



# A Novel Routing Algorithm with Bernoulli Sampling-based Link Quality Estimation in Wireless Sensor Networks

Chao Meng<sup>1</sup>

Accepted: 28 May 2022 / Published online: 24 June 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Link quality is important and can greatly affect the performance of wireless transmission algorithms and protocols. Currently, researchers have proposed a variety of approaches to implement link quality estimation. However, the estimated result of link quality is not accurate enough and the error is large, so they may lead to the failure of routing algorithm and protocol. In this paper, a novel method is proposed to achieve the more accurate estimation of link quality than before. This method employs Bernoulli sampling-based algorithm to complete the estimation of link quality. The problem is modeled as calculation of estimators based on Bernoulli sampling data. The authors further prove that the calculation results are accurate by probability theory. Furthermore, according to link quality estimation, the authors also provide a centralized routing algorithm and a distributed improvement algorithm in order to switch the data transmission on the better quality link. Finally, the extensive experiment results indicate that the proposed methods obtain high performance in terms of energy consumption and accuracy.

**Keywords** Wireless sensor networks · Bernoulli sampling · Link quality · Approximate algorithm

## 1 Introduction

Wireless Sensor Networks have been widely used in many applications for gathering data from physical environment. In most of the applications, the scale of sensory data has already reached several petabytes each year. A large amount of sensory data needs to be transferred from sensor nodes to sink in order to make a deep analysis. However, data transmission depends on wireless link quality, which is affected by external conditions and node power. Therefore, it has become a research focus on how to estimate link quality accurately and select high-quality wireless links to ensure the efficient transmission of data. This is also the core topic of this paper.

---

✉ Chao Meng  
mz848@163.com

<sup>1</sup> School of Modern Posts, Nanjing University of Posts and Telecommunications Nanjing, Jiangsu 210003, China

Data packet reception ratio (PRR) [1] is an important symbol of link quality. At present, many network transmission algorithms and protocols are proposed to evaluate the link PRR. In the early studies [2–4], data dissemination protocols and algorithms can broadcast data to multiple neighbors at once [5–7]. In [5], the authors propose two algorithms, which rely only on local two-hop topology information to reduce the number of transmission. In [6], the corresponding link PRR is obtained by receiving data from neighbor nodes and then used to determine the transmission route [7].

Among these researches, the bitmap schemes [1–3, 8–10] have demonstrated their effectiveness in achieving communication efficiency and reliability. These schemes can avoid network congestion and improve transmission efficiency. However, they are based on direct measurement method, which is the random experiments characterized by uncertainty. Specifically, in the experiments, a small amount of sensory data is transferred from a node to its neighbor. The authors observe the proportion of data received and lost, and obtain the measured value of PRR. However, according to the principle of probability theory, a small number of random experiments cannot obtain accurate results. Only when a large number of repeated experiments are done can accurate results be obtained.

Consider the coin toss experiments as an example, if a few randomized coin toss experiments are performed, it is difficult to ensure that half of the experimental results are heads and the other half are tails. Only when a large number of repeated experiments are performed can the frequency of heads be maintained at about 0.5 [11, 12]. Likewise, in a small amount of direct data transmissions such as the bitmap schemes [1–3, 8–13], if the actual PRR is 0.7, it is almost impossible to achieve that seventy percent of packets are correctly accepted and thirty percent of packets are lost. There is a large deviation between the measured PRR and the actual PRR. Only when a large amount of direct transmission experiments are implemented can the accurate result be obtained. Unfortunately, the node energy is limited, and the energy is almost exhausted after a large amount of transmission. Hence, the measurement of the link PRR loses meaning.

In order to obtain accurate link PRR, a lot of direct data transmissions are done, which leads to too much energy consumption. However, a small number of direct transmissions are difficult to achieve accurate PRR. Furthermore, the routing protocols and algorithms often choose the optimal path based on link quality. Inaccurate link quality directly causes wrong results of routing protocols. Therefore, it seems like a dilemma.

In order to solve this problem, in this paper, the authors use the Bernoulli-sampling theory [13] to accurately estimate the link PRR by sampling sensory data and transmitting a small amount of sampling data instead of direct data transmissions. Because the energy consumption of transmission is much larger than that of instruction execution and data sampling [13, 14], the energy consumption of data sampling is almost negligible. Therefore, a small amount of sample data transmissions bring low energy consumption. In addition, the application of sampling theory can accurately estimate the link quality and ensure that the estimation results meet the requirements of high accuracy [15–17]. Because the link quality is obtained accurately, the subsequent routing protocols can also get accurate and reliable results [18].

Specifically, for arbitrary  $\epsilon (\epsilon \geq 0)$  and  $\delta (0 \leq \delta \leq 1)$ , the method proposed based on Bernoulli-sampling [19–21] theory computes  $(\epsilon, \delta)$ -approximate link PRR satisfying that the probability of the PRR's relative error being larger than  $\epsilon$  is less than  $\delta$ . Both  $\epsilon$  and  $\delta$  represent the accuracy requirements of estimation results. Therefore, as long as  $\epsilon$  and  $\delta$  are set small enough, an appropriate sampling probability can be determined to ensure that the estimation result of link PRR meets the requirements of high accuracy.

The purpose of link quality estimation is to choose the optimal transmission path. Because the sensor nodes are usually powered by batteries, energy consumption is the primary consideration. The authors use the expected transmission count (ETC) as the important indicator of energy consumption. The least link ETC means the lowest energy consumption. Therefore, the optimal transmission path means that the energy consumption of data transmission in this path is minimum and the sum of link ETC in this path is minimum.

In this paper, the authors also propose an algorithm to calculate the ETC of all links according to  $(\epsilon, \delta)$ -approximate link PRR. Then, an optimal path algorithm is presented, which takes the sum of all links ETC in the path as the optimization objective, and finds the path with the smallest sum as the optimal path [23, 23, 24]. Finally, due to the distributed architecture of sensor network, a distributed improvement algorithm is proposed in order to decrease the time complexity. Moreover, this algorithm can find the optimal path from all sensing nodes to sink. The research of this paper has great practical significance for battery-powered WSNs.

The contributions of this paper are as follows:

1. A mathematical method to determine a sampling probability based on given  $(\epsilon, \delta)$  is proposed.
2. An approach to estimate the link PRRs and ETCs of all the links in WSNs based on the sampling probability is provided.
3. A centralized algorithm is presented to seek the most optimal path which guarantees that the sum of the link ETC from sensor node to sink is minimal.
4. A distributed improvement scheme is proposed in order to reduce the time complexity.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 gives the problem definition. Section 4 describes the mathematic foundations of  $(\epsilon, \delta)$ -approximate link PRR. Section 5 proposes an algorithm to calculate the PRRs and ETCs of all the links in WSNs and also presents the search algorithms of the optimal path. The experimental results are shown in Sect. 6. Section 7 concludes the paper.

## 2 Related Work

Using sampling algorithm to evaluate link quality is a major innovation in this paper. In fact, the sampling-based approximate algorithms have been presented in several field, such as tradition database, aggregation analysis, P2P network and so on. However, none of them are about data transmission in WSNs.

In data transmission of wireless link, direct measurement [1–3, 8–12] is often used to obtain link PRR. In [1], the authors present the design of supporting layer for energy-efficient reliable broadcast, in which direct measurement is used to obtain link PRR. In [9], the authors propose collective flooding, which achieves flooding reliability using the concept of collective ACK. Zhao et al. [8] improve collective flooding architecture. Direct measurement is also used in both of these documents.

Currently, for the acquisition of link quality, direct measurement is simple and convenient. Therefore, it has been applied in many literatures. However, because direct measurement which can be regarded as a random experiment is always based on a small amount of direct data transmissions, there is a big deviation between the measured value

and the actual value according to the principle of probability theory. As a result, the inaccurate measurement of link quality directly leads to the failure of routing algorithm. On the other hand, although a large number of transmission can obtain accurate measurement values, it consumes a lot of energy, so it has no practical significance.

To solve these problems, the authors present the estimation of link PRR based on Bernoulli-sampling theory. The transmission of a small amount of sampled data ensures a low level of energy consumption. Meanwhile, the application of the sampling theory can also ensure the accuracy of the estimation results. Precise link PRR makes routing algorithm reliable [25–27].

### 3 Problem Definitions and Assumptions

In this section, the authors firstly describe the mathematical model and related parameters, then the authors present the definition of the problem which are tackled in this paper.

Assuming that the sensor network is relatively stable. Therefore, the link quality remains almost unchanged for a small period of time. In this period, the author considers the problem of link quality estimation and data transmission.

In WSN, the clocks of sensor nodes are synchronized, which can be obtained by some technologies [18].

A sensor node may have several neighbor nodes. The authors obtain the link PRR from arbitrary node to one of the neighbor nodes.

Let  $p$  and  $N$  be the link PRR from arbitrary node  $u$  to its neighbor node  $v$  and the overall amount of data transmissions from node  $u$  to  $v$ , respectively.  $\hat{p}$  is the corresponding estimation value of  $p$  based on the sampling algorithm.

$X(i)$  is a random variable. If the  $i$ -th sensed data of node  $u$  is successfully received by  $v$ ,  $X(i)$  equals 1, otherwise 0. If there are two or more nodes with the same  $X$ , the link quality may be almost the same, which has no impact on the scheme in this paper.

$X = \{X(1), X(2), \dots, X(N)\}$  is the set of random variables. The exact link PRR is

$$p = \frac{1}{N} \sum_{i=1}^N X(i)$$

The above formula is the direct measurement method in order to achieve the PRR [1–3, 8–12]. The value of  $N$  cannot be too large. Otherwise, it is meaningless to transmit a large amount of data and consume a lot of energy for the sake of precise link PRR. Therefore, the value of  $N$  is small. However, in this case, the results of direct measurement are not accurate enough according to probability theory.

To solve this problem, the authors propose the improved scheme, which uses the Bernoulli sampling theory to transmit a small amount of sampling data instead of direct data transmission. This scheme can not only estimate link PRR accurately, but also ensure low energy consumption.

The estimated definition of link PRR is as follows:

**Definition 1** ( $(\epsilon, \delta)$ -approximate value).  $\hat{I}$  is called as an  $(\epsilon, \delta)$ -estimated value of  $I$  if and only if.

$$\Pr \left\{ \left| \frac{\hat{I} - I}{I} \right| \leq \epsilon \right\} \geq 1 - \delta \quad \text{or} \quad \Pr \left\{ \left| \frac{\hat{I} - I}{I} \right| \geq \epsilon \right\} \leq \delta \tag{1}$$

for any  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , where  $\Pr \{Y\}$  is the probability of random event  $Y$ .

**Definition 2** Let  $I$  and  $\hat{I}$  be the value  $p$  and  $\hat{p}$ , respectively.  $\hat{p}$  is called as  $(\epsilon, \delta)$ -estimated value of  $p$  if and only if  $\hat{p}$  and  $p$  meet (1).

If the estimated value  $\hat{p}$  satisfy (1), it also indicates that the relative errors of  $\hat{p}$  satisfy (1).

In definition 1,  $\epsilon$  and  $\delta$  are accuracy requirements. The smaller the values of  $\epsilon$  and  $\delta$  are, the higher the estimation accuracy is.

**Definition 3** (*Bernoulli sampling*) Bernoulli sampling is a sampling method in which all elements of the population have the same probability to be independently selected [19, 20]. Hence, for a given sampling probability  $q$ , a Bernoulli sample satisfies that every data is excluded in sampling set with probability  $1-q$  and included with probability  $q$  independently.

In our design, every node acquires the Bernoulli sampling set, and then broadcasts the sampling data to its neighbors so as to estimate the link PRR.

Therefore, the determination of sampling probability is a key problem. Since the relative error of  $\hat{p}$  is also related to the sampling probability, the next section will discuss how to use  $\epsilon$  and  $\delta$  to give an optimal sampling probability and ensure that the estimator meets the requirements of high accuracy.

$N$  is the overall amount of data transmissions, and  $N$  is also the number of elements in the sensory data set. Bernoulli sample set is a sample of the sensory data set. Let  $n$  denote the number of elements in the sample set. Because the value of  $n$  is small and  $n \ll N$ , a small amount of sampling data transmissions are used to estimate the link PRR. This method not only estimates the link PRR accurately but also saves energy consumption.

Let  $\hat{X}(i)$  be a random variable. If the  $i$ -th data in sampling set is successfully received by the neighbor node,  $\hat{X}(i)$  equals 1. Otherwise 0.  $\hat{X} = \{\hat{X}(1), \hat{X}(2), \dots, \hat{X}(n)\}$  is the set of random variables (Table 1)

**Definition 4** (*Problem Definition 1*) The problem of computing  $(\epsilon, \delta)$ -approximate is defined as follows:

Input:

- ① Bernoulli sampling probability  $q$
- ②  $\epsilon(\epsilon \geq 0)$  and  $\delta (0 \leq \delta \leq 1)$ .
- ③  $\hat{X} = \{\hat{X}(1), \hat{X}(2), \dots, \hat{X}(n)\}$ .

**Result** : The estimation value  $\hat{p}$  from node  $u$  to its neighbor node  $v$ .

## 4 Mathematic Foundations

In this section, the authors firstly describe the estimation of PRR, and then present the acquisition of sampling probability in order to achieve precise estimation of PRR. Finally, the authors propose an approach to calculate the link ETC based on PRR, which lays the foundation for the presentation of the transmission algorithms.

**Table 1** Symbols and notations

Notation	Description
PRR	Data packet reception ratio
ETC	Expected transmission count
$N$	The overall amount of data transmissions from arbitrary node $u$ to neighbor node $v$
$p$	Exact link PRR
$\hat{p}$	Estimation value of $p$
$\Pr\{Y\}$	The probability of random event $Y$
$n$	The number of elements in sample set
$n_{\max}$	The number of elements in the maximum sample set
$X(i)$	A random variable that indicates whether the $i$ -th sensed data of the node is accepted by the neighbor node
$\hat{X}(i)$	A random variable that indicates whether the $i$ -th data in sampling set is accepted by the neighbor node
$q$	sampling probability
$r$	The lower bound of link PRR
$(\epsilon, \delta)$	Accuracy requirement
$B(q, N)$	Bernoulli distribution
$\Phi_{\delta/2}$	The $\delta/2$ fractile of the standard normal distribution
$ec_s$	The energy consumption for sending a data packet by a sensor node
$ec_r$	The energy consumption for receiving a data packet by a sensor node
$nbor(u_i)$	The set of $u_i$ 's neighboring nodes
$G = (V, E)$	Wireless Sensor Network $G$ , $V$ is the set of nodes, and $E$ is the set of links
$ed_{ik}$	The edge or link from node $u_i$ to $u_k$
$wgh(ed_{ik})$	The corresponding weight of $ed_{ik}$
$m$	The number of sensor nodes in WSNs
$nbor_{\max}$	The maximum value of the number of neighbor nodes
Relative error of PRR	$ \hat{p} - p /p$

### 4.1 Computational Model of PRR

**Definition 5 (Unbiased Estimation).**  $\hat{I}$  is an unbiased estimation of  $I$  if and only if the expectation of  $\hat{I}$  is equal to  $I$ , that is,

$$E(\hat{I}) = I \tag{2}$$

Otherwise,  $\hat{I}$  is a biased estimation.

**Definition 6 (Estimation of PRR).** The estimator of link PRR can be calculated by.

$$\hat{p} = \frac{1}{qn} \sum_{i=1}^n \hat{X}(i) \tag{3}$$

where  $q$  represents the sampling probability.

The following theorem 1 indicates that  $\hat{p}$  is an unbiased estimator of the exact  $p$ .

**Theorem 1**  $E(\hat{p})$  and  $Var(\hat{p})$  represent expectation and variance of  $\hat{p}$ , respectively.

$$E(\hat{p}) = p \tag{4}$$

and

$$\text{Var}(\hat{p}) \leq \frac{p(1 - pq)}{q} \tag{5}$$

**Proof**  $\hat{X}(i)$  follows the (0–1) distribution. When  $\hat{X}(i)=1$ , it is represented that the data is sampled and successfully transmitted. Therefore,  $\text{Pr}\{\hat{X}(i) = 1\} = pq$ .

$\hat{X}(i)$	1	0
Pr	$pq$	$1 - pq$

According to Bernoulli sampling,  $n$  can be considered as a random variable and obeys the Bernoulli distribution of the parameter  $(q, N)$ , that is,  $n \sim B(q, N)$ .

According to (3), this paper has

$$\begin{aligned} E(\hat{p}) &= \sum_{n=1}^N E(\hat{p}/Y = n) \text{Pr}(Y = n) \\ &= \sum_{n=1}^N \frac{1}{qn} npq \cdot C_N^n q^n (1 - q)^{N-n} \\ &= p \sum_{n=1}^N C_N^n q^n (1 - q)^{N-n} = p \end{aligned}$$

Due to  $n \sim B(q, N)$ , the deriving process of this formula employs the conditional expectation.

Because the value of  $N$  is large,  $n = 0$  is almost impossible to appear and has no practical significance. This paper assumes  $\text{Pr}(n = 0) = 0$ , so

$$E(\hat{p}) = p$$

Similarly,

$$\begin{aligned}
\text{Var}(\hat{p}) &= \sum_{n=1}^N \text{Var}(\hat{p}/Y = n) \Pr(Y = n) \\
&= \sum_{n=1}^N \text{Var}\left(\frac{1}{qn} \sum_{i=1}^n \hat{X}(i)\right) \Pr(Y = n) \\
&= \sum_{n=1}^N \frac{p(1-pq)}{qn} \Pr(Y = n) \\
&= \frac{p(1-pq)}{q} \sum_{n=1}^N \frac{1}{n} C_N^n q^n (1-q)^{N-n} \\
&\leq \frac{p(1-pq)}{q} \sum_{n=1}^N C_N^n q^n (1-q)^{N-n} \\
&= \frac{p(1-pq)}{q}
\end{aligned}$$

## 5 End of Proof

Theorem 1 shows that  $\hat{p}$  is an unbiased estimator of exact  $p$ .

### 5.1 Bernoulli Sampling Probability

The estimation method of link quality has been given according to (3). However, the estimation accuracy is also related to the sampling probability in addition to obtaining sufficiently small value of  $(\varepsilon, \delta)$ .

This section discusses how to calculate a sampling probability to insure that the estimator is the  $(\varepsilon, \delta)$ -approximate value and meets the requirements of high accuracy.

In general, because of the large size of WSNs, a sample usually includes more than 30 sensory data. According to (3),  $\hat{p}$  follows the normal distribution based on central limit theorem [21]. Theorem 2 gives the sampling probability.

**Theorem 2** *If the sampling probability satisfies the following inequality:*

$$q \geq \frac{1}{r} \cdot \frac{\Phi_{\delta/2}^2}{\varepsilon^2}$$

$\hat{p}$  is an  $(\varepsilon, \delta)$ -approximate value of  $p$ , where  $\Phi_{\delta/2}$  is the  $\delta/2$  fractile of the standard normal distribution, and  $r$  is the lower bound of PRR. If the link PRR is less than  $r$ , it is considered that the link is disconnected.

**Proof** Because of  $r \leq p$ , the authors have.



$$\begin{aligned}
 q &\geq \frac{1}{r} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2} \geq \frac{1}{p} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2} \\
 \Rightarrow q &\geq \frac{1}{p} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2} \\
 \Rightarrow p^2 &\geq \frac{p}{q} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2} \geq \frac{p(1-pq)}{q} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2}
 \end{aligned}$$

According to (5), the authors have

$$\begin{aligned}
 p^2 &\geq \frac{p(1-pq)}{q} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2} \geq \frac{\Phi_{\delta/2}^2}{\epsilon^2} \cdot \text{Var}(\hat{p}) \\
 \Rightarrow |p| \cdot \epsilon &\geq \Phi_{\delta/2} \cdot \sqrt{\text{Var}(\hat{p})}
 \end{aligned}$$

Because  $\hat{p}$  follows the normal distribution and  $E(\hat{p}) = p$ , according to theorems 1, the authors have

$$\begin{aligned}
 \Pr\{|\hat{p} - p| \geq \Phi_{\delta/2} \cdot \sqrt{\text{Var}(\hat{p})}\} &\leq \delta \\
 \Rightarrow \Pr\{|\hat{p} - p| \leq \Phi_{\delta/2} \cdot \sqrt{\text{Var}(\hat{p})}\} &\geq 1 - \delta \\
 \Rightarrow \Pr\{|\hat{p} - p| \leq |p| \cdot \epsilon\} &\geq 1 - \delta
 \end{aligned}$$

According to Definition 1,  $\hat{p}$  is an estimated value of  $p$ .

### 6 End of Proof

According to the theorem 2, the optimal sampling probability for calculating the  $(\epsilon, \delta)$ -approximate results of  $p$  is greater than or equal to  $\frac{1}{r} \cdot \frac{\Phi_{\delta/2}^2}{\epsilon^2}$ .

According to the link PRR, the authors can calculate the link ETC.

**Definition 7** (*expected transmission count(ETC)*). The expected transmission count (ETC) of link is determined by.

$$etc = \frac{1}{\hat{p}} \tag{7}$$

According to the above theorems, in next section, the authors can propose an algorithm which obtains the PRRs and ETCs of all the links in WSNs. This algorithm lays the foundation for putting forward the transmission algorithms in WSNs.

## 7 Acquisition Algorithm of Link etc. and Search Algorithm of the Optimal Path

In Sect. 4, the authors point out the mathematical basis for the PRR and ETC estimation methods.

In this section, the authors firstly give the network model and parameters, and then propose an algorithm to calculate the PRRs and ETCs of all the links according to the mathematical theorems proposed in the previous section. Finally, according to the ETC of all the links, the authors propose a centralized algorithm and an improved distributed algorithm to find out the optimal path from all sensor nodes to sink.

Because the energy consumption of transmission is much greater than that of execution instructions, energy consumption in transmission becomes our top issues. The optimal path algorithms always take the sum of all the link ETCs in the path as the optimization goal and find the path with the minimum sum of link ETC as the optimal path for data transmission.

In the optimal path, the sum of all the link ETCs reaches the minimum, which means the least energy consumption. This is particularly important for the sensor nodes powered by battery.

### 7.1 Wireless Sensor Network Model and Parameters

**Definition 8** (*WSN Graph*). A Wireless Sensor Network can be considered as a directed graph  $G = (V, E)$ , in which  $V$  is the set of sensor nodes, and  $E$  is the set of transmission links. Each link has a unique weight value which represents the ETC of this link.

In this paper, the authors assume that all the nodes are in a two-dimensional plane and have the completely different transmission power. If the node A is within the transmission radius of node B, there exists a link from B to A. This paper also assumes that every node has a unique ID and knows the IDs of all the one-hop neighbors.

Let  $m$  be the number of sensor nodes. For the arbitrary node  $u_i \in V$ , the set of  $u_i$ 's neighbors is represented by  $nbor(u_i)$ .

$$nbor_{\max} = \max_{1 \leq i \leq m} |nbor(u_i)|$$

Suppose that the edge or link from node  $u_i$  to  $u_k$  is denoted by  $ed_{ik} \in E$ . The corresponding weight of  $ed_{ik}$  is  $wgh(ed_{ik})$ , which also represents the link ETC from node  $u_i$  to  $u_k$ . Let  $ec_s$  and  $ec_r$  be the energy consumption for sending and receiving a data packet, respectively.

### 7.2 Acquisition Algorithm of Link PRR and ETC

According to the theorems 1 and 2, the authors can propose an algorithm which obtains the PRRs and ETCs of all the links in WSNs. In this paper, the authors use the sampling data to estimate the link PRR.

**Algorithm 1** (obtaining the PRR and ETC of all the links in WSNs)

Input: sample data transmissions of each node

Output: the PRRs and ETCs of all the links

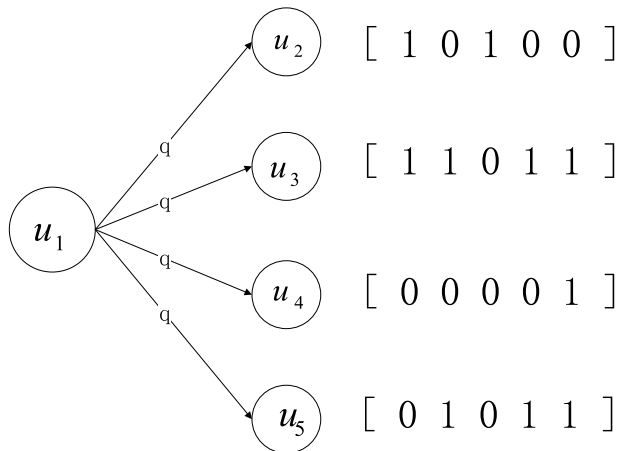
- 1 The sink calculates the sampling probability  $q$  according to (6) and broadcasts  $q$  to all nodes in WSNs.
- 2 After receiving  $q$ , each node determines whether to broadcast a packet to the neighbor nodes according to the probability  $q$ .
- 3 According to the packet reception of the neighbor nodes, each node uses (3) to calculate the link PRR  $\hat{p}$  from current node to the neighbor node.
- 4 Each node uses (7) to calculate the link ETC.

In the algorithm1, each node broadcasts the packets to its neighbors in the initial stage. Each packet is identified by the packet sequence number and the node ID.

For example, in Fig. 1, node  $u_1$  has 4 neighbors. Suppose the set of random variables of  $u_2$  is  $\hat{X} = \{1, 0, 1, 0, 0\}$ , which indicates that  $u_2$  receives the first and third packets, and misses the rest of the packets.

Suppose that the sampling probability is  $q$ . According to (3), the link PRR from  $u_1$  to  $u_2$  can be calculated. In the same way, other link PRR can also be obtained. According to (7), the link ETC can be calculated.

**Fig. 1** Example of calculating the link PRR



In the first phase of algorithm1, since sink needs to calculate sampling probability and broadcast it to other nodes in the network, the computation complexity of sink is  $O(1)$  and the maximum communication complexity of each node is  $O(ec_s \times nbor_{\max} + m \times ec_r)$ .

In the second phase, since each node performs Bernoulli sampling, the computation complexity of each node is  $O(1)$ . Because the number of elements in the sample data set is different, the authors can assume that the number of elements in the maximum sample set is  $n_{\max}$ . The maximum communication complexities of each node is  $O(nbor_{\max} \times n_{\max} \times ec_s + m \times n_{\max} \times ec_r)$ .

In the third phase, since each node only needs to calculate  $\hat{p}$ , the computation complexity of each node is  $O(1)$ . Similarly, in the fourth phase, the computation complexity of each node is  $O(1)$ .

In summary, the total computation complexities of each node is  $O(1)$ , and the total communication complexities of each node is  $O(nbor_{\max} \times n_{\max} \times ec_s + m \times n_{\max} \times ec_r)$ .

So far, the authors have solved the problem definition 1 and give all link ETCs. Next, the authors propose the routing optimization algorithms based on ETCs.

### 7.3 Centralized Optimal Path Search Algorithm

In algorithm 1, the authors use sampling data to obtain the PRRs and ETCs of all the links, which represent the quality of all the links.

According to the link ETCs, the authors propose the optimal path algorithms to ensure that the sum of all the link ETCs in the optimal path from the sensor node to sink is minimum.

In this section, the authors give the problem definition of the optimal path, and present a centralized optimal path algorithm.

Firstly, the authors take the WSN Graph displayed in Fig. 2a as an example. Suppose that the set of nodes is  $V = \{u_1, u_2, u_3, u_4, u_5\}$  and  $u_5$  is sink.  $u_1, u_2, u_3, u_4$  are sensor nodes. The number associated with each edge indicates the weight or ETC calculated by algorithm1. According to Fig. 2a, the authors seek the optimal path from  $u_1$  to sink. Figure 2b shows each possible path from nodes  $u_1$  to sink and corresponding sum of weights in the path.

Obviously,  $u_1 \rightarrow u_2 \rightarrow u_3 \rightarrow u_5$  (sink) is an optimal path, because the sum of the weights of all links in this path is minimum.

Because the energy consumption of transmission is far greater than that of execution instructions [13, 14], the energy consumption of data forwarding can be ignored. Therefore, although the path  $u_1 \rightarrow u_2 \rightarrow u_3 \rightarrow u_5$  forwards data many times, the sum of the weights in this path is the minimum, which also means the minimum energy consumption.

From the above example, the authors introduce the concept of "optimal path".

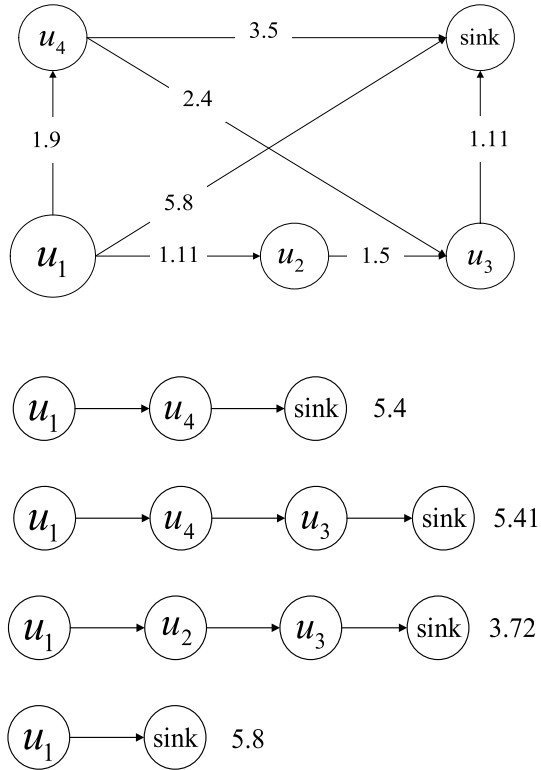
**Definition 9** (*optimal path*). Among all the paths from a perceptive node to sink, the optimal path must satisfy that the sum of all the link weights in this path reach the minimum.

**Definition 10** (*problem definition 2*). The problem of the optimal path is defined as follows:

Input:

- ① WSN Graph G
- ② The weights (ETCs) of all the wireless link, which are calculated in algorithm 1.

**Fig. 2** a WSN Graph G b All possible paths from  $u_1$  to sink



**Result** The optimal paths from all perceived nodes to sink.

According to Definition 10, the authors propose an optimal algorithm [23, 23, 24].

The algorithm maintains a set  $P$  of nodes, whose optimal path weights from the starting node have been determined. The authors repeatedly select the node  $u_k \in V - P$  with the minimum weight and add  $u_k$  to the set  $P$ .

In the following algorithm implementation, the authors denote by array  $E[k]$  the optimal path weights from the starting node to  $u_k$ .

**Algorithm 2(Centralized Algorithm). Finding the optimal path from arbitrary node  $u_i \in V$  to sink**

Input: Network topology diagram  $G$  and the ETC of each link

Output: the optimal path

1.  $P \leftarrow \emptyset$
2.  $E[k] \leftarrow wgh(ed_{ik})$  of link from the starting node  $u_i$  to node  $u_k$  ( $u_k$  is an arbitrary node except the starting node  $u_i$ )
3. While(all the nodes except the starting node)
  - {
  - 4. Find  $\min(E[j], 1 \leq j \leq n)$ ;  $P \leftarrow j$ ;
  - 5. for (all the nodes  $k$  except the nodes in the set  $P$ )
  - 6. if( $E[j] + wgh(ed_{jk}) < E[k]$ )
  - 7.  $E[k] = E[j] + wgh(ed_{jk})$ ;
  - }

Lines 1 and 2 are initialization stages. Line 1 initializes the set  $P$  to the empty set. Line 2 initializes array  $E[k]$ . If there is a link from the starting node to  $u_k$ ,  $E[k]$  is the weight value of link. If there is not a link,  $E[k] = \infty$ .

Lines 3–7 are the while loop. Line 4 finds the smallest weight in the array  $E[m]$  and adds the corresponding node  $u_j$  to the set  $P$ . Lines 5–7 is the adjustment phase of weights. For each node  $u_k$  that is not be added to the set  $P$ , the algorithm calculates whether the sum of  $E[j]$  and  $wgh(ed_{jk})$  which is the ETC of link from  $u_j$  to  $u_k$  is less than  $E[k]$ . If true,  $E[k]$  is modified. If false, it has not been modified.

Consider the architecture of Fig. 2a as an example, the execution process of algorithm 2 is shown in Fig. 3.

The authors find the optimal path from  $u_1$  to sink. In Fig. 3, the dotted line indicates that there is such a link, but the corresponding node is not added to the set  $P$ . After the corresponding node is added to the set  $P$ , the dotted line becomes the solid line.

The sub-graph (a) of Fig. 3 is initialization. The circle of the thick line represents the node that has been added to the set  $P$ , in which there is only one starting node. The number indicates the weight of link from  $u_1$  to the corresponding node. In the sub-graph (a), because there does not exist a link from  $u_1$  to  $u_3$ ,  $E[3] = \infty$ . The smallest weight is  $E[2]$  which is equal to 1.11, so the sub-graph (b) adds  $u_2$  to the set  $P$ . For the remaining nodes, the sub-graph (c) shows adjustment of weights. Because the sum of  $E[2]$  and  $wgh(ed_{23})$  is equal to 2.61 which is less than  $E[3] = \infty$ , the algorithm changes  $E[3]$  to 2.61. According to the algorithm 2, the weights of remaining node cannot be modified. The sub-graph (d) finds the smallest weight  $E[4] = 1.9$  and adds  $u_4$  to the  $P$ . The sub-graph (e) is adjustment. The following sub-graphs are like this.

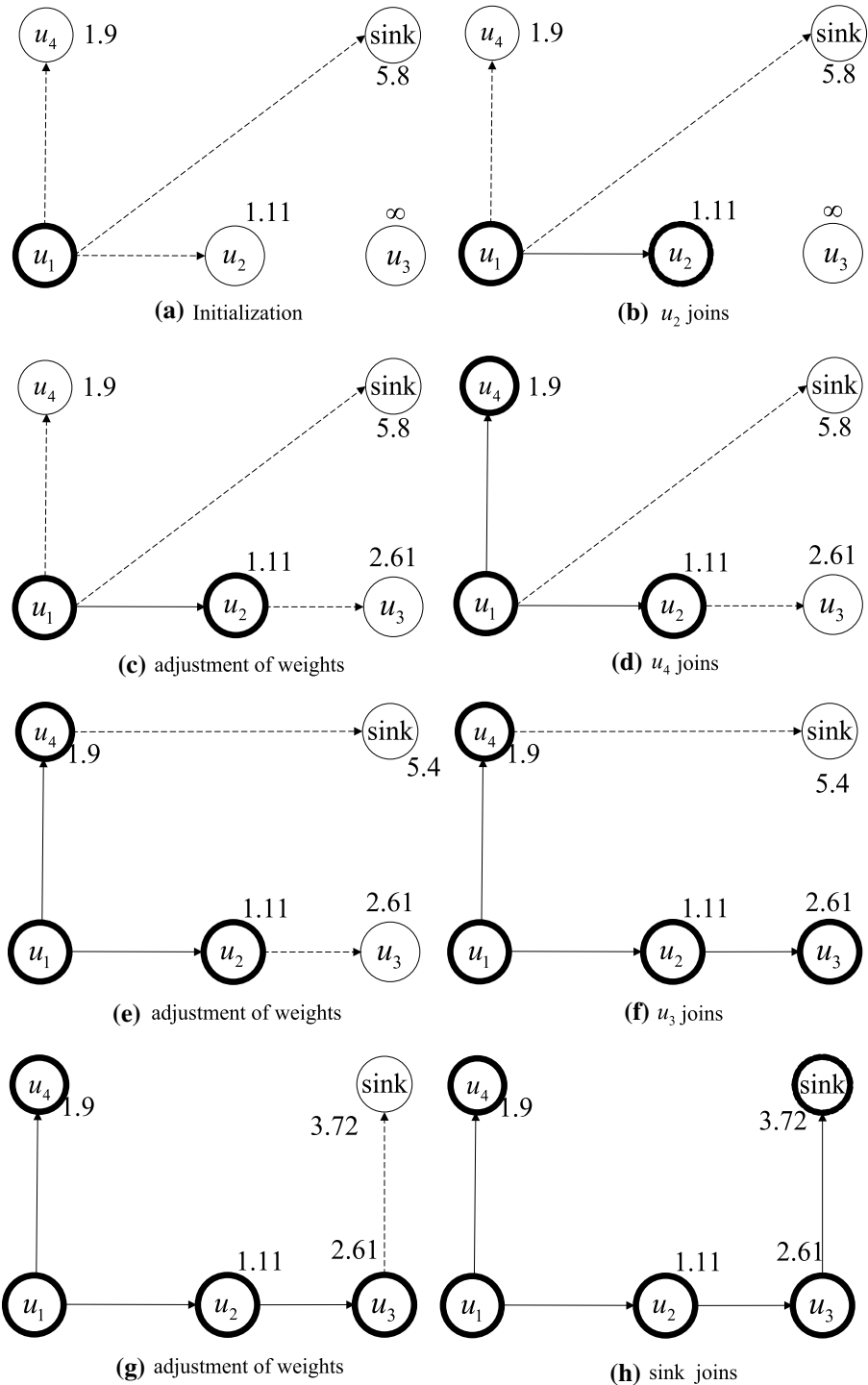


Fig. 3 The execution of algorithm 2 according to the framework of Fig. 2a

If the rest of nodes are considered as the starting nodes, the optimal path from all the nodes to sink can be obtained.

The time complexity of the algorithm is mainly concentrated in the while loop. The time complexity of algorithm 2 is  $o(m^2)$ . If the optimal path from all the nodes to sink is obtained, the total time complexity is  $o(m^3)$ .

The implementation of the algorithm 2 is very simple. Even if the number of nodes is large, the algorithm can still be applied. However, the shortcomings are very obvious. Because of the centralized algorithm, it is easy to cause overloading of local nodes, while the rest of the nodes are very idle. In addition, the algorithm needs to know the topology of the entire network before running, which is almost impossible. The algorithm is often used as a theoretical basis.

#### **7.4 Distributed Improvement Scheme for Optimal Path Algorithm**

According to the shortcomings of the centralized algorithm, the authors improve the algorithm 2 and propose a distributed implementation scheme, which can adapt to the distributed environment of sensor network. Besides, the distributed implementation also greatly reduces the requirement of time complexity.

Each node knows the neighbor nodes and the link weights or ETCs from itself to neighbors. Algorithm 3 is a distributed improvement scheme running in arbitrary node  $u_i$ .



**Algorithm 3(distributed scheme). Finding the optimal path**

Input: Network  $G$  and the ETC of each link  
Output: the optimal path from  $u_i$  to sink

```

var  $N_i$  : the set of nodes
     $W_i$  : array of weights
     $Nb_i$ : array of nodes

begin
1  $N_i \leftarrow \emptyset$ 
2 for all  $u_j \in V$  do
    if  $u_j = u_i$ 
        then begin  $W_i[j] = 0$ ;  $Nb_i[j] = undef$  end
    else if  $u_j \in N(v_i)$ 
        then begin  $W_i[j] = wgh(e_{i,j})$ ;  $Nb_i[j] = j$  end
        else begin  $W_i[j] = \infty$ ;  $Nb_i[j] = undef$  end
3 while  $N_i \neq V$  do
    begin select  $u_x$  from  $V/N_i$ 
        if  $u_i = u_x$ 
            then "Broadcast  $W_x$ "
        else "accept  $W_x$ "
        forall  $u_j \in V$  do
            if  $W_i[x] + W_x[j] < W_i[j]$ 
                then begin  $W_i[j] = W_i[x] + W_x[j]$ 
                     $Nb_i[j] = Nb_i[x]$ 
                end
            end
         $N_i := N_i \cup \{u_x\}$ 
    end
end

```

Stage 1 and 2 are initialization. In node  $u_i$ ,  $N_i$  is the set of nodes which have been considered whether they are intermediate nodes in the optimal path. In the initialization phase,  $N_i$  is an empty set.  $Nb_i$  and  $W_i$  are  $u_i$ 's successor node in the optimal path and the sum of the link weights in the optimal path, respectively.

In Stage 3, the distributed algorithm firstly broadcasts the pivot node  $u_x$  which is likely to be the relay node of the optimal path. Then, for all the nodes in WSN, the distributed algorithm continues to insert  $u_x$  into the optimal path from  $u_i$  to all the nodes. Finally,  $W_i$  saves the sum of the link weights in the optimal path from  $u_i$  to all the nodes.  $Nb_i$  preserves the successor node of  $u_i$  in the optimal path.

In this paper, the authors only consider the maximum energy cost. In Stage 1–2, the computation complexity is  $O(m)$ . In Stage 3, the computation complexity is  $O(m^2)$  and the communication complexity is  $O(ec_s)$ . Because  $ec_s \gg ec_r$ , the authors neglect  $ec_r$ .

So far, the authors solve the problem definition 2 and obtain the optimal paths from all the nodes to sink.

## 8 Experimental Evaluation

### 8.1 Testbed Experimentation

Data packet reception ratio (PRR) is of great significance across network environments. Because the wireless link is completely open, PRR is affected by environmental conditions and external interference. In order to evaluate the performance of our algorithms in WSNs, the authors use a testbed to perform the several algorithms mentioned in this paper. The testbed environment is an office building, in which a total of 50 TelosB nodes are randomly deployed on the walls of corridors and staircases, as shown in Fig. 4.

In the testbed environment, the transmission power is set at  $-20$  dBm in order that the perceptive nodes might form the multi-hop wireless networks. The default channel is 20. For broadcast, two adjacent nodes are considered as neighbors if the link PRR between them is greater than 0.2.

After deployment, all the nodes in WSNs are synchronized and begin to find out the neighbor nodes by sending out packets according to sampling probability, based on which the authors estimate the link PRR between two adjacent nodes.

Because the energy consumption of transmission is much larger than that of instruction execution and data sampling [13, 14], the energy consumption of data sampling is almost negligible. Therefore, the authors concentrate on the energy consumption of network communication.

In the following experiments, firstly, each node sends out a small amount of packets based on sampling probability  $q$  in order to obtain the link PRR and ETC which is the input parameter of the optimal path algorithm. Then, they send out 30 data packets with a time interval of 1 s to carry out performance comparison with several algorithms. For the convenience of analysis, every data packet contains hop count, timestamp and previous hop's node ID. Once the packet is accepted, the relay node may record the number of transmission for each packet.

#### 8.1.1 Performance Analysis of PRR Acquisition Algorithm

The purpose of this experiment is to investigate the relationship between the maximal relative error and the accuracy requirements. When  $\epsilon$  and  $\delta$  were increased from 0.01 to 0.1, respectively, the relative error of link PRR is calculated. The experimental results are shown in Fig. 5.

The figure shows that the maximal relative error of calculating PRR is less than 0.1. When  $\epsilon$  and  $\delta$  are less than 0.1, all the approximations are close to the actual results. Figure 5 also shows that our acquisition algorithm of link PRR can meet the requirements of high precision.

Then, in the next experiment, the authors investigate the sampling probability and the energy consumption. When the sampling probability is increased from 0.7 to 0.8, the maximal



(a)



(b)

**Fig. 4** **a** on the walls of corridors **b** on the walls of the staircase

relative error is calculated. As shown in Fig. 6, when the sampling probability is above 0.7, acquisition algorithm of link PRR only needs a small amount of sample data to obtain high precision results.

For instance, when the sampling probability is 0.75, the relative errors of results are less than 0.03. Figure 6 indicates that the accuracy of calculation increases as the sampling probability increases, but this increase will result in more data sampling and greater energy consumption. When the sampling probability is greater than 0.75, the accuracy is not improved obviously. So the sampling probability is about 0.75.

According to Fig. 6, when  $\delta$  and  $\delta$  decrease, the estimation accuracy increases gradually and the relative error decreases gradually. This result is consistent with definition 1.

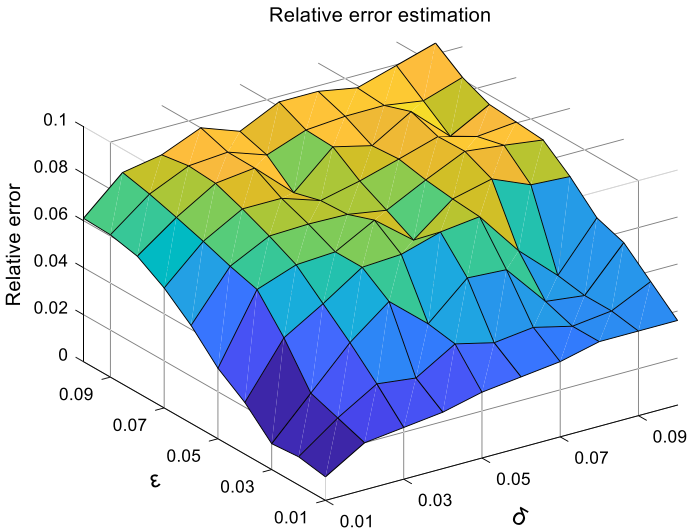


Fig. 5 Accuracy requirements

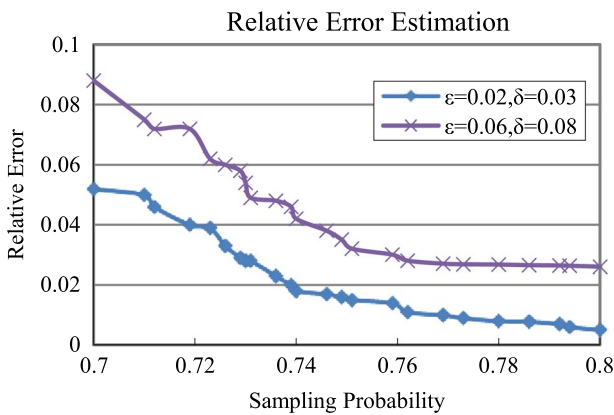


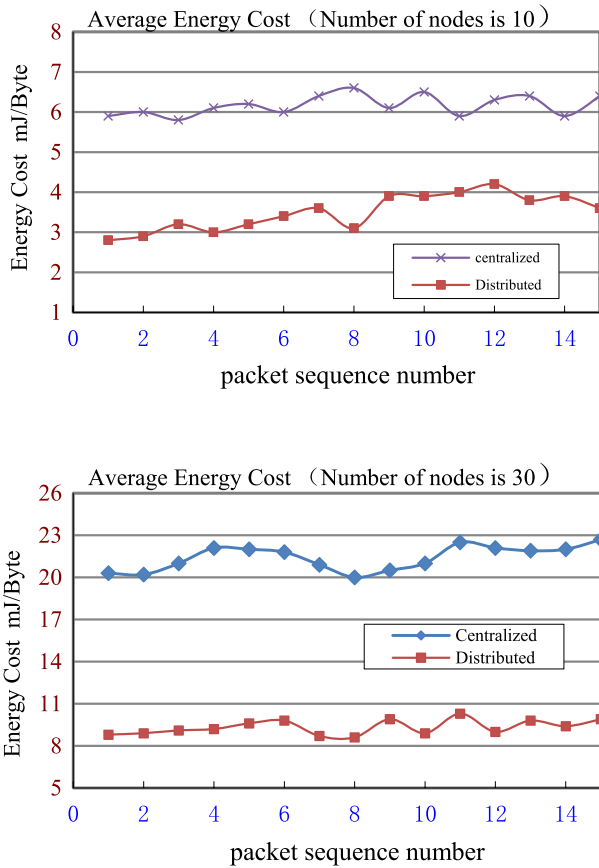
Fig. 6 Sampling probability and relative error

### 8.1.2 Comparison of Several Optimal Path Search Algorithms

The energy consumption of these algorithms proposed in this paper varies greatly under different network scale. In this experiment, the authors explore the impact of network scale on these algorithms.

The authors use the data from the testbed in office building. The results of the experiment are shown in Fig. 7, which indicates the maximum energy consumption in all nodes. When each packet is transmitted in a network of 10 nodes and 30 nodes, respectively, the energy consumption mainly refers to the energy consumption of data transmission.

Figure 7 shows that when there are 10 nodes in the network, the difference of energy consumption is not significant. However, when the number of nodes is increased to 30, the energy consumption of centralized algorithm is much larger than that of distributed



**Fig. 7** a the maximum energy consumption in all nodes b the maximum energy consumption in all nodes

algorithm. Because the distributed algorithm can distribute computation load to each node, the energy consumption is greatly reduced. Figure 7 also indicates that the distributed algorithm can be applied to networks with a large number of sensor nodes.

In the following experiment, the authors compare the distributed algorithm with several representative algorithms proposed in the literature. The excellent transmission algorithms always consume less energy. The fundamental reason is that there is less link ETC.

Hence, the authors use the total number of transmission from sensor node to sink as the metric for evaluating the energy consumption of algorithms. Because of space constraints, the authors only choose four representative transmission algorithms as comparison objects, which named as CODEB algorithm [5], CorLayer Cluster(CC) [1], Multipoint Relay (MPR) [6] and Dominating Pruning (PRUN) [7] for the rest of this paper.

Firstly, the authors examine the effect of the algorithm with different network sizes. Figure 8 shows the total number of transmission for a packet.

It can be seen that the distributed algorithm has obvious advantages compared with the other algorithms. No matter how many nodes exist in the network, the distributed algorithm has the smallest number of transmission, which indicates that the distributed algorithm can consume the least energy. Furthermore, this advantage will become more

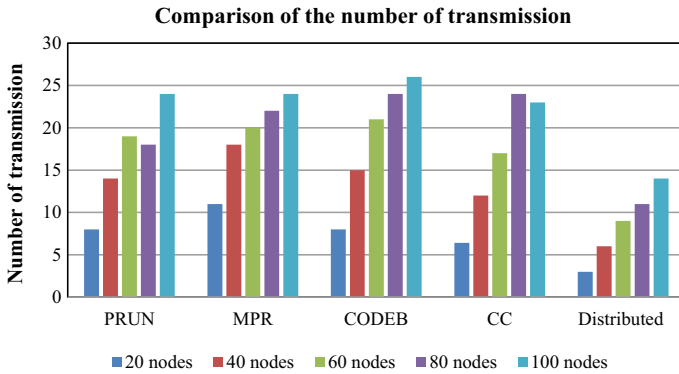


Fig. 8 Impact of network sizes

and more obvious as the scale of the network grows. This is mainly because the method proposed in this paper can accurately estimate the link quality. On this basis, the routing algorithm has achieved the best optimization effect.

Then, the authors explore the influence of different channels. The authors use 50 TelosB nodes to conduct experiments in channel 12 and channel 26, respectively. The experiment results are in Fig. 9. The transmission power is set at  $-20$  dBm. As shown in Fig. 9, when the authors choose channel 12 for data transmission, the expected transmission count (ETC) is obviously higher than that of channel 26. The reason is that WiFi signal interferes with the ZigBee protocol communication.

In China, the most frequently used channels in Wifi are 1, 6, and 11, which overlaps seriously with the ZigBee channel 12, so that the more data packets could be lost. However, channel 26 can be free from WiFi interference, and the transmission quality has been greatly improved.

No matter which channel is used for the experiment, the method in this paper can obtain accurate link quality estimation. Therefore, the distributed routing algorithm has excellent performance.

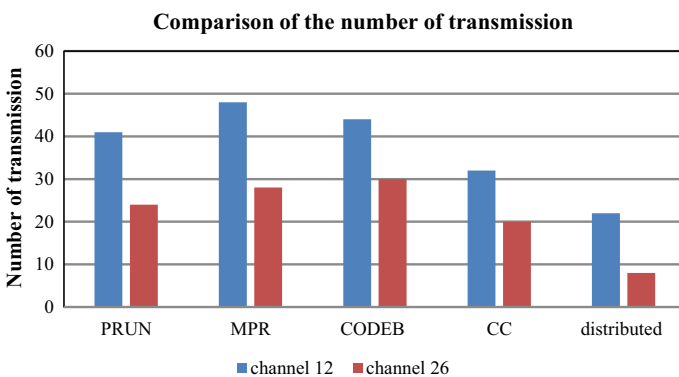


Fig. 9 Impact of channels

## 8.2 Simulation Experimentation

In order to evaluate the performance of the algorithms in larger-scale networks, the authors use the NS2 to simulate the sensor network with 10,000 nodes. All the nodes are randomly deployed into a 1,000 m  $\times$  1,000 m rectangular region. The transmission range of each node is 50 m. For the convenience of comparison, the authors assume that when the Bernoulli sampling algorithm is adopted, the sink node knows the distribution of all the sensor nodes in advance.

### 8.2.1 PRR Acquisition Algorithm in Larger Scale Networks

In this experiment, the authors investigate the influence of network size on the sampling probability. The parameter  $(\epsilon, \delta)$  is: (0.02, 0.04) and (0.14, 0.08), respectively. When the network size is changed from 4000 to 10,000, the authors examine the sampling probability required by the acquisition algorithm. The results of the experiment are in Fig. 10. It can be seen from the figure that the sampling probability of the acquisition algorithm is obviously reduced when the network size becomes larger. The reduction of sampling probability means that the packets that are required to estimate the PRR are also reduced, and the burden of the network is reduced. The experimental results show that the larger the network scale, the better the performance of the acquisition algorithm.

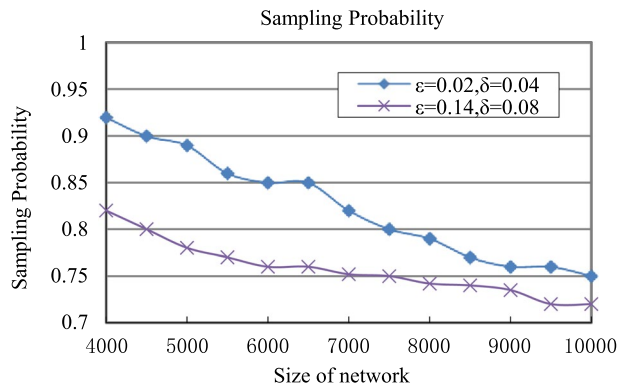
In the next experiment, the authors examine the impact of the network size and the sampling probability on the accuracy of the acquisition algorithm.

When the network size increases from 1000 to 4000, and the sampling probability changes from 0.7 to 0.9, the authors calculate the relative error of the acquisition algorithm. The results of the experiment are shown in Fig. 11. When the size of the network is larger and the sampling probability is gradually improved, the relative error of the approximate results will be significantly reduced. The experimental results also show that the larger the network scale is, the better the performance of the acquisition algorithm is.

### 8.2.2 Comparison of Distributed Optimal Algorithm and Several Representative Algorithms

Due to the shortcomings of the centralized algorithm, it is not suitable for large-scale networks. Therefore, in the following experiments, the authors use the distributed optimal

**Fig. 10** Relationship between sampling probability and network size



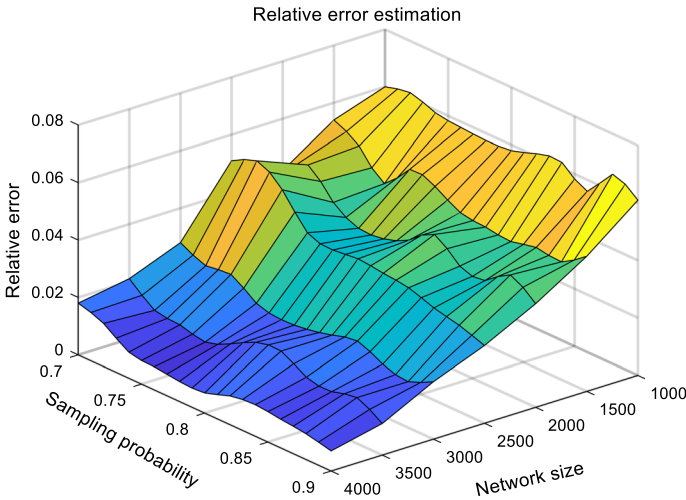


Fig. 11 Accuracy affected by the scale of the network and the sampling probability

algorithm to compare with several representative algorithms in large-scale network simulation. The authors also use the total number of transmission from sensor node to sink as the metric for evaluating the energy consumption of algorithms.

Firstly, the authors consider the total number of transmission with different link qualities. In the testbed experiment, the link quality cannot be set arbitrarily, but it can be changed arbitrarily in the simulation. In this experimental scene, the authors choose 200 sensor nodes. The experimental results are shown in Fig. 12. The expected transmission count of distributed algorithm varies from 50.8 to 31 when the average PRR varies from 0.3 to 0.9. When the link quality is poor, the distributed algorithm reduces transmissions by at least 40% compared with other algorithms under the same condition. With the improvement of link quality, our design can save at least 30% energy consumption when the link PRR reaches 0.9.

Then, the authors explore the impact of network density on our algorithms. In this experiment, the authors use 400 sensor nodes. The average link quality is about 0.8. The

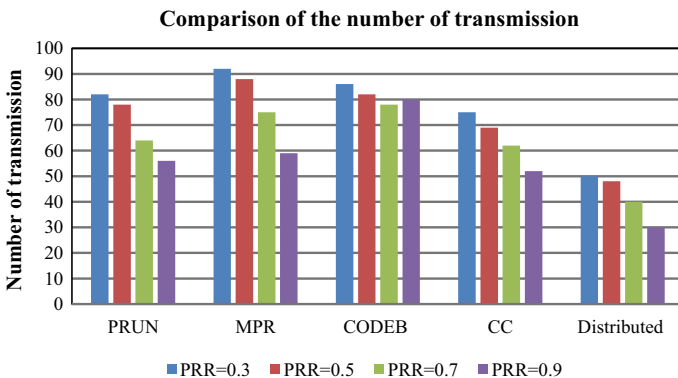
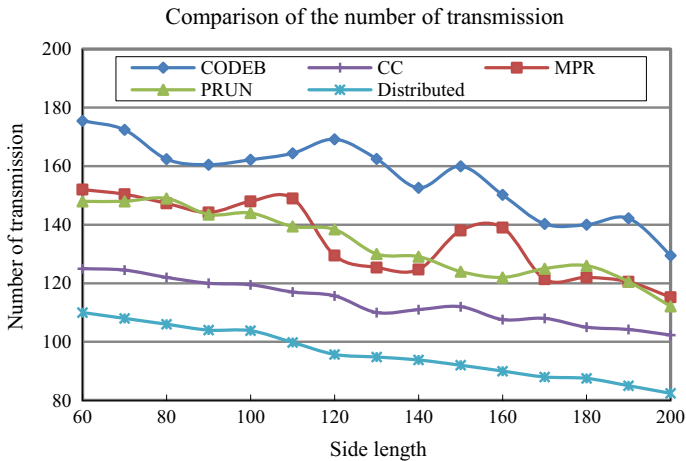


Fig. 12 Impact of link qualities





**Fig. 13** Impact of network densities

nodes are randomly deployed in a rectangular area. The smaller the perception area is, the larger the network density is. Figure 13 shows the number of transmission with different network densities. The x-axis represents the side length of the perceptual area. The increase of side length indicates the reduction of network density. When the perceptual area becomes large and network becomes sparse, the number of neighbors around sensor nodes decreases gradually. As a result, the number of transmission is correspondingly reduced.

From Fig. 13, the number of transmission does not change monotonically. The reason is that the nodes are randomly distributed in the area of perception. This non-uniform distribution may cause a large number of nodes to gather in the local area, which increases the number of local transmission. Figure 13 shows that our algorithm saves at least 30% of the energy consumption.

In Figs. 12 and 13, because the link quality can be accurately estimated, the routing algorithm can always achieve excellent performance advantages regardless of the network environment.

## 9 Conclusion

In this paper, the authors propose a novel link quality estimation method and a PRR and ETC acquisition algorithm based on it, which can obtain the PRRs of all links in the networks. It is proved that the algorithm can meet the requirements of high precision. According to the link PRR and ETCs, the authors also propose a centralized path algorithm and a distributed improvement algorithm in order to find the optimal path from all the node to sink. Finally, the experiment results indicate that the performance advantage of the proposed algorithms is obvious in terms of the algorithms accuracy and energy consumption.

## Declarations

**Conflict of interest** We declare that we do not have any commercial or associative interest that represents a

conflict of interest in connection with the work submitted.

## References

1. Wang, S., Kim, S. M., Liu, Y., Tan, G., & He, T. (2013). CorLayer: A transparent link correlation layer for energy efficient broadcast. In *Proceedings of ACM MobiCom*.
2. Srinivasan, K., Jain, M., Choi, J. I., Azim, T., Kim, E. S., Levis, P., & Krishnamachari, B. (2010). The K-factor: Inferring protocol performance using inter-link reception correlation. In *Proceedings MobiCom* (pp. 317–328).
3. Zhu, T., Zhong, Z., He, T., & Zhang, Z. L. (2010). Exploring link correlation for efficient flooding in wireless sensor networks. In *Proceedings NSDI*.
4. Thangaramya, K., Kulothungan, K., Logambigai, R., et al. (2019). Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT. *Computer Networks*, *151*(14), 211–223.
5. Yue, G., Yang, K., Zhao, S., & Poor, H. V. (2018). Design of network coding for wireless broadcast and multicast with optimal decoders. *IEEE Transactions on Wireless Communications*, *17*(10), 6944–6957.
6. Wu, J., & Lou, W. (2006). Extended multipoint relays to determine connected dominating sets in MANETs. *IEEE Transactions on Computers*, *55*(3), 334–347.
7. Liu, C. (2016). Performance-guaranteed strongly connected dominating sets in heterogeneous wireless sensor networks. In *Proceedings IEEE INFOCOM* (pp. 1–9).
8. Zhao, Z., Dong, W., & Guan, G. (2015). Modeling link correlation in low-power wireless networks. In *Proceedings of IEEE INFOCOM*.
9. Zhu, T., Zhong, Z., & Zhang, Z.-L. (2013). Achieving efficient flooding by utilizing link correlation in wireless sensor networks. *IEEE/ACM Translation on Networks*, *21*(1), 121–134.
10. Alam, S. I., Sultana, S., Hu, Y. C., & Fahmy, S. (2013). SYREN: Synergistic link correlation-aware and network coding-based dissemination in wireless sensor networks. In *Proceedings of MASCOTS*.
11. Zhao, Z., Dong, W., Bu, J., Gu, T., & Chen, C. (2014). Exploiting link correlation for core-based dissemination in wireless sensor networks. In *Proceedings of IEEE SECON*.
12. Wang, S., Basalamah, A., Kim, S., Guo, S., Tode, Y., & He, T. (2014). Link correlation aware opportunistic routing in wireless networks. *IEEE Transactions on Wireless Communications*, *14*(99), 1–1.
13. Cheng, S. Y., & Li, J. Z. (2009). Sample based  $(\epsilon, \delta)$ -approximate aggregation in sensor networks. In *Proceedings IEEE 29th int'l conference distributed computing systems (ICDCS)* (pp. 273–280).
14. Li, J., & Cheng, S. (2012).  $(\epsilon, \delta)$ -approximate aggregation algorithms in dynamic sensor networks. *IEEE Transactions Parallel and Distributed Systems*, *23*(3), 385–396.
15. Wu, J., Lou, W., & Dai, F. (2006). Extended multipoint relays to determine connected dominating sets in MANETs. *IEEE Transactions on Computers*, *55*(3), 334–347.
16. Guo, S., Kim, S. M., Zhu, T., Gu, Y., & He, T. (2011). Correlated flooding in low-duty-cycle wireless sensor networks. In *Proceedings ICNP* (pp. 383–392).
17. Zhong, Z., Zhu, T., Wang, D., & He, T. (2009). Tracking with unreliable node sequence. In *Proceedings IEEE INFOCOM* pp. (1215–1233).
18. Ben-El-Kezadri, R., Pau, G., & Claveirole, T. (2011). TurboSync: Clock synchronization for shared media networks via principal component analysis with missing data. In *INFOCOM* (pp. 1170–1178).
19. Tile, Y. (2006). *Sampling algorithms*. New York: Springer.
20. Lind, D. A., Marchal, W. G., & Wathen, S. A. (2011). *Basic statistics for business and economics*. Irwin: McGraw-Hill.
21. Bernstein, S., & Bernstein, R. (2004). *Elements of statistics II: inferential statistics*. New York: McGraw-Hill.
22. Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C. (2001). Section 24.3: Dijkstra's algorithm. *Introduction to algorithms*. MIT Press (pp: 595–601).
23. Knuth, D. E. (1977). A generalization of Dijkstra's algorithm. *Information Processing Letters*, *6*(1), 1–5.
24. Thorup, M. (1999). Undirected single-source shortest paths with positive integer weights in linear time. *Journal of the ACM*, *46*(3), 362–394.
25. Xiaoyong Yan, J., Cao, L., Sun, J., Zhou, S. W., & Song, A. (2020). Accurate analytical-based multi-hop localization with low energy consumption for irregular networks. *IEEE Transactions on Vehicular Technology*, *69*(2), 2021–2033.

26. Xiaoyong Yan, L., Sun, J. Z., & Song, A. (2018). DV-hop localisation algorithm based on optimal weighted least square in irregular areas. *Electronics Letters*, *54*(21), 1243–1245.
27. Xiaoyong Yan, L., Sun, Z., Sun, J. Z., & Song, A. (2019). Improved hop-based localisation algorithm for irregular networks. *IET Communications*, *13*(5), 520–527.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Chao Meng** received the Ph.D. degree in computer science from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014. His research interests include wireless network protocol.