sensors

MDPI

Article

Pseudo-Random Encryption for Security Data Transmission in Wireless Sensor Networks

Liang Liu [1,2], Wen Chen [1,*], Tao Li [1] and Yuling Liu [1]

[1] College of Cybersecurity, Sichuan University, Chengdu 610065, China; liangzhai118@163.com (L.L.); litao@scu.edu.cn (T.L.); xiaye524@163.com (Y.L.)
[2] College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China
* Correspondence: wenchen@scu.edu.cn

Received: 13 May 2019; Accepted: 27 May 2019; Published: 29 May 2019

Abstract: The security of wireless sensor networks (WSN) has become a great challenge due to the transmission of sensor data through an open and wireless network with limited resources. In the paper, we discussed a lightweight security scheme to protect the confidentiality of data transmission between sensors and an ally fusion center (AFC) over insecure links. For the typical security problem of WSN's binary hypothesis testing of a target's state, sensors were divided into flipping and non-flipping groups according to the outputs of a pseudo-random function which was held by sensors and the AFC. Then in order to prevent an enemy fusion center (EFC) from eavesdropping, the binary outputs from the flipping group were intentionally flipped to hinder the EFC's data fusion. Accordingly, the AFC performed inverse flipping to recover the flipped data before data fusion. We extended the scheme to a more common scenario with multiple scales of sensor quantification and candidate states. The underlying idea was that the sensor measurements were randomly mapped to other quantification scales using a mapping matrix, which ensured that as long as the EFC was not aware of the matrix, it could not distract any useful information from the captured data, while the AFC could appropriately perform data fusion based on the inverse mapping of the sensor outputs.

Keywords: wireless sensor network; network security; pseudo-random function; distributed detection; likelihood ratio test

1. Introduction

Wireless sensor networks (WSNs) are commonly deployed in various practical applications to gather information from a hostile area or placements where human intervention is impossible, and thus can support many new and important areas, such as customer satisfaction survey [1], unmanned aerial vehicles [2], military surveillance [3], and fire detection [4]. However, due to the broadcast nature of the wireless network, sensor data are prone to be intercepted by unauthorized receivers [5,6]. The security of WSN mainly involve decentralized detection whereby the sensors send their measurements to an ally fusion center (AFC) which attempts to detect the state of nature using the data received from all the sensors, meanwhile an enemy fusion center (EFC), also located in a vicinity of the AFC, tries to eavesdrop on the wireless communications between sensors and the AFC [7].

The data confidentiality of WSN has been the subject of many studies using different security strategies [8,9]. To facilitate data encryption using symmetric cryptographic algorithms, the authors in [10,11] proposed novel lightweight anonymous authentication and key agreement (AKA) protocols for WSN, which combines multi-security stages: User registration, sensor node registration, login, authentication, key agreement, and password change. However, the issue of scalability remains a great challenge, since a sensor's life span is largely determined by its energy supply, which is difficult

to satisfy when the resource requirements of traditional encryption methods are at a high security level [12]. Therefore, the security solutions which demand excessive processing costs are not suitable for WSN [13].

With these challenges, several novel efficient approaches have been explored. Aysal and et al. [14] proposed that sensors can intentionally generate errors in the transmitted data to confuse eavesdroppers. They assume that the statistics of the induced error pattern is only available to the AFC, and the eavesdroppers who do not know this, fail to detect the real state from the eavesdropped data. In [15], the sensor outputs are randomly flipped through XOR operations on the local observations and independent binary variables. It is shown that using carefully selected operation parameters, we can hinder the EFC from making correct decisions while maintaining an acceptable error rate at the AFC. In [16], the sensor outputs are randomly mapped to other possible quantification levels according to a probability matrix. It is assumed that the AFC is aware of the matrix but the EFC is not. Therefore, the AFC can optimize its decision result along with the mapping probabilities while imposing a high error rate on the EFC. In [17], the authors developed a security scheme based on sensor censorship to maximize the J-divergence of the EFC while ensuring that the divergence of the AFC is zero. In [18], the authors proposed a lightweight encryption scheme whereby sensors flip their local decisions based on the instantaneous channel (to the AFC) fading gains which are unknown to the EFC due to the independence of the physical communication channels. They showed that information-theoretic perfect secrecy can be achieved in the condition that the global flipping and non-flipping data have the same size. However, the channel state information (CSI) changes from time to time, if the data flipping of each sensor is independently decided by its local CSI statistic, it will be difficult to satisfy the required equivalent condition. Moreover, currently most of the security schemes are designed for the binary hypothesis testing problems in which only two opposite candidate states are considered. However, in many practically-oriented applications, the target systems may contain multiple states and quantification scales. Thus a more general security model with a high efficiency is required to ensure the data confidentiality of WSN.

In the paper, we proposed a security data propagation method based on pseudo-random encryption (SDPR) for WSN. The underlying idea is that since most of the random functions generate stochastic sequence in a pseudo-random way, such that as long as the input initial random seed is the same, the functions will always return a same random digital sequence; it is nearly impossible to guess the generated sequence without the seed, even when the employed function is known. Furthermore, the time complexity of the pseudo-random function is much lower than traditional high security encryption algorithms, which makes them more applicable to the resource constrained environment of WSN. It should be noted that pseudo-functions are utilized to control the data interfering process in our method SDPR. The detailed discussion of pseudo-functions is out of the scope of the paper, but anyone interested in it can refer to [19–21] .

Firstly the typical security problem of distributed binary hypothesis testing in WSN is considered. It is assumed that the sensors report their measurements in the form of binary bits and sequentially transmit them to the AFC. At the beginning of a sensing cycle, the AFC and each sensor s_i select an unique seed (for pseudo-random functions) based on the channel state information. In each duplex time, s_i is randomly assigned to the flipping or non-flipping group based on the output of its random function. The sensors in the non-flipping group directly transmit their binary results to the AFC, while the others perform data encryption in a way to flip their quantized bits (zero turns to one and vice versa). Therefore, the AFC can recover the flipped data while the EFC can not, and the flipping will only prevent the EFC from fusing the correct result. Note that the seeds can be safely decided by the sensors and AFC based on the observation of the main channel state information. In [22–24], the wireless channel gains are considered as common randomness exclusively shared by the transmitter and legitimate receiver using the time-division duplexing (TDD) protocol. Specifically, in [18], the magnitude of the channel gain (MCG) is applied to the negotiation of secret keys. As the main channels (sensors to the AFC) and eavesdropping channels (sensors to the EFC) are statistically

independent when the EFC is located more than a half wavelength apart from both the sensors and AFC, it is impossible for the EFC to eavesdrop the seed information.

The scheme is then extended to a more common scenario with multiple quantification scales and candidate states. Our goal is to achieve data confidentiality of WSN by ensuring that the data observed by the EFC is useless for detection purposes. It should be noted that the proposed approach has minimal processing requirements and does not introduce any communication overhead.

2. The Description of the Security Model

2.1. System Model

The proposed security scheme SDPR for binary hypothesis testing is referred to as SDPR_B. The sensors observe a common target state which can be θ_0 or θ_1, and the measurement of each sensor s_i is quantized into binaries (local decision u_i) based on the measurement and decision threshold: If the measurement is less than the decision threshold, then $u_i = 1$ (s_i detected θ_1), otherwise $u_i = -1$ (s_i detected θ_0). Let $p(u_i = 1|\theta_1)$ and $p(u_i = 1|\theta_0)$ denote the local detection rate pd_i and false alarm rate pf_i of the i-th sensor respectively. Usually, the environment sensing performance pd_i and pf_i can be known in advance both by the AFC and EFC. In order to prevent the EFC from detecting the real state, the sensors in the flipping group flipped their binary outputs before sending to the AFC.

2.2. Flipping Encryption

A sensor is assigned to the flipping group if the output of its local pseudo-random function $rand(\cdot)$ is in the flipping domain $[\tau_4, \tau_3]$. Usually the output sequence of $rand(\cdot)$ is decided by its initial $seed$. Therefore, in order to make sure that the AFC and sensors can synchronously generate the same random sequences at the beginning of each sensing cycle, the AFC and each corresponding sensor s_i decide $seed_i$ based on the observation of the instantaneous channel state of the i-th main channel (s_i to AFC), such that $seed_i = g_i$, where g_i can be the magnitude of the channel gain or the CQI (channel quality indicator) level. As g_i is only known by the AFC and s_i due to the channel independence between the main channels and the eavesdropping channels [18], the EFC cannot eavesdrop $seed_i$ through the captured data in WSN.

Before the data transmission in each duplex time, the AFC first broadcasts pilot signals with flipping thresholds: $\{\tau_1, \tau_2, \tau_3, \tau_4\}$, where $\{\tau_1 > \tau_2 > \tau_3 > \tau_4\}$, to trigger the sensors to report their local decisions in a time-division duplex (TDD) manner. In particular, the thresholds remain constant during a duplexing time for pilot transmission and sensors' transmissions, say one block, and changes independently across blocks and sensors. Once the i-th sensor received the pilot signal, it immediately generates a random value $\varphi_i^{(t)}$ by $rand(\cdot)$ as shown in Equation (1):

$$\varphi_i^{(t)} = \begin{cases} rand(seed_i), & if \quad t = 1 \\ rand(\varphi_i^{(t-1)}), & if \quad t > 1 \end{cases} \tag{1}$$

The encrypted data (output) of the i-th sensor s_i is denoted by x_i. If $\tau_2 < \varphi_i^{(t)} < \tau_1$, then s_i is put into the non-flipping group and output $x_i = u_i$. On the contrary, s_i is assigned to the flipping group when $\tau_4 < \varphi_i^{(t)} < \tau_3$, and its output is $x_i = -u_i$, which is a bit-flipping version of the quantized data. The remaining sensors are dormant. Note that the EFC can also receive the sensor outputs through eavesdropping channels due to the broadcast nature of WSN. We emphasize that since the encryption is based on the pseudo-random functions' outputs, which are different between sensors, the EFC cannot distinguish the flipped data from the original outputs even if it has eavesdropped the pilot packets from the AFC. As a result, we expect that the EFC fails to detect the target state. In the next section, we will show that the AFC can correctly fuse the received data whereas the EFC totally failed to utilize the captured data.

2.3. The Fusion Result of the AFC

In this section, the fusion result of the AFC based on log-likelihood ratio (LLR) is analyzed. Suppose the number of activated sensors in one duplexing time is $k(k \leq N)$, then the input vector, which contains data from all the activated sensors, for the decision fusion is $z^A = [z_1^A ... z_k^A]$, and according to [25] the LLR-based fusion rule at the AFC is given by:

$$\Lambda = log\frac{P(z^A|\theta_1)}{P(z^A|\theta_0)} = \sum_{i=1}^{k} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} \underset{\theta_0}{\overset{\theta_1}{\gtrless}} 0 \tag{2}$$

$f(.|.)$ is the conditional probability density function. As the sensor outputs have been randomly flipped, z_i^A may not be the original sensor quantification. Note that the AFC holds the initial $seed_i$ of each sensor. Then the original quantized measurement u_i^A of $z_i^A, i = 1...k$, can be recovered by Equation (3):

$$u_i^A = \begin{cases} sign(z_i^A), & if \quad \tau_2 < \varphi_i^{(t)} = rand(\varphi_i^{(t-1)}) < \tau_1 \\ -sign(z_i^A), & if \quad \tau_4 < \varphi_i^{(t)} = rand(\varphi_i^{(t-1)}) < \tau_3 \end{cases} \tag{3}$$

where sign(z_i^A) represents the sign of z_i^A. Let $S_1 = \{i|z_i^A = 1, \tau_2 < \varphi_i^{(t)} < \tau_1\}$, $S_2 = \{i|z_i^A = -1, \tau_2 < \varphi_i^{(t)} < \tau_1\}$, $S_3 = \{i|z_i^A = 1, \tau_4 < \varphi_i^{(t)} < \tau_3\}$, and $S_4 = \{i|z_i^A = -1, \tau_4 < \varphi_i^{(t)} < \tau_3\}$. Then the left side of Equation (2) can be rewritten as:

$$\begin{aligned} \Lambda = &\sum_{i\in S_1} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} + \sum_{i\in S_2} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} + \\ &\sum_{i\in S_3} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} + \sum_{i\in S_4} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} \end{aligned} \tag{4}$$

When $i \in S_1$, it means that the original quantification u_i^A is not flipped and $u_i^A = z_i^A = 1$, therefore the LLR value of the i-th sensor in S_1 becomes:

$$\sum_{i\in S_1} log\frac{f(z_i^A|\theta_1)}{f(z_i^A|\theta_0)} = \sum_{i\in S_1} log\frac{f(u_i^A = 1|\theta_1)}{f(u_i^A = 1|\theta_0)} = \sum_{i\in S_1} log\frac{pd_i}{pf_i} \tag{5}$$

In the same manner, we can compute the LLR values in different set $S_2 \sim S_4$, finally the LLR based fusion rule can be reduced to the following one:

$$\begin{aligned} \Lambda &= log\frac{f(z^A|\theta_1)}{f(z^A|\theta_0)} \\ &= \sum_{i\in S_1} log\frac{pd_i}{pf_i} + \sum_{i\in S_2} log\frac{1-pd_i}{1-pf_i} + \\ &\sum_{i\in S_3} log\frac{1-pd_i}{1-pf_i} + \sum_{i\in S_4} log\frac{pd_i}{pf_i} \underset{\theta_0}{\overset{\theta_1}{\gtrless}} 0 \end{aligned} \tag{6}$$

Note that this approximation can be viewed as a modified version of the Chair–Varshney fusion rule, which only utilizes local false alarm and detection probabilities [25].

3. Security Analysis

In the security analysis of the proposed scheme, the main concern is whether the EFC can get useful information from the eavesdropped data. At the EFC, the captured data from the i-th sensor is

denoted by z_i^E and the input vector for EFC's fusion rule is $z^E = [z_1^E, z_2^E ... z_k^E]$. And the final decision of EFC is is shown in Equation (7).

$$L = log\frac{f(z^E|\theta_1)}{f(z^E|\theta_0)} \underset{\theta_0}{\overset{\theta_1}{\gtrless}} 0 \tag{7}$$

From Equation (7) we can see that data confidentiality can be achieved by deriving $f(z^E|\theta_1)$ equals $f(z^E|\theta_0)$, which makes the LLR value at the EFC always equals to zero, and the EFC will totally ignore the data since it cannot make a final decision for the binary hypothesis testing problem when $L = 0$.

Obviously, the sensor outputs are independent with each other, and $f(z^E|\theta_1) = \prod_{i=1}^{k} f(z_i^E|\theta_1)$. Then $f(z_i^E|\theta_1)$ is computed with the total probability theorem, as shown in Equation (8).

$$\begin{aligned} f(z_i^E|\theta_1) &= \sum_{u_i}\sum_{x_i} f(z_i^E, x_i, u_i|\theta_1) \\ &= \sum_{u_i} p(u_i|\theta_1) \sum_{x_i} \int f(z_i^E, \varphi_i, x_i|u_i, \theta_1) d\varphi_i \\ &= \sum_{u_i} p(u_i|\theta_1) \sum_{x_i} \int f(z_i^E|\varphi_i, x_i, u_i, \theta_1) \cdot \\ &\quad f(\varphi_i) \cdot p(x_i|\varphi_i, u_i, \theta_1) d\varphi_i \end{aligned} \tag{8}$$

where φ_i denotes the value of the pseudo-random function $rand(\cdot)$ with probability density function $f(\varphi_i)$, u_i and x_i represent the original measurement and final output of the i-th sensor s_i respectively. z_i^E is conditionally independent of u_i, φ_i, and θ_1 when x_i is known, while x_i is conditionally independent of θ_1 when φ_i, u_i are known. Remember that only the sensors whose random values φ_i belong to $[\tau_2, \tau_1]$ or $[\tau_4, \tau_3]$ are activated in a duplex time, and we can derive $f(z_i^E|\theta_1)$ as follows:

$$\begin{aligned} f(z_i^E|\theta_1) = &\sum_{u_i} p(u_i|\theta_1) \sum_{x_i} f(z_i^E|x_i) \cdot \\ &(\int_{\tau_2}^{\tau_1} f(\varphi_i) \cdot p(x_i|\varphi_i, u_i) d\varphi_i + \\ &\int_{\tau_4}^{\tau_3} f(\varphi_i) \cdot p(x_i|\varphi_i, u_i) d\varphi_i) \end{aligned} \tag{9}$$

Since the EFC cannot be aware of φ_i and its probability density function $f(\varphi_i)$, it should take into account all the flipping cases: (1) $\varphi_i \in [\tau_2, \tau_1], u_i = \pm 1$, $x_i = u_i$; (2) $\varphi_i \in [\tau_4, \tau_3], u_i = \pm 1$, $x_i = -u_i$. To simplify the expression, let λ_1 and λ_2 represent $\int_{\tau_2}^{\tau_1} f(\varphi_i) \cdot p(x_i|\varphi_i, u_i) d\varphi_i$ and $\int_{\tau_4}^{\tau_3} f(\varphi_i) \cdot p(x_i|\varphi_i, u_i) d\varphi_i$ respectively, then we have:

$$\begin{aligned} f(z_i^E|\theta_1) = &f(z_i^E|x_i = -1)(1 - pd_i)\lambda_1 \\ &+ f(z_i^E|x_i = 1) pd_i \lambda_1 \\ &+ f(z_i^E|x_i = 1)(1 - pd_i)\lambda_2 \\ &+ f(z_i^E|x_i = -1) pd_i \lambda_2 \end{aligned} \tag{10}$$

If λ_1 and λ_2 satisfy $\lambda_1 = \lambda_2 = \lambda$, then,

$$f(z_i^E|\theta_1) = f(z_i^E|x_i = 1)\lambda + f(z_i^E|x_i = -1)\lambda \tag{11}$$

In the same manner, we can compute $f(z_i^E|\theta_0)$ as follows,

$$f(z_i^E|\theta_0) = f(z_i^E|x_i = 1)\lambda + f(z_i^E|x_i = -1)\lambda \tag{12}$$

From Equation (9) we can see that the AFC can easily ensure $\lambda_1 = \lambda_2$ as long as the width $\tau_1 - \tau_2 = \tau_3 - \tau_4$. And when $\lambda_1 = \lambda_2 = \lambda$, it means that the transmitted data from the flipping and non-flipping groups arrive at the EFC with the same size, and the proposed transmission scheme achieves $f(z_i^E|\theta_1) = f(z_i^E|\theta_0)$. Consequently, the EFC cannot make the final decision for the nature state θ_1 or θ_0, and it has to totally ignore the captured data.

4. Generalization to the Multiple Decisions

Besides the binary hypothesis testing problem discussed in previous sections, there are many WSN applications involved multiple states $\Theta = \{\theta_1, \theta_2 ... \theta_m\}$, and the sensor quantifications also have multiple scales $\Xi = \{v_1, v_2 ... v_n\}$. Usually each scale corresponds to an unique conditional probability in different state. Then the AFC needs to make a decision from the multiple candidate states while the EFC also tries to eavesdrop the natural state based on the captured data.

Suppose the received vector is $Z = [z_1, z_2 ... z_k], z_i \in \Xi$, according to Bayes rules [26], the state which has the maximum posterior probability is selected as the final decision.

$$\max_j \frac{P(Z|\theta_j)P(\theta_j)}{\sum_{i=1}^{n} P(Z|\theta_i)P(\theta_i)} \tag{13}$$

Without loss of generality, we assume that the prior probability of each state is equal, therefore Equation (13) can be simplified to:

$$\max_j \Pi_{i=1}^{k} P(z_i|\theta_j) \tag{14}$$

In order to prevent the EFC from forming the correct fusion result of Equation (14), the proposed SDPR is extended for the case of multiple scales (referred to SDPR_M). As shown in Figure 1, the measurement results are automatically mapped to other quantification scales based on a $1 \times n$ mapping matrix ϕ before sending to the AFC. Consequently, the AFC recovers the received data using an inverse mapping before data fusion. To make sure that the AFC and each sensor s_i keep the same mapping matrix ϕ_i, firstly, they select an initial $seed_i$ in a similar manner as described in Section 2 for a pseudo-random function $randInt$ (usually, $randInt(seed_i) = \lfloor min + rand(seed_i) * (max - min) \rfloor$), which uniformly generates random integer numbers in a given range [min, max] based on the $seed_i$, and $randInt$ always outputs the same sequence of numbers with the same $seed_i$. And then, in the t-th time-division sensor s_i generates a random sequence and forms ϕ_i using the following process:

- Step 1 initialize $\phi_i = null$, $R = null$, $j = 2$, n = number of scales;
- Step 2 $R(1) = randInt(seed_i)$, $temp = R(1)$;
- Step 3 if $j > n$ update; $seed_i = randInt(R(n))$, go to step 5, else $temp = randInt(temp)$, endif;
- Step 4 if $\forall k < j, R(k) \neq temp, R(j) = temp, j = j + 1$, endif, go to Step 3;
- Step 5 set $\phi_i = \{\Xi(R(1)), \Xi(R(2)) ... \Xi(R(n))\}$.

As shown in Equation (15), if the quantified result of sensor s_i is $u_i = v_j$, then the final output x_i is mapped to another scale $x_i = \phi_i(j)$ based on the mapping matrix.

$$x_i = \phi_i(j),\ if\ u_i = v_j, where\ \phi_i(j),\ v_i \in \Xi \tag{15}$$

After the data mapping, the output is sent to the AFC to form a decision using Equation (14). Then we analyze the security of the mapping schemes for multiple states and scales.

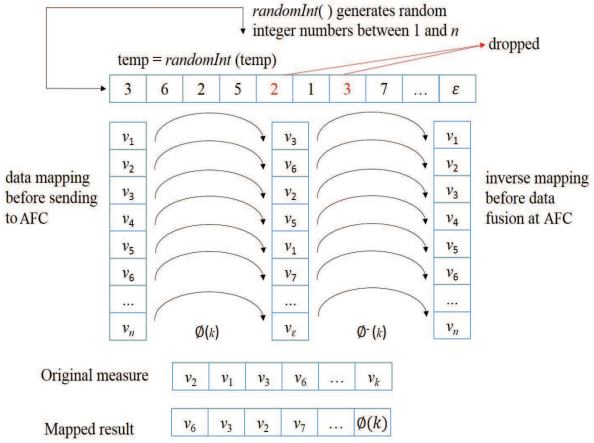


Figure 1. Procedure of pseudo-random encryption (SDPR_M).

Firstly, it is assumed that the outputs are transmitted through error-free channels, and the received data from the i-th sensor is $z_i = x_i$. Let us consider the probability of $P(z_i|\theta_j)$ in Equation (14).

$$\begin{aligned} P(z_i|\theta_j) &= \sum_{u_i} P(z_i, u_i|\theta_j) \\ &= \sum_{u_i} P(u_i|\theta_j) P(z_i|u_i, \theta_j) \\ &\overset{(a)}{=} \sum_{u_i} P(u_i|\theta_j) P(\Phi(u_i) = z_i|u_i, \theta_j) \\ &\overset{(b)}{=} \sum_{u_i} P(u_i|\theta_j) P(\Phi(u_i) = z_i) \end{aligned} \tag{16}$$

where $\Phi(.)$ is a mapping function defined on matrix ϕ_i, such that $\Phi(v_k) = v_t$ if $\phi_i(k) = v_t$ and $P(\Phi(u_i) = z_i)$ represents the probability of mapping scale u_i to z_i. (a) Follows from the fact that the final output of s_i is decided by the local measurement u_i and the mapping matrix; (b) follows the fact that the mapping probability of $P(\Phi(u_i) = z_i)$ is independent of u_i and θ_j when the mapping matrix Φ is known. Let $P(\Phi(u_i) = z_i) = \lambda_i$, then $P(z_i|\theta_j) = \sum_{u_i} P(u_i|\theta_j)\lambda_i$.

As the random function *randInt* generate integers uniformly between 1 and n, it can be assumed that the probability of mapping v_i to any v_j, $i, j = 1, 2...n$ is the same and $\lambda_1 = \lambda_2 ... = \lambda_n = \lambda$. Therefore, Equation (16) is simplified as:

$$\begin{aligned} P(z_i|\theta_j) &= \sum_{u_i} P(u_i|\theta_j)\lambda \\ &= \lambda \end{aligned} \tag{17}$$

Then we discuss a more complex scenario that the channel is interfered by noise, and for the i-th sensor, the received data is given by $z_i = h_i \cdot x_i + n_i$, where h_i is the channel gain and n_i is the signal noise. If the channel is not error-free, z_i may not equal x_i, and $f(z_i|\theta_j)$ in Equation (14) can be derived as follows:

$$\begin{aligned} P(z_i|\theta_j) &= \sum_{u_i}\sum_{x_i} P(z_i, x_i, u_i|\theta_j) \\ &= \sum_{u_i} P(u_i|\theta_j) \sum_{x_i} P(x_i|u_i, \theta_j) P(z_i|x_i, u_i, \theta_j) \\ &\overset{(a)}{=} \sum_{u_i} P(u_i|\theta_j) \sum_{x_i} P(x_i|u_i) P(z_i|x_i) \\ &\overset{(b)}{=} \sum_{u_i} P(u_i|\theta_j) \sum_{x_i} P(\Phi(u_i) = x_i) P(z_i|x_i) \end{aligned} \tag{18}$$

where (a) follows the fact that z_i is independent of u_i and θ_j when x_i is known, and (b) follows from the fact that the output of x_i is decided by u_i and local mapping matrix. As we have discussed in Equation (17), the probability of mapping u_i to any quantification scale can be the same, and $P(z_i|\theta_j)$ is further derived as:

$$P(z_i|\theta_j) = \sum_{u_i} P(u_i|\theta_j) \sum_{x_i} \lambda P(z_i|x_i) \tag{19}$$

The probability of $P(z_i|x_i)$ in Equation (19) is decided by the physical state of the wireless channel, including channel gains, noise level, etc. However, the physical channel can be assumed to be stable during one duplex time, and $\sum_{x_i} P(z_i|x_i)$ is a constant value P_i, which is not related to the candidate state θ_j. Finally we can have:

$$\begin{aligned} P(z_i|\theta_j) &= \sum_{u_i} P(u_i|\theta_j) \sum_{x_i} \lambda P(z_i|x_i) \\ &= \lambda \cdot P_i \end{aligned} \tag{20}$$

The AFC can recover the original quantifications based on reverse mapping of ϕ_i, but the EFC dose not know the mapping matrix, and has to fuse the raw data using Equation (14). According to Equations (17) and (19), the EFC for any input vector $Z = [z_1, z_2 ... z_k]$ and candidate state $\theta_j, j = 1, 2...m$, 1) in error-free channels $P(Z|\theta_j) = \Pi_{i=1}^{k} P(z_i|\theta_j) = \lambda^k$; 2) in noise channels $P(Z|\theta_j) = \Pi_{i=1}^{k} P(z_i|\theta_j) = \lambda^k \cdot \Pi_{t=1}^{k} P_t$. For both of the two cases, the conditional probability of the fusion rules in Equation (14) is all the same no matter what the candidate state θ_j is. Therefore, it will be nearly impossible for the EFC to generate correct decisions based on the captured sensor data.

5. Simulation Results

5.1. Experiments on Binary Hypothesis Testing

In this section, a group of simulations were carried out to test whether SDPR_B could prevent the EFC from eavesdropping the real state, while the AFC could correctly fuse the sensor outputs when there were only two candidate states.

The comparison objects include security fusion rules proposed in [18] (referred to Jeon), in which the sensors' binary decisions are flipped based on instantaneous channel gains in the main channels to the AFC. In addition, the performance of the Optimum-LLR proposed in [26], where the optimum LLR based fusion rule is derived without the presence of the EFC, is considered as the lower bounds of the error probability.

For ease of comparisons, we adopted a similar condition as in [18], the sensors were deployed into a star-like topology and the main channel gains were assumed to follow a Rayleigh distribution. Furthermore, the sensors have the same local detection performances $p_f = 0.2$, $p_d = 0.9$. The total error probability by $P_\varepsilon = \Delta(1 - P_d) + (1 - \Delta)P_f$ is taken as the criterion of fusion performance, where P_d

and P_f are the detection and false alarm probabilities at the fusion center, respectively, and $\Delta = 0.5$ is a weighting factor.

The first round of comparisons were carried out in ideal conditions that the channel gains strictly followed the Rayleigh distribution. The results are shown in Figures 2–4, which depict the weighted error probability of the AFC and EFC with different signal to noise rate (SNR) and sensor number.

From Figures 2 and 4, we can see that Optimum-LLR achieved the lowest error probability (lower bound). However, Optimum-LLR does not take any security mechanism against eavesdropping, thus the EFC can get the same low error rate as that achieved by the AFC. For SDPR_B, its error probability at the AFC was near to that of Optimum-LLR, and obviously better than Jeon. That is because the AFC could recover the flipped data using inverse flipping before data fusion. On the other hand, from Figures 3 and 5, we can see that the error probability of the EFC was always near 50% even with high SNR and many sensors, because it could not distinguish the flipped data from the original outputs, which completely interfered its data fusion. Therefore, in the ideal condition, both SDPR_B and Jeon achieved information-theoretic perfect secrecy.

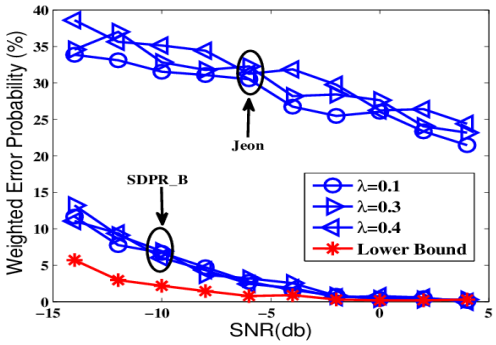


Figure 2. The error probabilities at the ally fusion center (AFC) as a function of signal to noise rate (SNR) in ideal channel conditions.

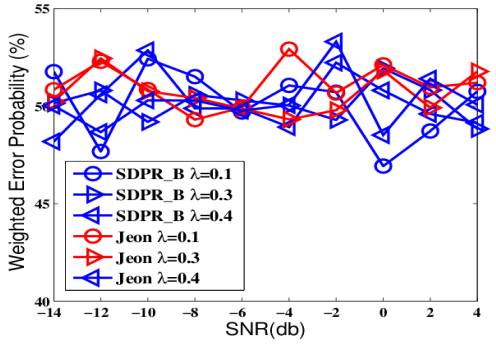


Figure 3. The error probabilities at the enemy fusion center (EFC) as a function of SNR in ideal channel conditions.

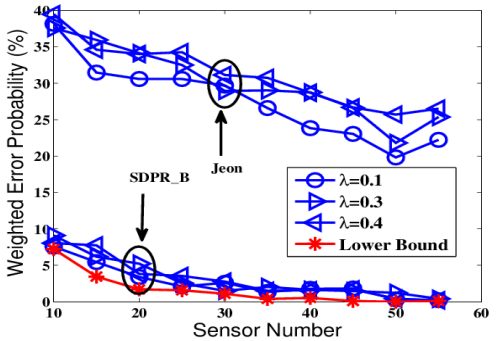


Figure 4. The error probabilities at the AFC with an increasing number of sensors in ideal channel conditions.

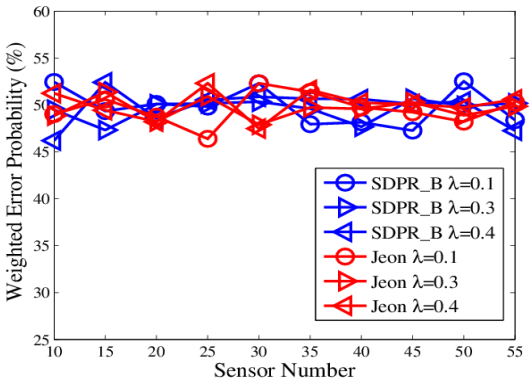


Figure 5. The error probabilities at the EFC with an increasing number of sensors in ideal channel conditions.

To further test the proposed schemes, we designed another group of comparisons that simulated a more realistic network environment where the channel gains continuously changed and we had no prior-knowledge of the probability distribution of the gains. The results are shown in Figures 6–9.

The figures depict that, at the AFC, the error probabilities of SDPR_B and Jeon decrease as the SNR and sensor number increased. Furthermore, for the EFC, the error probability of SDPR_B is always near to 50%, but compared with the results in the ideal condition, the error probability of Jeon apparently decreased. In the real situations, the channel gain, which affects the data flipping of Jeon, dynamically change due to power consumption, external disturbance, etc. Therefore, the numbers of flipping data and non-flipping data of Jeon are not equal any more, resulting in less disturbance to the EFC. As shown in Figures 7 and 9, the average error probability is reduced to 40%, which means the confidentiality of Jeon is compromised. In addition, the data flipping of SDPR_B is controlled by the pre-deployed random function, and when the AFC sets $\tau_1 - \tau_2 = \tau_3 - \tau_4$ in Equation (9), it ensures that the sizes of the flipped and non-flipped data in the EFC are the same and the two sets of data are self-contradictory, resulting in the error probability of the EFC to be always near 50%, which is a necessary condition of ideal confidentiality for binary hypothesis testing in WSN.

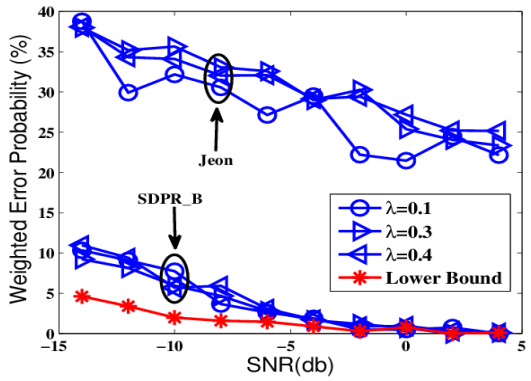


Figure 6. The error probabilities at the AFC as a function of SNR in time-varying channel conditions.

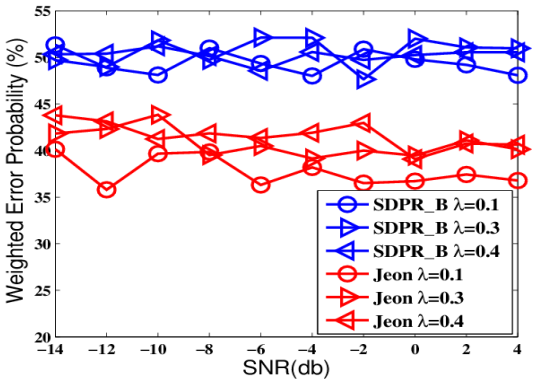


Figure 7. The error probabilities at the EFC as a function of SNR in time-varying channel conditions.

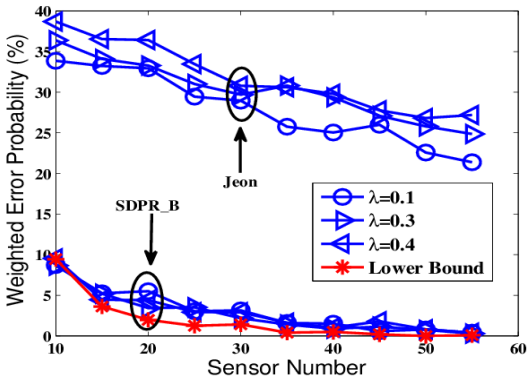


Figure 8. The error probabilities at the AFC with an increasing number of sensors in time-varying channel conditions.

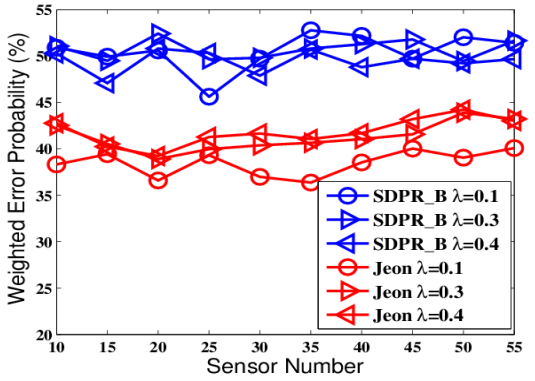


Figure 9. The error probabilities at the EFC with an increasing number of sensors in time-varying channel conditions.

5.2. Experiment on Multiple Scales

In this section, we tested the security schemes involved in multiple scales and states. Our goal is to demonstrate that SDPR_M proposed in the paper can ensure that the fusion result of EFC provides no more information than the priori-knowledge of the natural state, while the AFC can make a correct decision.

Our experiment simulated a common scenario in the industry area: n sensors quantize the vibration velocity of a mechanical system and periodically report the quantifications to the fusion center through WSN. The velocity scale ranges from $S = \{1, 2...100\}$, while the candidate states consists of $\theta_1, \theta_2, \theta_3, \theta_4$, corresponding to 'Idle', 'Busy', 'High Load', and 'Broken' respectively. As shown in Figure 10, the sensor quantifications in different state are assumed to follow Gauss distributions,

such that $p(x|\theta_1) \sim N(12.5, 8)$, $p(x|\theta_2) \sim N(37.5, 8)$, $p(x|\theta_3) \sim N(62.5, 8)$, $p(x|\theta_4) \sim N(87.5, 8)$ and, $p(x|\theta_i) = \int_{x-1}^{x} f(x)_{\theta_i} dx$. Without loss of generality, the priori-probability of each state is assumed to be the same.

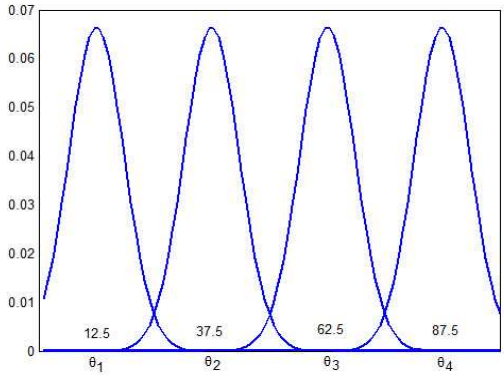


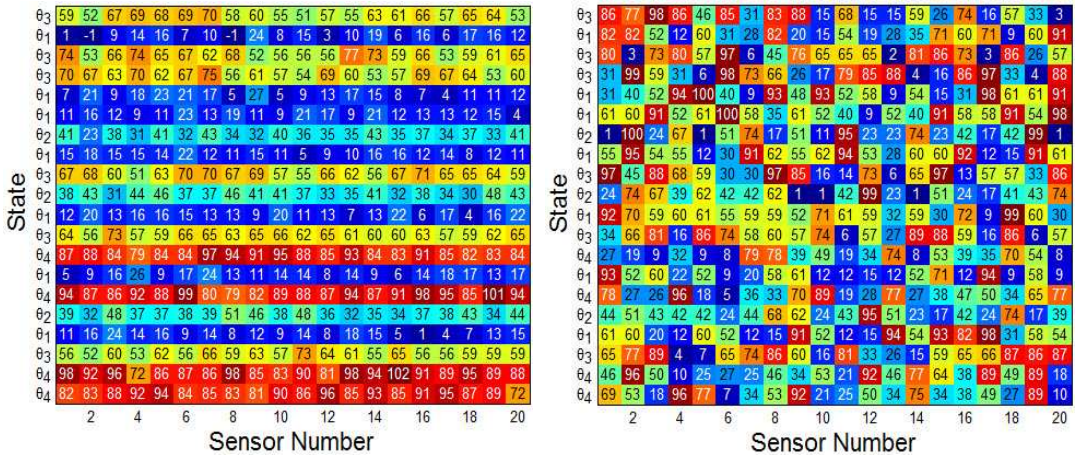
Figure 10. The distribution of sensor quantification in different states.

In each round of the simulation, one state $\theta_i, i \in \{1, 2, 3, 4\}$ is randomly selected and then the sensor measurements $[u_1, u_2 ... u_n]$, $u_i \in S$ are generated based on the probability density function $p(x|\theta_i)$. As described in Section 4, before data transmission, sensor $s_i, i = 1...n$ mapped its measurement u_i to another scale based on its mapping matrix to confuse the EFC. An example of data mapping is shown in Figure 11, each rows in (a) represent the original measurements of 20 sensors for a given state θ_i, while the rows in (b) are the corresponding mapped results which are totally different from the original ones. Note that the probability of mapping a scale to any one in S is the same. The theoretical number of possible mapping matrix is $100! = 9.3326215443944e + 157$ and it is nearly impossible for the EFC to recover the original quantifications through brute-force analyzing of the captured data (it will cost more than $3e + 132$ years, even if America's newest top supercomputer 'SUMMIT' is employed to the analysis).

For comparisons, we employed a native Bayes decision fusion rules [27] which is derived by using the principle of maximum posterior probability without data mapping.

In the first group of comparisons, 30% of the sensor outputs are interfered by White Gaussian noise. Figure 12 depicts the average error probability with different sensor number for 10 rounds of simulations. We see that in [27] as a lower error bound of the AFC, however the EFC has a similar performance too due to the lack of data encryption. Therefore, the EFC can easily eavesdrop the target state based on the captured data. The figure shows that the AFC of SDPR_M has a similar performance with that of [27]. Meanwhile, the EFC of SDPR_M has an apparently higher error probability and the corresponding true detection rate is near 25%, which is the priori-knowledge of the distribution of the 4 states. That is because in SDPR_M, the AFC and sensors deployed same pseudo-random functions and initial seeds, which ensures that they also have same mapping matrices and the AFC can recover the original quantifications through inverse mapping of the received data. In addition, the initial seeds are selected based on the instantaneous state information of the main channels (sensors to AFC). The EFC is unaware of the mapping matrix because of the independence between the main and eavesdropper channels. Therefore the data mapping only causes interference to the EFC's fusion results.

We can see similar results in Figure 13, which depicts the error probability with different SNRs. The results present a similar trend that the AFC of SDPR_M has a performance close to [27], while the EFC can still not get useful information about the state even with high SNR. The experiment confirmed the capabilities of SDPR_M in that it ensures the security of WSN's open transmissions while reserving the high performance of the AFC.



(**a**) sensor quantifications (**b**) Mapped results

Figure 11. An example of data mapping with 20 sensors and four candidate states θ_1 to θ_4.

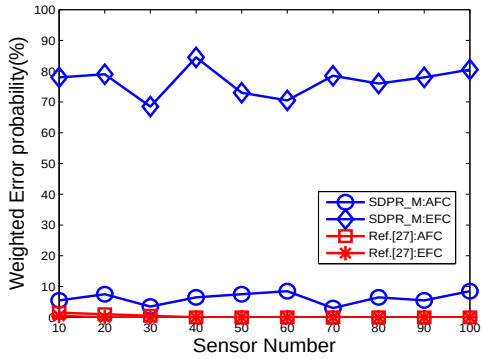


Figure 12. The error probabilities as a function of the number of deployed sensors for SNR = –5 dB.

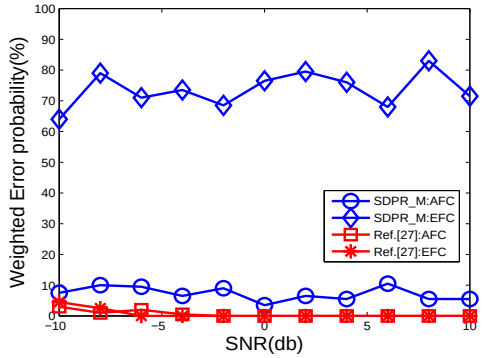


Figure 13. The error probabilities as a function of various SNRs for N = 20.

To see the performance superiority of the proposed scheme, another group of simulations were carried out to compare SDPR_M with the scheme proposed in [16] where the sensor outputs are randomly mapped to other scales using an optimized stochastic cipher matrix Φ. Its encryption is converted to solve the following optimization problem of Φ:

$$\Phi^* = \underset{\Phi}{arg\,max}\ J_A(\alpha_1\Phi\Delta||\alpha_2\Phi\Delta)$$
$$\text{subject to: 1) } 0 \le \phi_{ij}^* \le 1,\ 2)\ \Phi^*\mathbf{1}_{m\times 1} = \mathbf{1}_{m\times 1},\ \text{and}$$
$$3)\ (q_1\alpha_1 + q_2\alpha_2)\Phi^*\Delta = 1/m\mathbf{1}_{1\times m}$$

where J_A is the detection gain of the AFC, q_1 and q_2 represent priori-probability of two candidate states

θ_i, i = 1,2, $\alpha_i = [p(x_1|\theta_i), p(x_2|\theta_i)...p(x_m|\theta_i)]$ represents the post-probability vector of multiple scales in state θ_i, ϕ_{ij} is the probability of mapping a scale i to another level j, and Δ is the transition probability matrix. In addition, the native Bayes fusion rule [27] without data encryption is also compared and its results are taken as a lower bound of the error probability.

The proposed scheme in [16] is limited to two candidate states. For fair comparisons, it is assumed that there are only two states θ_1, θ_2, and the quantification scales $\{1, 2...100\}$ are assumed to follow Gauss distributions in different state: $p(x|\theta_1) \sim N(37.5, 12)$, and $p(x|\theta_2) \sim N(62.5, 12)$.

The comparison results are depicted in Figures 14 and 15. From the figures, we can see that both SDPR_M and the scheme proposed in [16] ensure that the error probability is around 50% at the EFC. Furthermore, the performance of SDPR_M at the AFC is quite similar to the lower bound shown as the red line. However, the performance at the AFC of [16] obviously deteriorated, especially when the sensor number is less than 20, the error probability is higher than 20%. That is because the optimization problem of the stochastic cipher matrix Φ in [16] is NP-hard, and usually its heuristical solutions are suboptimal, which results in the non-trivial degeneration of performance across a wide range of sensor numbers (SNR values). Although the scheme in [16] achieved good confidentiality, it sacrificed the performance of the AFC. Whereas SDPR_M realized a similar security level that maintained data availability of the AFC. In addition, it should be noted that in many applications, both the data confidentiality and the data reliability should be taken into consideration. Therefore, from this point of view our method can be more applicable.

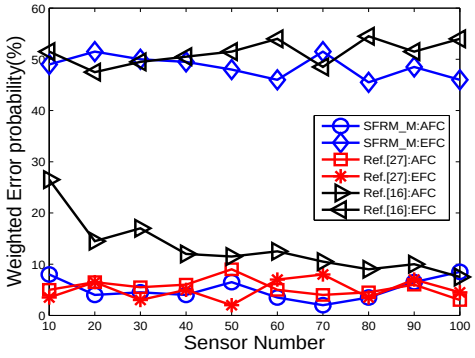


Figure 14. The error probabilities as a function of the number of deployed sensors for SNR = –5 dB.

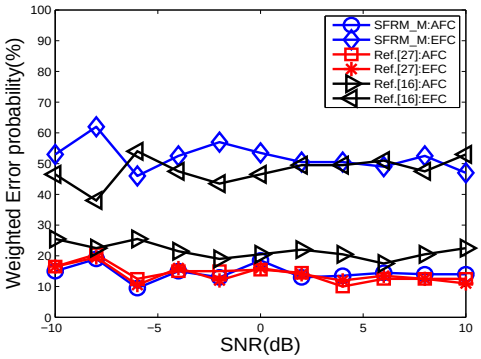


Figure 15. The error probabilities as a function of various SNRs for N = 20.

6. Conclusions

In the paper, a lightweight scheme was proposed to protect data confidentiality in a distributed sensor network. The data was supposed to be transmitted over open and insecure channels with the presence of an enemy fusion center EFC, which tried to gather all the transmitted data to form its own decision regarding the state of nature. To prevent the EFC from eavesdropping, the security scheme

was designed by exploiting randomness of data flipping (mapping). The main idea was that the activated sensors change their quantized outputs in a random way based on pseudo-random functions known by sensors and the AFC. Thus the EFC who captured the sensor data over the open channels failed to perform data decryption since it could not distinguish the original output from the flipped data. The theory analysis and experimental results demonstrated that the AFC could appropriately decrypt the data, but for the EFC, even with high SNR and large number of sensors, it still could not make right decisions on the state.

We claim that due to the simplicity and low complexity, the proposed solution could be deployed in many resource-limited applications of WSN, including natural disaster monitoring, battlefield situation awareness, remote control of unmanned aerial vehicle, etc. Furthermore, the new pseudo-encryption model opens several future research lines. Being a generalization of data-flipping encryption, we can expect stochastic data flipping to allow better confidentiality while reserving data utility. Exploring the capability limit of pseudo random flipping for data confidentiality is another possible follow-up of this article. Finally, it should be noted that our method is preferred in the condition that the number of sensor quantification scales were stable and known in advance, and how to improve the proposed method for the unpredictable and unstable environment of WSN will be taken into consideration in our future work.

Author Contributions: Conceptualization, W.C. and L.L.; methodology, W.C.; software, W.C., Y.L.; validation, Y.L.; formal analysis W.C. and T.L.; investigation, W.C. and T.L.; data curation, Y.L.; writing—original draft preparation, L.L.; writing—review and editing, W.C.

Funding: This work was supported by the National Key Research and Development Program of China (Grant no. 2016YFB0800605 and 2016YFB0800604) and the Natural Science Foundation of China (Grant no. 61872255, U1736212 and 61572334).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reyes-Menendez, A.; Palos-Sanchez, P.; Saura, J.R.; Martin-Velicia, F. Understanding the Influence of Wireless Communications and Wi-Fi Access on Customer Loyalty: A Behavioral Model System. *Wirel. Commun. Mob. Com.* **2018**, *2018*, 1–16. [CrossRef]
2. Saura, J.R.; Reyes-Menendez, A.; Palos-Sanchez, P. Mapping multispectral Digital Images using a Cloud Computing software: Applications from UAV images. *Heliyon* **2019**, *5*, 1–22. [CrossRef] [PubMed]
3. Shigen, S.; Hongjie, L.; Risheng, H.; Athanasios, V.V.; Yihan, W.; Qiying, C. Differential game-based strategies for preventing malware propagation in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1962–1973. [CrossRef]
4. Murat, D.; Yunus, O.; Cevat, B. Fire detection systems in wireless sensor networks. *Procedia Soc. Behav. Sci.* **2015**, *195*, 1846–1850.
5. Huang, D.J.; Teng, W.C. A defense against clock skew replication attacks in wireless sensor networksoriginal research article. *J. Netw. Comput. Appl.* **2015**, *39*, 26–37. [CrossRef]
6. He, D.; Chan, S.; Guizani, M. Accountable and privacy–enhanced access control in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2015**, *14*, 389–398. [CrossRef]
7. Soltanmohammadi, E.; Oroojiand, M.; Naraghi–Pour, M. Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 205–215. [CrossRef]
8. Moosavi, H.; Bui, F.M. A game–theoretic framework for robust optimal intrusion detection in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1367–1379. [CrossRef]
9. Uluagac, A. S.; Beyah, R.A.; Copeland, J.A. Secure source–based loose synchronization (sobas) for wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 803–813. [CrossRef]
10. Yanrong, L.; Lixiang, L.; Haipeng, P.; Yixian, Y. An Energy Efficient Mutual Authentication and Key Agreement Scheme Preserving Anonymity for Wireless Sensor Networks. *Sensors* **2016**, *16*, 837–8576.
11. Dawei, Z.; Haipeng, P.; Lixiang, L.; Yixian, Y. A secure and effective anonymous authentication scheme for roaming service in global mobility networks. *Wirel. Pers. Commun.* **2014**, *78*, 247–269.

12. Akyildiz, I.F.; Su, Y.S.W.; Cayirci, E. Wireless sensor networks: A survey. *Comput. Netw.* **2002**, *38*, 393–442. [CrossRef]
13. Incebacak, D.; Bicakci, K.; Tavli, B. Evaluating energy cost of route diversity for security in wireless sensor networksoriginal. *Comput. Standa. Interfaces* **2015**, *39*, 44–57. [CrossRef]
14. Aysal, T.C.; Barner, K. Sensor data cryptography in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2008**, *3*, 273–289. [CrossRef]
15. Pour, M.N.; Nadendla, V. Secure detection in wireless sensor networks using a simple encryption method. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Cancun, Mexico, 28–31 March 2011; pp. 114–119.
16. Soosahabi, R.; Naraghi–Pour, M.; Perkins, D.; Bayoumi, M.A. Optimal probabilistic encryption for secure detection in wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 375–385. [CrossRef]
17. Marano, S.; Matta, V.; Willett, P.K. Distributed detection with censoring sensors under physical layer secrecy. *IEEE Trans. Signal Process.* **2009**, *57*, 1976–1986. [CrossRef]
18. Jeon, H.; Choit, J.; McLaughlin, S.W.; Ha, J. Channel aware encryption and decision fusion for wireless sensor networks. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 619–625. [CrossRef]
19. Matsumoto, M.; Nishimura, T. Mersenne Twister: A 623–Dimensionally Equidistributed Uniform Pseudo–Random Number Generator. *ACM Trans. Model. Comput. Simul.* **1998**, *8*, 3–30. [CrossRef]
20. Naor, M. Number–Theoretic Constructions of Efficient Pseudo–Random Functions. *J. ACM* **2004**, *51*, 231–262. [CrossRef]
21. Wichmann, B.A.; Hill, I.D. An Efficient and Portable Pseudo–random Number Generator. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1982**, *31*, 188–190.
22. Liu, Y.; Draper, S.C.; Sayeed, A.M. Exploiting channel diversity in secret key generation from multipath fading randomness. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1484–1497. [CrossRef]
23. Wallace, J.; Sharma, R. Automatic secret keys from reciprocal MIMO wireless channels: Measurement and analysis. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 381–392.
24. Ye, C.; Mathur, S.; Reznik, A.; Shah, Y.; Trappe, W.; Mandayam, N. Information–theoretically secret key generation for fading wireless channels. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 240–254.
25. Varshney, P. *Distributed Detection and Data Fusion*, 1st ed.; Springer: Berlin, Germany, 2005; pp. 180–189.
26. Chen, B.; Jiang, R.; Kasetkasem, T.; Varshney, P.K. Channel aware decision fusion in wireless sensor networks. *IEEE Trans. Signal Process.* **2006**, *52*, 3454–3458. [CrossRef]
27. Guerriero, M.; Svensson, L.; Willett, P. Bayesian Data Fusion for Distributed Target Detection in Sensor Networks. *IEEE Trans. Signal Process.* **2010**, *58*, 3417–3421. [CrossRef]