Trade-Off between Traffic Overhead and Reliability in Multipath Routing for Wireless Sensor Networks

Stefan Dulman, Tim Nieberg, Jian Wu, Paul Havinga
Faculty of EEMCS
University of Twente
Enschede, the Netherlands
Emails: {dulman,tim,jian,havinga}@cs.utwente.nl

***Abstract*— In wireless sensor networks (WSN) data produced by one or more sources usually has to be routed through several intermediate nodes to reach the destination. Problems arise when intermediate nodes fail to forward the incoming messages. The reliability of the system can be increased by providing several paths from source to destination and sending the same packet through each of them (the algorithm is known as *multipath routing*). Using this technique, the traffic increases significantly. In this paper, we analyze a new mechanism that enables the trade-off between the amount of traffic and the reliability. The data packet is split in k subpackets (k = number of disjoined paths from source to destination). If only E_k subpackets ($E_k < k$) are necessary to rebuild the original data packet (condition obtained by adding redundancy to each subpacket), then the trade-off between traffic and reliability can be controlled.**

I. Introduction

Sensor nodes have many failure modes [1]. Each failure decreases the performance of the network. Our approach to assure that the gathered data will reach its destination in the network is by using an algorithm that assumes as a regular fact that nodes may be not available during the routing procedure. Additional energy will be required only for a small amount of computations; this is almost negligible compared with the energy used for communications [2].

The algorithm starts by discovering k multiple paths from the source to the destination [3] [4] [5]. These paths are calculated taking into account the reputation coefficient of the nodes (a coefficient that shows the way a specific node behaved in the past). A low reputation coefficient of a node means that the node failed to route messages in the network in the past and it will be avoided from the routing scheme. The reputation coefficient is increased with each successfully routed packet.

Sending the same data over multiple paths is a solution in case of node failures but it requires large quantities of network resources (such as bandwidth and energy). Our contribution is to split the data packet into k parts (further referred to as *subpackets*) and to send these subpackets instead of the whole packet. The subpackets are created using *erasure codes* (or *forward error correcting codes* FEC) that add redundancy to the original source data. The basic principle is to transmit a sequence of k subpackets, out of which only E_k subpackets are necessary to reconstruct the original packet. The receiver's robustness to missing packets is increased, which also implies that a return feedback channel is not the critical element.

For the underlying network, we assume low mobility, that is, the mean time to transmit a data package (including the route discovery process) is significantly lower than the time between major variations in the topology. This assumption justifies a rather constant failing probability for the nodes and thus for the transmission along each of the different paths. This, of course, requires an on-demand routing and no static routing table.

This work is performed as a part of the European EYES project (IST-2001-34734) on self-organizing and collaborative energy-efficient sensor networks [6]. It addresses the convergence of distributed information processing, wireless communication and mobile computing.

This paper focuses on the way E_k has to be constructed. Multipath degree, amount of traffic and failure probability of the nodes are the constraints that are taken into consideration. Section II presents an overview of the multipath routing problem and the already existent solutions. The way of how *braided multipath* can use the results of this paper is also analyzed. Section III focuses on how to compute E_k and what approximations can be made. The obtained theoretical results are verified by several simulations described in Section IV. The paper ends with conclusions and directions for the future work.

II. Multipath Routing

This section describes several aspects of multipath routing. The computation of the E_k value assumes multiple *disjoint paths* from the source to the destination. The case of *braided paths* is also taken into consideration and we explain how our results can be used also in this case.

A. Multipath Routing Overview

Multipath Routing allows the establishment of multiple paths between source and destination, which provides an easy mechanism to increase the likelihood of reliable data delivery by sending multiple copies of data along different paths.

Several different multipath routing algorithms have been studied by the prior work. The *Temporally Ordered Routing Algorithm* [5] provides loop free multiple alternate paths for mobile wireless network by maintaining a "destination oriented" directed acyclic graph (DAG) from the source. It rapidly adapts to topological changes, and has the ability to detect network partitions and erase all invalid routes within a finite time.

Dynamic Source Routing (DSR) depends on query floods to discover routes whenever a new routed is needed. Intelligent multipath extensions by Napsipuri and Das [4] have been added to reduce the frequency of routing discovery flooding, while maintaining several disjoint alternate paths between source and destination.

Another candidate for multipath routing is *Directed Diffusion* [3], which features data centric dissemination and in network data aggregation. It can realize robust multipath delivery, empirically adapt to a small subset of network paths, and achieve significant energy savings when intermediate nodes aggregate responses to queries. Based on directed diffusion, a novel braided multipath routing scheme, which results in several partially disjoint paths, is studied by D.Ganesan [7]. Results show it is viable alternative for energy efficient recovery from failures in WSN.

B. Disjoint and Braided Multipath

Out of many possible designs for multipath routing protocols, two distinct mechanisms exist: disjoint and braided.

- *disjoint multipath* routing tries to construct alternate paths which are node disjoint with the primary path, and with each other. Thus they are unaffected by failure on the primary path. But those alternate paths could potentially have much longer latency than the primary path and therefore expend significantly more energy than that on the primary path.
- *braided multipath* routing relaxes the requirement for node disjointness, which means alternate paths in a braid are partially overlay with the primary path, not completely node disjoint.

In this paper, we only consider the disjoint multipath cases, because the braided multipath can easily be transformed into disjoint multipath. As shown in Figure 1, there exists four multipaths between the source and destination and three of them are overlaid. The data splitting algorithm uses only the failing probability of each path. This means that we can consider the 3 paths from source to node X as a separate disjoint multipath problem, apply the algorithm presented in the next section to it and get a failing probability for it. This way we can reduce the group of braided paths to a single path. It is easy to see that the simple example presented here can be generalized for any type of configuration.

III. Data Splitting Across Multiple Paths

In this section we present some considerations on how to predict the number of paths that will successfully deliver a message among the multiple disjoint paths obtained from the route discovery process. Furthermore, we will give an approximation that allows for a term that can be used to predict the successful delivery. The increase of the probability for this successful delivery comes at the trade-off of added redundancy.

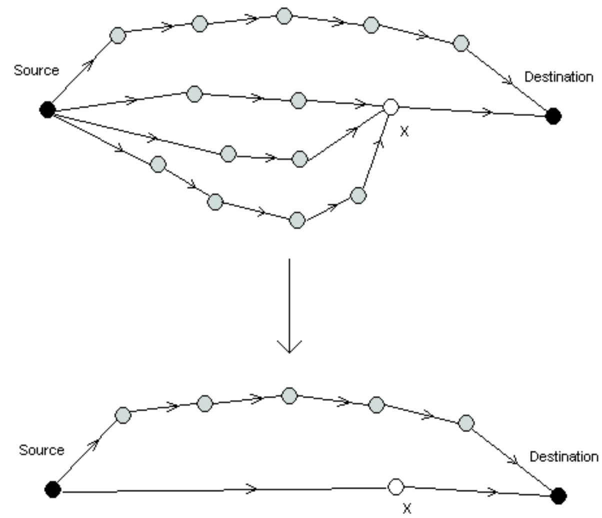

Fig. 1. Braided multipath case

Of course, one can send the whole message along each of the available paths, but the overhead induced by this will be too high. The entire data package to be sent from the source to the destination over the available k disjoint paths will be split up into smaller subpackets of equal size with added redundancy. The number of created subpackets corresponds to the number of available paths. Only a smaller number of these subpackets will then be needed at the destination to reconstruct the original message. There exist several fast and simple (i.e. linear) forward error correcting codes (or erasure codes) that allow the reconstruction of an original message that has been split up, with added redundancy, and of which not all parts arrive at the destination [8]. In the following, we will focus on approximating a value E_k that gives, with high probability, the number of successful paths. This value will then be used to determine the amount of redundancy to be added for the split message transmission.

The total number of subpackets as well as the added redundancy is a function dependent on the multipath degree and on the failing probabilities of the available paths. As these values change according to the positions of the source and the destination in the network, each source must be able to decide on the parameters for the error correcting codes before the transmission of the actual data subpackets.

A. Expected Number of Successful Paths (k given)

Suppose we want to send a data packet from a source to a destination and the process of route construction is finished, resulting in k different paths that are to be used. Each path has some rate $p_i (i = 1, \dots, k)$ that corresponds to the probability of successfully delivering a message to the destination.

This setting corresponds to a repeated Bernoulli experiment, the i-th subrun corresponding to the message transmission along the i-th path. Note that we consider node-disjoint paths and therefore can assume these experiments to be independent of each other. Let $S_k : \{0, 1\}^k \to \mathbb{N}$ be the random variable

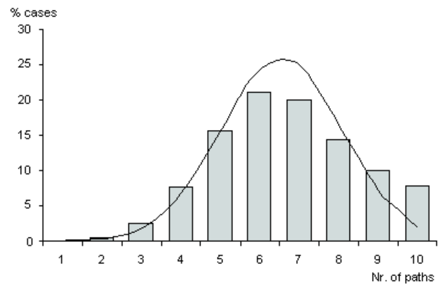

Fig. 2. Histogram for multipath of degree 10

corresponding to the number of successfully delivering paths. (For S_k, each subrun is assigned a 1 if the transmission was a success along the respective path, and 0 if it failed. Then, S_k represents the sum of these for the subruns, and clearly $S_k \leq k$.) Then, the expectation for the total number of successful paths is given by

$$E(S_k) = \sum_{i=1}^{k} p_i.$$

The distribution of the above repeated Bernoulli experiment can be approximated by normal distribution. This will be used to obtain a good estimator for E_k for a given bound α, the overall probability of successfully reconstructing the original message at the destination. More formal, we want to deliver a good estimation for the value of E_k for a desired bound α such that $P(S_k \geq E_k) \geq \alpha$ holds.

In order to approximate by normal distribution $N(\mu, \sigma)$, the mean μ and the standard deviation σ are needed. In our case, μ will be given by the expectation for S_k, i.e. by the sum of the probabilities of successful delivery along each path, and thus we set

$$\mu := E(S_k) = \sum_{i=1}^{k} p_i.$$

Accordingly, we obtain the standard deviation by setting

$$\sigma^2 := \sum_{i=1}^{k} p_i(1 - p_i).$$

Figure 2 shows a 100,000 run histogram for a multipath of degree 10, each path having 5 to 15 intermediate nodes between source and destination (avg. 10), and each node having a failing for 20% of the simulation time with a probability of 0.2. This corresponds to an average probability of a successful delivery of $p_i = 0.96$ for each path. Additionally, the graph for the normal distribution with the estimated values of $\mu = 6.65$ and $\sigma = 1.5$ is given.

Obviously, each combination of the multipath degree k and different probabilities $p_1, \ldots, p_k$ will yield a different normal distribution. To overcome this problem, we transform to the standard normal distribution $N(0, 1)$. The random variable

$$S_k^* := \frac{S_k - \mu}{\sigma}$$

α	95%	90%	85%	80%	50%
x_α	-1.65	-1.28	-1.03	-0.85	0

TABLE I

SOME VALUES FOR THE BOUND α

is $N(0, 1)$-distributed.

Now, consider a given bound α for the desired probability of being able to reconstruct the original message at the destination after being sent along the different paths. For the standard normal distribution, the values of the bound x_α for any given α such that the probability

$$P(S_k^* \geq x_\alpha) \geq \alpha$$

holds are known. Some value-pairs are presented in table I. Note that these values are independent from the number of paths k used to send data.

Using the above estimations, we transform the arguement

$$S_k^* = \frac{S_k - \mu}{\sigma} \geq x_\alpha$$

and obtain the following probability

$$P(S_k \geq x_\alpha \cdot \sigma + \mu) \geq \alpha.$$

We can therefore use $x_\alpha \cdot \sigma + \mu$ as E_k for the forward error correction code and set

$$E_k := \max\{\lfloor x_\alpha \cdot \sigma + \mu \rfloor, 1\}.$$

In terms of the input for the decision algorithm, the value for E_k is given by the following expression

$$E_k = \max\{\lfloor x_\alpha \cdot \sqrt{\sum_{i=1}^{k} p_i(1 - p_i)} + \sum_{i=1}^{k} p_i \rfloor, 1\},$$

which gives the number of successfully delivering paths with the overall success probability of α.

IV. SIMULATION RESULTS

We have performed several simulations (the code was written in C++) in order to verify the theoretical results and to get a better understanding on how well our solution works. The obtained results confirmed the presented theory and the main conclusions are presented bellow.

We have assumed that in a given sensor network there exists a routing algorithm that returns several paths from a source to a destination. We have also assumed that the multipath degree varies between 2 and 10 and each path has a length of 5 to 15 intermediate nodes (6 to 16 hops) between the source and the destination. For each simulation we considered that the sensor nodes have a failing probability between 0 and an upper bound. This bound took values from 0.01 up to 0.25 with a step of 0.01. For each multipath degree and node failing probability bound we have performed 100000 simulations.

A *failed transmission* is the case when the destination could not reconstruct the original message from the source

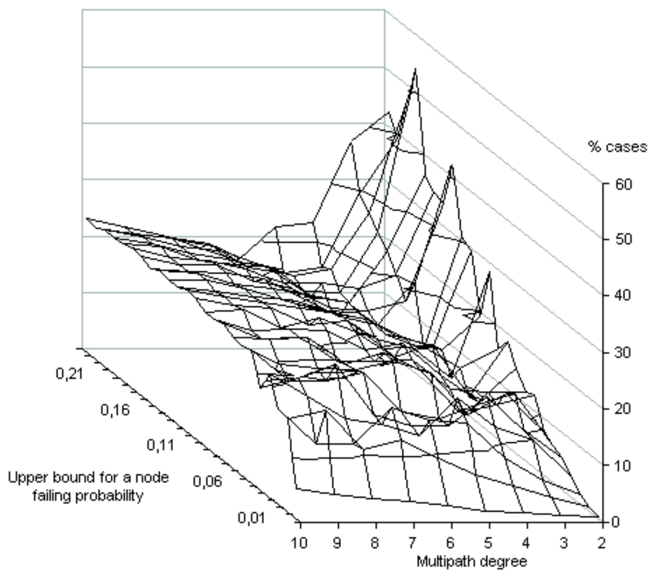

Fig. 3. Failed transmissions percentage

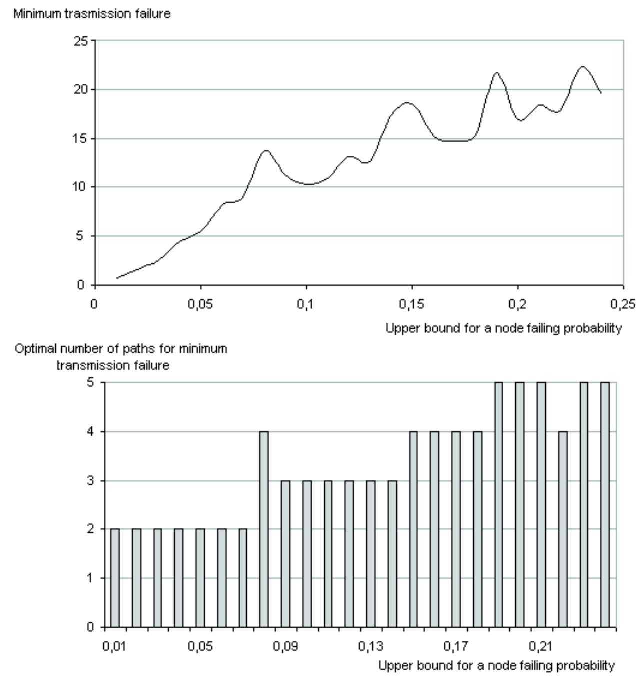

Fig. 4. Minimum transmission failure and number of paths used

(when it received a smaller number of subpackets than E_k). Figure 3 presents the total failure probability of the algorithm function of the upper bound of the node failing probability and the multipath degree. In other words, this is the percentage of failed transmissions out of all transmissions. E_k can be calculated as shown in Section III or in an approximative way in order to reduce the amount of computation in the sensor nodes. The difference between the cases when E_k is computed in an exact way and when the approximation is used is very small - we have found that the relative error is less than 5%. The approximation is:

$$p_i = \prod_{j=1}^{n} (1 - q_{ij}) \approx (1 - \bar{q}_i)^n$$

where $\bar{q}_i = \frac{\sum_{j=1}^{n} q_{ij}}{n}$. q_{ij} represents the failing probability of the j-th node from the i-th path, while p_i is the probability of the i-th path to be successful.

As one can notice in Figure 3, for each upper bound of node failing probability there is an optimal number of paths, for which the failed transmission probability gets to a minimum. With the increase in the number of paths, the probability of error also increases! This can be explained by the fact that E_k is not constant, it also varies. The upper graphic in Figure 4 presents the minimum transmission failure that occurred in our simulations while the lower one gives the corresponding number of paths to show where this minimum was achieved.

The next set of results concerns the traffic used. The packet that the source needs to transmit to the destination across the multipath is considered to have the size 1. The total traffic (the case when the full data packet is sent through each subpath) and the traffic used when applying data splitting across the multiple paths is given in Figure 5. The big difference between the two cases is the reason for which this algorithm should be employed: we reduce the used traffic significantly with an increase of failed transmissions (see also Figure 6 and Figure 7).

Finally, we have analyzed what happens when changing the x_α coefficient. Actually we use a coefficient γ such that: $E_k = \gamma \cdot E_{k0}$, where $E_{k0} = \sum_{i=1}^{k} p_i$. The variation of x_α maps on the variation of γ. Figure 6 and Figure 7 show how the traffic and the number of failed transmissions vary for different values of the γ coefficient (in this example we have chosen a given upper bound of the failing probability of the nodes of 0.15). For all the others upper bounds of node failing probability the values respect the same proportions. Based on these curves one can choose the multipath degree for a maximum failing probability, taking into consideration how much traffic the network can support or how many failures can be handled.

V. Conclusions

One way of assuring that a data packet reaches the destination in a sensor network is by using multipath routing algorithms. The method has the disadvantage of increasing the overall traffic substantially. In this paper we introduce a new idea of splitting the original data packet in subpackets and sending each one of them through one of the multiple paths. Even if some of them are lost, the original message can be reconstructed.

Because of the failures that might appear (mainly due to mobility of the nodes and to the wireless transmissions) a path rater mechanism needes to be used. After the rating process, it will be decided on the amount of redundancy to be added in order to use the available resources efficiently.

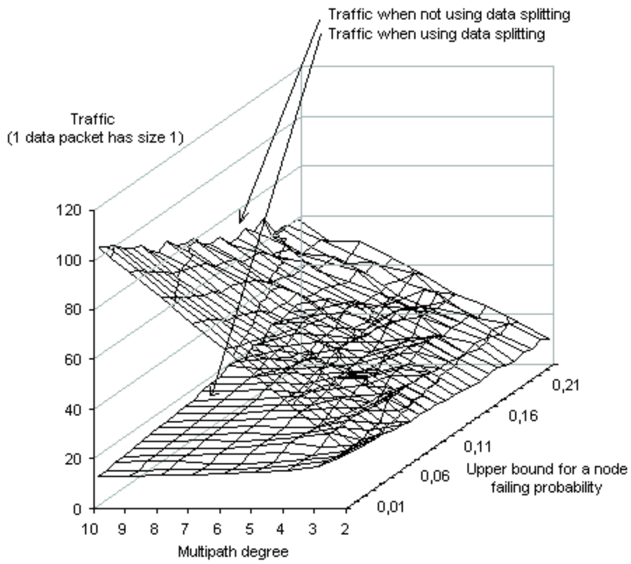

Fig. 5. Traffic when using data splitting and when not

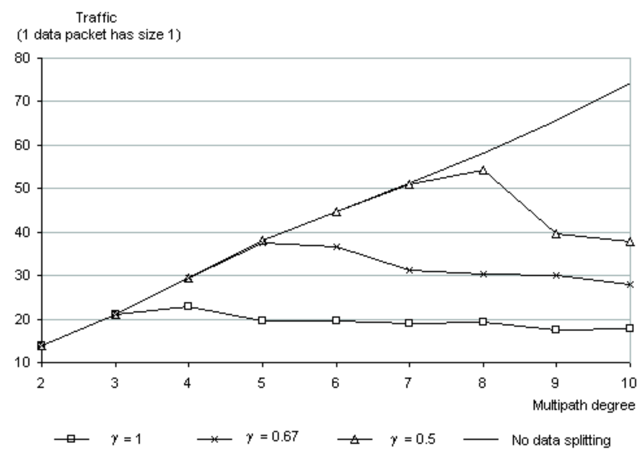

Fig. 6. Traffic for different γ values

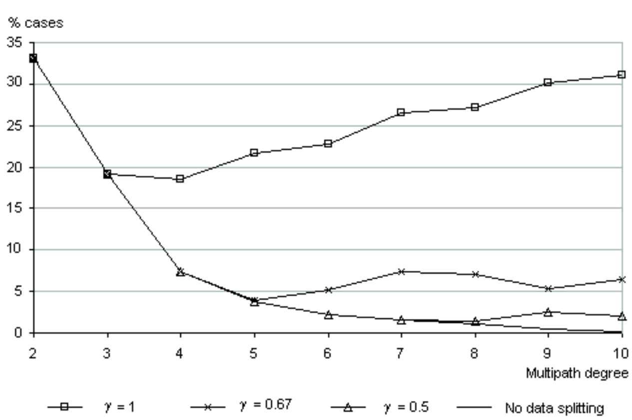

Fig. 7. Failed transmissions percentage for different γ values

By splitting the data across the multiple paths, the percentage of errors increases while the amount of traffic goes to lower values. This gives us a way to adjust the reliability while keeping the data traffic low. When using a lower value for E_k than the calculated one, the traffic increases but the percentage of failures decreases (down to the case of all paths failing) (see Figure 6 and Figure 7). The variables that the user can tune are: the amount of traffic and the multipath degree versus the message delivery failure.

One of the interesting results of this paper shows that, contrary to one's intuition, for a given maximum node failure probability, using a higher multipath degree than the optimum one actually increases the total probability of failure (see Figure 7). This happens due to the way the algorithm computes E_k.

The proposed scheme is useful for delivering data in unreliable environments. The amount of redundancy is low and so, the traffic overhead stays at low values as well. The latency of transmitting the data decreases proportionaly with the increase of the multipath degree. It also decreases because no feedback mechanism is required.

The future work focuses on finding ways of estimating the failing probabilities of each node. We consider it to be related to the failure of the hardware itself as well as the properties of the medium. The history of each node behaviour will have an important role in estimating its failure probability.

Our scheme, although focused on WSNs, can be incorporated into any routing scheme to improve reliable packet delivery in the face of a dynamic (wireless) environment where nodes move and connections break. The network can be fully distributed because the decisions are taken only at the source node and at the destination (a global view is not needed).

References

[1] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehaviour in mobile ad hoc networks," in *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, 2000, pp. 255–265.

[2] G. Pottie and R. Molva, "Embedding the internet: wireless integrated network sensors," in *Communications of ACM, 43(5), 51–58*, 2000.

[3] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks," in *Proc. Sixth Annual International Conference on Mobile Computing and Networks*, 2000.

[4] Nasipuri and S. Das, "On-Demand Multipath Routing for Mobile Ad Hoc Networks," in *8th Intl. Conference on Computer Communications and Networks (IC3N 99)*, 1999.

[5] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proceedings of IEEE INFOCOM'97 Conf.*, April 1997.

[6] EYES project, "website," http://eyes.eu.org.

[7] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 4, pp. 11–25, 2001.

[8] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," in *ITW 1998, San Diego Feb. 1998*, 1998, pp. 150–159. [Online]. Available: citeseer.nj.nec.com/luby98practical.html