A Novel Energy-Efficient Routing Probabilistic Strategies For Distributed And Localized Heterogeneous Wsn

yassine sabri (✉ sabriyassino@gmail.com)  
Universite Chouaib Doukkali  https://orcid.org/0000-0002-2083-5422

Elkamoun Najib  
ParisTech

Research Article

Keywords: Wireless sensor networks Probabilistic, Energy-Efficient, Heterogeneous, Network Lifetime

Posted Date: June 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1608865/v1

License: ☒  This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
A Novel Energy-Efficient Routing
Probabilistic Strategies For Distributed And
Localized Heterogeneous Wsn

Yassine Sabri$^{1*}$ and Najib El kamoun$^{2†}$

$^1$Laboratory of Innovation in Management and Engineering for Enterprise (LIMIE), ISGA Rabat, 27 Avenue Oqba, Rabat, 10587, MAR, Morocco.
$^2$Paris-Sud University, Experiments and Digital Approaches, molino, Paris, 100190, Fr, France.

*Corresponding author(s). E-mail(s): yassine.sabri@isga.ma; elkamoun@gmail.com; †These authors contributed equally to this work.

Abstract

In the context of geographic routing in wireless sensor networks linked by probabilistic communication channels, energy efficient transmission is important to extend the network lifetime. presents a novel probabilistic forwarding approach for directed data transmission without route discovery. In this model, each message is required to reach the base station successfully with a predefined probability, hence sensor nodes which are located nearer to the base station need to relay messages with a certain relay probability. The relationship between the number of relaying nodes and relay probability is analyzed and the condition for relay probability to guarantee the success probability is obtained. We propose a novel methods ProFor and EnProFor protocols are distributed and localized protocols. They can adapt to a topological change of the network and work robustly against individual node failure. Simulation examples comparing the energy costs for the different strategies illustrate the theoretical analysis in the cases. With the method proposed it is possible to obtain a significant energy savings (up to ten times) with probabilistic transmission rate and power.

Keywords: Wireless sensor networks Probabilistic, Energy-Efficient, Heterogeneous, Network Lifetime
1 Introduction

Data transmission has been an important research topic in WSN. By radio frequency communication, the energy consumption is proportional to the nth power of the transmission distance. To save energy, short distance multiple hop communication becomes preferable to long distance direct communication. Data is transmitted from its source node to the base station or the sink node via relaying by intermediate sensor nodes. Usually there exist multiple routes from the source to the base station. Routing is to find the proper one for data transmission. Data transmission protocols are based on the topological structure of networks. In general, topological structures of WSNs are classified as cluster topology, tree topology, and gradient topology, as shown in Fig. 1.

![Fig. 1 Three kinds of topological structure. (A) Cluster topology. (B) Tree topology. (C) Gradient topology.](image)

In a cluster topology (see Fig. 1A), nodes are divided into different clusters (dotted ovals). In each cluster, all data first converge to the cluster head (gray circle). Cluster heads usually directly communicate with the base station. If one cluster head shows failure, then data from the rest of the nodes in the cluster will be unable to reach the base station. Cluster head switching can efficiently balance energy consumption in the network. However, head switching may need extra communication for negotiation among nodes.

In a tree topology (see Fig. 1B), each node in a hierarchy level has point-to-point links with each adjacent node on its below level. The base station is at the top level (or root). Data from any node in the tree needs to be relayed by the intermediate nodes on the route from the node to the base station. If one intermediate node shows failure, all data from the nodes which are located at its lower levels cannot reach the base station. Tree structure also results in uneven energy consumption in the network. In a gradient topology
(see Fig. 1C), all nodes create their gradients based on the minimal number of hops to the base station. Data travels along the gradient descent direction and finally converges to the base station. The gradient topology can support more flexible and multi-path routing protocols, which may spend more energy on transmission but enhance the reliability of data reaching the base station.

Data transmission protocols in WSNs may be roughly classified into two categories according to the way of routing. In one category, the transmission route is predetermined for packets and intermediate nodes have to relay packets from the source to the base station. Route tables are determined by the topology of the network and need to be updated when the topology changes. For example, LEACH [8], as a clustering-based protocol, lets each node maintain its route table by communicating with neighboring nodes. In the other category, packets are transmitted without any specific route. Flooding is such a kind of protocol, and so is Gossiping. They do not require a packet to follow any specific route to the base station; hence no route table is necessary any more.

Haas et al. [7] develop a gossiping-based approach, where each node forwards a packet with some probability. Combined with variations of flooding, the gossiping-based approach can reduce the overhead of routing protocols. This work first introduces forward probability into routing protocols.

Wu et al. [14] present a selective forwarding probability. Neighbors as the next hop are selected with some probability. The selective forwarding probability takes into account of the node degree and link loss in order to enhance the reliability of selective forwarding. Barrett et al. [3] propose a family of light-weight and robust multi-path routing protocols in which an intermediate sensor node forwards a message with a probability which depends on various parameters, such as the distance of the source node to the destination, the distance of the node to the destination, or the number of hops a packet has already traveled. In this paper, a probabilistic forwarding approach for data transmission is presented. With this approach, each message may not reach the base station with certainty but with a predefined success probability. There are many reasons for such a relaxation. First of all, in a large scale WSN, data from different sensor nodes may be similar or duplicated, the absence of any small part of data can hardly affect the network performance. Secondly, in some real applications, a high success probability is acceptable and good enough. Thirdly, the cost of demanding absolute certainty is usually unaffordable.

Notice that in [3, 7, 14], all probabilistic approaches are used in a heuristic manner. In this paper, we analyze and obtain the condition for the relay probability to meet the requirement on data transmission.
2 Methods/Experimental: Probabilistic forwarding (ProFor)

We study the way of data transmission and develop a probabilistic forwarding protocol. This protocol lets relay nodes transmit packets with a certain probability to let messages reach the base station with a predefined probability [5].

2.1 Model description

The following assumptions are imposed in this paper.

- Sensor nodes are randomly and densely deployed in the area of interest.
- Each sensor node knows its gradient.
- Each sensor node has an unique ID.
- Each sensor node adopts the communication radius $R_0$.

In addition, some definitions on nodes and messages are as follows.

- A sensor node is called an $h$-hop node if its gradient is $h$.
- A message is called an $h-hop$ message if its source is an $h-hop$ node.
- If an $(h - 1)-hop$ node and an $h-hop$ node can communicate with communication radius $R_0$, then the former one is called a $1-hop$ downstream neighbor of the latter one; and the latter one is called a $1-hop$ upstream neighbor of the former one.
- If different $(h \neq 1)-hop$ nodes have a common h-hop node as their $1-hop$ upstream node, then they are called sibling nodes with respect to the $h-hop$ node.

2.1.1 Generating the gradient for each node

Suppose the network has only one stationary base station and it is located at the border or inside the area of interest and all nodes would not move after the deployment.

The network performs an initialization to generate the gradient for each node by following a similar way to [11]: the base station initiates a gradient by sending its neighboring sensor nodes a message with communication radius $R_0$. The message includes a count set to 1. Each node which receives the message remembers the value of the count and forwards the message to its neighbors with communication radius $R_0$ but with the count incremented by 1. Therefore, a wave of messages propagates outwards from the base station. Each node keeps the minimum count value it has received and ignores messages which contain larger count values. A node, say $O_j$, gets the minimum hop count, say $h_j$. $h_j$ represents the length of shortest path from $O_j$ to the base station in terms of communication hops. $h_j$ is also called the gradient of $O_j$ with respect to the base station.
2.2 Analysis of relay probability

In this paper, messages and packets are two related but different concepts. A message refers to a set of data which is carried by packets to be sent to the base station. Different packets may carry the same message when the message is traveling in the network.

Now we analyze and derive the relay probability with which intermediate nodes relaying messages to let them reach the base station with probability \( P^* \). \( P^* \) is called the success probability.

Suppose there is no packet loss during data transmission.

**Case 1:** when SourceNode-Gradient=1, which means the source node is one hop from the base station. Once the node broadcasts the message, the base station will receive it immediately. There is no need to relay the message.

**Case 2:** when SourceNode-Gradient=2, it is a 2-hop message. The source node is two hops from the base station; therefore its 1-hop downstream neighbors need to relay the message.

Suppose the source node has \( M_1 (M_1 \geq 1) \) 1-hop nodes as its 1-hop downstream neighbors. These \( M_1 \) nodes are sibling nodes as the source node is their common 1-hop upstream neighbor. If each of them forwards the message with probability \( \eta_1 \), the probability that the message reaches the base station is \( 1 - (1 - \eta_1)^{M_1} \). This probability must satisfy \( 1 - (1 - \eta_1)^{M_1} \geq P^* \). Hence we have

\[
\eta_1 \geq 1 - (1 - P^*)^{1/M_1} \tag{1}
\]

For convenience, we set

\[
\kappa_1 = 1 - (1 - \eta_1)^{M_1} \tag{2}
\]

Eq. (1) yields the condition of relay probability for the 1-hop downstream neighbors of the source node. Essentially, the relay probability \( \eta_1 \) is determined by \( M_1 \).

Case 3: when SourceNode-Gradient = 3, the message’s source node is 3 hops from the base station. The message needs to be relayed by intermediate 2-hop nodes and 1-hop nodes.

Suppose the source node has \( M_2 (M_2 \geq 1) \) 2-hop nodes as its 1-hop downstream neighbors, say \( O_1, O_2, \ldots, O_{M_2} \), which also are sibling nodes as they have the same source node as their 1-hop upstream neighbor. Suppose each of them will relay the message with probability \( \eta_2 \). Moreover, suppose each of these \( M_2 \) nodes independently has \( M_1^{(j)} \) 1-hop downstream neighbors, respectively. Here, \( M_1^{(j)} \geq 1, j = 1, 2, \ldots, M_2 \).
Provided $O_j$ forwards the message, its $M_1^{(j)}$ 1-hop downstream neighbors will receive and relay the message with probability $\eta_1^{(j)} \cdot \eta_2^{(j)}$ satisfies Eq. (1) for Case 2, that is,

$$\eta_1^{(j)} \geq 1 - (1 - P^*)^{1/M_1^{(j)}} \quad (3)$$

Hence, via $O_j$, the probability that the message successfully reaches the base station is $\kappa_1^{(j)} \eta_2$, where $\kappa_1^{(j)} = 1 - \left(1 - \eta_1^{(j)}\right)^{M_1^{(j)}}$.

As the source node has $M_2$ 1-hop downstream neighbors, the relay probability $\eta_2$ should satisfy the following condition:

$$\begin{cases}
1 - \prod_{j=1}^{M_2} \left[1 - \kappa_1^{(j)} \eta_2\right] \geq P^* \\
\kappa_1^{(j)} = 1 - \left(1 - \eta_1^{(j)}\right)^{M_1^{(j)}} \quad (4)
\end{cases}$$

As shown in Eq. (4), the relay probability $\eta_2$ is dependent on $\left\{\eta_1^{(j)}\right\}_{j=1}^{M_2}$. According to Eq. (3), $\kappa_1^{(j)} \geq P^*$ holds for $j = 1, 2, \cdots, M_2$. Replacing $\kappa_1^{(j)}$ with $P^*$ in Eq. (4), we have

$$1 - \prod_{j=1}^{M_2} \left(1 - \kappa_1^{(j)} \eta_2\right) \geq 1 - (1 - P^* \eta_2)^{M_2}$$

Let $1 - (1 - P^* \eta_2)^{M_2} \geq P^*$. We get another condition for $\eta_2$, that is,

$$\eta_2 \geq \frac{1 - (1 - P^*)^{1/M_2}}{P^*} \quad (5)$$

If $\eta_2$ satisfies Eq. (5), then Eq. (4) will hold, namely, Eq. (5) is more conservative than Eq. (4). However, Eq. (5) has an obvious advantage as $\eta_2$ is independent on $\kappa_1^{(j)}$ ($j = 1, 2, \cdots, M_2$). In other words, Eq. (5) delivers a localized solution for $\eta_2$.

Now we consider the generalized case.

Case 4: when SourceNode-Gradient $= k + 1$. Suppose the source node has $M_k$ 1-hop downstream neighbors, which are $k$-hop nodes and will get involved in relaying the $(k + 1)$-hop message. Denote by $\eta_k$ the relay probability. With the same inductive reasoning and the trick of removing the dependence of $\eta_k$ on $\eta_{k-1}^{(j)}$, we have the condition for $\eta_k$, that is,
\( \eta_k \geq \frac{1-(1-P^*)^{1/M_k}}{P^*}. \) (6)

To guarantee a success probability \( P^* \) for a \((k+1)\)-hop messages, the relay probability of the intermediate \( i \)-hop nodes, \( \eta_i \), is

\[
\begin{align*}
\eta_i & \geq 1 - (1 - P^*)^{1/M_i}, \quad i = 1 \\
\eta_i & \geq \frac{1-(1-P^*)^{1/M_i}}{P^*}, \quad i = 2, 3, \cdots, k.
\end{align*}
\] (7)

Eq. (7) is a localized and conservative solution for \( \eta_i \) which depends on \( M_i \) only. Such a feature makes the implementation of ProFor much easier.

With Eq. (7), we set the relay probability as

\[
\eta_i = \begin{cases} 
1 - (1 - P^*)^{1/M_i}, & i = 1 \\
\frac{1-(1-P^*)^{1/M_i}}{P^*}, & i = 2, 3, \cdots, k.
\end{cases}
\] (8)

![Fig. 2 Relay probability.](image)

Given \( P^* \), the relationship between \( \eta_i \) and \( M_i \) is shown in Fig. 2. A larger \( M_i \) means more paths for transmission and results in a smaller \( \eta_i \). This observation is consistent with the logic of the real situation. In particular, if there is only one path from the source to the base station which means that \( M_1 = 1 \) and \( M_i = 1(i = 2, 3, \cdots, k) \), then \( \eta_1 = P^* \) and \( \eta_i = 1(i = 2, 3, \cdots, k) \), that is, the 1-hop node relays it with probability \( P^* \) and the other intermediate nodes must relay the message. Obviously there exist multiple solutions to make the message reach the base station.
with probability $P^*$. For instance, let one of the intermediate nodes relay
the message with $P^*$ and each of the rest relay it with probability 1.
Nevertheless, Eq. (8) yields one feasible solution $\eta_i$ is determined by $M_i$
and $P^* \cdot P^*$ is predefined.

Each $(i+1)$-hop node counts its 1-hop downstream neighbors,
which are $i$-hop nodes, at the initialization stage and then broadcasts
its $M_i$ (and $P^*$ if needed) together with the message it relays or
initiates so that its 1-hop downstream neighbors can compute $\eta_i$
and decide whether or not to relay the message. The procedure for
an intermediate $i$-hop node is described as follows.

**Algorithm 1** Relaying a packet with ProFor protocol

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Receiving one packet.</td>
</tr>
<tr>
<td>2</td>
<td>Checking if this packet being sent from a $(i+1)$-hop node. If not, going to Step 7.</td>
</tr>
<tr>
<td>3</td>
<td>Taking $P^*$ and $M_i$ from the packet and computing relay probability $\eta_i$ with Eq. (8).</td>
</tr>
<tr>
<td>4</td>
<td>Generating a random number $\beta$ which follows uniform distribution $U(0,1)$.</td>
</tr>
<tr>
<td>5</td>
<td>If $\beta &gt; \eta_i$, going to Step 7.</td>
</tr>
<tr>
<td>6</td>
<td>Updating the packet, i.e. leaving the message unchanged but replacing the gradient and the number of 1-hop downstream neighbors with its own. Then relaying the updated packet and going to Step 8.</td>
</tr>
<tr>
<td>7</td>
<td>Discarding the packet and going to Step 8.</td>
</tr>
<tr>
<td>8</td>
<td>Termination.</td>
</tr>
</tbody>
</table>

### 3 Enhanced probabilistic forwarding (EnProFor)

For WSNs, routing protocols should be distributed, have low energy consumption, and be able to cope with frequently changing network topologies [9]. Meanwhile, routing protocols have to tolerate packet loss due to bad radio communication, congestion, packet collision, memory capacity, node failures, etc. In this section, an enhanced probabilistic forwarding (EnProFor) method is presented for data transmission. Different from ProFor, EnProFor takes into account the packet loss. We derive the conditions for relay probability and explore the asymptotic property of the networks.

#### 3.1 Analysis of relay probability

In reality, communication with radio frequency is sensitive to many environmental factors. Due to the existence of a gray zone for links between nodes (see [1] for details as regards the communication zone), there is a possibility
that a node has some 1-hop downstream neighbors located in its gray zone. If so, the connectivity between the node and its neighbors may become unreliable, so that its neighbor may not receive the packet or receive the packet with error codes. We assume that sensor nodes will simply discard any corrupted packet. For simplicity, the packet loss probability between neighboring nodes is assumed to be $1 - q (0 < 1 - q < 1)$.

The network performs an initialization to generate the gradient for each node in the same way as described in Section 1. Of course, packet loss may also happen on gradient establishment packets. Hence packet loss may let a node get a smaller number of 1-hop downstream neighbors and hence EnProFor could be a little more conservative.

Due to packet loss in the network, we assume the predefined success probability $P^* \leq q$. Now we analyze and derive the sufficient condition for relay probability [6].

**Case 1:** when $\text{SourceNode-Gradient} = 1$, which means the source node is one hop from the base station. When it broadcasts a message, the base station can receive it immediately with probability $q$. Since $P^* \leq q$, no relay is needed.

**Case 2:** when $\text{SourceNode-Gradient} = 2$, the source node is two hops from the base station so that relaying by intermediate 1-hop node(s) is indispensable.

Suppose the source node has $\tilde{M}_1 \left( \tilde{M}_1 \geq 1 \right)$ 1-hop downstream neighbors. If each of them relays the message with probability $\tilde{\eta}_1$, then the base station receives it with probability $1 - (1 - \tilde{\eta}_1 q)^{\tilde{M}_1}$. Let $1 - (1 - \tilde{\eta}_1 q)^{\tilde{M}_1} \geq P^*$. Hence

$$1 \geq 1 - (1 - P^*)^{\tilde{M}_1} / q . \quad (9)$$

Eq. (9) stipulates the condition for the relay probability for 1-hop downstream neighbors. $\tilde{\eta}_1$ is dependent on $\tilde{M}_1$ and $q$.

Moreover, we set

$$1 = 1 - (1 - \tilde{\eta}_1 q)^{\tilde{M}_1} . \quad (10)$$

**Case 3:** when $\text{SourceNode-Gradient} = 3$, the 3-hop message needs to be relayed by 2-hop nodes and 1-hop nodes.

Suppose the source node has $\tilde{M}_2$ 1-hop downstream neighbors, say $O_1, O_2, \ldots$, and $O_{\tilde{M}_2}$, which are 2-hop nodes. Suppose $O_j \left( j = 1, 2, \ldots, \tilde{M}_2 \right)$ will relay the message with probability $\tilde{\eta}_2$ and $O_j$ has $\tilde{M}_1^{(j)}$ 1-hop downstream neighbors, which are 1-hop nodes. Provided $O_j$
sends out the message, its $\bar{M}_1^{(j)}$-hop downstream neighbors can receive and relay it with probability $\bar{\eta}_1^{(j)}$. $\bar{\eta}_1^{(j)}$ is defined by Eq. (9), namely

$$
\bar{\eta}_1^{(j)} \geq \frac{1 - (1 - P^*)^q / \bar{M}_1^{(j)}}{q}.
$$

Then the probability for the message successful arriving at the base station via $O_j$ is $\bar{\kappa}_1^{(j)} \bar{\eta}_2$, where $\bar{\kappa}_1^{(j)} = 1 - \left(1 - \bar{\eta}_1^{(j)} q^2\right)^{\bar{M}_1^{(j)}}$.

The probability that the message fails to arrive at the base station is $\prod_{j=1}^{\bar{M}_2} (1 - \bar{\kappa}_1^{(j)} \bar{\eta}_2 q)$. Hence the success probability is $1 - \prod_{j=1}^{\bar{M}_2} (1 - \bar{\kappa}_1^{(j)} \bar{\eta}_2 q) \cdot \eta_2$ should satisfy

$$
\begin{cases}
1 - \prod_{j=1}^{\bar{M}_2} (1 - \bar{\kappa}_1^{(j)} \bar{\eta}_2 q) \geq P^* \\
\bar{\kappa}_1^{(j)} = 1 - \left(1 - \bar{\eta}_1^{(j)} q^2\right)^{\bar{M}_1^{(j)}}
\end{cases}
$$

(11)

With (5.10), $\bar{\kappa}_1^{(j)} = 1 - \left(1 - \bar{\eta}_1^{(j)} q\right)^{\bar{M}_1^{(j)}} \geq P^*$ holds for $j = 1, 2, \cdots, \bar{M}_2$. In Eq. (11), by replacing $\bar{\kappa}_1^{(j)}$ with $P^*$, we have

$$
1 - \prod_{j=1}^{\bar{M}_2} (1 - \bar{\kappa}_1^{(j)} \bar{\eta}_2 q) \geq 1 - (1 - P^* \bar{\eta}_2 q)^{\bar{M}_2}
$$

(12)

Let $1 - (1 - P^* \bar{\eta}_2 q)^{\bar{M}_2} \geq P^*$, hence we have the condition for $\bar{\eta}_2$ as

$$
2 \geq \frac{1 - (1 - P^*)^q / \bar{M}_2}{P^* q}.
$$

(13)

Eq. (12) is not dependent on $\left\{\bar{\eta}_1^{(j)}, j = 1, 2, \cdots, \bar{M}_2\right\}$. It provides a localized solution for $\bar{\eta}_2$.

Case 4: when SourceNode-Gradient = $k+1$ in which a($k+1$)-hop message relaying by $\bar{M}_k \left(\bar{M}_k \geq 1\right)$ $k$-hop nodes. Each of them relays the message with probability $\bar{\eta}_k$. Similar to the analysis in Case 3, the condition for $\bar{\eta}_k$ is
\[ k \geq \frac{1-(1-P^*)^i}{P^* q} / \tilde{M}_i. \quad (14) \]

In general, to let a \((k+1)\)-hop message successfully arrive at the base station with probability \(P^*\), the condition for the relay probability \(\tilde{\eta}_i\) is

\[
\begin{cases} 
\tilde{\eta}_i \geq \frac{1-(1-P^*)^i}{q} / \tilde{M}_i, & i = 1 \\
\tilde{\eta}_i \geq \frac{1-(1-P^*)^i}{P^* q} / \tilde{M}_i, & i = 2, 3, \ldots, k
\end{cases} \quad (15)
\]

This probability determined by \(\tilde{M}_i\), success rate \(P^*\) and \(q\). We can set

\[
i = \begin{cases} 
\frac{1-(1-P^*)^i}{q} / \tilde{M}_i, & i = 1 \\
\frac{1-(1-P^*)^i}{P^* q} / \tilde{M}_i, & i = 2, 3, \ldots, k
\end{cases} \quad (16)
\]

This probabilistic forwarding protocol takes into account the possibility of packet loss. Hence it is an enhanced version of ProFor. With EnProFor, the procedure for an intermediate \(i\)-hop node is described as follows Algorithm 1.

\textbf{Algorithm 2} (Relaying a packet with EnProFor protocol ( \(q\) and \(P^*\) are given)).

\begin{itemize}
  \item \textbf{Step 1:} Receiving one packet.
  \item \textbf{Step 2:} Checking if this packet has been sent from a \((i+1)\)-hop node. If not, going to Step 7.
  \item \textbf{Step 3:} Taking \(P^*\) and \(M_i\) from the packet and computing relay probability \(\tilde{\eta}_i\) with Eq. (15).
  \item \textbf{Step 4:} Generating a random number \(\beta\) which follows the uniform distribution \(U(0, 1)\).
  \item \textbf{Step 5:} If \(\beta > \tilde{\eta}_i\), going to Step 7.
  \item \textbf{Step 6:} Updating the packet, i.e. leaving the message unchanged but replacing the gradient and the number of 1-hop downstream neighbors with its own. Then relaying the updated packet and going to Step 8.
  \item \textbf{Step 7:} Discarding the packet and going to Step 8.
  \item \textbf{Step 8:} Termination.
\end{itemize}

\section{4 Results and Discussion}

In this section, we analyze the relay procedure to estimate the number of relays.
4.1 Simulation and analysis of probabilistic forwarding (ProFor)

We use network simulator ns2 [10] to simulate ProFor. The simulator ns2 is a discrete event simulator targeted at networking research. It provides substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless networks.

![Fig. 3](image.png)

**Fig. 3** Left lower area of the simulation area (BS: base station).

Within an area of 100 × 100 square meters, 4000 nodes are randomly and uniformly distributed. The communication radius is 10 meters. The base station is located at the center of the area (50,50). Fig. 3 shows the left lower quarter of the area, in which sensor nodes are marked with different patterns and the same pattern indicates the same gradient. In Fig. 3, nodes which are located at the lower left corner are farthest from the base station and their gradient is 10°.

When a message is being transmitted in the network, a node may receive duplicate packets which carry an identical message. Fig. 4 provides an example of this situation, in which both $O_2$ and $O_3$ are 1-hop downstream neighbors of $O_4$; and $O_1$ is 1-hop downstream neighbor of $O_2$ and $O_3$. A packet sent by $O_4$ is relayed by both node $O_2$ and $O_3$. Then $O_1$ will receive two packets separately. According to FroFor, $O_1$ has some probability to relay both of them. Such situations will happen more frequently on nodes which are close to the base station. The FroFor-1 protocol aims to alleviate the traffic load by preventing any node from ever forwarding a message more than once.
To analyze and illustrate the simulation results of ProFor, we compare it with ALL-1. ALL-1 means that a node will rebroadcast any message it receives from its upstream nodes but will not relay an identical message more than once. ALL-1 provides a benchmark to analysis of the performance of ProFor.

We arbitrarily choose one 10-hop node as a source node. Messages are transmitted downstream from the source to the base station. Let the source node send $N_{message}$ messages in total, here $N_{message} = 100$. Having fixed the value of $P^*$, we simulate protocols ALL-1, ProFor, and ProFor separately, and we calculate the total number of relays and the number of messages received by the base station and compute the total number of relays in the network to estimate the energy consumption. In Fig. 5 A, the success rate is ratio of the number of messages received by the base station to $N_{message}$, and in Fig. 5 B, the average number of relays is the total number of relays divided by $N_{message}$. The average number of relays is also proportional to the energy consumption on data transmission per message.

In Fig. 5 A and 5 B, the two lines at the top demonstrate that ALL-1 has a 100% success rate with 200 relays per message; ProFor performs well to secure success probability $P^*$. As $P^*$ is increasing, the success rate with ProFor is even bigger than $P^*$, while the average number of relays also increases. When $P^*$ approaches 100%, the average number of relays increases rapidly and approaches the number incurred by ALL-1. This property indicates that ProFor can save much energy with a small loss of $P^*$. On the other hand, with ProFor-1, the average number of relays has a moderate increase when $P^*$ approaches 1, but the success rate always falls behind $P^*$. Therefore, when $P^* \leq 0.90$, ProFor is a good choice; when $P^*$ is getting larger or approaching 1, we may let a part of the nodes adopt ProFor and the rest adopt ProFor-1 to get a good success rate with less energy consumption.

By simulation experiments, we further explore the relationship between the gradient of the source node and the number of relays per message. Having fixed $P^* = 0.80$, we replicate the above simulations for an arbitrary $i$-hop node, $i = 4, 5, 6, 7, 8, 9$, respectively. We plot the resulting success rate and the average number of relays per message in Fig. 6. As the gradient of the source node increases (i.e. the source node is getting farther from the base

---

**Fig. 4** Intermediate node receiving duplicate packets.
Fig. 5 Success rate and average number of relays (gradient of source = 10).

Fig. 6 Success rate and average number of relays ($P^* = 0.8$).
station), more nodes will be involved in relaying so that the number of relays should increase. Fig. 6 shows that ProFor delivers a success rate close to $P^*$ and the average number of relays increases moderately while the performance of ProFor-1 gets worse. Compared with ProFor-1, ProFor provides enough packets for each message to guarantee the success probability.

### 4.2 Simulation and analysis of enhanced probabilistic forwarding (EnProFor)

We use the same settings and follow the same way as described in Section 3 to carry out simulations for EnProFor. Similar to ProFor-1, EnProFor-1 means that if a relay node receives multiple packets of one message, it will forward the message at most once.

![Graph showing success rate and average number of relays](image)

**Fig. 7** Success rate and average number of relays (gradient of source = 10 and $q = 0.95$).

Figs. 7 A and 7 B show that, when $q < 1$, ProFor fails to achieve the predefined $P^*$ but EnProFor can make it. Of course, EnProFor needs more relays. Compared to EnProFor, EnForFor-1 needs less relays but its success rate is smaller. As $P^*$ gets larger, the success rate lags much further behind.

Figs. 8 A and 8 B show that EnProFor performs very well and meets the requirement. As the gradient of the source node increases, there are more nodes getting involved in relaying the message. The average number of relays required by ALL-1 increases sharply, but the number required by EnProFor
increases mildly. Hence, for a large scale WSN, EnProFor is much more energy efficient.

Since the success rate by EnProFor-1 is much smaller than $P^*$, we introduce EnProFor- $n$, which lets a node forward packets of one message at most $n$ times.

By simulation, we compare EnProFor, EnProFor-1, EnProFor-2 and EnProFor-3. As shown in Figs. 9 A and 9 B, with a larger $n$, the success rate obtained with EnProFor- $n$ is getting closer to the rate with EnProFor. And the corresponding number of relays also approaches the number required by EnProFor. However, on average, EnProFor-3 needs less relays than EnProFor does, especially when $P^*$ goes to 1. Figs. 10 A and 10 B show the same features of EProFor- $n$ as discussed above and the growing gap of success rate and average number of relays between EnProFor and EnProFor-1, 2, 3 as the gradient of source node increases.

The results of the simulation indicate that EnProFor achieves the predefined success rate $P^*$ when considering packet loss. The success rate obtained with EnProFor3 is almost the same as the rate with EnProFor but EnProFor-3 needs less relays.

### 4.3 Relaying by sibling nodes

Consider the procedure of relaying message. The number of $i$-hop sibling nodes which receive the packet sent by their common 1-hop upstream neighbor is $M_i$. Each of these $M_i$ sibling nodes computes the relay probability $\eta_i$ and then generates a random number following the uniform distribution.
Fig. 9 Success rate and average number of relays (gradient of source = 10 and $q = 0.95$).

Fig. 10 Success rate and average number of relays ($P_* = 0.8$ and $q = 0.95$).
$U(0, 1)$; if the number falls in $(0, \eta_i]$, then the node will relay the packet. Suppose there are $m_i$ of these $M_i$ sibling nodes that eventually relay the packet and the rest, $M_i - m_i$ nodes, discard the packet. $m_i$ is a random variable with the mean of

$$E(m_i) = M_i \eta_i = \begin{cases} \frac{M_i(1-(1-P_i^*)^{1/M_i})}{M_i}, & i = 1 \\ \min \left\{ q \frac{M_i(1-(1-P_i^*)^{1/M_i})}{P_i^* q}, M_i \right\}, & i \geq 2 \end{cases}$$

(17)

where operator $E$ is for taking the expectation. We compute $E(m_i)$ with Eq. (16) and plot it in Fig. 11, which shows that $E(m_i)$ increases monotonically with $M_i$; and when $M_i$ goes to infinity, $E(m_i)$ will converge to a finite number instead of growing unlimitedly. In other words, there exists a finite upper bound for $E(m_i)$. In fact, when $M_i \to \infty$, with the Taylor series we have

$$\left(1 - P_i^* \right)^{1/M_i} = 1 + \frac{\ln (1-P_i^*)}{1! M_i} + \frac{[\ln (1-P_i^*)]^2}{2! M_i^2} + o \left( \frac{1}{M_i^2} \right)$$

(18)

$$1-(1 - P_i^*)^{1/M_i} = -\frac{\ln (1-P_i^*)}{1! M_i} - \frac{[\ln (1-P_i^*)]^2}{2! M_i^2} + o \left( \frac{1}{M_i^2} \right)$$

(19)
If we denote by $E(m_i)$ the upper bound of $E(m_i)$, with Eq. (17) and Eq. (16), by taking the limit we obtain

$$E(m_i) = \lim_{M_i \to \infty} E(m_i) = \begin{cases} -\ln (1-P^*) / q, & i = 1 \\ \min \left\{ -\ln \left( \frac{1-P^*}{P^*} \right), M_i \right\}, & i \geq 2 \end{cases}$$

(20)

If $q = 0.95$, when $P^* = 0.80, 0.90, 0.95$, the corresponding $E(m_i)$ is 2.12, 2.69, 3.32, respectively. They are shown with three horizontal broken lines in Fig. 11. As $M_i$ is getting large, $m_i$ is approaching its upper bound $E(m_i)$.

### 4.4 Simulation and analysis of the number of relays

We use the same settings and follow the same way as described in Section 3 to carry out simulations for EnProFor.

Let $q = 0.95$ and $P^* = 0.9$. We simulate the EnProFor protocol and collect statistics about the relay procedure of intermediate nodes and list them in Table 1. For each S-node (abbr. of source node), we count the number of its 1-hop downstream neighbors, $M_i$, the number of sibling nodes. Each S-node broadcasts a number of packets (in the third column), we calculate the total number of relays by its 1-hop downstream neighbors (in the fourth column) and the average number of relays per packet sent by S-node (in the fifth column).

We observe that, no matter how big the value of $M_i$ is, the average number of relays is less than 3.32. This agrees with our analysis as regards $E(m_i)$. For example, S-node 131 has 23 downstream neighbors while S-node 424 has 12 downstream neighbors. The former is nearly twice as much as the latter. S-node 131 broadcasts 8 packets while S-node 424 broadcasts 7 packets but the average number of relays is 3 for both of them.

As one message is relayed to 1-hop downstream nodes, the number of packets of the message increases at most $E(m_i)$ times, which is less than $M_i$. Hence, for a large scale WSN in which sensor nodes are densely deployed, EnProFor costs much less energy than ALL-1 does (see Section 2).

For a relay node, its relay probability is determined by Eq. (15). Fig. 12 shows that when there are fewer sibling nodes the relay probability will increase; when $P^*$ is fixed, a larger $q$ leads to a smaller relay probability. The relay probability is sensitive to $M_i$ when $M_i$ is small and it converges when $M_i$ is getting larger. Given $q = 0.95$ and $P^* = 0.9$, we simulate the EnProFor protocol and collect statistics of the relay procedure and list a part of the data in Table 2, in which the R-node ID is the ID of the relay node. The R-node follows Algorithm 2. The second column records the gradient of the R-node; the third column is the number of packets received by the R-node from its 1-hop upstream neighbor. The fourth column records the total number of packets that the R-node has relayed. The last column is the
Fig. 12 Relay probability.

Table 1 Partial statistics for relay nodes \((q = 0.95, \ P^* = 0.9)\).

<table>
<thead>
<tr>
<th>R-node ID</th>
<th>gradient</th>
<th>Num. of packets received</th>
<th>Num. of relays</th>
<th>Relay rate</th>
<th>Average of (M_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>50%</td>
<td>36</td>
</tr>
<tr>
<td>1271</td>
<td>7</td>
<td>87</td>
<td>21</td>
<td>33.33%</td>
<td>14.14</td>
</tr>
<tr>
<td>1255</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>50%</td>
<td>9</td>
</tr>
<tr>
<td>1254</td>
<td>4</td>
<td>95</td>
<td>30</td>
<td>31.57%</td>
<td>6.25</td>
</tr>
<tr>
<td>1241</td>
<td>5</td>
<td>93</td>
<td>19</td>
<td>20.43%</td>
<td>4.67</td>
</tr>
<tr>
<td>124</td>
<td>1</td>
<td>11</td>
<td>4</td>
<td>36.36%</td>
<td>6</td>
</tr>
<tr>
<td>1220</td>
<td>6</td>
<td>25</td>
<td>7</td>
<td>28%</td>
<td>6.17</td>
</tr>
<tr>
<td>1189</td>
<td>6</td>
<td>61</td>
<td>16</td>
<td>26.22%</td>
<td>20</td>
</tr>
<tr>
<td>1171</td>
<td>7</td>
<td>21</td>
<td>9</td>
<td>19.05%</td>
<td>20</td>
</tr>
<tr>
<td>1163</td>
<td>6</td>
<td>29</td>
<td>2</td>
<td>31.03%</td>
<td>6.75</td>
</tr>
<tr>
<td>1112</td>
<td>6</td>
<td>5</td>
<td>40%</td>
<td>4</td>
<td>3.5</td>
</tr>
<tr>
<td>1066</td>
<td>5</td>
<td>127</td>
<td>32</td>
<td>33.07%</td>
<td>5.25</td>
</tr>
<tr>
<td>1034</td>
<td>4</td>
<td>25</td>
<td>32%</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1018</td>
<td>7</td>
<td>93</td>
<td>23.65%</td>
<td>8.57</td>
<td></td>
</tr>
</tbody>
</table>

average number of the R-node’s sibling nodes, i.e. the average of \(M_i\) encoded in packets received by the R-node.

Nodes which are close to the base station usually have to relay more packets and hence spend more energy. For example, the gradient of R-node 1112 is 2; it receives a large number of packets but it has less sibling nodes. With Eq. (15), we know that its relay probability is 0.5638 as \(q = 0.95, \ P^* = 0.9\) and \(M_i = 3.5\). In the simulation, it relays 253 of 493 packets, the relay rate is 51.31\%.
Nodes with more sibling nodes are usually less active in relaying as they share relay tasks with their sibling nodes. For example, the relay rate for $R$-node 1171 is smaller than that for $R$-node 1271 and $R$-node 1018.

Table 2 Partial statistics for source nodes ($q = 0.95, P^* = 0.9$).

<table>
<thead>
<tr>
<th>S-node ID</th>
<th>$M_i$</th>
<th>Num. of packets</th>
<th>Num. of relays</th>
<th>Average num. of relays</th>
</tr>
</thead>
<tbody>
<tr>
<td>43 17</td>
<td>10</td>
<td>24</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>60 17</td>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>131 23</td>
<td>8</td>
<td>24</td>
<td><strong>3.00</strong></td>
<td></td>
</tr>
<tr>
<td>152 2</td>
<td>151</td>
<td>171</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>156 4</td>
<td>20</td>
<td>26</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>190 2</td>
<td>13</td>
<td>12</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>200 18</td>
<td>1</td>
<td>3</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>201 1</td>
<td>18</td>
<td>9</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>213 1</td>
<td>126</td>
<td>98</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>235 5</td>
<td>13</td>
<td>20</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>261 7</td>
<td>5</td>
<td>9</td>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>317 7</td>
<td>11</td>
<td>23</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>352 17</td>
<td>11</td>
<td>16</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>357 14</td>
<td>1</td>
<td>3</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>365 8</td>
<td>14</td>
<td>15</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>377 3</td>
<td>4</td>
<td>6</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>389 2</td>
<td>48</td>
<td>56</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>424 12</td>
<td>7</td>
<td>21</td>
<td><strong>3.00</strong></td>
<td></td>
</tr>
<tr>
<td>461 2</td>
<td>185</td>
<td>231</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>493 29</td>
<td>3</td>
<td>1</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

Even nodes are in idle mode, they also consume much energy (e.g. see Table 1 of the radio power characterization in Ref. [13]). Hence redundant nodes should be turned OFF to save energy [2,12,13].

According to the earlier analysis, Eq. (18) provides the theoretical upper bound of $\mathbb{E}(m_i), \mathbb{E}(m_i)$. Hence the rest, $M_i - \mathbb{E}(m_i)$ sibling nodes, are redundant. To balance energy consumption among the sibling nodes, these $M_i$ nodes should take turns to relay messages.

The state switching scheme presented in Section 1 may be applied with $\rho = \left\lceil \mathbb{E}(m_i) / M_i \right\rceil$ as the working probability, where $\left\lceil x \right\rceil$ represents the smallest integer which is equal to or bigger than $x$. Accordingly, an intermediate node should use $\left\lceil \mathbb{E}(m_i) \right\rceil$ instead of $M_i$ to determine its relay probability or simply relay any packet received from its 1-hop upstream neighbors.

4.5 Multiple base stations

In a WSN, nodes close to the base station have to spend much more energy on relaying messages. Once their energy is used up, the network cannot work any more. Such a situation generally exists in various network topologies [4]. Deploying more nodes around the base station should be helpful. However, it may be impossible or too expensive to deploy more nodes in a particular region. As discussed in previous papers, a randomly and uniformly deployed
WSN usually has a good coverage performance. Moreover, one static base station may be unable to efficiently collect data for the long term. An intuitive solution is configuring multiple base stations such that nodes can send data to the nearest one, or base stations take turns working to balance the energy consumption among all nodes.

When implementing ProFor or EnProFor with multiple base stations, every node’s gradient should be initialized with respect to each base station separately. Hence every node knows the corresponding 1-hop upstream neighbor and the number of 1-hop downstream neighbors.

4.6 Message priority

Messages generated by different nodes or with different contents may be of higher or lower importance. ProFor and EnProFor can distinguish the rank of importance with different success probabilities $P^*$. A bigger $P^*$ implies a higher importance.

5 Conclusion

ProFor and EnProFor protocols are distributed and localized protocols. They can adapt to a topological change of the network and work robustly against individual node failure. Of course, probabilistic forwarding inherently runs the risk of losing a message. However, both analysis and simulations show that ProFor and EnProFor work very well to guarantee a high success probability for data transmission.

In practical situations, ProFor and EnProFor are much easier and better for implementation. The results in this paper are widely applicable to different settings such as multiple base stations and message priority. The basic idea and framework of ProFor or EnProFor are also applicable to heterogeneous WSNs.

The theoretical analysis and simulation demonstration not only reveal the underlying features but also provide deep insights of a large scale WSN. They have great potential to improve the scheduling at a medium access control layer to gain better energy-efficiency.

References


Glossaries

IoT Internet-of-Things

WSN Wireless sensor networks

LEACH Adaptive Hierarchy with Low Power Consumption

ProFor Probabilistic forwarding

EnProFor Enhanced probabilistic forwarding

Declarations

1. We understand that this author is the sole contact for the Editorial process.

2. Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

3. Consent for publication: This is to state that I give my full permission for the publication, reproduction, broadcast and other use of photographs.

4. Authors’ contributions: We attest that all authors contributed significantly to the creation of this manuscript.

5. Authors’ information: Phd Yassine Sabri, Phd Najib Elkamoun (Moroccan University of Science and Technology, 2022–28) in Information and Communication Technology (ICT) aims at training scientists and engineers to design and develop innovative methodologies towards the design of systems, data processing, modeling, and applications, in the fields of electronic engineering, remote sensing, applied electromagnetics.

Editorial Policies for:

Springer journals and proceedings:
https://www.springer.com/gp/editorial-policies

Nature Portfolio journals:
https://www.nature.com/nature-research/editorial-policies

Scientific Reports:
https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals:
https://www.biomedcentral.com/getpublished/editorial-policies