IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 25, 2018, accepted March 24, 2018, date of publication April 20, 2018, date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2825441

Evaluation Models for the Nearest Closer Routing Protocol in Wireless Sensor Networks

NING CAO[1,2], PINGZENG LIU[3], GUOFU LI[4], CE ZHANG[5], SHAOHUA CAO[6], GUANGSHENG CAO[1], MAOLING YAN[3], AND BRIJ BHOOSHAN GUPTA[7]

[1]College of Information Engineering, Sanming University, Sanming 365004, China
[2]College of Information Engineering, Qingdao Binhai University, Qingdao 266555, China
[3]College of Information Science and Engineering, Shandong Agricultural University, Tai'an, China
[4]College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China
[5]School of Computer Science and Technology, Harbin Institute of Technology at Weihai, Weihai, China
[6]College of Computer and Communication Engineering, China University of Petroleum, Qingdao 102200, China
[7]Department of Computer Engineering, National Institute of Technology at Kurukshetra, Kurukshetra 136119, India

Corresponding author: Pingzeng Liu (pingzengliu2017@126.com)

This work was supported in part by the Grant Shandong Education Department under Grant J16LN73, in part by the Shandong Independent Innovation and Achievements Transformation Project under Grant 2014ZZCX07106, in part by the Natural Science Foundation of China under Grant 61572022, and in part by the Fujian Province University Key Lab for Industry Big Data Analysis and Application.

ABSTRACT Wireless sensor networks (WSNs) have been pushed to the forefront in the past decade, owing to the advert of the Internet of Things. Our research suggests that the reliability and lifetime performance of a typical application in WSNs depends crucially on a set of parameters. In this paper, we implemented our experiments on the nearest closer protocol with the J-Sim simulation tool. We then analyze the closure relationships among the density, reliability and lifetime, and reveal the trade-off among them based on our analysis on the experiment results. Next, we propose five intelligent evaluation models that are applicable to such situations. Our research allows the WSN users to predict the significant evaluation parameters directly from the settings while costly simulations are no longer necessary.

INDEX TERMS Evaluation model, nearest closer, routing protocol.

I. INTRODUCTION

The next generation networks have the potential to offer heterogeneous connectivity, especially for the IoT systems. When the transmission speed reaches 10Gbps, 10 to 100 times faster than that of the 4G networks, massive types of wireless communications demand can be fulfilled [1]–[3].

IoT systems must allow things to be inexpensive, highly power efficient, ubiquitous, safe and reliable [4]–[6]. Research related to sensors is a significant research point for IoT. The research of Wireless Sensor Networks (WSNs) is a subset research area of IoT. The sensors should be assigned with a routing protocol, so that instead of transmitting data directly to the end-user, they choose to transmit their data via a number of other sensors subject to the condition that the data will eventually arrive at the end-user (multi-hop protocol). 5G can provide a reliable backhaul infrastructure for many IoT systems. In addition, the communication security [7]–[9] and security of sensor networks [10]–[13] are also very important. It is necessary to prevent the monitoring data from being stolen and to obtain forged monitoring information.

The main problems in the primary research area of WSNs are conserving sensor energy [14]–[17] and improving data accuracy [18], [19]. For instance, Raghunathan *et al.* discuss several key factors, including architecture and protocols, which are related to energy-efficient design of sensor nodes in a WSN [20]–[22]. The properties of data security and energy consumptions in wireless networks are highly depend on the physical neighborhood and the transmission power, which makes previous theories for wired networks no longer viable.

Although many of the researches realize that energy consumption and reliability are the important parameters for a WSN, which are also influenced by the features of the network, there have been very few research works focus on the trade-off between the important factors of reliability, lifetime, and node density empirical. They also haven't proposed effective evaluation models among these parameters.

Maity and Gupta [23] considered a randomly distributed wireless sensor network covering a large area. They wished to find an estimate for the number of nodes required with the minimum critical communication distance to ensure network connectivity and stability. Using results from graph theory certain mathematical formula-based algorithms already existed for a relatively small number of sensors; however, the authors proposed a new formula based on mathematical simulation, which minimized the inter-node critical radius prediction for a large number of nodes. They chose MATLAB as their simulation tool and constructed a regression equation between the radius and number of nodes. The theoretical model provided better results if the number of nodes is less than 250, but their regression equation provided smaller answers for the radius to preserve connectivity for larger numbers of nodes. In our research, we will combine three parameters together and propose the evaluation model among these parameters. The model can also provide good results even if the number of nodes is greater than 250.

This research seeks to develop some evaluation models, which will serve to inform the organization of a high efficiency wireless sensor network without prior use of simulations. Such mathematical models will avoid the necessity to undertake simulations for the deployment of sensor nodes resulting in a saving of both time and money.

In this paper, we choose the Nearest Closer protocol as a typical routing protocol for location-based routing protocols [24]–[27]. After obtaining several useful results via simulations, several evaluation models based on the Nearest Closer routing protocol [28] in wireless sensor networks will be proposed.

II. MATERIALS AND METHODS

A. SIMULATION TOOLS

Before the real deployment of sensors to the real area, to do simulation is really significant. From the simulation results, users can conclude how the routing protocols work. After comparisons of simulation results, users can select the most suitable routing protocol for their applications.

There exist several simulation tools, such as: NS-2 [29], Omnet++ [30], TOSSIM [31], J-Sim [32]–[33]. J-Sim will be selected as the simulation tool for this paper. The reason can be explained as follows:

J-Sim is an open-source simulation tool.

J-Sim has provided a strong energy model to the users.

J-Sim has a good user interface and is convenient for the users to invoke the existing methods.

J-Sim is a Java-based simulation tool, it will be possible to combine the energy model with the Java-based sensor's energy model in future design.

The authors of J-Sim have performed detailed performance comparisons in simulating several typical WSN scenarios in J-Sim and NS-2. The simulation results indicate J-Sim and NS-2 incur comparable execution time, but the memory allocated to carry out simulation in J-Sim is at least two orders of magnitude lower than that in NS-2. As a result, while NS-2 often suffers from out-of-memory exceptions and was unable to carry out large-scale WSN simulations, the proposed WSN framework in J-Sim exhibits good scalability.

J-Sim models are easily reusable, so users can combine the components in the framework freely.

B. NEAREST CLOSER PROTOCOL

The Nearest Closer protocol is both a typical location-based routing protocol and a multi-hop routing protocol. Consequently, this work has implemented the protocol in J-Sim as an example. To implement this protocol each node has to know its own position, the position of its neighbors within its transmission range, and the position of the sink node. The main idea in the Nearest Closer protocol is that the transmitter sensor will transmit to its nearest neighbor that is closer to the sink node (the distance between the neighbor and the sink node is less than the distance between the sensor node and the sink node; choose the nearest neighbor from the sensor node).

Slotted ALOHA protocol has been selected as the MAC layer protocol. Slotted ALOHA is a type of TDMA transmission system and it improves contention management through the use of beaconing. Slotted ALOHA can make a single active sensor nearly continuously transmit at full channel rate, thus better results can be obtained for the Nearest Closer protocol.

C. REASON TO SELECT NEAREST CLOSER PROTOCOL

The location-based routing protocols, of which there are a large number, but difficult to be implemented in J-Sim Greedy Perimeter Stateless Routing (GPSR) essentially allows backtracking if a dead end is reached, so within any implementation program for GPSR and other similar routing protocols there must be a large number of nested conditional statements of the 'IF' and 'THEN' form. This means that the behavior of the protocol can fundamentally change during a simulation depending on the geographical and/or power status of the network, so that trends in results obtained from simulations using such protocols will be difficult (if not impossible) to analyze. The choice in this thesis, to avoid such conditionality and the apparent complexity of any resultant mathematical model, came down to Nearest with Forward Progress (NFP) or Nearest Closer (NC) for the work. The routes in NFP along which sensors transmit data to the sink node will be very jagged (with lots of 'ups' and 'downs') compared to those in NC by definition. Thus NC presents, at face value at least, a more sensible approach to transmitting data from a sensor to the sink node and hence our choice of it as a representative for location-based protocols.

D. EXPERIMENTAL SET-UP

The simulated area for the following experiments is defined as a 10 meter by 10 meter square. All the sensor nodes are randomly deployed. The routing protocol is the Nearest Closer. The sink node is in the middle of this simulation area. There exists a fixed target node in each simulation experiment, and this target node will generate a stimulus every one second. All the sensors are active at the start of the simulation (For the NC protocol, the sensor nodes can only transmit its data to the nearest neighbor sensor as the routing map will be generated before data transmission.). The simulation time is enough for obtaining data in each separate experiment. All the points (in the figures) in this paper are the average value from 5 separate experiments. In each experiment the sensors are static once placed, and the sensor nodes are re-positioned randomly for the 5 experiments. The Network Lifetime will be measured in seconds. The maximum simulation time for all experiments in this paper is 100,000 seconds.

III. RESULTS

A. EVALUATION PARAMETERS

1) RELIABILITY

This paper bases its experiment on the concept of Reliability, which is defined as:

$$Reliability = \frac{\text{the number of packets received by the sink node}}{\text{the number of packets sent directly to the sink node}} \quad (1)$$

2) LIFETIME

Network *Lifetime* has various definitions. Some researchers may select the time when the all the sensors run out their energy to define the Lifetime, however, if the key sensors in this WSN are dead, this sensor network will never continue to work anymore. So this is not a good choice for the definition of Lifetime. Some researchers may select the time when the first sensor is dead to define the parameter of Lifetime. But this is not a suitable moment for Lifetime as the sensor network may work well. In this paper, we will select the time-point when the last packet sent from sensors is received by the sink node to define the Lifetime.

3) DENSITY

In a fixed area, the number of sensors deployed can be used to denote the Density.

B. RESULTS AND ANALYSIS

A series of experiments [34] was carried out with the number of sensors starting at 10, and increasing in increments of 10 up to 300 sensors. The transmission radius for each sensor was fixed at 15 meters as this is large enough to transmit data anywhere within the square.

In Figure 1, we observe a clear relationship between the number of sensors and the Reliability, which, in this application, increased as the number of sensors increased to 40, when it reached its highest value. It then essentially decreased as the number of sensors increased from 40 to 300, when it reached its lowest value. This may be explained by observing that: as the Density increases, more sensors will join the data transmission process and consequently communication among the sensors will become more and more complex. So, dropped data due to data collision and latency cannot be ignored. Consequently, it is reasonable to expect the Reliability to decrease with the Density.

In Figure 2, the Lifetime meets its minimum value when the number of sensors equals to 220, whereas the highest value of the Lifetime occurred when there were 10 sensors.

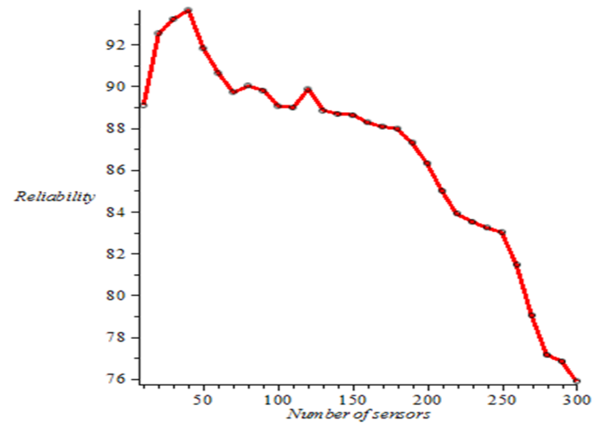


FIGURE 1. Density, reliability relationship for NC.

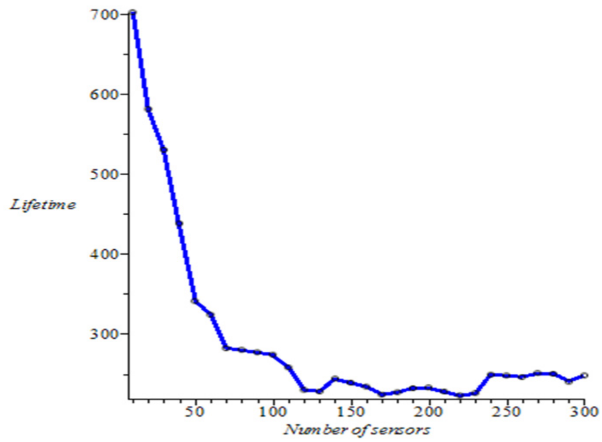


FIGURE 2. Density, lifetime relationship for NC.

In the NC protocol, sending a packet to the sink node will require each sensor to transmit its data down transmission trees to the nearest sensors to the sink node. These nearest sensors (only about two or three) will be receiving all the data in the network and transmitting it all to the sink node, using a large amount of energy. Thus, a NC network will generally die fairly rapidly from the sink node outwards.

In Figure 1, note that when the number of sensors equals 40, the Reliability was 93.64%, its highest level, but with this number the Lifetime is fairly short. This illustrates that it is possible for users to choose an optimum Density value for this application depending on the Reliability and Lifetime required.

IV. BASE EVALUATION MODELS

A. LIFETIME MODEL

The energy requirement of a routing task can be largely approximated by the hop count, under which circumstance we assume a constant metric per hop. However, if nodes can adjust their transmission power (knowing the location of their neighbors) the constant metric can be replaced by a power metric $u(d) = e + d\alpha$ (or some variation of this) for some constant value of α and e that depend on the distance d between nodes. The value of e, which includes energy loss due to start up, collisions, retransmissions, and acknowledgements, is relatively significant, and protocols using any kind of periodic hello messages are extremely energy inefficient.

Our motivation of proposing the NC protocol is based on the consideration that the average energy consumption of a sensor for data to its nearest neighbor can be calculated by a simple formulae $w(s)\,E\left(d^2\right)$, in which $E\left(d^2\right)$ is the expectation of the squared distance between the sensors, and $w(s)$ is the average number (weight) of packets of data to be sent to its nearest neighbor. In the NC protocol one possible parameter that could affect the Lifetime is the number of paths or trees formed by the sensors. This is not the case as demonstrated by the following experiments, which are of independent interest. Using J-Sim it is difficult to extract the paths that sensors transmit along to the sink node. For this reason, a program was written in the algebraic software package Magma to not only compute the individual paths, but to marry them together in order to form the initial distinct trees. During the Lifetime of the network various sensors will die, normally from the center outwards, and new trees will be formed (The sink node is in the middle. Tree roots are close to the sink node and they will die quickly as they do receive and transmit data frequently. In this paper, all the common sensors are randomly deployed and have the same power.). The mean number of trees is shown in Table 1. The effect of randomly placed sensors in each experiment can be ignored as the value was averaged from the results of 1000 experiments.

According to Table 1, we conclude that the average number of trees is independent of the number of sensors, and thus can be discounted as having any effect on the Reliability or Lifetime. The reason that two or three trees may appear on average can be fully explained by considering the diagram below.

In the case of 300 sensors, the number of trees with frequencies was one with frequency 1, two with frequency 42, three with frequency 53 and four with frequency 3. The distributions for two trees took various values from (299, 1) to (152, 148). For the cases of two trees the average length of the two trees for 10 sensors was 6.69 and 3.31 with a sample standard deviation of 1.199. The ratio between these average lengths is 2.02. For 40 sensors the corresponding figures are 26.667 and 13.333 with a sample standard deviation of 4.646. The ratio between these average lengths is 2. For 300 sensors the corresponding figures are 213.929 and 86.071 with a sample std. value of 39.38. The ratio between these average lengths is 2.49.

The reason for using the average tree length as a test statistic is that it can be generalized to more than two trees. In the latter part of this paper it is assumed that the sensors are equally distributed amongst the trees. The results above indicate in the two-tree case that the ratio between the average lengths of the trees is about 2 or 2.5 to 1. The reason for this may be that the shape of the square influences the sensors into one large and one small tree, this merits further investigation from a mathematical or statistical perspective.

TABLE 1. Mean number of trees.

No. of Sensors	Trees
10	2.537
20	2.527
30	2.561
40	2.559
50	2.531
60	2.550
70	2.542
80	2.534
90	2.553
100	2.549
110	2.568
120	2.578
130	2.555
140	2.543
150	2.574
160	2.550
170	2.557
180	2.552
190	2.575
200	2.582
210	2.574
220	2.600
230	2.571
240	2.573
250	2.569
260	2.551
270	2.575
280	2.543
290	2.563
300	2.562

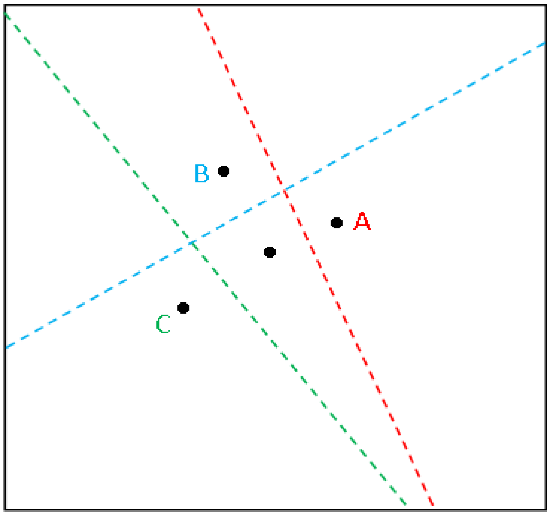


FIGURE 3. Typical tree zones.

For the NC protocol the relation 'is the closest neighbor to' is not symmetric. In Figure 3, B might be the closest neighbor to A, but without A being the closest neighbor to B. Thus, for this protocol it is necessary to find the closest neighbor to a sensor A, but with the stipulation that the neighbor is closer to the sink node than A. It is also the case that the sink node could be the closest neighbor to A, but the sink node is a fixed point and therefore does not fit in with the assumption that this work is dealing with a random distribution of sensors.

The second column of Table 2 lists the expectation of the distance d between a sensor node and its closest neighbor that should be closer to the sink, and the third column shows the expectation of the squared distance d^2. Finally, the fourth column shows the average weight of the sensors, where each branch in the various trees count as weight one. This is a measure of the number of packets of data an average sensor has to transmit to its nearest neighbor, and it is also the average number of hops from the sensors to the sink node. The results of these simulations are tabulated in Table 2 (In Table 2, $E(d)$ denotes the expectation of the distance between sensors, $E\left(d^2\right)$ denotes the expectation of the squared distance between sensors, and $w(s)$ denotes the average number (weight) of packets of data to be sent to its nearest neighbor.)

Note that, for those sensors whose nearest neighbor is the sink node that the distribution of the distance and squared distance to the sink node is slightly different to those between sensors, and the averages are always slightly higher than those given in Table 2, with the exception of the 10 sensor case when there is a significant difference.

We induce two conclusions from this table. First, the value of $w(s)E(d)$ is nearly constant which varies between 5.16 and 5.45, and is a measure of the average distance from a sensor to the sink node measured along the hop path. Second, the value of $w(s)E\left(d^2\right)$ decreases when the number of sensors growth, which shows that the energy expended by a sensor in receiving and sending received packets to its nearest neighbor decreases when the number of sensors increases. It also represents a measure of the average amount of energy

TABLE 2. Expected distances and weights, variable sensors.

Sensors	$E(d)$	$E(d^2)$	$w(s)$
10	2.1945	6.0533	2.3526
20	1.5985	3.2761	3.3421
30	1.3223	2.2444	4.0821
40	1.1512	1.7023	4.7152
50	1.0264	1.3535	5.2218
60	0.9402	1.1350	5.7941
70	0.8672	0.9679	6.2249
80	0.8104	0.8450	6.6531
90	0.7657	0.7545	7.0685
100	0.7256	0.6759	7.4110
110	0.6926	0.6155	7.7859
120	0.6618	0.5634	8.1589
130	0.6360	0.5200	8.4701
140	0.6124	0.4817	8.7956
150	0.5910	0.4485	9.0999
160	0.5724	0.4210	9.4502
170	0.5553	0.3957	9.7218
180	0.5394	0.3734	9.9967
190	0.5240	0.3525	10.3059
200	0.5108	0.3346	10.5684
210	0.4983	0.3183	10.7823
220	0.4874	0.3046	11.0709
230	0.4759	0.2904	11.2847
240	0.4658	0.2785	11.5108
250	0.4562	0.2670	11.8269
260	0.4466	0.2559	12.0056
270	0.4393	0.2475	12.2536
280	0.4308	0.2378	12.4573
290	0.4231	0.2296	12.7054
300	0.4160	0.2216	12.9399

required to transmit one packet of data to the sink node. Two theories are proposed to explain why the Lifetime decreases as the number of sensors goes up.

In general, the Lifetime can be modelled by considering only the roots of the forest, which normally consists of just two or three trees. If there exists just one tree, which is an interesting case for a small number of sensors, the system should run until the only root sensor dies naturally, when the system might interpret the death as a transmission break or the next sensor(s) out will take over the role of the root(s). In either case, the Lifetime is expected to be much longer for just one tree. The basic model constructed here is of the form

$$\frac{E\left(T\right)K}{nE\left(d^2\right)}, \tag{2}$$

where n is the number of sensors, K is the initial energy of a sensor and $E\left(T\right)$ is the expected number of trees. Here $n/E(T)$ represents the average weight per tree and since $E\left(T\right)$ is nearly constant, the model should roughly resemble the curve $1/n$. Particularly, this model assumes that the sensors are evenly distributed amongst the trees and is ignorant of data build-up, so it can be deemed as an upper bound for the Lifetime if the system runs at full capacity. The implication is that the Lifetime decreases monotonically as the number of sensors increases; however, this basic model assumes that the system runs at full capacity from the start and ignores idle slots in the slotted ALOHA design. Thus there is a base value below which the Lifetime will not fall, as the system will behave in a manner akin to a sparser system during its initial phase and there will furthermore always be a number of idle slots. This base value will be taken to be the smallest Lifetime value of 223 that occurs when $n = 220$. So the revised model is:

$$\left(1-\lambda\left(n\right)\right)\times 223+\lambda(n)\times\frac{E\left(T\right)K}{nE\left(d^2\right)}, \tag{3}$$

where $\lambda\left(n\right)$ is a proportionality measure with $\lambda\left(220\right) = 0$. Now from the Lifetime data $\lambda\left(n\right)$ will be very close to 0 if $n \geq 120$. Setting $\lambda\left(10\right) = 1$ yields the value $K = 16,750$.

The parameter $\lambda\left(x\right)$ can be piecewise approximated by

$$1.1438571428571-0.01501071428514x, \quad \text{for } 10\leq x\leq 70, \tag{4}$$

and

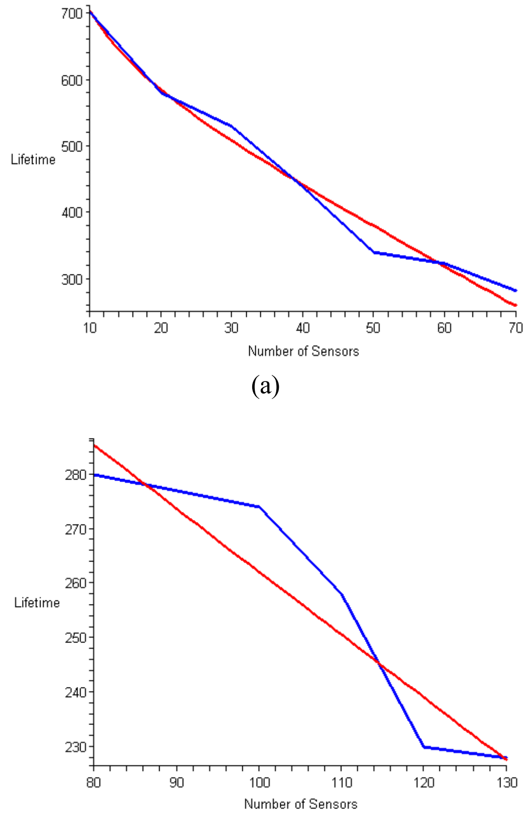
$$0.3954-0.0029514285714286x, \quad \text{for } 80\leq x\leq 130 \tag{5}$$

by fitting straight lines through the given values. An alternative that will yield better local approximations is to interpolate between the known values of λ. Thus, if n≤x<n+10, then

$$\lambda\left(x\right)\cong\left(1-\frac{x-n}{10}\right)\lambda(n)+\frac{x-n}{10}\lambda(10+n) \tag{6}$$

The general equation obtained for $10 \leq x \leq 133$ is

$$\left(1-\lambda\left(x\right)\right)\times 223+\lambda\left(x\right)\times\frac{42712.5}{xE\left(d^2\right)} \tag{7}$$



(a)

(b)

FIGURE 4. (a) Lifetime model (red), lifetime (blue), 10 – 70 sensors; (b) lifetime model (red), lifetime (blue), 80 – 130 sensors.

and the value of $E\left(d^2\right)$ for x sensors can again either be interpolated between the known values in Table 5 or by using the fitted formula

$$\frac{4918141600.24}{6653x^2+68674970x+125057800}, \quad \text{for } 10\leq x\leq 300. \tag{8}$$

Using the latter formula (which is a good fit) yields at worst an implicit equation for the Lifetime for $10 \leq x \leq 130$. Finally, it could be stated that the Lifetime is essentially constant for $x \geq 140$ with a value between 223 and 250 depending on the topology of the sensors.

In both Figure 4 (a) and (b) the Lifetime model is nearly a straight line and is clearly a better fit for $\leq$70 sensors.

B. RELIABILITY MODEL

Using the NC protocol three principal causes for data collision can be identified:

1. Out of range or connectivity: A sensor may have insufficient transmission radius to reach its nearest neighbor at some stage, and from that time all its data (including data that it receives from other sensors) will not reach the sink node. This type of data loss does not occur in the 10 by 10 meter simulated area considered in this paper.

2. Sensor complexity: Collisions may occur at the receiving end, when two or more sensors are sending data to the same nearest neighbor, if they use the same slot or slots. The larger the average weight of a sensor (with the weight being a measure of the packets to be transmitted) the more likely that such collision may occur, since the number of transmission slots increases accordingly.
3. Data collision at the sink node: The data collision at the sink node is also a major concern since different sensors may transmit packets to the sink node simultaneously. From Table 1 the number of trees for this application is around 2.5. In other words, the number of sensors in this application doesn't affect the number of trees very much. The larger the number of sensors, the longer the time required to transmit data from sensors to the sink node. So the sensors will contact the sink node more frequently. The smaller the number of sensors, the more data could be sent to the sink node in one application. Thus, data collision at the sink node is more complex.

The Reliability obtained in the experiments measures the data loss that arises from out of range sensors, sensor complexity and data collision at the sink node. For a large radius there are no out of range sensors, but then it is a question of which out of complexity and data collision at the sink node is the dominant cause of data loss, with the former increasing and the latter decreasing as the number of sensors increases.

When the transmission radius is 15 meters the data collision at the sink node is the dominant factor for less than or equal to 40 sensors and from then on complexity is the dominant factor. This would then explain why the Reliability increases as the number of sensors increases to 40 sensors, and then decreases as the number of sensors increases beyond 40.

There is a known result for the slotted ALOHA design using the Single-hop protocol that is relevant here. This measures the efficiency of the system, that is the long run fraction of successful slots when there are n (assumedly a large value) sensors each with many packets to send. The efficiency is $q\,(1-q)^{(n-1)}$, where each sensor transmits a packet in a slot with probability q. This result will apply at each receiving sensor (and the sink node) individually in the NC protocol although the result will depend on the number of sensors transmitting to the receiving sensor in question. However, trying to marry these results together for the whole system under the NC protocol would involve a detailed analysis of the trees and is certainly impractical to analyze.

We divide the Reliability per unit Lifetime $R\,(n)\,/L\,(n)$ into three sections according to the value range of n, which are: $10 \leq n \leq 120$, $130 \leq n \leq 210$ and $220 \leq n \leq 300$. In the first of these regions the function $R\,(n)\,/L\,(n)$ may be approximated by

$$0.12609 + 0.002226n \tag{9}$$

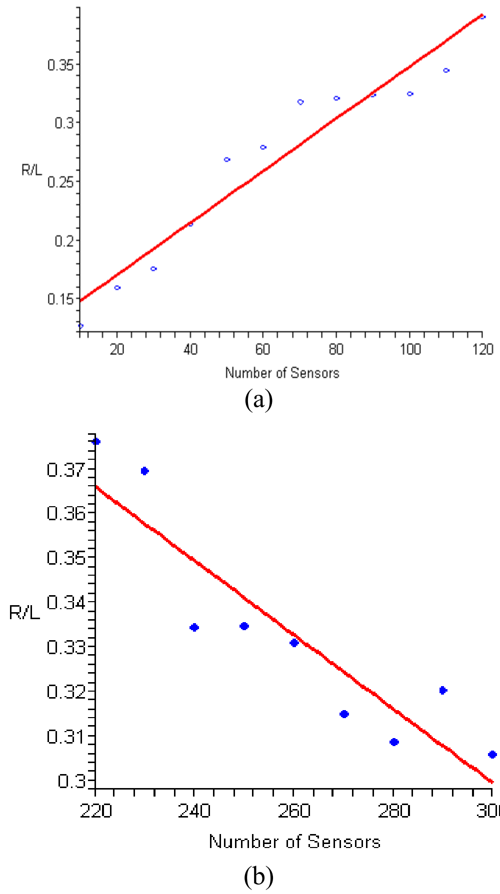


FIGURE 5. (a) R/L model (red), R/L data points (blue), 10 – 120 sensors; (b) R/L model (red), R/L data points (blue), 220 – 300 sensors.

or equivalently since this work has determined L(n) in the previous section

$$R(n) = 0.12609 \times L(n) + 0.002226 \times n \times L(n), \quad \text{for } 10 \leq n \leq 120. \tag{10}$$

This indicates that 12.5% of the Lifetime figure represents a lower bound for the Reliability, but that the Reliability per unit Lifetime will increase by about 2% of the Lifetime as the number of sensors is increased by 10.

In Figure 5, the rate $R\,(n)\,/L\,(n)$ reaches its highest level and stays constant at about 0.3779, so that the Reliability is about 38% of the Lifetime figure for $130 \leq n \leq 210$.

In the last region the rate decreases, this can be satisfactorily explained by the fact that the Lifetime may be considered to be constant in this region and thus increasing the number of sensors just increases the complexity of communication to the sink node. The rate $R\,(n)\,/L\,(n)$ can be approximated by

$$0.549282 - 0.000832666n, \quad \text{for } 220 \leq n \leq 300. \tag{11}$$

So, in this region 55% of the Lifetime is an upper bound for the Reliability, but this decreases by about 0.8% for each extra 10 sensors.

The base evaluation model provides a simple analytic relationship between Lifetime and Reliability. But also due

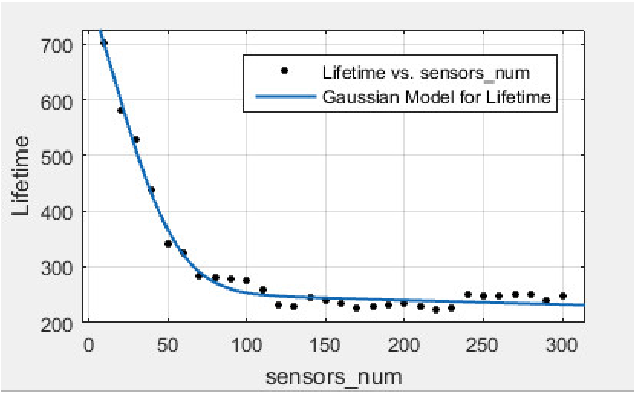


FIGURE 6. **A diagram of gaussian model for lifetime, obtained by the gaussian fitting.**

to its simplicity, the prediction ability is not satisfying, and leaves more space for us to examine more modeling methods.

V. GAUSSIAN EVALUATION MODELS

A. THE GAUSSIAN FITTING

The Gaussian distribution (a.k.a Normal distribution) is a continuous distribution within the exponential family, which is widely used in the scientific and engineering areas. An one-dimensional Gaussian distribution is parameterized by its mean and variance, and is usually denoted as $N(\mu, \sigma^2)$, whose probability density can be expressed as:

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{12}$$

The Gaussian distribution has a lot of good properties, among which the most important one is that it has the maximal entropy for a given variance within the continuous distribution, i.e., it encodes the least prior information about the data generation process. Moreover, the central limit theorem ensures that when a large number of independent random variables are summed up, their normalized sum tends toward a normal distribution, even when their original variables are not normally distributed.

In this section, we choose a more general form of the Gaussian function to fit our observations. The key difference between the Gaussian distribution and this more general Gaussian function is that the Gaussian distribution has to be rescaled by a multiplier, commonly known as the partition constant Z, to ensure it integrate to one. When our observed data are not bounded to follow a probability distribution, we can express the Gaussian function in this more general form:

$$G_i(x) = A_i \times \exp\left(\frac{(x - B_i)^2}{C_i^2}\right) \tag{13}$$

Fitting a Gaussian function involves finding the best values for parameters A, B, and C that best fit the data we observed. A technique called Maximum Likelihood Estimation (MLE) is most widely used if the fitted function is a probability density, which involves constructing the likelihood function of product form and transform it into sum of log. Here, we are fitting a Gaussian distribution with a similar idea, which can be analogous to fitting a polynomial equation, but using Gaussian function instead.

Assuming a dataset contains a list of number pairs (x_i, y_i), for $i = 1, 2, \ldots, N$, which can be described by a Gaussian function:

$$y_i = y_{\max} \times \exp\left(-\frac{(x_i - x_{\max})^2}{S}\right) \tag{14}$$

where $x_{\max}, y_{\max}$, and S denote the peak-height, peak-position, and half-width of the Gaussian curve respectively, which are the three parameters of interest. Then, we can turn both sides of the equation into the logarithm scale, and get the following equation:

$$\begin{aligned} \ln y_i &= \ln y_{\max} - \frac{(x_i - x_{\max})^2}{S} 5 \\ &= \left(\ln y_{\max} - \frac{x_{\max}^2}{S}\right) + \frac{2x_i x_{\max}}{S} - \frac{x_i^2}{S} \end{aligned} \tag{15}$$

Now, let

$$z_i = \ln y_i, \quad b_0 = \ln y_{\max} - \frac{x_{\max}^2}{S}, \; b_1 = \frac{2x_{\max}}{S}, \; b_2 = -\frac{1}{S} \tag{16}$$

Thus, we can reduce the problem of fitting a Gaussian distribution to the problem of fitting a polynomial equation.

$$z_i = b_0 + b_1 x_i + b_2 x_i^2 = \begin{bmatrix} 1 & x_i & x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \tag{17}$$

B. GAUSSIAN MODEL FOR LIFETIME

Based on the Gaussian model, we first experiment with the Lifetime value with respective to the number of sensors. In our baseline evaluation model, we break the whole function between the number of sensors and Lifetime into two parts. However, we model this by simply creating a hard cut in the domain, and fit the two separate sets of parameters for each. A more sensible way is to introduce a ''mixture'' model that consists of two sub-components, whose domains are both the entire interval of the whole function, both contributing to the final outcome by adding-up together their individual values. When the Gaussian function is chosen, we have a well-known Gaussian Mixture Model (GMM). According to the definition of the Gaussian fitting, the function family our fitting procedure searches through is in the following form:

$$y = a_1 \times \exp\left(-\left(\frac{x - b_1}{c_1}\right)^2\right) + a_2 \times \exp\left(-\left(\frac{x - b_2}{c_2}\right)^2\right) \tag{18}$$

After fitting the data subject to the confidence interval of 95%, we get

$$y = 525.2\exp\left(-\left(\frac{x+1.896}{0.5519}\right)^2\right) + (1.504e+17) \times \exp\left(-\left(\frac{x+1464}{251}\right)^2\right) \quad (19)$$

where x is normalized by mean 155 and std. 88.03. Then, several significant parameters can be obtained as follows:

Goodness of fit:

SSE:	6524,
R-square:	0.9835,
Adjusted R-square:	0.98,
RMSE:	16.49.

From the points of the Lifetime, we find that these points form a trend of decline in straight lines with the increase of the number of sensors when the number of sensors is less than 50, which shows that the Lifetime is more sensitive to the changes in the number of sensors in this phase.

When the number of sensors ranges from 50 to 100, the speed of the Lifetime drop slows down, implying that the sensitivity of the Lifetime on the number of sensors starts to decrease, but its negative impact has not yet vanished. When the number of sensors is greater than or equal to 100, the Lifetime curve tends to keep horizontal with slight negative slope, until the function gradually converges to a value between 200 and 230, meaning that the number of sensors no longer has impact on the Lifetime. Telling from this curve, we may induce the conclusion that the Lifetime will continue to keep stable at this value, when the number of sensors is above 300.

C. GAUSSIAN MODEL FOR RELIABILITY

Similarly, we also assume that the function between the number of sensor and the Reliability is a mixture of two Gaussian functions, which can be written as:

$$y=a_1\times\exp\left(-\left(\frac{x-b_1}{c_1}\right)^2\right)+a_2\times\exp\left(-\left(\frac{x-b_2}{c_2}\right)^2\right) \quad (20)$$

Subject to the 95% confidence bounds, the fitting result is

$$y = 7.26\times\exp\left(-\left(\frac{x+1.417}{0.3511}\right)^2\right) + 89.12 \times \exp\left(-\left(\frac{x+0.4427}{5.132}\right)^2\right), \quad (21)$$

where x is normalized by mean 155 and std. 88.03. The fitting result gives us a diagram shown in Figure 8, as well as an analytic result shown below:

Goodness of fit:

SSE:	8.779,
R-square:	0.9871,
Adjusted R-square:	0.9844,
RMSE:	0.6048.

After study the experimental diagram, we observe that the Reliability shows a trend of increase with a linear proportion to the number of sensors when the number of sensors is less than 40, and reaches its maximum value when the number of sensors is 40, which implies that the number of sensors plays a positive role in Reliability. When the number of sensors is between 40 and 100, this number starts to play a negative role in Reliability, which means that more sensors will cause the decrease of the Reliability. When the number of sensors is between 100 and 150, the curve of Reliability tends to keep constant, whose value mostly lies in the range between 87 and 90. This phenomenon proves that the number of sensors is the main factor affecting Reliability. Therefore, we can predict that Reliability will keep declining trend when the number of sensors is over 300.

VI. MULTIVARIATE EVALUATION MODEL

In this part of the work, we use a three-dimensional polynomial equation to model the relationship with regard to Reliability, Lifetime and the number of sensors, using a surface function in the form of:

$$\begin{aligned} f(x,y) = p_{0,0} + p_{1,0}x + p_{0,1}y \\ + p_{2,0}x^2 + p_{1,1}xy + p_{0,2}y^2 \\ + p_{3,0}x^3 + p_{2,1}x^2y + p_{1,2}xy^2 \end{aligned} \quad (22)$$

Subject to the confidence interval of 95%, we obtain that:

$$\begin{aligned} p_{00} &= 1.329e+04\ (-4.774e+04,\ 7.433e+04) \\ p_{10} &= -298.4\ (-645.5,\ 48.79) \\ p_{01} &= -253.3\ (-1581,\ 1074) \\ p_{20} &= 0.4937\ (0.05945,\ 0.928) \\ p_{11} &= 5.523\ (-1.468,\ 12.51) \\ p_{02} &= 1.272\ (-5.942,\ 8.486) \\ p_{30} &= -0.0002371\ (-0.0003863,\ -8.791e-05) \\ p_{21} &= -0.004449\ (-0.008888,\ -1.006e-05) \\ p_{12} &= -0.02606\ (-0.0611,\ 0.008984) \end{aligned}$$

Goodness of fit:

SSE: 8743

R-square: 0.9778

Adjusted R-square: 0.9694

RMSE: 20.4

Meanwhile, we can also come up with a set of three-dimensional distribution diagrams, as shown in Figure 8.

In Figure 8 (a), we notice that the Lifetime and Reliability show a trend of decline in straight lines and relevant function becomes convergent with the increase of the number of sensors. From the picture, we can screen out the range of the number of sensors to achieve a compromise between Lifetime and Reliability. For example, when Lifetime is greater than 260 and Reliability is greater than 85, the range of the number of ranges from 50 to 100.

Mapping a surface to a two-dimensional direction will get the distribution diagram of any two elements. In Figure 8(b),

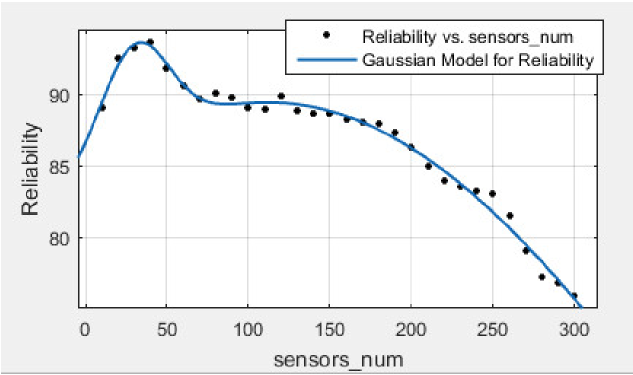


FIGURE 7. A diagram of gaussian model for reliability, obtained by the gaussian fitting.

we see that 40 is a key point for the number of sensors. In Figure 8 (c), when the number of sensors is between 20 and 60, Reliability shows an insensitive phenomenon. However, when the number of sensors is over 60, there is a significant change in Reliability, that is, positive in pre-stage and negative in post-stage.

The overall analysis shows that the number of sensors affects the performance of Lifetime and Reliability directly. The number of sensors less than 50 is an important indicator of building a sensor system, which ensures higher Lifetime and Reliability.

This approach combines three significant evaluation parameters together. This is great contribution of this paper as we can discover the number of sensors if we set both a good value of Lifetime and Reliability. According to the above evaluation model, the suitable number of sensors can be obtained. Thus, the simulation will not be necessary before deploying real sensors.

VII. CRITIQUE

The main drawbacks of the approach can be explained as follows:

The first disadvantage is that there are no experiments on real sensors within this paper. The reason why such experiments have not been undertaken is due to the high cost for the large number of sensors that would be required. All the equations obtained in this work are based on simulation results. There probably exist differences between simulated and real results, which can only be rectified by performing some experiments on a test bed of real sensors.

Some real-life applications, such as detection of temperature and moisture in a vineyard, would have sensors placed rather than randomly positioned. The simulations in this thesis do not cover this scenario, although placed sensors should generally perform better than randomly distributed one from the viewpoint of the user.

The setting of the simulation area is fixed and the sink node is located in the middle of this area. The reason for this setup is that it has military (and other) applications, such as a warship at sea (the sink node) deploying sensors to detect

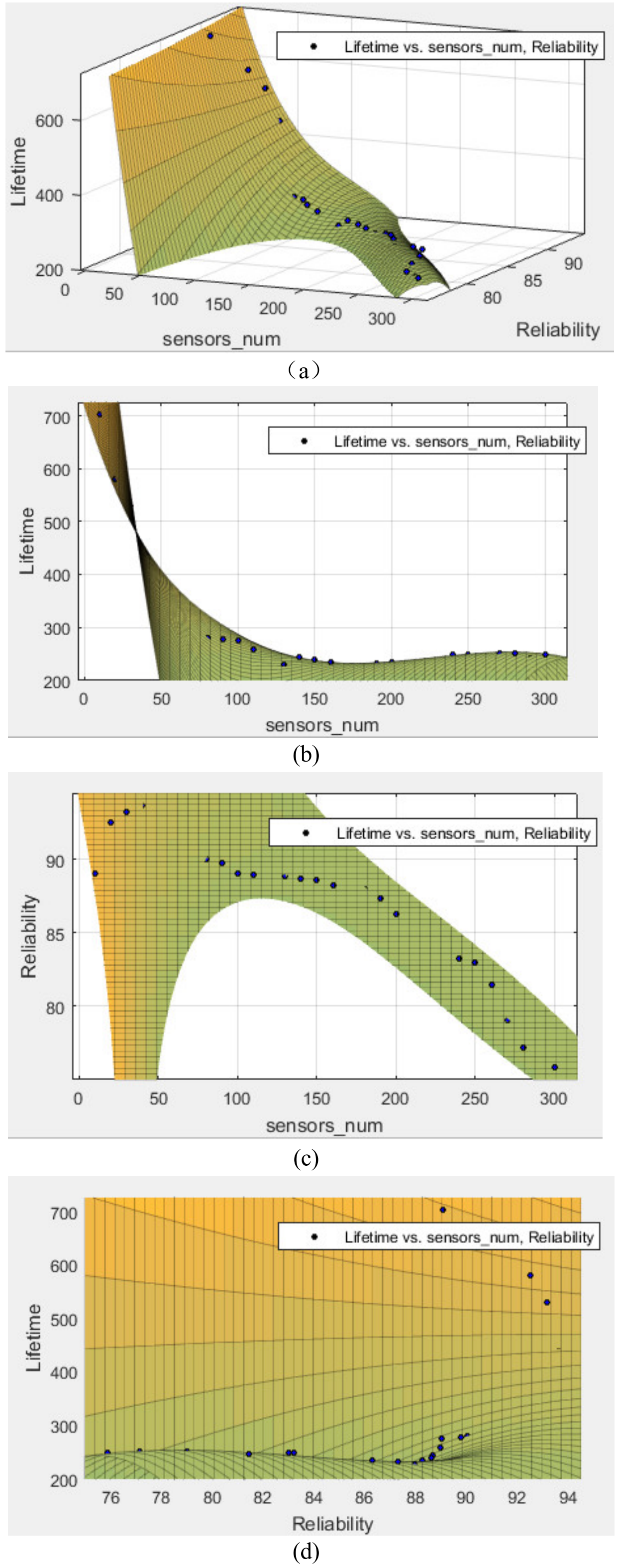


FIGURE 8. A three-dimensional distribution diagram obtained by fitting the polynomial multivariate model.

another vessel. Another obvious choice for the position of the sink node is in one of the corners, but placing the sink node in the center utilizes the symmetry of the square deployment area. So, the equations obtained in thesis have limitations as well.

There may be limitations to which the results are scalable, if the network area was increased to a 100 by 100 meter-square, the results for a small number of sensors in such a large area might demonstrate some new behavior.

VIII. CONCLUSIONS

In this paper, deriving from a series of experiments undertaken in the J-Sim simulation tool several conclusions emerge for NC:

The Reliability for this application increases as the number of sensors increases to 40, when it reached its highest value. It then essentially decreases as the number of sensors increases from 40 to 300.

The Lifetime decreases as the number of sensors increases to 130, and then it fluctuates slightly as the number of sensors increases to 300. The Lifetime reaches its lowest value when the number of sensors is 220, whereas the highest value of the Lifetime occurs when there exist 10 sensors.

Base evaluation models have been proposed in this paper. It provides a simple relationship between Lifetime and Reliability. However, the prediction ability of it is limited largely due to its simplicity.

Gaussian fitting models have been proposed in this paper. These models show significantly better fit result than our base models. Based on these equations, the parameters of Lifetime and Reliability can be predicted directly without further simulations.

A multivariate evaluation model has been constructed in this paper. Three significant parameters, Lifetime, Reliability and Density respectively, are combined in this evaluation model. Based on this evaluation model, the suitable number of sensors can be obtained if the users set the value of Reliability and Lifetime.

ACKNOWLEDGMENT

The authors would also need to say thanks for the assistance from Prof. Chaoyi Pang, who is working at Hebei University of Economics and Business.

REFERENCES

[1] H. S. Maghdid, I. A. Lami, K. Z. Ghafoor, and J. Lloret, "Seamless outdoors-indoors localization solutions on smartphones: Implementation and challenges," *ACM Comput. Surv. (CSUR)*, vol. 48, no. 4, 2016, Art. no. 53.

[2] H. S. Maghdid, A. Al-Sherbaz, N. Aljawad, and I. A. Lami, "UNILS: Unconstrained indoors localization scheme based on cooperative smartphones networking with onboard inertial, Bluetooth and GNSS devices," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Savannah, GA, USA, Apr. 2016, pp. 129–136.

[3] I. A. Lami, H. S. Maghdid, and T. Kuseler, "SILS: A smart indoors localisation scheme based on on-the-go cooperative smartphones networking using on-board Bluetooth, WiFi and GNSS," in *Proc. ION GNSS+*, Tampa, FL, USA, 2014, pp. 503–509.

[4] J. Li, Y. Zhang, X. Chen, and Y. Xiang, "Secure attribute-based data sharing for resource-limited users in cloud computing," *Comput. Secur.*, vol. 72, pp. 1–12, Jan. 2018.

[5] P. Li *et al.*, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generat. Comput. Syst.*, vol. 74, pp. 76–85, Sep. 2017.

[6] P. Li, J. Li, Z. Huang, C.-Z. Gao, W.-B. Chen, and K. Chen, "Privacy-preserving outsourced classification in cloud computing," *Cluster Comput.*, pp. 1–10, Apr. 2017.

[7] Z. Huang, S. Liu, X. Mao, K. Chen, and J. Li, "Insight of the protection for data security under selective opening attacks," *Inf. Sci.*, vols. 412–413, pp. 223–241, Oct. 2017.

[8] J. Li, X. Chen, M. Li, J. Li, P. P. C. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 6, pp. 1615–1625, Jun. 2014.

[9] J. Li, X. Huang, J. Li, X. Chen, and Y. Xiang, "Securely outsourcing attribute-based encryption with checkability," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 2201–2210, Aug. 2014.

[10] J. Li, J. Li, X. Chen, C. Jia, and W. Lou, "Identity-based encryption with outsourced revocation in cloud computing," *IEEE Trans. Comput.*, vol. 64, no. 2, pp. 425–437, Feb. 2015.

[11] J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1206–1216, May 2015.

[12] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannidis, "Secure multiple amplify-and-forward relaying with cochannel interference," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1494–1505, Dec. 2016.

[13] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannidis, "Secrecy cooperative networks with outdated relay selection over correlated fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7599–7603, Aug. 2017.

[14] M. Garcia, S. Sendra, J. Lloret, and A. Canovas, "Saving energy and improving communications using cooperative group-based wireless sensor networks," *Telecommun. Syst.*, vol. 52, no. 4, pp. 2489–2502, 2013.

[15] C.-F. Wang, J.-D. Shih, B.-H. Pan, and T.-Y. Wu, "A network lifetime enhancement method for sink relocation and its analysis in wireless sensor networks," *IEEE Sensors J.*, vol. 14, no. 6, pp. 1932–1943, Jun. 2014.

[16] Y. Yang, M. I. Fonoage, and M. Cardei, "Improving network lifetime with mobile wireless sensor networks," *Comput. Commun.*, vol. 33, no. 4, pp. 409–419, 2010.

[17] X. Wang, J.-J. Ma, S. Wang, and D.-W. Bi, "Prediction-based dynamic energy management in wireless sensor networks," *Sensors*, vol. 7, no. 3, pp. 251–266, 2007.

[18] D. P. Man, G. D. Qin, W. Yang, W. Wang, and S. C. Xuan, "Improved DV-Hop algorithm for enhancing localization accuracy in WSN," *Appl. Mech. Mater.*, vols. 543–547, pp. 3256–3259, Mar. 2014.

[19] S. Phu, M. Tran, and T. A. Yang, "Evaluations of target tracking in wireless sensor networks," in *Proc. ACM 37th SIGCSE Tech. Symp. Comput. Sci. Edu.*, Austin, TX, USA, Mar. 2006, vol. 38. no. 1, pp. 97–101.

[20] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 40–50, Mar. 2002.

[21] A. Sharma and V. Goyal, "Performance analysis during transmission in WSN with distributed energy efficient clustering by modified threshold protocol," *Int. J. Electron. Commun. Comput. Eng.*, vol. 8, no. 1, 2017.

[22] J. Sudharsana and M. R. N. Arumugam, "A survey on energy efficient sensor node design for IoT applications," *Int. J. Comput. Math. Sci.*, vol. 6, no. 8, 2017.

[23] C. Maity and A. Gupta, "Critical communication radius prediction with random distributed nodes in wireless sensor network," in *Proc. Annu. Seminar C-DAC Noida Technol.*, New Delhi, India, 2010, pp. 31–38.

[24] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. ACM 6th Annu. Int. Conf. Mobile Comput. Netw.*, Boston, MA, USA, Aug. 2000, pp. 243–254.

[25] A. A. Kadhim and M. W. Abdulrazzaq, "Efficient routing techniques for wireless sensor networks," *J. Appl. Sci.*, vol. 14, no. 24, pp. 3479–3485, 2014.

[26] B. Musznicki, M. Tomczak, and P. Zwierzykowski, "Dijkstra-based localized multicast routing in wireless sensor networks," in *Proc. IEEE Commun. Syst., Netw. Digit. Signal Process.*, Jul. 2012, pp. 1–6.

[27] H. Frey, S. Rührup, and I. Stojmenović, "Routing in wireless sensor networks," in *Guide to Wireless Sensor Networks*. Springer, 2009, pp. 81–111.

[28] I. Stojmenovic and X. Lin, "Power-aware localized routing in wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 11, pp. 1122–1133, Nov. 2001.

[29] I. T. Downard, "Simulating sensor networks in NS-2," Naval Res. Lab., Washington, DC, USA, Tech. Rep. NRL/FR/5522-04-10073, 2004.

[30] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. ACM 1st Int. Conf. Simulation Tools Tech. Commun., Netw. Syst. Workshops (ICST)*, Marseille, France, Mar. 2008, Art. no. 60.

[31] P. Levis, N. Lee, M. Welsh, and D. Culler, "TOSSIM: Accurate and scalable simulation of entire TinyOS applications," in *Proc. ACM 1st Int. Conf. Embedded Netw. Sensor Syst.*, New York, NY, USA, 2003, pp. 126–137.

[32] A. Sobeih *et al.*, "J-Sim: A simulation and emulation environment for wireless sensor networks," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 104–119, Aug. 2006.

[33] A. Sobeih *et al.*, "J-Sim: A simulation environment for wireless sensor networks," in *Proc. 38th Annu. Symp. Simulation*, Washington, DC, USA, Apr. 2005, pp. 175–187.

[34] N. Cao, R. Higgs, G. M. P. O'Hare, and R. Wu, "Lifetime and reliability evaluation models based on the nearest closer protocol in wireless sensor networks," in *Proc. IEEE 3rd Int. Conf. Wireless Sensor Netw.*, Warsaw, Poland, Sep. 2014, pp. 993–1000.

NING CAO received the B.S. degree in software engineering from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer science from University College Dublin in 2015. He is an Academic Leader with the College of Information Engineering, Sanming University, and with the College of Information Engineering, Qingdao Binhai University. He has published over 40 papers. His research interests include IoT and wireless sensor networks.

PINGZENG LIU is currently a Professor with the College of Information Science and Engineering, Shandong Agricultural University. He is also the Director of the Center for Agricultural Big Data Research. He successively published over 70 papers, applied for over 40 patents, and obtained 26 software copyrights. His research interests include the Internet of Things architecture, agricultural Internet of Things promotion and application, and agricultural big data processing.

GUOFU LI received the B.S. degree in software engineering from Fudan University in 2007 and the Ph.D. degree in computer science and informatics from University College Dublin in 2014. He is currently a Lecturer with the College of Communication and Art Design, University of Shanghai for Science and Technology. His research interests include digital communication, machine learning, and natural language processing.

CE ZHANG born in 1978. He received the Ph.D. degree in computer architecture from the Harbin Institute of Technology, Harbin, China, in 2015. He is currently with the Harbin Institute of Technology at Weihai, China. His research interests include stochastic process, reliability modeling, and trusted computing.

SHAOHUA CAO is currently an Associate Professor with the College of Computer and Communication, China University of Petroleum. His main research interests include: network performance optimization, software-defined network, cloud computing, IoT, Internet technology, computer hardware technology, computer software, and computer applications.

GUANGSHENG CAO is currently a Lecturer with the College of Information Engineering, Qingdao Binhai University. He is also a Senior Engineer in Software Engineering. His research interests include computer communication technology, wireless mobile communications applications, and data security technology.

MAOLING YAN is currently pursuing the degree with the College of Information Science and Engineering, Shandong Agricultural University. Her major is computer application technology. Her research interests include the application of agricultural Internet of Things and agricultural big data processing technology.

BRIJ BHOOSHAN GUPTA is currently an Assistant Professor with the National Institute of Technology at Kurukshetra, Kurukshetra, India. His research interests include cyber security, intrusion detection, IoT, cloud security, mobile security, Web Security, and big data analytics.