sensors

MDPI

Article

A Probabilistically Weakly Secure Network Coding Scheme in Multipath Routing for WSNs

Xiang Liu *, Jie Huang and Xiang Gao

School of Information Science and Engineering, Southeast University, Nanjing 210096, China; jhuang@seu.edu.cn (J.H.); xianggao@seu.edu.cn (X.G.)

* Correspondence: krys_liu@seu.edu.cn; Tel.: +86-158-5187-8909

Academic Editor: Kemal Akkaya

Received: 20 January 2017; Accepted: 13 May 2017; Published: 16 May 2017

Abstract: In wireless sensor networks, nodes are mostly deployed in unsupervised areas and are vulnerable to a variety of attacks. Therefore, data security is a vital aspect to be considered. However, due to the limited computation capability and memory of sensor nodes, it is difficult to perform the complex encryption algorithm, as well as the key distribution and management algorithm. Toward this end, a low-complexity algorithm for security in wireless sensor networks is of significant importance. In this article, a weakly secure network coding based multipath routing scheme is proposed, which can guarantee the data confidentiality in transmission probabilistically, and can improve the energy efficiency in the meantime. Then the simulations of the probability of transmission being secure are performed. The results show that with the increase of the number of hops k, the probability of transmission being secure suffers from a rapid decrease. On the contrary, with the increase of multicast capacity h it undergoes a slight growth. Therefore, the weak security can be achieved with probability approaching 1 by limiting the number of hops and increasing the multicast capacity. Meanwhile, the simulations of energy consumption are performed and the comparison between the energy consumption of the scheme in this article and the multipath routing scheme without network coding is conducted. The results show that by employing network coding, the scheme in this article can improve the energy efficiency, and the more packets transmitted, the more energy consumption can be reduced.

Keywords: probability of transmission being secure; weakly secure network coding; multipath routing; energy efficiency

1. Introduction

Due to the complex working environment, wireless sensor networks (WSNs) can suffer from a variety of attacks. Therefore, transmission security, including data confidentiality, data integrity, and data availability, is a vital aspect to be considered [1–4]. Existing researches on security of WSNs are mostly based on encryption/decryption. In [5], a symmetric encryption algorithm is proposed, which is an amalgamation of two different encryption algorithms in randomized method. In [6], the authors analyzed the security challenges in WSNs and smart home systems, then proposed a security evaluation technique based on attack graph generation. SNEP protocol is one of most maturely applied security protocols in WSNs [7]. Since the communication among the nodes requires the involvement of the base station, the SNEP protocol is of rather low efficiency and is not applicable in large-scale networks. Since the low-cost wireless sensors which are battery-powered have limited computational capability and memory, it is difficult to perform complicated encryption algorithms in WSNs. Moreover, achieving key distribution and key management is also a great challenge in large-scale WSNs. In [8], a random key distribution based key management protocol is proposed. In this protocol, a key pool with size of S is established firstly and each node in the network stores m

keys of the key pool. Once any two nodes in the network have the same key, a secure channel can be established between them. However, the biggest problem is that if there are no identical keys between a pair of nodes, the secure data transmission between them cannot be completed. To overcome the disadvantages in the aforementioned techniques, a weakly-secure network coding based multipath routing scheme is proposed in this article to guarantee the confidentiality of messages. The superiority of the scheme in this article over encryption is that it gets rid of the complex encryption algorithm and key distribution/management, so it can greatly reduce the computation overhead.

Compared with Shannon security [9–13], the requirement of weak security in this article is not information-theoretically perfect [14]. While the Shannon security requires the attacker cannot get any information about the source message, in the paradigm of weak security, the attacker cannot get any useful information about the source message, which means the attacker cannot decode any part of the source message correctly. As a result, weak security can bring higher transmission capacity than Shannon security. Therefore, the weakly secure network coding based scheme is adopted in this article to improve the transmission efficiency and reduce the transmission overhead.

The main contributions of this article can be listed as follows. Firstly, the relationship between the probability of transmission being secure and some network parameters is presented through mathematical derivations and simulations. Then it is shown that by setting the parameters appropriately, the weak security can be achieved with the probability approaching 1. Secondly, by employing network coding in the intermediate nodes and calculating the least number of communication nodes to satisfy the network requirements, the method proposed in this article can largely reduce the transmission overhead compared with the multipath routing scheme without network coding.

The remainder of this article is organized as follows. In Section 2, the network model is introduced, including the adversary model and the algorithms to get the number of paths as well as the least number of communication nodes. Then the simulation results about these algorithms are presented as well. In Section 3, the pre-coding scheme and random network coding scheme are presented. In Section 4, the security analysis is conducted, and the probability of the transmission being weakly secure is derived. In Section 5, the analysis of energy consumption of two different schemes is performed, namely, the network coding based multipath routing scheme and the multipath routing scheme without network coding. In Section 6, the simulations based on the above analysis are performed and the results are presented. In Section 7, the conclusions based on the simulations are presented.

2. Network Model

2.1. Adversary Model

In this article, the communication network can be described as a directed acyclic graph $G = (V, E)$, where V is the set of nodes and E is the set of links. For each link $e \in E$, we define $tail(e)$ and $head(e)$ as the tail and head of e, respectively. In the set of nodes V, it is denoted by s the source node, and by T the sink node, which is the base station in the practical WSNs. For each node $v \in V$, let $\mathrm{Out}(v)$ and $\mathrm{In}(v)$ denote the set of outgoing channels and incoming channels of v, respectively. That is, $\mathrm{Out}(v) = \{(v, u) : (v, u) \in E\}$ and $\mathrm{In}(v) = \{(u, v) : (u, v) \in E\}$. For the eavesdropper, let V_{eav} and E_{eav} denote the set of nodes and channels being eavesdropped, respectively. The multi-cast capacity H is the minimum number of edges in any cut between the source node and sink node. Each channel $e \in E$ contains a message packet, whose elements are selected from a finite field F_q, where q is the size of the finite field. For the source node, we introduce H artificial channels which carry the H source packets that the source node transmits to the base station.

It is supposed that there exist some nodes that are randomly deployed in a specified district, and the number of nodes is denoted by N. After the routing procedure is completed, there exist N_c intermediate nodes that are involved in transmission. It is assumed that there is an adversary which is randomly located in the district, and for each intermediate node, there is a probability that it may be attacked by the adversary and the figure of the probability is up to the range that the adversary can

control. Once the intermediate node is attacked, it can be controlled by the adversary and the messages it receives and transmits can be observed completely by the adversary. For each transmission, whether an individual node is controlled is independent. Notably, the source node and the base station cannot be attacked, otherwise the malicious node can get the message without any loss.

2.2. Calculation of Number of Paths

Suppose that the successful delivery ratio (SDR) is denoted by r. For every single link between any two nodes, the link failure probability is e. In addition, the average number of hops of the paths from the source node to the base station is k, the desired multicast capacity is h. To simplify the question, it is assumed that the number of hops of each path exactly equals to k. Transmission of packets on each hop can be regarded as dependent event, hence, for each path, the probability of successfully delivering one packet is

$$p_k = (1-e)^k \tag{1}$$

For one transmission, only if the number of successful paths is at least h do we call it a successful transmission. Since the desired multicast capacity is h, it requires $H \geq h$ paths to guarantee the expected successful delivery ratio R. Under the condition that link failure probability is e and the number of hops is k, let $H_{h,e,k,R}$ denote the least number of paths to implement to achieve capacity of h and SDR of R. Among the H paths, the number of successful paths should be at least h to guarantee the correct recovery of original message at the base station. Hence, the SDR can be represented as

$$r = \sum_{i=h}^{H} \mathrm{C}_H^i p_k^i (1-p_k)^{H-i} \tag{2}$$

where C_H^i is the binomial coefficient, defined as $\mathrm{C}_H^i = \frac{H!}{(H-i)!(i)!}$.

In this article, it is essential to determine the least number of H, referring to as $H_{h,e,k,R}$, to guarantee that the successful delivery ratio satisfies $r \geq R$, i.e.

$$\sum_{i=h}^{H_{h,e,k,R}} \mathrm{C}_{H_{h,e,k,R}}^i p_k^i (1-p_k)^{H_{h,e,k,R}-i} \geq R \tag{3}$$

In practice, it can be impossible to find the analytical solution of formula (3). So we need to find the numeric solution by the iteration algorithm. However, when h gets larger, it can be of great computational complexity to perform the iteration algorithm to get H. Consequently, it is necessary to adopt another algorithm with light complexity to get the approximated solution of (3), which is also presented in [15,16]. Let H_s denote the number of the successful paths in the H paths, then H_s satisfies the Binomial distribution B(H,p_k), and the mean value and the variance of H_s can be written as

$$\mu = E(H_s) = Hp_k = H(1-e)^k \tag{4}$$

$$\sigma^2 = Hp_k(1-p_k) = H(1-e)^k(1-(1-e)^k) \tag{5}$$

The successful delivery ratio r can be rewritten as

$$r = \mathrm{P}(H_s \geq h) \tag{6}$$

Thus, the question can be described as finding the least H that guarantees $\mathrm{P}(H_s \geq h) \geq R$. According to the central-limit theorem, the Binomial distribution can be regarded as Normal distribution approximately, i.e.,

$$H_s \sim N(\mu, \sigma^2) \tag{7}$$

Let

$$H_s^* = \frac{H_s - \mu}{\sigma} \tag{8}$$

then H_s^* satisfies the standard normal distribution, i.e.

$$H_s^* \sim N(0,1) \tag{9}$$

In the standard normal distribution, for the given R, the value of x_R that satisfies $\mathrm{P}(H_s^* \geq x_R) \geq R$ can be obtained from the probability density function of standard normal distribution, which means

$$h = x_R\sigma + \mu \tag{10}$$

After that, the value of H can be calculated through the Equation (11).

$$h = x_R\sigma + \mu = x_R\sqrt{H(1-e)^k(1-(1-e)^k)} + H(1-e)^k \tag{11}$$

Hereto, given the desired successful delivery ration R, the link failure probability e, the average number of hops k, and the expected multi-cast capacity h, the least number of paths can be calculated according to Algorithm 1:

Algorithm 1 approximated algorithm of Calculating H

1: Initializing with R, e, k, and h
2: Calculating x_R
3: Calculating H

According to the aforementioned algorithms, the simulations with different initializing values are performed, and the results are shown in Figures 1–3:

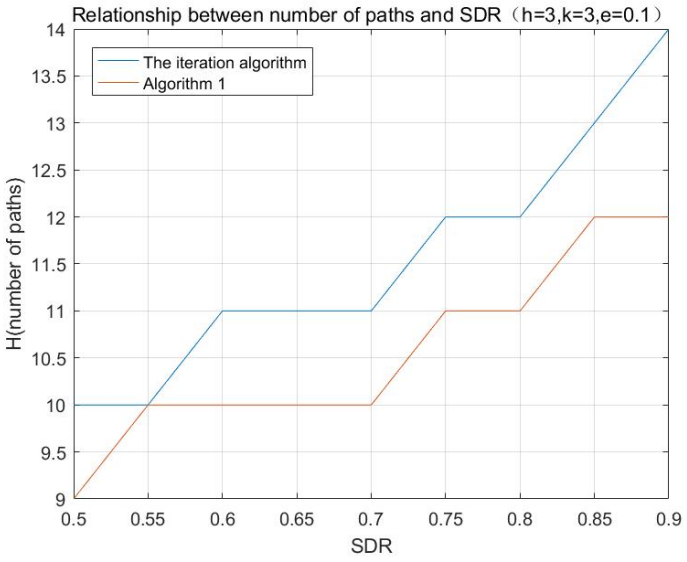


Figure 1. Relationship between number of paths and SDR.

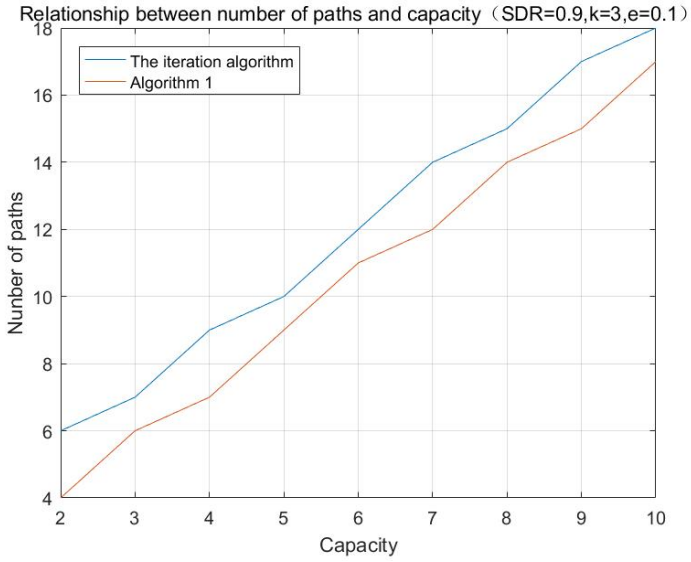


Figure 2. Relationship between number of paths and multi-cast capacity.

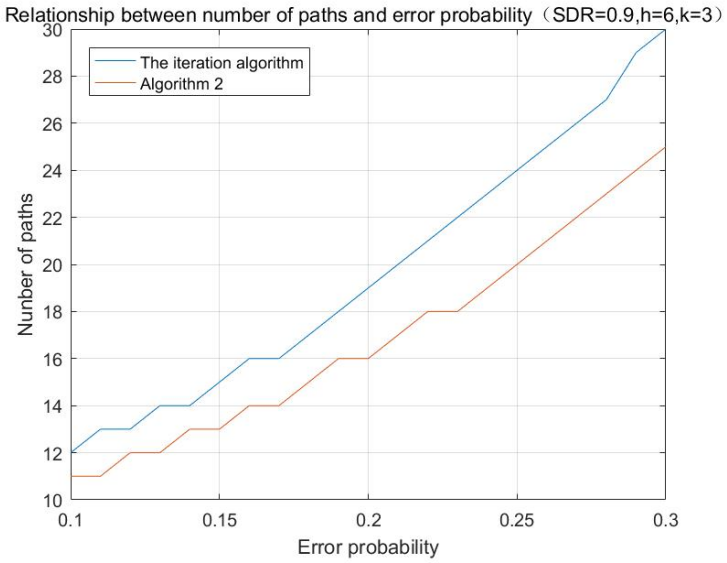


Figure 3. Relationship between number of paths and error probability.

2.3. Least Number of Communication Nodes

After getting the least number of paths H according to the algorithms mentioned above, the routing procedure can be proceeded to establish those H paths to build the communication network. For the sake of energy efficiency and transmission security, it is essential to involve as few nodes as possible in communication under the network conditions. Given the number of hops of each path, and the number of paths, we define N_l as the least number of communication nodes, a.k.a the least number of nodes that need to be involved in communication to satisfy the conditions.

Algorithm 2 The algorithm to calculate the least number of communication nodes

Initiate with the parameters H, k
Create the first path and set $num\_node \leftarrow k-1, s\_path \leftarrow 1$
while $s\_path < H$ **do**
    \\ **Create a new path**
    $p\_flag \leftarrow 0$
    while $p\_flag == 0$ **do**
        $c\_hop \leftarrow 0$
        while $c\_hop < k-1$ **do**
            for $i = 1 : num\_node$ **do**
                $new\_node \leftarrow 1$
                if the $i-th$ node is the current hop || exists parallel channels || exists a loop **then**
                    Continue
                else
                    pick the $i-th$ node as the next hop
                    $new\_node \leftarrow 0$
                    break
                end if
            end for
            if $new\_node == 1$ **then**
                pick up a new node as the next hop
                $num\_node \leftarrow num\_node + 1$
            end if
            $c\_hop \leftarrow c\_hop + 1$
        end while
        $p\_flag \leftarrow 1$
        $s\_path \leftarrow s\_path + 1$
    end while
end while
$N_l \leftarrow num\_node$

In Algorithm 2, the first path with k hops is established initially. To do so, there should be $k-1$ intermediate nodes involved. Then, it can be proceeded to establish the rest $H-1$ paths. What is notable is that the network topology is node-braided and edge-disjoint, which means that between any two paths, there may exist common nodes but cannot exist any common edges. Therefore, when executing the routing, it should be firstly determined whether the existing intermediate nodes can be used as the next hop of the current path, if not, then it is necessary to introduce a new node. Meanwhile, it is crucial that there cannot exist parallel links between any two nodes, either loops in the whole network.

According to Algorithm 2, some simulations are performed and the results are shown in Figure 4.

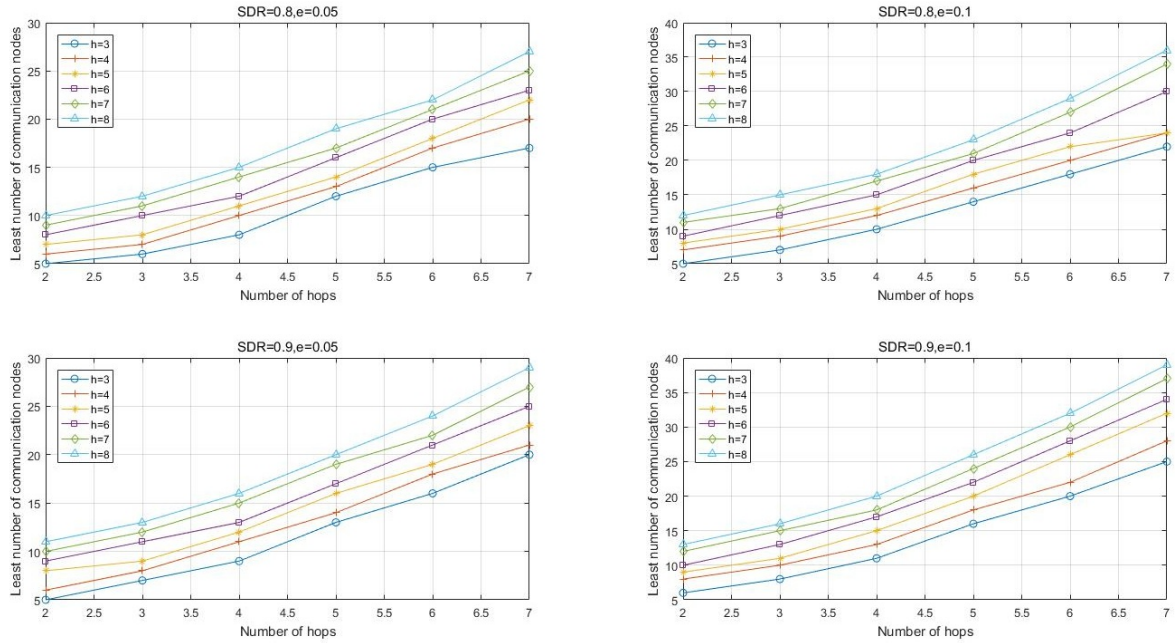


Figure 4. The least number of communication nodes.

3. Weakly Secure Network Coding

In this section, a packet format which can get rid of centralized knowledge of network topology is proposed, which is shown in Figure 5.

Source ID	Dest ID	Generation ID	Packet ID	Coding Vector	Coded Data

Figure 5. Packet Format.

Here the 'Source ID' is the ID of the transmitting node, the 'Dest ID' is the ID of the receiving node, and the 'Generation ID' is the identifier of a generation. The source node categories all the source packets into some groups and each group includes h source packets. Such a group is called one generation. In each generation, the h packets are assigned with a packet ID ranging from 1 to h respectively. In addition, the source sends one generation in each transmission. The 'Coding Vector' is the vector of combination coefficients of the coded packet.

To achieve weak security, the source node needs to encode the data before transmitting it and this process is called pre-coding. The pre-coding algorithm in this article is generated from the algorithm in [17]. Compared with the algorithm in [17], the algorithm in this article introduces more non-linear property to the coded packets.

According to the hypothesis, the source node can transmit h packets in one transmission. Therefore, without loss of generality, the source message can be denoted as

$$\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_h]^T \tag{12}$$

Then the pre-coded message can be written as

$$\mathbf{M}' = [\mathbf{m}_1', \mathbf{m}_2', \cdots, \mathbf{m}_h']^T \tag{13}$$

where

$$\begin{cases} \mathbf{m}_1^{'} = f(\mathbf{m}_1) + \mathbf{m}_2 \\ \mathbf{m}_2^{'} = f(\mathbf{m}_1 + \mathbf{m}_2) + \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_{n-1}^{'} = f(\mathbf{m}_1 + \mathbf{m}_2 + \cdots + \mathbf{m}_{n-1}) + \mathbf{m}_n \\ \mathbf{m}_n^{'} = \mathbf{m}_1 + f(\mathbf{m}_1^{'} + \cdots + \mathbf{m}_{n-1}^{'}) \end{cases} \tag{14}$$

The function f is a permutation function and both its input and output are vectors which consist of elements in the finite field. Note that the construction of function f is public to all nodes, even including the adversary.

After the pre-coding procedure is completed, the source node applies the generating matrix $\mathbf{G}$ to the coded message to generate H packets to transmit along the H outgoing channels of the source node s,

$$\mathbf{X} = \mathbf{G}\mathbf{M}^{'} = [\mathbf{x}_1, \cdots, \mathbf{x}_H]^T \tag{15}$$

where $\mathbf{G}$ is a H-by-h matrix whose elements are chosen randomly from the finite field.

For the sink node, the received packets can be denoted as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m]^T = \mathbf{C}\mathbf{M}^{'}$, where $\mathbf{C}$ is the coding matrix of $\mathbf{Y}$, and the $i-th$ row of $\mathbf{C}$ is the coding vector of packet $\mathbf{y}_i$. Then by using Gaussian elimination method, $\mathbf{M}^{'}$ can be calculated. After that, $\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_h$ can be calculated iteratively according to formula (16):

$$\begin{cases} \mathbf{m}_1 = \mathbf{m}_n^{'} - f(\mathbf{m}_1^{'} + \cdots + \mathbf{m}_{n-1}^{'}) \\ \mathbf{m}_2 = \mathbf{m}_1^{'} - f(\mathbf{m}_1) \\ \vdots \\ \mathbf{m}_{n-1} = \mathbf{m}_{n-2}^{'} - f(\mathbf{m}_1 + \cdots + \mathbf{m}_{n-2}) \\ \mathbf{m}_n = \mathbf{m}_{n-1}^{'} - f(\mathbf{m}_1 + \mathbf{m}_2 + \cdots + \mathbf{m}_{n-1}) \end{cases} \tag{16}$$

In this way, the sink node can decode all the h packets in one generation.

4. Security Analysis

Since each node is deployed randomly in the district and the adversary is randomly located in the district, then for every single node, the probability of being located in the overhearing zone (being controlled by the adversary) is

$$p_o = \frac{S_{overhear}}{S_{total}} \tag{17}$$

where S_{total} is the size of the whole district wherein the wireless sensors are deployed, and $S_{overhear}$ is range that the adversary can control. Once a node is located in that range, it will be controlled by the adversary. Therefore, in the whole district, the number of nodes which are overheard N_o satisfies the binomial distribution

$$N_o \sim B(N, p_o) \tag{18}$$

Given the average number of hops of each path and the number of paths, define N_{co} as the number of communication nodes which be overheard by the attacker.

Theorem 1. *The probability of $N_{co} = m$ for all integers $0 \leq m \leq N_l$ can be denoted as*

$$p(N_{co} = m) = \mathrm{C}_N^{N_{co}} (p_c p_o)^{N_{co}} (1 - p_c p_o)^{N - N_{co}} \tag{19}$$

where $p_c = \frac{N_l}{N}$ is the probability of a node being involved in the communication.

Proof.

$$\begin{aligned} p(N_{co}=m) &= \sum_{k=m}^{N} p(N_o=k)p(N_{co}=m|N_o=k) \\ &= \sum_{k=m}^{N} \mathrm{C}_k^{N_{co}} p_c^{N_{co}}(1-p_c)^{k-N_{co}}(1-p_c)^{N-k}\mathrm{C}_N^k p_o^k(1-p_o)^{N-k} \\ &= \mathrm{C}_N^{N_{co}}(p_c p_o)^{N_{co}}(1-p_c p_o)^{N-N_{co}} \end{aligned} \tag{20}$$

□

Theorem 2. *Let E_o denote the number of channels that being overheard by the attacker, and E_{vo} denote the number of valid channels that being overheard, then for all integers $0 \le c \le kH$*

$$\begin{aligned} p(E_{vo}=c) &= \textstyle\sum_{i=c}^{kH} p(E_o=i)p(E_{vo}=c|E_o=i) \\ &= \textstyle\sum_{i=c}^{kH} p(E_o=i)(1-e)^c e^{(i-c)} \end{aligned} \tag{21}$$

And $p(E_o=i)=\sum_{j=0}^{N_l} p(N_{co}=j)p(E_o=i|N_{co}=j)$, so equation (27) can be rewritten as

$$p(E_{vo}=c)=\sum_{i=c}^{kH}\sum_{j=0}^{N_l} p(N_{co}=j)p(E_o=i|N_{co}=j)(1-e)^c e^{(i-c)} \tag{22}$$

Let $\mathbf{\Gamma_w}$ be the overhearing matrix, referring to as the matrix that consists of the coding vectors of the valid channels being overheard.

Theorem 3. *The attacker cannot get any useful information of the original messages given that $R(\mathbf{\Gamma_w}) < h$, where $R(\mathbf{\Gamma_w})$ is the rank of matrix $\mathbf{\Gamma_w}$, i.e., $\mathbf{\Gamma_w}$ is not a full-rank matrix.*

Proof. Let $\mathbf{X}=(x_1,x_2,\cdots,x_h)^T$ be the original message that is sent over the network. Then after the pre-coding on the source node, the input message can be written as $\mathbf{X}^{'}=[x_1^{'},x_2^{'},\cdots,x_h^{'}]$, i.e., the source transmits $\mathbf{X}^{'}$ instead of $\mathbf{X}$. Since a linear random network code is used ,the message on each channel e_j can be written as $\mathbf{\Gamma}_{e_j}\mathbf{X}^{'}$. The message obtained by the attacker is $\mathbf{W}$=$\mathbf{\Gamma_w}\mathbf{X}^{'}$. Since $R(\mathbf{\Gamma_w}) < h$, which means the adversary can obtain at most $h-1$ linearly independent equations, which means that it cannot resolve for all the packets in $\mathbf{X}^{'}$. Then the attacker cannot solve any packets through formula (16). Hence, we have $I(x_i;\mathbf{B})=I(x_i;\mathbf{W})=0$ and by so we can achieve weak security. □

Let p_e be the probability of transmission being insecure, which means

$$p_e = p(R(\mathbf{\Gamma_w})=h) \tag{23}$$

Then $p_s=1-p_e$ is the probability of transmission being secure.

Theorem 4.

$$\begin{aligned} p_e &= \textstyle\sum_{m=0}^{kH} p(E_{vo}=m)p(R(\mathbf{\Gamma_w})=h|E_{vo}=m) \\ &= \textstyle\sum_{m=h}^{kH} p(E_{vo}=m)p(R(\mathbf{\Gamma_w})=h|E_{vo}=m) \end{aligned} \tag{24}$$

Proof. When $E_{vo} \le h-1$, it is obvious that the rank of overhearing matrix $\mathbf{\Gamma_w}$ cannot be h since the maximum value of $R(\mathbf{\Gamma_w})$ is $E_{vo} \le h-1$. Hence, for $0 \le m \le h-1$, $p(R(\mathbf{\Gamma_w})=h|E_{vo}=m)=0$. When $h \le m \le kH$, the probability of $R(\mathbf{C})=h$ under the condition of $E_{vo}=m$ is $p(R(\mathbf{\Gamma_w})=h|E_{vo}=m)$. In summary, $p_e=\sum_{m=h}^{kH} p(E_{vo}=m)p(R(\mathbf{\Gamma_w})=h|E_{vo}=m)$, then Theorem 4 proved. □

Lemma 1. *For $h \leq m \leq kH$, we have*

$$p(R(\mathbf{\Gamma}_{\boldsymbol{w}} = h | E_{vo} = m) = \prod_{i=0}^{h-1}(1 - q^{i-m}) \tag{25}$$

where q is the size of the finite field.

Proof. Let $N_{m,h}(h)$ denote the number of $m - by - h$ matrices that have a rank of $h(m \geq h)$. Then form [18], we have

$$\begin{array}{ll} N_{m,h}(h) & = q^{mh} \prod_{i=0}^{h-1} \frac{(1-q^{i-h})(1-q^{i-m})}{(1-q^{i-h})} \\ & = q^{mh} \prod_{i=0}^{h-1}(1 - q^{i-m}) \end{array} \tag{26}$$

Hence, we have

$$\begin{array}{ll} p(R(\mathbf{\Gamma}_{\boldsymbol{w}}) = h | E_{vo} = m) & = \frac{N_{m,h}(h)}{q^{mh}} \\ & = \prod_{i=0}^{h-1}(1 - q^{i-m}) \end{array} \tag{27}$$

This, we complete the proof of Lemma 1. □

Theorem 5. *According to Lemma 1, we have*

$$\begin{array}{ll} p_e & = \sum_{m=h}^{kH} p(E_{vo} = m) p(R(\mathbf{\Gamma}_{\boldsymbol{w}}) = h | E_{vo} = m) \\ & = \sum_{m=h}^{kH} p(E_{vo} = m) \prod_{i=0}^{h-1}(1 - q^{i-m}) \end{array} \tag{28}$$

Hence, the probability of transmission being secure can be written as

$$\begin{array}{ll} p_s & = 1 - p_e \\ & = 1 - \sum_{m=h}^{kH} p(E_{vo} = m) \prod_{i=0}^{h-1}(1 - q^{i-m}) \end{array} \tag{29}$$

5. Power Consumption Analysis

Figure 6 shows the power consumption of different components in WSNs, which is proposed by Estrin [19]. It indicates that compared with the energy consumed by data transmitting and receiving, the energy consumed by other components, including sensing, computing and sleeping can be negligible. Meanwhile, the energy consumption of idling is always allocated by the nodes to avoid collisions and does not affect the energy analysis in network layer since avoiding collisions is a function in MAC layer [20]. Therefore, in this article, the total energy consumption of one transmission can be written as

$$E = E_{TX} + E_{RX} \tag{30}$$

Specifically, in one transmission, the energy consumption of transmitting and receiving can be written as

$$E_{TX}(B, d) = E_{TXElec} * B + E_{amp} * B * d^{\gamma} \tag{31}$$

$$E_{RX}(B, d) = E_{RXElec} * B \tag{32}$$

where B is the number of bits of data, E_{TXElec} and E_{RXElec} are the energy consumption of transmitting and receiving a bit of data respectively. In addition, E_{amp} is the amplification factor of the amplifier, d is the distance between transmitting node and receiving node, γ is the path-loss factor. Therefore, when a source node needs to transmit a packet of B bits to a sink node through a path of k hops, the total energy consumption is

$$\begin{array}{ll} E_{B,d,k} & = E_{TX}(B, d) * k + E_{RX}(B, d) * k \\ & = k * B * (E_{TXElec} + E_{RXElec}) + k * E_{amp} * B * d^{\gamma} \end{array} \tag{33}$$

In multipath routing scheme without network coding, to achieve a desired successful delivery ratio of R, the number of paths should be employed is

$$P = \lceil \frac{\log(1-R)}{\log(1-(1-e)^k)} \rceil \tag{34}$$

Therefore, to transmit N_p packets successfully from the source to the node, the number of packets that source node needs to transmit totally is

$$N_{total}^1 = N_p * \lceil \frac{\log(1-R)}{\log(1-(1-e)^k)} \rceil \tag{35}$$

Hence, the total energy consumption is

$$E_{h,d,k}^1 = k * N_p * \lceil \frac{\log(1-R)}{\log(1-(1-e)^k)} \rceil * b * (E_{TXElec} + E_{RXElec}) + k * E_{amp} * N_p * \lceil \frac{\log(1-R)}{\log(1-(1-e)^k)} \rceil * b * d^\gamma \tag{36}$$

where b is the number of bits per packet. On the other hand, in the network coding based multipath routing scheme, to achieve a desired successful delivery ratio of R with multicast capacity of h, the number of path needs to be employed H can be calculated by Algorithm 1. Then to transmit N_p packets successfully from the source to the node, the number of packets that source node needs to transmit totally is

$$N_{total}^2 = N_p * \frac{H}{h}; \tag{37}$$

Hence, the total energy consumption is

$$E_{h,d,k}^2 = k * N_p * \frac{H}{h} * b * (E_{TXElec} + E_{RXElec}) + k * E_{amp} * N_p * \frac{H}{h} * b * d^\gamma \tag{38}$$

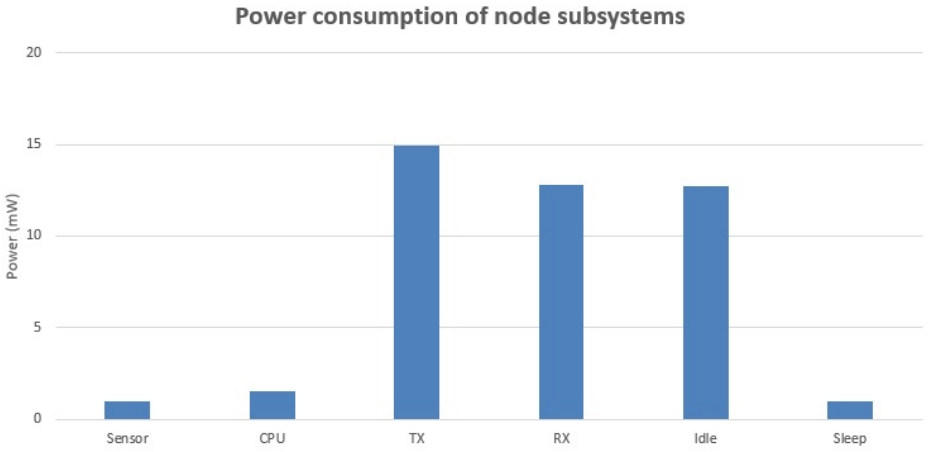
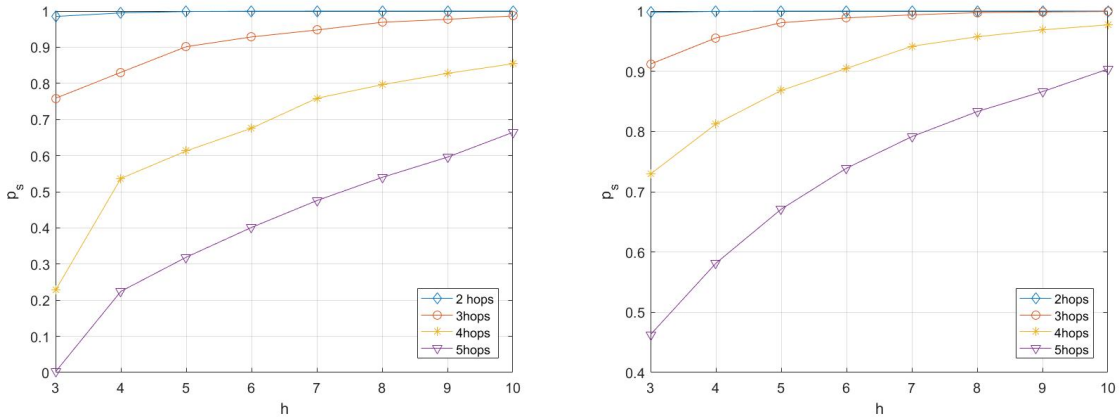


Figure 6. Power consumption of different components in wireless sensor nodes.

6. Simulations Results

6.1. Simulations of Security

Basing on the analysis in Section 4, the simulations on the probability of transmission being secure are conducted with different network parameters, including h, e, k, p_o, etc. Here the simulation results are presented in Figures 7–10.



(a) **(b)**

Figure 7. This is a figure that shows the relationship between the probability of transmission being secure and the multicast capacity with different p_o. (**a**) relationship between the probability of transmission being secure and capacity with $p_o = 0.1$; (**b**) relationship between the probability of transmission being secure and capacity with $p_o = 0.05$.

It can be concluded from Figure 7 that with the increase of multicast capacity h, the probability of transmission being secure p_s increases slightly. However, With the increase of number of hops k, the probability of transmission being secure decreases rapidly. If the desired probability of transmission being secure is $p_S \geq 0.99$, the number of hops needs to be limited with $k \leq 4$.

From Figure 8, it can be concluded that the attacking capability of adversary has a relatively significant impact on the probability of transmission being secure, and the larger the number of hops is, the greater the impact is.

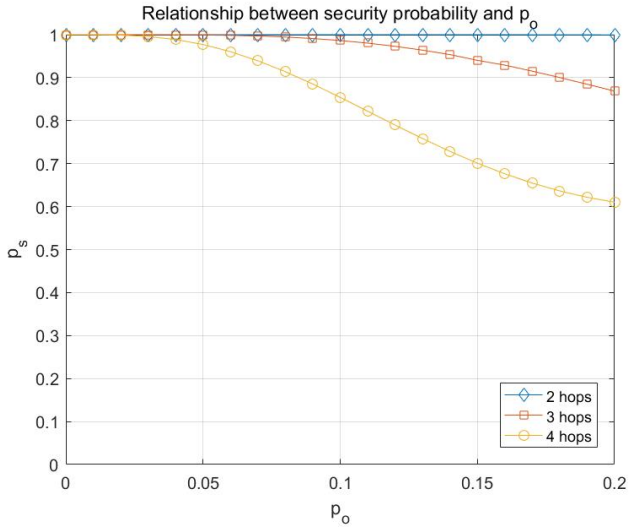


Figure 8. Relationship between the probability of transmission being secure p_s and p_o with multicast capacity is $h = 10$.

From Figures 9 and 10, a conclusion can be drawn that the probability of transmission being secure is irrelevant with the number of nodes or the size of finite field.

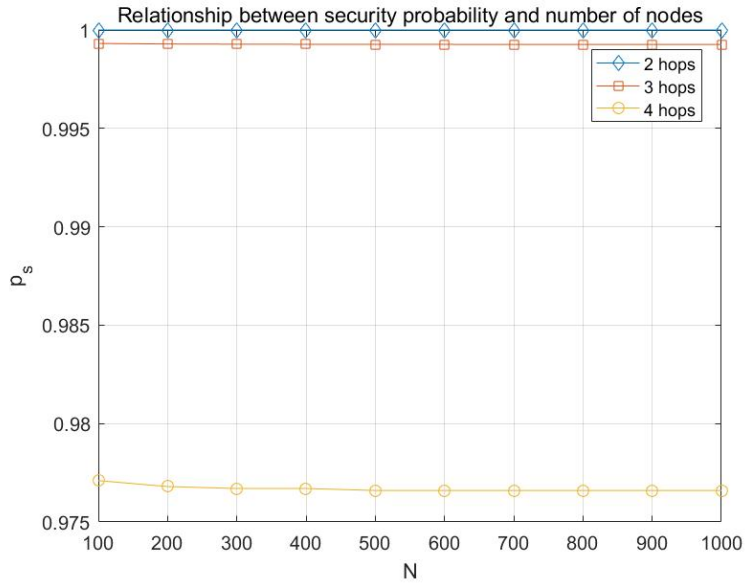


Figure 9. Relationship between the probability of transmission being secure p_s and number of nodes with multicast capacity is $h = 10$ and and $p_o = 0.05$.

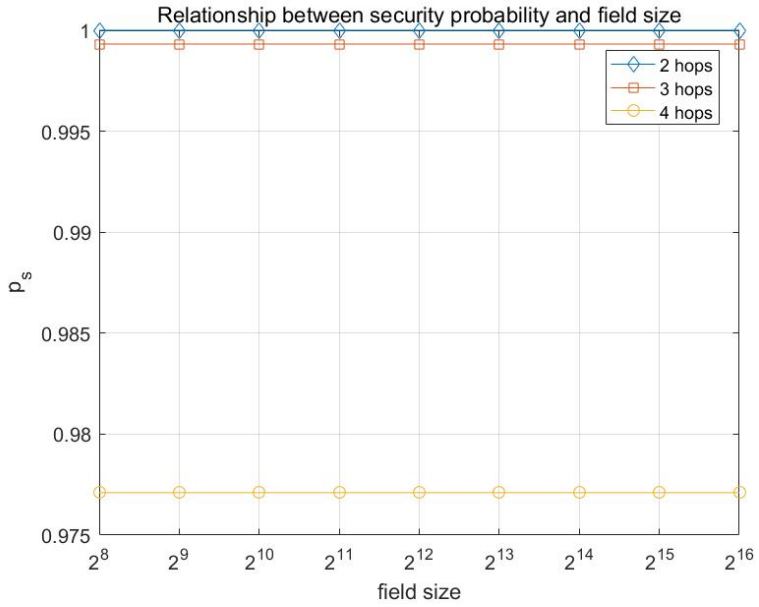


Figure 10. Relationship between the probability of transmission being secure p_s and the size of the finite field with multicast capacity is $h = 10$ and $p_o = 0.05$.

6.2. Simulations of Energy Consumption

In Section 5, the analysis of energy consumption of the network coding based multipath routing scheme and the multipath routing scheme without network coding is conducted. Basing on that, simulations on energy consumption are performed and the results are presented in Figure 11.

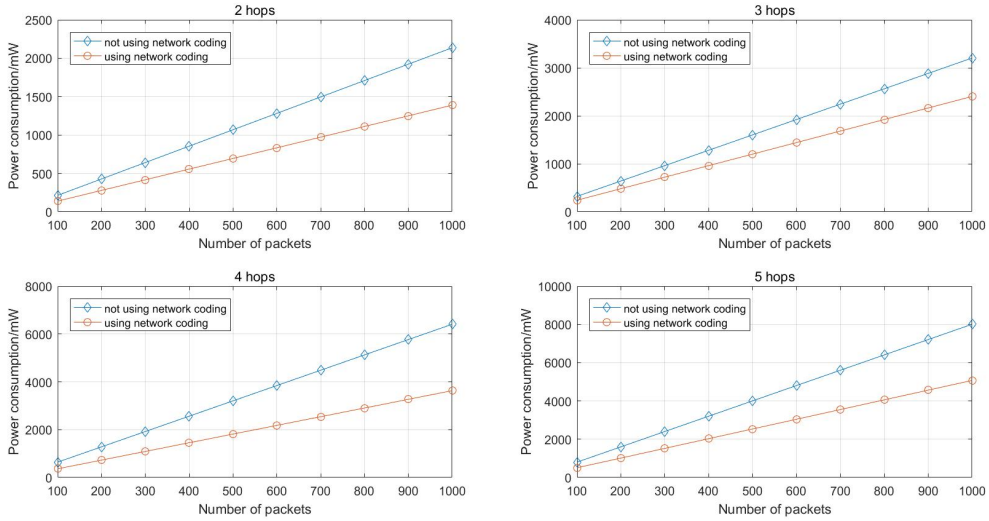


Figure 11. Energy consumed for transmitting $N_p = 100 : 100 : 1000$ packets with different path length.

As Figure 11 shows, it is clear that the network coding based scheme has a better energy efficiency and the more packets transmitted, the more energy consumption can be reduced.

7. Conclusions

In this article, a weakly secure network coding based multipath routing scheme is proposed. Based on that, the analysis on security and power consumption is conducted. Accordingly, the simulations on the probability of transmission being secure and power consumption are performed and the comparison of power consumption between two different schemes is performed as well. According to the analysis and simulation results, some conclusions can be drawn and are listed as follows:

1. As the number of hops of each path in the network increases, the probability of transmission being secure decreases rapidly, especially under the condition of low communication capacity. For example, when the the capacity is $h = 3$ and $p_o = 0.1$, with k being 2, 3, 4, and 5, the probability of transmission being secure is 0.9851, 0.7586, 0.2277, and 0.0020 correspondingly. Toward this end, if the desired probability of transmission being secure is $p_s \geq 0.99$, the number of hops should be limited with $k \leq 4$. To do this, it is necessary to deploy nodes with larger communication distance,especially when the nodes are deployed in a rather vast area.
2. When the number of hops $k \geq 3$, with the increase of multicast capacity h, the probability of transmission being secure increases and approaches 1 gradually. When $k = 2$, with the increase of multicast capacity h, the probability of transmission being secure almost keeps unchanged and satisfies $p_s \approx 1$.
3. When the number of hops $k \geq 3$, the overhearing ability, which can be reflected by the figure of p_o, has a relatively significant impact on the probability of transmission being secure. However, when $k = 2$, the probability of transmission being secure almost keeps unchanged and is approximately equal to 1.
4. Compared with the multipath routing scheme without network coding, the network coding based scheme in this article has a better energy efficiency. According to the simulation results, when 1000 packets are transmitted and the multicast capacity is $h = 10$, the power consumption of network coding based multipath routing scheme is 36.67% less than the scheme without network coding.

Acknowledgments: This work has been supported by the National High Technology Research and Development Program of China (863 Program).

Author Contributions: Xiang Liu conceived the idea of probabilistically weakly secure network coding scheme and conducted the derivations and proofs. Xiang Liu and Jie Huang designed the simulations. Xiang Gao performed the simulations and analyzed the data. Xiang Liu and Jie Huang wrote the paper.

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

References

1. Shaikh, R. A.; Lee, S.; Song, Y. J.; Zhung, Y. Securing distributed wireless sensor networks: Issues and guidelines. In Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, Taichung, Taiwan, 5–7 June 2006; pp. 226–231.
2. Gaware, A.; Dhonde, S.B. A survey on security attacks in wireless seneor networks. In Proceedings of the International conference on Computing for Sustainable Global Development, New Delhi, India, 16–18 March 2016; pp. 536–539.
3. Abd El Dayem, Sanaa. S.; Rizk, M.R.M. Security for wireless seneor network. In Proceedings of the International Conference on Information Communication and Management, Hertfordshire, UK, 29–31 October 2016; pp. 173–177.
4. Sharma, S. Wireless sensor network and security. In Proceedings of the International Conference on Computing for Sustainable Global Development, New Delhi, India, 16–18 March 2016; pp. 3301–3304.
5. Praveena, A.; Smys, S. Efficient cryptographic approach for data security in wireless sensor networks using MES V-U. In Proceedings of the International Conference on Intelligent Systems and Control, Coimbatore, TN, India, 7–8 January 2016; pp. 1–6.
6. Zhang, M.; Liu, Y. A New Approach to security Analysis of Wireless Sensor Networks for Smart Home Systems. In Proceedings of the International Conference on Intelligent Networking and Collaborative Systems, Ostrava, Czech Republic, 7–9 September 2016; pp. 318–323.
7. Perrig, A.; Szewczyk, R.; Wen, V.; Culler, D.; Tygar, J.D. SPINS: Security Protocols for Sensor Networks. *ACM Wirel. Netw.* **2002**, *8*, 521–534.
8. Eschenauer, L; Gligor, V.D. A Key-Management Scheme for Distributed Sensor Networks. In Proceedings of the 9th ACM Conference on Computer and Communications Security, Washington, DC, USA, 18–22 November 2002; pp. 41–47.
9. Cai, N.; Yeung, R.W. Secure Network Coding. In Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland, 30 June–5 July 2002.
10. Shannon, C.E. Communication theory of secrecy system. *Bell Syst. Technol. J.* **1949**, *28*, 656–715.
11. Ozarow, L.H.; Wyner, A.D. Wire-tap channel II. *AT&T Bell Labs Technol. J.* **1984**, *63*, 2135–2157.
12. Blakley, G.R. Safeguarding cryptographic keys. In Proceedings of the International Workshop on Managing Requirements Knowledge, New York, NY, USA, 4–7 June 1979; pp. 313–317.
13. Shamir, A. How to share a secret. *Commun. ACM* **1979**, *22*, 612–613.
14. Bhattad, K.; Narayanan, K.R. Weakly secure network coding. In Proceedings of the 1st Workshop on Network Coding, Theory and Applications, Riva del Garda, Italy, 7 April 2005.
15. Yang, Y.; Zhong, C.; Sun, Y.; Yang, J. Network coding based reliable disjoint and braided multipath routing for sensor networks. *J. Netw. Comput. Appl.* **2010**, *33*, 422–432.
16. Li, S.-S.; Zhu, P.-D.; Liao, X.-K.; Cheng, W.-F.; Peng, S.-L. Energy efficient multipath routing using network coding for sensor networks. In Proceeding of the ADHOC-NOW'06, Ottawa, ON, Canada, 17–19 August 2006; pp. 114–127.
17. Wei, Y.; Yu, Z.; Guan, Y. Efficient Weakly-Secure Network Coding Schemes against Wiretapping Attacks. In Proceedings of the IEEE International Symposium on Network Coding, Toronto, ON, Canada, 9–11 June 2010; pp. 1–6.
18. Laksov, D.; Thorup, A. Counting matrices with coordinates in finite fields and of fixed rank. *Math. Scand.* **1994**, *74*, 19–33.

19. Estrin, D. Wireless sensor networks tutorial part IV: Sensor network protocols. In Proceedings of the 8th Annual International Conferece on Mobile Computing and Networking, Altlanta, GA, USA, 23–28 September 2002; pp. 23–28.
20. Wang, L.; Yang, Y.; Zhao, W. Network coding-based multipath routing for energy efficiency in wireless sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2012**, *2012*, 115.