Distributed Detection with Multiple Sensors in the Presence of Sybil Attacks

Wael Hashlamoun
Department of ECE
Birzeit University, Palestine
Email: hwael@birzeit.edu

Swastik Brahma
Department of CS
University of Cincinnati, USA
Email: brahmask@ucmail.uc.edu

Pramod K. Varshney
Department of EECS
Syracuse University, USA
Email: varshney@syr.edu

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001514

***Abstract*—This paper considers the problem of distributed detection in the presence of a Sybil attack where a malicious sensor node can send multiple falsified decisions using multiple fake identities to a Fusion Center (FC) to degrade its decision-making performance. We study the problem under the Neyman–Pearson (NP) setup. We find that, due to the Sybil attack, the decisions received at the FC become correlated and that the degree of correlation is dependent on the number of fake identities used. The paper characterizes the optimal Sybil attack that blinds the FC, i.e., makes the FC incapable of making an informed decision. We find that if the sum of the local detection and false alarm probabilities of the sensor nodes is 1, the FC can be made blind when at least 50% of the decisions are sent using fake identities. However, if this condition is not met, then all decisions would have to be sent using fake identities in order to blind the FC. The paper also investigates strategic interactions between the FC and the Sybil attacker using Game Theory and proves the existence of a Nash Equilibrium (NE). Numerical results are presented to gain important insights.**

***Index Terms*—Sensor Networks, Distributed Detection, Sybil Attack, Data Falsification, Game Theory.**

I. Introduction

The problem of performing distributed detection by fusing data from multiple sensors has been a well-studied topic [1]–[3]. In distributed detection systems, multiple sensor nodes observe a signal from a phenomenon of interest, make local decisions regarding the state of the phenomenon based on their observations, and then send their local decisions to a fusion center (FC) which fuses the received decisions to make a global decision regarding the phenomenon's state. Due to resource constraints, the local decisions made by the nodes are often 1-bit in nature. Distributed detection was originally motivated by its applications in military surveillance [2], but with the advent of the Internet-of-Things (IoT), is now being employed in a wide variety of applications, such as for inferring road and traffic conditions [4], societal-scale environmental monitoring [5], [6], and inferring dietary patterns [7]. Further, [8], [9] view social networks as sensing systems where humans act as sensors for enabling the detection of the state of a phenomenon, such as that of a sound event. Distributed spectrum sensing (DSS) in cognitive radio networks [10] is yet another example application of a distributed detection system.

This work was supported in part by the U.S. NSF under Award Number CCF-2047701 and in part by the Zamalah Program, Bank of Palestine.

The distributed nature of the sensor nodes in such systems makes them vulnerable to different types of cyber attacks [11], such as jamming attacks [12], Byzantine attacks [13]–[17], and eavesdropping attacks [18]. In fact, the resource constrained nature of the sensor nodes, such as their limited energy and computational capabilities, inhibits the use of sophisticated security solutions, like cryptographic techniques, which exacerbates security concerns. In recent years, security issues of such distributed networks are increasingly being studied within the networking [19], signal processing [20] and information theory communities [21].

The type of attack on distributed detection that we consider in this paper is a Sybil attack, originally described by [22] in the context of peer-to-peer networks. In a Sybil attack, one physical entity can present itself using multiple identities, thereby controlling a substantial fraction of system resources and undermining its performance. While past work has identified the vulnerability of sensory systems to Sybil attacks [23], [24], *to the best of our knowledge, the problem of performing distributed detection using such systems in the presence of Sybil attacks remains unexplored with a lack of analytical results on the topic. We aim to fill this void in this paper.*

In our model, we consider that a Sybil sensor node can launch an attack by simultaneously exploiting two degrees-of-freedom. First, we consider that a Sybil node can assume fake identities either by *fabricating* new identities or disabling legitimate nodes and *stealing* their identities [23], [24]. Second, we consider that the Sybil node can falsify its own local decision regarding the phenomenon's state to send multiple falsified decisions (using the fake identities) to the FC. We find that a distinctive feature of the Sybil attack is that it introduces correlations among elements of the decision vector received at the FC, with the degree of correlation being dependent on the number of fake identities used. While accounting for such correlations, under the Neyman-Pearson (NP) framework, we analytically characterize the optimal Sybil attack that makes the FC incapable of making an informed decision, in which case we say that the FC is blind, as well as analyze strategic interactions between the FC and the Sybil attacker using game theoretic tools. Our results indicate that Sybil attacks pose a severe threat to distributed detection systems, since even a single Sybil attacker can blind the FC. Following are the main contributions of the paper:

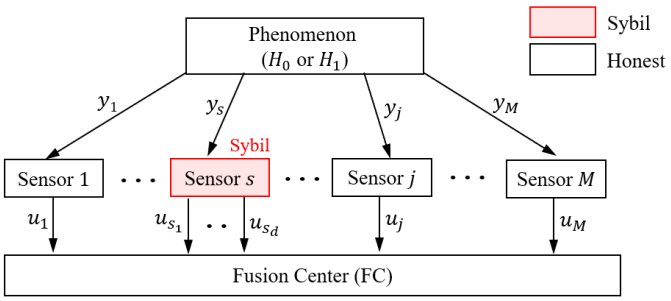

Fig. 1. Sybil attack on a distributed detection system.

- We study the problem of a Sybil attack on distributed detection, where a malicious sensor can assume multiple fake identities to transmit multiple potentially falsified local decisions to the FC to degrade its performance.
- We present a correlation structure to model the dependencies that the Sybil attack introduces into the decision vector received at the FC and characterize the degree of correlation that the attack introduces.
- We analytically characterize the optimal Sybil attack that blinds the FC to prevent it from making an informed decision.
- We perform game theoretic analysis of strategic attack-defense between the Sybil attacker and the FC and analytically prove the existence of an NE.

The rest of the paper is organized as follows. Section II models Sybil attacks on distributed detection. Section III presents a correlation structure for modeling the dependencies introduced by the Sybil attack. The optimal Sybil attack that blinds the FC is characterized in Section IV. Section V analyzes strategic attack-defense between the Sybil attacker and the FC using game theoretic tools. Finally, Section VI concludes the paper.

II. System Model

Consider a binary hypothesis testing problem with a phenomenon which can be in either one of two states, viz. H_0 or H_1. Consider also a sensor network, comprised of M sensors (nodes) and an FC. First, the sensors observe the phenomenon, carry out local computations to decide the phenomenon's state, and then send their local decisions to the FC. Finally, the FC makes a global decision regarding the phenomenon's state after processing the local decisions received from the sensors. Observations at the sensors are assumed to be conditionally independent and identically distributed (i.i.d) given the phenomenon's state. Communication channels are considered to be error-free.

A. Modus Operandi of the Nodes

Based on its observation y_i, each sensor $i \in \{1, \cdots, M\}$ makes a 1-bit local decision $v_i \in \{0, 1\}$ regarding the phenomenon's state using the likelihood ratio test (LRT) [1]:

$$\frac{p_{Y_i}^{(1)}(y_i)}{p_{Y_i}^{(0)}(y_i)} \underset{v_i=0}{\overset{v_i=1}{\gtrless}} \lambda, \tag{1}$$

where $p_{Y_i}^{(k)}(y_i)$ is the conditional probability density function (PDF) of observation y_i under state H_k, $k = 0, 1$, and λ is the identical threshold used at all the sensors for the LRT (use of identical thresholds is asymptotically optimal [25]). We denote the probabilities of detection and false alarm of each sensor i in the network by $P_d = P(v_i = 1|H_1)$ and $P_f = P(v_i = 1|H_0)$, respectively, which hold for every sensor in the network, irrespective or whether it is malicious or not.

In such a system, we consider the presence of a malicious Sybil sensor (denoted as sensor s) which assumes d fake identities as well as falsifies its local decision with a probability p to send d replicas of its potentially falsified decision to degrade the FC's decision-making performance (see Fig. 1). Thus, denoting the decision vector received at the FC as $\mathbf{u} = [u_1, u_2, \cdots, u_N]$, $\mathbf{u}$ contains $N = M - 1 + d$ decisions, out of which $M-1$ decisions are from non-malicious (honest) sensors and d decisions (viz. $[u_{s_1}, \cdots, u_{s_d}]$) are from the Sybil sensor, with $u_i = v_i$ if i is an honest node. We also consider that the d decisions in $\mathbf{u}$ from the Sybil node are flipped with a probability p, such that

$$\begin{aligned} p &= \text{Prob}(u_{s_1} = \cdots = u_{s_d} = b | v_s = a) \\ &= 1 - \text{Prob}(u_{s_1} = \cdots = u_{s_d} = a | v_s = a) \end{aligned} \tag{2}$$

is the probability that the Sybil node s sends $u_{s_1} = \cdots = u_{s_d} = b$ to the FC when its actual decision was a $(a, b = 0, 1)$.

B. The Fusion Center

Based on $\mathbf{u}$, the FC makes a global decision regarding the state of the phenomenon. From a practical viewpoint, we consider that the FC does *not* know which decisions in $\mathbf{u}$ belong to the Sybil node, but rather views every decision to have come from the Sybil node with the probability $\alpha = d/N$ (following the law of large numbers). Now, for an honest sensor i, we have

$$P(u_i = 1|H_1) = P_d \tag{3a}$$

$$P(u_i = 1|H_0) = P_f \tag{3b}$$

Further, since the Sybil node s sends d replicas of its potentially falsified decision to the FC, the joint probability mass functions (PMFs) of any subset containing w of the d decisions are

$$P(u_{s_1} = \cdots = u_{s_w} = 1|H_1) = P_d(1-p) + p(1-P_d) = P_d^{(s)} \tag{4a}$$

$$P(u_{s_1} = \cdots = u_{s_w} = 1|H_0) = P_f(1-p) + p(1-P_f) = P_f^{(s)} \tag{4b}$$

for $1 \leq w \leq d$, where p is the flipping probability of the Sybil node (2). To be able to make the decision optimally, employing (3) and (4), the FC has to construct the conditional PMFs of $\mathbf{u}$ under H_0 and H_1, respectively. This is, however, challenging due to the fact that, although the sensors' observations, viz. $y_1, y_2, \cdots, y_M$, are considered independent under both hypotheses, the Sybil attack, as described above, renders the elements of $\mathbf{u}$ dependent. We model the dependency among the elements of $\mathbf{u}$ in the next section.

III. The Correlation Model for the Sybil Attack

The Sybil attack, as discussed above, introduces correlation among the decisions in $\mathbf{u}$. In this section, we investigate the degree of correlation, its dependence on p and α, and its effect on the detection capability of the system. In addition, we introduce the Kullback-Leibler divergence, as a surrogate for the probability of error, to evaluate system performance.

A. The Pairwise Correlation Coefficient at the FC

We start by characterizing the correlation among pairs of decisions in $\mathbf{u}$.

LEMMA 1: *Let u_i and u_j be two decisions in $\mathbf{u}$. Then, the correlation coefficient, $\rho^{(k)}(u_i, u_j)$, between u_i and u_j under the phenomenon's state H_k, $k = 0, 1$, is given by*

$$\rho^{(k)} = \frac{\epsilon_2^{(k)} - (\epsilon_1^{(k)})^2}{\epsilon_1^{(k)}(1 - \epsilon_1^{(k)})} \tag{5}$$

where, $\epsilon_q^{(k)}$ is the q^{th} order joint moment over decisions in $\mathbf{u}$ under H_k with

$$\epsilon_2^{(1)} = (1-\alpha)^2(P_d)^2 + 2\alpha(1-\alpha)P_d P_d^{(s)} + \alpha^2 P_d^{(s)} \tag{6a}$$

$$\epsilon_1^{(1)} = (1-\alpha)P_d + \alpha P_d^{(s)} \tag{6b}$$

$$\epsilon_2^{(0)} = (1-\alpha)^2(P_f)^2 + 2\alpha(1-\alpha)P_f P_f^{(s)} + \alpha^2 P_f^{(s)} \tag{6c}$$

$$\epsilon_1^{(0)} = (1-\alpha)P_f + \alpha P_f^{(s)} \tag{6d}$$

Proof: The correlation coefficient between u_i and u_j in $\mathbf{u}$ is defined by [26]

$$\rho^{(k)} = \frac{E^{(k)}(u_i u_j) - E^{(k)}(u_i)E^{(k)}(u_j)}{\sqrt{\mathrm{Var}^{(k)}(u_i)}\sqrt{\mathrm{Var}^{(k)}(u_j)}} \tag{7}$$

where $E^{(k)}(\cdot)$ and $\mathrm{Var}^{(k)}(\cdot)$ denote expectation and variance, respectively, under H_k, $k = 0, 1$. Let $k = 1$ in (7). Now, using (3a) and (4a), we can express $P(u_i = 1|H_1)$ as

$$P(u_i = 1|H_1) = (1-\alpha)P_d + \alpha P_d^{(s)} \tag{8}$$

To find $P(u_i = 1, u_j = 1|H_1)$, denote by $i = H$ and $i = S$ the state that decision u_i came from an honest node and the state that it came from the Sybil node, respectively. Now, using (3a) and (4a), we have

$$\begin{aligned} P(u_i = 1, u_j = 1|H_1) = &\sum_{X\in\{H,S\}}\sum_{Y\in\{H,S\}} \\ &P(i = X, j = Y)P(u_i = 1, u_j = 1|i = X, j = Y) \\ =&(1-\alpha)^2(P_d)^2 + 2\alpha(1-\alpha)P_d P_d^{(s)} + \alpha^2 P_d^{(s)} \end{aligned} \tag{9}$$

Next, we find the expectation terms, $E^{(1)}(\cdot)$, that appear in (7). First, using (8), we have

$$E^{(1)}(u_i) = \sum_{u_i\in\{0,1\}} u_i P(u_i) = \epsilon_1^{(1)} \tag{10}$$

where $\epsilon_1^{(1)}$ is the first order moment under H_1. Further,

$$\mathrm{Var}^{(1)}(u_i) = \sum_{u_i\in\{0,1\}} (u_i)^2 P(u_i) - (\epsilon_1^{(1)})^2 = \epsilon_1^{(1)}(1 - \epsilon_1^{(1)}) \tag{11}$$

Next, using (9), $E^{(1)}(u_i u_j)$ can be shown to be

$$E^{(1)}(u_i u_j) = \sum_{u_j\in\{0,1\}}\sum_{u_i\in\{0,1\}} u_i u_j P(u_i u_j) = \epsilon_2^{(1)} \tag{12}$$

where $\epsilon_2^{(1)}$ is the second order joint moment under H_1.

Substituting (10), (11), and (12) into (7), we get the result stated in the lemma for $k = 1$. Similarly, the lemma can be shown to hold true for $k = 0$. ∎

Next, we prove some properties of the correlation coefficient presented in Lemma 1.

LEMMA 2: *For the correlation coefficient in* (5)*, $\rho^{(k)} = 0$ when $\alpha = 0$; and $\rho^{(k)} = 1$ when $\alpha = 1$, $k = 0, 1$.*

Proof: Let $k = 1$. First, considering $\alpha = 0$, using (8), (11), and (12), we get $\epsilon_1^{(1)} = P_d$, $\mathrm{Var}^{(1)}(u_i) = P_d(1 - P_d)$, and $\epsilon_2^{(1)} = (P_d)^2$. Substituting these into (5), we get $\rho^{(1)} = 0$. Next, considering $\alpha = 1$, using (8), (11), and (12), we get $\epsilon_1^{(1)} = P_d^{(s)}$, $\mathrm{Var}^{(1)}(u_i) = P_d^{(s)}(1 - P_d^{(s)})$, and $\epsilon_2^{(1)} = P_d^{(s)}$. Substituting these into (5), we get $\rho^{(1)} = 1$. Similarly, the lemma can be shown to hold true for $k = 0$. ∎

LEMMA 3: *The degree of correlation between any pair of decisions u_i and u_j in $\mathbf{u}$ increases monotonically as the fraction of decisions from the Sybil node, α, increases.*

Proof: For $\rho^{(k)}$ (5), we have $\frac{\partial \rho^{(k)}}{\partial \alpha} > 0$, $0 \le \alpha \le 1$. ∎

B. General Correlation Structure over the Decision Vector $\mathbf{u}$

We now characterize the correlation structure among a set of q decisions in $\mathbf{u}$ as a function of the pairwise correlation coefficient in (5) and the first order moment over $\mathbf{u}$. Note that, for correlated binary decisions, the conditional PMFs of $\mathbf{u}$, when the local decision rules are given, and all local decisions have the same characteristics, can be expressed in terms of the joint moments of the elements of $\mathbf{u}$ as follows:

$$P_{\mathbf{u}0}(m) = \binom{N}{m}\sum_{i=0}^{m}(-1)^i\binom{m}{i}\epsilon_{N-m+1}^{(0)} \tag{13a}$$

$$P_{\mathbf{u}1}(m) = \binom{N}{m}\sum_{i=0}^{m}(-1)^i\binom{m}{i}\epsilon_{N-m+1}^{(1)} \tag{13b}$$

where, under H_k, $k = 0, 1$, $P_{\mathbf{u}k}(m)$ is the PMF of $\mathbf{u}$ giving the probability of having $0 \le m \le N$ decisions in favor of H_0, and $\epsilon_q^{(k)}$ is the q^{th} order joint moment given by

$$\epsilon_q^{(k)} = E^{(k)}(u_{i_1} u_{i_2} \cdots u_{i_q}), \qquad q \in \{1, 2, \cdots, N\} \tag{14}$$

Characterization of the PMFs in (13a) and (13b) requires the evaluation of the set of all N joint moments $\epsilon_q^{(k)}$ (14), $q \in \{1, 2, \cdots, N\}$, $k = 0, 1$. Specifically, the q^{th} order joint moment, under the Sybil attack, following a process similar to the one used for the derivation of (12), can be found as

$$\epsilon_q^{(1)} = (1-\alpha)^q(P_d)^q + \sum_{k=1}^{q}\binom{q}{q-k}(1-\alpha)^{q-k}(P_d)^{q-k}\alpha^k P_d^{(s)} \tag{15a}$$

$$\epsilon_q^{(0)} = (1-\alpha)^q(P_f)^q + \sum_{k=1}^{q}\binom{q}{q-k}(1-\alpha)^{q-k}(P_f)^{q-k}\alpha^k P_f^{(s)} \tag{15b}$$

Clearly, characterization of the joint PMFs in (13) using (15) becomes intractable as N increases. Thus, for tractability, we leverage a correlation structure presented in [27], which we present in our context in the following remark.

REMARK 1: As proven in Lemma 2, the pairwise correlation coefficient presented in (5), $\rho^{(k)} \in [0,1]$, $k = 0,1$, with $\rho^{(k)} = 0$ (which happens when $\alpha = 0$) implying that the received decisions are uncorrelated and $\rho^{(k)} = 1$ (which happens when $\alpha = 1$) implying that the received decisions are maximally correlated. In such a scenario, higher order joint moments, $\epsilon_q^{(k)}$ for $q \geq 2$, can be obtained recursively using [27] as

$$\epsilon_q^{(k)} = \epsilon_1^{(k)} \prod_{l=0}^{q-2} \frac{\rho^{(k)}(l+1-\epsilon_1^{(k)}) + \epsilon_1^{(k)}}{1 + l\rho^{(k)}} \tag{16}$$

where $\rho^{(k)}$ is the correlation coefficient given in (5) and $\epsilon_1^{(k)}$ is the first order moment under H_k given by (6b) and (6d).

In the rest of the paper, we employ the above correlation structure in (16) to model dependencies and perform analysis.

C. The Kullback-Leibler Divergence

We characterize the performance of the FC in the NP framework using the Kullback-Leibler Divergence (KLD) between the PMFs of $\mathbf{u}$ under H_0 (13a) and H_1 (13b). In the NP framework, the objective is to minimize the global missed detection probability (P_M) while keeping the global false alarm probability (P_F) below a threshold. According to Stein's lemma [28], KLD represents the best error exponent of the missed detection error probability in the NP setup, implying that the FC's decision-making performance improves with KLD. Specifically, KLD between the two conditional PMFs of $\mathbf{u}$, viz., $P_{\mathbf{u}0}(m)$ (13a) and $P_{\mathbf{u}1}(m)$ (13b), at the FC is:

$$D(\alpha, p) = \sum_{m=0}^{N} P_{\mathbf{u}1}(m) \log \frac{P_{\mathbf{u}1}(m)}{P_{\mathbf{u}0}(m)} \tag{17}$$

Next, using (17) as the FC's performance metric, we study the optimal Sybil attack against a distributed detection system.

IV. Optimal Sybil Attack that Blinds the FC

In this section, we characterize the optimal Sybil attack that "blinds" the FC, i.e, makes it incapable of making an informed decision. Specifically, we say that the FC is blind if the Sybil node can manipulate the received decision vector at the FC such that it does not convey any information regarding the phenomenon's state. Based on Stein's lemma [28], since lower KLD (D) is, worse will be the performance of the FC, and since D is always non-negative, to maximally degrade the FC's decision-making performance, the Sybil node would have to adopt strategies that make $D = 0$ (in which case no information reaches the FC and we say that it is blind).

In general, to make $D = 0$ in (17), we must have $P_{\mathbf{u}0}(m) = P_{\mathbf{u}1}(m)$, $0 \leq m \leq N$. Now, note that in (16), all higher order joint moments over decisions in $\mathbf{u}$ are expressed in terms of the first and second order moments. Thus, making the first and second order moments under H_0 to be equal to their corresponding moments under H_1 will yield $\epsilon_q^{(1)} = \epsilon_q^{(0)}$, $1 \leq q \leq N$, which would make $P_{\mathbf{u}0}(m) = P_{\mathbf{u}1}(m)$, $0 \leq m \leq N$ (as can be noted from (13a) and (13b)). Leveraging this approach, we characterize the optimal Sybil attack that blinds the FC in the next theorem.

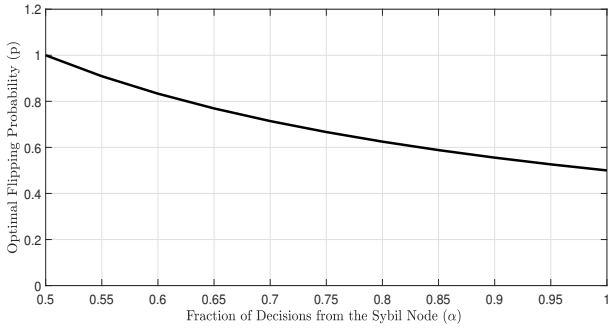

Fig. 2. Optimal flipping probability (p) that blinds the FC versus the fraction of decisions from the Sybil node (α).

THEOREM 1: *The FC becomes blind under the Sybil attack when either one of the following two conditions is satisfied:*

- *All decisions are sent by the same Sybil node, i.e., $\alpha = 1$, with the Sybil node using the flipping probability $p = 1/2$.*
- *The sum of the local detection and false alarm probabilities of the nodes is 1, i.e., $P_d + P_f = 1$, with $\alpha p = 1/2$.*

Proof: Denote by Δ_1 the difference between the first order moments under H_1 and H_0 of a decision u_i in $\mathbf{u}$. Then, using (4a), (4b), (6b), and (6d), and simplifying, we get

$$\Delta_1 = \epsilon_1^{(1)} - \epsilon_1^{(0)} = (P_d - P_f)(1 - 2\alpha p) \tag{18}$$

Denote by Δ_2 the difference between the second order joint moments under H_1 and H_0 of two decisions, u_i and u_j, in $\mathbf{u}$. Then, using (4a), (4b), (6a), and (6c) and simplifying, we get

$$\begin{aligned}\Delta_2 = \epsilon_2^{(1)} - \epsilon_2^{(0)} &= (P_d - P_f)[\alpha^2 + (1-\alpha^2)(P_d + P_f)] \\ &\quad - 2\alpha p(P_d - P_f)[2(P_d + P_f)(1-\alpha) + 2\alpha - 1]\end{aligned} \tag{19}$$

As discussed earlier, to blind the FC, the Sybil attacker would have to make $\Delta_1 = 0$ and $\Delta_2 = 0$. Now, since $P_d > P_f$ for each point on the receiver operating characteristic of an optimum detector, to make $\Delta_1 = 0$ in (18), we must have

$$\alpha p = 1/2 \tag{20}$$

Substituting (20) into (19), we get

$$\Delta_2 = (P_d - P_f)(1-\alpha)^2[1 - (P_d + P_f)] \tag{21}$$

Clearly, to have $\Delta_2 = 0$ in (21), we must have $\alpha = 1$ or $P_d + P_f = 1$. Thus, we conclude that $\Delta_1 = \Delta_2 = 0$ when either one of the following is true: *a*) $\alpha = 1$ and $p = 1/2$, *or*, *b*) $P_d + P_f = 1$ and $\alpha p = 1/2$. This completes the proof. ■

COROLLARY 1: *The minimum fraction of decisions in $\mathbf{u}$ that must come from the Sybil node to blind the FC is $\alpha = 1/2$.*

Proof: From the second condition in Theorem 1, it can be noted that the minimum value of α needed to blind the FC is $1/2$, with $\alpha = 1/2$ when $p = 1$. ■

In Fig. 2, considering $P_d = 0.9$ in (3a) and $P_f = 0.1$ in (3b), i.e., $P_d + P_f = 1$, we numerically found the flipping probability (p) that makes $D = 0$ in (17) and plotted them with varying fraction of decisions from the Sybil node (α). As

can be noted from the figure, the condition for blinding the FC follows $\alpha p = 1/2$, which corroborates the blinding criteria presented in Theorem 1 for $P_d + P_f = 1$. It can also be shown that, for $P_d + P_f \neq 1$, numerically solving $D = 0$ yields $\alpha = 1$ and $p = 1/2$. These results corroborate Theorem 1.

Next, we use game theory to investigate strategic attack-defense techniques for the regime where the attacker does not have sufficient resources to blind the FC.

V. Game Theoretic Attack-Defense

In this section, we use game theory [29] to investigate strategic interactions between the Sybil node and the FC, with the Sybil node aiming to minimize, and the FC aiming to maximize, the FC's decision-making performance. Use of KLD (17) as the FC's performance metric in such a scenario, however, becomes mathematically intractable for analysis without relying on numerical techniques. Therefore, we adopt a surrogate function, which is the sum of the squares of $\Delta_1(p, \lambda)$ (18) and $\Delta_2(p, \lambda)$ (19), as the performance metric. Note, it can be shown that D given in (17) increases with $\Delta_1(p, \lambda)$ (18) as well as with $\Delta_2(p, \lambda)$ (19). However, since $\Delta_1(p, \lambda)$ and $\Delta_2(p, \lambda)$ can be negative or positive, taking the squares of the two quantities in the surrogate function becomes necessary. Specifically, the surrogate function is defined as

$$\Delta(p, \lambda) = (\Delta_1(p, \lambda))^2 + (\Delta_2(p, \lambda))^2 \tag{22}$$

where

$$\Delta_1(p, \lambda) = [P_d(\lambda) - P_f(\lambda)][1 - 2\alpha p] \tag{23a}$$

$$\Delta_2(p, \lambda) = [P_d(\lambda) - P_f(\lambda)][\alpha^2 + (1 - \alpha^2)\{P_d(\lambda) + P_f(\lambda)\}] - 2\alpha p\{P_d(\lambda) - P_f(\lambda)\}[2\{P_d(\lambda) + P_f(\lambda)\}(1 - \alpha) + 2\alpha - 1] \tag{23b}$$

In the game, we consider that the FC chooses the threshold λ of the LRT (1) conducted at the sensors, and that the Sybil node chooses the flipping probability p (for a given α), with the FC aiming to maximize (22), and the Sybil node aiming to minimize (22). The game is clearly a zero-sum game. Therefore, the Nash Equilibrium (NE) of the game (which would coincide with its saddle point) corresponds to choosing p^* (for the Sybil) and λ^* (for the FC) that solve the following optimization problem:

$$\max_{\lambda} \min_{p} \Delta(p, \lambda) = \min_{p} \max_{\lambda} \Delta(p, \lambda) \tag{24}$$

for a given α. Next, we investigate the NE of the game and show that a pure strategy NE exists.

Lemma 4: *For a fixed λ, Δ (22) is a convex function of p.*

Proof: The second partial derivative of Δ (22) w.r.t p can be shown to be $\partial^2\Delta/\partial p^2 = 8\alpha^2(a^2 + b^2) \geq 0$, where $a = (P_d - P_f)$ and $b = (P_d - P_f)[2(P_d + P_f)(1 - \alpha) + 2\alpha - 1]$, implying that Δ is a convex function of p. ∎

Lemma 5: *For a fixed p, Δ (22) is a quasi-concave function of λ attaining its maximum value at the critical point*

$$\lambda^* = \frac{a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_f}{a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_d} \tag{25}$$

where

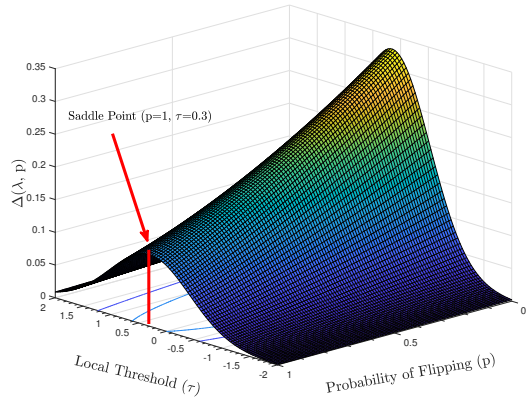

Fig. 3. Utility function ($\Delta(p, \lambda)$) versus flipping probability (p) and local sensor threshold (τ), where $\tau = \log\lambda + 1/2$, for $\alpha = 0.2$.

$$a_1(\alpha, p) = 1 - 2\alpha p \tag{26a}$$

$$b_1(\alpha, p) = \alpha^2 - 2\alpha p(2\alpha - 1) \tag{26b}$$

$$c_1(\alpha, p) = 1 - \alpha^2 - 4\alpha p(1 - \alpha) \tag{26c}$$

Proof: Let us rewrite Δ_1 (23a) and Δ_2 (23b) as

$$\Delta_1 = a_1(P_d - P_f) \tag{27a}$$

$$\Delta_2 = b_1(P_d - P_f) + c_1(P_d^2 - P_f^2) \tag{27b}$$

where a_1 (26a), b_1 (26b), and c_1 (26c) are constants that do not involve λ. The partial derivative of Δ (22) w.r.t λ is

$$\frac{\partial\Delta}{\partial\lambda} = 2\Delta_1\frac{\partial P_f}{\partial\lambda}\left[a_1\left(\frac{\partial P_d}{\partial P_f} - 1\right)\right] + 2\Delta_2\frac{\partial P_f}{\partial\lambda}\left[b_1\left(\frac{\partial P_d}{\partial P_f} - 1\right) + 2c_1\left(P_d\frac{\partial P_d}{\partial P_f} - P_f\right)\right] \tag{28}$$

Recognizing that $\partial P_d/\partial P_f = \lambda$, we can simplify (28) as

$$\frac{\partial\Delta}{\partial\lambda} = 2\frac{\partial P_f}{\partial\lambda}[\lambda(a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_d) - (a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_f)] \tag{29}$$

Setting $\partial\Delta/\partial\lambda = 0$ in (29), we get λ^* as given in (25).

Next, we investigate the sign of $\partial\Delta/\partial\lambda$ when $\lambda < \lambda^*$ and when $\lambda > \lambda^*$. Dividing both sides of (29) by $(a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_f)$, and using (25), we get

$$\frac{\partial\Delta/\partial\lambda}{a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_f} = 2\frac{\partial P_f}{\partial\lambda}\left(\frac{\lambda - \lambda^*}{\lambda^*}\right) \tag{30}$$

Using (26a), (26b), and (26c), it can be shown that $a_1\Delta_1 + b_1\Delta_2 + 2c_1\Delta_2 P_f > 0$ when $P_f < 0.5$ (which is the case in practice). Also, we note from the properties of the receiver operating characteristic of an optimum detector that $\partial P_f/\partial\lambda < 0$. Thus, when $\lambda < \lambda^*$, $\partial\Delta/\partial\lambda > 0$, and when $\lambda > \lambda^*$, $\partial\Delta/\partial\lambda < 0$. Hence, Δ is a quasi-concave function of λ attaining its maximum value at $\lambda = \lambda^*$. ∎

Theorem 2: *A pure strategy NE which solves (24) exists.*

Proof: Since we have proven that, given λ, Δ is a convex function of p (Lemma 4), and given p, Δ is a quasi-concave function of λ (Lemma 5), we conclude using the Debreu-Fan-Glicksberg theorem [29] that a pure strategy NE which solves (24) exists. ∎

Next, we provide numerical results to corroborate our game theoretic results. Consider a phenomenon which can be in state H_0 or H_1, with the sensors' observations under each state following a Gaussian distribution, viz. $H_0 \sim \mathcal{N}(0,1)$ and $H_1 \sim \mathcal{N}(1,1)$. Consider $\alpha = 0.2$. In Fig. 3, we plot $\Delta(p,\lambda)$ (22) versus both p and τ, where $\tau = \log\lambda + 1/2$ for λ defined in (1). From the figure, we can observe the convexity of $\Delta(p,\lambda)$ w.r.t p for a fixed τ (which corroborates Lemma 4) as well as the quasi-concavity of $\Delta(p,\lambda)$ w.r.t τ for a fixed p (which corroborates Lemma 5). Such a nature of $\Delta(p,\lambda)$ indicates the existence of a saddle-point as mentioned in Theorem 2, which occurs at $(p^*,\tau^*) = (1, 0.3)$ as marked in the figure.

VI. Conclusion

The problem of distributed detection pertains to the employment of data from multiple distributed sensors for inferring the state of a phenomenon of interest. Such systems have historically found applications in diverse areas with the importance of such system having become further bolstered by the advent of IoT. This paper investigated the problem of a Sybil attack on a distributed detection system where a malicious Sybil sensor can send falsified decisions using multiple fake identities to degrade an FC's decision-making performance. The paper studied the problem in the Neyman-Pearson setup. The paper showed that such an attack introduces correlations among the elements of the received decision vector at the FC and presented a correlation structure to model such dependencies. Accounting for the correlations introduced, the paper characterized the optimal Sybil attack that would blind the FC, i.e., would prevent the FC from receiving any information, under different scenarios. Further, using game theory, the paper also analyzed strategic attack-defense between the Sybil attacker and the FC and proved the existence of an NE. Numerical results were presented to provide important insights.

References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York:Springer-Verlag, 1997.

[2] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part i - fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54 –63, Jan 1997.

[3] 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, pp. 100–117, 2012.

[4] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys. New York, NY, USA: ACM, 2008, pp. 323–336.

[5] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: Participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '09. New York, NY, USA: ACM, 2009, pp. 349–350.

[6] N. Maisonneuve, M. Stevens, M. E. Niessen, P. Hanappe, and L. Steels, "Citizen noise pollution monitoring," in *Proceedings of the 10th Annual International Conference on Digital Gov. Research: Social Networks: Making Connections Between Citizens, Data and Government*, ser. dg.o '09. Digital Government Society of North America, 2009, pp. 96–103.

[7] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen, "Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype," in *Proceedings of the 4th Workshop on Embedded Networked Sensors*, ser. EmNets '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 13–17.

[8] M. B. Srivastava, T. F. Abdelzaher, and B. K. Szymanski, "Human-centric sensing," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, pp. 176 – 197, 2012.

[9] D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti, X. Wang, P. Mohapatra, B. Szymanski, and H. Le, "Using humans as sensors: An estimation-theoretic perspective," in *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN)*, 2014, pp. 35–46.

[10] P. J. Smith, R. Senanayake, P. A. Dmochowski, and J. S. Evans, "Distributed spectrum sensing for cognitive radio networks based on the sphericity test," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1831–1844, 2019.

[11] Y. Chen, S. Kar, and J. M. Moura, "The internet of things: Secure distributed inference," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 64–75, 2018.

[12] V. S. S. Nadendla, V. Sharma, and P. K. Varshney, "On strategic multi-antenna jamming in centralized detection networks," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 186–190, Feb 2017.

[13] A. Vempaty, L. Tong, and P. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *Signal Processing Magazine, IEEE*, vol. 30, no. 5, pp. 65–75, 2013.

[14] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Distributed detection in tree topologies with byzantines," *IEEE Trans. Signal Process.*, vol. 62, pp. 3208–3219, June 2014.

[15] H. Lin, P. 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.

[16] W. A. Hashlamoun, S. Brahma, and P. K. Varshney, "Mitigation of byzantine attacks on distributed detection systems using audit bits," *IEEE Trans. Signal Inf. Process. over Networks*, vol. 4, no. 1, pp. 18–32, 2018.

[17] B. Kailkhura, S. Brahma, and P. K. Varshney, "Data falsification attacks on consensus-based detection systems," *IEEE Trans. on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 145–158, 2017.

[18] B. Kailkhura, V. S. S. Nadendla, and P. K. Varshney, "Distributed inference in the presence of eavesdroppers: a survey," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 40–46, June 2015.

[19] B. Wu, J. Chen, J. Wu, and M. Cardei, "A survey of attacks and countermeasures in mobile ad hoc networks," *Wireless/Mobile Network Security, Springer*, vol. 17, pp. 103–135, 2007.

[20] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[21] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, "Resilient network coding in the presence of byzantine adversaries," in *Proc. 26th IEEE Int. Conf. on Computer Commun., INFOCOM, (Anchorage, AK)*, 2007, pp. 616–624.

[22] J. R. Douceur, "The sybil attack," in *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, ser. IPTPS '01. London, UK, UK: Springer-Verlag, 2002, pp. 251–260. [Online]. Available: http://dl.acm.org/citation.cfm?id=646334.687813

[23] J. Newsome, E. Shi, D. Song, and A. Perrig, "The sybil attack in sensor networks: analysis and defenses," in *Third International Symposium on Information Processing in Sensor Networks (IPSN)*, 2004, pp. 259–268.

[24] G. Wang, B. Wang, T. Wang, A. Nika, H. Zheng, and B. Y. Zhao, "Defending against sybil devices in crowdsourced mapping services," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 179–191. [Online]. Available: https://doi.org/10.1145/2906388.2906420

[25] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. control, Signals, and Systems*, vol. 1, pp. 167–182, 1988.

[26] R. E. Ziemer, *Elements of Engineering Probability and Statistics*. New Jersey: Prentice Hall, 1997.

[27] E. Drakopoulos and C. C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 593–606, Jul 1991.

[28] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[29] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.