Asymptotic Analysis of a New Low Complexity Encryption Approach for the Internet of Things, Smart Cities and Smart Grid

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*— Parameter estimation in wireless sensor networks (WSN) using encrypted non-binary quantized data is studied. In a WSN, sensors transmit their observations to a fusion center through a wireless medium where the observations are susceptible to unauthorized eavesdropping. Encryption approaches for WSNs with fixed threshold binary quantization were previously explored. However, fixed threshold binary quantization limits parameter estimation to scalar parameters. In this paper, we propose a stochastic encryption approach for WSNs that can operate on non-binary quantized observations and has the capability for vector parameter estimation. We extend a binary stochastic encryption approach proposed previously, to a non-binary generalized case. Sensor outputs are quantized using a quantizer with $R+1$ levels, where $R \in \{1, 2, 3, \ldots\}$, encrypted by flipping them with certain flipping probabilities, and then transmitted. Optimal estimators using maximum-likelihood estimation are derived for both a legitimate fusion center (LFC) and a third party fusion center (TPFC) perspectives. We assume the TPFC is unaware of the encryption. Asymptotic analysis of the estimators is performed by deriving the Cramer-Rao lower bound for LFC estimation, and the asymptotic bias and variance for TPFC estimation. Numerical results validating the asymptotic analysis are presented.**

***Keywords*— Information security, sensor networks, low complexity encryption, secure estimation, stochastic encryption.**

I. Introduction

Recent advancements in wireless communications, digital electronics and Micro Electro-Mechanical Systems (MEMS) have led to the emergence of infrastructure systems such as smart grid, smart homes, smart water networks that connect our world intricately. Together such systems are associated with a single concept called, the internet of things (IoT), where networks of embedded devices called sensors are used to perform the required tasks of monitoring and information exchange. A WSN is composed of a large number of low-cost and low-power devices called sensors that are capable of monitoring information from an environment, processing and transmitting data. Sensor networks for parameter estimation have been successfully employed in many applications ranging from commercial systems to military systems [1]. In many of these applications, the sensors monitor critical information like patient health or military data and communicate the information to a location, referred to here as a legitimate fusion center (LFC). Such wireless communication makes the systems susceptible to a passive eavesdropper, referred to as a third party fusion center (TPFC). Protection against eavesdropping is thus an important requirement in the design of sensor networks. However, the limited processing power, energy and memory size of the sensors, along with the small required maximum delays 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 [2].

As is the case of all digital communications the sensors must employ quantized data. The sensors quantize their observations and transmit the quantized symbols to a LFC. Much of the previous work has focused on scalar parametric estimation using the most restrictive case of binary quantization [3]-[5]. In [5], the authors propose a low-complexity encryption approach considering fixed binary quantization (with a single fixed threshold for all sensors). They propose a binary channel-like encryption scheme at each sensor that flips the quantized binary symbols with given probabilities. The encrypted symbols are then transmitted to the LFC . The encryption key is the bit flipping probabilities. By deriving the appropriate asymptoic analysis, these authors show that a significant bias can be introduced to a TPFC with minimum cost to the LFC estimation variance.

A significant drawback of the approach of [5] is that recent work shows it is limited to estimation of only scalar parameters. In practice, we might need to estimate vector parameters. Estimating the position of an object is one example. In [6]-[8], it was shown that only scalar parameter estimation is possible with fixed threshold binary quantization. With the objective of making vector parameter estimation possible, we extend the binary stochastic encryption proposed in [5] to an encryption scheme that can operate on non-binary quantized data. Each sensor employs a quantizer with $R+1$ levels, $R \in \{1, 2, 3, \ldots\}$. Each quantized symbol is

This work was supported by the U. S. Army Research Laboratory and the U.S. Army Research Office under Agreement Number W911NF-14-1-0245 and was also supported by the NSF under grant CNS 1702555.

then flipped into one of the $R+1$ symbols with given flipping probabilities, which is the encryption scheme. The encrypted symbols are then transmitted. A LFC, with the knowledge of the flipping probabilities (encryption key) would be able to estimate the unknown vector parameter but a TPFC unaware of the encryption key would suffer from estimation bias. We derive the ML estimator of the LFC that has the encryption key, and the ML estimator of the TPFC, assuming the TPFC is not aware of the encryption key. Primarily, our focus in this study is providing an asymptotic analysis of these estimators. We present the Cramer-Rao lower bound (CRLB) for the LFC since it is asymptoically unbiased with known variance [9], and the asymptotic bias and variance for the TPFC estimators which are new.

The paper is structured as follows. In Section II, the encryption scheme is formulated. Section III, discusses ML estimation of the unknown parameter from a LFC perspective while Section IV, discusses ML estimation of the unknown parameter from a TPFC perspective. In Section V, we illustrate numerical results validating the asymptotic analysis and also highlight limitations of the proposed encryption scheme. Finally, conclusions are drawn in Section VI.

II. Encrypted Wireless Sensor Network Model

Consider a set of N distributed sensors, each making observations of an unknown deterministic vector parameter θ. The observation at k^{th} sensor $\mathbf{x}_k$, corrupted by additive noise $\mathbf{n}_k$, is denoted by

$$\mathbf{x}_k = \theta + \mathbf{n}_k. \tag{1}$$

for $k = 1, 2, \ldots, N$. The probability density function of the additive noise is denoted by $\mathbf{n}_k \sim f()$. We assume that each sensor employs a common quantizer with $R+1$ levels. Each sample $\mathbf{x}_k$ is quantized to u_k, where $u_k \in \{0, 1, 2, \ldots, R\}$. The quantizer at each sensor is described by a set of non-overlapping regions $\{A_0, A_1, ..., A_R\}$, such that the quantizer will assign the symbol $u_k = \ell_k$ to any input $\mathbf{x}_k \in A_{\ell_k}$ for $\ell_k \in \{0, 1, 2, \ldots, R\}$. In accordance with the proposed encryption scheme, at each sensor the quantizer output u_k is flipped to $\tilde{u}_k$ with a flipping probability $pr(\tilde{u}_k = \tilde{\ell_k}|u_k = \ell_k)$. $pr(\tilde{u}_k = \tilde{\ell_k}|u_k = \ell_k)$ is the conditional probability that $\tilde{u}_k = \tilde{\ell_k}$ given that $u_k = \ell_k$, for $\ell_k, \tilde{\ell_k} \in \{0, 1, 2, \ldots, R\}$. The encrypted symbols $\tilde{u}_k$ are then transmitted to the LFC, and at the same time could possibly be received by an unauthorized TPFC. The problem of interest at both the LFC and TPFC, is to estimate θ from the received $\tilde{u}_k$. The system model is illustrated in Fig. 1.

III. LFC Estimation Performance

A. LFC Maximum Likelihood Estimator

In this section, we derive the ML estimator of θ at a LFC. In our analysis in this paper, we make the following assumptions.

Assumption 1: The probability density function of $\mathbf{x}_k$ is $f(\mathbf{x}_k|\theta)$, which depends on the parameter θ. We assume that $f(\mathbf{x}_k|\theta)$ obeys regularity (smoothness) conditions [9] such that interchanges involving derivatives with respect to θ and integrals with respect to $\mathbf{x}_k$ are valid.

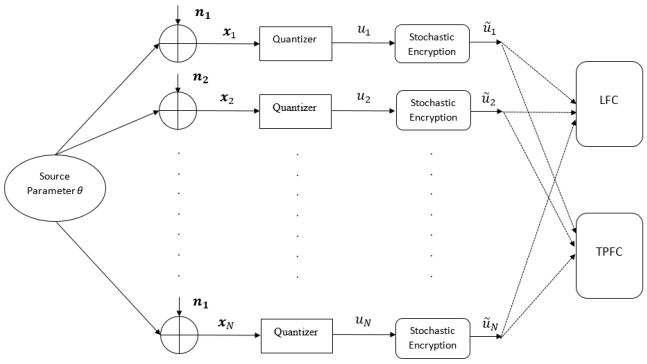

Figure 1. Encrypted wireless sensor network model

Assumption 2: The observations at different sensors $\mathbf{x}_k$ are statistically independent and identically distributed (i.i.d.)

We define the probability that the quantizer output is $u_k = \ell_k$ for $\ell_k \in \{0, 1, 2, \ldots, R\}$ as

$$pr(u_k = \ell_k|\theta) \quad = \quad \int_{\mathbf{x}_k \in A_{\ell_k}} f(\mathbf{x}_k|\theta) d\mathbf{x}_k. \tag{2}$$

We denote the indicator function $I(\ell_k = \ell'_k)$ as taking on the value unity when $\ell_k = \ell'_k$ and zero otherwise, where $\ell'_k \in \{0, 1, 2, \ldots, R\}$. Using the law of total probability [10], we write the probability that the encrypted symbol is $\tilde{u}_k = \ell_k$ as

$$pr\left(\tilde{u}_k = \ell_k|\theta\right) = \sum_{\ell'_k=0}^{R} \Big(pr(\tilde{u}_k = \ell_k|u_k = \ell'_k) \times pr(u_k = \ell'_k|\theta) \Big). \tag{3}$$

Under Assumption 2, the joint probability mass function of the encrypted symbols $\tilde{\mathbf{u}} = (\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_N)^T$ evaluated at $\bar{\ell} = (\ell_1, \ell_2, \ldots, \ell_N)^T$ is

$$pr(\tilde{\mathbf{u}} = \bar{\ell}|\theta) = \prod_{k=1}^{N} \prod_{\ell'_k=0}^{R} pr(\tilde{u}_k = \ell_k|\theta)^{I(\ell_k = \ell'_k)}. \tag{4}$$

Taking the logarithm of (4) we obtain the log-likelihood function evaluated at $\tilde{\mathbf{u}} = \bar{\ell}$ as

$$\begin{aligned} L_{LFC}(\theta) &= \ln pr(\tilde{\mathbf{u}} = \bar{\ell}|\theta) \\ &= \sum_{k=1}^{N} \sum_{\ell'_k=0}^{R} I(\ell_k = \ell'_k) \ln pr(\tilde{u}_k = \ell'_k|\theta). \end{aligned} \tag{5}$$

The ML estimate of θ for a LFC is the solution to max_θ $L_{LFC}(\theta)$, and is represented as $\hat{\theta}(\bar{\ell})_{LFC}$. Setting the derivative of (5) to zero yields a necessary condition for the ML estimate as that θ satisfying the following equation.

$$\sum_{k=1}^{N}\sum_{\ell'_k=0}^{R} I(\ell_k=\ell'_k)\frac{\frac{d}{d\theta}pr(\tilde{u}_k=\ell'_k|\theta)}{pr(\tilde{u}_k=\ell'_k|\theta)}=0. \quad (6)$$

Using (2), (3) in (6) and employing Assumption 2 we obtain a more explicit equation as

$$\frac{1}{N}\sum_{k=1}^{N}\sum_{\ell'_k=0}^{R}\Bigg(I(\ell_k=\ell'_k)\times \frac{\sum_{\tilde{\ell}_k=0}^{R} pr(\tilde{u}_k=\ell'_k|u_k=\tilde{\ell}_k)\int_{\mathbf{x}_k\in A_{\tilde{\ell}_k}}\frac{d}{d\theta}f(\mathbf{x}_k|\theta)d\mathbf{x}_k}{\sum_{\tilde{\ell}_k=0}^{R} pr(\tilde{u}_k=\ell'_k|u_k=\tilde{\ell}_k)\int_{\mathbf{x}_k\in A_{\tilde{\ell}_k}} f(\mathbf{x}_k|\theta)d\mathbf{x}_k}\Bigg)=0. \quad (7)$$

We solve (7) iteratively to find the ML estimate θ_{LFC} for a received $\bar{\ell}$ after choosing an appropriate initialization.

B. LFC Cramer-Rao Lower Bound (CRLB)

Being a standard ML estimation problem with known distribution, the LFC ML estimate is asymptotically unbiased and asymptotically approaches a Gaussian random variable with a variance approaching the CRLB [9]. In this section, we derive the CRLB for a LFC. To find the CRLB we first compute the Fisher information $J(\theta)$. Let the true value of θ be θ_0. Using Assumption 2, the Fisher information is given by

$$\begin{aligned}J(\theta) &= E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\left(\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)\right)^2|_{\theta=\theta_0}\right\}\\ &= \sum_{k=1}^{N} E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\left(\frac{d}{d\theta}\ln pr(\tilde{u}_k=\ell_k|\theta)\right)^2|_{\theta=\theta_0}\right\}\\ &= \sum_{k=1}^{N} E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\left(\frac{\frac{d}{d\theta}pr(\tilde{u}_k=\ell_k|\theta)}{pr(\tilde{u}_k=\ell_k|\theta)}\right)^2|_{\theta=\theta_0}\right\} \quad (8)\end{aligned}$$

Applying Assumption 2 again, (8) can be simplified as

$$J(\theta) \quad = N\sum_{r=0}^{R}\frac{\left(\frac{d}{d\theta}pr(\tilde{u}_1=r|\theta)\right)^2}{pr(\tilde{u}_1=r|\theta)}|_{\theta=\theta_0} \quad (9)$$

Then, the CRLB is given by $\psi_{LFC}(\theta)=\frac{1}{I(\theta)}$.

IV. TPFC Estimation Performance

A. TPFC Maximum Likelihood Estimator

Now, we consider ML estimation at a TPFC. We assume that the TPFC is unaware of the encryption and hence believes the received symbols $\tilde{u}_k$ are unencrypted. Thus, it will perform calculations which would be ML estimation if the symbols were unencrypted, but the symbols are actually encrypted. Thus, the effect of this mismatched estimation is quite different from standard ML. It is assumed that the TPFC has knowledge of the quantizer design and the parameters of the pdf characterizing the WSN environment. Note that TPFC performance degrades further if it is unaware of these parameters. TPFC ML processing is exactly (7) with $pr(\tilde{u}_k=\ell'_k|u_k=\tilde{\ell}_k)=1$ if $\ell'_k=\tilde{\ell}_k$ and zero otherwise. Thus the TPFC will use the estimate $\hat{\theta}_{TPFC}$ which is the θ satisfying

$$\frac{1}{N}\sum_{k=1}^{N}\sum_{\ell'_k=0}^{R} I(\ell_k=\ell'_k)\frac{\int_{\mathbf{x}_k\in A_{\ell'_k}}\frac{d}{d\theta}f(\mathbf{x}_k|\theta)d\mathbf{x}_k}{\int_{\mathbf{x}_k\in A_{\ell'_k}} f(\mathbf{x}_k|\theta)d\mathbf{x}_k}=0. \quad (10)$$

Similar to the LFC ML estimation, an iterative algorithm can be employed to solve (10) to find $\hat{\theta}_{TPFC}$. The TPFC log-likelihood function is the log-likelihood function (5) with $pr(\tilde{u}_k=\ell'_k|u_k=\tilde{\ell}_k)=1$ if $\ell'_k=\tilde{\ell}_k$ and zero otherwise.

B. TPFC ML Estimator Asymptotic Bias

In this section, we present asymptotic bias of the TPFC ML estimator. Let the true value of θ be θ_0. As $N\to\infty$, the strong law of large numbers [10] and Assumption 2 imply that the sums over k in (10) scaled by $\frac{1}{N}$ approach the common expected value of the term being summed, where the expectation is taken with respect to the true distribution of the observations $pr(\tilde{u}_1=\ell_1|\theta_0)\cdots pr(\tilde{u}_N=\ell_N|\theta_0)$ (the flipped symbols), so the limit of (10) becomes

$$\sum_{r=0}^{R} pr(\tilde{u}_1=r|\theta_0)\frac{\int_{\mathbf{x}_1\in A_r}\frac{d}{d\theta}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}{\int_{\mathbf{x}_1\in A_r} f(\mathbf{x}_1|\theta)d\mathbf{x}_1}=0. \quad (11)$$

where (3) can be inserted. As before, we solve (11) iteratively and the solution gives the asymptotic mean, represented as $\hat{\theta}_{ATPFC}$. The asymptotic bias of the estimator is then computed as, $\beta_{TPFC}(\theta)=\hat{\theta}_{ATPFC}-\theta_0$. The TPFC asymptotic bias is the most important performance metric since the limiting distribution of the TPFC ML estimate is Gaussian with a variance that shrinks to zero as N gets very large.

C. TPFC ML Estimator Asymptotic Variance

In this section, we derive the asymptotic variance of the TPFC ML estimator, denoted by $\psi_{ATPFC}(\theta)$. As $N\to\infty$ we know that $\hat{\theta}_{TPFC}\to\hat{\theta}_{ATPFC}$. For large N, we employ a Taylor series, about $\hat{\theta}_{ATPFC}$, for the derivative of the log-likelihood function to obtain the equation describing the ML estimate (10) for a specific observed $\tilde{\mathbf{u}}=\bar{\ell}$ as

$$\begin{aligned}&\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}_{TPFC}} =\\ &\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}_{ATPFC}}+\\ &(\hat{\theta}_{TPFC}-\hat{\theta}_{ATPFC})\ \frac{d^2}{d\theta^2}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}'}=0. \quad (12)\end{aligned}$$

where the equality follows from the mean value theorem, for some $\hat{\theta}_{ATPFC}<\hat{\theta}'<\hat{\theta}_{TPFC}$. Re-arranging (12) yields

$$\begin{aligned}&\sqrt{N}(\hat{\theta}_{TPFC}-\hat{\theta}_{ATPFC})=\\ &\frac{\frac{1}{\sqrt{N}}\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}_{ATPFC}}}{-\frac{1}{N}\frac{d^2}{d\theta^2}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}'}}. \quad (13)\end{aligned}$$

As $N \to \infty$ and $\hat{\theta}_{ATPFC} < \hat{\theta}' < \hat{\theta}_{TPFC}$, for sufficiently large N it must be that $\hat{\theta}' \to \hat{\theta}_{ATPFC}$, so the denominator of (13) becomes

$$\begin{aligned}
&-\frac{1}{N}\frac{d^2}{d\theta^2}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}'} \\
&\to -\frac{1}{N}\sum_{k=1}^{N}\frac{d^2}{d\theta^2}\ln pr(\tilde{u}_k=\ell_k|\theta)|_{\theta=\hat{\theta}_{ATPFC}} \\
&\to -E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\frac{d^2}{d\theta^2}\ln pr(\tilde{u}_k=\ell_k|\theta)|_{\theta=\hat{\theta}_{ATPFC}}\right\} \\
&= -\sum_{r=0}^{R} pr(\tilde{u}_1=r|\theta_0)\Bigg(\frac{\int_{\mathbf{x}_1\in A_r}\frac{d^2}{d\theta^2}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}{\int_{\mathbf{x}_1\in A_r}f(\mathbf{x}_1|\theta)d\mathbf{x}_1} \\
&\quad -\left(\frac{\int_{\mathbf{x}_1\in A_r}\frac{d}{d\theta}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}{\int_{\mathbf{x}_1\in A_r}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}\right)^2\Bigg)|_{\theta=\hat{\theta}_{ATPFC}} = c \qquad (14)
\end{aligned}$$

where convergence to the constant c is with probability one from the strong law of large numbers. Given Assumption 2, from the central limit theorem the numerator of (13) approaches a Gaussian distribution with mean a and variance b as in

$$\frac{1}{\sqrt{N}}\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}_{ATPFC}} \to N(a,b). \qquad (15)$$

Using (11) we find a as

$$\begin{aligned}
a &= E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\frac{1}{\sqrt{N}}\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)|_{\theta=\hat{\theta}_{ATPFC}}\right\} \\
&= E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\frac{d}{d\theta}\ln pr(\tilde{u}_k=\ell_k|\theta)\right\}|_{\theta=\hat{\theta}_{ATPFC}} \\
&= \sqrt{N}\left(E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\frac{d}{d\theta}\ln pr(\tilde{u}_1=\ell_1|\theta)\right\}|_{\theta=\hat{\theta}_{ATPFC}}\right) \\
&= 0. \qquad (16)
\end{aligned}$$

Since $a = 0$ from (16), using the definition of variance we compute b as

$$\begin{aligned}
b &= E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}\left\{\left(\frac{1}{\sqrt{N}}\frac{d}{d\theta}\ln pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta)\right)^2|_{\theta=\hat{\theta}_{ATPFC}}\right\} \\
&= \sum_{k=1}^{N}\frac{E_{pr(\tilde{\mathbf{u}}=\bar{\ell}|\theta_0)}}{N}\left\{\left(\frac{d}{d\theta}\ln pr(\tilde{u}_k=\ell_k|\theta)\right)^2|_{\theta=\hat{\theta}_{ATPFC}}\right\} \\
&= \sum_{r=0}^{R} pr(\tilde{u}_1=r|\theta_0)\left(\frac{\int_{\mathbf{x}_1\in A_r}\frac{d}{d\theta}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}{\int_{\mathbf{x}_1\in A_r}f(\mathbf{x}_1|\theta)d\mathbf{x}_1}\right)^2\Bigg|_{\theta=\hat{\theta}_{ATPFC}} \qquad (17)
\end{aligned}$$

Next, we employ Slutsky's Theorem which states that if $x(N) \to x$ in distribution as $N \to \infty$ and $y(N) \to c$ as $N \to \infty$ with probability one, where c is a constant, then $x(N)/y(N) \to x/c$ in distribution as $N \to \infty$. Applying this to (13), we have that $\sqrt{N}(\hat{\theta}_{TPFC} - \hat{\theta}_{ATPFC})$ approaches a Gaussian distribution with mean 0 and variance b/c^2. Thus, $\hat{\theta}_{TPFC}$ asymptotically approaches a Gaussian distribution with mean $\hat{\theta}_{ATPFC}$ and variance $\psi_{ATPFC}(\theta) = \frac{b}{Nc^2}$.

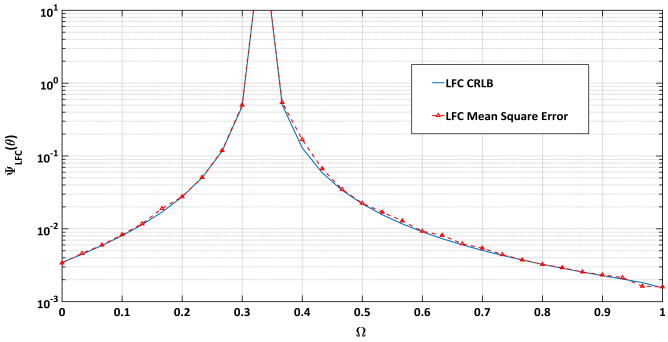

Fig. 2. LFC CRLB as a function of Ω

V. Simulation Results

In this section, we present simulations results illustrating the asymptotic behaviour of the LFC and TPFC ML estimators. We simulate a WSN of $N = 1000$ sensors with Additive White Gaussian noise. A 3-level quantizer with quantizer outputs $u_k \in \{0, 1, 2\}$ is considered. The quantization regions are $A_0 = (-\infty, -1)$, $A_1 = [-1, 1]$ and $A_2 = (1, \infty)$. We consider stochastic encryption with symmetric encryption key, i.e., $pr(\tilde{u}_k = \ell_k | u_k = \ell'_k) = \Omega$ if $\ell_k = \ell'_k$ and $pr(\tilde{u}_k = \ell_k | u_k = \ell'_k) = (1-\Omega)/2$ if $\ell_k \neq \ell'_k$, for $\ell_k, \ell'_k \in \{0, 1, 2\}$. For simplicity, we simulate a scalar parameter estimation with $\theta = -1.3$. By varying Ω, we change the encryption and simulate the corresponding LFC and TPFC estimation performance. Monte-Carlo simulations were run, and the bias and variance were plotted against Ω.

The first performance metric to be illustrated is the LFC CRLB $\psi_{LFC}(\theta)$. In Fig. 2, the LFC CRLB $\psi_{LFC}(\theta)$ as well as the simulated mean-square-error of the LFC ML estimator are plotted as a function of Ω. It is clear that the simulated mean-square-error approaches the $\psi_{LFC}(\theta)$ for every Ω. When $\Omega = 1/3$, $pr(\tilde{u}_k = \tilde{\ell}_k) = 1/3, \forall \tilde{\ell}_k \in \{0, 1, 2\}$, i.e., the encrypted symbols $\tilde{u}_k$ are independent of u_k and hence carry no information about θ. Hence, as $\Omega \to 1/3$ it can be observed in Fig. 2 that $\psi_{LFC}(\theta) \to \infty$. This is similar to the binary quantization case in [5], in which $\psi_{LFC}(\theta) \to \infty$ as $\Omega \to 1/2$. Note that $\Omega = 1/3$ produces $\psi_{LFC}(\theta) \to \infty$ for a symmetric encryption key case. For an asymmetric encryption key case, different values of Ω produce $\psi_{LFC}(\theta) \to \infty$.

Next, we simulate the performance for a TPFC. First, we look at the bias, $\beta_{TPFC}(\theta)$. Fig. 3 illustrates the TPFC ML estimator asymptotic bias $\beta_{TPFC}(\theta)$ as well as the simulated estimator bias as functions of Ω. It can be clearly seen that the bias is maximum for $\Omega = 0$ and reduces linearly to zero bias at $\Omega = 1$. This is intuitive, as $\Omega = 0$ corresponds to the case where the symbols u_k are most often flipped into $\tilde{u}_k$ with $\tilde{u}_k \neq u_k$ thereby introducing a large bias. Similarly $\Omega = 1$ represents the case of no encryption thereby introducing zero bias. Note that the linearly decreasing behavior of the $\beta_{TPFC}(\theta)$ with increasing Ω, is true only for a symmetric encryption key. Similar linear behavior with a symmetric key is illustrated for binary quantization case in [5]. If a different encryption key

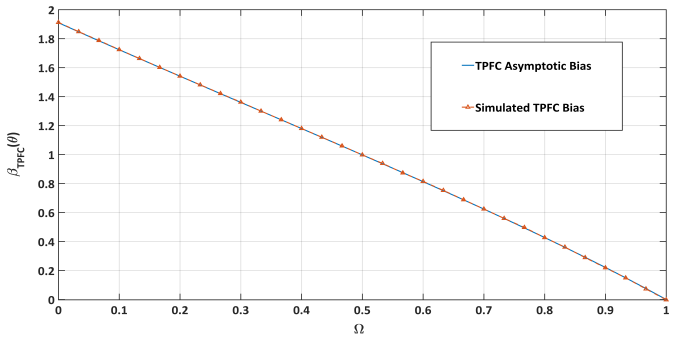

Figure 3. TPFC asymptotic bias as a function ofΩ

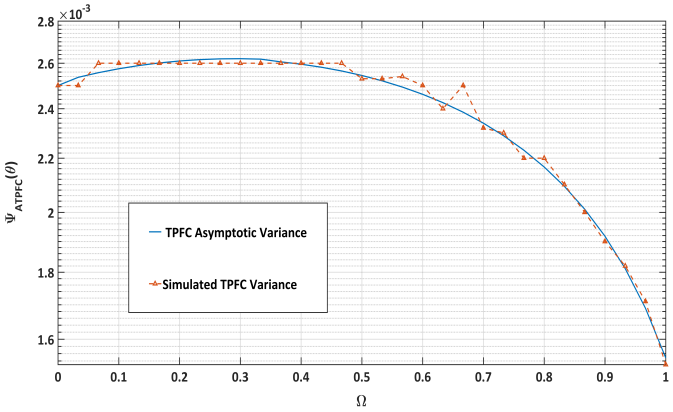

Figure 4. TPFC asymptotic variance as a function ofΩ

is employed, the behavior would be different. It is important to note here that depending on the design of quantizer and the choice of the flipping probabilities, it is possible that the bias $\beta_{TPFC}(\theta) = 0$ for certain values of θ. For example, in the current case of a symmetric encryption key, if θ is the midpoint of the quantization region A_1, then the bias $\beta_{TPFC}(\theta) = 0$. Similar behavior was observed for the binary stochastic encryption in [5]. Hence, while designing a system using the proposed encryption, care has to be taken to avoid the zero-bias points.

Finally, the TPFC ML estimator asymptotic variance $\psi_{ATPFC}(\theta)$ is plotted in Fig. 4 as a function of Ω, along with the simulated variance. It can be observed that $\psi_{ATPFC}(\theta)$ is very small, and hence TPFC is mainly degraded by it's estimator bias.

VI. Conclusions and Future Work

In this paper, we introduced a stochastic encryption scheme for a wireless sensor network that can operate on non-binary quantized observations and is capable of estimation of vector parameters. The stochastic encryption was achieved by flipping the quantized symbols, where the flipping probabilities act as the encryption key. The optimal ML estimator for a LFC and a TPFC were derived, assuming that the TPFC has no knowledge of the encryption key. Asymptotic behavior of the estimators was analyzed by deriving the CRLB for the LFC, and the asymptotic bias and variance for the TPFC. Numerical results validating the asymptotic analysis were presented. Currently, we are investigating efficient methods to deploy the proposed approach in practical situations.

References

[1] I. F. Akyildiz, Weilian Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," in *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114, Aug 2002.

[2] 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.

[3] 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.

[4] K. Agrawal, A. Vempaty, H. Chen and P. K. Varshney, "Target localization in Wireless Sensor Networks with quantized data in the presence of Byzantine attacks," *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, CA, 2011, pp. 1669-1673.

[5] 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.

[6] J. Zhang, R. S. Blum, X. Lu and D. Conus, "Asymptotically Optimum Distributed Estimation in the Presence of Attacks," in *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1086-1101, March1, 2015.

[7] 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*

[8] 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

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

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