Front. Electr. Electron. Eng. China 2011, 6(3): 481–485
DOI 10.1007/s11460-011-0163-x

RESEARCH ARTICLE

Xi CHEN, Xingshi WANG

An enhanced probabilistic scheme for data transmission in large-scale sensor networks

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

Abstract In this paper, a probabilistic scheme is presented for directed data transmission without maintaining route tables. In the model, each message is required to reach the base station (BS) successfully with a certain probability. We analyze the relationship between the number of the intermediate nodes, link reliability and relay probability. We obtain the condition for relay probability which can guarantee the performance of the networks. This scheme is robust and adaptable to the change of topology of the sensor networks. Simulation with Ns-2 helps to illustrate the main results of the analysis.

Keywords wireless sensor networks, relay probability, analysis, simulation

1 Introduction

Large-scale sensor networks can be used in many applications such as military sensing, environment monitoring and traffic surveillance, etc. Sensor node has limited energy. The energy consumption for communication is proportional to k $(2 \leqslant k \leqslant 4)$ power of the transmission distance. Communication is considerably more expensive than computation [1]. To save energy, short distance multiple hop communication becomes preferable to long distance direct communication in sensor networks. Data packet is transmitted from its source node to the base station (BS) via relaying by intermediate nodes. Usually, there exist multiple routes from the source to the BS.

Data transmission protocols for sensor networks may be roughly classified into two categories according to routing. In one category, transmission route is predetermined for packets and intermediate nodes have to relay packets from the source to the BS according to the route tables. Route tables are determined by the topology of the sensor networks and need to be updated while topology changes, for example, LEACH [2] lets each node maintain its route table by communicating with its neighboring nodes and such maintenance may consume more energy. In the other category, packets are transmitted without any specific route. Both Flooding and Gossiping are such kind of protocols. They do not require a packet to follow any specific route to the base so that route table becomes unnecessary.

For sensor networks, routing protocols should be distributed, low energy consumption, and be able to cope with frequently changing network topologies [1]. Meanwhile, each routing protocol has to tolerate packet loss, which may be due to bad radio communication, congestion, packet collision, full memory capacity, and node failures [3].

A gossiping-based approach is presented in Ref. [4], where each node forwards a packet with some probability. Gossiping-based approach can also be combined with various optimizations of flooding to reduce the overhead of the routing protocols. This work first introduced the forward probability into routing protocols.

Reference [5] presents a selective forwarding probability. It selects neighbors as the next hop with some probability. The selective forwarding probability is based on the node degree and link loss to increase the reliability of selective forwarding.

Reference [6] proposes a family of light-weight and robust multi-path routing protocols in which an intermediate sensor decides to forward a message with a probability which depends on various parameters, such as the distance of the sensor to the destination, the distance of the source sensor to the destination, or the number of hops a packet has already traveled.

A probabilistic forwarding (ProFor) approach for directed data transmission without maintaining route tables is presented in Ref. [7]. With ProFor, each message may not reach the BS for sure but with a predefined

Received March 18, 2011; accepted May 10, 2011

Xi CHEN (✉), Xingshi WANG
Department of Automation, Tsinghua University, Beijing 100084, China
E-mail: bjchenxi@mail.tsinghua.edu.cn

success probability. There are many reasons for such relaxation. First, in large-scale sensor network, data from different sensor nodes may be similar or duplicated; absence of any small part can hardly seriously affect the performance of the large-scale sensor networks. Second, in some real applications, a high success probability is acceptable and good enough. Third, the cost of demanding absolute certainty is usually unaffordable. With ProFor, intermediate nodes relay messages with a certain relay probability. ProFor requires little information of the global networks and works robustly against individual node failure.

However, ProFor ignores packet loss in transmission which inherently exists for radio frequency communication. In this paper, following the basic idea of ProFor, we present an enhanced probabilistic scheme (EPro) for data transmission. Different from the work in Ref. [7], we take into account the packet loss. By theoretical analysis, we obtain the conditions for relay probability which can adjust to packet loss and explore asymptotic property of the networks. Notice that in Refs. [4–6], all probabilistic approaches are used in a heuristic manner and neither of them has obtained the analytical relationship between the forwarding probability and the successful transmission probability.

The rest of the paper is organized as follows. The problem is formulated in Sect. 2. Relay probability and asymptotic property of the networks are presented in Sect. 3. Simulation results are recorded in Sect. 4. Section 5 concludes this paper.

2 Problem formulation

In our model, we assume that sensor nodes are densely distributed in the area of interest. Without loss of generality, we suppose there is only stationary BS in the area of interest. After the networks are deployed, each node no longer moves and networks proceed to initialization stage in which nodes get their gradients. A node's gradient is defined as the shortest path to the BS which is represented by the number of hops. Moreover, each node can generate a random number as its ID since the probability of different node sharing a common ID is very low. We have the following definitions.

Definition 1 A node is h-hop node if the node's gradient is h.

Definition 2 A packet is h-hop packet if its source is an h-hop node.

Definition 3 If an h-hop node and an $(h$–$1)$-hop node can communicate, then the $(h$–$1)$-hop node is called a 1-hop downstream neighbor of the h-hop node and the h-hop node is called a 1-hop upstream neighbor of the $(h$–$1)$-hop node.

Theoretically, if an h-hop node broadcasts a packet to its 1-hop downstream neighbor, its neighbor should receive the packet successfully. However, in reality, communication with radio frequency is sensitive to many environmental factors. Thus, its neighbor may not receive the packet or does receive the packet but with error codes. We assume that sensor node simply drops such corrupted packet as if it does not receive the packet at all. Let $1-q$ $(0<1-q<1)$ be packet loss probability between neighboring nodes. For simplicity, we assume $1-q$ is same for all nodes in the networks.

The network performs an initialization to generate the gradient for each node by following a similar way as presented in Ref. [8]: the BS initiates a gradient by sending its neighbors, with radio radius R_0, a message with a count which is set to 1. Each recipient remembers the value of the count and, with radio radius R_0, forwards one new message to its neighbors; the new message consists of the count incremented by 1 and the ID of the sending node. Hence a wave of messages propagates outwards from the BS. Each node maintains the smallest count it receives. For example, the smallest hop count value of Node i, h_i, will eventually be the length of the shortest path to the BS in communication hops. The value of h_i is called the gradient of Node i (with respect to the BS) and it roughly indicates the distance between Node i and the BS. According to all messages it receives, Node i can obtain the number of its 1-hop downstream nodes. Of course, packet loss may happen during the transmission of gradient establishment packets. Packet loss may let a node's number of downstream nodes be smaller than the real one. This difference can only make EPro a little more conservative. Figure 1 shows a network in which each circle represents a sensor node and the number in it is the node's gradient after the initialization stage.

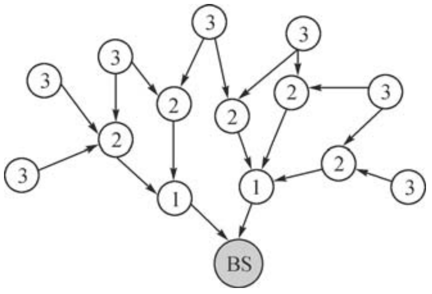

Fig. 1 Gradients of sensor nodes

3 Analysis of relay probability

In this section, we first derive the condition for relay probability. Then we discuss the asymptotic property of the number of nodes which relay packets. At last, we present a variation of EPro to save more energy and reduce the number of duplicated packets.

3.1 Sufficient condition for relay probability

We analyze the probability with which intermediate nodes relaying packets to guarantee the success probability P^*. We assume $P^* \leqslant q$. If there is any error code in a packet, a sensor will simply drop it.

Case when *SourceNode_Gradient* = 1 It means the source node is one hop distant from the BS. When the source node broadcasts the packet, the BS can receive it with probability q. Since $P^* \leqslant q$, no more action is necessary.

Case when *SourceNode_Gradient* = 2 This is a 2-hop packet. Its source is two hops distant from the BS so that relaying by intermediate 1-hop node(s) becomes indispensable.

Suppose the source node has K_1 ($K_1 \geqslant 1$) 1-hop downstream nodes. If each of them re-broadcasts the message with probability p_1, as $1-q$ is packet loss probability, the probability of this message arriving in the BS is $1-(1-p_1q)^{K_1}$. Let $1-(1-p_1q)^{K_1} \geqslant P^*$, then we have $1 \geqslant p_1q \geqslant 1-\sqrt[K_1]{1-P^*}$, such that

$$p_1 \geqslant \frac{1-\sqrt[K_1]{1-P^*}}{q}. \tag{1}$$

Furthermore, we set

$$\alpha_1 = 1-(1-p_1q)^{K_1}. \tag{2}$$

Equation (1) stipulates the condition for the relay probability for 1-hop nodes. In fact, this condition is dependent on K_1.

Case when *SourceNode_Gradient* = 3 It means the source is three hops distant from the BS. To reach the BS, the 3-hop packet needs 2-hop nodes and 1-hop nodes to relay it.

Suppose the source node has K_2 2-hop nodes as its 1-hop downstream nodes which are Node 1, Node 2, $\ldots$, Node K_2. Each of them will relay the message with probability p_2. Suppose Node j ($1 \leqslant j \leqslant K_2$) has $K_1^{(j)}$ 1-hop nodes as 1-hop downstream nodes, respectively. Provided Node j sends out the packet, its $K_1^{(j)}$ 1-hop downstream nodes will relay the message with probability $p_1^{(j)}$ which satisfies Eq. (1). Hence we have

$$p_1^{(j)} \geqslant \frac{1-\sqrt[K_1^{(j)}]{1-P^*}}{q}. \tag{3}$$

Then the probability for the packet successful arriving at the BS via Node j is $\alpha_1^{(j)}p_2q$, where $\alpha_1^{(j)} = 1-(1-p_1^{(j)}q)^{K_1^{(j)}}$. The probability for the message not arriving at the BS is $\prod_{j=1}^{K_2}(1-\alpha_1^{(j)}p_2q)$. Therefore, the probability for this 3-hop packet arriving at the BS is $1-\prod_{j=1}^{K_2}(1-\alpha_1^{(j)}p_2q)$ and p_2 should satisfy

$$1-\prod_{j=1}^{K_2}(1-\alpha_1^{(j)}p_2q) \geqslant P^*, \tag{4}$$

$$\alpha_1^{(j)} = 1-(1-p_1^{(j)}q)^{K_1^{(j)}}. \tag{5}$$

By Eq. (2), we know $\alpha_1^{(j)} \geqslant P^*$, $j=1,2,\ldots,K_2$, then $\prod_{j=1}^{K_2}(1-\alpha_1^{(j)}p_2q) \leqslant (1-P^*p_2q)^{K_2}$. In Eq. (4), we replace $\prod_{j=1}^{K_2}(1-\alpha_1^{(j)}p_2q)$ with $(1-P^*p_2q)^{K_2}$, then we get $1 \geqslant 1-(1-P^*p_2q)^{K_2} \geqslant P^*$. Thus

$$p_2 \geqslant \frac{1-\sqrt[K_2]{1-P^*}}{P^*q}. \tag{6}$$

p_2 in Eq. (6) is a little more conservative than in Eq. (4). However, p_2 in Eq. (6) no longer hinges on $\{p_1^{(j)},\ j=1,2,\ldots,K_2\}$ and is determined merely by K_2. Equation (6) provides a localized solution for p_2.

In general, we can obtain the constraint of relay probability for m-hop nodes.

Case when *SourceNode_Gradient* = *m*+1 Suppose there are K_m m-hop nodes that get involved in relaying the (m+1)-hop message and the relay probability is p_m. By similar reasoning and applying the same trick of removing the dependence of p_m on $\{p_{m-1}^{(j)},\ j=1,2,\ldots,K_m\}$ which are carried out in the case when *SourceNode_Gradient* = 3, we have

$$p_m \geqslant \frac{1-\sqrt[K_m]{1-P^*}}{P^*q}. \tag{7}$$

Notice that, when P^* is close to q and $K_i = 1$, p_i may be bigger than 1 although in large-scale sensor networks, $K_i = 1$ can hardly happen. To secure the success probability P^* for $(m+1)$-hop packet, p_i should be

$$p_i = \begin{cases} \dfrac{1-\sqrt[K_i]{1-P^*}}{q}, & i = 1, \\ \min\left\{1, \dfrac{1-\sqrt[K_i]{1-P^*}}{P^*q}\right\}, & i \geqslant 2. \end{cases} \tag{8}$$

Equation (8) shows that p_i decreases as K_i increases. Equation (8) provides a sufficient condition on p_i. We assume each packet has the information of the gradate and ID of intermediate node. Thus K_i is updated and node failure can be tolerant.

3.2 Asymptotic property

With EPro, when receiving one packet, each of K_i nodes will independently generate a random number which follows uniform distribution $U(0,1)$. If the number falls in $(0,p_i)$, the node forwards this packet; otherwise, the node ignores it. Suppose that there are N_i nodes to relay the packet. The rest K_i-N_i nodes simply ignore this packet to save energy. Hence, the value of N_i is important because only these nodes make concrete contribution to data transmission. N_i is a random variable and its mean can be calculated by the following formula:

$$E(N_i) = K_i p_i = \begin{cases} \dfrac{K_i\left(1-\sqrt[K_i]{1-P^*}\right)}{q}, & i=1, \\ \dfrac{K_i\left(1-\sqrt[K_i]{1-P^*}\right)}{P^*q}, & i \geqslant 2. \end{cases} \tag{9}$$

In Eq. (9) we assume $p_i \leqslant 1$. $E(N_i)$ will converge when K_i goes to infinity as shown in Fig. 2. In fact, when $K_i \to \infty$, we have

$$\sqrt[K_i]{1-P^*} = 1 + \frac{\ln(1-P^*)}{1!K_i} + \frac{[\ln(1-P^*)]^2}{1!K_i^2} + o\left(\frac{1}{K_i^2}\right),$$

therefore,

$$\lim_{K_i\to\infty} E(N_i) = \begin{cases} \dfrac{-\ln(1-P^*)}{q}, & i=1, \\ \dfrac{-\ln(1-P^*)}{P^*q}, & i \geqslant 2. \end{cases} \tag{10}$$

From Eq. (10), we can see that $E(N_i)$ does not increase as K_i increases and is bounded by $-\ln(1-P^*)/(P^*q)$. This feature makes EPro feasible to large-scale networks.

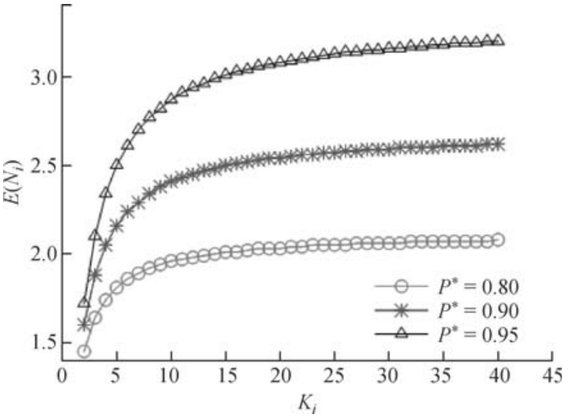

Fig. 2 $E(N_i)$ vs. K_i according to Eq. (9) with $q = 0.95$

3.3 EPro-n: Variation of EPro

The procedure of EPro is as follows. Suppose the gradient of the source node is h and the h-hop packet it generates is denoted by M. The source node puts the number of its 1-hop downstream nodes K_{h-1} in M so that its 1-hop downstream nodes can calculate their relay probability with K_{h-1} with the number of its own 1-hop downstream nodes before relaying M. As M travels downstream to the BS, intermediate nodes which have a lower gradient may produce multiple copies of M and these duplicated packets may arrive at the BS and let the BS bear more burdens of receiving and processing packets. Duplicated packets also waste the energy of intermediate nodes. To reduce the number of duplicated packets, we can set a number $n, n \geqslant 1$, as an upper bound of relay times for intermediate nodes. We denote this variation of EPro as EPro-n. In particular, EPro-1 means any intermediate node will not relay M more than once even if it receives M multiple times. EPro-1 can greatly reduce the duplicated copies of M but may not guarantee the success probability P^*. By increasing n, we may get better performance than EPro-1. n is the key to balance the performance and the cost. We will compare EPro, EPro-1, EPro-2 and EPro-3 by simulation.

4 Simulation results

In simulation, 4000 nodes are randomly and uniformly deployed in a 300 m × 300 m area. The BS is located at the center of the square. Communication radius R_0 is 24 m by adjusting transmission power in Ns-2. After initialization, the maximal gradient is 10.

With Ns-2, we compare four approaches in terms of packets received rate and average relay times for each packet in the same scenario. Besides EPro, approach ‘All’ means that every node will rebroadcast any packet it receives from its upstream nodes, but same packet will not be relayed once more by the same node. Similarly, in EPro-n, a node will not forward any packet more than n times. But in EPro, there is no bound on the number of relay times. The basic idea of gossiping-based approach is: when a node first receives a packet, with probability p it broadcasts the packet to its neighbors; if the node receives the same packet again, the packet will be discarded. A modification of the basic gossip protocol can be denoted as Gossip(p,k), which means nodes gossip with probability 1 for the first k hops before continuing to gossip with probability p. In the simulation, we randomly choose four different nodes as source nodes.

The results when *SourceNode_Gradient* = 4, 6, 8, 10 are shown in Fig. 3, respectively. Simulation results show that packet received rate with EPro can guarantee $P^* = 0.85$. The rate with EPro-3 is close to 0.85 but relay times per packet with EPro-3 is much less than EPro.

Figure 4 shows the simulation results when $P^* = 0.75, 0.80, 0.85, 0.90, 0.95$, respectively. With EPro, when $P^* = 0.95$ and $q = 0.95$, the relay probability for intermediate nodes goes to 1 so that relay times per packet increase dramatically and exceed the number by ‘All’. However, EPro-n can largely reduce relay times. Both of the two figures show that EPro or EPro-n can outperform Gossip(p,k).

5 Conclusion

EPro and EPro-n do not require global information but

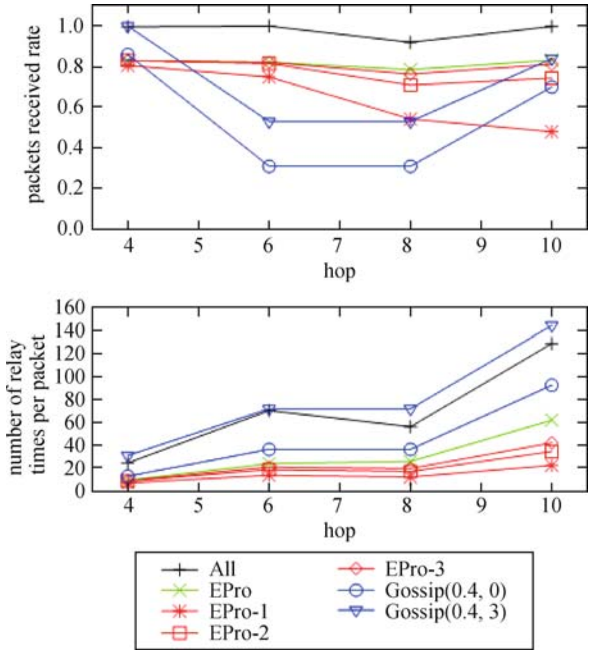

Fig. 3 Receive rate and relay times when $P^* = 0.85; q = 0.95;$ $SourceNode\_Gradient = 4, 6, 8, 10$

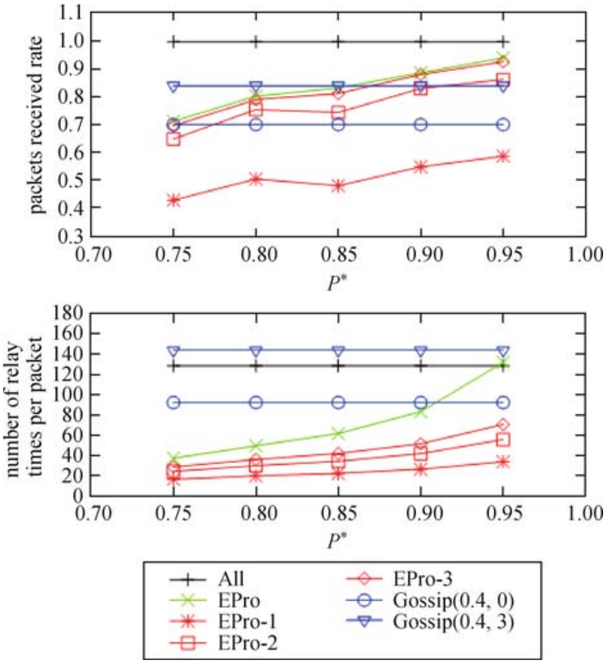

Fig. 4 Receive rate and relay times when $SourceNode\_Gradient = 10$; $P^* = 0.75, 0.80, 0.85, 0.90, 0.95$; $q = 0.95$

merely rely on local information of the networks. They are more robust under node failure and can adjust to the change of the topology of the networks. They can be easily extended for networks with multiple BSs, packets with different priorities, etc.

The theoretical analysis in this paper provides us deep insights to understand large-scale sensor networks and the results obtained are informative and helpful to network designers.

References

1. Karl H, Willig A. Protocols and Architectures for Wireless Sensor Networks. New York: John Wiley & Sons, 2005
2. Heinzelman W R, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless micro sensor networks. In: Proceedings of the 33rd Hawaii International Conference on System Sciences. 2000, 1 – 10
3. Yick J, Mukherjee B, Ghosal D. Wireless sensor network survey. Computer Networks, 2008, 52(12): 2292 – 2330
4. Haas Z J, Halpern J Y, Li L. Gossip-based ad hoc routing. In: Proceedings of the Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. 2002, 3: 1707 – 1716
5. Wu J, Chen L J, Yan P L, Zhou J G, Jiang H. A new reliable routing method based on probabilistic forwarding in wireless sensor network. In: Proceedings of the Fifth International Conference on Computer and Information Technology. 2005, 524 – 529
6. Barrett C L, Eidenbenz S J, Kroc L, Marathe M, Smith J P. Parametric probabilistic routing in sensor networks. Mobile Networks and Applications, 2005, 10(4): 529 – 544
7. Chen X, Ho Y C, Zhang J S, Bai H X. Probabilistic forwarding (ProFor) for large scale sensor networks. In: Proceedings of the 2006 IEEE International Conference on Networking, Sensing and Control. 2006, 963 – 968
8. Nagpal R, Shrobe H, Bachrach J. Organizing a global coordinate system from local information on an ad hoc sensor network. In: Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks. 2003, 333 – 348

Xi CHEN received her B.Sc. and M.Eng. Degrees from Nankai University, Tianjin, China, in 1986 and 1989, respectively. She received her Ph.D from the Chinese University of Hong Kong, Hong Kong, China, in 2000. She is now working in the Center for Intelligent and Networked Systems (CFINS), Department of Automation, Tsinghua University, Beijing, China. Her research interests include wireless sensor networks and stochastic control.

Xingshi WANG got his B.Eng. degree and M.Eng. degree in automation from Tsinghua University, Beijing, China, in 2007 and 2010. His research interest is wireless sensor networks.