H-SPREAD: A Hybrid Multipath Scheme for Secure and Reliable Data Collection in Wireless Sensor Networks

Wenjing Lou, *Member, IEEE*, Younggoo Kwon, *Member, IEEE*

***Abstract*— Communication security and reliability are two important issues in any network. A typical communication task in a wireless sensor network is for every sensor node to sense its local environment and, upon request, sends data of interest back to a base station. In this paper, we propose a hybrid multipath scheme (H-SPREAD) to improve both security and reliability of this task in a potentially hostile and unreliable wireless sensor network. The new scheme is based on a distributed *N-to-1* multipath discovery protocol which is able to find multiple node-disjoint paths from every sensor node to the base station simultaneously in one route discovery process. Then, a hybrid multipath data collection scheme is proposed. On the one hand, end-to-end multipath data dispersion, combined with secret sharing, enhances the security of end-to-end data delivery in the sense that the compromise of a small number of paths will not result in the compromise of a data message in the face of adversarial nodes. On the other hand, in the face of unreliable wireless links and/or sensor nodes, alternate path routing available at each sensor node improves reliability of each packet transmission significantly. The extensive simulation results show that our hybrid multipath scheme is very efficient in improving both security and reliability of the data collection service seamlessly.**

***Index Terms*— Sensor networks, Multipath routing protocol, Reliability, Security**

I. Introduction

Recent advancement in microprocessor, memory, and wireless networking and communication technologies have paved the way for the deployment of wireless sensor networks. A wireless sensor network (WSN) typically is composed of a large number of low-cost sensor nodes which work collectively to carry out some real-time sensing and monitoring tasks within a designated area. This emerging technology has drawn growing attention recently since it provides a promising solution to many challenging tasks, such as military sensing and tracking in a hostile ground, remote sensing in nuclear plants, mines, and other hazardous industrial venues, real-time traffic monitoring, realtime weather monitoring, wild animal monitoring and tracking, etc.

Realization of a WSN faces many challenges. Although some of the wireless ad hoc networking techniques are applicable to WSNs, a WSN differs from a mobile ad hoc network (MANET) in many aspects [1]. For example, the number of nodes in a WSN is usually much larger than that in a MANET. Typical sensor nodes are more resource constrained in terms of power, computational capabilities, and memory. Sensor nodes are typically randomly and densely deployed (e.g., by aerial scattering) within the target sensing area. The post-deployment topology is not predetermined. Although in many cases the nodes are static, the topology might change frequently because of unreliable wireless links and failure-prone sensor nodes. Moreover, a MANET is typically infrastructureless, end-to-end communications are the common communication pattern. While a WSN is typically formed around one (or more) *base station* (BS, a.k.a. *sink*). All the sensor nodes are usually designed to sense its local environment and, upon request, send data of interest back to the base station which is generally several magnitudes more powerful than sensor nodes and serves as a concentration point of the WSN and at the same time the nexus connecting the WSN to the rest of the world. Reliable and secure data collection is an important task in a WSN.

Reliability, defined as the successful end-to-end data delivery ratio, has been an issue in WSNs since nodes are prone to failure and wireless transmission between nodes are susceptible to all kinds of interferences. Security is another issue since sensor nodes, when deployed in a hostile ground, are subject to compromise. Generally it is not economically feasible to make sensor nodes tamper-proof, which means that once a node is compromised, all the secrets stored in that node, including cryptographic keys, may be compromised too, which jeopardizes information relayed by that node. Multipath traffic dispersion has been known as an effective strategy to improve reliability in the face of path failures caused by unreliable links and frequent topological changes [2]. However, improved reliability can be achieved only at the cost of excessive redundancy, that is, sending more data than necessary along multiple paths such that the reconstruction of original information can tolerate up to a certain amount of path failure/packet loss. In [3], we proposed a Secure Protocol for REliable dAta Delivery (SPREAD) for end-to-end message delivery in a MANET. In stead of using the single shortest path to route data from one node to the other, SPREAD splits a message into multiple shares using the secret sharing scheme and then delivers message shares to the destination via multiple independent paths. The SPREAD idea was shown to be effective in improving security in the sense that it is more resistant to collusion attacks of up to a certain number of compromised nodes. However, from the security perspective, little or none redundancy should be added to the information transmitted. The amount of information redundancy required

W. Lou is with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute. Email:wjlou@ece.wpi.edu. Y. Kwon is with the Department of Electronic Engineering, Konkuk University, Korea. Email: kwonyg@konkuk.ac.kr.

makes security and reliability a seemingly contradicting objectives for schemes based on multipath routing.

In this paper, we first propose a distributed *N-to-1* multipath discovery protocol, based on which we then propose a hybrid multipath scheme to achieve more reliable and more secure data collection task in WSNs. While most of multipath routing protocols are source-initiated and aim to find multiple disjoint or partially disjoint paths between a single source-destination pair [4]–[11], the distinct feature of our *N-to-1* multipath discovery protocol is that it is receiver-initiated (i.e., BS initiated) and at the end of one route discovery process, the protocol finds every sensor node a set of node-disjoint paths to the BS simultaneously. It is highly efficient, with an average overhead of less than one routing message per path. Then we investigate the hybrid multipath data collection scheme, which combines concurrent multipath dispersion for end-to-end data collection and alternate path routing for each individual packet transmission. The simulation results show that our hybrid scheme, Hybrid-SPREAD or in short H-SPREAD, can achieve significantly better reliability and better security seamlessly with little or even none redundancy. The proposed scheme is extremely suitable for WSNs where the major task is for the base station to collect sensor readings from all the sensor nodes simultaneously.

The rest of the paper is organized as follows. The SPREAD idea is briefly reviewed in section II. The distributed *N-to-1* multipath discovery protocol and its evaluation are presented in section III. In section IV, an framework for security and reliability analysis is provided and the active packet salvaging strategy is proposed. The overall performance of H-SPREAD is evaluated in section V. Finally, related work is reviewed in section VI and conclusion is drawn in section VII.

II. A Brief Review of SPREAD

In [3], we proposed the SPREAD scheme as a complementary mechanism to enhance data confidentiality in a MANET. The basic idea and operation of SPREAD is as follows. A secret message m is transformed into multiple shares, $S_1, S_2, \cdots$, by secret sharing scheme, and then delivered to the destination via multiple independent paths. Due to the salient features of secret sharing and the distributed fashion of multipath delivery, the SPREAD has been shown to be more resilient to a collusive attack by up to a certain number of compromised nodes, namely, even if a small number of paths/nodes/shares are compromised, the message as a whole is not compromised.

A number of coding schemes can be used to split the traffic for multipath routing in order to enhance reliability. Examples include well-known Reed-Solomon codes, diversity coding [12], multiple description coding, etc. In the SPREAD scheme [3], we used the threshold secret sharing scheme to split the information. A (T, N) threshold secret sharing scheme could transform a secret into N pieces, called *shares* or *shadows*. The nice property of the N shares is that form any less than T shares one cannot learn anything about the secret, while with an effective algorithm, one can reconstruct the secret from any T out of N shares. The generation of shares is very simple - by evaluating a polynomial of degree $(T-1)$

$$f(x) = (a_0 + a_1 x + \cdots + a_{T-1} x^{T-1}) \bmod p$$

at point $x = i$ to obtain the $i-$th share:

$$S_i = f(i)$$

where $a_0, a_1, a_2, \ldots, a_{T-1}$ are secret bits while p is a large prime number greater than any of the coefficients and can be made public. Note when all the coefficients are used to carry secret bits, the fraction of redundant information is $\frac{N-T}{N}$. For $T = N$, there is no redundant information resulting from secret sharing[1].

According to the fundamental theorem of algebra, T values of a polynomial of degree $(T-1)$ can completely determine the polynomial (i.e., all its coefficients), while any fewer values cannot determine the polynomial (at least computationally difficult). Thus, any T shares can reconstruct the original secret bits, but any fewer shares cannot. Efficient $(O(T \log^2 T))$ algorithms have been developed for polynomial evaluation and interpolation [13]. In addition, the reconstruction is done in the base station, which is not computationally constrained very much. Therefore, in our H-SPREAD scheme, we still choose secret sharing as the coding scheme.

III. N-to-1 Multipath Discovery Protocol

A challenging job in any multipath routing scheme is the efficient and effective multipath routing protocols. In [3], we discussed multipath finding techniques between a single source-destination pair. In fact, most of current multipath routing protocols fall into this category (refer to Section VI). In response to the communication pattern in a WSN, in this paper we propose a novel *N-to-1* multipath discovery protocol. Instead of finding multiple paths between a specific source and a specific destination, the proposed *N-to-1* multipath discovery protocol takes advantage of flooding in a typical route discovery process and find multiple node-disjoint paths from every sensor node to the common destination (i.e., the sink node) simultaneously. The proposed *N-to-1* multipath discovery protocol is essentially the enabling technique for our hybrid data collection scheme. Therefore we present the distributed protocol and evaluate its path finding capability in this section first.

A. Motivation and Overview

A typical task of a WSN is data collection where the base station broadcasts the request for the data of interest and every sensor node (or nodes that have the data of interest) sends its readings back to the base station. For this purpose, Berkeley's TinyOS sensor platform utilizes a flooding-based beaconing protocol. The base station periodically broadcasts a route update. Each sensor node when receiving the update for the first time rebroadcasts the update and marks the node from which it receives the update as its parent. The algorithm continues recursively till every node in the network has rebroadcasted the update once and finds its parent. What

[1]The length of coefficients needs to be one bit shorter than that of p.

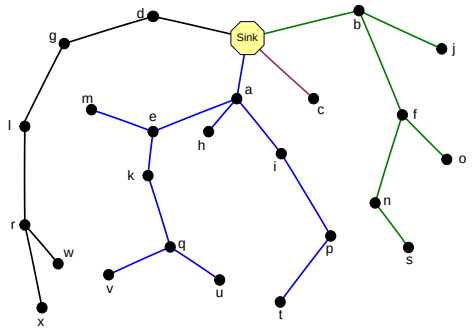


Fig. 1. Spanning tree created by flooding

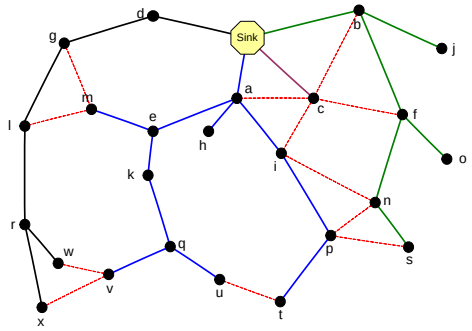


Fig. 2. A simple multipath extension of flooding

follows is that every node forwards the packets it received or generated to its parent until the packets reach the base station [14]. As illustrated in Fig. 1, the beaconing protocol essentially constructs a breadth first spanning tree rooted at a base station. It finds every sensor node a single path back to the base station efficiently. However, reliability and security suffer from the single path routing. The failure of a single node or link will disrupt data flow from the node itself and all its children. Similarly, compromise of a single node will cause the information leakage from the node and all its children.

The proposed *N-to-1* multipath discovery protocol is based on the simple flooding initiated at the BS. Then, by carefully incorporating two other mechanisms into the protocol design, it is able to find every sensor node multiple node-disjoint paths back to the BS at the end of a distributed path discovery process. To facilitate the understanding, we present the multipath discovery procedure in two phases, with each phase implementing one of the mechanisms. In fact, the second phase can be started at each individual node in a distributed fashion without considering the completion of phase one at other nodes. The mechanism used in phase one, termed *branch aware flooding*, takes advantage of the simple flooding technique. Without introducing additional routing messages, the mechanism is able to find a certain number of node-disjoint paths, depending on the density of the network topology. The mechanism used in phase two, termed *multipath extension of flooding*, helps to exchange the node-disjoint paths found in phase one among nodes on different branches. At the cost of some more message exchanges, it is able to increase the number of paths found at each sensor node.

B. Phase One: Branch Aware Flooding

The general form of routing messages in both phases is $\{mtype, mid, nid, bid, cst, path\}$, where *mtype* indicates the type of message. We define *mtype*="RPRI" for phase one, which refers to "primary" because primary paths (i.e., on the shortest path tree) are found by this type of messages; *mid* is the sequence number of the current routing update; *nid* is the identifier of the node sending out the message; *bid* is the identifier of the branch defined as *nid* of the node closest to the BS in the branch; *path* contains a sequence of nodes which the message has travelled; and *cst* is the cost of the *path*.

The propagation of RPRI messages follows exactly the same way as the TinyOS beaconing protocol. The BS initializes a routing update periodically (or on demand) by broadcasting message $\{RPRI, mid, Sink, \emptyset, 0, (Sink)\}$. Every node, say z, when hearing a message $\{RPRI, mid, nid, bid, cst, path\}$ for the first time, marks node nid as its parent, and it also learns the primary path back to the BS by following the reverse order of $p = path + (z)$. It then forms a new routing message $\{RPRI, mid, z, (bid == \emptyset)?z : bid, cst + cost(z, parent(z)), path + (z)\}$ according to the following rules: replacing nid field with its own ID; if bid field is $\emptyset$, replacing bid field with its own ID, otherwise keeping the original bid intact; updating cst field by adding the cost from z to the node from which this message is received; and updating $path$ field by appending its own ID at the end of the old path. Node z then rebroadcasts the new message in the neighborhood.

In the simple flooding protocol (such as the beaconing protocol), a node simply ignores the duplicate route update messages from other nodes. However, in our branch aware flooding, when a node z hears the same message (i.e., identified by the same mid) from a neighbor, it will check the content of the message and mark the neighbor accordingly. If the message has the same bid as node z itself, z will mark that neighbor as a *child* or *sibling*, according to the $path$ contained in the message; if the message has a different bid, which means the message is from another branch, z will mark that neighbor as a *cousin*. Node z maintains an alternate path set $\mathcal{Q}_z$. Once receiving a message from a cousin node, z will further examine the path contained in the message. If the new path $q = path + (z)$ is disjoint from the primary path p and any other alternate path with lower cost in $\mathcal{Q}_z$, the new path q will be included into the $\mathcal{Q}_z$, while at the same time, paths with higher cost than q that share common nodes with q will be removed from $\mathcal{Q}_z$. Same as the beaconing protocol, the propagation of RPRI messages is terminated at the leaf nodes when each node has rebroadcasted the message once and only once.

The branch aware routing technique is actually based on the following observation. As show in Fig. 2, the number of branches a tree has depends on the number of immediate neighbors the base station has (e.g., 4 branches in the example). The maximum number of node-disjoint paths from any node to the base station is thus bounded by the number of branches. We notice that while each node has a *primary* path to the base station by following its tree links up, a link

between two nodes that belong to two different branches will provide each node an alternate disjoint path to the base station through the other. For example, as shown in Fig. 2, while node w has the primary path $(w - r - l - g - d - Sink)$ back to the base station, it learns another alternate path $(w - v - q - k - e - a - Sink)$ from node v which is not in the same branch as w when overhearing v's broadcast. The branch aware flooding is therefore designed to allow nodes to go across a cousin link thereby finding disjoint paths in other branches. This mechanism takes advantage of the broadcast nature of wireless communication. Without introducing extra routing messages, nodes that have cousin neighbors are able to find a few disjoint paths.

C. Phase Two: Multipath Extension of Flooding

The ability of finding extra paths by branch aware flooding is limited to nodes that have cousin neighbors. In what follows, we present a second mechanism/phase proposed for our *N-to-1* multipath discovery protocol - a multipath extension to flooding technique, which is able to find more node-disjoint paths at each sensor node at the cost of some extra message exchanges.

Phase two message exchange uses the same message format but with $mtype$ field set to "RALT", which refers to "alternate" because this type of messages find alternate paths. The RALT messages are used to further propagate the alternate paths found at one node to its parent and sibling/cousin neighbors[2]. The propagation of RALT messages is initiated distributively and independently at each node where an alternate disjoint path(s) is found during branch aware flooding. For each alternate path q, node z forms a RALT message $\{RALT, mid, z, q.bid, q.cst, q\}$ and broadcasts it in its neighborhood.

Upon receiving a RALT message $\{RALT, mid, nid, bid, cst, path\}$, node z will ignore it if it is from its parent. Otherwise, it will check and see if itself is already in the path contained in the message. If not, node z learns about a new path $q = path + (z)$. Again, node z includes the new path q into its alternate path set $\mathcal{Q}_z$ if q is disjoint from any other paths in $\mathcal{Q}_z$ of lower cost. If q is included, node z excludes from the path set $\mathcal{Q}_z$ paths of higher cost and intersecting with q. Whenever a new path q is added to $\mathcal{Q}_z$, node z forms a new RALT message $\{RALT, mid, z, q.bid, q.cst, q\}$ and broadcasts it in the neighborhood.

The propagation of RALT messages terminates when no new disjoint path is added to any path set. At this time, each node has found a set of disjoint paths to the BS.

The rationale behind the design of phase two mechanism is to maximize the number of disjoint paths at each node by further propagating alternate paths found at phase one across multiple branches. Using the same example as shown in Fig. 2, notice that if w further propagates the disjoint paths it learned to its neighbors, its parent or siblings/cousins might learn a new disjoint path as well. For example, node r has the primary path $(r - l - g - d - Sink)$. When it hears a disjoint path $(w - v - q - k - e - a - Sink)$ from w and it does not yet know a path through branch a, it learns a new disjoint path $(r - w - v - q - k - e - a - Sink)$. The tradeoff of the second phase is that it finds more disjoint paths with additional routing messages.

[2]Not intended for children because the parent node must be in the primary path of the child node.

TABLE I

PARAMETERS OF NETWORKS SIMULATED

Transmission range (TR)	15 m	20 m	25 m	30 m
Average node degree (*d*)	6.05	10.19	15.29	21
Average network diameter (*D*)	10.56	6.09	4.72	3.85

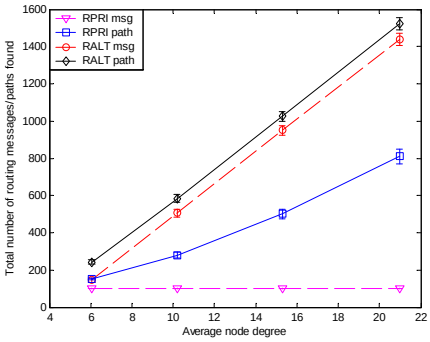


Fig. 3. Path Finding Capability

D. Performance Evaluation

We use simulations to evaluate the performance of the proposed *N-to-1* multipath routing protocol. We simulate a WSN consisting of 100 nodes randomly deployed in a field of 100m $\times$ 100m square area. The base station is located in the middle of one edge. Nodes have same transmission range in one experiment. In order to evaluate the impact of edge density on the performance, we vary transmission range in different experiments to adjust edge density in the network. We tried four different transmission ranges, 15, 20, 25, and 30 meters[3]. Table I summarizes some topological parameters of the networks simulated when using different transmission ranges, including the average node degree d (i.e., the number of neighbors a node has) and the average network diameter D (i.e., the maximum hop count from any sensor node to the BS based on the shortest path routing). The simulation results are averaged over 60 random network deployments. The 95% confidence intervals are shown in the figure.

Fig. 3 shows the total number of routing messages and the total number of disjoint paths found in the simulated networks. We observe that the branch-aware flooding mechanism find disjoint paths without introducing any extra message exchanges. When edge density is high, say when average node degree is 22, this simple modification could find an average of 8 node-disjoint paths per node. Our multipath extension of flooding mechanism, although requiring more

[3]Randomly generated networks sometimes are not connected if the edge density is not high. In our simulation, 70% of networks are not connected when average node degree is 7 (e.g., TR=15m) and 5% of networks are not connected when average node degree is 11 (e.g., TR=20m). Only the results from connected networks are considered here.

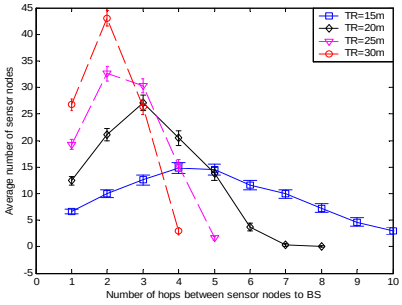


Fig. 4. The distribution of nodes in terms of distance

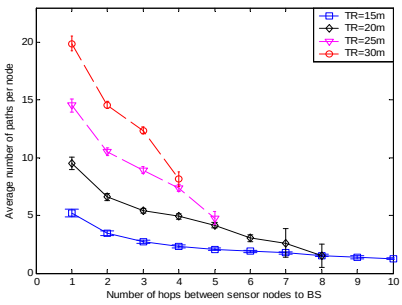


Fig. 5. The distribution of paths in terms of distance

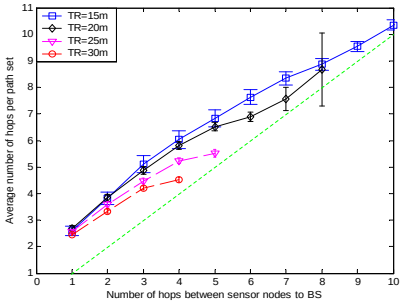


Fig. 6. The quality of the alternate paths

message exchange, is able to find more paths. The results show that, in general, the routing algorithm is very efficient in terms of path finding - the per path cost is less than one message.

The next three figures reveal some more characteristics of the paths found. Fig. 4 shows the distribution of nodes in terms of their distances from the BS. Correspondingly, Fig. 5 depicts the average number of node-disjoint paths found per node with regard to distance between sensor nodes and the BS. It is observed that the closer the node to the BS, the more paths from that node to the BS. This is reasonable because it is harder to find node-disjoint paths when nodes are far away from the BS as each alternate path has to find more unused nodes to reach the destination. This property is actually desirable for the tasks we are considering. Since the typical task of the network is to collect data from all sensor nodes, packets are travelling from everywhere toward the BS. Nodes that are closer to the BS would be used more for forwarding traffic thus it is more desirable for those nodes to be more reliable. More alternate paths gives a node more choices in the face of node or link failures thus inferring better reliability. Fig. 6 shows the average hop count per path correspondingly. The dotted line is plotted as the reference, indicating the shortest distance to the BS. It is observed that the average path length is typically 1 or 2 hops longer than the shortest path, regardless the length of the shortest path.

The typical data collection in a WSN involves the following communication patterns: (a) broadcast from the base station to sensor nodes (e.g., requests of data of interest); (b) from sensor nodes to the base station (e.g., sending back the sensor readings); and (c) node to node communication (e.g., if aggregation of sensor readings are applied). In this section, we described our multipath discovery protocol which, similar to any on-demand routing protocol, starts with a route update initialized at the base station. This route update is a network wide broadcast thus can be used to fulfill the above mentioned type-(a) communication. Then at the end of the discovery, each node will be able to find a set of multiple node disjoint paths to the base station with which our hybrid data collection scheme can be implemented for the type-(b) communication. We do not consider data aggregation explicitly. However, if data aggregation centers are to be applied, a hierarchical routing structure can be constructed: from each sensor node to the aggregation center forms the lower layer and from each aggregation center to the base station forms the higher layer. Our proposed algorithm could be made applicable to each layer separately.

IV. Security and Reliability Analysis of H-SPREAD

In this section, we provide an analytical framework to analyze security and reliability achievable from the proposed H-SPREAD scheme. More specifically, by "reliability" we mean the probability that a message generated at one sensor node can actually be routed to the BS, and by "security" we mean the probability that adversaries/compromised nodes can intercept the message. The purpose of the security and reliability study is twofold: (a) provide analytical measures to evaluate the security and reliability of the data delivery when applying our scheme and (b) provide insights to design a share allocation scheme that will tolerate failures and compromised nodes in the best possible way (probabilistically).

A. Security Analysis

We first consider security. Assume that M disjoint paths have been selected to deliver the message. Vector $\underline{q} = [q_1, q_2, \cdots, q_M]$ denotes the security characteristics of the paths where q_i denoting the probability that the i-th path might be compromised[4]. Let vector $\underline{n} = [n_1, n_2, \cdots, n_M]$ denote the share allocation, meaning that n_i packets are assigned to the i-th path for delivery, n_i is an positive integer, and $\sum_{i=1}^{M} n_i = N$. We assume that if one path is compromised, all the shares travelling along that path are compromised. Since paths are mutually node-disjoint, we further assume that the compromising of each path is independent. Let $\psi(i)$ denote the indicator function on path i: with probability q_i, $\psi(i) = 1$ indicating the i-th path is compromised, and with probability $(1-q_i)$, $\psi(i) = 0$ indicating that the ith path is not compromised. Then, the number of shares compromised given the allocation $\underline{n}$ will be given by $\sum_{i=1}^{M} \psi(i) n_i$. Notice that the adversaries have to compromise a minimum of T shares

[4]As discussed in [3], here we consider the information leakage problem from the compromised nodes. The eavesdropping on the wireless links can be solved by physical layer measures or link layer encryption. Therefore, the probability that a path is compromised here is the probability that any one or more nodes on that path is compromised.

in order to compromise the message, the probability that the message might be compromised (P_C) is then given by

$$P_C(\underline{n}) = \Pr\{\mathcal{J} \geq T\} = \Pr\left\{\sum_{i=1}^{M} \psi(i) n_i \geq T\right\} \quad (1)$$

where $\mathcal{J}$ is the total number of shares compromised among the N packets delivered across the M paths.

In [3], we have shown that, depending on the number of paths selected to deliver the message (M)[5], the maximum security can be achieved when allocating the N shares onto the M paths in such a way that the adversaries must compromise all the M paths to compromise the message. The maximum security, measured as the minimum probability that the message might be compromised (P_C) is

$$P_C(\underline{n}) = \prod_{i=1}^{M} q_i$$

We have also shown in [3] that the SPREAD scheme can be designed in such a way that a certain degree of redundancy can be added without sacrificing the security. The maximum redundancy that can be added without scarifying security is bounded by

$$r < 1/M \quad (2)$$

where the redundancy factor r is defined as $r = 1 - T/N$ and M is the number of paths selected to deliver the message shares[6]. Then proper (T, N) values (i.e., integers) could be selected and optimal share allocations could be designed to achieve the maximum security and at the same time be able to tolerate a certain degree of packet losses. For example, for $M = 3,\ T = 8,\ N = 10$, the optimal share allocation would be $[4\ 3\ 3]$ by which the adversaries must compromise all the three paths to get enough shares while the scheme can tolerate the loss of 2 packets. Notice that the optimal share allocations proposed are in terms of security, although achieving maximum security, they actually do not tolerate the loss of a complete path. In fact, improved reliability obtained from the this type of traffic dispersion essentially comes from the redundancy added to the data traffic which should be much enough to tolerate one or more complete path failure. This makes security and reliability two contradicting objectives when using concurrent multipath routing approach, meaning that reliability relies on excessive redundancy while security requires no or little redundancy.

The direct calculation of P_C is hard since it needs to enumerate all possible (2^M) combinations of ψ_i and sum up the probabilities of those satisfying the condition. Fortunately, efficient algorithms have been developed for this type of problem in the reliability evaluation of *k-out-of-n* system models. Particularly our problem is similar to the reliability evaluation of *weighted-k-out-of-n* systems which were proposed in [15] and where an efficient algorithm for computing the reliability of the system was presented. The calculation of P_C can be derived as follows. Without loss of generality, we assume that $q_1 \leq q_2 \leq \cdots, \leq q_m$. Let $R(j, i)$ denotes the probability that a minimum of j packets are compromised from the first i paths. Then P_C can be calculated as

$$P_C(\underline{n}) = R(T, M)$$

where $R(T, M)$ can be calculated using the following recursive equation

$$R(j, i) = q_i R(j - n_i, i - 1) + (1 - q_i) R(j, i - 1) \quad (3)$$

which requires the following boundary conditions:

$$R(j, i) = 1, \qquad for\ j \leq 0, i \geq 0$$
$$R(j, 0) = 0, \qquad for\ j > 0$$

The computational complexity of Eq. (3) for $R(T, M)$ is $O(T(M-T+1))$ when $n_i = 1$ for all i. However, when $n_i > 1$ for all $1 \leq i \leq M$, the number of terms to be computed may be much less than that. In addition, the intermediate values are meaningful that represent $R(j, i)$ for $1 \leq j \leq T$ and $1 \leq i \leq M$. These values are helpful in understanding the assessment of the probability accumulatively.

B. Reliability Analysis

Next we evaluate the reliability of proposed data collection scheme, namely, how "reliable" a message generated at one sensor node can actually be routed to the BS. In fact, in a WSN, both sensor nodes and wireless links are error-prone. Node failure might be caused by physical node failure (e.g., physical damage or depletion of the battery) or heavy congestion at the node which causes packet drop due to buffer overflow. Link failure might be caused by media access contention, multiuser interference, or any interference which causes the radio signal not being correctly decoded at the intended receiver. If we assume that each node has equal probability p_{n0} to reliably relay a packet and each link has equal probability p_{l0} to reliably deliver a packet, the probability that a path consisting of H hops can successfully deliver a packet will be $p = {p_{n0}}^H {p_{l0}}^H$ (we assume the destination is reliable).

In [2], the authors proposed an analytical model for the evaluation of the reliability of multipath routing in mobile ad hoc networks. They considered packet loss due to topological changes therefore they modelled each path as a pure erasure channel, namely, if a path fails, all the segments transmitted on that path will be lost, otherwise, all the segments on that path is assumed received. This model is similar to the model we described for security analysis and can be viewed as a model of node failure problem in our case. The authors proposed to use a Gaussian approximation to calculate the reliability. Their analytical results showed that multipath routing is more resistant to node failure problem and the packet delivery ratio can be improved. However, the improved reliability highly relies on the excessive redundancy added (actually redundant paths) which will impair the security purpose as we just discussed.

In what follows, we evaluate the reliability from link failure perspective. In other words, we assume a network with

[5]The number of paths found between a source to a destination can be larger than M. Depends on the required security level, it is not necessary to use all the paths. Only the first M most secure paths that meet the desired security level are selected.

[6]A tighter bound is $r \leq 1/M - 1/N$

perfectly reliable nodes but unreliable links. The failure of a packet's transmission is caused by the link failure and is independent of other transmissions. We make this assumption in order to gain some insights on the impact of link failure problem on multipath routing. Moreover, in the combinatorics of network reliability an undirected problem with node failures can be transformed into a directed problem without node failures [16]. In some cases this assumption can often be avoided. For example, in the above example, the path reliability with node failures is just the probability that all nodes are operational times the path reliability without node failures.

Similar to the problem formulation for security analysis, we assume that M disjoint paths have been selected, each of which is reliably delivering packets with probability p_i ($i = 1, 2, \ldots, M$). Let again use vector $\underline{n} = [n_1, n_2, \cdots, n_M]$ denote the share allocation, which allocates n_i packets onto the i-th path. Let $\chi_i(j)$ denote the indicator function on path i: $\chi_i(j) = 0$ indicates the j-th packet is delivered successfully, and $\chi_i(j) = 1$ indicates that the j-th packet is lost. Thus, $p_i = \Pr\{\chi_i(j) = 0\}$, and $\sum_{j=1}^{n_i} \chi_i(j)$ is the number of packets lost over the i-th path. Based on this assumption and the fact that as long as the total number of lost packets among the N packets is less than or equal to $N - T$, the original information can be correctly recovered by the BS, we obtain that the probability of successful delivery (P_R) is given by

$$P_R(\underline{n}) = \Pr\{\mathcal{L} \leq N - T\} = \Pr\left\{\sum_{i=1}^{M}\sum_{j=1}^{n_i} \chi_i(j) \leq N - T\right\} \quad (4)$$

where $\mathcal{L}$ is the total number of lost packets among the N packets delivered by the M paths.

In fact, an intuitive solution to this problem is that, given $\underline{p}$, we can maximize this probability by assigning as many as possible shares to the most reliable paths and assign as few as possible shares to the least reliable paths. This result actually implies that, given the intermittent packet failure model described here, the best reliability is achieved by allocating all the packets to the most reliable path. The multipath traffic dispersion is not helpful in improving reliability in this situation.

C. Alternate Path Packet Salvaging

From the above analysis we notice that, multipath routing is effective in improving reliability in the face of persistent errors, such as node failure or persistent link errors (i.e., pure erasure channel model), while it does not help in the case that lost of packets is due to intermittent link failures. This observation indicates that security and reliability are two contradictive design objectives with regard to redundancy added - reliability requires more redundancy while security requires less or no redundancy.

We also notice that, the above analysis is based on the end-to-end concurrent multipath routing, meaning, the information is split at the source and the segments are spread onto multiple paths between the source and destination pair. However, single path routing is assumed for the delivery of each packet/share. This is true for most of multipath routing approaches proposed in the literature where the proposed multipath routing protocols aim to find multiple disjoint paths between a single source and destination pair on-demand. The end-to-end multipath routing approach essentially relies on the redundant paths to improve reliability, while the unreliability of each path remains unimproved. The results are therefore the low efficiency of reliability improvement and the excessive redundancy which deteriorates the foundations of our security enhancement.

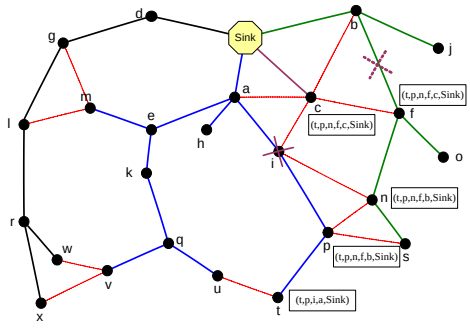


Fig. 7. Alternate path packet salvaging

A distinct feature of our multipath discovery protocol is that it finds multiple node-disjoint paths at each sensor node. If the WSN uses a reliable MAC protocol, such as IEEE 802.11 which acknowledges the successful transmission of each frame, each node knows whether the transmission is successful or not before it removes the frame from its transmission buffer. Therefore, taking advantage of the multiple paths available at each hop, we adopt an active per-hop packet salvaging strategy so that the reliability of each packet delivery (or each path) can be greatly improved. It works as follows. Each packet carries with it the source routing option. At an intermediate node *z*, if the transmission to the next hop is not successful, *z* actively salvages the packet by sending it to another randomly selected available route to the destination rather than dropping the packet. Only when all the next hops from node *z* to the BS fail should the packet be dropped. Fig. 7 shows an example that a packet originated at node *t* is salvaged twice at node *p* and *f* respectively and finally reaches the destination. One potential problem here is when a node salvages a packet with a new path but that new path consists of a node that the packet has already travelled. In this case, a routing loop would result. This problem can be easily solved by the source routing option. Notice that our multipath discovery protocol guarantees the loop freedom for all the paths selected. Each packet carries the source routing option when it is sent out. At an intermediate node, when the salvaging needs to be done, the node makes sure no loop would form by comparing the partial route the packet already travelled and the candidate path it would use to salvage the packet. Only when there is no common node would the candidate path be selected. Then the intermediate node modifies the source routing option carried in the packet by replacing the rest of the source route with the newly selected salvaging path. When a node reaches the BS, what it carries is the actual path it travelled through.

The per-hop alternate path packet salvaging is an effec-

tive and efficient way to improve reliability. Particularly, it improves reliability on a per packet/path basis without imposing redundant information. This property complements our necessity of H-SPREAD in improving reliability with little or no redundant information. Alternate path routing is not a new idea and it has been adopted in many single path routing protocols, such as DSR and AODV, whenever possible to improve unreliable packet delivery in MANETs [7]–[9]. However, the availability of alternate paths is the key and also the challenge to fully enjoy the performance enhancement. Obviously our N-to-1 multipath discovery protocol has paved the way to exploit this technique and has made it particularly suitable to apply this technique to improve the reliability of each path in our H-SPREAD data collection scheme.

V. Simulation Results

In this section, we evaluate the overall security and reliability performance of the proposed H-SPREAD scheme, namely, the combination of the concurrent multipath routing on the end-to-end data collection task and the alternate path routing on each packet delivery along the designated path.

We run the *N-to-1* multipath discovery protocol we proposed in section III. Then we consider the impact of node failure, link failure, as well as compromised node problem. We assume that node failure is persistent. Once a node fails, it cannot be used to forward packets. Link failure is intermittent and is independent of each packet transmission. When a link error occurs to a packet, no retransmission is performed for the same packet. Node compromise is persistent too. If a node is compromised, all the shares/packets relayed by that node are considered compromised.

Due to space limitation, we only report results in networks where the transmission range is 20m (refer to section III-D for network parameters). Each node which is at least 2 hops away from the Sink node initiates 100 messages. Each message is divided into $N = 10$ shares and spread onto M paths ($M = 1, \cdots, 7$). For $M = 1$, the shall allocation vector is $\underline{n} = [10]$, namely, all the 10 shares go through the primary path. For $M = 2$, $\underline{n} = [5\ 5]$; $M = 3$, $\underline{n} = [4\ 3\ 3]$; $M = 4$, $\underline{n} = [3\ 3\ 2\ 2]$; $M = 5$, $\underline{n} = [2\ 2\ 2\ 2\ 2]$; $M = 6$, $\underline{n} = [2\ 2\ 2\ 2\ 1\ 1]$; $M = 7$, $\underline{n} = [2\ 2\ 2\ 1\ 1\ 1\ 1]$.

The simulation results are averaged over 300 randomly generated networks. The 95% confidence intervals are shown in the figures.

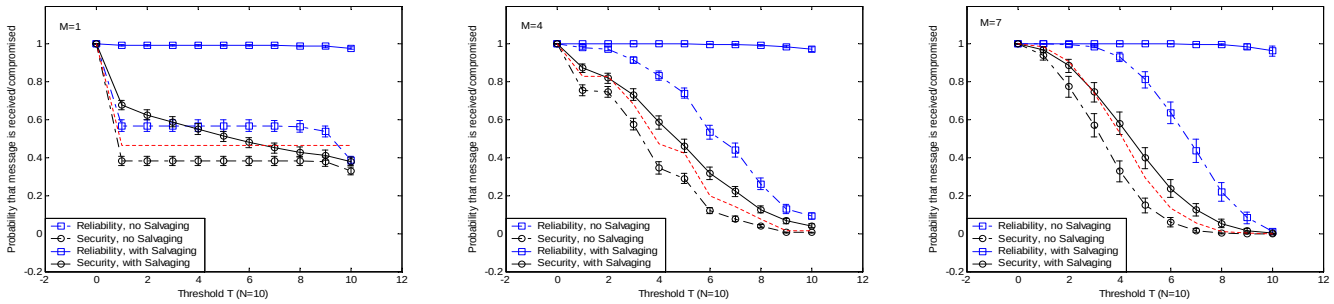
Fig. 8 shows the significant improvement in reliability when salvaging is used. The results for $M = 1, 4, 7$ are shown here. Figures for other M values show the same trend therefore are omitted. The X axis is the threshold T (with N set to 10 in all simulations) which can be interpreted as the level of redundancy. Here reliability is represented by the probability that message is successfully delivered which is calculated as the total number of messages received at the Sink node over the total number of messages initiated from all the sensor nodes. A message is received when at least T shares of the message reach the Sink. Similarly, security is represented by the probability that message is compromised, which is calculated as the total number of messages compromised over the total number of messages initiated by all the sensor nodes. A message is compromised when at least T shares are compromised by the compromised nodes collectively. Therefore, $T = 10$ means no redundancy and either the BS or adversaries must receive/intercept all the 10 shares to recover a message. It is observed that without salvaging, the packet loss rate is sensitive to the redundancy level and is unacceptably high even with excessive redundancy (small T values). However, our alternate path packet salvaging effectively maintains a very high (close to 100%) delivery ratio at all redundancy levels, even with zero redundancy. On the other hand, we observe that security is very sensitive to the redundancy - the less redundancy, the more secure the scheme is. The (red) dotted line is drawn as the reference. It represents security level achieved when all the nodes and links are reliable therefore no salvaging is performed[7]. As expected, salvaging weakens security a little bit because of possible overlapping of the paths. However, the impact is not significant compared with the significantly improved reliability. This is the most desirable property that enables our proposed scheme to improve both security and reliability at the same time.

Fig. 9 and Fig. 10 plot security and reliability performance as a function of number of paths used respectively with various network error and threat conditions. It is clear that our scheme is effective in reducing the probability that a message might be compromised. We observe that although the active packet salvaging breaks the independence of the paths, the probability that the message might be compromised decreases with the increase of the number of paths used to spread the information. Notice that the situations we studied are very challenging, with 10%, 20%, and 30% of nodes compromised. In fact, in less challenging situations, the improvement would be more significant (i.e., curves dropping more steeply). The results confirm the effectiveness of the proposed scheme - it is more resistant to the collusive attacks of compromised nodes. Correspondingly, the reliability performance shows that the proposed scheme is able to maintain pretty good message delivery ratio in the face of both link and node failures. Again, the situations we studied are very stressful, with 10%, 20%, and 30% of nodes failed. Therefore, we conclude that the proposed scheme with active alternate path salvaging is more robust to node failure problem too.

VI. Related Work

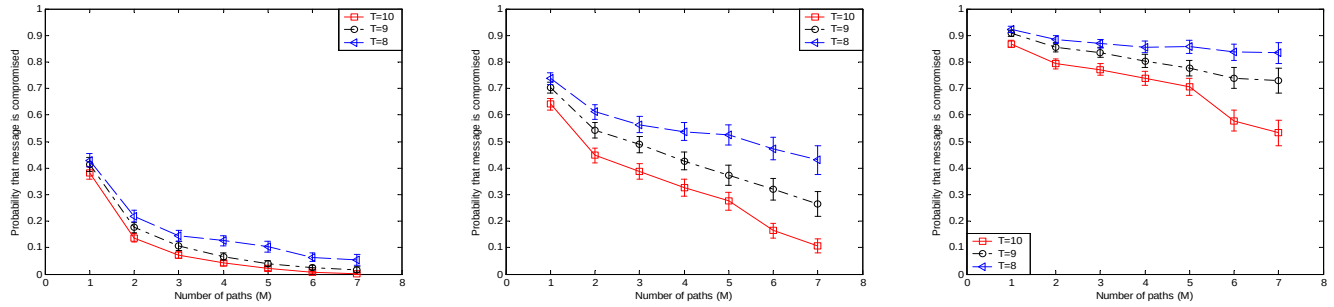
Efficient data delivery in WSNs is a challenging task. Direct diffusion [17], [18] and SPIN [19] are two exemplary data dissemination paradigms. As a data-centric approach, direct diffusion employs low rate flooding to establish gradients and uses gradual reinforcement of better paths to accommodate certain levels of network and sink dynamics. SPIN adopts meta-data negotiation to eliminate the redundant data transmission and is suitable for the scenarios where an individual sensor disseminates its observations to all sensors in a network. Some other approaches for data dissemination in WSNs include the flooding based Gossiping [20], probabilistic-based

[7]The lower compromise probability for the no salvaging case does not indicate more security. It is lower than the reference because of the loss of packet.



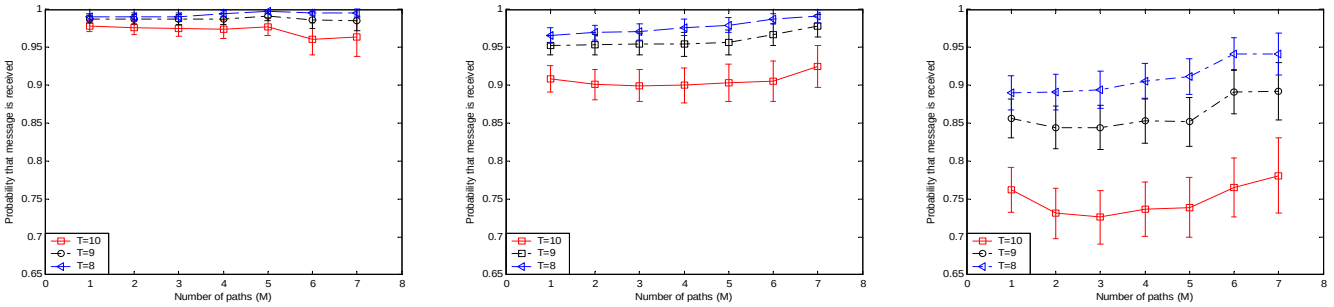
(a) M=1 (b) M=4 (c) M=7

Fig. 8. Security and reliability performance with or without packet salvaging (10% faulty nodes, 10% compromised nodes, link failure probability 1%)



(a) 10 compromised nodes (b) 20 compromised nodes (c) 30 compromised nodes

Fig. 9. Security performance



(a) 10 faulty nodes, 1% Link failure rate (b) 20 faulty nodes, 2% Link failure rate (c) 30 faulty nodes, 3% Link failure rate

Fig. 10. Reliability performance

flooding [21], [22], geometry-based flooding [23], cluster-based LEACH [24], hierarchical-based TTDD [25], etc.

Multipath routing has been a promising technique in MANETs in order to aggregate limited bandwidth, to smooth the burstiness of traffic, to alleviate network congestion, and to improve fault tolerance, and most importantly, to improve reliability. Several multipath routing protocols have been proposed to find multiple disjoint or partially disjoint paths between a single source and destination pair [5]–[11], [18], [23]. These multiple paths can be used in different ways. One way is to use them alternatively, namely, use the primary path first, when the primary one fails, switch to the secondary one, and so on. The other type of usage is to use the multiple paths simultaneously.

Our approach distinguishes from most previous work in that (a) Our *N-to-1* multipath discovery protocol is receiver-initiated (in contrast to the common source-initiated route discovery) and the protocol is efficient in that it finds multipath from every sensor node to the base station, which fits the special communication pattern of the WSN very well [26]–[28]; (b) We adopt a hybrid multipath approach for data delivery. We use concurrent multipath scheme to spread traffic onto multiple disjoint paths for end-to-end data delivery. Meanwhile, taking advantage of the multiple paths available at each node, the per-hop alternate path packet salvaging uses the multiple paths alternately and helps to improve the reliability of each packet delivery/path significantly; and (c) The overall scheme improves both security and reliability.

VII. Conclusions

Data collection is an important task in a WSN. Reliable and secure techniques are desired to perform the task efficiently. In this paper, we consider a WSN where the typical task is to disseminate data requests from a base station to all sensor nodes and to collect sensor readings from every sensor node

back to the base station. We first propose an efficient *N-to-1* multipath discovery protocol which initiates a route update periodically or on demand at the base station and at the end of each discovery process, finds every sensor node a set of node-disjoint paths back to the base station. Then based on the availability of the multiple paths at each node, we propose a hybrid multipath scheme for secure and reliable data collection task. The simulation results show that the proposed multipath discovery protocol is very efficient, with less than one message per path found. The proposed hybrid multipath data collection scheme is more resilient to node/link failures and a collusive attack of compromised nodes. It is effective in improving both reliability and security at the same time.

REFERENCES

[1] A. F. Akyildiz, W. Su, Y. Sankarasubramainiam, E. Cayirci, "A survey on sensor networks", *IEEE Communications Magazine*, August 2002

[2] A. Tsirigos, Z.J. Haas, "Multipath routing in the presence of frequent topological changes", *IEEE Communication Magazine*, Nov 2001

[3] W. Lou, W. Liu, Y. Fang, "SPREAD: Enhancing data confidentiality in mobile ad hoc networks, *IEEE INFOCOM 2004*, HongKong, China, March 2004

[4] S.K. Das, A. Mukherjee, et al, "An adaptive framework for QoS routing through multiple paths in ad hoc wireless newtorks", *J. Parallel Distributed Computing*, 63(2003)141-153

[5] S. De, C. Qiao, H. Wu, "Meshed multipath routing: an efficient strategy in sensor networks", *IEEE Wireless Communications and Networking Conference (WCNC'03)*, New Orleans, LA, Mar 2003

[6] M. K. Marina, S. R. Das, "On-demand multipath distance vector routing in ad hoc networks", *9th International Conference on Network Protocols*, Riverside, CA, November, 2001

[7] M.R. Pearlman, Z.J. Haas, P. Sholander, S. S. Tabrizi, "On the impact of alternate path routing for load balancing in mobile ad hoc networks", *The ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC'00)*, Boston, MA, August 2000.

[8] S.-J. Lee, M. Gerla, "AODV-BR: backup routing in ad hoc networks", *IEEE Wireless Communications and Networking Conference (WCNC'00)*, Sep 2000

[9] S.-J. Lee, M. Gerla, "Split multipath routing with maximally disjoint paths in ad hoc networks", *International Conference on Communications (ICC'01)*, Helsinki, Finland, June 2001.

[10] K. Zeng, K. Ren, W. Lou, "Geographic on-demand disjoint multipath routing in wireless ad hoc networks", *IEEE MILCOM*, Oct 2005

[11] Z. Ye, S. V. Krishnamurthy, S. K. Tripathi, "A framework for reliable routing in mobile ad hoc networks", *IEEE INFOCOM 2003*, Sanfrancisco CA, Mar 2003

[12] E. Ayanoglu, C-L. I, R.D. Gitlin, J.E Mazo, "Diversity coding for transparent self-healing and fault-tolerant communication networks", *IEEE Transactions on Communications*,41(11):1677-1686, Nov 1993

[13] T. Cormen, C. Leiserson, R. Rivest, *Introduction to algorithms*, MIT Press, 1990

[14] C. Karlof, D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures", *Elsevier's AdHoc Networks Journal, Special Issue on Sensor Network Applications and Protocols*, 1(2-3):293-315, September 2003

[15] J-S. Wu, R-J. Chen, "An algorithm for computing the reliability of weighted-k-out-of-n systems", *IEEE Transactions on Reliability*, 43(2):327-328, June 1994

[16] C.J. Colbourn, *The Combinatorics of Network Reliability*, Oxford University Press, 1987

[17] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, "Directed diffusion for wireless sensor networks", *IEEE/ACM Transactions on Networking*, 11(1):2-16, Feb 2003

[18] D. Ganesan, R. Govindan, S. Shenker, D. Estrin, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks", *Mobile Computing and Communication Review*, 5(4):10-24, Oct 2001

[19] J. Kulik, W. Heinzelman, H. Balakrishnan, " Negotiation-based protocols for disseminating information in wireless sensor networks", *Wireless Networks*, 8(2-3):169-185, March/May 2002

[20] Z. Hass, J. Halpern, and L. Li. Gossip-based ad hoc routing. *proc. of IEEE INFOCOM02*, New York, June 2002

[21] C. Barrett, S. Eidenbenz, L. Kroc. Parametric probabilistic sensor network routing. *Proc. 2nd International Workshop on Wireless Sensor Networks and Applications (WSNA 2003)*, San Diego, Sept. 2003

[22] Y. Sasson, D. Cavin, and A. Schiper. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'03)*, New Orleans, Louisiana, Mar. 2003

[23] W. Liu, Y. Zhang, W. Lou, Y. Fang, "Scalable and Robust Data Dissemination in Wireless Sensor Networks, *IEEE GLOBECOM 2004*, Dallas, TX, Dec 2004

[24] W. Heinzelman, A. Chandrakasan, and H. balakrishnan. Energy-effcient rotuing protocols for wireless microsensor networks. *Proc. 33rd hawaii Intenational Conference on System Sciences (HICSS'00)*, Hawaii, Jan. 2000

[25] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang. A Two-tier Data Dissemination Model for Large-scale Wireless Sensor Networks. *Proc. of the Ninth Anuual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom02)*, Atlanta, Georgia, Sept. 2002

[26] D. Sidhu, R. Nair, S. Abdallah, "Finding disjoint paths in networks", *Proc. of ACM SIGCOMM*, 1991

[27] W. Lou, Y. Fang, "A multipath routing approach for secure data delivery", *IEEE MILCOM*, McLean, VA, Oct 2001

[28] W. Lou, "An efficient N-to-1 multipath routing protocol in wireless sensor networks", 2nd IEEE International Confernce on Mobile Ad-hoc and Sensor System (MASS 2005), Washington D.C., Nov 2005

Wenjing Lou Wenjing Lou is an assistant professor in the Electrical and Computer Engineering department at Worcester Polytechnic Institute. She obtained her Ph.D degree in Electrical and Computer Engineering from University of Florida in 2003. She received the M.A.Sc degree from Nanyang Technological University, Singapore, in 1998, the M.E degree and the B.E degree in Computer Science and Engineering from Xi'an Jiaotong University, China, in 1996 and 1993 respectively. From December 1997 to July 1999, she worked as a Research Engineer in Network Technology Research Center, Nanyang Technological University. Her current research interests are in the areas of ad hoc and sensor networks, with emphases on network security and routing issues.

Younggoo Kwon Younggoo Kwon received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, Korea, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the Department of Electrical and Computer Engineering, University of Florida, Gainesville, in 2002. From 2002, he had been a Senior Member of the Research Staff at Samsung Electro-Mechanics Central R&D Center. Currently, he is an Assistant Professor in the Department of Electronic Engineering, Konkuk University, Seoul, Korea. He has authored/ coauthored technical papers and book chapters in the areas of wireless/mobile networks and RFID/Ubiquitous Sensor Networks. His current research interests include the area of RFID/USN systems with emphasis on the energy efficient network protocols, and the multimedia networking based on wireless LANs and wireless PANs.