Behavioral Utility-Based Distributed Detection in the Presence of Byzantines

Ahsan Yousaf, Berkan Dulek, and Sinan Gezici

***Abstract*—In this work, we consider a distributed detection system for a binary hypothesis testing problem with conditionally independent observations at the local decision agents (DAs) that are binary quantized and transmitted to a fusion center (FC). Some of the DAs act as malicious attackers, called Byzantines, with the intention of degrading the overall system performance. Unlike classical distributed detection, the FC is modeled as a human operator whose decisions reflect the nonlinear probability distortions and asymmetric gain-loss valuations as characterized by prospect theory. The FC employs a fixed fusion rule, while Byzantine DAs choose their local decision rules to minimize the FC's behavioral utility and honest DAs choose theirs to maximize it. This leads to a Stackelberg formulation in which the FC commits to its fusion rule, the Byzantine nodes respond adversarially, and the honest nodes optimize accordingly. The optimal Byzantine strategy is shown to exist and characterized to be a randomization between two single-threshold likelihood ratio-based decision rules.**

***Index Terms*—Distributed detection, behavioral utility, likelihood ratio test, prospect theory, Byzantine attacks.**

I. Introduction

Distributed detection in sensor networks has been a central topic in statistical signal processing, with a rich body of literature establishing optimal strategies for local decision making and data fusion [1], [2], [3], [4]. A practical challenge that has received considerable attention in recent literature is the presence of Byzantine sensors, namely compromised, malicious, or unreliable sensing nodes that may falsify their local decisions, intentionally or otherwise, thereby undermining overall system performance [5], [6], [7], [8], [9], [10], [11]. For a binary hypothesis testing setup, it has been shown that when the fraction of Byzantine nodes, or decision agents (DAs), exceeds a critical threshold, the detection scheme at the fusion center (FC) can be rendered completely uninformative, in the sense that the Kullback-Leibler divergence between the distributions of the received data under the two hypotheses vanishes [5], [8].

Received 4 June 2026; revised 10 July 2026; accepted 11 July 2026. Date of publication 20 July 2026; date of current version 7 August 2026. This work was supported by Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 122E493. The associate editor coordinating the review of this article and approving it for publication was Prof. Xuebo Zhang. *(Corresponding author: Sinan Gezici.)*

Ahsan Yousaf and Sinan Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Türkiye (e-mail: ahsan.yousaf@bilkent.edu.tr; gezici@ee.bilkent.edu.tr).

Berkan Dulek is with the Department of Electrical and Electronics Engineering, Hacettepe University, 06800 Ankara, Türkiye (e-mail: berkan@ee.hacettepe.edu.tr).

This article has supplementary downloadable material available at https://doi.org/10.1109/LSP.2026.3715140, provided by the authors.

Digital Object Identifier 10.1109/LSP.2026.3715140

The strategic interaction between the FC and the adversary has also been studied through a Stackelberg game formulation, where the FC acts as the leader committing to a defensive strategy and the Byzantines respond optimally [8]. Using the notion of blinding probability, [7] showed that uniform falsification can render the fusion center blind, although this requires compromising a very large portion of the sensors. To counteract this drawback, [9] instead analyzed the problem at the individual sensor level and demonstrated that, when attackers possess local output statistics and operate under symmetric falsification, an optimal strategy exists that achieves blinding with minimal compromise probability. Recent work in [11] investigates dynamic Byzantine attackers in reconfigurable intelligent surface (RIS)-enhanced cooperative spectrum sensing, where attack strategies are derived under both optimal and suboptimal fusion rules, with all cases yielding identical optimal strategies such that in practice no knowledge of the fusion rule is required.

Classical detection frameworks typically model the FC as a rational decision maker operating under the Bayesian risk criterion. However, when the FC is a human operator, empirical evidence suggests that decisions systematically deviate from this ideal, motivating the use of Prospect Theory (PT) to account for cognitive biases such as probability distortion and loss aversion [12], [13]. Common biases include the overweighting of small probability events and the underweighting of large probability events [14]. Recent work has incorporated these behavioral considerations into the distributed detection setting under fully cooperative DAs [15], [16], [17], [18]. In particular, [16] introduced a behavioral utility-based detection criterion by incorporating key elements of PT to capture several characteristic tendencies of human decision making. The optimal solution for their M-ary hypothesis testing setup involved randomization among likelihood-ratio quantizers. Combining Byzantine attacks with behavioral detection is nontrivial because the prospect-theoretic objective differs from conventional detection criteria.

In this letter, a subset of DAs minimizes the FC's prospect-theoretic utility, while the honest DAs maximize it under a fixed fusion rule. In contrast to classical Byzantine models that specify explicit bit-flipping and additional constraints to model the Byzantine nodes [8], we adopt the following structure: the FC employs a fixed fusion rule, while a subset of DAs act adversarially by choosing their local decision rules to *minimize* the FC's behavioral utility. The remaining (honest) DAs choose their local decision rules to *maximize* the FC's behavioral utility. This yields a min-max (Stackelberg) formulation over the local operating points of the DAs under a fixed fusion rule where the Byzantine nodes move first and the honest nodes move in response. We establish existence of an optimal Byzantine strategy and show that it can be implemented by randomizing between at most two single-threshold likelihood-ratio rules. Numerical examples illustrate its performance impact.

II. Problem Formulation

A. System Model and Behavioral Utility

Consider a distributed detection network consisting of N sensors (local DAs) indexed by $\mathcal{N} \triangleq \{1, \ldots, N\}$ and a single FC. The local DAs observe a phenomenon of interest (POI) modeled by a binary hypothesis testing problem with states $\mathcal{H}_0$ (null hypothesis) and $\mathcal{H}_1$ (alternative hypothesis), having prior probabilities π_0 and π_1, respectively [1].

The n-th local DA observes a random variable $Y_n \in \mathcal{Y}_n$. Conditioned on $\mathcal{H}_j$, local DAs' observations $\boldsymbol{Y} \triangleq (Y_1, Y_2, \ldots, Y_N)$ are independent and their joint probability density function (PDF) $f(\boldsymbol{y} \mid \mathcal{H}_j)$ is given by

$$f(\boldsymbol{y} \mid \mathcal{H}_j) = \prod_{n=1}^{N} f_n(y_n \mid \mathcal{H}_j), \qquad j \in \{0, 1\}, \tag{1}$$

where $f_n(y_n \mid \mathcal{H}_j)$ denotes the PDF of Y_n under $\mathcal{H}_j$. The n-th local DA observes $Y_n = y_n$ and generates a binary message $Q_n \in \{0, 1\}$ according to a (possibly randomized) local decision rule, which is specified by a measurable function $\delta_n : \mathcal{Y}_n \to [0, 1]$ defined as

$$\delta_n(y_n) \triangleq \Pr(Q_n = 1 \mid Y_n = y_n). \tag{2}$$

That is, having observed $Y_n = y_n$, the n-th local DA outputs $Q_n = 1$ with probability $\delta_n(y_n)$ and $Q_n = 0$ with probability $1 - \delta_n(y_n)$. The local DAs transmit their binary messages, denoted by $\mathbf{Q} = (Q_1, \ldots, Q_N)$, to the FC over error-free channels. Then, the FC applies a (possibly randomized) fusion rule $\delta_0 : \{0, 1\}^N \to [0, 1]$ based on a given realization of the received message vector, represented by $\mathbf{q} = (q_1, \ldots, q_N)$, and produces the final decision by selecting $\mathcal{H}_1$ (i.e., $Q_0 = 1$) with probability $\delta_0(\mathbf{q})$ and $\mathcal{H}_0$ (i.e., $Q_0 = 0$) with probability $1 - \delta_0(\mathbf{q})$. In this work, the fusion rule δ_0 is assumed fixed, and the strategic interaction concerns only the local decision rules.

The FC is modeled as a human decision maker whose overall performance is measured by the following behavioral utility-based criterion [16]:

$$\mathcal{U} \; = \; \sum_{i=0}^{1} \sum_{j=0}^{1} w(\pi_j \; \Pr(Q_0 = i \mid \mathcal{H}_j)) \; v(u_{i,j}), \tag{3}$$

where $u_{i,j}$ is the outcome utility of deciding $\mathcal{H}_i$ when $\mathcal{H}_j$ is true, $w(\cdot)$ is the probability weighting function, and $v(\cdot)$ is the value function computed as follows:

$$w(p) = \frac{p^\theta}{(p^\theta + (1-p)^\theta)^{1/\theta}}, \qquad \theta > 0, \tag{4}$$

$$v(u) = \begin{cases} u^\tau, & u \geq 0, \\ -\beta(-u)^\tau, & u < 0, \end{cases} \qquad \tau \in (0, 1], \; \beta > 0. \tag{5}$$

Here, θ, τ, and β model probability distortion, diminishing sensitivity, and loss aversion, respectively [13]; they are fixed user-specific parameters.

B. Adversarial Min-Max Problem Formulation

A subset of the DAs is honest and cooperative, whereas the complementary subset is Byzantine, i.e., adversarial. It is assumed that all DAs possess full knowledge of the underlying observation model, the prior probabilities, and the fixed fusion rule employed by the FC. Honest and Byzantine DAs cooperate within themselves but differ in their objectives. Let

$$\mathcal{N}_H \subseteq \mathcal{N} \; \text{ denote the set of honest DAs,} \tag{6}$$

$$\mathcal{B} \triangleq \mathcal{N} \setminus \mathcal{N}_H \; \text{ denote the set of Byzantine DAs,} \tag{7}$$

with $|\mathcal{B}| = B$. Unlike formulations that parameterize Byzantine behavior through explicit falsification or bit-flipping [9], we model Byzantine behavior directly through the choice of local decision rules: Byzantine DAs select $\{\delta_n\}_{n \in \mathcal{B}}$ to *minimize* the behavioral utility $\mathcal{U}$, while honest DAs select $\{\delta_n\}_{n \in \mathcal{N}_H}$ to *maximize* $\mathcal{U}$, under the fixed fusion rule δ_0. Let $\boldsymbol{\delta}_{\mathcal{N}_H} \triangleq \{\delta_n\}_{n \in \mathcal{N}_H}$ and $\boldsymbol{\delta}_{\mathcal{B}} \triangleq \{\delta_n\}_{n \in \mathcal{B}}$. We define the resulting utility under the fixed fusion rule as $\mathcal{U}(\boldsymbol{\delta}_{\mathcal{N}_H}, \boldsymbol{\delta}_{\mathcal{B}})$. Let Δ_n denote the admissible class of (possibly randomized) binary decision rules at DA n. In the Stackelberg (attacker-first) min-max formulation, the Byzantine DAs move first:

$$V \; \triangleq \; \min_{\boldsymbol{\delta}_{\mathcal{B}} \in \Delta_{\mathcal{B}}} \; \max_{\boldsymbol{\delta}_{\mathcal{N}_H} \in \Delta_{\mathcal{N}_H}} \; \mathcal{U}(\boldsymbol{\delta}_{\mathcal{N}_H}, \boldsymbol{\delta}_{\mathcal{B}}), \tag{8}$$

where $\Delta_{\mathcal{N}_H}$ and $\Delta_{\mathcal{B}}$ denote the admissible sets of local decision rules for honest and Byzantine DAs, respectively. A Stackelberg solution is any pair $(\boldsymbol{\delta}^\star_{\mathcal{B}}, \boldsymbol{\delta}^\star_{\mathcal{N}_H})$ satisfying [19]

$$\boldsymbol{\delta}^\star_{\mathcal{B}} \in \arg \min_{\boldsymbol{\delta}_{\mathcal{B}} \in \Delta_{\mathcal{B}}} \; \max_{\boldsymbol{\delta}_{\mathcal{N}_H} \in \Delta_{\mathcal{N}_H}} \; \mathcal{U}(\boldsymbol{\delta}_{\mathcal{N}_H}, \boldsymbol{\delta}_{\mathcal{B}}), \tag{9}$$

$$\boldsymbol{\delta}^\star_{\mathcal{N}_H} \in \arg \max_{\boldsymbol{\delta}_{\mathcal{N}_H} \in \Delta_{\mathcal{N}_H}} \; \mathcal{U}(\boldsymbol{\delta}_{\mathcal{N}_H}, \boldsymbol{\delta}^\star_{\mathcal{B}}). \tag{10}$$

III. Problem Geometry and Optimal Byzantine Strategy

A. Local Decision Rules and Achievable Operating Points

For each DA $n \in \mathcal{N}$, we define its pairwise operating point as

$$\mathbf{p}_n(\delta_n) \; \triangleq \; \begin{bmatrix} P_{\text{FA}}^{(n)} \\ P_{\text{D}}^{(n)} \end{bmatrix} \; = \; \begin{bmatrix} \Pr(Q_n = 1 \mid \mathcal{H}_0) \\ \Pr(Q_n = 1 \mid \mathcal{H}_1) \end{bmatrix}. \tag{11}$$

Under conditional independence, the distribution of $\mathbf{Q}$ (the vector of all DAs' binary messages) under each hypothesis depends on the local decision rules $\{\delta_n\}_{n=1}^N$ only through the pairwise probabilities $\{(P_{\text{FA}}^{(n)}, P_{\text{D}}^{(n)})\}_{n=1}^N$. Hence, for a fixed fusion rule δ_0, optimizing over decision rules is equivalent to optimizing over the corresponding set of achievable operating points [16]. The set of achievable operating points for the n-th DA is

$$\mathcal{P}_n \; \triangleq \; \{\mathbf{p}_n(\delta_n) : \delta_n \in \Delta_n\} \subset [0, 1]^2, \tag{12}$$

where Δ_n denotes the admissible class of (possibly randomized) binary decision rules at the n-th DA. The set $\mathcal{P}_n$ is compact and convex [16, Lemma 1]. Consequently, any finite Cartesian product of these sets is compact and convex [20].

Under binary hypotheses and binary decisions at the local DAs, $\mathcal{P}_n$ in (12) can be viewed as the compact convex set lying between the ROC curve of Neyman-Pearson (NP) rule and that of the "flipped" NP (FNP) rule. Let the likelihood ratio be defined as $\mathcal{L}(y_n) \triangleq f_n(y_n \mid \mathcal{H}_1)/f_n(y_n \mid \mathcal{H}_0)$. Then, for a target false-alarm level $\alpha_n \in [0, 1]$, a (possibly randomized) NP rule of size α_n is given by

$$\delta_n^{\text{NP}}(y_n) = \begin{cases} 1, & \mathcal{L}(y_n) > \eta_n, \\ \rho_n, & \mathcal{L}(y_n) = \eta_n, \\ 0, & \mathcal{L}(y_n) < \eta_n, \end{cases} \tag{13}$$

where $\eta_n \geq 0$ and $\rho_n \in [0,1]$ are chosen so that $P_{\text{FA}}^{(n)}(\delta_n^{\text{NP}}) = \alpha_n$ [21]. A (possibly randomized) FNP rule of the same size α_n is expressed as

$$\delta_n^{\text{FNP}}(y_n) = \begin{cases} 1, & \mathcal{L}(y_n) < \zeta_n, \\ \gamma_n, & \mathcal{L}(y_n) = \zeta_n, \\ 0, & \mathcal{L}(y_n) > \zeta_n, \end{cases} \tag{14}$$

where $\zeta_n \geq 0$ and $\gamma_n \in [0,1]$ are chosen to satisfy $P_{\text{FA}}^{(n)}(\delta_n^{\text{FNP}}) = \alpha_n$ [16, Lemma 2]. Moreover, interior points can be achieved via randomization. In particular, to traverse the interior, each DA randomizes between at most two single-threshold likelihood-ratio quantizers: one NP rule and one FNP rule operating at the same false-alarm probability [16]. Hence, any interior operating point on the vertical line segment between the corresponding ROC boundary points can be achieved by

$$\delta_n \ = \ \lambda_n \delta_n^{\text{NP}} + (1-\lambda_n)\delta_n^{\text{FNP}}, \qquad \lambda_n \in [0,1], \tag{15}$$

and the induced pairwise probabilities satisfy $\mathbf{p}_n(\delta_n) = \lambda_n \mathbf{p}_n(\delta_n^{\text{NP}}) + (1-\lambda_n)\mathbf{p}_n(\delta_n^{\text{FNP}})$. The randomization/time-sharing operation governed by the parameter λ_n can be realized as follows. For a given observation y_n, the n-th local DA randomly selects the α_n-level NP detector with probability λ_n or the α_n-level FNP detector with probability $1-\lambda_n$. Alternatively, for a distributed detection task spanning T time units, the NP detector is employed for a period of duration $\lambda_n T$ and the FNP detector is used in the remaining time. The values of λ_n and α_n in (15) are unique for any given feasible operating point in $\mathcal{P}_n$.

B. Existence and Characterization of the Optimal Byzantine Strategy

For a given fusion rule δ_0, we define the product achievable sets as

$$\mathcal{P}_{\mathcal{N}_H} \triangleq \prod_{n\in\mathcal{N}_H} \mathcal{P}_n, \qquad \mathcal{P}_{\mathcal{B}} \triangleq \prod_{n\in\mathcal{B}} \mathcal{P}_n. \tag{16}$$

For notational convenience, we also define the corresponding vectors of honest and Byzantine DAs' operating points, respectively, as

$$\mathbf{p}_{\mathcal{N}_H} \triangleq (\mathbf{p}_n)_{n\in\mathcal{N}_H} \in \mathcal{P}_{\mathcal{N}_H}, \quad \mathbf{p}_{\mathcal{B}} \triangleq (\mathbf{p}_n)_{n\in\mathcal{B}} \in \mathcal{P}_{\mathcal{B}}. \tag{17}$$

Under conditional independence, $\Pr(Q_0 = i \mid \mathcal{H}_j)$ is a finite sum of products of the local probabilities, hence it is continuous with respect to $(\mathbf{p}_1, \ldots, \mathbf{p}_N) \in [0,1]^{2N}$. Since $w(\cdot)$ and $v(\cdot)$ are also continuous, the behavioral utility $\mathcal{U}(\mathbf{p}_{\mathcal{N}_H}, \mathbf{p}_{\mathcal{B}})$ is continuous on $\mathcal{P}_{\mathcal{N}_H} \times \mathcal{P}_{\mathcal{B}}$. Define the honest-response behavioral utility function (inner problem of (8)) as

$$J(\mathbf{p}_{\mathcal{B}}) \ \triangleq \ \max_{\mathbf{p}_{\mathcal{N}_H}\in\mathcal{P}_{\mathcal{N}_H}} \mathcal{U}(\mathbf{p}_{\mathcal{N}_H}, \mathbf{p}_{\mathcal{B}}), \quad \mathbf{p}_{\mathcal{B}} \in \mathcal{P}_{\mathcal{B}}. \tag{18}$$

Then we have the following lemma that provides a result for the existence of a solution to (18).

Lemma 1: For every fixed $\mathbf{p}_{\mathcal{B}} \in \mathcal{P}_{\mathcal{B}}$, the maximum in (18) is attained and the argmax set is nonempty. Moreover, J is continuous on $\mathcal{P}_{\mathcal{B}}$.

Proof: Fix $\mathbf{p}_{\mathcal{B}}$. The map $\mathbf{p}_{\mathcal{N}_H} \mapsto \mathcal{U}(\mathbf{p}_{\mathcal{N}_H}, \mathbf{p}_{\mathcal{B}})$ is continuous on the compact set $\mathcal{P}_{\mathcal{N}_H}$, hence it attains its maximum by the extreme value theorem. The continuity of J follows from the Maximum theorem [22, p. 570], since $\mathcal{P}_{\mathcal{N}_H}$ is compact and independent of $\mathbf{p}_{\mathcal{B}}$, and $\mathcal{U}$ is continuous. ∎

Given that the inner optimization problem has a solution, we can now consider the outer problem which concerns the Byzantine attack strategy (see (8)). The following theorem provides an existence result for the optimal Byzantine strategy and characterizes its structure.

Theorem 1: There exists a $\mathbf{p}_{\mathcal{B}}^\star \in \mathcal{P}_{\mathcal{B}}$ such that

$$V \ = \ \min_{\mathbf{p}_{\mathcal{B}}\in\mathcal{P}_{\mathcal{B}}} J(\mathbf{p}_{\mathcal{B}}) \ = \ J(\mathbf{p}_{\mathcal{B}}^\star). \tag{19}$$

Furthermore, there exists an honest best response $\mathbf{p}_{\mathcal{N}_H}^\star \in \mathcal{P}_{\mathcal{N}_H}$ attaining the inner maximum at $\mathbf{p}_{\mathcal{B}}^\star$. Moreover, each DA operating point in $\mathbf{p}_{\mathcal{N}_H}^\star$ and $\mathbf{p}_{\mathcal{B}}^\star$ is achievable by randomizing between at most two single-threshold likelihood-ratio tests as given in (15).

Proof: By Lemma 1, J is continuous on the compact set $\mathcal{P}_{\mathcal{B}}$, hence it attains its minimum due to the extreme value theorem, proving existence of $\mathbf{p}_{\mathcal{B}}^\star$. Existence of $\mathbf{p}_{\mathcal{N}_H}^\star$ follows from attainment of the inner maximization which was shown in Lemma 1. The achievability claim follows from the characterization established in the preceding subsection, that is every operating point in $\mathcal{P}_n$ is achieved by a randomized rule of the form (15), with common false-alarm probability α_n and mixing coefficient λ_n. Therefore, each coordinate of $\mathbf{p}_{\mathcal{N}_H}^\star$ and $\mathbf{p}_{\mathcal{B}}^\star$ is implementable by such a rule. For points lying on the NP or FNP boundary, a degenerate randomization suffices, and one may take $\lambda_n \in \{0,1\}$ accordingly. ∎

This completes the theoretical solution of the proposed behavioral utility-based min-max problem, establishing both the existence of optimal strategies and their implementability via LRT-based decision rules. The optimal operating points can be found by searching over only two parameters at each local DA (e.g., via a fine grid search), namely the common false-alarm level α_n, which is shared by the NP and FNP rules, and the time-sharing parameter λ_n. This alleviates the need for an exhaustive search over all possible decision rules .

IV. Numerical Results and Conclusion

In this section, numerical examples are presented for a distributed binary hypothesis testing problem with $N=3$ local DAs, one of which is Byzantine (i.e., $B=1$ and $|\mathcal{N}_H| = 2$), a fixed k-out-of-N fusion rule at the FC, and equal priors $\pi_0 = \pi_1 = 0.5$. The FC rule declares $\mathcal{H}_1$ whenever at least $k=2$ local DAs report a local decision in favor of $\mathcal{H}_1$. Mathematically, the deterministic fusion rule is given by

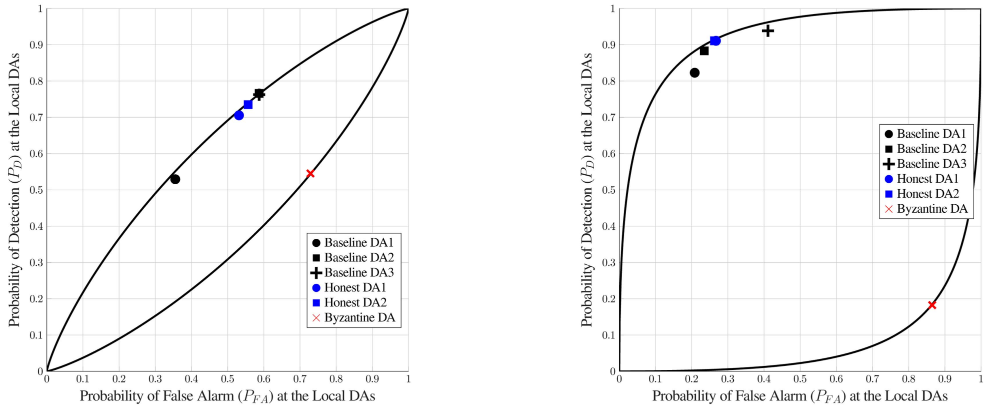
$$q_0 \ = \ \delta_0^{(k)}(\mathbf{q}) \ \triangleq \ \mathbb{I}\left\{ \sum_{n=1}^{N} q_n \geq k \right\}. \tag{20}$$

To compute the FC decision probabilities $\Pr(Q_0 = 1 \mid \mathcal{H}_j)$ required for the behavioral utility, we evaluate the Poisson-binomial tail probability. Letting $p_0^{(n)} = P_{\text{FA}}^{(n)}$ and $p_1^{(n)} = P_{\text{D}}^{(n)}$, we have

$$\Pr(Q_0 = 1 \mid \mathcal{H}_j) = \sum_{\substack{S\subseteq\mathcal{N}:\\ |S|\geq k}} \left(\prod_{n\in S} p_j^{(n)} \right) \left(\prod_{m\in\mathcal{N}\setminus S} (1 - p_j^{(m)}) \right) \tag{21}$$

where S denotes the subset of DAs transmitting 1. Each local operating point, whether chosen by an honest or a Byzantine DA, lies in its corresponding set of achievable points, $\mathcal{P}_n$, induced by the Gaussian observation model

$$\mathcal{H}_0 : y_n \sim \mathcal{N}(0, \sigma^2), \quad \mathcal{H}_1 : y_n \sim \mathcal{N}(d, \sigma^2), \tag{22}$$



(a) $d\!=\!0.5$ (b) $d\!=\!2.0$

Fig. 1. Operating points of the local DAs. The baseline case with three honest DAs is marked along with the adversarial setting with two honest DAs and one Byzantine DA. (a) $d = 0.5$. (b) $d = 2.0$.

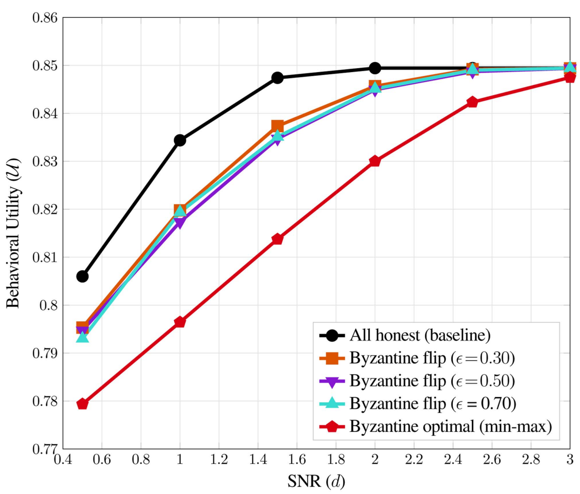


Fig. 2. Behavioral utility versus SNR under different attack scenarios: an all-honest baseline, a single Byzantine node employing a symmetric flipping strategy at three different levels ($\epsilon = 0.30, 0.50, 0.70$), and a single Byzantine node employing the optimal min-max strategy.

where d controls the local SNR. Unless stated otherwise, $d \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$, and the prospect-theoretic parameters defined in (4) and (5) are set to $\tau = 0.88$, $\beta = 2.25$, and $\theta = 0.70$. Also, the utility values in (3) are given by $(u_{00}, u_{01}, u_{10}, u_{11}) = (0.80, 0.25, 0.45, 0.95)$.

The different operating points are shown in Fig. 1 for $d \in \{0.5, 2.0\}$. The baseline case is when all three DAs are cooperative, which yields the largest behavioral utility. In the adversarial case, the two honest DAs operate at distinct points when $d = 0.5$, whereas they coincide at the same operating point when $d = 2.0$. In both cases, the Byzantine point is close to the lower FNP rule boundary but does not sit exactly on the boundary. The exact local decision rules for the adversarial setting when $d = 2.0$ are given below. An optimal local decision rule (with false-alarm rate 0.2652) at each honest DA is given by

$$\delta_s = 0.9955\, \delta_s^{\mathrm{NP}} + 0.0045\, \delta_s^{\mathrm{FNP}}, \tag{23}$$

where

$$\delta_s^{\mathrm{NP}}(y_s) = \begin{cases} 1, & \text{if } \mathcal{L}(y_s) \geq 0.4747, \\ 0, & \text{if } \mathcal{L}(y_s) < 0.4747, \end{cases} \tag{24}$$

$$\delta_s^{\mathrm{FNP}}(y_s) = \begin{cases} 1, & \text{if } \mathcal{L}(y_s) \leq 0.0386, \\ 0, & \text{if } \mathcal{L}(y_s) > 0.0386. \end{cases} \tag{25}$$

for $s \in \{1, 2\}$ since the honest DAs in this case are operating with identical rules. The Byzantine DA adopts the operating point $(P_{\mathrm{FA}}^{(3)}, P_{\mathrm{D}}^{(3)}) = (0.8629, 0.1825)$ which is almost exactly at the lower boundary. Its decision rule is given by

$$\delta_3 = 0.0002\, \delta_3^{\mathrm{NP}} + 0.9998\, \delta_3^{\mathrm{FNP}}, \tag{26}$$

where

$$\delta_3^{\mathrm{NP}}(y_3) = \begin{cases} 1, & \text{if } \mathcal{L}(y_3) \geq 0.0152, \\ 0, & \text{if } \mathcal{L}(y_3) < 0.0152, \end{cases} \tag{27}$$

$$\delta_3^{\mathrm{FNP}}(y_3) = \begin{cases} 1, & \text{if } \mathcal{L}(y_3) \leq 1.2056, \\ 0, & \text{if } \mathcal{L}(y_3) > 1.2056. \end{cases} \tag{28}$$

Note that although in this case the honest DAs are choosing the same operating point, forcing them to pick identical rules typically degrades the overall performance [16]. Also in the baseline case all three points are further away from the boundary but the presence of a Byzantine node forces the honest nodes to move closer to the upper NP boundary.

In Fig. 2, we plot the behavioral utility as a function of the SNR level. To benchmark the proposed approach against traditional methods commonly used in the distributed detection literature involving adversarial agents, we include two reference cases. First, we consider the all-honest scenario, in which all three DAs are cooperative and no Byzantine DA is present, thereby providing a performance upper bound. Second, we compare the proposed strategy with the conventional Byzantine flipping attack model, in which a Byzantine node flips its local decision with a prescribed probability. Depending on whether the flipping probabilities are symmetric or asymmetric, this model is equivalent to transmission through a binary symmetric channel (BSC) or binary asymmetric channel (BAC), respectively. Thus, while formulated as an adversarial attack, this specific benchmark concurrently provides insight into how channel noise on an individual link degrades the system's behavioral utility. For the present numerical setup, the proposed min-max optimal Byzantine strategy is significantly more effective in degrading the performance than the conventional flipping attack strategy ($\epsilon = 0.30, 0.50, 0.70$).

References

[1] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, vol. 2, H. V. Poor and B. Thomas, Eds., London, U.K.: JAI Press, 1993, pp. 297–344.

[2] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES- 22, no. 1, pp. 98–101, Jan. 1986.

[3] I. Hoballah and P. Varshney, "Distributed bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 995–1000, Sep. 1989.

[4] S. Barbarossa, G. Scutari, and A. Swami, "Distributed detection and estimation in decentralized sensor networks: An overview," in *Proc. 14th Eur. Signal Process. Conf.*, 2006, pp. 1–5.

[5] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16–29, Jan. 2009.

[6] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Distributed detection in tree topologies with Byzantines," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3208–3219, Jun. 2014.

[7] V. S. S. Nadendla, Y. S. Han, and P. K. Varshney, "Distributed inference with M-ary quantized data in the presence of Byzantine attacks," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2681–2695, May 2014.

[8] B. Kailkhura, S. Brahma, B. Dulek, Y. S. Han, and P. K. Varshney, "Distributed detection in tree networks: Byzantines and mitigation techniques," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 7, pp. 1499–1512, Jul. 2015.

[9] H.-Y. Lin, P.-N. Chen, Y. S. Han, and P. K. Varshney, "Minimum Byzantine effort for blinding distributed detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 647–661, 2020.

[10] G. Zhang et al., "An information-theoretic approach to distributed detection for mobile wireless sensor networks under Byzantine attack in entirely unknown or complicated environment: Design, analysis, and evaluation of the attack strategy," *IEEE Internet Things J.*, vol. 12, no. 4, pp. 3689–3706, Feb. 2025.

[11] G. Zhang et al., "Byzantine attacks in RIS-enhanced cooperative spectrum sensing: A decision fusion perspective," 2025, *arXiv:2510.27175*.

[12] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[13] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[14] D. Kahneman, *Thinking, Fast and Slow*. New York, NY, USA: Farrar, Straus and Giroux, 2011.

[15] S. Gezici and P. K. Varshney, "On the optimality of likelihood ratio test for prospect theory-based binary hypothesis testing," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1845–1849, Dec. 2018.

[16] B. Dulek, E. Efendi, and P. K. Varshney, "Behavioral utility-based distributed detection with conditionally independent observations," *IEEE Trans. Signal Process.*, vol. 72, pp. 3717–3730, 2024.

[17] B. Dulek, C. Ozturk, and S. Gezici, "Optimal decision rules for simple hypothesis testing under general criterion involving error probabilities," *IEEE Signal Process. Lett.*, vol. 27, pp. 261–265, 2020.

[18] A. Berber and B. Dulek, "Optimal binary hypothesis testing based on the behavioral Kullback–Leibler divergence criterion," *IEEE Signal Process. Lett.*, vol. 33, pp. 161–165, 2026.

[19] T. Basar and G. J. Olsder, "Dynamic noncooperative game theory," in *Classics in Applied Mathematics*, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1999.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K: Cambridge Univ. Press, 2004.

[21] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Berlin, Germany: Springer, 1994.

[22] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Berlin, Germany: Springer, 2006.