An Enhanced Probabilistic Scheme for Data Transmission in Large Scale Sensor Networks

Xi Chen, Member, IEEE
Center for Intelligent and Networked Systems
Department of Automation
TNList, Tsinghua University
Beijing, China, 100084
bjchenxi@mail.tsinghua.edu.cn

Xingshi Wang
Center for Intelligent and Networked Systems
Department of Automation
TNList, Tsinghua University
Beijing, China, 100084
wangxs07@mails.tsinghua.edu.cn

***Abstract*—In this paper, an enhanced probabilistic scheme is presented for directed data transmission without route discovery. In our model, we require each message can reach the base station successfully with a certain probability. We analyze the relationship between the number of intermediate nodes, reliability of links 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**

I. Introduction

Large scale sensor networks can be used in many applications such as military sensing, environment monitoring and traffic surveillance, etc. By radio frequency communication, the energy consumption is proportional to k $(2 \leq k \leq 4)$ power of the transmission distance. Sensor node has limited energy and communication is energy- consuming. 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 via relaying by intermediate nodes. Usually, there exist multiple routes from the source to the base station.

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 base station according to the route tables. In the other category, packets are transmitted without any specific route. Both Flooding and Gossiping are such kind of protocols.

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[2].

A gossiping-based approach is presented in [3], where each node forwards a packet with some probability. Gossiping-based approach can also be combined with various optimizations of flooding such as AODV, to reduce the overhead of the routing protocols. This work first introduced the forward probability into routing protocols.

Reference [4] 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 [5] 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 route discovery is presented in [6]. With ProFor, each message may not reach the base station for sure but with a predefined success probability. There are many reasons for such relaxation. First, in large scale sensor networks, data from different sensor nodes may be similar or duplicated, absence of any small part can hardly seriously affects the performance of the large scale sensor networks; Secondly, in some real applications, a high success probability is acceptable and good enough; Thirdly, 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 networks and works robustly against individual node failure.

In this paper, following the basic idea of ProFor, we present an enhanced probabilistic scheme (EPro) for data transmission. Different from the work in [6], we take into account the packet loss caused by various factors. 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 [3][4][5], all probabilistic approaches are used in a heuristic manner. Rigorous analysis of the problem is the main contribution of this paper.

The rest of the paper is organized as follows. The problem is formulated in Section II. Relay probability and asymptotic property of the networks are presented in Section III. Simulation results are recorded in Section IV. Section V concludes the paper.

II. Problem Formulation

In our model, we assume sensor nodes are densely distributed in the area of interest. Without loss of generality, we suppose there is only one stationary base station (BS) at the center of the area of interest. After the networks are deployed, each node no longer moves and the 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 nodes sharing a common ID is very low. We have the following definitions.

Definition 1 A node is ***h-hop node*** if its 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 a (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 broadcast 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 sensor node simply drops such corrupted packet as if it does not receive the packet at all. Let $1-q(0<1-q<1)$ to be packet loss probability between neighboring nodes. For simplicity, we assume $1-q$ is same for all nodes in the networks.

The networks perform an initialization to generate the gradient for each node by following a similar way as presented in [7]: 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 the message to its neighbors with the count incremented by 1. Hence a wave of messages propagates outwards from the BS. Each node maintains the minimum counter value received and ignores messages containing larger values. The minimum hop count value, h_i , that Node i maintains 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.

III. Analysis of the Problem

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.

A. Sufficient condition for Relay probability

We analyze the probability with which intermediate nodes relaying packets to guarantee the success probability P^*. We assume $P^* \leq q$. When locating error codes 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 $P^* \leq q$, 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 $K_1 (K_1 \geq 1)$ 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 the BS is $1-(1-p_1 q)^{K_1}$.

Let$1-(1-p_1 q)^{K_1} \geq P^*$then we have $1 \geq p_1 q \geq 1 - \sqrt[K_1]{1-P^*}$, such that

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

Furthermore, we set

$$\alpha_1 = 1-(1-p_1 q)^{K_1}, \tag{2}$$

(1) stipulates the condition for the relay probability for 1-hop nodes. In fact, this condition is dependent on K_1 and $p_1(K_1)$ describes such dependence more clearly than p_1 does. In case there is no confusion caused, we use p_1 for simplicity. We also use p_i to represent $p_i(K_i)$ in the sequel.

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, $\cdots$, Node K_2. Each of them will relay the message with probability p_2. Suppose Node $j, (1 \leq j \leq K_2)$ 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 (1). Hence we have

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

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

$$1-\prod_{j=1}^{K_2}(1-\alpha_1^{(j)} p_2 q) \geq P^*, \tag{4}$$

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

By (2), we know $\alpha_1^{(j)} \geq P^*$, $j = 1, \cdots, K_2$, then $\prod_{j=1}^{K_2}(1 - \alpha_1^{(j)} p_2 q) \leq (1-P^* p_2 q)^{K_2}$. In (4), we replace $\prod_{j=1}^{K_2}(1-\alpha_1^{(j)} p_2 q)$ with $(1-P^* p_2 q)^{K_2}$ then we get $1 \geq 1-(1-P^* p_2 q)^{K_2} \geq P^*$. Thus

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

p_2 in (6) is a little more conservative than in (4). However, p_2 in (6) no longer hinges on $\{p_1^{(j)}, j = 1, 2, \cdots, K_2\}$ and is determined merely by K_2. (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 get involved in relaying the (m+1)-hop message and the relay probability is p_m. By same reasoning and adopting the same trick of removing the dependence of p_m on $\{p_{m-1}^{(j)}, j = 1, 2, \cdots, K_m\}$ which are carried out in the case when $SourceNode\_Gradient = 3$, we have

$$p_m \geq \frac{1 - \sqrt[K_1]{1 - P^*}}{P^* q} \tag{7}$$

We 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 happens. To secure the success probability P^* for (m+1)-hop packet, p_i should be

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

(8) shows that p_i decreases as K_i increases. (8) provides a sufficient condition on p_i.

B. 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 forward this packet; otherwise, the node ignore it. Suppose that there are N_i nodes do 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.

$$E(N_i) = K_i p_i = \begin{cases} \dfrac{K_1(1 - \sqrt[K_1]{1 - P^*})}{q}, & i = 1; \\ \dfrac{K_i(1 - \sqrt[K_i]{1 - P^*})}{P^* q}, & i \geq 2. \end{cases} \tag{9}$$

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

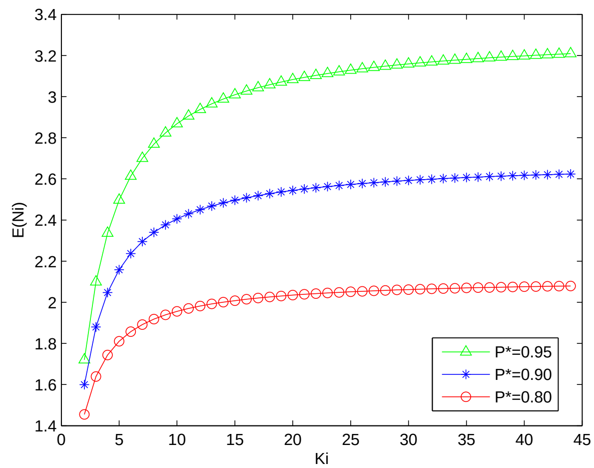

Fig. 1. $E(N_i)$ v.s. K_i according to (9) with $q = 0.95$.

we have

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

therefore,

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

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

C. 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}. Similarly, any intermediate relaying node need replace 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 $n, n \geq 1$ 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 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.

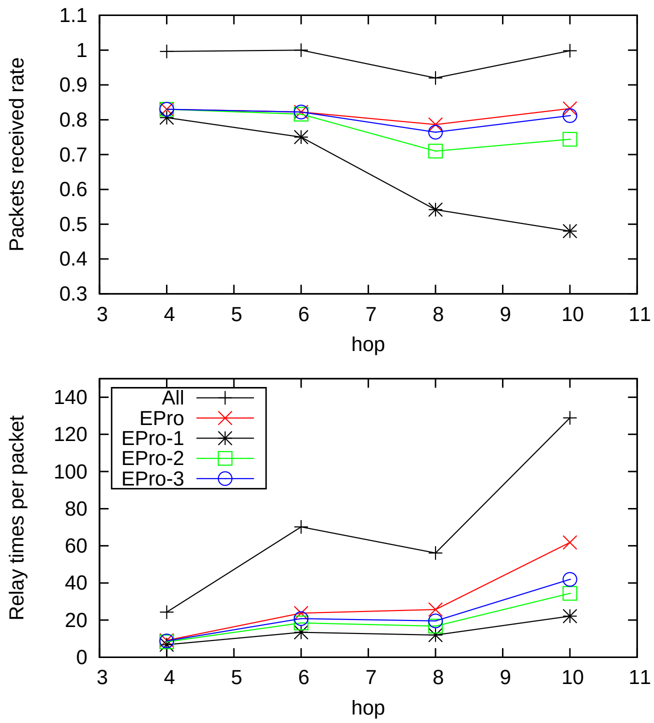

Fig. 2. Receive rate and relay times when $P^* = 0.85$, $q = 0.95$, $SourceNode\_Gradient = 4, 6, 8, 10$.

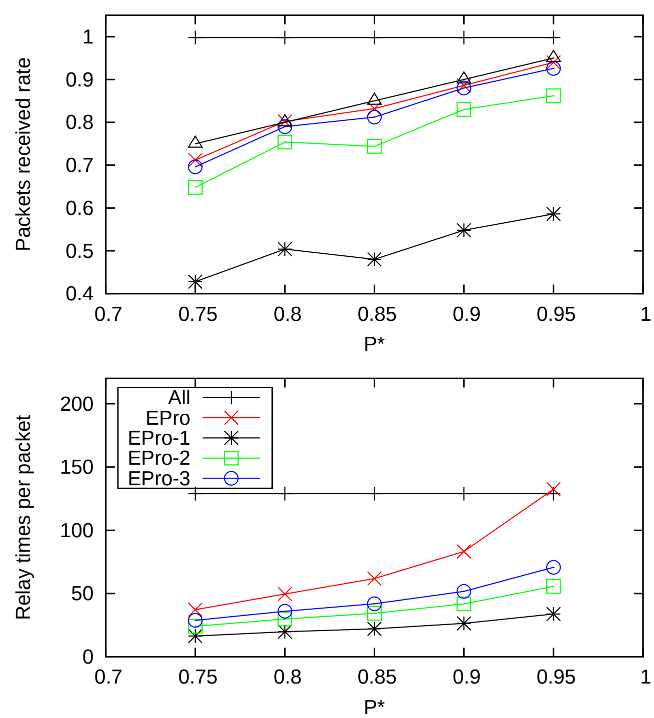

Fig. 3. Receive rate and relay times when $SourceNode\_Gradient = 10$, $q = 0.95$, $P^* = 0.75, 0.90, 0.98$,

IV. Simulation Results

In simulation, 4000 nodes are randomly and uniformly deployed in a $100m \times 100m$ area. The BS is located at the center of the square. Communication radius R_0 is $8m$. After initialization, the maximal gradient is 10.

With Ns-2, we compare three approaches in terms of packets received rate and average relay times for each packet. Besides EPro, approach 'All' means that node will rebroadcast any packet it receives from its upstream nodes, but same packet will not be relayed more than once 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 relay times. We randomly choose 4 different nodes as source nodes. Fig. 2 shows the results when $SourceNode\_Gradient = 4, 6, 8, 10$, respectively. Simulation results show that packet received rate with EPro can guarantee $P^* = 0.85$. The rate with EPro-3 is very close to 0.85 but relay times per packet with EPro-3 is much less than it with EPro.

For different value of P^*, Fig.3 shows the simulation results when $P* = 0.75, 0.90, 0.95$ respectively. As shown in Fig.3, with EPro, when $P^* = 0.95$ equals to q, the relay probability for intermediate nodes goes to 1 so that relay times per packet increases dramatically and exceeds the number by 'All'. However, EPro-n can largely reduce relay times.

V. Conclusion

EPro and EPro-n do not require global information but merely rely on local information of the networks. They are more robust under node failure and can adjust to the topology change of the networks. EPro and EPro-n have a great potential to cope with sensor networks in various settings. They are easily to be expanded for networks with multiple base stations, 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] H. Karl, and A. Willig, "Protocols and architectures for wireless sensor network", *John Wiley & Sons, Ltd*, 2005.

[2] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey", *Computer Network*, vol. 52, pp.2292-2330, 2008.

[3] Z. Haas, J. Y.Halpern, and L. Li, "Gossip-Based Ad Hoc Routing", *In Proceedings of the IEEE INFOCOM*, New York, June, 2002.

[4] J. Wu, L. J. Chen, P. L. Yan, J. G. Zhou, and H. Jiang, "A New Reliable Routing Method Based on Probabilisitc Forwarding in Wireless Sensor Network", *In Proceedings of the 2005 The Fifth International Conference on Computer and Information Technology(CIT'05)*, Shanghai, September, 2005.

[5] C. L.Barrett, S. J.Eidenbenz, and L.Kroc, Madhav Marathe and James P.Smith, "Parametric Probabilistic Routing in Sensor Networks", *Mobile Networks and Applications*,vol. 10, pp.529-544, 2005.

[6] X. Chen, Y. C. Ho, J. Zhang, and H. Bai, "Probabilistic Forwarding (ProFor) for large scale sensor networks", *In Proceedings of the 2006 IEEE international conference on networking, sensing and control*, Ft. Lauderdale, FL, April, 2006.

[7] R. Nagpal, H. Shrobe, and J. Bachrach, "Organizing a global coordinate system from local information on an ad hoc sensor network". *2nd International Workshop on Information Processing in Sensor Network(IPSN'03)*,Palo Alto, CA, April, 2003