Enhanced Audit Bit Based Distributed Bayesian Detection in the Presence of Strategic Attacks

Chen Quan, Baocheng Geng, Yunghsiang S. Han, *Fellow, IEEE* and Pramod K. Varshney, *Life Fellow, IEEE*

***Abstract*—This paper employs an audit bit based mechanism to mitigate the effect of Byzantine attacks on distributed Bayesian detection systems. In this framework, the optimal attacking strategy for strategic attackers is investigated for the traditional audit bit based scheme (TAS) to evaluate the robustness of the system. We show that it is possible for a strategic attacker to degrade the performance of TAS to the system without audit bits. To enhance the robustness of the system in the presence of strategic attackers, we propose an enhanced audit bit based scheme (EAS). The optimal fusion rule for the proposed scheme is derived and the detection performance of the system is evaluated via the probability of error for the system. Simulation results show that the proposed EAS improves the robustness and the detection performance of the system. Moreover, based on EAS, another new scheme called the reduced audit bit based scheme (RAS) is proposed which further improves system performance. We derive the new optimal fusion rule and the simulation results show that RAS outperforms EAS and TAS in terms of both robustness and detection performance of the system. Then, we extend the proposed RAS for a wide-area cluster based distributed wireless sensor networks (CWSNs). Simulation results show that the proposed RAS significantly reduces the communication overhead between the sensors and the FC, which prolongs the lifetime of the network.**

I. Introduction

Distributed detection in wireless sensor networks (WSNs) has been studied over the last few decades [1], [2]. In distributed WSNs, instead of sending raw observations, the sensors send their quantized observations or their hard/soft decisions regarding the presence or absence of the phenomenon of interest (PoI) to the fusion center (FC) to make the final decision. This distributed framework is attractive for sensor networks that employ battery-limited sensors in bandwidth-limited environments. Because of the advantages of the distributed mechanism, it is widely used in many applications, such as IoT, cognitive radio networks, object detection networks, distributed spectrum sensing and military surveillance systems [3].

Security is an important issue for the distributed WSNs. The openness of the wireless networks and the distributed nature of such networks make the distributed system more vulnerable to various attacks. The security issues associated with distributed networks are increasingly being studied, e.g., jamming, wiretap, spoofing [4]–[7], advanced persistent threats [8] and Byzantine attacks [9], [10]. In this paper, we focus on Byzantine attacks. When the system suffers from Byzantine attacks, some sensors in the network might be compromised and fully controlled by strategic adversaries. We refer to these compromised sensors as Byzantine nodes. They may send falsified information to the FC. There are several types of Byzantine attacks, such as independent probabilistic attack [11], dependent probabilistic attack [12] and non-probabilistic attack [13]. In probabilistic attacks, the Byzantine nodes are in pursuit of long-term profits by launching attacks with a certain probability. In non-probabilistic attacks, the Byzantine nodes decide to launch attacks only when the observations satisfy some specific conditions. For example, a Byzantine node decides to launch attacks only when its observations are higher than threshold λ_1 or lower than threshold λ_0, where $\lambda_1 > \lambda_0$.

There are several works that have studied Byzantine attack issues in distributed detection systems. In [14], optimal strategic data falsification attacks on distributed detection systems are studied. The smart attackers attempt to constrain their exposure to the defense mechanism and maximize the attacking efficacy. In [15], an adaptive algorithm at the FC is proposed to mitigate the impact of Byzantine attacks in the false discovery rate based distributed detection system when the Byzantine nodes know the true hypothesis. In [16]–[18], distributed detection problems are investigated in the context of collaborative spectrum sensing under Byzantine attacks. An abnormality-detection-based algorithm for the detection of attackers in collaborative spectrum sensing is proposed [16]. In [17], the condition under which the Byzantine attackers totally blind the FC is investigated and an algorithm is proposed to detect Byzantine attacks by counting the mismatches between the local decisions and the global decision at the FC. Authors of [18] proposed a Byzantine attacker identification framework in collaborative spectrum sensing where two cases are considered: with and without the prior knowledge of attacker behavior. Good identification performances are achieved in both homogeneous and heterogeneous scenarios even when Byzantine nodes are in a majority. Similarly, in [19], the optimal attacking strategies are analyzed in general distributed network for the cases where the FC has the knowledge of the attackers' strategy and where the FC does not know the attackers' strategy. Audit based mechanisms are proposed to mitigate the effect of Byzantine attacks on the distributed WSNs [20], [21]. In [20], the audit bit based distributed detection scheme is proposed in the Neyman-Pearson framework

C. Quan and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: {chquan,varshney}@syr.edu).

B. Geng is with the Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL 35294 USA (e-mail: bgeng@uab.edu).

Y. S. Han is with the Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, China (e-mail: yunghsiangh@gmail.com).

by utilizing Kullback-Leibler divergence (KLD) to characterize the detection performance of the system. Each sensor sends one additional audit bit to the FC which gives some information about the behavioral identity of each sensor and improves the detection and security performance of the system. Improved system robustness to Byzantine attacks is achieved at the expense of increased communication overhead. In [21], the audit bit based mechanism is utilized in the Bayesian setting. The detection performance of the system is evaluated in terms of the probability of error of the global decision at the FC, and the mitigation scheme over time is proposed by using the information coming from the audit bits.

Our work is most related to the works in [20] and [21]. In [20] and [21], all the sensors in the network are divided into groups of two. Each sensor sends its local decisions to the FC via two paths, one is direct path and another is through the sensor in the same group (indirect path). The indirect decision bits that reach the FC via indirect path are referred to as audit bits which gives us extra information about the behavioral identity of each sensor. In [15] and [16], it is assumed that each Byzantine node falsifies its own local decisions and the decisions coming from its group member with the same probability. However, different from the existing works in [20] and [21], we consider a more realistic case in which the strong assumption of Byzantine nodes' attack behavior made in [20] and [21], namely of equal probability, is relaxed. We call this type of Byzantine nodes as strategic attackers. We show that the traditional audit bit based scheme (TAS) is not robust enough in the presence of strategic attackers. Two new schemes, which are the enhanced audit bit based scheme (EAS) and the reduced audit bit based scheme (RAS), are proposed to improve the robustness and the detection performance of the system under strategic attacks. Then, we extend the above RAS for cluster based wide-area wireless sensor networks (CWSNs) [22], [23]. The cluster based framework has been proposed to deal with the significantly increased energy consumption of the sensors due to the long distance transmission in wide-area networks [24], [25]. This framework not only ensures higher data transmission efficiency, larger network scale, lower bandwidth consumption and prolonged network lifetime, but also efficiently reduces the amount of information transmission in the entire network and mitigates energy dissipation due to collisions. In CWSNs, sensors are divided into several clusters and each cluster is equipped with one cluster head (CH) which has ample energy and computation capacities for operation purposes. The CHs are responsible for collecting the data in the cluster and sending it to the FC. In this work, the sensors in each cluster are further divided into groups of two. Each sensor sends its own decisions via direct and indirect path to the corresponding CH just like the previously proposed audit-based system [20] and [21]. The data aggregation rule for the CHs are designed according to RAS which prolongs the lifetime of the networks with the improved detection performance of the system.[1] We assume that CHs have ample energy to support the long distance transmission[2] and some protections against the attacks so that they can be trusted by the FC, e.g., tamper-resistant security module [28], [29]. The main contributions of this work are summarized as follows:

- We derive the detection performance of the system that employs TAS in the presence of strategic attackers. Instead of considering an identical attacking strategy in which each sensor utilizes the same attacking probability to falsify its own decisions and the decisions coming from their group member [20], [21], we consider attackers that can use different attacking strategies. The optimal attacking strategy of strategic attackers is investigated and we show that it is possible to degrade the performance of TAS to the system without audit bits.
- An EAS is proposed to deal with the security issues arising from the strategic attackers that may use different attacking strategies. We derive the optimal decision rule at the FC and evaluate its detection performance. Simulation results show that the proposed EAS outperforms TAS and the direct scheme under both strategic attacks and non-strategic attacks.
- The scheme EAS is further extended and a new scheme namely RAS is proposed based on our newly proposed EAS. We show that RAS is able to further improve the robustness and the detection performance of the system.
- A wide-area cluster-based WSN is considered. We extend the proposed RAS and design the data aggeration rule for the CHs. Simulation results show a significant reduction in the overall communication overhead between the FC and the CHs.

The rest of the paper is organized as follows. Section II presents the system model of TAS. The optimal attacking strategy is investigated for strategic attackers and the detection performance of the system is evaluated under strategic attacks. Section III presents the proposed EAS and evaluates the detection performance and the robustness of the system. Section IV presents the proposed RAS and extends it for the wide-area networks with several clusters. Section V presents some concluding remarks. The key notations and symbols used in this paper are listed in Table 1 for the convenience of readers. Some intermediate steps of derivations in this paper are omitted due to page limitation. Detailed derivations can be found in [30].

II. Traditional Audit Bit Based Scheme Under strategic Attacks

We consider the binary hypothesis testing problem assuming that there are two possible hypotheses, H_0 (signal is absent) and H_1 (signal is present), regarding a PoI. Consider that we deploy a cluster of N sensors to determine which of the two hypotheses is true. Based on the local observations, each sensor $i \in \{1, \ldots, N\}$ makes a binary decision $v_i \in \{0, 1\}$

[1] This framework is also suitable for sensor networks with mobile access points (SENMA) where the CHs traverse the network to collect information directly from the sensors [26].

[2] The CHs are assumed to be small base stations that can be charged or be unmanned aerial vehicles (UAVs) that are equipped with energy harvesting (EH) circuits which enable the CHs to harvest energy from renewable sources, e.g., vibration, solar and wind, to replenish their energy buffers [27].

TABLE I: Glossary

N	number of sensors
G	number of sensor groups
For any sensor $i \in \{1, 2, \ldots, N\}$:	
v_i	the true local decision made by sensor i
u_i	the local decision sent to MMSD (or FC) by sensor i
z_i	the decision sent to MMSD (or FC) by sensor i which represents the decision made by its group member
w_i	the decision sent to the sensor in the same group by sensor i
P_d	the probability of detection for sensor i
P_f	the probability of false alarm for sensor i
For any sensor pair $i \in \{1, \ldots, N\}$ **and** $j \in \{1, \ldots, i-1, i+1, \ldots, N\}$**:**	
p_1	the probability of flipping v_i
p_2	the probability of flipping w_j
d_i	the status indicator which represents the MMS status of sensor i
$\underline{S}$	the set contains all the sensors whose status indicators are equal to 1
$\overline{S}$	the set contains all the sensors whose status indicators are equal to 0
$\underline{SS}$	the set contains all the sensors whose status indicators and group members' status indicators are both equal to 1
$\underline{S}\overline{S}$	the set contains all the sensors whose status indicators is equal to 1 and group members' status indicators is equal to 0
$\overline{S}\underline{S}$	the set contains all the sensors whose status indicators is equal to 0 and group members' status indicators is equal to 1
$\overline{SS}$	the set contains all the sensors whose status indicators and group members' status indicators are both equal to 0
$\underline{\mathcal{M}}$	the set contains all the sensors whose local decisions and group members' local decisions satisfy $u_i = u_j$
Key acronyms:	
TAS	traditional audit bit based scheme
EAS	enhanced audit bit based scheme
RAS	reduced audit bit based scheme
WSNs	wireless sensor networks
CWSNs	cluster based wide-area wireless sensor networks
MMS	match and mismatch
MMSD	match and mismatch detector

regarding the true hypothesis using the likelihood ratio (LR) test

$$\frac{P(y_i|\mathcal{H}_1)}{P(y_i|\mathcal{H}_0)} \underset{v_i=0}{\overset{v_i=1}{\gtrless}} \lambda, \qquad (1)$$

where λ is the identical threshold used by all the sensors [31], and, $P(y_i|\mathcal{H}_m)$ denotes the conditional probability density function (PDF) of observation y_i under the hypothesis $\mathcal{H}_m$, for $m = 0, 1$. In the audit bit based framework [20] [21], the N sensors are partitioned into G groups where each group $g \in \{1, \ldots, G\}$ is composed of two sensors.[3] Let i and j represent the sensors in the same group, where $i \in \{1, 2, \ldots, N\}$ and $j \in \{1, 2, \ldots, i-1, i+1, \ldots, N\}$. Each sensor i sends its local binary decision to the FC via two paths, one is direct and the other is through sensor j in the same group. At the FC, we design a match and mismatch detector (MMSD) module that detects if the sensor's direct decision matches or mismatches the decision sent through sensor j (indirect decision).

The architecture of each group is shown as Fig. 1(a) and the overall detection network for TAS is shown as Fig. 1(b). As shown in Fig. 1(a), after making its own decision v_i, sensor i sends (i) u_i directly to the MMSD; (ii) w_i to the sensor j in the same group; (iii) z_j, corresponding to w_j coming from the sensor j in the same group, to the MMSD. Similarly, sensor j also sends two decisions u_j and z_i to the MMSD. If the sensor i is a Byzantine node, i.e., $i = B$, the decisions v_i, w_i and u_i are not necessarily the same and z_j are also not necessarily equal to u_j. Let $p(v_i \neq u_i|i = B)$, $p(v_i \neq w_i|i = B)$ and $p(w_j \neq z_j|i = B)$ denote the probabilities that the Byzantine node i flips its own decision, flips the decision sent to its group member and flips the decision coming from its group member, respectively. The probabilities $p_2 = p(w_j \neq z_j|i = B)$ and $p_1 = p(v_i \neq u_i|i = B) = p(v_i \neq w_i|i = B)$ are the attacking parameters the attackers want to optimize. If the sensor i is honest, i.e., $i = H$, we have $v_i = w_i = u_i$ and $z_j = u_j$. In other words, $p(v_i \neq u_i|i = H) = p(v_i \neq w_i|i = H) = 0$. We assume that a fraction α_0 of the N sensors are Byzantine nodes and the FC is not aware of the identity of Byzantine nodes in the network. Hence, each node has the probability of α_0 to be a Byzantine node. We also assume that each Byzantine node attacks the network independently with a certain probability and all the sensors are able to successfully receive the packets from their group members.

After collecting all the local decisions, the MMSD makes binary decisions regarding the match and mismatch (MMS) status of the two decisions corresponding to the same sensor received over different paths, i.e., whether or not the decisions sent via different paths are the same, for all the sensors. Let d_i represent the MMS status of sensor i which is called the status indicator of sensor i. To give a concrete illustration, take one group of sensors (i, j) as an example. The MMSD sets $d_j = 1$ when $u_i = z_i$ and $d_j = 0$ when $u_i \neq z_i$. Similarly, the MMSD sets $d_i = 1$ if $u_j = z_j$ and $d_i = 0$ if $u_j \neq z_j$. The decisions d_i and d_j are the status indicators of sensor i and sensor j, respectively. According to the status indicator for each sensor, the FC places the sensors into two sets $\underline{S}$ and $\overline{S}$. Set $\underline{S}$ contains the sensors whose status indicators are equal to 1 and Set $\overline{S}$ contains the sensors whose status indicators are equal to 0. By employing the extra information coming from these status indicators, we are able to improve the detection performance of the system.

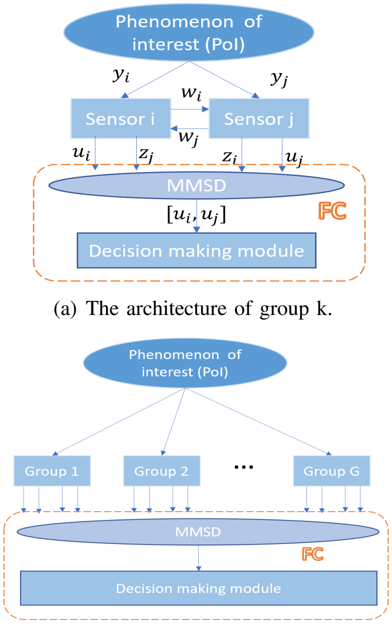
In the following two subsections, we discuss two different attack models and investigate the robustness of the traditional audit bit based mechanism under these two types of attacks. One attack model[4] is that the Byzantine nodes are assumed to flip their own decisions and all the decisions they received with the same probability p, i.e., $p_1 = p_2 = p$. The other model is that the Byzantine nodes use different probabilities to flip their own decisions and all the decisions they receive, i.e., $p_1 \neq p_2$. It is more general and practical to consider Byzantine nodes which relax the assumption of $p_1 = p_2 = p$ made in the traditional audit bit based mechanism. This allows the Byzantines to be strategic by optimally employing unequal probabilities p_1 and p_2.

A. TAS

In the traditional audit bit based mechanism, the Byzantine nodes are assumed to flip their own decisions and all the decisions they receive with the same probability p, i.e.,

[3] The sensors are divided into groups of two based on certain criteria, e.g., according to their distances from each other.

[4] This attack model follows the work in [20] and [21].



(a) The architecture of group k.

(b) The overall detection network for audit bit based scheme.

Fig. 1: (a) The architecture of a group $k \in \{1, 2, \dots, G\}$ and (b) the overall system model.

$p_1 = p_2 = p$. Based on the status indicators $\{d_i\}_{i=1}^{N}$, we have the following two cases [20].

a) If $d_i = 1$: i is a Byzantine node with probability

$$\begin{aligned}\underline{\alpha} &= P(i = B|d_i = 1)\\ &= \frac{\alpha_0(1-p)[1-2\alpha_0 p(1-2p)]}{1-\alpha_0(3-2p)p+4\alpha_0^2(1-p)p^2}\end{aligned} \tag{2}$$

and the sensor i is placed in set $\underline{S}$.

b) If $d_i = 0$: i is a Byzantine node with probability

$$\begin{aligned}\overline{\alpha} &= P(i = B|d_i = 0)\\ &= \frac{1+2(1-p)(\alpha_0-2\alpha_0 p))}{1+2(1-p)(1-2\alpha_0 p))}\end{aligned} \tag{3}$$

and the sensor i is placed in set $\overline{S}$.

It has been proved in [21] (Lemma 1) that $\underline{\alpha} \leq \alpha_0 \leq \overline{\alpha}$. In other words, all the sensors are divided into two sets $\underline{S}$ and $\overline{S}$ in which the sensors have lower probability $\underline{\alpha}$ and higher probability $\overline{\alpha}$ of being Byzantine nodes, respectively, according to status indicators $\mathbf{d} = [d_1, d_2, \dots, d_N]$. Let P_d, P_f be the probability of detection and the probability of false alarm for any sensor $i \in \{1, \dots, N\}$, respectively, i.e., $P_d = P(v_i = 1|\mathcal{H}_1)$ and $P_f = P(v_i = 1|\mathcal{H}_0)$. Thus, the probability mass function (pmf) of local decision u_i is expressed as

$$P(u_i|\mathcal{H}_q) = \begin{cases} \underline{\pi}_{1q}^{u_i}(1-\underline{\pi}_{1q})^{1-u_i} & \text{for } i \in \underline{S} \\ \overline{\pi}_{1q}^{u_i}(1-\overline{\pi}_{1q})^{1-u_i} & \text{for } i \in \overline{S} \end{cases} \tag{4}$$

for q=0,1, where, for $i \in \underline{S}$,

$$\underline{\pi}_{11} = 1 - \underline{\pi}_{01} = P(u_i = 1|\mathcal{H}_1) = P_d(1-\underline{\alpha}p) + \underline{\alpha}p(1-P_d) \tag{5a}$$

$$\underline{\pi}_{10} = 1 - \underline{\pi}_{00} = P(u_i = 1|\mathcal{H}_0) = P_f(1-\underline{\alpha}p) + \underline{\alpha}p(1-P_f) \tag{5b}$$

and, for $i \in \overline{S}$,

$$\overline{\pi}_{11} = 1 - \overline{\pi}_{01} = P(u_i = 1|\mathcal{H}_1) = P_d(1-\overline{\alpha}p) + \overline{\alpha}p(1-P_d) \tag{6a}$$

$$\overline{\pi}_{10} = 1 - \overline{\pi}_{00} = P(u_i = 1|\mathcal{H}_0) = P_f(1-\overline{\alpha}p) + \overline{\alpha}p(1-P_f). \tag{6b}$$

Then the optimal decision rule when the attacking strategy p is assumed to be known is given as

$$\underline{WU} + \overline{WU} \gtrless \eta^{(A)}, \tag{7}$$

where $\underline{U} = \sum_{i \in \underline{S}} u_i$, $\overline{U} = \sum_{i \in \overline{S}} u_i$, $\underline{W} = \log(\frac{\underline{\pi}_{11}(1-\underline{\pi}_{10})}{\underline{\pi}_{10}(1-\underline{\pi}_{11})})$, $\overline{W} = \log(\frac{\overline{\pi}_{11}(1-\overline{\pi}_{10})}{\overline{\pi}_{10}(1-\overline{\pi}_{11})})$, $\eta^{(A)} = \log(\frac{\pi_0}{\pi_1}) + \underline{N}\log(\frac{1-\underline{\pi}_{10}}{1-\underline{\pi}_{11}}) + \overline{N}\log(\frac{1-\overline{\pi}_{10}}{1-\overline{\pi}_{11}})$, $\underline{N} = |\underline{S}|$, and $\overline{N} = |\overline{S}|$. Note that $\underline{U}$ and $\overline{U}$ are binomial distributed random variables with parameters $(N, \underline{\pi}_{10})$ and $(N, \overline{\pi}_{10})$, respectively, under $\mathcal{H}_0$, and with parameters $(N, \underline{\pi}_{11})$ and $(N, \overline{\pi}_{11})$, respectively, under $\mathcal{H}_1$. When N is large, $\underline{N}$ and $\overline{N}$ can be approximated by their expected value $NP(u_i = z_i)$ and $NP(u_i \neq z_i)$. $\eta^{(A)}$ is the threshold used by the FC for the traditional audit bit based system, where $\eta^{(A)} = \log(\frac{\pi_0}{\pi_1}) + NP(u_i = z_i)\log(\frac{1-\underline{\pi}_{10}}{1-\underline{\pi}_{11}}) + NP(u_i \neq z_i)\log(\frac{1-\overline{\pi}_{10}}{1-\overline{\pi}_{11}})$. Moreover, $\underline{U}$ and $\overline{U}$ can be approximated by the Gaussian distribution with parameters given as follows:

$$\begin{aligned}\mu_m^{(A)} =& E[U|\mathcal{H}_m]\\ =& N[P(u_i = z_i)\underline{\pi}_{1m}\underline{W} + P(u_i \neq z_i)\overline{\pi}_{1m}\overline{W}]\end{aligned} \tag{8a}$$

$$\begin{aligned}(\sigma_m^{(A)})^2 =& Var[U|\mathcal{H}_m] = N[P(u_i = z_i)\underline{\pi}_{1m}(1-\underline{\pi}_{1m})\underline{W}^2\\ &+ P(u_i \neq z_i)\overline{\pi}_{1m}(1-\overline{\pi}_{1m})\overline{W}^2],\end{aligned} \tag{8b}$$

for $m = 0, 1$. The detection performance, characterized by the probability of error $P_e^{(A)}$ for the system with TAS, is given as

$$P_e^{(A)} = \pi_0 Q\left(\gamma_f^{(A)}\right) + \pi_1 Q\left(\gamma_m^{(A)}\right), \tag{9}$$

where $\gamma_f^{(A)} = \frac{\eta^{(A)} - \mu_0^{(A)}}{\sigma_0^{(A)}}$ and $\gamma_m^{(A)} = \frac{\mu_1^{(A)} - \eta^{(A)}}{\sigma_1^{(A)}}$. Let $P_e^{(D)}$ denote the probability of error for the system with direct scheme, which is expressed as (53). It has been shown in [21] (Theorem 3) that the probability of error of the traditional audit based system given any α_0 and p is always less than or equal to that of the system which relies only on direct decisions, i.e, $P_e^{(A)} \leq P_e^{(D)}$.

However, due to the strong assumption of $p_1 = p_2 = p$, TAS can accurately assess the behavioral identity of each sensor in the network so that it can improve the detection and security performances of the system. It is obvious that a higher p means a higher probability that the Byzantine nodes flip their own decisions and the decisions coming from their group members. Thus, the Byzantine nodes have a higher probability of being placed in the Set $\overline{S}$. In the next subsection, we relax the the assumption of $p_1 = p_2 = p$ and investigate the detection performance of the traditional audit bit based system under the relaxed assumption.

B. The strategic Attacker under Traditional Audit Bit Based System

To make the model more general, we assume that the attackers are more strategic in that they can employ different values of p_1 and p_2 that are not necessarily equal. In this subsection, we analyze the detection performance of the traditional audit bit based system under such strategic attacks.

When the FC under strategic attacks makes use of the status indicators to place all the sensors into two sets, we have the following two cases.

a) If $d_i = 1$: i is a Byzantine node with probability

$$\underline{\alpha}^I = P(i = B|d_i = 1) = \frac{P(d_i = 1|i = B)P(i = B)}{P(d_i = 1)}, \quad (10)$$

where

$$\begin{aligned} &P(d_i = 1|i = B) = P(u_j = z_j|i = B) \\ &= -4\alpha_0 p_1^2 p_2 + 4\alpha_0 p_1 p_2 - 2\alpha_0 p_1 + 2\alpha_0 p_1^2 - p_2 + 1 \end{aligned} \quad (11)$$

and

$$\begin{aligned} &P(d_i = 1|i = H) = P(u_j = z_j|i = H) \\ &= 2\alpha_0 p_1^2 - 2\alpha_0 p_1 + 1. \end{aligned} \quad (12)$$

Thus, the unconditional probability of matching $p(u_j = z_j)$ is given as

$$\begin{aligned} &P(d_i = 1) \\ &= P(u_j = z_j|i = H)P(i = H) + P(u_j = z_j|i = B)P(i = B) \\ &= -4\alpha_0^2 p_1^2 p_2 + 4\alpha_0^2 p_1 p_2 + 2\alpha_0 p_1^2 - \alpha_0 p_2 - 2\alpha_0 p_1 + 1. \end{aligned} \quad (13)$$

In this case, the sensor i is placed in set $\underline{S}$ with

$$\underline{\alpha}^I = \frac{4\alpha_0^2 p_1^2 p_2 + 4\alpha_0^2 p_1 p_2 - 2\alpha_0^2 p_1 + 2\alpha_0^2 p_1^2 - \alpha_0 p_2 + \alpha_0}{4\alpha_0^2 p_1^2 p_2 + 4\alpha_0^2 p_1 p_2 + 2\alpha_0 p_1^2 - \alpha_0 p_2 - 2\alpha_0 p_1 + 1}. \quad (14)$$

b) If $d_i = 0$: i is a Byzantine node with probability

$$\begin{aligned} \overline{\alpha}^I &= P(i = B|d_i = 0) = P(i = B|u_j \neq z_j) \\ &= \frac{P(u_j \neq z_j|i = B)P(i = B)}{P(u_j \neq z_j)} \\ &= \frac{4\alpha_0 p_1^2 p_2 - 4\alpha_0 p_1 p_2 + 2\alpha_0 p_1 - 2\alpha_0 p_1^2 + p_2}{4\alpha_0 p_1^2 p_2 - 4\alpha_0 p_1 p_2 - 2p_1^2 + p_2 + 2p_1}, \end{aligned} \quad (15)$$

where $p(u_j \neq z_j|i = B) = 1 - p(u_j = z_j|i = B)$ and $p(u_j \neq z_j) = 1 - p(u_j = z_j)$. In this case, the sensor i is placed in set $\overline{S}$. We show two important properties of $\overline{\alpha}^I$ and $\underline{\alpha}^I$ in the next lemma.

Lemma 1: We have the following two relationships in terms of $\underline{\alpha}^I$, $\overline{\alpha}^I$, and α_0.

1) Under strategic attacks, the probability of being a Byzantine node given the sensor in Set $\underline{S}$ is smaller than or equal to the one given the sensor in Set $\overline{S}$, i.e., $\underline{\alpha}^I \leq \alpha_0 \leq \overline{\alpha}^I$.
2) $\underline{\alpha}^I = \overline{\alpha}^I = \alpha_0$ when $p_2 = 0$.

Proof: According to (14) and (15), we show that $\frac{\partial \underline{\alpha}^I}{\mathrm{d}p_2} \leq 0$, and $\frac{\partial \underline{\alpha}^I}{\mathrm{d}p_1} \leq 0$. Due to the fact that $\alpha_0 \in [0,1]$, $p_1 \in [0,1]$, and $p_2 \in [0,1]$, we have

$$\begin{aligned} \frac{\partial \underline{\alpha}^I}{\mathrm{d}p_2} &= \frac{(4\alpha_0^2 p_1(1-p_1) - \alpha_0)(1-\alpha_0)(2\alpha_0 p_1(p_1-1)+1)}{(4\alpha_0^2 p_1^2 p_2 + 4\alpha_0^2 p_1 p_2 + 2\alpha_0 p_1^2 - \alpha_0 p_2 - 2\alpha_0 p_1 + 1)^2} \\ &\overset{(a)}{\leq} \frac{\alpha_0(\alpha_0 - 1)(1-\alpha_0)(2\alpha_0 p_1(p_1-1)+1)}{(4\alpha_0^2 p_1^2 p_2 + 4\alpha_0^2 p_1 p_2 + 2\alpha_0 p_1^2 - \alpha_0 p_2 - 2\alpha_0 p_1 + 1)^2} \\ &\overset{(b)}{\leq} 0 \end{aligned} \quad (16a)$$

$$\frac{\partial \underline{\alpha}^I}{\mathrm{d}p_1} = -2\alpha_0^2(1 - \alpha_0 p_2)(1 - 2p_2)^2 \leq 0. \quad (16b)$$

The equality in (a) is achieved when $p_1 = \frac{1}{2}$. (b) is due to the fact that $2\alpha_0 p_1(p_1 - 1) + 1 \geq 1 - \frac{\alpha_0}{2} > 0$ and the equality in (b) is achieved when $\alpha_0 = 1$. Thus, according to (16), we have the maximum value of $\underline{\alpha}^I$ when $p_1 = 0$ and $p_2 = 0$, i.e., $\underline{\alpha}^I(p_1, p_2) \leq \underline{\alpha}^I(p_1 = 0, p_2 = 0) = \alpha_0$. Since $p(u_i = z_i)\underline{\alpha}^I + p(u_i \neq z_i)\overline{\alpha}^I = \alpha_0$, we have

$$\begin{aligned} P(u_i = z_i)\alpha_0 + P(u_i \neq z_i)\overline{\alpha}^I &\geq \alpha_0 \\ P(u_i \neq z_i)\overline{\alpha}^I &\geq \alpha_0(1 - P(u_i = z_i)) \\ \overline{\alpha}^I &\geq \alpha_0 \end{aligned} \quad (17)$$

Based on the analysis above, we conclude that $\underline{\alpha}^I \leq \alpha_0 \leq \overline{\alpha}^I$. Note that the equality on both sides can be achieved when $p_2 = 0$. Hence, we get the results stated in Lemma 1. ■

Substituting $\underline{\alpha}$, $\overline{\alpha}$ and p with $\underline{\alpha}^I$, $\overline{\alpha}^I$ and p_1, respectively, in (5) and (6), we can obtain $\underline{\pi}_{10}^I$, $\underline{\pi}_{11}^I$, $\overline{\pi}_{10}^I$, $\overline{\pi}_{11}^I$. After getting $\underline{\pi}_{10}^I$, $\underline{\pi}_{11}^I$ and $\overline{\pi}_{10}^I$, $\overline{\pi}_{11}^I$, we can calculate the pmfs of u_i according to (4). Hence, the probability of error for the system under strategic attack is given by $P_e^I = \pi_0 Q\left(\gamma_f^I\right) + \pi_1 Q\left(\gamma_m^I\right)$. γ_f^I and γ_m^I are shown in (18), where $D_0(\underline{\alpha}^I, p_1, p_2) = \underline{\pi}_{10}^{(I)} \log(\frac{\underline{\pi}_{10}^{(I)}}{\underline{\pi}_{11}^{(I)}}) + (1 - \underline{\pi}_{10}^{(I)}) \log(\frac{1-\underline{\pi}_{10}^{(I)}}{1-\underline{\pi}_{11}^{(I)}})$, $D_0(\overline{\alpha}^I, p_1, p_2) = \overline{\pi}_{10} \log(\frac{\overline{\pi}_{10}^{(I)}}{\overline{\pi}_{11}^{(I)}}) + (1 - \overline{\pi}_{10}^{(I)}) \log(\frac{1-\overline{\pi}_{10}^{(I)}}{1-\overline{\pi}_{11}^{(I)}})$, $D_1(\overline{\alpha}^I, p_1, p_2) = \overline{\pi}_{11}^{(I)} \log(\frac{\overline{\pi}_{11}^{(I)}}{\overline{\pi}_{10}^{(I)}}) + (1 - \overline{\pi}_{11}^{(I)}) \log(\frac{1-\overline{\pi}_{11}^{(I)}}{1-\overline{\pi}_{10}^{(I)}})$ and $D_1(\underline{\alpha}^I, p_1, p_2) = \underline{\pi}_{11}^{(I)} \log(\frac{\underline{\pi}_{11}^{(I)}}{\underline{\pi}_{10}^{(I)}}) + (1 - \underline{\pi}_{11}^{(I)}) \log(\frac{1-\underline{\pi}_{11}^{(I)}}{1-\underline{\pi}_{10}^{(I)}})$. We also have $g_0(\underline{\alpha}^I, p_1, p_2) = \underline{\pi}_{10}^{(I)}(1-\underline{\pi}_{10}^{(I)})\underline{W}^2$, $g_0(\overline{\alpha}^I, p_1, p_2) = \overline{\pi}_{10}^{(I)}(1 - \overline{\pi}_{10}^{(I)})(\overline{W}^I)^2$ and $g_1(\underline{\alpha}^I, p_1, p_2) = \underline{\pi}_{11}^{(I)}(1 - \underline{\pi}_{11}^{(I)})\underline{W}^2$, $g_1(\overline{\alpha}^I, p_1, p_2) = \overline{\pi}_{11}^{(I)}(1 - \overline{\pi}_{11}^{(I)})(\overline{W}^I)^2$ where $\underline{W}^I = \log(\frac{\underline{\pi}_{11}^I(1-\underline{\pi}_{10}^I)}{\underline{\pi}_{10}^I(1-\underline{\pi}_{11}^I)})$ and $\overline{W}^I = \log(\frac{\overline{\pi}_{11}^I(1-\overline{\pi}_{10}^I)}{\overline{\pi}_{10}^I(1-\overline{\pi}_{11}^I)})$. The optimal attacking strategy is stated based on (18) in the following theorem.

Theorem 1: In the traditional audit based system, if the strategic Byzantine attackers adopt the strategy given by $p_2 = 0$ when $\alpha_0 \in [0,1]$, the system reduces to the one without audit bits and it can always be made blind by choosing p_1 such that $\alpha_0 p_1 = \frac{1}{2}$ if $\alpha_0 \geq 0.5$.

Proof: Please see Appendix A. ■

Note that the probability of error for the system under strategic attack is $P_e^I = \pi_0 Q\left(\gamma_f^I\right) + \pi_1 Q\left(\gamma_m^I\right)$. γ_f^I and γ_m^I are the arguments of function $Q(.)$ for the probability of false alarm and the argument of function $Q(.)$ for the probability of miss detection, respectively such that larger arguments mean

$$\gamma_f^I = \frac{\log(\frac{\pi_0}{\pi_1})/\sqrt{N} + \sqrt{N}(D_0(\overline{\alpha}^I, p_1, p_2)p(u_n = z_n) + D_0(\underline{\alpha}^I, p_1, p_2)p(u_n \neq z_n))}{\sqrt{p(u_i \neq z_i)g_0(\overline{\alpha}^I, p_1, p_2) + p(u_i = z_i)g_0(\underline{\alpha}^I, p_1, p_2)}} \tag{18a}$$

$$\gamma_m^I = \frac{\log(\frac{\pi_0}{\pi_1})/\sqrt{N} + \sqrt{N}(D_1(\overline{\alpha}^I, p_1, p_2)p(u_n = z_n) + D_1(\underline{\alpha}^I, p_1, p_2)p(u_n \neq z_n))}{\sqrt{p(u_i \neq z_i)g_1(\overline{\alpha}^I, p_1, p_2) + p(u_i = z_i)g_1(\underline{\alpha}^I, p_1, p_2)}}, \tag{18b}$$

better detection performance. Fig. 2 shows how γ_f^I and γ_m^I change with p_2. We can observe that both γ_f^I and γ_m^I achieve the minimum when $p_2 = 0$, which means that P_e^I achieves the maximum. We can also observe that arguments that attain this are equal to the ones in the system that does not use audit bits and thus P_e^I reduces to the probability of error of the system that does not use audit bits. Hence, Fig. 2 is in accordance with the result given in Theorem 1.

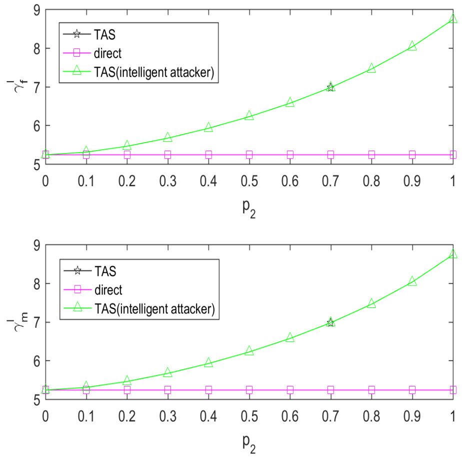


Fig. 2: γ_f^I and γ_m^I versus p_2 given $p_1 = 0.7$ and $\alpha_0 = 0.3$. Note that $p_1 = p_2 = 0.7$ in TAS.

Based on the analysis above, the assumption $p_1 = p_2$ given in [21] is not the optimal choice for the attackers in practice. The attackers can launch stronger attacks when they set $p_2 = 0$. Under this attacking strategy, there is no improvement in the detection performance of TAS compared with the direct scheme. Thus, we conclude that the strategic attackers can hide themselves by not flipping the decisions from their group members, i.e., $p_2 = 0$, according to Theorem 1 and Fig. 2. Moreover, when $p_2 = 0$, the detection error for TAS is the same as the one for the direct scheme. To enhance the robustness of the system, we propose a new scheme called enhanced audit bit based scheme (EAS) in next section.

III. Enhanced Audit Bit Based Scheme

In this section, an enhanced audit bit based scheme (EAS) is proposed to improve the robustness of the system under strategic attacks. In TAS, the behavioral identity of each sensor is characterized by $\underline{\alpha}$ and $\overline{\alpha}$. The evaluations of the value of $\underline{\alpha}$ and $\overline{\alpha}$ only depends on its own status indicator as discussed in Section II. However, in the newly proposed scheme, we utilize both the status indicators of the sensors in the same group to more accurately infer the behavioral identities of sensors in the network compared with TAS.

The status indicators $\{d_i\}_{i=1}^N$ are again made by the MMSD. However, the sensors are no longer partitioned into two sets ($\underline{S}$ and $\overline{S}$). They are partitioned into four sets which are $\underline{SS}$ $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$ based on both status indicators of sensor i and sensor j in the same group. If $d_i = d_j = 1$, sensor i and sensor j are both placed in the set $\underline{SS}$. If $d_i = 0$ and $d_j = 1$, sensor i is placed in the set $\underline{S}\overline{S}$ and sensor j is placed in the set $\overline{S}\underline{S}$. If $d_i = d_j = 0$, sensor i and sensor j are both placed in the set $\overline{SS}$. We still assume a general attacking strategy which is $p_1 \neq p_2$. Then, we have the following four cases.

a) If $i \in \underline{SS}$: i is a Byzantine node with probability

$$\begin{aligned}\alpha_1 =& P(i = B | i, j \in \underline{SS}) \\ =& P(i = B, j = H | i, j \in \underline{SS}) + P(i = B, j = B | i, j \in \underline{SS}) \\ =& \frac{\alpha_0^2 f_{BB}^{(1)} + \alpha_0(1-\alpha_0) f_{BH}^{(1)}}{P(i, j \in \underline{SS})},\end{aligned} \tag{19}$$

where

$$\begin{aligned}P(i, j \in \underline{SS}) =& \alpha_0^2 f_{BB}^{(1)} + \alpha_0(1-\alpha_0)(f_{HB}^{(1)} + f_{BH}^{(1)}) \\ &+ (1-\alpha_0)^2 f_{HH}^{(1)}\end{aligned} \tag{20}$$

and $f_{BB}^{(1)} = [2p_1p_2(1-p_1) + (1-2p_1+2p_1^2)(1-p_2)]^2$, $f_{HB}^{(1)} = f_{BH}^{(1)} = (1-p_2)(1-2p_1+2p_1^2)$ and $f_{HH}^{(1)} = 1$.

b) If $i \in \underline{S}\overline{S}$: i is a Byzantine node with probability

$$\begin{aligned}\alpha_2 =& P(i = B | i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S}) \\ =& P(i = B, j = H | i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S}) \\ &+ P(i = B, j = B | i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S}) \\ =& \frac{\alpha_0^2 f_{BB}^{(2)} + \alpha_0(1-\alpha_0) f_{BH}^{(2)}}{P(i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S})},\end{aligned} \tag{21}$$

where

$$\begin{aligned}P(i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S}) =& \alpha_0^2 f_{BB}^{(2)} + \alpha_0(1-\alpha_0)(f_{HB}^{(2)} + f_{BH}^{(2)}) \\ &+ (1-\alpha_0)^2 f_{HH}^{(2)}\end{aligned} \tag{22}$$

and $f_{BB}^{(2)} = [2p_1p_2(1-p_1) + (1-2p_1+2p_1^2)(1-p_2)][1 - 2p_1p_2(1-p_1) - (1-2p_1+2p_1^2)(1-p_2)]$, $f_{HB}^{(2)} = p_2(1-2p_1+2p_1^2)$,$f_{BH}^{(2)} = 2p_1(1-p_2)(1-p_1)$ and $f_{HH}^{(2)} = 0$.

c) If $i \in \overline{S}\underline{S}$: i is a Byzantine node with probability

$$\begin{aligned}\alpha_3 =& P(i = B|i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S}) \\ =& P(i = B, j = H|i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S}) \\ &+ P(i = B, j = B|i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S}) \\ =& \frac{\alpha_0^2 f_{BB}^{(3)} + \alpha_0(1-\alpha_0)f_{BH}^{(3)}}{P(i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S})},\end{aligned} \tag{23}$$

where

$$\begin{aligned}P(i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S}) =& \alpha_0^2 f_{BB}^{(3)} + \alpha_0(1-\alpha_0)(f_{HB}^{(3)} + f_{BH}^{(3)}) \\ &+ (1-\alpha_0)^2 f_{HH}^{(3)}\end{aligned} \tag{24}$$

and $f_{BB}^{(3)} = [2p_1p_2(1-p_1) + (1-2p_1+2p_1^2)(1-p_2)][1 - 2p_1p_2(1-p_1) - (1-2p_1+2p_1^2)(1-p_2)]$, $f_{HB}^{(3)} = 2p_1(1-p_2)(1-p_1)$,$f_{BH}^{(3)} = p_2(1-2p_1+2p_1^2)$ and $f_{HH}^{(3)} = 0$.

d) If $i \in \overline{SS}$: i is a Byzantine node with probability

$$\begin{aligned}\alpha_4 =& P(i = B|i, j \in \overline{SS}) \\ =& P(i = B, j = H|i, j \in \overline{SS}) \\ &+ P(i = B, j = B|i, j \in \overline{SS}) \\ =& \frac{\alpha_0^2 f_{BB}^{(4)} + \alpha_0(1-\alpha_0)f_{BH}^{(4)}}{P(i, j \in \overline{SS})},\end{aligned} \tag{25}$$

where

$$\begin{aligned}P(i, j \in \overline{SS}) =& \alpha_0^2 f_{BB}^{(4)} + \alpha_0(1-\alpha_0)(f_{HB}^{(4)} + f_{BH}^{(4)}) \\ &+ (1-\alpha_0)^2 f_{HH}^{(4)}\end{aligned} \tag{26}$$

and $f_{BB}^{(4)} = [2p_1(1-p_2)(1-p_1) + p_2p_1^2]^2$, $f_{HB}^{(4)} = f_{BH} = 2p_1p_2(1-p_1)$ and $f_{HH}^{(4)} = 0$.

The next lemma shows that our proposed EAS performs a more accurate evaluation of the behavioral identity of each sensor compared with TAS.

Lemma 2: The probability of sensor i being a Byzantine node when $i \in \underline{S}$ in TAS is equal to the weighted average of the probabilities of sensor i being a Byzantine node when $i, j \in \underline{SS}$ and $i \in \underline{S}\overline{S}, j \in \overline{S}\underline{S}$, respectively. That is

$$\begin{aligned}&P(i = B|i \in \underline{S}) \\ &= \alpha_1 P(d_j = 1|d_i = 1) + \alpha_2 P(d_j = 0|d_i = 1)\end{aligned} \tag{27}$$

A similar result can be obtained for sensor $i \in \overline{S}$.

Proof: The right hand side (RHS) of (27) is the same as $P(i = B|d_i = 1, d_j = 1)P(d_j = 1|d_i = 1) + P(i = B|d_i = 1, d_j = 0)P(d_j = 0|d_i = 1)$. According to the Bayes' rule, we have

$$\begin{aligned}&\sum_{x=0,1} P(i = B|d_i = 1, d_j = x)P(d_j = x|d_i = 1) \\ &= \sum_{x=0,1} P(i = B|d_i = 1, d_j = x)\frac{P(d_i = 1, d_j = x)}{P(d_i = 1)} \\ &= \sum_{x=0,1} \frac{\alpha_0^2 f_{BB}^{(x)} + \alpha_0(1-\alpha_0)f_{BH}^{(x)}}{P(d_i = 1)} \\ &= P(i = B|i \in \underline{S})\end{aligned} \tag{28}$$

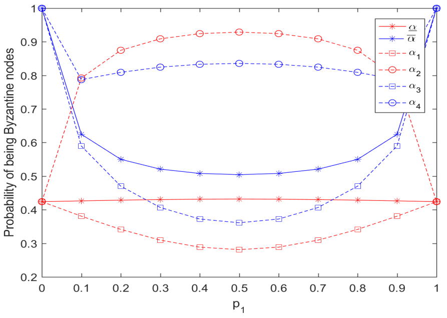


Fig. 3: The probability of being Byzantine nodes for sensors in sets $\underline{S}$, $\overline{S}$, $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$ when $p_2 = 0.1$.

We can also show that $i \in \overline{S}$ is the weighted average of the probabilities of sensor i being a Byzantine node when $i, j \in \overline{SS}$ and $i \in \overline{S}\underline{S}, j \in \underline{S}\overline{S}$ by following a similar procedure and, therefore, the details of its proof are omitted here. ■

Fig. 3 corroborates the results in Lemma 2. Note that each sensor placed in $\underline{S}$ (or $\overline{S}$) is a Byzantine node with probability of $\underline{\alpha}$ (or $\overline{\alpha}$) for TAS. We can observe that the value of $\underline{\alpha}$ (or $\overline{\alpha}$) is in the middle of the values of α_1 and α_2 (or α_3 and α_4) for the proposed scheme. It shows that taking both the status indicators from the same group into consideration can give us more information about the the behavioral identities of the sensors in the network. Hence, our proposed EAS outperforms TAS that only utilizes the averaged probabilities ($\underline{\alpha}$ or $\overline{\alpha}$) to assess the behavioral identity for each sensor. Thus, the pmf of local decision u_i for our proposed EAS is expressed as

$$P(u_i|\mathcal{H}_q) = \begin{cases} \pi_{1q,1}^{u_i}(1-\pi_{1q,1})^{1-u_i} & \text{for } i \in \underline{SS} \\ \pi_{1q,2}^{u_i}(1-\pi_{1q,2})^{1-u_i} & \text{for } i \in \underline{S}\overline{S} \\ \pi_{1q,3}^{u_i}(1-\pi_{1q,3})^{1-u_i} & \text{for } i \in \overline{S}\underline{S} \\ \pi_{1q,4}^{u_i}(1-\pi_{1q,4})^{1-u_i} & \text{for } i \in \overline{SS} \end{cases} \tag{29}$$

for $q = 0, 1$, where

$$\pi_{11,e} = 1 - \pi_{10,e} = P_d(1-\alpha_e p_1) + \alpha_e p_1(1-P_d) \tag{30a}$$

$$\pi_{10,e} = 1 - \pi_{00,e} = P_f(1-\alpha_e p_1) + \alpha_e p_1(1-P_f) \tag{30b}$$

for $e = 1, 2, 3, 4$. $\pi_{11,e}$ and $\pi_{10,e}$ are the probabilities of sending the local decision $u_i = 1$ given hypothesis $\mathcal{H}_1$ and given hypothesis $\mathcal{H}_0$, respectively, for $e = 1, 2, 3, 4$ which are corresponding to the sensors being in $\underline{SS}$ $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$. The new optimal decision rule is provided in Theorem 2.

Theorem 2: The new decision rule for the proposed EAS, given the Byzantine flipping probabilities p_1, p_2 and α_0 fraction of Byzantine nodes, is expressed as

$$\sum_{e=1}^{4} W_e U_e \gtrless \eta^{(En)}, \tag{31}$$

where $U_1 = \sum_{i \in \underline{SS}} u_i$, $U_2 = \sum_{i \in \underline{S}\overline{S}} u_i$, $U_3 = \sum_{i \in \overline{S}\underline{S}} u_i$, $U_4 = \sum_{i \in \overline{SS}} u_i$ and $W_e = \log(\frac{\pi_{11,e}(1-\pi_{10,e})}{\pi_{10,e}(1-\pi_{11,e})})$ for $e = 1, 2, 3, 4$. $\eta^{(En)}$ is the threshold used by the FC for EAS, where $\eta^{(En)} = \log(\frac{\pi_0}{\pi_1}) + \sum_{e=1}^{4} N_e \log(\frac{1-\pi_{10,e}}{1-\pi_{11,e}})$. N_1, N_2, N_3 and N_4 are the cardinalities of sets $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$, respectively, where $N_1 = |\underline{SS}|$,$N_2 = |\underline{S}\overline{S}|$, $N_3 = |\overline{S}\underline{S}|$ and $N_4 = |\overline{SS}|$.

$$\prod_{i\in\underline{SS}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\prod_{i\in\underline{S}\overline{S}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\prod_{i\in\overline{S}\underline{S}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\prod_{i\in\overline{SS}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\gtrless\frac{\pi_0}{\pi_1} \tag{32}$$

Proof: We know that the local decisions are independent given the hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$ and the information about the sets where all the sensors are placed in. Hence, the optimal decision rule, which is given in (32), can be further simplified. Substituting (29) in (32), and taking the logarithm on both sides, we obtain the fusion rule in the theorem. ■

Note that U_e is binomial distributed random variables with parameters $(N, \pi_{11,e})$ under $\mathcal{H}_1$, and with parameters $(N, \pi_{10,e})$ under $\mathcal{H}_0$ for $e = 1, 2, 3, 4$. When N is large, N_1, N_2, N_3 and N_4 can be approximated by their expected value $NP(i \in \underline{SS})$, $NP(i \in \underline{S}\overline{S})$, $NP(i \in \overline{S}\underline{S})$ and $NP(i \in \overline{SS})$, respectively. For any sensor $i \in \{1, 2, \ldots, N\}$, the probability of being placed in $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$ are $P(i \in \underline{SS}) = P(d_i = d_j = 1)$, $P(i \in \underline{S}\overline{S}) = P(i \in \overline{S}\underline{S}) = P(d_i = 1, d_j = 0) = P(d_i = 0, d_j = 1)$ and $P(i \in \overline{SS}) = P(d_i = d_j = 0)$, respectively. The threshold used by the FC becomes $\eta^{(En)} = \log(\frac{\pi_0}{\pi_1}) + NP(i \in \underline{SS})\log(\frac{1-\pi_{10,1}}{1-\pi_{11,1}}) + NP(i \in \underline{S}\overline{S})\log(\frac{1-\pi_{10,2}}{1-\pi_{11,2}}) + NP(i \in \overline{S}\underline{S})\log(\frac{1-\pi_{10,3}}{1-\pi_{11,3}}) + NP(i \in \overline{SS})\log(\frac{1-\pi_{10,4}}{1-\pi_{11,4}})$. Thus, the PDF of the global static $U = \sum_{e=1}^{4} W_e U_e$ can be approximated by the Gaussian distribution with parameters given as follows.

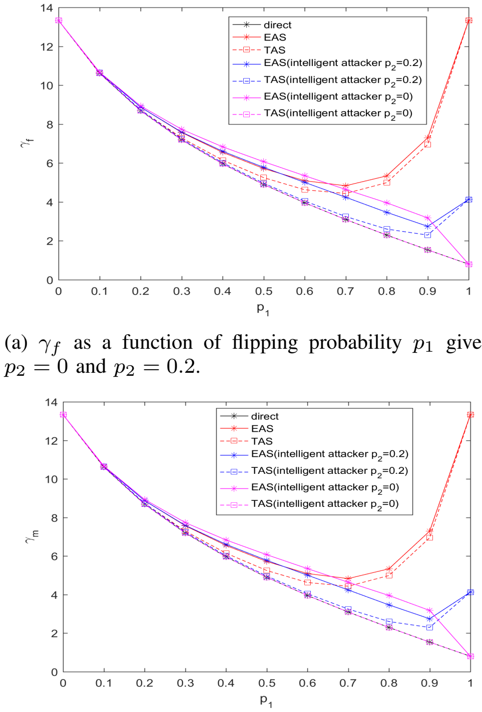
$$\begin{aligned}\mu_m^{(En)} =& E[U|\mathcal{H}_m]\\ =& N(P(i \in \underline{SS})\pi_{1m,1}W_1 + P(i \in \underline{S}\overline{S})\pi_{1m,2}W_2\\ &+ P(i \in \overline{S}\underline{S})\pi_{1m,3}W_3 + P(i, j \in \overline{SS})\pi_{1m,4}W_4)\end{aligned} \tag{33a}$$

$$\begin{aligned}(\sigma_m^{(En)})^2 =& Var[U|\mathcal{H}_m]\\ =& N(P(i \in \underline{SS})\pi_{1m,1}(1-\pi_{1m,1})W_1^2\\ &+ P(i \in \underline{S}\overline{S})\pi_{1m,2}(1-\pi_{1m,2})W_2^2\\ &+ P(i \in \overline{S}\underline{S})\pi_{1m,3}(1-\pi_{1m,3})W_3^2\\ &+ P(i \in \overline{SS})\pi_{1m,4}(1-\pi_{10,4})W_4^2),\end{aligned} \tag{33b}$$

for $m = 0, 1$. The detection performance, characterized by the probability of error of the system, is given as

$$P_e^{(En)} = \pi_0 Q\left(\gamma_f^{(En)}\right) + \pi_1 Q\left(\gamma_m^{(En)}\right), \tag{34}$$

where $\gamma_f^{(En)} = \frac{\eta^{(En)} - \mu_0^{(En)}}{\sigma_0^{(En)}}$ and $\gamma_m^{(En)} = \frac{\mu_1^{(En)} - \eta^{(En)}}{\sigma_1^{(En)}}$. Fig. 4 shows that the detection performance of the proposed scheme in terms of $\gamma_f^{(En)}$ and $\gamma_m^{(En)}$ is better than the detection performance of the traditional one, TAS, under both strategic attacks and non-strategic attacks. We can observe that the detection performance of TAS is the same as the direct scheme when the system is under strategic attacks ($p_2 = 0$). This is in accordance with the results shown in Theorem 1. However, the proposed EAS prevents it from happening. As shown in Fig. 4, the worst case from the perspective of the FC is that the strategic attackers take the attacking strategy of $p_1 = 1$ and $p_2 = 0$, i.e. , the Byzantine nodes always send falsified data to the MMSD and their group members and do not forge data from their group members. In this case, the proposed EAS has the same detection performance as the direct scheme. In the next section, another new scheme is proposed which achieves better detection performance and higher robustness compared with EAS.



(a) γ_f as a function of flipping probability p_1 given $p_2 = 0$ and $p_2 = 0.2$.

(b) γ_m as a function of flipping probability p_1 given $p_2 = 0$ and $p_2 = 0.2$.

Fig. 4: The probability of error is characterized by the argument of function $Q(.)$ for the probability of false alarm shown in (a) and the argument of function $Q(.)$ for the probability of miss detection shown in (b). Smaller values of the argument result in higher probabilities of error.

IV. Proposed Optimal Bayesian Fusion Rule

In this section, we propose a new framework and a new fusion rule for the audit bit based system. In this framework, we focus on the practical scenario in which the Byzantine nodes are in a minority due to the limited attacking resources, i.e., $\alpha_0 \leq 1/2$. We will first start with a network with one cluster, then we will move on to a wide-area network with multiple clusters.

A. A single-cluster network

As before, the sensors are partitioned into sets $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$ by the MMSD based on both status indicators of sensor i and sensor j in the same group. Moreover, the local decisions (u_i, u_j) sent from the same group are also compared to give us additional information about the behavioral identity of sensors in the networks. Each sensor again transmits its decision to the MMSD via two paths, namely the direct path and indirect path to the FC. After collecting all the local decisions, the MMSD places the sensors into sets $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$. These steps are the same as the ones in EAS. However, the MMSD also considers the MMS of the decisions u_i and u_j from the same group: if the sensor decisions for sensors i and j are the same, i. e., $u_i = u_j$, they are placed in the Set $\underline{\mathcal{M}}$ and the others are placed in the Set $\overline{\mathcal{M}}$. The MMSD only transmits the local decisions of the sensors with the sensor index i given by $\{i : (\underline{SS}\bigcap\underline{\mathcal{M}})\bigcup\underline{S}\overline{S}\bigcup\overline{S}\underline{S}\}$ to the decision making module to make the final decision. In other words, the local decisions from the sensors in Set $\underline{SS}\bigcap\overline{\mathcal{M}}$ or Set $\overline{SS}$ are not used to make the final decision which correspond to the two conditions stated as below.

Condition 1: The sensor i and its group member j are both in the set $\overline{SS}$.

Condition 2: The sensor i and its group member j are both in the set $\underline{SS}$ and $u_i \neq u_j$.

In the next lemma, we show the reasons why not using the decisions of sensors that satisfy one of the above two conditions improves the detection performance of the system.

Lemma 3:

1) When the sensor pair (i, j) satisfies Condition 1, i. e., sensors i and$j$ belong to $\overline{SS}$, removing this sensor pair results in the removal of two Byzantine nodes when $p_2 = 0$.
2) When we remove the sensor pairs that satisfy Condition 2, the ability of removing the Byzantine nodes for the proposed RAS increases with the increase of p_1 given specific p_2 and α_0.

Proof:

1) Let E be the event that at least one node in sensor pair (i, j) is a Byzantine node. When $i, j \in \overline{SS}$, it is obvious that $P(E|i, j \in \overline{SS}) = 1$. Thus, we can obtain $P(i, j \notin \overline{SS}|\overline{E}) = 1$ due to the fact that the contrapositive of the conditional statement is also true. So we can conclude that there is at least one Byzantine node in the sensor pair. Moreover, it is easy to conclude that all the sensors are Byzantine nodes in the Set $\overline{SS}$ when the attackers take the strategy of $p_2 = 0$ according to (25). Thus, removing the decisions of sensors in this set can remove at least one Byzantine node in each pair, and it can even remove two Byzantine nodes in each pair when the attackers employ the strategy of $p_2 = 0$.
2) To evaluate the impact of removing the unequal local decisions of sensor pairs on the performance of removing Byzantine nodes, we utilize the ratio $F = \frac{P(E, u_i = u_j | i, j \in \underline{SS})}{P(E|i, j \in \underline{SS})}$ to characterize that performance. The numerator of ratio F is the probability of the joint event that there exists at least one Byzantine node and the event $u_i = u_j$ given $i, j \in \underline{SS}$. The denominator is the probability of at least one Byzantine node given $i, j \in \underline{SS}$. The ratio $F = P(u_i = u_j | i, j \in \underline{SS}, E)$ gives the probability of $u_i = u_j$ given event E and $i, j \in \underline{SS}$. We have

$$\begin{aligned} & P(E, u_i = u_j | i, j \in \underline{SS}) \\ & = P(E | u_i = u_j, i, j \in \underline{SS}) P(u_i = u_j | i, j \in \underline{SS}) && \text{(35a)} \\ & = (1 - P(i = H, j = H | i, j \in \underline{SS}, u_i = u_j)) \\ & \quad \times P(u_i = u_j, | i, j \in \underline{SS}) && \text{(35b)} \\ & = P(u_i = u_j | i, j \in \underline{SS}) - \frac{(1 - \alpha_0)^2}{P(i, j \in \underline{SS})} \\ & \quad \times P(u_i = u_j | i, j \in \underline{SS}, i = H, j = H) && \text{(35c)} \end{aligned}$$

and

$$\begin{aligned} & P(E | i, j \in \underline{SS}) \\ & = 1 - P(i = H, j = H | i, j \in \underline{SS}) && \text{(36a)} \\ & = 1 - \frac{(1 - \alpha_0)^2}{P(i, j \in \underline{SS})}, && \text{(36b)} \end{aligned}$$

where $P(u_i = u_j | i, j \in \underline{SS}) = P(u_i = u_j | i, j \in \underline{SS}, \mathcal{H}_0) P(\mathcal{H}_0) + P(u_i = u_j | i, j \in \underline{SS}, \mathcal{H}_1) P(\mathcal{H}_1) = \pi_1[\pi_{11,1}^2 + (1 - \pi_{11,1})^2] + \pi_0[\pi_{10,1}^2 + (1 - \pi_{10,1})^2]$ and $P(u_i = u_j | i, j \in \underline{SS}, i = H, j = H) = [P_d^2 + (1 - P_d)^2]\pi_1 + [P_f^2 + (1 - P_f)^2]\pi_0$.

■

The relationship among p_1, p_2, α_0 and F is shown in Fig. 5. Note that a small F means a lower probability of existence of Byzantine nodes in the sensor pair given $i, j \in \underline{\mathcal{M}}\bigcap\underline{SS}$. We can observe from Fig. 5 that the value of F has a significant decrease when p_1 is large. It can also be observed that the value of F decreases with the increase of α_0 given $p_1 \geq 0.5$ and p_2, and it slightly changes with different α_0 and p_2 given $p_1 < 0.5$. From the analysis in Section II, it is evident that p_1 affects the final decision making by mainly affecting the local decisions (u_i) used to make the final decision, while both α_0 and p_2 only affect the final decision making by affecting the evaluated probability of one sensor being a Byzantine node (α_1, α_2, α_3 or α_4). Intuitively, changing p_1 has greater effect on the final decision making. Hence, when $p_1 < 0.5$, p_1 is not large enough to enable us to observe a distinct difference in F for different p_2 and α_0. In general, Fig. 5 shows that the ability of removing the Byzantine nodes increases with the increase of p_1 for a given p_2 by removing the sensor pairs which satisfy Condition 2.

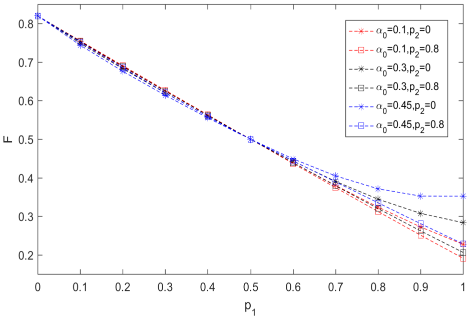


Fig. 5: F versus p_1 given $p_2 = 0.1$ for different α_0 and $N = 100$.

According to Theorem 1, the attackers' optimal attacking strategy in TAS is to choose $p_2 = 0$. In the scenario where p_2 is very small (close to 0), however, Fig. 2 has shown that the detection performance of TAS significantly degrades for a large value of p_1. The proposed scheme in this section achieves better detection performance compared with TAS when the attackers adopt the strategy of $p_2 = 0$ with $\forall p_1 \in [0, 1]$. It is because when p_2 is small, the Byzantine nodes have high probabilities of being placed in the set $\underline{SS}$ in our proposed scheme. If the attacker chooses p_1 to be large, there is a high probability that the group containing a Byzantine node satisfies Condition 2. Hence, the decision of the Byzantine node is likely to be blocked by the MMSD and not transmitted to the FC. As a result, our scheme prevents the attacker from designing p_1 to be very large and p_2 to be very small. On the other hand, when p_1 is not so large, each Byzantine node has a relatively higher probability, i.e., $1 - p_1$, to act honestly. Through such a trade off, the detection accuracy of the proposed scheme outperforms TAS under strategic attacks.

Based on the analysis above, we can show that the proposed scheme can effectively remove the decisions coming from Byzantine nodes. Hence, in the proposed RAS, we have the following relations for sensor i.

$$P(u_i = 1 | i \in \underline{S}\overline{S}, \mathcal{H}_q) = \pi_{1q,2} \tag{37a}$$

$$P(u_i = 1 | i \in \overline{S}\underline{S}, \mathcal{H}_q) = \pi_{1q,3} \tag{37b}$$

where $q = 0, 1$. Although u_i and u_j are dependent given $i, j \in \underline{SS} \bigcap \underline{\mathcal{M}}$, they are independent given $i, j \in \underline{SS}$. Hence, we have

$$\begin{aligned} &P(u_i = 1, u_j = 1 | i, j \in \underline{SS} \bigcap \underline{\mathcal{M}}, \mathcal{H}_q) \\ =& \frac{P(u_i = 1 | i, j \in \underline{SS}, \mathcal{H}_q) P(u_j = 1 | i, j \in \underline{SS}, \mathcal{H}_q)}{P(u_i = u_j | i, j \in \underline{SS}, \mathcal{H}_q)} \\ =& \frac{\underline{\pi}_{1q}^2}{\underline{\pi}_{1q}^2 + (1 - \underline{\pi}_{1q})^2} = \pi_{1q,5} \end{aligned} \tag{38}$$

for $q = 0, 1$. To simplify the analysis, we consider the group votes instead of the individual votes for the sensors in set $\underline{SS} \bigcap \underline{\mathcal{M}}$. Let z_g denote the group vote for group $g \in \underline{T}$, where $\underline{T}$ is the set of group whose sensors are in set $\underline{SS} \bigcap \underline{\mathcal{M}}$. Due to the fact that the sensors in the same group in set $\underline{SS} \bigcap \underline{\mathcal{M}}$ has the same decisions, we have $z_g = \{0, 2\}$. Hence, we obtain the following pdfs

$$f(u_i | \mathcal{H}_q) = \begin{cases} \pi_{1q,2}^{u_i} (1 - \pi_{1q,2})^{1 - u_i} & \text{for } i \in \underline{S}\overline{S} \\ \pi_{1q,3}^{u_i} (1 - \pi_{1q,3})^{1 - u_i} & \text{for } i \in \overline{S}\underline{S} \end{cases} \tag{39}$$

for sensor $i \in \underline{S}\overline{S} \bigcup \underline{S}\overline{S}$, and

$$f(z_g | \mathcal{H}_q) = \pi_{1q,5}^{z_g/2} (1 - \pi_{1q,5})^{1 - z_g/2} \tag{40}$$

for group $g \in \underline{T}$, where $q = 0, 1$. Thus, the proposed new decision rule is shown in Theorem 2.

Theorem 2: The new optimal decision rule, given the Byzantine flipping probabilities p_1, p_2 and α_0 fraction of Byzantine nodes, is expressed as

$$W_5 \sum_{g \in \underline{T}} \frac{z_g}{2} + W_2 \sum_{i \in \underline{S}\overline{S}} u_i + W_3 \sum_{i \in \overline{S}\underline{S}} u_i \gtrless \eta^{(RA)}, \tag{41}$$

where $W_2 = \log(\frac{\pi_{11,2}(1-\pi_{10,2})}{\pi_{10,2}(1-\pi_{11,2})})$, $W_3 = \log(\frac{\pi_{11,3}(1-\pi_{10,3})}{\pi_{10,3}(1-\pi_{11,3})})$, $\eta^{(RA)} = \log(\frac{\pi_0}{\pi_1}) + N_{re}^{LL} \log(\frac{1-\pi_{10,5}}{1-\pi_{11,5}}) + N_{re}^{L} \log(\frac{1-\pi_{10,2}}{1-\pi_{11,2}}) + N_{re}^{U} \log(\frac{1-\pi_{10,3}}{1-\pi_{11,3}})$. N_{re}^L, N_{re}^U and N_{re}^{LL} are the cardinalities of sets $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\underline{T}$, respectively, where $N_{re}^L = |\underline{S}\overline{S}|$,$N_{re}^U = |\overline{S}\underline{S}|$, and $N_{re}^{LL} = |\underline{T}|$. W_5 denotes the rearranged weight for group decisions in set $\underline{T}$ which is given as

$$W_5 = \frac{\pi_{11,5}(1 - \pi_{10,5})}{\pi_{10,5}(1 - \pi_{11,5})}. \tag{42}$$

Proof: We know that all groups of sensors whose decisions are sent to the FC are elements of one of the three sets $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\underline{SS} \bigcap \underline{\mathcal{M}}$. Thus, the optimal decision rule is given as (43) due to the fact that the sensors in sets $\underline{S}\overline{S}$ or $\overline{S}\underline{S}$ independently send their local decisions to the FC given the hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$. Even though the decisions coming from the sensors in the same group in set $\underline{SS} \bigcap \underline{\mathcal{M}}$ are dependent, the group votes are independent of each other. Hence, the optimal decision rule can be reformulated as (44). Substituting (37), (38), (39), (40) in (44), and taking the logarithm on both sides, we can get the fusion rule stated in the theorem.

$$\prod_{g \in \underline{T}} \frac{P(z_g | \mathcal{H}_1)}{P(z_g | \mathcal{H}_0)} \prod_{i \in \underline{S}\overline{S}} \frac{P(u_i | \mathcal{H}_1)}{P(u_i | \mathcal{H}_0)} \prod_{i \in \overline{S}\underline{S}} \frac{P(u_i | \mathcal{H}_1)}{P(u_i | \mathcal{H}_0)} \gtrless \frac{\pi_0}{\pi_1} \tag{44}$$

■

Let U denote the left-hand side of the optimal decision rule in (41) which is given as

$$U = W_5 U_5 + W_2 U_2 + W_3 U_3, \tag{45}$$

where $U_5 = \sum_{g \in \underline{T}} z_g/2$, $U_2 = \sum_{i \in \underline{S}\overline{S}} u_i$ and $U_3 = \sum_{i \in \overline{S}\underline{S}} u_i$. U_2 and U_3 are all Binomial distributed variables and U_5 is equivalent to a Binomial distributed variable. When N is large, the expected number of sensors in $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and the expected number of groups in $\underline{T}$ are $NP(i \in \underline{S}\overline{S})$, $NP(i \in \overline{S}\underline{S})$ and $GP(u_i = u_j | i, \underline{j} \in \underline{SS}) P(i, j \in \underline{SS})$, respectively. $P(i \in \underline{S}\overline{S})$ and $P(i \in \overline{S}\underline{S})$ are defined in (33), and $P(i, j \in \underline{SS})$ is defined in (20). $P(u_i = u_j | i, j \in \underline{SS})$ is given as

$$\begin{aligned} P(u_i = u_j | i, j \in \underline{SS}) = \sum_{q=0,1} P(\mathcal{H}_q) \sum_{t=0,1} & P(u_i = t | i \in \underline{SS}, \mathcal{H}_q) \\ & P(u_j = t | j \in \underline{SS}, \mathcal{H}_q) && \text{(46a)} \\ =& (\underline{\pi}_{11}^2 + (1 - \underline{\pi}_{11})^2) \pi_1 \\ & + (\underline{\pi}_{10}^2 + (1 - \underline{\pi}_{10})^2) \pi_0 && \text{(46b)} \end{aligned}$$

Hence, U, which is the sum of Binomial distributed variables, can be approximated as the Gaussian distribution with parameters as follows:

$$\begin{aligned} \mu_m^{(RA)} =& E[U | \mathcal{H}_m] \\ =& GP(u_i = u_j | i, j \in \underline{SS}) P(i, j \in \underline{SS}) \pi_{1m,5} W_5 \\ & + N(P(i \in \overline{S}\underline{S}) \pi_{1m,3} W_3 + P(i \in \underline{S}\overline{S}) \pi_{1m,2} W_2) \end{aligned} \tag{47a}$$

$$\begin{aligned} (\sigma_m^2)^{(RA)} =& Var[U | \mathcal{H}_m] \\ =& GP(u_i = u_j | i, j \in \underline{SS}) P(i, j \in \underline{SS}) \pi_{1m,5} \\ & (1 - \pi_{1m,5}) W_5^2 + N(P(i \in \overline{S}\underline{S}) \pi_{1m,3} (1 - \pi_{1m,3}) W_3^2 \\ & + P(i \in \underline{S}\overline{S}) \pi_{1m,2} (1 - \pi_{1m,2}) W_2^2), \end{aligned} \tag{47b}$$

$$\prod_{i,j\in\underline{SS}\bigcap\underline{\mathcal{M}}}\frac{P(u_i,u_j|\mathcal{H}_1)}{P(u_i,u_j|\mathcal{H}_0)}\prod_{i\in\underline{S}\overline{S}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\prod_{i\in\overline{S}\underline{S}}\frac{P(u_i|\mathcal{H}_1)}{P(u_i|\mathcal{H}_0)}\gtrless\frac{\pi_0}{\pi_1} \quad (43)$$

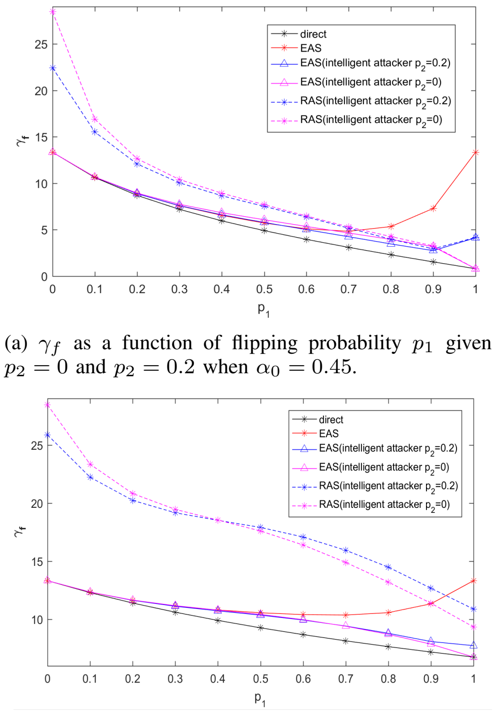
for $m = 0, 1$. The threshold η for large N is given as

$$\eta^{(RA)} = \log(\frac{\pi_0}{\pi_1}) + E(N_{re}^{LL})\log(\frac{1-\pi_{10,5}}{1-\pi_{11,5}}) + E(N_{re}^{L})\log(\frac{1-\pi_{10,2}}{1-\pi_{11,2}}) + E(N_{re}^{U})\log(\frac{1-\pi_{10,3}}{1-\pi_{11,3}}), \quad (48)$$

where $E(N_{re}^L) = NP(i \in \underline{S}\overline{S})$, $E(N_{re}^U) = NP(i \in \overline{S}\underline{S})$ and $E(N_{re}^{LL}) = GP(u_i = u_j | i, j \in \underline{SS})$. Thus, the probability of error $P_e^{(RA)}$ for the system is expressed as

$$P_e^{(RA)} = \pi_0 Q\left(\gamma_f^{(RA)}\right) + \pi_1 Q\left(\gamma_m^{(RA)}\right), \quad (49)$$

where $\gamma_f^{(RA)} = \frac{\eta^{(RA)}-\mu_0^{(RA)}}{\sigma_0^{(RA)}}$ and $\gamma_m^{(RA)} = \frac{\mu_1^{(RA)}-\eta^{(RA)}}{\sigma_1^{(RA)}}$ is the argument of function $Q(.)$ for the probability of false alarm and the argument of function $Q(.)$ for the probability of miss detection for the new proposed fusion rule. Fig. 6 shows how argument $\gamma_f^{(RA)}$ changes with p_1 given specific p_2 and α_0 when $N = 100$, $P_d = 0.9$ and $P_f = 0.1$. We can observe that the argument $\gamma_f^{(RA)}$ of RAS is larger than that of EAS under strategic attacks. Since the argument $\gamma_m^{(RA)}$ has similar properties, we only include the simulation results of $\gamma_f^{(RA)}$ in the paper. Note that the larger arguments mean better detection performance. Fig. 7 shows how the probabilities of detection and false alarm of the system change with p_1 given specific $p_2 = 0$ and $\alpha_0 = 0.3$ when $N = 10$, $P_d = 0.9$ and $P_f = 0.1$. From Fig. 6 and Fig. 7, we can observe that our proposed RAS has a significant improvement on the detection performance of the system when α_0 is small. Even though the detection performance of the proposed scheme gets close to EAS when α_0 approaches 0.5 and p_1 is large, the proposed RAS still outperforms EAS and the direct scheme. This improvement becomes more prominent when p_1 is relatively small. Moreover, in both EAS and RAS, a large p_1 can always make it harder for Byzantine nodes to evade the detection system. In this case, the FC has the history of all the local decisions it received in the past to identify Byzantine nodes. And some reputation-based schemes can help the FC to identify the Byzantine nodes [17] [31].



(a) γ_f as a function of flipping probability p_1 given $p_2 = 0$ and $p_2 = 0.2$ when $\alpha_0 = 0.45$.

(b) γ_f as a function of flipping probability p_1 given $p_2 = 0$ and $p_2 = 0.2$ when $\alpha_0 = 0.15$.

Fig. 6: The argument for the probability of false alarm function for different values of α_0.

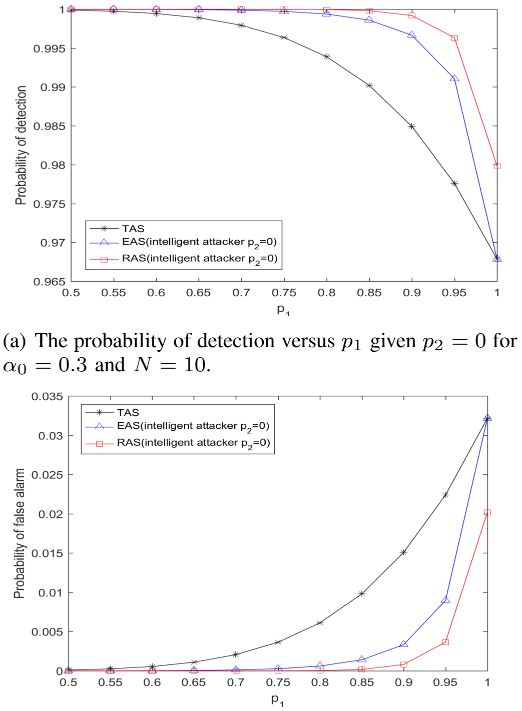
B. The network with multiple clusters

In this subsection, we extend our work from the single cluster case to the case of multiple clusters in the wide-area network. We show that the proposed RAS can not only improve the detection performance of the system, but also reduce the communication overhead[5] between the clusters and the FC. In a cluster based network as shown in Fig. 8, the N sensors in the network are grouped into T clusters and the sensors in each cluster are further divided into groups of two. Each cluster is equipped with one MMSD which serves as a data integration processor for this cluster. Note that the MMSD is no longer a part of the FC.

Based on the local observations, each sensor makes a binary decision regarding the absence or presence of the PoI. Then, the sensors send both their own decisions and their group member's decision to the corresponding MMSDs. By comparing the MMS of the direct and indirect decisions, the MMSDs are able to obtain the status indicators for all the sensors in the corresponding clusters. Based on these status indicators, each MMSD partitions the sensors in the cluster into sets $\underline{SS}$, $\underline{S}\overline{S}$, $\overline{S}\underline{S}$ and $\overline{SS}$. In addition, the sensors are placed into $\underline{\mathcal{M}}$ if the local decisions of the sensors in the same group are the same. The flow chart to illustrate the decision making and communication process of a cluster $t \in \{1, \ldots, T\}$ is shown in Fig. 9.

Let $N_t^{(RA)}$ and $N_t^{(A)}$ denote the number of local decisions sent by the MMSDs to the FC for the proposed RAS and the number of local decisions sent by the sensors to the FC, respectively. Note that the MMSDs only transmit the direct decisions, and they do not transmit the ones that satisfy Condition 1 or Condition 2. Thus, the number of direct decisions $N_t^{(RA)}$ sent by the MMSDs to the FC is smaller than that of TAS $N_t^{(A)}$, where $N_t^{(RA)} = |\underline{SS}\bigcap\underline{\mathcal{M}}| + |\underline{S}\overline{S}| + |\overline{S}\underline{S}|$

[5] In this paper, we measure the overall communication overhead of the system by the number of bits in all communication messages sent.



(a) The probability of detection versus p_1 given $p_2 = 0$ for $\alpha_0 = 0.3$ and $N = 10$.

(b) The probability of false alarm versus p_1 given $p_2 = 0$ for $\alpha_0 = 0.3$ and $N = 10$.

Fig. 7: The probability of false alarm and the probability of detection for the system.

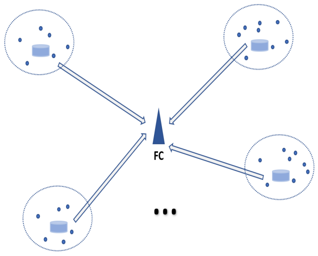


Fig. 8: System model of a distributed CWSN. The blue cylinders represent MMSDs in each cluster and the small blue circles represent low-cost sensors.

and $N_t^{(A)} = 2N$. Let r represent different sets as follows. If $r = 00$, it refers to the set $\underline{SS} \bigcap \underline{\mathcal{M}}$; If $r = 01$ it refers to the set $\underline{S}\overline{S}$; If $r = 10$, it refers to the set $\overline{S}\underline{S}$. Each MMSD sends three data packets which contain r and the direct decisions from the sensors in the sets $\underline{SS} \bigcap \underline{\mathcal{M}}$, $\underline{S}\overline{S}$ and $\overline{S}\underline{S}$, respectively. For example, if sensor 1 to sensor 4 are in $\underline{SS} \bigcap \underline{\mathcal{M}}$, sensor 5 to sensor 8 are in $\overline{S}\underline{S}$ and sensor 9 to sensor 12 are in $\underline{S}\overline{S}$. The three data packets contain $[r = 00, u_1, \ldots, u_4]$, $[r = 10, u_5, \ldots, u_8]$ and $[r = 01, u_9, \ldots, u_{12}]$. Upon receiving these data packets, the FC is able to determine which sets those sensors belong to so that it can make the final decision based on those transmitted direct decisions.

When N is large, we are able to calculate the expected number of bits transmitted to the FC from all the MMSDs, which is $E(N_t^{(RA)}) = E(N_{re}^{LL}) + E(N_{re}^{L}) + E(N_{re}^{U})$, according to (48). Fig. 10 shows the expected number of bits transmitted to the FC when $N = 100$ and $N_t^{(A)} = 2N = 200$. We can observe that the expected number of bits transmitted to the FC for the proposed RAS significantly decreases compared with the one for TAS. It is due to fact that the MMSDs only send the direct decisions of sensors which do not satisfy Condition 1 or Condition2. We can also observed that the expected number of bits decreases with an increased α_0 given a specific p_2. It is due to the fact that the number of sensors temporarily removed by the MMSDs increases when the fraction of Byzantine nodes α_0 increases with a given attacking probability p_2. Hence, the proposed new fusion rule is able to reduce the energy cost of the sensors to half of the traditional case which prolongs the lifetime of the network, especially for the wide area network.

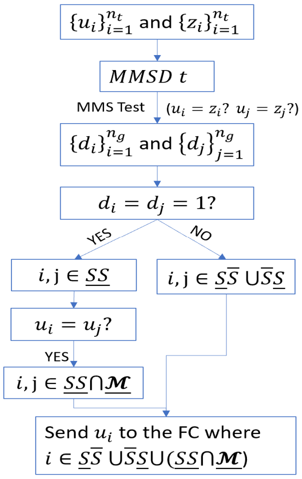

$\{u_i\}_{i=1}^{n_t}$ and $\{z_i\}_{i=1}^{n_t}$
$MMSD\ t$
MMS Test
$(u_i = z_i?\ u_j = z_j?)$
$\{d_i\}_{i=1}^{n_g}$ and $\{d_j\}_{j=1}^{n_g}$
$d_i = d_j = 1?$
YES
NO
$i, j \in \underline{SS}$
$i, j \in \underline{S}\overline{S} \cup \overline{S}\underline{S}$
$u_i = u_j?$
YES
$i, j \in \underline{SS} \cap \underline{\mathcal{M}}$
Send u_i to the FC where
$i \in \underline{S}\overline{S} \cup \overline{S}\underline{S} \cup (\underline{SS} \cap \underline{\mathcal{M}})$

Fig. 9: The flow chart of the decision making and communication processes of cluster t. n_t is the number of sensors in cluster t. Sensor j is the group member of sensor i.

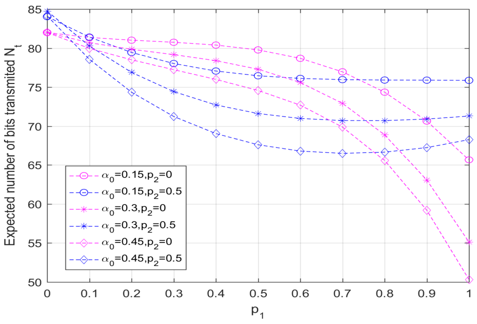


Fig. 10: The expected number of bits transmitted to the FC N_t versus p_1 given different value of α_0 and p_2.

V. CONCLUSION

In this work, an audit based mechanism was utilized to mitigate the effect of Byzantine attacks in the networks. Instead of employing the identical attacking strategy of TAS where each sensor utilizes the same attacking probability to falsify the decisions coming from their group member and its own decision, we considered strategic attackers that can use different attacking strategies. We showed that it was possible for the strategic attackers to blind the FC as far as the information conveyed by the audit bits in TAS is concerned. To overcome this problem, we proposed an enhanced audit bit based scheme, namely EAS. Our results showed that the proposed scheme outperforms TAS. Furthermore, we proposed a reduced audit bit based scheme (RAS) based on our proposed EAS. We showed that RAS is able to further improve the robustness and the detection performance of the system. We extended our work for the wide-area CWSNs. In wide-area cluster-based WSNs, we showed that the proposed RAS is able to significantly reduce the communication overhead between the clusters and the FC.

In the future, we intend to consider the scenarios where the FC has no prior knowledge of the attacking strategy of Byzantine nodes, i.e. p_1 and p_2, and propose algorithms to defend against Byzantine attacks under this assumption. Unlike the assumption made in this work that all the CHs are trustworthy, we will also consider the scenarios where the CHs (or MMSDs) could also be compromised.

Appendix A
Proof of Theorem 1

Instead of directly analyzing the property of P_e^I in terms of p_2, we utilize Bhattacharyya distance $\mathcal{BD}$ as a surrogate to asymptotically characterize the detection performance of the system for simplicity. The relationship between Bhattacharyya distance and the probability of error P_e^I is $\lim_{N\to\infty} \frac{ln(P_e^I)}{N} \leq \mathcal{BD}$. For discrete probability distribution, $\mathcal{BD} = \sum_{\mathbf{u}\in\mathcal{U}} -ln\sqrt{P(\mathbf{u}|\mathcal{H}_1)P(\mathbf{u}|\mathcal{H}_0)}$, where $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{2^N}\}$ is the set of all the possible realizations of vector $\mathbf{u} = [u_1, u_2, \ldots, u_N]$. Let $f_i(u_i|i \in \underline{S}) = P(u_i|\mathcal{H}_1, i \in \underline{S})P(u_i|\mathcal{H}_0, i \in \underline{S})$ and $f_i(u_i|i \in \overline{S}) = P(u_i|\mathcal{H}_1, i \in \overline{S})P(u_i|\mathcal{H}_0, i \in \overline{S})$. Due to the fact that sensors independently send their local decisions, $\mathcal{BD}$ is given as

$$\begin{aligned}
\mathcal{BD} &= \sum_{\mathbf{u}\in\mathcal{U}} -ln\sqrt{\prod_{i\in\underline{S}} f_i(u_i|i\in\underline{S}) \prod_{i\in\overline{S}} f_i(u_i|i\in\overline{S})} \\
&= \sum_{\mathbf{u}\in\mathcal{U}} -ln\sqrt{\prod_{i=1}^{N} \mathcal{F}_i(u_i)} \\
&= \sum_{\mathbf{u}\in\mathcal{U}} -ln\sqrt{\prod_{i=1}^{N} \left(\sum_{d_i\in\mathcal{Q}} \mathcal{F}_i(u_i|d_i)P(d_i) \right)} \\
&= \sum_{\mathbf{u}\in\mathcal{U}} -ln\sqrt{\prod_{i=1}^{N} E_{d_i}\{\mathcal{F}_i(u_i|d_i)\}}
\end{aligned} \tag{50}$$

where $\mathcal{Q} = \{0, 1\}$, $\mathbf{d} = [d_1, d_2, \ldots, d_N]$ and $d_i \in \mathcal{Q}$. $\mathcal{F}_i(u_i|d_i) = (\overline{\pi}_{11}^{u_i}(1-\overline{\pi}_{11})^{1-u_i}\overline{\pi}_{10}^{u_i}(1-\overline{\pi}_{10})^{1-u_i})^{1-d_i}(\underline{\pi}_{11}^{u_i}(1-\underline{\pi}_{11})^{1-u_i}\underline{\pi}_{10}^{u_i}(1-\underline{\pi}_{10})^{1-u_i})^{d_i}$. $d_i = 1$ indicates that the sensor i is placed in Set $\underline{S}$, otherwise, it is placed in Set $\overline{S}$. For sensor i, $E_{d_i}\{\mathcal{F}(u_i|d_i)\}$ is given as

$$\begin{aligned}
&E_{d_i}\{\mathcal{F}(u_i|d_i)\} \\
&\quad = \sum_{q=0,1} \mathcal{F}(u_i|d_i = q)P(d_i = q) \\
&\quad = \overline{\pi}_{11}^{u_i}(1-\overline{\pi}_{11})^{1-u_i}\overline{\pi}_{10}^{u_i}(1-\overline{\pi}_{10})^{1-u_i}P(d_i = 1) \\
&\qquad + \underline{\pi}_{11}^{u_i}(1-\underline{\pi}_{11})^{1-u_i}\underline{\pi}_{10}^{u_i}(1-\underline{\pi}_{10})^{1-u_i}P(d_i = 0).
\end{aligned} \tag{51}$$

We now have following two cases:

a) $u_i = 1$: In this case, $E_{d_i}\{\mathcal{F}(u_i|d_i)\} = \overline{\pi}_{11}\overline{\pi}_{10}P(d_i = 1) + \underline{\pi}_{11}\underline{\pi}_{10}P(d_i = 0)$. We know that $P(d_i = 1) + P(d_i = 0) = 1$ and $\underline{\alpha}^I \leq \alpha_0 \leq \overline{\alpha}^I$. Let $h(t) = \pi_{11}\pi_{10}$ where $t = \alpha p_1$ is the random variable here. We can obtain $\frac{\partial^2 h(t)}{t^2} = 2(1-2P_d)(1-2P_f) < 0$. Hence, $h(t)$ is a concave function and has the property as following.

$$\begin{aligned}
&P(d_i = 1)h(t_1) + P(d_i = 0)h(t_2) \\
&\leq h(P(d_i = 1)t_1 + P(d_i = 0)t_2) = h(t_0)
\end{aligned} \tag{52}$$

where $t_1 = \overline{\alpha}^I p_1$, $t_2 = \underline{\alpha}^I p_1$ and $t_0 = \alpha_0 p_1$.

b) $u_i = 0$: In this case, $E_{d_i}\{\mathcal{F}(u_i|d_i)\} = (1-\overline{\pi}_{11})(1-\overline{\pi}_{10})P(d_i = 1) + (1-\underline{\pi}_{11})(1-\underline{\pi}_{10})P(d_i = 0)$. Let $g(t) = (1-\pi_{11})(1-\pi_{10})$ where $t = \alpha p_1$ is the random variable here. We can obtain $\frac{\partial^2 g(t)}{t^2} = 2(1-2P_d)(1-2P_f) < 0$. Hence, $g(t)$ is also a concave function and follows the similar property as (52).

Note that we have $\underline{\pi}_{11} = \overline{\pi}_{11} = \pi_{11}$ and $\underline{\pi}_{10} = \overline{\pi}_{10} = \pi_{10}$ when $p_2 = 0$ according to Lemma 1. We can conclude that $E_{d_i}\{\mathcal{F}(u_i|d_i)\} \leq \mathcal{F}^0(u_i)$, where $\mathcal{F}^0(u_i) = \pi_{11}^{u_i}(1-\pi_{11})^{1-u_i}\pi_{10}^{u_i}(1-\pi_{10})^{1-u_i}$. We call the grouping in TAS with $p_2 = 0$ as non-effective grouping which is the same as the direct scheme, i.e., $\underline{\alpha}^I = \alpha_0 = \overline{\alpha}^I$, and the grouping in TAS with $p_2 \neq 0$ as effective grouping. According to (52), We show that the Bhattacharyya distance of the effective grouping is always larger than that of the non-effective grouping. According to the analysis above, the detection error $P_e^{(I)}$ can achieve the maximum value when $p_2 = 0$ given specific α_0, P_d, P_f and p_1. The probability of error for the system with direct scheme is

$$P_e^{(D)} = \pi_0 Q\left(\gamma_f^{(D)}\right) + \pi_1 Q\left(\gamma_m^{(D)}\right), \tag{53}$$

where $\gamma_f^{(D)}$ and $\gamma_m^{(D)}$ are expressed, respectively, as

$$\gamma_f^{(D)} = Q\left(\frac{\log(\frac{\pi_0}{\pi_1})/\sqrt{N} + \sqrt{N}D_0(\alpha_0, p)}{\sqrt{\pi_{10}(1-\pi_{10})W_d^2}}\right) \tag{54a}$$

$$\gamma_m^{(D)} = Q\left(\frac{\log(\frac{\pi_0}{\pi_1})/\sqrt{N} + \sqrt{N}D_1(\alpha_0, p)}{\sqrt{\pi_{11}(1-\pi_{11})W_d^2}}\right), \tag{54b}$$

and, $D_0(\alpha_0, p) = \pi_{10}\log(\frac{\pi_{10}}{\pi_{11}}) + (1-\pi_{10})\log(\frac{1-\pi_{10}}{1-\pi_{11}})$, $D_1(\alpha_0, p) = \pi_{11}\log(\frac{\pi_{11}}{\pi_{10}}) + (1-\pi_{11})\log(\frac{1-\pi_{11}}{1-\pi_{10}})$ and $W_d = \log(\frac{\pi_{11}(1-\pi_{10})}{\pi_{10}(1-\pi_{11})})$. Thus, for the non-effective grouping, according to (54), $D_0(\alpha_0, p) = 0$ can make the system be totally blind when N is large enough. We can easily obtain that $D_0(\alpha_0, p) = 0$ when $\alpha_0 p = \frac{1}{2}$.

References

[1] P. K. Varshney, *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[2] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1958, pp. 100–117, 2012.

[3] D. Ciuonzo, P. S. Rossi, and P. K. Varshney, "Distributed detection in wireless sensor networks under multiplicative fading via generalized score tests," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9059–9071, 2021.

[4] R. P. Jover, J. Lackey, and A. Raghavan, "Enhancing the security of lte networks against jamming attacks," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–14, 2014.

[5] K. Gai, M. Qiu, Z. Ming, H. Zhao, and L. Qiu, "Spoofing-jamming attack strategy using optimal power distributions in wireless smart grid networks," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2431–2439, 2017.

[6] D. Ciuonzo, A. Aubry, and V. Carotenuto, "Rician MIMO channel- and jamming-aware decision fusion," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3866–3880, 2017.

[7] C. Quan, B. Geng, and P. K. Varshney, "On strategic jamming in distributed detection networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4760–4764.

[8] L. Xiao, D. Xu, N. B. Mandayam, and H. V. Poor, "Attacker-centric view of a detection game against advanced persistent threats," *IEEE Transactions on Mobile Computing*, vol. 17, no. 11, pp. 2512–2523, 2018.

[9] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, "Byzantine attack and defense in cognitive radio networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1342–1363, 2015.

[10] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," in *Concurrency: the Works of Leslie Lamport*, 2019, pp. 203–226.

[11] F. Penna, Y. Sun, L. Dolecek, and D. Cabric, "Detecting and counteracting statistical attacks in cooperative spectrum sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1806–1822, 2011.

[12] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of Byzantines," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 2925–2929.

[13] W. Wang, L. Chen, K. G. Shin, and L. Duan, "Secure cooperative spectrum sensing and access against intelligent malicious behaviors," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 1267–1275.

[14] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "On covert data falsification attacks on distributed detection systems," in *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2013, pp. 412–417.

[15] A. Vempaty, P. Ray, and P. K. Varshney, "False discovery rate based distributed detection in the presence of Byzantines," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 1826–1840, 2014.

[16] H. Li and Z. Han, "Catch me if you can: An abnormality detection approach for collaborative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3554–3565, 2010.

[17] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, 2010.

[18] L. Zhang, G. Nie, G. Ding, Q. Wu, Z. Zhang, and Z. Han, "Byzantine attacker identification in collaborative spectrum sensing: A robust defense framework," *IEEE Transactions on Mobile Computing*, vol. 18, no. 9, pp. 1992–2004, 2018.

[19] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Distributed Bayesian detection in the presence of Byzantine data," *IEEE transactions on signal processing*, vol. 63, no. 19, pp. 5250–5263, 2015.

[20] W. Hashlamoun, S. Brahma, and P. K. Varshney, "Mitigation of Byzantine attacks on distributed detection systems using audit bits," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 18–32, 2017.

[21] ——, "Audit bit based distributed Bayesian detection in the presence of Byzantines," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 4, pp. 643–655, 2018.

[22] S. Lindsey and C. S. Raghavendra, "Pegasis: Power-efficient gathering in sensor information systems," in *Proceedings, IEEE aerospace conference*, vol. 3. IEEE, 2002, pp. 3–3.

[23] A. Manjeshwar and D. P. Agrawal, "Teen: Arouting protocol for enhanced efficiency in wireless sensor networks." in *ipdps*, vol. 1, no. 2001, 2001, p. 189.

[24] E. Masazade, R. Niu, and P. K. Varshney, "Dynamic bit allocation for object tracking in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5048–5063, 2012.

[25] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 1–11, 2005.

[26] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *IEEE Military Communications Conference, 2003. MILCOM 2003.*, vol. 1. IEEE, 2003, pp. 688–693.

[27] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2010.

[28] J. P. Walters, Z. Liang, W. Shi, and V. Chaudhary, "Wireless sensor network security: A survey," *Security in distributed, grid, mobile, and pervasive computing*, vol. 1, no. 367, p. 6, 2007.

[29] A. Perrig, J. Stankovic, and D. Wagner, "Security in wireless sensor networks," *Communications of the ACM*, vol. 47, no. 6, pp. 53–57, 2004.

[30] C. Quan, B. Geng, Y. S. Han, and P. K. Varshney, "Enhanced audit bit based distributed Bayesian detection in the presence of strategic attacks," p. arXiv:2109.13325, 2021.

[31] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals and Systems*, vol. 1, no. 2, pp. 167–182, 1988.