On the Estimation and Secrecy Capabilities of Stochastic Encryption for Parameter Estimation in IoT

Ananth Narayan Samudrala* and Rick S. Blum†
Department of Electrical and Computer Engineering
Lehigh University
Bethlehem, PA 18015
Email: ans416@lehigh.edu*, rblum@lehigh.edu†

***Abstract*—In Internet of Things (IoT) applications requiring parameter estimation, sensors often transmit quantized observations to a fusion center through a wireless medium where the observations are susceptible to unauthorized eavesdropping. The fusion center uses the received data to estimate desired parameters. To provide security to such networks, some low complexity encryption approaches have been proposed. In this paper, we generalize those approaches and present an analysis of their estimation and secrecy capabilities. We show that the dimension of the unknown parameter that can be efficiently estimated using an unbiased estimator when using these approaches, is upper bounded. Assuming that an unauthorized eavesdropper is aware of the low complexity encryption process but is unaware of the encryption key, we show successful eavesdropping, even with a large number of observations, is impossible with unbiased estimators and independent observations for these approaches. Numerical results validating our analysis are presented.**

I. Introduction

Advancements in technology have led to the development of the popular concept called the Internet of things (IoT). IoT consists of networks of sensors that are used to extract information useful to control and command machines and devices. Often this requires estimation of parameters from the sensor data. In many of these applications, the sensors monitor and communicate critical information to a central location, called a legitimate fusion center (LFC). An LFC uses the received sensor data to estimate a desired unknown parameter $\boldsymbol{\theta}$. A drawback of such parameter estimation networks is that they are vulnerable to passive eavesdropping by an intruder, called a third party fusion center (TPFC). Protection against eavesdropping is thus an important requirement. However, the limited processing power, energy and memory size of the sensors, along with the small required maximum latencies for the applications, make it difficult to employ traditional encryption schemes. For this purpose, low-complexity encryption schemes were proposed by several authors and much of this work is described in the survey paper [1].

As is the case of all digital communications, sensors quantize their observations. In parameter estimation networks, the quantized observations are encrypted and transmitted to an LFC. For such networks a low complexity encryption approach was proposed in [2] called binary stochastic encryption (BSE) for the case where each observation is quantized to a single bit (binary quantization). In BSE, all employed sensors employ identical binary quantization and encryption. In BSE, quantized binary symbols are flipped to a 0 or a 1 with certain probabilities, which serve as an encryption key. An LFC with knowledge of the encryption key can estimate $\boldsymbol{\theta}$ from received encrypted data. The Maximum likelihood (ML) estimator and the Cramér Rao lower bound (CRLB) for the LFC estimation were derived in [2].

Recent work [3][4] has shown that the use of identical binary quantization at all sensors limits parameter estimation to scalar parameters for independent and identically distributed (i.i.d.) observations, i.e., $\boldsymbol{\theta}$ has to be a scalar parameter for BSE. However, in practice we might need to estimate vector parameters like the position of an object. With the objective of making vector parameter estimation possible, we extended BSE to a non-binary stochastic encryption (NBSE) [5] since the use of higher order quantization enables vector parameter estimation. In NBSE, all sensors employ identical non-binary quantizers and the quantized symbols are flipped to different non-binary symbols with certain probabilities (encryption key). An LFC with knowledge of the encryption key would be able to estimate $\boldsymbol{\theta}$ from the encrypted data. For NBSE, we described the LFC ML estimator along with the CRLB.

An analysis of any encryption scheme must include a study of it's secrecy capabilities against passive eavesdropping. Both [2] and [5] present an analysis of the protection provided by stochastic encryption under the same two assumptions. Firstly, it is assumed that an eavesdropping TPFC has complete knowledge of the quantizer at each sensor, i.e., the TPFC knows the quantizer regions. Secondly, it is assumed that the TPFC is unaware that encryption is being employed at each sensor and hence considers the intercepted data to be unencrypted. Under these assumptions, by deriving the asymptotic bias and variance of the TPFC ML estimator, both the papers show that the TPFC ML estimate with a proper stochastic encryption would be biased and the estimation mean square error is large.

In this paper we present an analysis of the estimation and secrecy capabilities of stochastic encryption. We make three contributions. Firstly, we generalize BSE and NBSE to generalized stochastic encryption (GSE). GSE is a stochastic encryption scheme in which different sensors can employ different quantization and stochastic encryption. Varying the quantizer design across sensors enables us to perform efficient

This work was partially supported by the National Science Foundation under grants CNS 1702555 and ECCS 1744129.

978-1-5386-0579-0/18/$31.00 ©2018 IEEE

vector parameter estimation with lower order quantizers, for example binary quantizers. Additionally, varying the encryption key across sensors enhances secrecy. We present the LFC ML estimator and the CRLB for GSE.

Second, under the assumption of unbiased estimators we present an analysis of the estimation capabilities of the three stochastic encryption schemes: BSE, NBSE and GSE at an LFC. Applying results from [3][4], we show that the maximum dimension of $\boldsymbol{\theta}$ that can be estimated, is upper bounded.

Finally, we analyze the secrecy provided by each stochastic encryption scheme. In presenting this analysis, we make three assumptions. First, we assume that the TPFC is aware of the stochastic encryption approach that was used to encrypt the data. For example, if the sensors employ BSE then we assume that the TPFC knows that data is encrypted using BSE. Second we assume that the TPFC uses an unbiased estimator. Third, we assume that the TPFC does not the know the stochastic encryption encryption key and the quantizer regions. Under these assumptions for a TPFC to estimate $\boldsymbol{\theta}$ from the encrypted data, it has to perform joint estimation of all the unknown parameters, which are the encryption keys, quantizer regions and $\boldsymbol{\theta}$. Generally, one would think that given enough observations a TPFC would be able to determine all the unknown parameters including $\boldsymbol{\theta}$. But using results from [3][4] we show that this is not true for any of the three stochastic encryption schemes since the Fisher information matrix (FIM) of the TPFC estimation is singular. Thus, there exists no unbiased estimator that a TPFC can use to estimate all the unknown parameters. Hence, stochastic encryption is secure in the domain of unbiased estimators.

In this paper, bold upper case letters and bold lower case letters are used to represent matrices and column vectors respectively. The symbol $\mathbf{0}$ stands for the all-zero column vector. The remainder of the paper is organized as follows. In section II, we present the system model of a parameter estimation system that uses GSE and derive the system models for BSE and NBSE as special cases. Section III presents the LFC ML estimators and the CRLBs for all three encryption approaches. Section IV presents results from [3][4] that we shall use in our analysis of estimation and secrecy capabilities of stochastic encryption. Estimation capabilities of each stochastic encryption approach at an LFC are discussed in section V. Section VI analyzes the secrecy provided by each stochastic encryption approach against an eavesdropping TPFC. Section VII provides numerical results and finally conclusions are drawn in section VIII.

II. System Model

A. GSE System Model

Consider a parameter estimation network comprised of M groups of sensors as shown in Fig. 1. The m^{th} group includes N_m sensors for $m = 1, 2, \ldots, M$. The total number of sensors is N with $\sum_{m=1}^{M} N_m = N$. The groups of sensors differ from each other in the quantization and encryption they use, but all the sensors in a group employ an identical quantizer and stochastic encryption algorithm. The m^{th} group employs

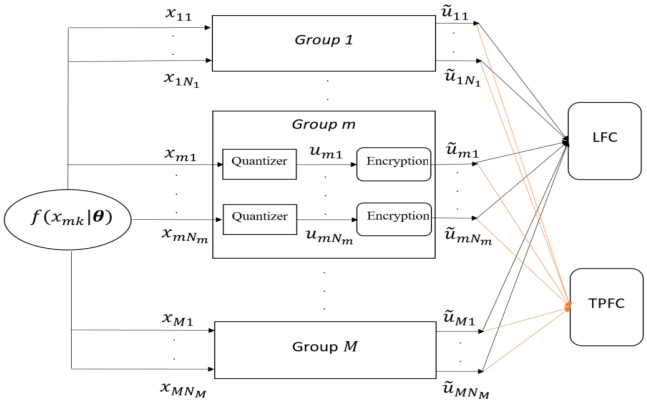

Figure 1. GSE system model

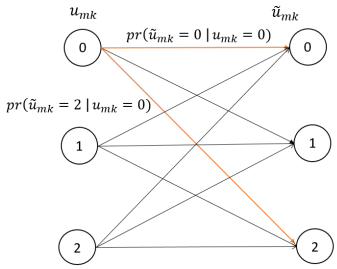

Figure 2. Stochastic Encryption for $R_m = 3$

R_m-level quantization and stochastic encryption with $\mathbf{c}_m$ as the stochastic encryption key.

The scalar observation of the k^{th} sensor in the m^{th} group is x_{mk}[1]. The probability density function (pdf) of x_{mk} depends on an underlying vector parameter $\boldsymbol{\theta}$ and is given by $f(x_{mk}|\boldsymbol{\theta})$ (all x_{mk} are i.i.d.). The dimension of $\boldsymbol{\theta}$ is $D_{\boldsymbol{\theta}}$. The observation x_{mk} is quantized to u_{mk}, where $u_{mk} \in \{0, 1, \ldots, (R_m - 1)\}$. For the m^{th} group consider the set of non-overlapping regions $\{A_{m0}, A_{m1}, ..., A_{m(R_m-1)}\}$ such that the R_m-level quantizer will assign the symbol $u_{mk} = i$ to any input $x_{mk} \in A_{mi}$ for $i \in \{0, 1, \ldots, (R_m - 1)\}$. Then, the probability that the quantizer output $u_{mk} = i$ is

$$pr(u_{mk} = i|\boldsymbol{\theta}) \quad = \quad \int_{x_{mk} \in A_{mi}} f(x_{mk}|\boldsymbol{\theta}) dx_{mk}. \qquad (1)$$

The quantizer output u_{mk} is encrypted to $\tilde{u}_{mk}$ with $\mathbf{S}_m$ as the stochastic encryption matrix. $\mathbf{S}_m$ is a $R_m \times R_m$ matrix with the $(i+1, j+1)^{th}$ entry as $pr(\tilde{u}_{mk} = j|u_{mk} = i)$ for $i, j \in \{0, 1, \ldots, (R_m - 1)\}$. The sum of the elements in each column of $\mathbf{S}_m$ is 1 and hence $\mathbf{S}_m$ has $R_m(R_m - 1)$ independent elements. These independent elements form the stochastic encryption key $\mathbf{c}_m$. For us, $\mathbf{c}_m$ consists of elements of first $R_m - 1$ rows of $\mathbf{S}_m$ in sequential order. Fig. 2 illustrates

[1]We consider a system model with scalar x_{mk} but all our results can be extended to include vector $\mathbf{x}_{mk}$.

stochastic encryption for $R_m = 3$. The probability that the encryption output $\tilde{u}_{mk} = j$ is

$$pr\left(\tilde{u}_{mk} = j|\boldsymbol{\theta}\right) = \sum_{i=0}^{R_m-1} \Big(pr(\tilde{u}_{mk} = j|u_{mk} = i) \times pr(u_{mk} = i|\boldsymbol{\theta}) \Big). \quad (2)$$

The encrypted symbols $\tilde{u}_{mk}$ are then transmitted to the LFC, and at the same time intercepted by an unauthorized TPFC. The objective at both the LFC and TPFC, is to estimate $\boldsymbol{\theta}$ from the received $\tilde{u}_{mk}$.

B. BSE System Model

BSE was proposed in [2] for scalar parameter estimation. According to BSE, all the sensors employ an identical binary quantizer and stocastic encryption key. Hence for BSE we have, $M = 1$, $N_1 = N$ and $R_1 = 2$. Since $M = 1$ for BSE, we discard the subscript m in all further discussions of BSE. As BSE was proposed for scalar parameter estimation using scalar observations we shall use θ and x_k in all further BSE discussions.

The GSE model discussed above is for a general estimation problem in which the pdf of the observations is dependent on the unknown parameter. However for BSE, the authors considered a restricted estimation problem in which $x_k = \theta + n_k, k = 1, 2, \ldots, N$ where $\{n_k, k = 1, 2, \ldots, N\}$ is an i.i.d zero-mean sequence. Hence, all further discussions of BSE will be for this restricted model.

Since we discarded the subscript m for BSE, for the binary quantizer regions we will replace A_{mi} with A_i for $i \in \{0, 1\}$. Hence, for BSE the quantization regions are $\{A_0, A_1\}$. In [2] the authors considered a binary quantizer design with connected quantization regions and hence we have $A_1 = \{x \geq \tau\}$ where τ is the binary quantizer threshold. The BSE matrix $\mathbf{S}_{BSE}$ is given as

$$\mathbf{S}_{BSE} = \begin{bmatrix} \Omega_0 & 1 - \Omega_1 \\ 1 - \Omega_0 & \Omega_1 \end{bmatrix}.$$

where $\Omega_1 = pr(\tilde{u}_k = 1|u_k = 1)$ and $\Omega_0 = pr(\tilde{u}_k = 0|u_k = 0)$. The encryption key is $\mathbf{c}_{BSE} = [\Omega_0, 1 - \Omega_1]^T$.

C. NBSE System Model

In [5] we proposed NBSE considering the general estimation problem in which the pdf of the observations is dependent on the parameter $\boldsymbol{\theta}$. According to NBSE, all the sensors employ an identical non-binary quantizer and encryption key. Therefore for NBSE we have, $M = 1$, $N_1 = N$ and $R_1 = R$ where $R \in \{3, 4, \ldots\}$. Since $M = 1$ for NBSE, we discard the subscript m in all further discussions of NBSE. Hence, similar to BSE the quantizer regions for NBSE are $\{A_0, A_1, \ldots, A_{(R-1)}\}$. The NBSE matrix is $\mathbf{S}_{NBSE}$ and the encryption key is $\mathbf{c}_{NBSE}$.

III. LFC ML Estimation

In this section, we present the LFC ML estimators and the CRLBs for each of the three stochastic encryption schemes. An LFC receiving the encrypted data can estimate $\boldsymbol{\theta}$ since it knows the encryption key and quantizer regions. In deriving the ML estimators and CRLBs, we make the following assumptions.

Assumption 1: We assume that $f(x_{mk}|\boldsymbol{\theta})$ obeys regularity conditions [6] such that interchanges involving derivatives with respect to $\boldsymbol{\theta}$ and integrals with respect to x_{mk} are valid.

Assumption 2: The LFC uses an unbiased estimator and the observations at different sensors x_{mk} are i.i.d.

Assumption 3: The stochastic encryption keys $\mathbf{c}_m \forall m$ for GSE are chosen such that $pr\left(\tilde{u}_{mk} = i|\boldsymbol{\theta}\right) \neq pr(u_{mk} = i|\boldsymbol{\theta}) \forall m \in \{1, 2, \ldots, M\}, i \in \{0, 1, \ldots, R_m - 1\}, k \in \{1, 2, \ldots, N_m\}$. Equivalent conditions for BSE and NBSE can be derived by removing the subscript m. This assumption implies that the probability distribution of quantized data before encryption and after encryption are different for effective encryption.

A. BSE ML Estimation

The BSE LFC ML estimate derived in [2] using the invariance property of the ML estimate [7] is given as

$$\hat{\theta}_{BSE} = \tau - F^{-1}\left(\frac{\frac{\sum_{k=1}^{N} \tilde{u}_k}{N} - \Omega_1}{1 - \Omega_1 - \Omega_0}\right). \quad (3)$$

where $F(.)$ is the cumulative distribution function of n_k. The ML estimate achieves the CRLB asymptotically. The CRLB for BSE is

$$\Psi_{BSE}\left(\theta\right) = \frac{1}{N(1 - \Omega_1 - \Omega_0)^2} \frac{\left[1 - p(\tilde{u}_k = 1)\right] p(\tilde{u}_k = 1)}{f^2(\tau - \theta)}. \quad (4)$$

where $f(.)$ is the pdf of n_k.

B. NBSE ML Estimation

Differentiating the log-likelihood function for LFC observations with respect to $\boldsymbol{\theta}$, we obtained a necessary condition for the NBSE LFC ML estimate in [5] as

$$\frac{1}{N}\sum_{k=1}^{N}\sum_{j=0}^{R-1}\Bigg(I(\tilde{u}_k = j) \times \frac{\sum_{i=0}^{R-1} pr(\tilde{u}_k = j|u_k = i) \int_{x_k \in A_i} \frac{d}{d\boldsymbol{\theta}} f(x_k|\boldsymbol{\theta}) dx_k}{\sum_{i=0}^{R-1} pr(\tilde{u}_k = j|u_k = i) \int_{x_k \in A_i} f(x_k|\boldsymbol{\theta}) dx_k} \Bigg) = \mathbf{0}. \quad (5)$$

where $I(.)$ is the indicator function. Note that (5) is a vector equation for $\boldsymbol{\theta}$. Using (5) with an appropriately initialized iterative algorithm such as the Newton-Raphson method we can obtain the ML estimate. The ML estimate achieves the CRLB asymptotically. The CRLB matrix $\boldsymbol{\Psi}_{NBSE}\left(\boldsymbol{\theta}\right)$ for NBSE is

$$\boldsymbol{\Psi}_{NBSE}\left(\boldsymbol{\theta}\right) = \left(N \sum_{i=0}^{R-1} \frac{\frac{d}{d\boldsymbol{\theta}} pr(\tilde{u}_1 = i|\boldsymbol{\theta}) (\frac{d}{d\boldsymbol{\theta}} pr(\tilde{u}_1 = i|\boldsymbol{\theta}))^T}{pr(\tilde{u}_1 = i|\boldsymbol{\theta})}|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}\right)^{-1} \quad (6)$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Proofs for (5) and (6) are available in [5].

C. GSE ML Estimation

Differentiating the log-likelihood function for LFC observations with respect to $\boldsymbol{\theta}$, we obtain a necessary condition for the GSE LFC ML estimate as

$$\frac{1}{N}\sum_{m=1}^{M}\sum_{k=1}^{N_m}\sum_{j=0}^{R_m-1}\left(I(\tilde{u}_{mk}=j)\times \frac{\sum_{i=0}^{R_m-1} pr(\tilde{u}_{mk}=j|u_{mk}=i)\int_{x_{mk}\in A_{mi}}\frac{d}{d\boldsymbol{\theta}}f(x_{mk}|\boldsymbol{\theta})dx_{mk}}{\sum_{i=0}^{R_m-1} pr(\tilde{u}_{mk}=j|u_{mk}=i)\int_{x_{mk}\in A_{mi}} f(x_{mk}|\boldsymbol{\theta})dx_{mk}}\right)=\mathbf{0} \quad (7)$$

Using the vector equation (7) with an appropriately initialized iterative algorithm we can obtain the ML estimate. The CRLB matrix $\boldsymbol{\Psi}_{GSE}(\boldsymbol{\theta})$ for GSE is

$$\boldsymbol{\Psi}_{GSE}(\boldsymbol{\theta})=\left(\sum_{m=1}^{M}N_m\sum_{i=0}^{R_m-1}\frac{\frac{d}{d\boldsymbol{\theta}}pr(\tilde{u}_{m1}=i|\boldsymbol{\theta})(\frac{d}{d\boldsymbol{\theta}}pr(\tilde{u}_{m1}=i|\boldsymbol{\theta}))^T}{pr(\tilde{u}_{m1}=i|\boldsymbol{\theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^{-1} \quad (8)$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Proofs for (7) and (8) follow the same steps as the proofs for (5) and (6).

IV. Limitation on Maximum Parameter Dimension for Estimation from Quantized data

The authors of [3], [4] state the following theorem that presents a limitation on the maximum dimension of a vector parameter $\boldsymbol{\theta}$ that can be estimated efficiently from quantized data.

Theorem 1: Let $D_{\boldsymbol{\theta}}$ be the dimension of a vector parameter $\boldsymbol{\theta}$ 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 H_j different pdfs. The FIM is singular if

$$D_{\boldsymbol{\theta}} > \sum_{j=1}^{Q} H_j(R_j - 1). \quad (9)$$

Applying Theorem 1, and Assumptions 1 and 2 to the GSE encryption model discussed in section II, we have $L = N$ independent observations from the N sensors, $Q = M$ since we have M groups each using a distinct quantizer and encryption, $R_j = R_m$ for the m^{th} group, and $H_j = 1$ since all the sensors in a group use an identical quantizer and encryption. Then (9) becomes

$$D_{\boldsymbol{\theta}} > \sum_{m=1}^{M}(R_m - 1) = D_{GSE}. \quad (10)$$

V. Stochastic Encryption Estimation Capability

In this section, we discuss the estimation capabilities of stochastic encryption at an LFC. Specifically, using (10) we present an upper bound on the maximum dimension of $\boldsymbol{\theta}$ that can be estimated using each of the three stochastic encryption schemes.

A. BSE Estimation Capability

For BSE we have, $M = 1$, $N_1 = N$ and $R_1 = 2$. Applying (10) to BSE we obtain $D_{\boldsymbol{\theta}} > 1 = D_{BSE}$ which implies that the maximum dimension of $\boldsymbol{\theta}$ that we can estimate using BSE is 1, i.e., scalar parameters under our assumptions.

B. NBSE Estimation Capability

For NBSE we have, $M = 1$, $N_1 = N$ and $R_1 = R$. Applying (10) to NBSE we obtain $D_{\boldsymbol{\theta}} > R - 1 = D_{NBSE}$. Hence, the maximum dimension of $\boldsymbol{\theta}$ that can be estimated with R-level NBSE is $R - 1$. For example, for NBSE with $R = 3$ we can estimate either scalar parameters or 2-dimensional parameters under our assumptions.

C. GSE Estimation Capability

For GSE, from (10) we can see that with GSE it is possible to estimate larger vector parameters with lower order quantizers. For example, to estimate a 2-dimensional $\boldsymbol{\theta}$ we can use $R = 3$ NBSE. Alternatively consider a GSE with $M = 2$ groups of sensors, each with a distinct $R_m = 2$ binary quantizer. Substituting $M = 2$ and $R_m = 2$ in (10) we obtain $D_{GSE} = 2$. Hence, using GSE with $M = 2$ we can estimate 2-dimensional $\boldsymbol{\theta}$ with binary quantizers.

In [8] it is shown that by using distinct binary quantizers for each of the M groups, it is possible to attain a lower mean estimation variance compared to the case when all the sensors use the same binary quantizer. Similarly, if $\boldsymbol{\theta}$ has a large dynamic range then by appropriately choosing distinct R_m level quantizers and encryption keys $\mathbf{c}_m$ for each of the M groups, it is possible for us to achieve a lower mean estimation variance for the entire range of $\boldsymbol{\theta}$ with GSE when compared to BSE and NBSE.

VI. Stochastic Encryption Secrecy Capability

In this section, we discuss the secrecy capabilities of each stochastic encryption scheme against an eavesdropping TPFC. In addition to previous assumptions we make the following three additional assumptions.

Assumption 4: Any eavesdropping TPFC uses an unbiased estimator and is aware of the stochastic encryption approach that was used to encrypt the data, i.e., BSE or NBSE or GSE.

Assumption 5: An eavesdropping TPFC is not aware of the actual stochastic encryption key and quantizer regions that were used to generate the encrypted data.

Under these assumptions, for a TPFC to estimate $\boldsymbol{\theta}$ from the encrypted data it must perform joint estimation of all the unknown parameters, which are the stochastic encryption keys, quantizer regions and $\boldsymbol{\theta}$. We shall first discuss the secrecy capabilities of GSE and then derive the secrecy capabilities of BSE and NBSE as special cases of GSE.

A. GSE Secrecy Capability

Consider a GSE system as described in section II. In illustrating the secrecy capabilities of GSE, we shall consider each R_m-level quantizer to have the following simple design with connected quantization regions. We quantize x_{mk} to u_{mk} directly by using the quantizer thresholds $\mathbf{t}_m = \left[t_{m0}, t_{m1}, \ldots, t_{m(R_m-2)}\right]^T$ with each $t_{mi} \in \mathbb{R}$. We have

$$u_{mk} = \begin{cases} 0 & \text{if } x_{mk} < t_{m0}, \\ R_m - 1 & \text{if } x_{mk} \geq t_{m(R_m-2)}, \\ i & \text{if } x_{mk} \in \left[t_{m(i-1)}, t_{mi}\right) \\ & \forall i \in \{1, 2, \ldots, (R_m - 2)\}. \end{cases} \quad (11)$$

Let us define $\boldsymbol{\phi}_m$ and $\boldsymbol{\phi}$ as

$$\boldsymbol{\phi}_m = [\mathbf{t}_m^T \ \mathbf{c}_m^T]^T \quad (12)$$

$$\boldsymbol{\phi} = [\boldsymbol{\theta}^T \ \boldsymbol{\phi}_1^T \ \boldsymbol{\phi}_2^T \ \ldots \boldsymbol{\phi}_M^T]^T \quad (13)$$

Each $\mathbf{t}_m$ has $R_m - 1$ independent elements. Each $\mathbf{c}_m$ has $R_m(R_m - 1)$ independent elements. Hence, each $\boldsymbol{\phi}_m$ has $R_m^2 - 1$ unknowns. Since we have M groups of sensors the dimension of $\boldsymbol{\phi}$ is $D_{\boldsymbol{\phi}} = D_{\boldsymbol{\theta}} + \sum_{m=1}^{M} R_m^2 - 1$. To determine $\boldsymbol{\theta}$ from the encrypted data, a TPFC has to estimate $\boldsymbol{\phi}$. From (10) we have $D_{GSE} = \sum_{m=1}^{M}(R_m - 1)$. Since $R_m \geq 2 \forall m \in \{1, 2, \ldots, M\}$ we will always have $D_{\boldsymbol{\phi}} > D_{GSE}$. Then, from Theorem 1 we can conclude that the FIM of TPFC estimation is always singular. This implies that there is no unbiased estimator of $\boldsymbol{\phi}$ that a TPFC can use to estimate $\boldsymbol{\phi}$ and therefore TPFC cannot determine $\boldsymbol{\phi}$ even if a large number of observations are available. Thus, GSE is secure in the domain of unbiased estimators.

Though we considered a simple quantizer design with connected quantization regions in our analysis, our conclusions hold true for any quantizer design with connected or disconnected quantization regions. We prove this by showing that the FIM remains singular even if the TPFC has knowledge of the quantizer designs. If the TPFC has knowledge of the quantizer regions, then we have $\boldsymbol{\phi}_m = \mathbf{c}_m$ which gives $D_{\boldsymbol{\phi}} = D_{\boldsymbol{\theta}} + \sum_{m=1}^{M} R_m(R_m - 1)$. Since $R_m \geq 2$ we will still have $D_{\boldsymbol{\phi}} > D_{GSE}$ resulting in a singular FIM.

B. BSE Secrecy Capability

For BSE we have, $M = 1$, $N_1 = N$, $R_1 = 2$ and $\mathbf{t}_1 = \tau$. Hence, we have $D_{\boldsymbol{\phi}} = D_{\theta} + 3$. Since $D_{BSE} = 1$ we have $D_{\boldsymbol{\phi}} > D_{BSE}$ always true. Thus, BSE is secure in the domain of unbiased estimators.

C. NBSE Secrecy Capability

For NBSE we have, $M = 1$, $N_1 = N$, $R_1 = R$ and $\mathbf{t}_{NBSE} = \mathbf{t}_1$. Hence, we have $D_{\boldsymbol{\phi}} = D_{\boldsymbol{\theta}} + R^2 - 1$. Since $R \geq 3$ and $D_{NBSE} = R - 1$ we have $D_{\boldsymbol{\phi}} > D_{NBSE}$ always true. Thus, NBSE is secure in the domain of unbiased estimators.

VII. Numerical Results

In this section we present numerical results supporting our analysis of the estimation and secrecy capabilities of stochastic encryption.

A. Comparison of BSE, NBSE and GSE

We first compare the three stochastic encryption schemes with respect to the LFC CRLB in Fig. 3 for $N = 1000$ and $x_{mk} = \theta + n_{mk}, m = 1, 2, \ldots, M, k = 1, 2, \ldots, N_m$. Let $\{n_{mk}, k = 1, 2, \ldots, N_m\}$ be a sequence of independent and identically distributed zero mean, unit variance Gaussian random variables. We consider $\boldsymbol{\theta} \in [-2, 2]$ and the quantizer design of (11). For GSE, we choose $M = 2$ with each $R_m = 2$ and $N_m = N/2$. For group $m = 1$, we choose $\mathbf{t}_1 = -0.7$ and $\mathbf{c}_1 = [0.3, 0.57]^T$. For group $m = 2$, we select $\mathbf{t}_2 = 0.7$ and $\mathbf{c}_2 = [0.64, 0.85]^T$. For BSE, we have $M = 1$, $N_1 = N$ and we choose $\tau = 0$, $\mathbf{c}_{BSE} = [0.2, 0.3]^T$. For NBSE, we have $M = 1$, $N_1 = N$ and we choose $R = 3$, $\mathbf{t}_{NBSE} = [-0.7, 0.7]^T$, and $\mathbf{c}_{NBSE} = [0.46, 0.33, 0.38, 0.39, 0.46, 0.32]^T$. With these parameter choices, we plot the LFC CRLB vs θ for each of the three approaches. Fig. 3 shows that NBSE has a lower CRLB than BSE. This is intuitive since the higher order quantization should improve estimation variance. We also see that GSE has the lowest CRLB compared to NBSE and BSE over the complete range of θ. Therefore, by designing the GSE system appropriately it is possible to achieve lower estimation variance and also improve the range of estimation. Note however that the performance improvement with GSE depends on the parameter choices. Hence, for other M, R_m, $\mathbf{t}_m$ and $\mathbf{c}_m$ choices the GSE performance could be different and possibly inferior to BSE and NBSE.

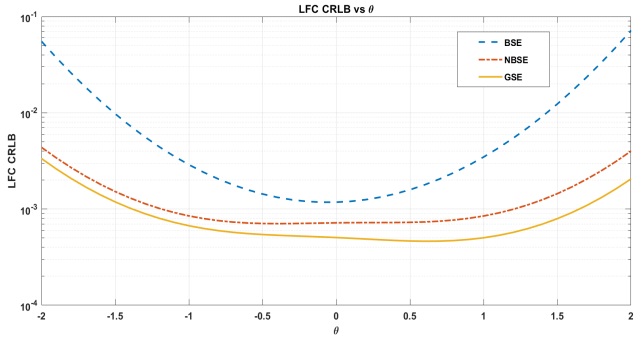

Figure 3. LFC CRLB as a function of $\boldsymbol{\theta}$

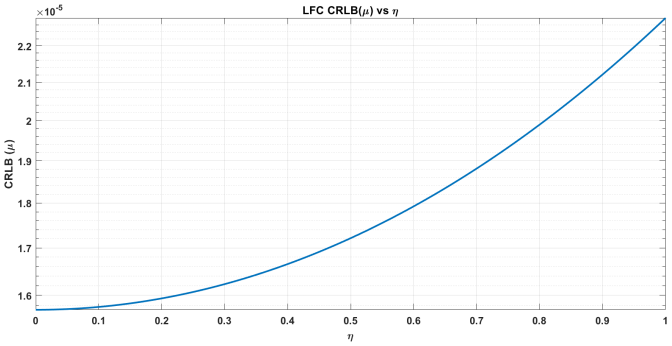

Figure 4. GSE LFC CRLB(μ) as a function of η

B. Estimation Capability

In section V we stated that the maximum dimension of $\boldsymbol{\theta}$ that can be estimated with stochastic encryption is upper-bounded. To support that result we will show that BSE can estimate only scalar parameters. We consider GSE with $M = 2$, $N = 1000$, $N_m = N/2$ and $x_{mk} = \mu + n_{mk}, m = 1, 2, k = 1, 2, \ldots, N_m$. Let $\{n_{mk}, k = 1, 2, \ldots, N_m\}$ be a sequence of independent and identically distributed zero mean, variance σ^2 Gaussian random variables. For each group we have $R_m = 2$ and $\mathbf{c}_m = [0.8, 0.8]^T$. We consider each binary quantizer design as per (11) and we choose, $\mathbf{t}_1 = -\eta$ and $\mathbf{t}_2 = \eta$ with η varying from $\eta = 0$ to $\eta = 1$. We consider

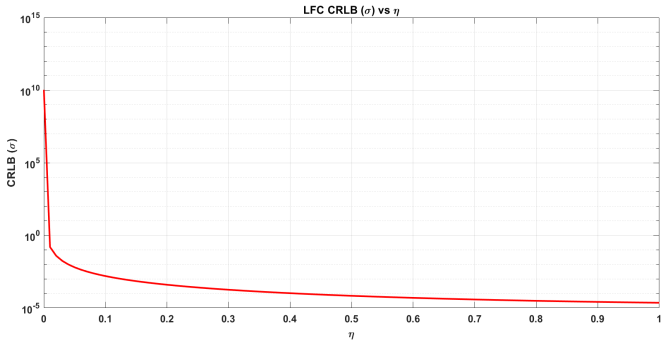

Figure 5. GSE LFC CRLB(σ) as a function of η

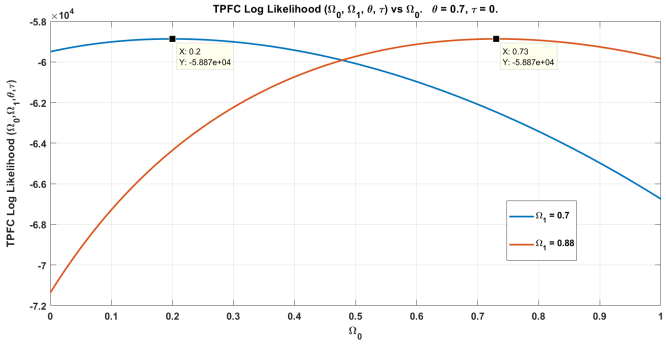

Figure 6. TPFC log likelihood function for BSE as a function of Ω_0 and Ω_1

vector parameter estimation with $\boldsymbol{\theta} = [\mu, \sigma]^T$. Fig. 4 illustrates the LFC CRLB for μ (CRLB(μ)) as a function of η while Fig. 5 illustrates the LFC CRLB for σ (CRLB(σ)) a function of η. CRLB(μ) and CRLB(σ) are the $(1,1)$ and $(2,2)$ elements of the of the 2×2 $\boldsymbol{\Psi}_{GSE}(\boldsymbol{\theta})$ respectively. In Fig. 5 we see that CRLB(σ)$\rightarrow \infty$ as $\eta \rightarrow 0$ and that it improves with increasing η. When $\eta = 0$ both the groups have the same quantizer threshold along with the same R_m, N_m and $\mathbf{c}_m$, and hence GSE becomes BSE. Since CRLB(σ)$\rightarrow \infty$ at $\eta = 0$ we see that both μ and σ cannot be estimated with BSE[2]. As η increases the system becomes GSE with $M = 2$ distinct groups and hence by (10) vector parameter estimation is possible.

C. Secrecy Capability

In section VI we showed that stochastic encryption is secure against eavesdropping in the domain of unbiased estimators. We augment this by illustrating the TPFC estimation log likelihood function for the simplest case of BSE. We consider a network of $N = 1000$ sensors with the additive i.i.d Gaussian noise samples having a standard normal distribution. To estimate θ we employ BSE with $\tau = 0$, $\Omega_0 = 0.2$ and $\Omega_1 = 0.7$. A TPFC intercepting the encrypted data has to estimate $\boldsymbol{\phi} = [\theta, \tau, \Omega_0, \Omega_1]^T$ to determine θ. Fixing $\theta = 0.7$ and $\tau = 0$, we plot the TPFC estimation log likelihood as a function of Ω_0 for two different values of Ω_1 which are 0.7 and 0.88 in Fig. 6. From the figure, we can see that the maximum TPFC log-likelihood for $\Omega_1 = 0.7$ is the same as the maximum TPFC log-likelihood for $\Omega_1 = 0.88$. Hence, the maxima $(\Omega_1 = 0.88, \Omega_0 = 0.73, \theta = 0.7, \tau = 0)$ has the same likelihood as the maxima $(\Omega_1 = 0.7, \Omega_0 = 0.2, \theta = 0.7, \tau = 0)$ (True values). Similarly, there are multiple other points of $(\Omega_1, \Omega_0, \theta, \tau)$ that have the same likelihood. This indicates the difficulty an estimator is faced with in this situation. See [4] for further discussion.

[2]The same singularity was observed with equal non-zero thresholds.

VIII. Conclusion

In this paper we presented three results. First, we generalized BSE and NBSE to GSE which has the capability to provide a lower mean estimation variance and improved range of estimation at a LFC depending on the system parameter choices. Next, we presented an analysis of estimation and secrecy capabilities of stochastic encryption. Applying results from [4], we showed that for each stochastic encryption scheme the dimension of $\boldsymbol{\theta}$ that can be estimated using it at a LFC, is upper bounded. Finally, we analyzed the secrecy provided by each stochastic encryption scheme against passive eavesdropping. Using results from [3][4] we showed that the FIM of a TPFC trying to estimate $\boldsymbol{\phi}$ from the encrypted data is singular under our assumptions. Thus, there exists no unbiased estimator that a TPFC can use to estimate $\boldsymbol{\phi}$. Hence, stochastic encryption is secure in the domain of unbiased estimators. Numerical results illustrating the three contributions were presented. Currently, we are investigating efficient methods to deploy the proposed approaches in practical situations.

References

[1] H. Hayouni, M. Hamdi and T. H. Kim, "A Survey on Encryption Schemes in Wireless Sensor Networks," *2014 7th International Conference on Advanced Software Engineering and Its Applications, Haikou*, 2014, pp. 39-43.

[2] T. C. Aysal and K. E. Barner, "Sensor Data Cryptography in Wireless Sensor Networks," in *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 273-289, June 2008.

[3] J. Zhang, R. S. Blum and L. Kaplan, "Cyber attacks on estimation sensor networks and IoTs: Impact, Mitigation and Implications to unattacked systems," 42^{nd} *IEEE International conference on Acoustics, Speech and Signal Processing*, 2017

[4] J. Zhang, R. S. Blum,L. Kaplan and X. Lu, "A fundamental limitation on maximum parameter dimension for accurate estimation using quantized data," https://arxiv.org/abs/1605.07679, submitted to *IEEE Transactions on Information Theory*

[5] A. N. Samudrala and R. S. Blum, "Asymptotic analysis of a new low complexity encryption approach for the Internet of Things, smart cities and smart grid," *2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), Singapore*, 2017, pp. 200-204.

[6] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[7] Rice, J. A. *Mathematical statistics and data analysis*. Thomson/Brooks/Cole, 2007.

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