Optimal Byzantine Attack for Distributed Inference with M-ary Quantized Data

Po-Ning Chen*, Yunghsiang S. Han†, Hsuan-Yin Lin* and Pramod K. Varshney‡
*Dept. of Electrical and Computer Eng., National Chiao-Tung Univ., Taiwan
(Emails: poning@faculty.nctu.edu.tw, lin.hsuanyin@ieee.org)
†Dept. of Elec. Eng., National Taiwan Univ. of Science and Technology, Taiwan (Email: yshan@mail.ntust.edu.tw)
‡Dept. of Electrical Engineering and Computer Science, Syracuse Univ., USA (Email: varshney@syr.edu)

***Abstract*—In many applications that employ wireless sensor networks (WSNs), robustness of distributed inference against Byzantine attacks is important. In this work, distributed inference is considered when local sensors send M-ary data to the fusion center. The optimal Byzantine attack policy is then derived under the assumption that the Byzantine adversary has the knowledge of the statistics of local quantization outputs. Our analysis indicates that the fusion center can be blinded such that the detection error is as poor as a random guess when an adequate fraction of sensors are compromised.**

I. Introduction

Wireless sensor networks (WSNs) have been studied for well over a decade [1]. For applications over WSNs, distributed inference plays an essential role and hence is one of the key problems to be investigated [2], [3]. In a WSN, simple inexpensive sensors are deployed to observe a phenomenon of interest (POI). Due to limited resources at each sensor, the observed local data is quantized into an M-ary symbol for transmission to a fusion center (FC), where $M \geq 2$. A local decision rule is thus required at each sensor to convert the observed data to one of the M symbols. In practice, simple threshold quantizers are commonly employed; and, therefore, the local decision rule can be characterized by a set of $M-1$ quantization thresholds. The sensors then send the resultant M-ary quantization output symbols to the FC, which yields the global inference regarding the POI.

In WSNs, an important issue is the robustness of inference against hostile actions. A lot of research has been devoted to resolving this issue [4], [5], [6]. While most of early works on data security with applications over WSNs focused on preventing or mitigating malicious threats, a recent trend has revolved around how to protect the global inference made by the FC when a fraction of sensors are compromised [7], [8], [9], [10]. In the literature, certain research publications, for analytical convenience, assume that only binary data are transmitted by local sensors [11]. Others such as Marano *el al.* [12] have dealt with non-binary transmissions from local sensors but consider asymptotic blindness of the FC as the number of sensors grows unbounded. In this work, we challenge the problem in a perhaps more practical scenario, where a fixed number of sensors transmit M-ary local decisions to the FC.

In principle, a competent Byzantine attacker may also collect the environmental data associated with compromised sensors and carry joint estimation on the POI. In the extreme case, the compromised nodes could be under a strong Byzantine attack, where with nearly complete knowledge regarding the identity of the compromised sensors as well as the POI estimated, data at local sensors can be accordingly altered by attackers.

In 2014, an optimal Byzantine attack policy for distributed inference sensor networks has been proposed in [13], where the attacker does not have full knowledge about the true state of the POI, or quantization thresholds of the sensors, or their statistics. Instead, the hostile actions at the sensors are restricted only to the modifications of the M-ary symbols transmitted to the FC. This surely limits the capability of a Byzantine adversary who can only conduct the so-called "man-in-the-middle" attack. In this paper, we further extend the work presented in [13] by assuming that the Byzantine adversary is endowed with the knowledge of the statistics of local quantization outputs. Under such circumstance, we found that blinding the FC becomes attainable when an adequate fraction of sensors are compromised even if the number of sensors considered is now fixed. Notably, by blinding the FC, we mean that the global inference at the FC is as poor as a random guess. Details will be given in subsequent sections.

II. System Model and Problem Formation

Consider a WSN, which is designed to estimate a particular phenomenon θ under the premise that local sensors acquire conditionally independent and identically distributed (i.i.d.) observations given $\theta \in \Theta$, where Θ is the sample space of the POI. Among N sensors in this network, we assume that α fraction of them are compromised by an adversary. In this paper, compromised sensors are referred to as *Byzantine sensors* and the remaining as *Honest sensors*. These Byzantine sensors transmit falsified data to the FC in order to deteriorate the performance of the global inference of the WSN.

Denote by r_i the local observation of the ith sensor. The local decision rule then converts r_i to one of the M symbols, denoted as $u_i \in \{1, \cdots, M\}$. Since the transmitted symbol may be different from the local quantization output u_i, we denote by v_i the symbol that is transmitted by the ith sensor. Accordingly, if node i is Honest, then $v_i = u_i$; otherwise, the ith sensor modifies $u_i = \ell$ to $v_i = m$ with probability $p_{\ell,m}(\theta)$.

978-1-5090-1806-2/16/$31.00 ©2016 IEEE

As a result, the Byzantine transition probability from u_i to v_i can be modeled using a row-stochastic matrix:

$$\mathbb{P}(\theta) \triangleq \begin{bmatrix} p_{1,1}(\theta) & p_{1,2}(\theta) & \dots & p_{1,M}(\theta) \\ p_{2,1}(\theta) & p_{2,2}(\theta) & \dots & p_{2,M}(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,1}(\theta) & p_{M,}(\theta) & \dots & p_{M,M}(\theta) \end{bmatrix}.$$

We assume that the adversary has the knowledge of the local decision rules employed by compromised sensors, or specifically the statistics of u_i; hence, he can adjust $\mathbb{P}(\theta)$ according to the probability mass function (pmf) of u_i, which is denoted as $c_m(\theta) \triangleq \Pr(u_i = m|\theta)$. Without loss of generality, we suppose $c_m(\theta) > 0$ for $1 \le m \le M$ since the attacker can exclude those rows and columns in $\mathbb{P}$ corresponding to zero-valued $c_m(\theta)$ and design a $\mathbb{P}(\theta)$ of smaller size to blind the FC. For notational convenience, we will ignore the parameter θ and simply write $p_{\ell,m}(\theta)$ and $c_m(\theta)$ as $p_{\ell,m}$ and c_m, respectively, in later derivations.

Additionally, we denote the transition probability of the discrete noisy link between a sensor and the FC by:

$$\mathbb{Q} \triangleq \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,M} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ q_{M,1} & q_{M,2} & \cdots & q_{M,M} \end{bmatrix},$$

where $q_{\ell,m}$ is the probability of $v_i = \ell$ being converted to symbol $z_i = m$ during the noisy transmission. It is reasonable to assume that the noisy link is independent of the POI and hence $\mathbb{Q}$ remains invariant when the value of θ varies. From elementary probability theory, $\mathbb{P}$, $\mathbb{Q}$ and $\mathbf{c} \triangleq [c_1\ c_2\ \cdots\ c_M]^\intercal$ must satisfy

$$\begin{cases} \mathbb{P}\mathbf{1} = \mathbf{1} \text{ with } 0 \le p_{\ell,m} \le 1 \text{ for } 1 \le \ell, m \le M; \\ \mathbb{Q}\mathbf{1} = \mathbf{1} \text{ with } 0 \le q_{\ell,m} \le 1 \text{ for } 1 \le \ell, m \le M; \\ \mathbf{1}^\intercal \mathbf{c} = 1 \text{ with } 0 < c_m < 1 \text{ for } 1 \le m \le M, \end{cases}$$

where superscript "$\intercal$" is the matrix transpose operation and $\mathbf{1}$ is the all-one column vector.

With the above setting, together with the assumption that α fraction of sensors are compromised, the conditional probability of $z_i = m$ given phenomenon θ can be obtained as follows:

$$\begin{aligned} \Pr(z_i = m|\theta) &= \sum_{j=1}^{M} q_{j,m} \Pr(v_i = j|\theta) \\ &= \sum_{j=1}^{M} q_{j,m} \Big(\alpha \Pr(v_i = j | i = \text{Byzantine}, \theta) \\ &\quad + (1-\alpha) \Pr(v_i = j | i = \text{Honest}, \theta) \Big) \\ &= \alpha \sum_{j=1}^{M} q_{j,m} \sum_{\ell=1}^{M} \Pr(v_i = j | u_i = \ell, \theta) \cdot \Pr(u_i = \ell|\theta) \\ &\quad + (1-\alpha) \sum_{j=1}^{M} q_{j,m} \Pr(u_i = j|\theta) \\ &= \alpha \left(\sum_{j=1}^{M} q_{j,m} \sum_{\ell=1}^{M} p_{\ell,j} c_\ell - \sum_{j=1}^{M} q_{j,m} c_j \right) + \sum_{j=1}^{M} q_{j,m} c_j. \end{aligned} \quad (1)$$

A Byzantine attack is targeted to make z_i and θ statistically independent (and hence blind the FC) with the least amount of effort (i.e., with minimum α). This can be characterized as:

$$\Pr(z_i = m|\theta) = b_m, \quad \forall\, 1 \le m \le M, \quad (2)$$

for some pmf $\mathbf{b} = [b_1\, b_2\, \cdots b_M]^\intercal$, independent of θ. With this objective in mind, together with (1), the problem that this paper focuses on is to find the minimum $\alpha = \alpha_{\text{blind}}$ subject to

$$\sum_{j=1}^{M} q_{j,m} c_j - b_m = \alpha \left(\sum_{j=1}^{M} q_{j,m} c_j - \sum_{j=1}^{M} q_{j,m} \sum_{\ell=1}^{M} p_{\ell,j} c_\ell \right)$$

for all $1 \le m \le M$. In matrix form, the above constraint can be expressed as:

$$\mathbb{Q}^\intercal \mathbf{c} - \mathbf{b} = \alpha \mathbb{Q}^\intercal \big(\mathbb{I} - \mathbb{P}^\intercal\big)\mathbf{c}, \quad (3)$$

where $\mathbb{I}$ is the identity matrix of proper size.

III. Main Theorem

Instead of determining directly the minimum α that satisfies (3) over all possible choices of $\mathbf{b}$ in the sense of (2), one can divide the task into two subtasks. First, find the minimum $\alpha = \alpha_{\text{blind}}(\mathbf{b})$ for a specific $\mathbf{b}$ and then minimize $\alpha_{\text{blind}}(\mathbf{b})$ over all possible choices of $\mathbf{b}$. Notably, any choice of $\mathbf{b}$ will yield the same "random guess" error performance $1 - 1/|\Theta|$ as long as the resultant $\alpha_{\text{blind}}(\mathbf{b})$ is attainable.

We now conduct the first subtask for a choice of uniform $\mathbf{b}$ that is perhaps common in the literature, i.e., $b_m = 1/M$, $\forall\, 1 \le m \le M$. Since in usual situations of practical interest, $\mathbb{Q}$ admits an inverse, we can simplify (3) to

$$\mathbf{c} - \mathbf{d} = \alpha\big(\mathbb{I} - \mathbb{P}^\intercal\big)\mathbf{c}, \quad (4)$$

where

$$\mathbf{d} \triangleq \frac{1}{M}(\mathbb{Q}^\intercal)^{-1}\mathbf{1}, \quad (5)$$

and hence transform the problem to determining the minimum $\alpha_{\text{blind}}(\mathbf{b})$ that satisfies (4) subject to all legitimate $\mathbb{P}$.

Before we state the main theorem, some preliminary derivations are necessary. We first note that $\mathbb{Q}\mathbf{1} = \mathbf{1}$ implies $\mathbf{1}^\intercal\big(\mathbb{Q}^\intercal\big)^{-1} = \mathbf{1}^\intercal$, based on which we derive $\mathbf{1}^\intercal\mathbf{d} = \frac{1}{M}\mathbf{1}^\intercal\big(\mathbb{Q}^\intercal\big)^{-1}\mathbf{1} = \frac{1}{M}\mathbf{1}^\intercal\mathbf{1} = 1$. Thus $\sum_{m=1}^{M}(c_m - d_m) = 1 - 1 = 0$. As a result, $\max_{1\le m\le M}\{c_m - d_m\} \ge 0$. For the trivial case of $\max_{1\le m\le M}\{c_m - d_m\} = 0$ (equivalently, $\mathbf{c} = \mathbf{d}$), (4) immediately implies $\alpha_{\text{blind}}(\mathbf{b}) = 0$. It remains to examine the case of $\max_{1\le m\le M}\{c_m - d_m\} > 0$.

Divide the index set $\{1, 2, \dots, M\}$ into two groups. The first group contains those satisfying $\max\{c_m - d_m, 0\} = 0$, while the remaining belong to the second group. Let $m^\diamond$ be the number of elements in the first group. The condition of $\max_{1\le m\le M}\{c_m - d_m\} > 0$ then implies that $1 \le m^\diamond < M$, which guarantees that none of the two groups are empty. Without loss of generality, we index the numbers in the first group as $1, 2, \dots, m^\diamond$, and for those in the second group, we

assume that $e_{m^\diamond+1} \le e_{m^\diamond+2} \le \cdots \le e_M$, where $e_m \triangleq 1-d_m/c_m$ for $m^\diamond < m \le M$. Note that $\max\{c_m-d_m,0\}>0$ implies $e_m > 0$ for those m's in the second group. We then have the following theorem.

Theorem 1: If the adversary knows the conditional pmf of local quantization output $\mathbf{c}$ for a given phenomenon θ and the transition probability of the noisy link $\mathbb{Q}$ admits an inverse, then for uniform $\mathbf{b}$ that induces $\mathbf{d}$ in (5),

$$\begin{aligned}\alpha_{\text{blind}}(\mathbf{b}) &\triangleq \min\big\{\alpha\in[0,1): \mathbf{c}-\mathbf{d}=\alpha\big(\mathbb{I}-\mathbb{P}^\intercal\big)\mathbf{c} \text{ for some } \mathbb{P}\big\}\\ &= e_M,\end{aligned} \tag{6}$$

provided that $e_M \le 1$. Furthermore, (6) can be achieved by $\mathbb{P}^*$ with its matrix entries defined as

$$p^*_{\ell,j} = \begin{cases} 1, & 1\le \ell = j \le m^\diamond;\\ 1-\frac{e_\ell}{e_M}, & m^\diamond < \ell = j < M;\\ \frac{(d_j-c_j)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_\ell}{e_M}, & 1\le j\le m^\diamond<\ell\le M;\\ 0, & \text{otherwise.}\end{cases} \tag{7}$$

Proof: The theorem can be proved in two steps. The first step shows that every α satisfying $\mathbf{c}-\mathbf{d}=\alpha\big(\mathbb{I}-\mathbb{P}^\intercal\big)\mathbf{c}$ for some $\mathbb{P}$ must be no less than e_M. The second step gives a feasible choice of $\mathbb{P}$ that verifies the achievability of the claimed $\alpha_{\text{blind}}(\mathbf{b})=e_M$.

A. $\alpha_{\text{blind}}(\mathbf{b}) \ge e_M$

From (4), we know that

$$\begin{aligned} c_M - d_M &= \alpha\bigg(c_M - \sum_{m=1}^{M} p_{m,M}c_m\bigg)\\ &= \alpha\bigg[(1-p_{M,M})c_M - \sum_{m=1}^{M-1} p_{m,M}c_m\bigg].\end{aligned}$$

This implies that

$$\begin{aligned}\alpha \ge \alpha(1-p_{M,M}) &= 1-\frac{d_M}{c_M} + \underbrace{\alpha\sum_{m=1}^{M-1} p_{m,M}\frac{c_m}{c_M}}_{\ge 0}\\ &\ge 1-\frac{d_M}{c_M} = e_M.\end{aligned} \tag{8}$$

B. $\alpha_{\text{blind}}(\mathbf{b}) = e_M$

By letting $\mathbb{I}$ and $\mathbb{O}$ be respectively the identity matrix and the all-zero matrix of proper sizes designated by their subscripts, the proposed $\mathbb{P}^*$ can be written as:

$$\mathbb{P}^* = \begin{bmatrix} \mathbb{I}_{m^\diamond\times m^\diamond} & \mathbb{O}_{m^\diamond\times(M-m^\diamond)}\\ \mathbb{K}_{(M-m^\diamond)\times m^\diamond} & \mathbb{L}_{(M-m^\diamond)\times(M-m^\diamond)}\end{bmatrix},$$

where

$$\mathbb{K} \triangleq \begin{bmatrix} \frac{(d_1-c_1)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+1}}{e_M} & \frac{(d_2-c_2)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+1}}{e_M}\\ \frac{(d_1-c_1)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+2}}{e_M} & \frac{(d_2-c_2)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+2}}{e_M}\\ \vdots & \vdots\\ \frac{(d_1-c_1)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_M}{e_M} & \frac{(d_2-c_2)}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_M}{e_M}\end{bmatrix}$$

$$\cdots \quad \frac{(d_{m^\diamond}-c_{m^\diamond})}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+1}}{e_M}$$
$$\cdots \quad \frac{(d_{m^\diamond}-c_{m^\diamond})}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_{m^\diamond+2}}{e_M}$$
$$\vdots \qquad \vdots$$
$$\cdots \quad \frac{(d_{m^\diamond}-c_{m^\diamond})}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{e_M}{e_M}\Bigg],$$

and

$$\mathbb{L} \triangleq \begin{bmatrix} 1-\frac{e_{m^\diamond+1}}{e_M} & 0 & \cdots & 0\\ 0 & 1-\frac{e_{m^\diamond+2}}{e_M} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1-\frac{e_M}{e_M}\end{bmatrix}.$$

We then validate:

$$\begin{aligned} e_M\big(\mathbb{P}^*\big)^\intercal\mathbf{c} &= e_M\begin{bmatrix} c_1 + \frac{\sum_{m=m^\diamond+1}^{M} c_m e_m}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{(d_1-c_1)}{e_M}\\ \vdots\\ c_{m^\diamond} + \frac{\sum_{m=m^\diamond+1}^{M} c_m e_m}{\sum_{m=1}^{m^\diamond}(d_m-c_m)}\frac{(d_{m^\diamond}-c_{m^\diamond})}{e_M}\\ c_{m^\diamond+1} - \frac{c_{m^\diamond+1}e_{m^\diamond+1}}{e_M}\\ \vdots\\ c_{M-1} - \frac{c_{M-1}e_{M-1}}{e_M}\\ c_M - \frac{c_M e_M}{e_M}\end{bmatrix}\\ &= \begin{bmatrix} c_1e_M + (d_1-c_1)\\ \vdots\\ c_{m^\diamond}e_M + (d_{m^\diamond}-c_{m^\diamond})\\ c_{m^\diamond+1}e_M - (c_{m^\diamond+1}-d_{m^\diamond+1})\\ \vdots\\ c_{M-1}e_M - (c_{M-1}-d_{M-1})\\ c_Me_M - (c_M-d_M)\end{bmatrix} \qquad (9)\\ &= e_M\mathbf{c} + \mathbf{d} - \mathbf{c},\end{aligned}$$

where (9) follows from

$$\begin{aligned} 0 = \sum_{m=1}^{M}(d_m-c_m) &= \sum_{m=1}^{m^\diamond}(d_m-c_m) + \sum_{m=m^\diamond+1}^{M}(d_m-c_m)\\ &= \sum_{m=1}^{m^\diamond}(d_m-c_m) - \sum_{m=m^\diamond+1}^{M} c_m e_m.\end{aligned}$$

The proof is accordingly completed. ∎

Several remarks are made based on the above theorem. First, for noiseless wireless links, we have $\mathbb{Q}=\mathbb{I}$; in this case, $\mathbf{d} \triangleq \frac{1}{M}\big(\mathbb{Q}^\intercal\big)^{-1}\mathbf{1} = \frac{1}{M}\mathbf{1}$ and hence

$$e_M = 1-\frac{d_M}{c_M} = 1-\frac{1}{Mc_M} = 1-\frac{1}{M\max_{1\le m\le M} c_m}$$

is no larger than unity. As a result, with additional knowledge of the POI θ and its associated $\mathbf{c}(\theta)$, the Byzantine adversary can equate the conditional pmf of z_i given θ to the targeted uniform distribution, and the sensor network system is blinded since for any $\theta_1 \ne \theta_2$ in Θ, $\Pr(z_i=m|\theta_1) = \Pr(z_i=m|\theta_2) = \frac{1}{M}$, $\forall\, 1\le m\le M$. In fact, it can be verified that $e_M \le 1$ as long as $\mathbf{d}$ consists of non-negative

components. It however could happen that $\mathbf{d} = \frac{1}{M}(\mathbb{Q}^\intercal)^{-1}\mathbf{1}$ has negative components for certain $\mathbb{Q}$, which may result in $e_M > 1$. As a consequence, (8) indicates that there exists no $\alpha \in [0,1)$ that fulfills (4). It is therefore impossible to blind the sensor network system in the sense of (2) for every $\theta \in \Theta$. In such case, an alternative $\mathbf{b}$ other than the uniform one may need to be employed.

Secondly, $\mathbf{d}$ in constraint (4) can be regarded as the target distribution that the Byzantine attacker intends to maliciously change $\mathbf{c}$ to. In an extreme situation, we have $\alpha_{\text{blind}}(\mathbf{b}) = 0$ when $\mathbf{d} = \mathbf{c}$, and hence no Byzantine effort is necessary.

Thirdly, the optimal $\mathbb{P}^*$ is not unique! It can be shown that any $\mathbb{P}^*$ satisfying

$$\begin{cases} p^*_{\ell,m} = 0, & \text{for } 1 \le \ell \le m^\diamond \text{ and } 1 \le m \le M \text{ and } \ell \ne m; \\ & \text{also for } m^\diamond < \ell < M \text{ and } m^\diamond < m \le M \text{ and } \ell \ne m; \\ \displaystyle\sum_{\ell=m^\diamond+1}^{M-1} p^*_{\ell,m} c_\ell \le \frac{d_m - c_m}{e_M}, & \text{for } 1 \le m \le m^\diamond; \\ \displaystyle\sum_{j=1}^{m^\diamond} p^*_{m,j} = \frac{e_m}{e_M}, & \text{for } m^\diamond < m < M \end{cases}$$

with the remaining terms obtained from $\mathbb{P}^*\mathbf{1} = 1$ (which leads to $p^*_{m,m} = 1$ for $1 \le m \le m^\diamond$ and $p^*_{m,m} = 1 - e_m/e_M$ for $m^\diamond < m < M$) and $\mathbf{c} - \mathbf{d} = \alpha\big(\mathbb{I} - \mathbb{P}^\intercal\big)\mathbf{c}$ (which leads to $p^*_{M,m} = 0$ for $m^\diamond < m \le M$) is a valid solution when $\mathbf{b}$ is a uniform pmf. This gives more freedom to the Byzantine attacker, in particular when a certain group of θ's can be made to correspond to a common system neutralizer $\mathbb{P}(\theta)$.

Last, after refining (5) as $\mathbf{d} \triangleq \big(\mathbb{Q}^\intercal\big)^{-1}\mathbf{b}$ for a general $\mathbf{b}$ (not necessarily uniform), Theorem 1 can still be applied with $e_M = \max_{1\le m\le M}\{1 - d_m/c_m\}$. Thus, our theorem can actually give a very general procedure about how to equate the conditional pmfs of z_i given θ for distinct θ. An example will be given in the next section.

IV. Numerical Results

To examine the performance deterioration of global inference due to Byzantine adversaries, we retain the inference detection model from [13]. Let the local observation of the ith sensor be modeled by

$$r_i = s(\theta) + a_i, \qquad i = 1, 2, \ldots, N,$$

where with $\theta \in \Theta = \{0,1\}$, $s(\theta) = \mu \cdot (-1)^{1+\theta}$ is an antipodally modulated signal to be estimated, and a_i is zero-mean Gaussian distributed with variance σ^2_{sen}. By using simple threshold quantizers, the conditional pmf of u_i given θ can be obtained by the rule that $u_i = m$ if $\eta_{m-1} < r_i \le \eta_m$ with

$$\eta_m \triangleq \begin{cases} -\infty, & m = 0; \\ A_M \cdot (2m - M), & 1 \le m < M; \\ \infty, & m = M, \end{cases}$$

where A_M is a finite number satisfying $(M-2)\cdot A_M = A$ and A is an *overloading* parameter [14]. Based on this information, the Byzantine adversary devises the flipping probability matrix $\mathbb{P}^*(\theta)$ according to (7) respectively for $\theta \in \{0,1\}$ and converts u_i to v_i probabilistically. The M-ary signal v_i is then sent to the FC via a noisy link characterized by the transition probability matrix $\mathbb{Q}$.

Subject to amplitude shift keying, z_i is assumed to be the quantization output due to input

$$y_i = (2v_i - M - 1) + n_i, \quad v_i = 1, 2, \ldots, M,$$

with thresholds

$$\lambda_m = \begin{cases} -\infty, & m = 0; \\ 2m - M, & 1 \le m < M; \\ \infty, & m = M, \end{cases}$$

where n_i is zero-mean Gaussian distributed with variance σ^2. In other words, $z_i = m$ if $\lambda_{m-1} < y_i \le \lambda_m$. For example, under $M = 4$, we have

$$\mathbb{Q} = \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 - \epsilon_3 & \epsilon_3 - \epsilon_5 & \epsilon_5 \\ \epsilon_1 & 1 - 2\epsilon_1 & \epsilon_1 - \epsilon_3 & \epsilon_3 \\ \epsilon_3 & \epsilon_1 - \epsilon_3 & 1 - 2\epsilon_1 & \epsilon_1 \\ \epsilon_5 & \epsilon_3 - \epsilon_5 & \epsilon_1 - \epsilon_3 & 1 - \epsilon_1 \end{bmatrix},$$

where $\epsilon_k \triangleq \Phi(-k/\sigma)$ and Φ is the standard normal cumulative distribution function. The Bayes detection error probability of the sensor network is equal to

$$P_{\text{e}} \triangleq \frac{1}{2}\Pr\Big(\hat\theta = 1 \Big| \theta = 0\Big) + \frac{1}{2}\Pr\Big(\hat\theta = 0 \Big| \theta = 1\Big),$$

where we assume $\Pr(\theta = 0) = \Pr(\theta = 1) = \frac{1}{2}$, and $\hat\theta$ is the global inference made by the FC based on the reception $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$. Note that since the FC is not aware of the presence of the Byzantine adversary, it makes the global inference, assuming that the conditional pmf of z_i given θ were $\sum_{j=1}^{M} q_{j,m} c_j$. The Byzantine then tries to blind the system with the uniform $\mathbf{b}$.

We illustrate the detection error P_{e} under $N = 10$, $\mu = 1$, $\sigma^2_{\text{sen}} = 1$, $\sigma^2 = 4$ and $A = 2$ as a function of α in Fig. 1. For convenience, we regard α as a real number rather than a multiple of $1/N$. Four local quantization resolutions respectively corresponding to $M = 2$, 4, 8 and 16 are examined. Since $\alpha_{\text{blind}}(\mathbf{b}) = \alpha_{\text{blind}}(\mathbf{b}|\theta)$ is a function of θ, what we present in this figure is $\alpha^*_{\text{blind}}(\mathbf{b}) \triangleq \max_{\theta\in\Theta} \alpha_{\text{blind}}(\mathbf{b}|\theta)$.[1] As expected, we observe that $\alpha^*_{\text{blind}}(\mathbf{b})$ increases as M grows. Thus a higher quantization resolution will increase the blinding effort of a Byzantine attacker. For example, only 4 sensors need to be compromised for perfect blinding when $M = 2$ as contrasted with 8 compromised sensors at $M = 8$ and $M = 16$.

Notably, we get $e_M > 1$ when $\sigma^2 = 9$ and $M \ge 4$; hence, making z_i conditionally uniform distributed given θ becomes impossible. Instead, the adversary should target $\Pr(z_i = m\,|\,\theta = 0) = \Pr(z_i = m\,|\,\theta = 1) = b_m$ for some $\mathbf{b} = [b_1\ b_2\ \cdots\ b_M]^\intercal$ other than the uniform distribution. With $\mathbf{b} = [0.2642\ 0.2004\ 0.2106\ 0.3248]^\intercal$, we plot the detection error in Fig. 2. It can be observed that the resulting new

[1] In this special example, we actually have $\alpha_{\text{blind}}(\mathbf{b}|0) = \alpha_{\text{blind}}(\mathbf{b}|1)$.

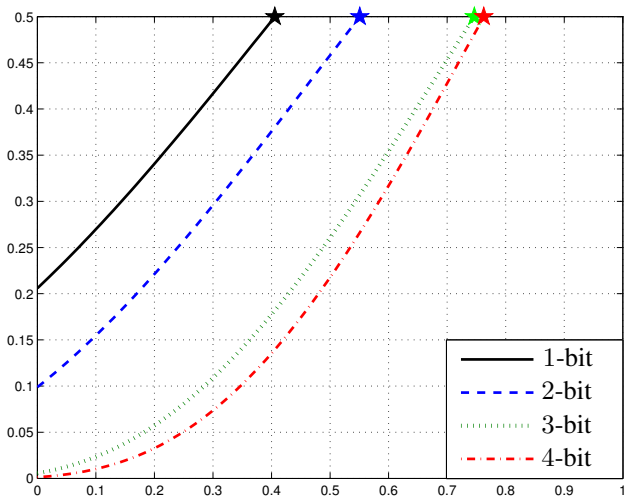

Fig. 1: Detection error probability P_{e} as a function of α under a fixed $\mathbb{P}^*(\theta)$ from (7). The pentagrams mark the values of $\alpha^*_{\mathrm{blind}}(\mathbf{b})$ for different local quantization resolutions, where $\mathbf{b}$ is a uniform pmf.

$\alpha^*_{\mathrm{blind}}(\mathbf{b}) \approx 0.9915$ is prohibitively high and the Byzantine adversary actually needs to compromise all of the ten sensors (as $\alpha^*_{\mathrm{blind}}(\mathbf{b})\, N \approx 9.915$) to blind the sensor network system.

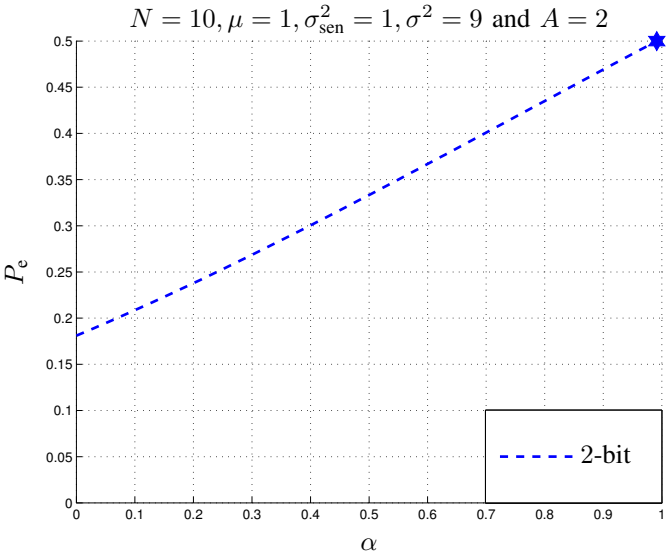

Fig. 2: Detection error probability P_{e} as a function of α under a fixed $\mathbb{P}^*$ from (7). The hexagram marks the value of $\alpha^*_{\mathrm{blind}}(\mathbf{b}) \approx 0.9915$ for local quantization resolution of $M = 4$, where $\mathbf{b} = [0.2642\ 0.2004\ 0.2106\ 0.3248]^{\mathsf{T}}$.

V. Conclusion

The optimal Byzantine attack policy for WSNs with M-ary quantized data was derived under the assumption that the adversary can acquire the statistics of local quantization outputs. Our analysis indicated that perfect blinding of global inference can be achieved when $\alpha^*_{\mathrm{blind}}(\mathbf{b})$ fraction of sensors are compromised. A numerical experiment showed that in certain situations, we may have $e_M > 1$ for the uniform $\mathbf{b}$, and perfect blinding of the sensor network system does not seem possible. Further investigations, however, reveals that by adopting an alternative $\mathbf{b}$, perfect blinding can still be achieved. A future work that would be important from a Byzantine adversary viewpoint is to know how to minimize the blinding effort by determining

$$\min_{\mathbf{b}\in\mathcal{R}^M\,:\,\text{each } b_i\geq 0 \text{ and } \sum_{i=1}^M b_i=1} \alpha^*_{\mathrm{blind}}(\mathbf{b}).$$

Acknowledgment

This work was supported by the Minister of Science and Technology (MOST) of Taiwan under Grants MOST 103-2911-I-011-515 and MOST 104-2911-I-011-503.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, New York, 1997.

[3] R. R. Brooks, P. Ramanathan, and A. M. Sayeed, "Distributed target classification and tracking in sensor networks," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1163–1171, 2003.

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

[5] C. Karlof, N. Sastry, and D. Wagner, "Tinysec: a link layer security architecture for wireless sensor networks," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, ser. SenSys '04, New York, NY, USA, 2004, pp. 162–175.

[6] W. Du, J. Deng, Y. S. Han, P. K. Varshney, J. Katz, and A. Khalili, "A pairwise key predistribution scheme for wireless sensor networks," *ACM Trans. Inf. Syst. Secur.*, vol. 8, no. 2, pp. 228–258, May 2005.

[7] 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 Trans. Signal Process.*, vol. 59, no. 2, pp. 774–786, 2011.

[8] A. Vempaty, K. Agrawal, H. Chen, and P. Varshney, "Adaptive learning of Byzantines' behavior in cooperative spectrum sensing," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, 2011, pp. 1310–1315.

[9] B. Kailkhura, S. Brahma, and P. K. Varshney, "Optimal Byzantine attack on distributed detection in tree based topologies," in *Proc. of International Conference on Computing, Networking and Communications Workshops (ICNC-CPS)*, San Diego, USA, January 2013.

[10] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 205–215, 2013.

[11] C. Yao, P.-N. Chen, T.-Y. Wang, Y. S. Han, and P. K. Varshney, "Performance analysis and code design for minimum hamming distance fusion in wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1716–1734, 2007.

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

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

[14] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.