of localized phenomena, information processing in sensor networks. In *4th International Workshop on Information Processing in Sensor Networks*, Los Angeles, CA, 2005.

[11] E. Ermis and V. Saligrama. Detection and localization in sensor networks using fdr. In *Conference on Information Sciences and Systems*, Princeton, NJ, 2006.

[12] Editor F. A. Sadjadi. *Selected Papers on Sensor and Data Fusion*. SPIE Milestone Series, 1996.

[13] G. Fleury, A. Hero, S. Yoshida, T. Carter, C. Barlow, and A. Swaroop. Clustering gene expression signals from retinal microarray data. In *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Process.*, Orlando, Fl, May 2002.

[14] C. Genovese and L Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society, Series B*, 64:479–498, 2002.

[15] T. Kasetkasem and P. K. Varshney. Communication structure planning for multisensor detection systems. *IEE Proc. Radar, Sonar Navig.*, 48:2–8, 2001.

[16] E. L. Lehmann. *Testing Statistical Hypothesis*. John Wiley and Sons, 1986.

[17] M. Longo, T. D. Lookabaugh, and Robert M. Gray. Quantization for decentralized hypothesis testing under communication constraints. *IEEE Trans. on Inform. Theory*, 36:241–255, 1990.

[18] M. Alanyali O. Savas, V. Saligrama. Randomized sequential algorithms for data aggregation in sensor networks. *DCOSS, San Francisco*, 2006.

[19] N. Patwari and A. Hero. Reducing transmissions from wireless sensors in distributed detection networks using hierarchical censoring. In *ICASSP*. IEEE, 2003.

[20] C. Rago, P. Willett, and Y. Bar-Shalom. Censoring sensors: a low-communication-rate scheme for distributed detection. *IEEE Trans. Aerosp. Electron. Syst.*, 32:554–568, 1996.

[21] R.J.Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrica*, 73:751–754, 1986.

[22] S. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, 1983.

[23] R.J. Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73:751–754, 1986.

[24] J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64:499–517, 2002.

[25] P.F. Swaszek and P. Willett. Parley as an approach to distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, pages 447–457, January 1995.

[26] J. N. Tsitsiklis. Decentralized detection. *in Advances in Statistical Signal Processing, H. V. Poor and J. B. Thomas Eds*, 2, 1993.

[27] V. Saligrama, M. Alanyali, O. Savas. Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Transactions on Signal Processing*, November 2006.

[28] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer, 1997.

[29] William C. Karl Venkatesh Saligrama, Yonggang Shi. Performance guarantees in sensor networks. *IEEE Int. Conf. on Acoust., Speech, and Sig. Process.*, 2004.
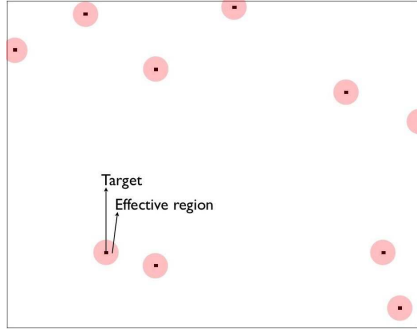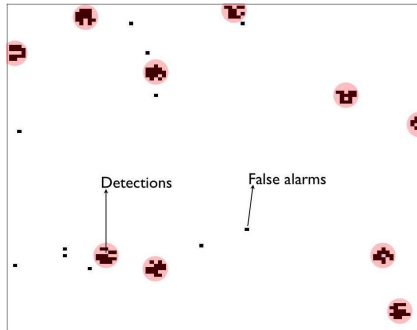
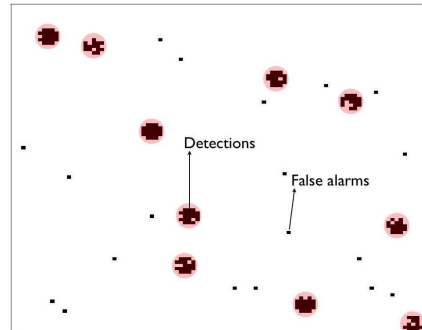(a) Target Locations
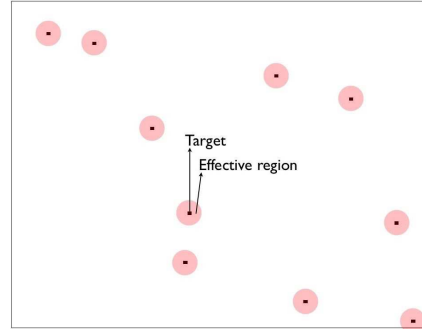
(b) Target Locations

(c) Detection with FDR

(d) Detection with FDR

(e) Detection with DTFDR

(f) Detection with DTFDR

Figure 7: Detection performance of distributed implementations under non ideal sensing model for $\alpha = 150$ bits (a,c,e), $\alpha = 200$ bits (b,d,f). (A purely gray plot indicates that no detection was made.)