# Efficient Ordered-Transmission Based Distributed Detection under Data Falsification Attacks

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

***Abstract*—In distributed detection systems, energy-efficient ordered transmission (EEOT) schemes are able to reduce the number of transmissions required to make a final decision. In this work, we investigate the effect of data falsification attacks on the performance of EEOT-based systems. We derive the probability of error for an EEOT-based system under attack and find an upper bound (UB) on the expected number of transmissions required to make the final decision. Moreover, we tighten this UB by solving an optimization problem via integer programming (IP). We also obtain the FC's optimal threshold which guarantees the optimal detection performance of the EEOT-based system. Numerical and simulation results indicate that it is possible to reduce transmissions while still ensuring the quality of the decision with an appropriately designed threshold.**

***Index Terms*—Ordered transmissions, data falsification attacks, wireless sensor networks, distributed detection.**

## I. Introduction

Energy efficiency is an important design consideration to enhance the lifetime of a wireless sensor network (WSN). There have been several notable schemes proposed to enhance the energy efficiency of WSNs [1]–[5]. In this paper, we consider the ordered transmission (OT) scheme. The OT scheme was first proposed in [4], where informative sensors[1] transmit their log-likelihood ratios (LLRs) to the FC prior to less informative ones. Once the fusion center (FC) has received enough observations to make a final decision of desired quality, the rest of the observations are no longer needed. A correlation-aware OT scheme is proposed in [6], where spatial correlation between the sensors is considered. The authors of [7] proposed an energy-efficient OT (EEOT) scheme in which informative sensors send binary decisions, rather than sending LLRs, to the FC to further improve the energy efficiency of the system. Studies in [4], [6]–[8] indicate that OT-based schemes can reduce the number of transmissions needed for decision-making.

The OT approach has also been used in some other systems to reduce communication costs and energy consumption. In [9], sequential detection along with OT was considered for cooperative spectrum sensing to obtain fast and reliable decisions regarding primary user activities over the spectrum. The authors of [10] applied the OT scheme to the distributed quickest change detection problem, where the number of transmissions to the FC are reduced without affecting the detection delay of the system. In [11], the OT scheme was incorporated into energy harvesting sensor networks in order to increase the energy efficiency of the sensors. The authors of [12] presented an ordered gradient approach to eliminate some sensor-to-FC uplink communications in distributed ADMM.

Thus, the OT scheme is a promising scheme for a significant improvement in the energy efficiency of distributed systems. However, due to the open nature of WSNs, the performance of the EEOT-based systems under attack by malicious sensors is an important aspect to consider. Here, we consider one type of attack, namely data falsification attacks. We consider the data falsification attack model used in [13], [14], where a sensor may be malicious that sends falsified data to the FC with the intention of degrading the detection performance of the system. To the best of our knowledge, this is the first work that investigates the impact of data falsification attacks on OT-based schemes. We evaluate the performance of the EEOT-based system via its detection performance and the average number of transmissions (ANT) required under data falsification attacks. We derive the probability of error for the EEOT-based distributed detection system along with the optimal decision threshold at the FC. Further, we derive a tight upper bound (UB) on the ANT required under attack. Numerical and simulation results indicate that it is possible to reduce transmissions while still ensuring the quality of the decision with an appropriately designed threshold.

## II. System Model

A distributed OT-based network consisting of $N$ sensors and one FC is considered in this work. A binary hypothesis testing problem is investigated where hypothesis $\mathcal{H}_1$ indicates the presence of the signal and $\mathcal{H}_0$ indicates the absence of the signal, and the goal is to determine which of the two hypotheses is true. Let $y_i$ be the received observation at sensor $i \in \{1, 2, \ldots, N\}$. We assume that all the observations are independent and identically distributed (i.i.d) conditioned on the hypotheses. For sensor $i$, the observation $y_i$ is modeled as

$$y_i = \begin{cases} n_i & \text{under } \mathcal{H}_0 \\ s + n_i & \text{under } \mathcal{H}_1, \end{cases} \tag{1}$$

C. Quan, Nandan Sriranga, Haodong Yang and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: {chquan, nsrirang, hyang85, varshney}@syr.edu). Funding from National Key Research and Development Program of China under Grant 2022YFA1004902.

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).

[1] Sensors with larger LLR magnitudes are more informative than the ones with smaller LLR magnitudes. In the rest of the paper, the sensors with larger LLR magnitudes are referred to as informative sensors.

where $n_i$ is the Gaussian noise with zero mean and variance $\sigma^2$ and $s$ is the signal strength at each sensor. $s$ and $n_i$ are assumed to be independent. Hence, we have $y_i|\mathcal{H}_0 \sim \mathcal{N}(0, \sigma^2)$ and $y_i|\mathcal{H}_1 \sim \mathcal{N}(s, \sigma^2)$. Based on the local observations $\{y_i\}_{i=1}^N$, each sensor $i \in \{1, \ldots, N\}$ makes a binary decision $v_i \in \{0, 1\}$ regarding the two hypotheses using the log-likelihood ratio test (LRT) $L_i = \log\left(\frac{f_{Y_i}(y_i|\mathcal{H}_1)}{f_{Y_i}(y_i|\mathcal{H}_0)}\right) \underset{v_i=0}{\overset{v_i=1}{\gtrless}} \lambda$, where $L_i$ denotes the LLR of sensor $i$, $\lambda$ is the identical threshold used by all the sensors and $v_i$ is the local decision made by sensor $i$. According to the OT-based framework presented in [4] and [7], the sensor transmissions are ordered based on the magnitudes of their LLRs. If the magnitudes of the LLRs are ordered as $|L_{[1]}| > |L_{[2]}| > \ldots > |L_{[N]}|$, the sensors transmit their local decisions in the order determined by their magnitude-ordered LLRs, i.e., in the order of $v_{[1]}, v_{[2]}, \ldots, v_{[N]}$, where $v_{[k]}$ is the local decision made by the $[k]^{th}$ sensor. [2]

The decision rule used by the FC is the same as in [7]:

$$\begin{cases} \sum_{i=1}^k v_{[k]} \geq T & \text{decide } \mathcal{H}_1 \\ \sum_{i=1}^k v_{[k]} < T - (N-k) & \text{decide } \mathcal{H}_0 \end{cases}, \tag{2}$$

where $T$ is the threshold used by the FC. When $\sum_{i=1}^k v_{[k]} \geq T$, the FC decides $\mathcal{H}_1$, even when the rest of the decisions that have not yet been transmitted are all 0's, and vice versa. As in [7], the following assumption is made in this paper.

*Assumption 1:* $P(v_i = 1|\mathcal{H}_1) \to 1$ and $P(v_i = 0|\mathcal{H}_0) \to 1$ when $s \to \infty$.

### A. Attack Model

In our setup, a sensor $i$ can be honest $(H)$ or malicious $(M)$. Assume that each sensor in the network has $\alpha$ probability of being malicious, i.e., $P(i = M) = \alpha$, and the malicious sensors falsify data by flipping their local decisions sent to the FC. Let $p = P(v_i \neq u_i|i = M)$ denote the probability that the malicious node $i$ flips its local decisions, where $v_i$ is the original unaltered local decision made by sensor $i$ and $u_i$ is the local decision sent to the FC by sensor $i$. Hence, we have $P(v_i = u_i) = 1 - \alpha p$ and $P(v_i \neq u_i) = \alpha p$.

### B. Detection Performance

First, we investigate the detection performance of the EEOT-based scheme in the presence of malicious nodes. We begin our analysis by first presenting the following Lemma. The proof is not included here due to space limitation. All the omitted proofs or derivations in this paper can be found in the extended version of this paper [15].

**Lemma II.1.** *When the FC follows the Bayesian decision rule, the detection performance of systems with and without the use of the EEOT-based scheme are the same under data falsification attacks.*

Thus, we evaluate the detection performance of the EEOT-based system under data falsification attacks by evaluating the detection performance of the corresponding distributed system without ordering. According to [16], for the system without ordering, the probabilities of $u_i = 1$ and $u_i = 0$ given $\mathcal{H}_h$ are expressed as $\widetilde{\pi}_{1,h} = \sum_{x \in \{M,H\}} P(u_i = 1|\mathcal{H}_h, i = x)P(i = x) = \alpha p \pi_{0,h} + (1 - \alpha p)\pi_{1,h}$ and $\widetilde{\pi}_{0,h} = P(u_i = 0|\mathcal{H}_h) = 1 - \widetilde{\pi}_{1,h}$, respectively, for $h = 0, 1$, where $\pi_{1,h} = P(v_i = 1|\mathcal{H}_h) = Q\left(\frac{\lambda - \mu_h}{\nu_h}\right)$, $\pi_{0,h} = P(v_i = 0|\mathcal{H}_h) = 1 - \pi_{1,h}$, $\mu_0 = 0$, $\mu_1 = s$ and $\nu_0 = \nu_1 = \sigma$. $Q(.)$ is the tail distribution function of the standard normal distribution. From Assumption 1, we have $\pi_{1,0} = \pi_{0,1} \approx 0$ and $\pi_{0,0} = \pi_{1,1} \approx 1$. Thus, we have $\widetilde{\pi}_{1,0} = \widetilde{\pi}_{0,1} \approx \alpha p$ and $\widetilde{\pi}_{1,1} = \widetilde{\pi}_{0,0} \approx 1 - \alpha p$.

The fusion rule for the distributed system when all of the sensor decisions are used is given by $\sum_{i=1}^N u_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T$, which follows from [17]. Thus, the error probability of the unordered system is $P_{e,EEOT}^{FC} = \pi_1(1 - P_{d,EEOT}^{FC}) + \pi_0 P_{f,EEOT}^{FC}$, where $\pi_1 = P(\mathcal{H}_1)$, $\pi_0 = P(\mathcal{H}_0)$, $P_{d,EEOT}^{FC} = \sum_{i=T+1}^N \binom{N}{i} \pi_{1,1}^i \pi_{0,1}^{N-i}$ and $P_{f,EEOT}^{FC} \sum_{i=T+1}^N \binom{N}{i} \pi_{1,0}^i \pi_{0,0}^{N-i}$ are the probabilities of detection and false alarm of the FC, respectively. According to Lemma II.1, the error probabilities of the ordered system are the same as that of the unordered system.

Next, we aim at finding the value of optimal $T$ which minimizes the probability of error of both the unordered system and the EEOT-based system. Let non-negative $Z = \sum_{i=1}^N u_i$ denote the number of local decisions that decided 1. The optimal decision rule at the FC, which is $\frac{\prod_{i=1}^N P(u_i|\mathcal{H}_1)}{\prod_{i=1}^N P(u_i|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\pi_0}{\pi_1}$, can be rewritten as

$$\left(\frac{\widetilde{\pi}_{1,1}}{\widetilde{\pi}_{1,0}}\right)^Z \left(\frac{1 - \widetilde{\pi}_{1,1}}{1 - \widetilde{\pi}_{1,0}}\right)^{N-Z} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\pi_0}{\pi_1}. \tag{3}$$

We make the reasonable assumption that the probability of a sensor being malicious is less than 0.5, i.e., $\alpha < 0.5$, and $0 \leq p \leq 1$ which implies that $\alpha p < 0.5$. This implies that $\widetilde{\pi}_{1,1} > \widetilde{\pi}_{1,0}$ (and $\widetilde{\pi}_{0,0} > \widetilde{\pi}_{0,1}$). Taking the logarithm of both sides of (3), the optimal decision rule can be rewritten as

$$Z \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \left[\log\left(\frac{\pi_0}{\pi_1}\right) + N \log\left(\frac{1 - \widetilde{\pi}_{1,0}}{1 - \widetilde{\pi}_{1,1}}\right)\right] / \log\left(\frac{\widetilde{\pi}_{1,1}(1 - \widetilde{\pi}_{1,0})}{\widetilde{\pi}_{1,0}(1 - \widetilde{\pi}_{1,1})}\right). \tag{4}$$

Therefore, the optimal threshold $T^*$ at the FC is equal to the right hand side of (4).

### C. Average Number of Transmissions Required for Energy-efficient OT-based System under Attack

Next, we consider the effect of malicious attacks on the average number of transmissions (ANT) required by the EEOT scheme. In order to simplify the computation, we find the upper bound (UB) of the ANT required by finding the lower bound (LB) of the ANT saved. We first consider the case when the FC decides $\mathcal{H}_1$. It has been derived in [7] that the ANT saved to make a final decision is lower bounded by $N/2$ in the absence of the data falsification attacks. Here, we investigate the effect that the attacks have on the lower bound of the ANT saved and finding a lower bound for the system under data falsification attacks. It is also shown in this paper that the lower bound we obtain is tight.

We define $k_L^* = \min_{1 \leq k \leq N} \left\{\sum_{i=1}^k u_{[i]} \geq T\right\}$ and $k_U^* = \min_{1 \leq k \leq N} \left\{\sum_{i=1}^k u_{[i]} < T - (N-k)\right\}$ as the minimum number

[2] Note that the magnitude-ordered LLRs do not imply that the local decisions are also magnitude-ordered, i.e., $|L_{[1]}| > |L_{[2]}| > \ldots > |L_{[N]}|$ does not imply $v_{[1]} \geq v_{[2]} \geq \ldots \geq v_{[N]}$.

of transmissions required to decide $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively, under data falsification attacks. Let $\lceil T\rceil$ denote the rounding up of $T$ to the closest integer that is greater than or equal to $T$. The ANT saved when the FC decides $\mathcal{H}_1$ is given as

$$\bar{N}_{s,1}(\beta) = E(N-k_L^*) = \sum_{k=1}^{N}(N-k)P(k_L^*=k) \tag{5a}$$

$$= \sum_{k=1}^{\lceil T\rceil+\beta}(N-k)P(k_L^*=k) + \sum_{k=\lceil T\rceil+\beta+1}^{N}(N-k)P(k_L^*=k) \tag{5b}$$

$$\geq \sum_{k=1}^{\lceil T\rceil+\beta}(N-k)P(k_L^*=k) \tag{5c}$$

$$\geq (N-\lceil T\rceil-\beta)P(k_L^*\leq\lceil T\rceil+\beta) \tag{5d}$$

$$= (N-\lceil T\rceil-\beta)P(\sum_{k=1}^{\lceil T\rceil+\beta}u_{[k]}\geq T|\Gamma_1\geq T,\mathcal{H}_1)\pi_1, \tag{5e}$$

where $\Gamma_1 = \sum_{i=1}^{\lceil T\rceil+\beta}v_{[i]}$ and $P(\sum_{k=1}^{\lceil T\rceil+\beta}u_{[k]} \geq T|\Gamma_1 \geq T,\mathcal{H}_1) = \sum_{i=0}^{\beta}\binom{\lceil T\rceil+\beta}{i}\widetilde{\pi}_{0,1}^{i}\widetilde{\pi}_{1,1}^{\lceil T\rceil+\beta-i}$. Note that the detailed derivation of (5) can be found in the extended version of this paper [15]. In going from (5b) to (5c), the second summation term, which is positive, is dropped. As the difference between the actual ANT saved and its LB is dependent on the number of terms in the dropped second summation term in (5b), an appropriate number of terms should be chosen in order to reduce that difference and tighten the LB. Thus, we introduce a variable $\beta$ in (5c) and try to find an appropriate $\beta$ later to prevent the dropped second part of (5b) from being too large so that the LB is tight when the FC decides $\mathcal{H}_1$.

Next, we consider the case when the FC decides $\mathcal{H}_0$. Similarly, the ANT saved when the FC decides $\mathcal{H}_0$ is given as

$$\bar{N}_{s,2}(\beta) \geq (\lceil T\rceil-\beta)P(\sum_{k=1}^{N-\lceil T\rceil+\beta}u_{[k]}<\kappa|\Gamma_2<T,\mathcal{H}_0)\pi_0, \tag{6}$$

where $\Gamma_2 = \sum_{i=1}^{N-\lceil T\rceil+\beta}v_{[i]}$, $\kappa = T-(\lceil T\rceil-\beta)$, $P(\sum_{k=1}^{N-\lceil T\rceil+\beta}u_{[k]} < \kappa|\Gamma_2 < T,\mathcal{H}_0) = \sum_{i=0}^{\lfloor T\rfloor-\lceil T\rceil+\beta}\binom{N-\lceil T\rceil+\beta}{i}\widetilde{\pi}_{1,0}^{i}\widetilde{\pi}_{0,0}^{N-\lceil T\rceil+\beta-i}$ and $\lfloor T\rfloor$ denote the rounding down of $T$ to the next lowest integer. The derivation is available in [15]. In a manner similar to the one employed earlier, the variable $\beta$ is introduced to ensure that the LB of the ANT saved is tight when the FC decides $\mathcal{H}_0$. Since only one of the two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$ can occur at any given time, the events $k_L^*=k$ and $k_U^*=k$ given hypothesis $\mathcal{H}_1$ or $\mathcal{H}_0$ are disjoint. Hence the total ANT saved is $N_{s,EEOT}(\beta) = \sum_{k=1}^{N}(N-k)\sum_{h=0}^{1}[P(k_U^*=k|\mathcal{H}_h)+P(k_L^*=k|\mathcal{H}_h)]P(\mathcal{H}_h) = \sum_{k=1}^{N}(N-k)[P(k_U^*=k)+P(k_L^*=k)]$ and the LB of the ANT saved is $N_{s,EEOT}^{L}(\beta) = \bar{N}_{s,1}(\beta)+\bar{N}_{s,2}(\beta)$. When $\beta=0$, the LB derived here reduces to the LB obtained in [7]. However, $\beta=0$ might not be an appropriate value that allows us to get a tight LB under attack. Thus, we aim at finding an optimal $\beta$ so that $N_{s,EEOT}^{L}(\beta)$ is maximized and the LB becomes tighter. Upon solving the optimization problem given in (7), we are able to find the optimal $\beta^*$. We denote the set of integers by $\mathbb{Z}$ and cast the optimization problem as:

$$\max_{\beta} \quad \bar{N}_{s,1}(\beta)+\bar{N}_{s,2}(\beta) \tag{7a}$$

$$\text{s.t.} \quad 0\leq\beta\leq min(N-\lceil T\rceil,\lceil T\rceil) \tag{7b}$$

$$\beta\in\mathbb{Z}, \tag{7c}$$

The constraint in (7b) is due to the fact that the value of $\beta$ must satisfy both $\lceil T\rceil+\beta\leq N$ and $N-\lceil T\rceil+\beta\leq N$, which are derived from (5e) and (6), respectively. This is due to the fact that the upper index of the summations in (5e) and (6) should be less or equal to $N$. As the optimization problem in (7) is an integer programming (IP) problem, it is a non-convex optimization problem. However, we have the following lemma which helps us obtain the optimal solution to (7).

**Theorem II.2.** *$N_{s,EEOT}^{L}(\beta)$ as a function of $\beta$ satisfies either*

1) *$N_{s,EEOT}^{L}(\beta)$ is a non-increasing function, $\forall\beta\in[0,\min(N-\lceil T\rceil,\lceil T\rceil)]$.*
   *or*
2) *There exists a $\beta_l\in\mathbb{Z}$ such that $N_{s,EEOT}^{L}(\beta)$ is an increasing function $\forall\beta\in[0,\beta_l-1]$ and a non-increasing function $\forall\beta\in[\beta_l,\min(N-\lceil T\rceil,\lceil T\rceil)]$.*

*Proof:* Let $g_1(\beta) = \sum_{i=0}^{\beta}\binom{\lceil T\rceil+\beta}{i}\widetilde{\pi}_{0,1}^{i}\widetilde{\pi}_{1,1}^{\lceil T\rceil+\beta-i}$ and $g_2(\beta) = \sum_{i=0}^{\lfloor T\rfloor-\lceil T\rceil+\beta}\binom{N-\lceil T\rceil+\beta}{i}\widetilde{\pi}_{1,0}^{i}\widetilde{\pi}_{0,0}^{N-\lceil T\rceil+\beta-i}$. Hence, we have $g_1(\beta+1) = \sum_{i=0}^{\beta+1}\binom{\lceil T\rceil+\beta+1}{i}\widetilde{\pi}_{0,1}^{i}\widetilde{\pi}_{1,1}^{\lceil T\rceil+\beta+1-i}$ and $g_2(\beta+1) = \sum_{i=0}^{T_d+\beta+1}\binom{N-\lceil T\rceil+\beta+1}{i}\widetilde{\pi}_{1,0}^{i}\widetilde{\pi}_{0,0}^{N-\lceil T\rceil+\beta+1-i}$, where $T_d = \lfloor T\rfloor-\lceil T\rceil = -1$, $\widetilde{\pi}_{0,1} = \widetilde{\pi}_{1,0} = \alpha p$ and $\widetilde{\pi}_{0,0} = \widetilde{\pi}_{1,1} = 1-\alpha p$ based on Assumption 1. $g_1(\beta+1)$ and $g_2(\beta+1)$ can be expressed in terms of $g_1(\beta)$ and $g_2(\beta)$ that are respectively given as

$$g_q(\beta+1) = g_q(\beta)+A_q(\beta), \tag{8}$$

for $q=1,2$, where $A_1(\beta) = \binom{\lceil T\rceil+\beta}{\beta+1}(\alpha p)^{\beta+1}(1-\alpha p)^{\lceil T\rceil}$ and $A_2(\beta) = \binom{N-\lceil T\rceil+\beta}{\beta+T_d+1}(\alpha p)^{\beta+T_d+1}(1-\alpha p)^{N-\lceil T\rceil-T_d}$. The derivation of (8) is available in [15]. It is evident that if $N_{s,EEOT}^{L}(\beta) = [\bar{N}_{s,1}(\beta+1)+\bar{N}_{s,2}(\beta+1)]-[\bar{N}_{s,1}(\beta)+\bar{N}_{s,2}(\beta)] < 0$, then $N_{s,EEOT}^{L}(\beta) > N_{s,EEOT}^{L}(\beta+1)$. By rewriting $N_{s,EEOT}^{L}(\beta) < 0$ using (8), we obtain the inequality

$$D(\beta) > D_2(\beta), \tag{9}$$

where $D(\beta) = \pi_1 g_1(\beta)+\pi_0 g_2(\beta)$, $D_2(\beta) = \pi_1 h_1(\beta)+\pi_0 h_2(\beta)$, $h_1(\beta) = (N-\lceil T\rceil-\beta-1)A_1(\beta)$ and $h_2(\beta) = (\lceil T\rceil-\beta-1)A_2(\beta)$.

We proceed to show that if $D(\beta) > D_2(\beta)$ is true, then $D(\beta+1) > D_2(\beta+1)$ is also true. Using the expressions in (8) and (9), we rewrite $D(\beta+1) > D_2(\beta+1)$ as

$$2[\pi_1 A_1(\beta)+\pi_0 A_2(\beta)] > D_2(\beta)-D(\beta). \tag{10}$$

Due to the assumption that $D(\beta) > D_2(\beta)$, $A_1(\beta)\geq 0$ and $A_2(\beta)\geq 0$, we have $\pi_1 A_1(\beta)+\pi_0 A_2(\beta)\geq 0$ and $D_2(\beta)-D(\beta) < 0$. Therefore, (10) is always true if $D(\beta) > D_2(\beta)$ is true. In other words, if $N_{s,EEOT}^{L}(\beta) > N_{s,EEOT}^{L}(\beta+1)$, we always have $N_{s,EEOT}^{L}(\beta+1) > N_{s,EEOT}^{L}(\beta+2)$. Let $\beta=\beta_l$ be the largest $\beta$ for which $N_{s,EEOT}^{L}(\beta) < N_{s,EEOT}^{L}(\beta+1)$. The above statement is equivalent to that made in Theorem II.2 about the monotonicity of $N_{s,EEOT}(\beta)$. ■

According to Theorem II.2, the optimal solution $\beta^*$ to the optimization problem in (7) is the smallest $\beta$ for which the inequality $D(\beta)\geq D_2(\beta)$ holds, and the LB of the ANT saved is then given as $N_{s,EEOT}^{L} = \bar{N}_{s,1}(\beta^*)+\bar{N}_{s,2}(\beta^*)$. Therefore, the tight UB of the ANT required is $N_{t,EEOT}^{U} = N-N_{s,EEOT}^{L}$.

*Remark.* Although we employ Assumption 1, which states that the signal strength $s$ tends to infinity, when solving the optimization problem in (7), a sufficiently large signal is adequate to exhibit a similar saving performance as that obtained by solving the optimization problem.

## III. Numerical and Simulation Results

In this section, we present some numerical and simulation results to support our theoretical analysis. We assume that $N = 100$ and $s = 20$. Fig. 1 shows the probability of error as a function of $p$ in the EEOT-based system. The system with the threshold closest to the optimal threshold $T^*$ ($T^*$ is roughly $N/2$), as compared to other systems, has the lowest error probability, which is in accordance with the conclusion that we obtained about the optimal threshold $T^*$.[3] We can also observe that for the same parameter values, both EEOT-based and unordered systems have the same probabilities of error. This is in accordance with Lemma II.1.

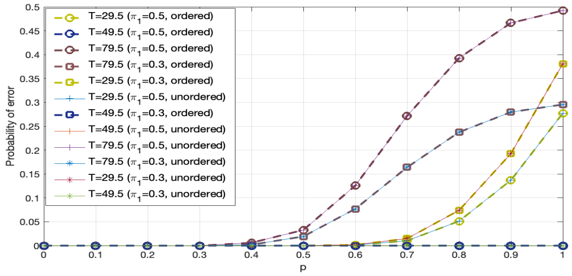

Fig. 1. $P_e$ as a function of $p$ with different values of $T$ for the EEOT-based system for $\pi_1 = 0.3$ and $\pi_1 = 0.5$.

Fig. 2 shows that the UB we obtained is a relatively tight UB compared with the UB obtained in [7] for the average fraction of number of transmissions required as a function of the attacking probability $p$ in the EEOT-based system. Fig. 3 presents the average fraction of transmissions required $N_{t,EEOT}/N$ in the EEOT-based system as a function of $p$ for different values of prior probability and $T$ when $\alpha = 0.3$. We observe from Fig. 3 that when $T \to T^*$ ($T^*$ here is roughly $N/2$), the system is most likely to have the highest transmissions required in the network if the prior probabilities of both hypotheses are 0.5. However, when the prior probabilities change, the value of $T$ that results in the highest transmissions required might also change. It is clear that a smaller $T$ results in a larger ANT required to decide $\mathcal{H}_0$ and a smaller ANT required to decide $\mathcal{H}_1$. For a relatively small $\pi_1 < 0.5$, the probability of the FC deciding $\mathcal{H}_0$ is higher. Consequently, the system that uses $T = 29.5$ has higher number of transmissions required when compared to the system that uses $T = 49.5$ given $\pi_1 = 0.3$. Thus, there is a relationship between the ANT needed and the detection performance of the system. With an appropriately designed threshold at the FC, it is possible to save transmissions while still guaranteeing the quality of the decision. In Fig. 4, we plot $N_{t,EEOT}/N$ (i.e., the ANT required) as a function of $s$ to show that a fairly small value of $s$ is sufficient for the derived result to serve as an UB on the ANT required.

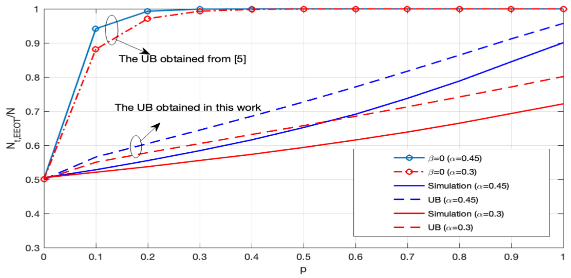

Fig. 2. Benchmarking upper bounds for the fraction of the number of transmissions required $N_{t,EEOT}/N$ as a function of $p$ with different values of $\alpha$ when $\pi_1 = 0.5$. The actual ANT for the system (simulation result in the figure) is obtained via Monte Carlo method given $s = 20$.

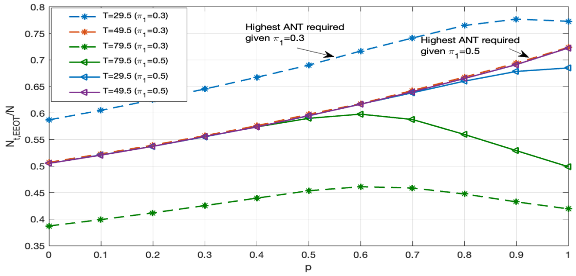

Fig. 3. $N_{t,EEOT}/N$ as a function of $p$ with different values of $T$ when $\alpha = 0.3$ and $\pi_1 = 0.3, 0.5$.

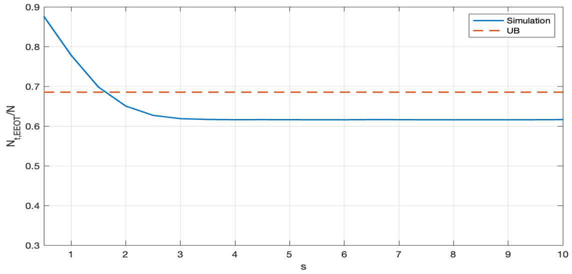

Fig. 4. $N_{t,EEOT}/N$ as a function of $s$ given $p = 0.6$, $N = 100$ and $\pi_1 = 0.5$. Red dashed line indicates the UB of ANT we obtained given $p = 0.6$. Note that for $s > 1.6$, the UB we obtained serves as a valid UB.

## IV. Conclusion

In this work, we considered the data falsification attack problem in an EEOT-based distributed detection system. We evaluated the performance of the EEOT-based system via the detection performance and the maximum number of transmissions required under data falsification attacks. We showed that the detection performance of a system using ordered transmissions is unaffected under data falsification attacks and the probability of error of the EEOT-based system was derived. We also found a tight UB on the ANT required under attack as well as the optimal threshold for the FC. Numerical and simulation results indicate that it is possible to reduce transmissions while still ensuring the quality of the decision with an appropriately designed threshold. In the future, we plan to consider different types of Byzantine attacks where Byzantines alter transmission order as well.

[3] The threshold closest to the optimal threshold $T^*$ is 49.5 in Fig. 1 when $\pi_1 = 0.5$ and $\pi_1 = 0.3$.

# References

[1] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, 1996.

[2] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, vol. 3. IEEE, 2003, pp. 1713–1723.

[3] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4938–4953, 2011.

[4] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3229–3235, 2008.

[5] A. Mohammadi, D. Ciuonzo, A. Khazaee, and P. S. Rossi, "Generalized locally most powerful tests for distributed sparse signal detection," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 528–542, 2022.

[6] S. S. Gupta and N. B. Mehta, "Ordered transmissions schemes for detection in spatially correlated wireless sensor networks," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1565–1577, 2020.

[7] N. Sriranga, K. G. Nagananda, R. S. Blum, A. Saucan, and P. K. Varshney, "Energy-efficient decision fusion for distributed detection in wireless sensor networks," in *2018 21st International conference on information fusion (FUSION)*. IEEE, 2018, pp. 1541–1547.

[8] P. Braca, S. Marano, and V. Matta, "Single-transmission distributed detection via order statistics," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2042–2048, 2011.

[9] L. Hesham, A. Sultan, M. Nafie, and F. Digham, "Distributed spectrum sensing with sequential ordered transmissions to a cognitive fusion center," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2524–2538, 2012.

[10] Y. Chen, R. S. Blum, and B. M. Sadler, "Optimal quickest change detection in sensor networks using ordered transmissions," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020, pp. 1–5.

[11] S. S. Gupta, S. K. Pallapothu, and N. B. Mehta, "Ordered transmissions for energy-efficient detection in energy harvesting wireless sensor networks," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2525–2537, 2020.

[12] Y. Chen, B. M. Sadler, and R. S. Blum, "Ordered gradient approach for communication-efficient distributed learning," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020, pp. 1–5.

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

[14] B. Kailkhura, S. Brahma, and P. K. Varshney, "Data falsification attacks on consensus-based detection systems," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 145–158, 2016.

[15] C. Quan, N. Sriranga, H. Yang, Y. S. Han, B. Geng, and P. K. Varshney, "Efficient ordered-transmission based distributed detection under data falsification attacks," 2022. [Online]. Available: https://arxiv.org/abs/2207.08870

[16] C. Quan, B. Geng, Y. S. Han, and P. K. Varshney, "Enhanced audit bit based distributed bayesian detection in the presence of strategic attacks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 49–62, 2022.

[17] 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.