# CYBER ATTACKS ON ESTIMATION SENSOR NETWORKS AND IOTS: IMPACT, MITIGATION AND IMPLICATIONS TO UNATTACKED SYSTEMS

*Jiangfan Zhang*[⋆] *, Rick S. Blum*[†] *, and Lance Kaplan*[‡]

[⋆]EE dapartment, Columbia University, New York, NY, 10027, USA
[†]ECE department, Lehigh University, Bethlehem, PA, 18015, USA
[‡]U.S. Army Research Lab., 2800 Powder Mill Road, Adelphi, MD, 20783, USA
Email: jiangfan.zhang@columbia.edu, rblum@ece.lehigh.edu, lance.m.kaplan.civ@mail.mil

## ABSTRACT

Estimation of an unknown deterministic vector from quantized sensor data is considered in the presence of spoofing and man-in-the-middle attacks. First, asymptotically optimum processing, which identifies and categorizes the attacked sensors into different groups according to distinct types of attacks, is outlined in the face of man-in-the-middle attacks. Necessary and sufficient conditions are provided under which utilizing the attacked sensor data will lead to better estimation performance when compared to approaches where the attacked sensors are ignored. Next, necessary and sufficient conditions are provided under which spoofing attacks provide a guaranteed attack performance in terms of the Cramer-Rao Bound regardless of the processing the estimation system employs. It is shown that it is always possible to construct such a highly desirable attack by properly employing an attack vector parameter having a sufficiently large dimension relative to the number of quantization levels employed, which was not observed previously. For unattacked quantized estimation systems, a general limitation on the dimension of a vector parameter which can be accurately estimated is uncovered.

***Index Terms***— Distributed parameter estimation, man-in-the-middle attack, spoofing attack, Cramer-Rao Bound, sensor network.

## 1. INTRODUCTION

The emerging revolution impacting the topics of sensor networking, the internet of things (IoT), enhanced data monitoring and data processing for enhanced situation awareness continues to magnify the impact that sensor data can have in our daily lives. Applications such as disaster prediction, security, smart cities, enhanced building operation for optimized energy usage, health monitoring, monitoring for assisted living, and smart transportation systems, among others, promise tremendous positive impact on our daily lives [1]. While the internet has been available for many years, the integration of sensing technology into the internet is still very immature and brings new problems that have not yet been addressed. The vision of inexpensive sensor nodes is being realized. For example powerful radar technology, including MIMO radar technology, is currently being incorporated in future products by virtually every car manufacturer. They are designing very inexpensive radar systems based on mass produced integrated circuits which will encourage the incorporation of radars in many commercial products and ultimately mass adoption. The products being offered through the internet of things roll out will accelerate this phenomenon and encourage similar inexpensive sensor system development of other currently expensive sensors. However, with all of this comes the increased risk of cyber attacks on theses systems.

The topic of cyber attacks on sensor systems has received much less attention than the topic of cyber attacks on other systems, but the increasing adoption of sensor and internet of things networks makes this a very important issue. We have already heard rumors of attacks on automotive radars and related GPS systems. Such attacks could lead to loss of human lives so these attacks are a very serious threat. In many sensor networking and internet of things applications, it is desired to estimate some quantity, possibly the position or velocity of a human. Here we focus on systems performing estimations and study the impact and mitigation of intruders altering the data entering or leaving the sensors. In fact, if the intruders modify the data entering a sensor, we call it a spoofing attack. If the intruders modify the data leaving a sensor node, we call it a man-in-the middle attack. Our focus here is somewhat on large sensor networks that seem to be coming and may be the most vulnerable, although lessons for smaller systems are also

This material is based upon work partially supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-14-1-0245 and by Pennsylvania Infrastructure Technology Alliance (PITA), a partnership of Carnegie Mellon University, Lehigh University, Bethlehem, PA, USA, and the Commonwealth of Pennsylvania, Department of Economic and Community Development (DCED). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, Army Research Office, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

uncovered. Typically, large sensor networks are comprised of low-cost and spatially distributed sensor nodes with limited battery power and low computing capacity, which makes them particularly vulnerable to cyberattacks by adversaries. This has led to great interest in studying the vulnerability of sensor networks in various applications and from different perspectives, see [2–9] and the references therein. Due to the dominance of digital technology, a great deal of attention has focused on parameter estimation using quantized data [10–14].

## 2. MAN-IN-THE-MIDDLE ATTACKS

Consider a set of $N$ distributed sensors, each making $K$ observations of a deterministic scalar parameter $\theta$ corrupted by additive noise. At the $j$-th sensor, the observation at the $k$-th time instant is described by

$$x_{jk} = \theta + n_{jk},\ \forall j = 1, 2, ..., N,\ \forall k = 1, 2, ..., K, \quad (1)$$

where $n_{jk}$ denotes the additive noise sample with common zero-mean probability density function (pdf) $f(n_{jk})$ and $\{n_{jk}\}$ is an independent and identically distributed sequence. Due to the stringent energy and bandwidth limitations in realistic sensor networks, each sensor is restricted to transmit a single bit per observation $x_{jk}$ to the fusion center (FC). In this section, to simplify the problem in terms of both implementation and analysis, all $x_{jk}$ are quantized to $u_{jk}$ by using threshold quantizers of the same design

$$u_{jk} = \mathbb{1}\left\{x_{jk} \in (\tau, \infty)\right\}. \quad (2)$$

We assume that the quantizer design and the threshold $\tau$ is known to the FC.

Let there be $P$ distinct types of malicious attacks, where each attack will modify some sensors' observations. Let $\mathcal{A}_p$ denote the set of sensors subjected to the $p$-th attack. Let $\tilde{u}_{jk}$ represent the after-attack quantized observation which is a modified version of $u_{jk}$. The statistical description of the $p$-th attack can be described by a probability transition matrix $\mathbf{\Psi}_p$,

$$\mathbf{\Psi}_p \triangleq \left[\begin{array}{cc} \psi_{p,0} & 1-\psi_{p,1} \\ 1-\psi_{p,0} & \psi_{p,1} \end{array}\right], \quad (3)$$

where $\psi_{p,0} \triangleq \Pr\left(\tilde{u}_{jk} = 0 \,|u_{jk} = 0\right)$ and $\psi_{p,1} \triangleq \Pr\left(\tilde{u}_{jk} = 1 \,|u_{jk} = 1\right)$ determine the modification probabilities. Due to the $p$-th attack, the after-attack probability mass function (pmf) of the observations can be related to the before-attack pmf using

$$\left[\begin{array}{c} 1-\tilde{p}\left(\mathbf{\Psi}_p, \theta\right) \\ \tilde{p}\left(\mathbf{\Psi}_p, \theta\right) \end{array}\right] \triangleq \left[\begin{array}{c} \Pr\left(\tilde{u}_{jk} = 0 \,|\theta\right) \\ \Pr\left(\tilde{u}_{jk} = 1 \,|\theta\right) \end{array}\right] = \mathbf{\Psi}_p \left[\begin{array}{c} \Pr\left(u_{jk} = 0 \,|\theta\right) \\ \Pr\left(u_{jk} = 1 \,|\theta\right) \end{array}\right] \quad (4)$$

For the sake of expressing the after-attack pmfs of observations in a uniform form for both attacked and unattacked sensors, the set $\mathcal{A}_0$ of unattacked sensors are considered "under attack" associated with probability transition matrix $\mathbf{\Psi}_0 = \mathbf{I}$. The following assumption is made through this paper.

**Assumption 1**

1. *Over the estimation time interval and for all $p$, the $p$-th attack is statistically described as in (4) for all the sensors in the set $\mathcal{A}_p$. The set $\mathcal{A}_p$ and $\mathbf{\Psi}_p$ are both unknown to the FC (except $\mathbf{\Psi}_0$), and for sufficiently large $N$ the number of sensors in $\mathcal{A}_p$, $|\mathcal{A}_p|$, is a fixed percentage $\mathcal{P}_p$ of the total number $N$ of sensors in the sensor network. Such an assumption is required so that as $N \to \infty$ the effect of an attack will not shrink to zero ($\mathcal{A}_p$ becoming a set of measure zero). Moreover, we assume that the group of unattacked sensors is the largest group and $\mathcal{P}_0 > \mathcal{P}_p + \Delta_0$ for all $p \geq 1$ where $\Delta_0$ is a positive constant. Further the sets $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_P$ are disjoint so that*

$$\mathcal{A}_p \cap \mathcal{A}_{p'} = \emptyset \quad \text{if} \quad p \neq p'. \quad (5)$$

2. Significant Attacks. *In order to give rise to sufficient impact on the statistical characterization of the outputs from attacked sensors, every attacker is required to guarantee a minimum distortion $d_{impact}$ on $\tilde{p}\left(\mathbf{\Psi}_0, \theta\right)$ and tamper with at least $\Delta$ percent of sensors so that the following relations should be satisfied*

$$\left|\tilde{p}\left(\mathbf{\Psi}_p, \theta\right) - \tilde{p}\left(\mathbf{\Psi}_0, \theta\right)\right| \geq d_{impact},\ \forall p = 1, 2, ..., P, \quad (6)$$

$$\mathcal{P}_p \geq \Delta > 0, \quad \forall p = 1, 2, ..., P. \quad (7)$$

3. Various Attacks. *The changes caused by two distinct types of attacks are considerably different, otherwise these two types of attacks can be treated as identical. To this end, we assume that*

$$\left|\tilde{p}\left(\mathbf{\Psi}_l, \theta\right) - \tilde{p}\left(\mathbf{\Psi}_m, \theta\right)\right| \geq d_{diff},\ \forall l \neq m. \quad (8)$$

4. Non-trivial Attacks. *If the FC perceives some sensor produces a constant value of 0 or 1, then the FC can easily recognize the sensor is under attack. For this reason, in order to reduce the probability of being detected, we assume that the adversaries ensure*

$$\tilde{p}\left(\mathbf{\Psi}_p, \theta\right) \neq 0 \text{ or } 1, \quad \forall p \geq 1. \quad (9)$$

It is worth mentioning that the adversary model assumed in (4) can change the after-attack pmf to have any desired valid values satisfying (6), (8), and (9) through proper choice of the two attack parameters $\psi_{p,0}$ and $\psi_{p,1}$. In this sense, it is a fairly general adversary model.

### 2.1. Identification and Categorization of Attacked Sensors

The following theorem gives our main results on the identification and categorization of the attacked sensors.

**Theorem 1** *Under* Assumption 1, *for any $N$ as $K \to \infty$, the FC can always identify from the observations, without further knowledge, a $\mathcal{P}_0$ percentage group of sensors which contains zero percent attacked sensors. Similarly as $K \to \infty$, the FC is also able to identify $P$ other groups of sensors, which respectively make up $\{\mathcal{P}_p\}_{p=1}^{P}$ percent of all sensors, such that for $p = 1, 2, ..., P$, group $p$ contains zero percent sensors not experiencing attack $p$.*

*On the other hand, assume each sensor observes a finite number $K$ of time samples such that*

$$K \geq -\frac{8 \ln 2}{\gamma^* \min\{\Delta\Delta_0, \Delta^2\}} + 1, \tag{10}$$

*where $\gamma^*$ is a constant defined in [6]. Under* Assumption 1, *as $N \to \infty$, the FC can determine $P$. Moreover, for the set of sensors $\mathcal{A}_p$ which are under the $p$-th attack $\forall p \geq 0$, the FC can identify a corresponding group of sensors $\tilde{\mathcal{A}}_p$ with $\tilde{\mathcal{P}}_p \triangleq |\tilde{\mathcal{A}}_p|/N$, $\mathcal{P}_p^* \triangleq |(\tilde{\mathcal{A}}_p \backslash \mathcal{A}_p) \cup (\mathcal{A}_p \backslash \tilde{\mathcal{A}}_p)|/N$, and*

$$\delta \triangleq -\frac{4 \ln 2}{\Delta(K-1)\gamma^*} \tag{11}$$

*which satisfy*

$$0 \leq |\tilde{\mathcal{P}}_p - \mathcal{P}_p| \leq \mathcal{P}_p^* < \delta. \tag{12}$$

In [6], it is shown that the quantization approach previously discussed leads to a singular Fisher Information Matrix (FIM) such that $\theta$ can not be estimated with increasing accuracy with larger $NK$. However, by splitting the observations into two or more groups at each sensor and employing different thresholds for each group, the FIM is always nonsingular. In the following theorem, we provide necessary and sufficient conditions under which the CRB performance of estimating $\theta$ can be improved by employing observations from an attacked sensor.

**Theorem 2** *The CRB for $\theta$ can be improved by utilizing the observations from the set of attacked sensors in our proposed fashion, if and only if the FIM for estimating $\theta$ based on the observations under the $p$-th attack has rank 3 for some $p$. Otherwise, there is no CRB improvement, but also no loss in performance, from utilizing the attacked observations.*

In [7] we provide a closed-form expression describing the increase in CRB obtained from using the data from attacked sensors. All the analysis presented can be extended to nonbinary and general estimation problems [7].

## 3. SPOOFING ATTACKS

Let $\mathcal{V} \subset \mathcal{S}_N$ denote the set of sensors undergoing spoofing attacks. The after-attack version $\tilde{x}_{jk}$ of $x_{jk}$ is an independent sequence over $j, k$ such that[1]

$$\tilde{x}_{jk} \sim \begin{cases} f_{jk}\left(\tilde{x}_{jk} \left| \boldsymbol{\theta} \right.\right), & \text{if } j \in \mathcal{U} \\ g_{jk}\left(\tilde{x}_{jk} \left| \boldsymbol{\theta}, \boldsymbol{\tau}^{(p)} \right.\right), & \text{if } j \in \mathcal{V} \text{ and } j \in \mathcal{A}_p \end{cases}, \tag{13}$$

where if $j \in \mathcal{V}$ and $j \in \mathcal{A}_p$, then the after-attack pdf $g_{jk}(x_{jk}|\boldsymbol{\theta}, \boldsymbol{\tau}^{(p)})$ is parameterized by the desired vector parameter $\boldsymbol{\theta}$ and the attack vector parameter $\boldsymbol{\tau}^{(p)}$. To conform to previous work, the functional forms of the attacks, thus $\{f_{jk}\}$ and $\{g_{jk}\}$, are assumed known to the attacked system but the desired and attack vector parameters are not.

We generalize the quantization model to allow nonbinary quantization. At the $j$-th sensor, each after-attack measurement $\tilde{x}_{jk}$ is quantized to $\tilde{u}_{jk}$ by using a $R_j$-level quantizer with quantization regions $\{I_j^{(r)}\}_{r=1}^{R_j}$, that is,

$$\tilde{u}_{jk} = \sum_{r=1}^{R_j} r \mathbb{1}\left\{\tilde{x}_{jk} \in I_j^{(r)}\right\}, \tag{14}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Let $\boldsymbol{\Theta}$ denote a vector containing the unknown vector parameter $\boldsymbol{\theta}$ along with all the unknown attack vector parameters which parameterize the spoofing attacks in the sensor network

$$\boldsymbol{\Theta} \triangleq \left[\boldsymbol{\theta}^T, \left(\boldsymbol{\tau}^{(1)}\right)^T, ..., \left(\boldsymbol{\tau}^{(P)}\right)^T\right]^T. \tag{15}$$

Now we define a highly desirable attack.

**Definition 1** *Consider attacks employing $\{f_{jk}(x_{jk}|\boldsymbol{\theta})\}$ and $\{g_{jk}(\tilde{x}_{jk}|\boldsymbol{\theta}, \boldsymbol{\tau}^{(p)})\}$. The optimal guaranteed degradation spoofing attack (OGDSA) maximizes the degradation of the Cramer-Rao Bound (CRB) for the vector parameter of interest at the FC when the attacked sensors are well identified and categorized according to distinct types of spoofing attacks by the FC. The CRB for the case where the attacked sensors are well identified and categorized provides a lower bound on the CRB for any case, including cases with unidentified and uncategorized attacked sensors, thus providing guaranteed sufficiently undesirable performance and justifying the name.*

One class of attacks that are OGDSA are called inestimable spoofing attacks, defined next and further illuminated by the subsequent theorem.

**Definition 2 (Inestimable spoofing attack (ISA))** *The $p$-th spoofing attack is referred to as an ISA if the corresponding FIM for estimating $\boldsymbol{\tau}^{(p)}$ is singular.*

[1] The notations $\tilde{x}_{jk}$ and $\tilde{u}_{jk}$ denote the after-attack analog measurements and the corresponding quantized measurements.

Such an attack can result from a sufficiently powerful attack relative to the number of quantification symbols employed by the quantizers as stated in the next theorem.

**Theorem 3** *For the $p$-th spoofing attack, if the dimension $D_p$ of the attack parameter $\boldsymbol{\tau}^{(p)}$ satisfies*

$$D_p > \sum_{j \in \mathcal{A}_p} K\,(R_j - 1), \tag{16}$$

*then the FIM for estimating $\boldsymbol{\tau}^{(p)}$ is singular, and furthermore, the FIM for estimating $\boldsymbol{\Theta}$ is also singular.*

The only other possible class of OGDSAs are called optimal estimable spoofing attacks (OESAs). The estimable spoofing attacks (ESAs) are defined next.

**Definition 3 (Estimable spoofing attack)** *The $p$-th spoofing attack is said to be estimable if the corresponding FIM for estimating $\boldsymbol{\tau}^{(p)}$ is nonsingular.*

**Theorem 4** *In the presence of ESAs, the CRB for $\boldsymbol{\theta}$ is bounded above as per*

$$CRB_{ESA}\,(\boldsymbol{\theta}) \triangleq \left[\mathbf{J}_{\boldsymbol{\Theta}}^{-1}\right]_{1:D_{\boldsymbol{\theta}}} \preceq \mathbf{J}_{\mathcal{A}_0}^{-1}. \tag{17}$$

*where $\mathbf{J}$ denotes the relevant FIM. If the $p$-th spoofing attack is an ESA and achieves the equality in (17), then the $p$-th spoofing attack is an OESA, and hence an OGSA.*

In [15] we give necessary and sufficient conditions for an optimal estimable spoofing attack in terms of a relationship between the subspaces spanned by the columns of certain matrices from the singular value decompositions of the FIMs for estimating $\boldsymbol{\theta}$ and $\boldsymbol{\tau}^{(p)}$ using data under the $p$-th attack. It is also shown [15] that a generalization of an additive shift in $\boldsymbol{\theta}$, thus the attack replaces $\boldsymbol{\theta}$ by $\boldsymbol{\theta} + \boldsymbol{\tau}^{(p)}$ is always an OESA.

## 4. IMPLICATIONS FOR UNATTACKED SYSTEMS

From Theorem 3, we can derive the following theorem describing a fundamental limitation on quantized estimation systems not under attack.

**Theorem 5** *Let $D_{\boldsymbol{\theta}}$ be the dimension of a vector parameter we want to estimate from $L$ independent observations quantized using $Q$ distinct quantizer designs with $R_j, j = 1, 2 \ldots, Q$ symbols. Assume the $j$-th group of observations, all facing an identical quantizer, are generated from $M_j$ different pdfs. The FIM is singular (accuracy of estimations does not increase with more observations) if*

$$D_{\boldsymbol{\theta}} > \sum_{j=1}^{Q} M_j\,(R_j - 1), \tag{18}$$

Generalized results can be found in [16].

## 5. CONCLUSIONS

Spoofing and man-in-the-middle attacks are studied for systems performing estimation of an unknown deterministic vector from quantized sensor data. For man-in-the-middle attacks, asymptotically optimum processing which identifies and categorizes the attacked sensors into different groups is described and necessary and sufficient conditions are provided under which utilizing the attacked sensor data will lead to better estimation performance. For spoofing attacks, necessary and sufficient conditions are provided under which an attack performance in terms of the Cramer-Rao Bound (CRB) is guaranteed regardless of the processing the estimation system employs. Theses conditions imply that, for any such attack when the attacked sensors can be perfectly identified by the estimation system, either the Fisher Information Matrix (FIM) for jointly estimating the desired and attack parameters is singular or the attacked system is unable to improve the CRB for the desired vector parameter through this joint estimation even though the joint FIM is nonsingular. If the attacker knows the number of quantization symbols and the number of different statistical models at each sensor, it is always possible to construct such a highly desirable attack by properly employing an attack vector parameter having a sufficiently large dimension relative to the number of quantization levels employed. For unattacked quantized estimation systems, a general limitation on the dimension of a vector parameter which can be accurately estimated is uncovered.

## 6. REFERENCES


[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, April 2005, pp. 91–98.

[3] J. H. Lee and R. Buehrer, "Characterization and detection of location spoofing attacks," *Communications and Networks, Journal of*, vol. 14, no. 4, pp. 396–409, Aug 2012.

[4] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 106–115, 2012.

[5] 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.

[6] J. Zhang, R. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *Signal Processing, IEEE Transactions on*, vol. 63, no. 5, pp. 1086–1101, March 2015.

[7] B. Alnajjab, J. Zhang, and R. S. Blum, "Attacks on sensor network parameter estimation with quantization: Performance and asymptotically optimum processing," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6659–6672, 2015.

[8] J. Zhang and R. S. Blum, "Distributed joint spoofing attack identification and estimation in sensor networks," in *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*. IEEE, 2015, pp. 701–705.

[9] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2015.

[10] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *Information Theory, IEEE Transactions on*, vol. 47, no. 3, pp. 978–1002, 2001.

[11] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *Signal Processing Magazine, IEEE*, vol. 23, no. 4, pp. 27–41, 2006.

[12] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1131–1143, 2006.

[13] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4519–4528, 2006.

[14] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5682–5699, 2009.

[15] J. Zhang, R. S. Blum, L. M. Kaplan, and X. Lu, "Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 705–720, Feb 2017.

[16] J. Zhang, R. S. Blum, L. Kaplan, and X. Lu, "A fundamental limitation on maximum parameter dimension for accurate estimation with quantized data," *arXiv preprint arXiv:1605.07679*, 2016.