Secure Data Collection in Wireless Sensor Networks Using Randomized Dispersive Routes

Tao Shu, Sisi Liu, and Marwan Krunz
Department of Electrical and Computer Engineering
University of Arizona

Abstract—Compromised-node and denial-of-service are two key attacks in wireless sensor networks (WSNs). In this paper, we study routing mechanisms that circumvent (bypass) black holes formed by these attacks. We argue that existing multi-path routing approaches are vulnerable to such attacks, mainly due to their deterministic nature. So once an adversary acquires the routing algorithm, it can compute the same routes known to the source, and hence endanger all information sent over these routes. In this paper, we develop mechanisms that generate randomized multi-path routes. Under our design, the routes taken by the "shares" of different packets change over time. So even if the routing algorithm becomes known to the adversary, the adversary still cannot pinpoint the routes traversed by each packet. Besides randomness, the routes generated by our mechanisms are also highly dispersive and energy-efficient, making them quite capable of bypassing black holes at low energy cost. Extensive simulations are conducted to verify the validity of our mechanisms.

I. Introduction

Of the various possible security threats that may be experienced by a wireless sensor network (WSN), in this paper we are specifically interested in combating two types of attacks: the compromised-node (CN) attack and the denial-of-service (DOS) attack [12]. The CN attack refers to the situation when an adversary physically compromises a subset of nodes to eavesdrop information, whereas in the DOS attack, the adversary interferes with the normal operation of the WSN by actively disrupting, changing, or even destroying the functionality of a subset of nodes in the system. These two attacks are similar in the sense that they both generate *black holes*: areas within which the adversary can either passively intercept or actively block information delivery. Due to the unattended nature of WSNs, adversaries can easily produce such black holes [1]. Severe CN and DOS attacks can disrupt normal data delivery between sensor nodes and the sink, or even partition the topology. A conventional cryptography-based security method cannot alone provide satisfactory solutions to these problems. This is because, by definition, once a node is compromised, the adversary can always acquire the encryption/decryption keys of that node, and thus can intercept any information passed through it. At the same time, an adversary can always perform certain form of DOS attack (e.g., jamming) even if it does not have any knowledge of the crypto-system used in the WSN.

One remedial solution to these attacks is to exploit the network's routing functionality. Specifically, if the locations of the black holes formed by the compromised (or jammed) nodes are known a priori, then information can be delivered over paths that circumvent (bypass) these holes, whenever possible. In practice, due to the difficulty of acquiring such location information, the above idea is implemented in a probabilistic manner, typically through a two-step process: secret sharing and multi-path routing. First, an information (e.g., a packet) is broken into M *shares* (i.e., components of a packet that carry partial information) using a (T, M)-threshold secret-sharing mechanism such as the Shamir's algorithm [10]. The original information can be recovered from a combination of at least T shares, but no information can be guessed from less than T shares. Then, multiple routes from the source to the destination are computed according to some multi-path routing algorithm (e.g., [7], [6], [4], [13]). These routes are node-disjoint or maximal node-disjoint subject to certain constraints (e.g., min-hop routes). The M shares are then distributed across these routes and delivered to the destination, following different paths. As long as at least $M - T + 1$ (or T) shares bypass the compromised (or jammed) nodes, the adversary cannot acquire (or deny the delivery of) the original information packet.

We argue that three security problems exist in the above counter-attack approach. First, this approach is no longer valid if the adversary can *selectively* compromise or jam nodes. This is because the route computation in the above multi-path routing algorithms is deterministic in the sense that for a fixed topology, a fixed set of routes are always computed by the routing algorithm for given source and destination. Therefore, even if the shares can be distributed over different routes, overall they are always delivered over the same set of routes that are computable by the algorithm. As a result, once the routing algorithm becomes open to the adversary (this can be done, e.g., through a memory interrogation of the compromised nodes), the adversary can by itself compute the set of routes for any given source and destination. Then the adversary can pinpoint to one particular node in each route and compromise (or jam) these nodes. Such an attack can intercept all shares of the information, rendering the above counter-attack approaches ineffective. Second, as pointed out in [13], actually very few node-disjoint routes can be found when node density is moderate and source and destination nodes are several hops apart. For example, for a node degree of 8, on average only two node-disjoint routes can be found between a source and a destination that are at least 7 hops apart. There is also a 30% possibility that no node-disjoint paths can be found between the source and the destination [13]. The lack of enough routes significantly undermines the security performance of this multi-path approach. Last, even worse, because the set of routes is computed under certain constraints, the routes may not be spatially dispersive enough to circumvent a moderate-sized black hole.

In this paper, we propose a randomized multi-path routing algorithm that can overcome the above problems. Instead of selecting paths from a pre-computed set of routes, this algorithm computes multiple paths in a randomized way each time an information packet needs to be sent, such that the set of routes taken by various shares of different packets keep

changing over time. As a result, a large number of routes can be potentially generated for each source and destination. To intercept different packets, the adversary has to compromise or jam all possible routes from the source to the destination, which is practically infeasible.

The key contributions of this work are as follows. We explore the potential of random dispersion for information delivery in WSNs. Depending on the type of information available to a sensor, we develop four distributed schemes for propagating information "shares": purely random propagation (PRP), directed random propagation (DRP), non-repetitive random propagation (NRRP), and multicast tree-assisted random propagation (MTRP). PRP utilizes only one-hop neighborhood information and provides baseline performance. DRP utilizes two-hop neighborhood information to improve the propagation efficiency, leading to a smaller packet interception probability. NRRP achieves the same effect, but in a different way: it records all traversed nodes to avoid traversing them again in the future. MTRP tries to propagate shares in the direction of the sink, making the entire delivery process more energy efficient. We conduct extensive simulations to study the performance of the proposed schemes under realistic settings. When their parameters are appropriately set, all four randomized schemes are shown to provide comparable or even better security and energy performance than their deterministic counterparts. At the same time, they do not suffer from pin-pointed node attacks of deterministic multi-path routing.

The remainder of this paper is organized as follows. In Section 2, we elaborate on the design of the randomized multi-path routing mechanism. Section 3 evaluates the performance of all four schemes using simulations. We provide an overview of the related work in Section 4 and conclude our work in Section 5.

II. RANDOMIZED MULTI-PATH DELIVERY

A. Overview

As illustrated in Figure 1, we consider a 3-phase approach for secure information delivery in a WSN: secret sharing of information, randomized propagation of each information share, and normal routing (e.g., min-hop routing) toward the sink. More specifically, when a sensor node wants to send a packet to the sink, it first breaks the packet into M shares according to a (T, M)-threshold secret sharing algorithm, e.g., the Shamir's algorithm [10]. Each share is then transmitted to some randomly picked neighbor. That neighbor will continue to relay the share it has received to other randomly picked neighbors, and so on. In each information share, there is a TTL field, whose initial value is set by the source node to control the total number of randomized relays. After each relay, the TTL field is reduced by 1. When the TTL count reaches 0, the final node receiving this share stops the random propagation phase and begins to route this share towards the sink using normal single-path routing. Once the sink collects at least T shares, it can inversely compute the original information. No information can be recovered from less than T shares.

Because routes are randomly generated, there is no guarantee that different routes are still node-disjoint. However, the algorithm should ensure that the randomly generated routes are as dispersive as possible, i.e., different routes are geographically separated as far as possible such that they have high likelihood of not simultaneously passing through a
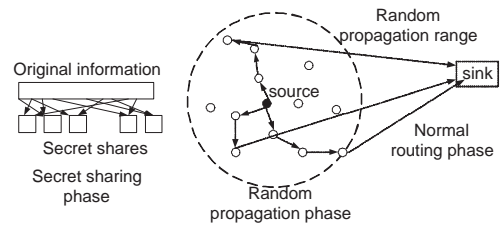

Fig. 1. Randomized dispersive routing in a WSN.

black hole. Considering the stringent requirement on energy consumptions in WSNs, the major challenge in our design is to generate highly dispersive random routes at low energy cost. As explained later, such a challenge is not trivial. A naive algorithm of generating random routes, such as Wanderer scheme [2] (a pure random-walk algorithm), only leads to long paths (containing many hops, and therefore, consuming much energy) without achieving good dispersiveness. Due to security considerations, we also require that the route computation be implemented in a distributed way, such that the final route represents the aggregate decision of all the nodes participating in route selection. As a result, a small number of colluding/compromised nodes cannot dominate the selection result. In addition, for efficiency purposes, we also require that the randomized route selection algorithm only incurs a small amount of communication overhead.

Needless to say, the random propagation phase is the key component that dictates the security and energy performance of the entire mechanism. We further elaborate on the design of this component in the following subsections.

B. Random propagation of Information Shares

To diversify routes, an ideal random propagation algorithm propagates information shares as dispersively as possible. Typically, this means propagating the share farther from its source. At the same time, it is highly desirable to have an energy-efficient propagation, which calls for limiting the number of randomly propagated hops. The challenge here lies in the random and distributed nature of the propagation: a share may be sent one-hop farther from its source in a given step, but may be sent back closer to the source in the next step, wasting both steps from the security's point of view. To tackle this issue, some control needs to be imposed on the random propagation process to ensure that in each step the share is more likely to be forwarded outwards from the source. We develop four distributed random propagation mechanisms, which approach this goal in various degrees.

1) Purely Random Propagation (Baseline Scheme): In PRP, information shares are propagated based on one-hop neighborhood information. More specifically, a sensor node maintains a neighbor list, which contains the ids of all the nodes that are within its receiving range. When a source node wants to send information shares to the sink, it includes a TTL of initial value N in each share. It then randomly selects a neighbor for each share, and unicasts the share to that neighbor. After receiving the share, the neighbor first decrements the TTL. If the new TTL is greater than 0, the neighbor randomly picks a node from its neighbor list (this node cannot be the source node) and relays the share to it, and so on. When the TTL reaches 0, the final node receiving this share stops the random propagation of this share, and starts routing this share towards the sink using normal min-hop routing. The WANDERER [2] scheme is a special case of PRP with $N = \infty$.

The main drawback of PRP is that its propagation efficiency can be low, because a share may be propagated back and forth multiple times between neighboring hops. As shown in the simulations in the next section, increasing the value of TTL does not fully address this problem. This is because the random propagation process reaches steady state under a large TTL, and its distribution will no longer change even if the TTL becomes larger.

2) Non-repetitive Random Propagation: NRRP is based on PRP, but it improves the propagation efficiency by recording all the nodes that the propagation has traversed so far. More specifically, NRRP adds a "node-in-route" (NIR) field to the header of each share. Initially, this field is empty. Starting from the source node, whenever a node propagates the share to the next hop, the id of the up-stream node is appended to the share's NIR field. Nodes included in NIR are excluded from the random pick of the next hop of propagation. This non-repetitive propagation guarantees that the share will be relayed to a different node in each step of random propagation, leading to better propagation efficiency.

3) Directed Random Propagation: DRP improves the propagation efficiency by using two-hop neighborhood information. More specifically, DRP adds a "last-hop neighbor list" (LHNL) field to the header of each share. Before a share is propagated to the next node, the relaying node first replaces the old content in the LHNL field of the share by its neighbor list. When the next node receives the share, it compares the LHNL field against its own neighbor list, and randomly picks one node from its neighbors that are not in the LHNL. It then decrements the TTL value, updates the LHNL field, and relays the share to the next hop, and so on. Whenever the LHNL fully overlaps with or contains the relaying node's neighbor list, a random neighbor is drawn, just as in the case of the PRP scheme. According to this propagation method, DRP reduces the chance of propagating a share back and forth by eliminating this type of propagation within any two immediate consecutive steps. Compared with PRP, DRP attempts to push a share outward away from the source, and thus leads to better propagation efficiency for a given TTL value.

4) Multicast Tree-assisted Random Propagation: The MTRP scheme aims at actively improving the energy efficiency of random propagation while preserving the dispersiveness of DRP. The basic idea comes from the following observation of Figure 1: Among the 3 different routes taken by the shares, the route on the bottom right is the most energy efficient because it has the shortest end-to-end path. So, in order to improve energy efficiency, the shares should be best propagated in the direction of the sink. In other words, their propagation should be restricted to the right half of the circle in Figure 1.

Conventionally, directional routing requires location information of both the source and the destination nodes, and sometimes the intermediate nodes. Examples of this type of location-based routing are GPSR (Greedy Perimeter Stateless Routing) and LAR (Location-Aided Routing). Location information mainly relies on GPS in each node, or on some distributed localization algorithms. The high cost and the low accuracy of localization are the main drawbacks of these two methods, respectively.

MTRP involves directionality in its propagation process without needing location information. More specifically, after the deployment of the WSN, MTRP requires that the sink constructs a multicast tree from itself to every node in the network. Such a tree-construction operation is not unusual in existing protocols, and is typically conducted via flooding a "hello" message from the sink to every node. Once this multicast tree is constructed, a node knows its distance (in number of hops) to the sink and the id of its parent node. We assume that each entry in the neighbor list maintained by a node has a field recording the number of hops to the sink from the corresponding neighbor. Under MTRP, the header of each share contains two additional fields: $\max_{hop}$ and $\min_{hop}$. The values of these two parameters are set by the source to $\max_{hop} = n_s + \alpha_1$ and $\min_{hop} = n_s - \alpha_2$, where n_s is the hop count from the source to the sink, and α_1 and α_2 are non-negative integers with $\alpha_1 \leq \alpha_2$. The parameter α_1 controls the limit that a share can be propagated away from the sink, i.e., to the left half of the circle in Figure 1. The parameter α_2 controls the propagation area toward the sink, i.e., the right half of the circle. A small α_2 makes the propagation of a share be dispersed away from the center line connecting the source and the link and forces them to take the side path, leading to better dispersion.

Before a node begins to pick the next relaying node from its neighbor list, it first filters out neighbors that are in the LHNL, just as in the case of DRP. Next, it filters out nodes that have a hop count to the sink greater than $\max_{hop}$ or smaller than $\min_{hop}$. The next relaying node will be randomly drawn from the remaining neighbors. In case the set of remaining nodes after the first step is empty, the second step will be directly applied to the entire set of neighbors.

III. SIMULATION STUDIES

A. Simulation Setup

In this section, we use simulation to evaluate the performance of PRP, NRRP, DRP, and MTRP. The performance metric of interest is the packet interception probability for a source, defined as the ratio of the number of intercepted packets to the total number of packets sent from that source. In addition, we also study the average number of hops of the end-to-end route generated by various schemes. The hop-count is indirectly related to the energy efficiency of the routes generated by a given scheme. To better understand the capability of these randomized multi-path routing algorithms in bypassing black holes, we also compare their performance against a deterministic counterpart, H-SPREAD [6], which generates node-disjoint multi-path routes to combat CN attack in WSNs.

We consider a a 200m $\times$ 200m field that is uniformly covered by sensors. The center of this square is the origin point. All coordinates are in the unit of meters. The black hole formed by compromised nodes is represented by its circumcircle, i.e., the smallest circle that encompasses the shape of the black hole. We denote the radius of the black hole by R_e. The sink and the center of the black hole are placed at (100, 0) and (50, 0), respectively. The transmission range of each sensor is $R_h = 10$ m. During network operation, any end-to-end path that goes through this circle is considered as vulnerable to an eavesdropper, i.e., the information shares delivered over this path are all intercepted by the adversary. We assume that a packet is intercepted if all its shares are intercepted by the adversary. For the MTRP scheme, we set $\alpha_1 = 0$ and $\alpha_2 = 5$. In all simulations, after the random propagation phase, each secret share is delivered to the sink using min-hop routing.

Each simulation result is averaged over 50 randomly generated topologies. For each topology, 1000 information packets are sent from the source node to the sink.

B. Simulation Results

We first fix the location of the source node at $(-50, 0)$. In Figures 2 and 3, we plot the packet interception probability as a function of the TTL value (N) and the number of shares (M) that each packet is broken into, respectively. These figures show that increasing N and M helps reduce the packet interception probability for all proposed schemes. However, for a sufficiently large N, (e.g., $N = 20$ in Figure 2), the interception probability does not change much with a further increase in N. This is because the random propagation process has reached steady state. It can also be observed that, in all cases, the packet interception probabilities under the DRP, NRRP, and MTRP schemes are much smaller than that of the baseline PRP scheme, because their random propagations are more efficient. In addition, when N and M are large, all four randomized algorithms achieve smaller packet interception probabilities than the deterministic H-SPREAD scheme. In many cases, the gap is more than one order of magnitude. The poor performance of H-SPREAD is due to the small number of node-disjoint routes that can be found by the algorithm when the source is far away from the sink (15 hops apart in our simulation), and the fact that these routes may not be dispersive enough. Increasing M does not change the number of routes the algorithm can find, so it does not help in reducing the interception probability for H-SPREAD.

We plot the packet interception probability as a function of the size of the black hole in Figure 4. It is clear that the interception probability increases with R_e. This trend is in line with our intuition.

In Figure 5 we study the impact of node connectivity. The number of nodes is changed from 1000 to 3000, corresponding to changing the average node connectivity degree from 8 to 24. It can be observed that, in general, the packet interception probabilities of the four proposed schemes do not change significantly with node connectivity. Such insensitivity to node connectivity/density is because the packet interception probability is mainly decided by how dispersive the shares can be geographically after random propagation. As long as nodes are uniformly distributed, a change in node density does not impact the geographic distribution of the shares after random propagation. In contrast, the packet interception probability of H-SPREAD decreases significantly with the increase in node density, because more node-disjoint routes can now be found.

In Figure 6, we slide the x-coordinate of the source node along the line $y = 0$ to evaluate the packet interception probabilities at different source locations in the network. A segmented trend can be observed: When the source is far away from the black hole ($-100 \leq x \leq 0$), the closer the source is to the black hole, the smaller the packet interception probability will be. This is because, when the source is far away from the black hole, shares are mainly intercepted during the normal routing phase. Note that during the normal routing phase, all paths start to converge geographically to the sink (see Figure 1). As a result, the closer the source is to the black hole, the less convergent the paths will be at the black hole, so the lower interception probability. When $x = -100$ (this is at the boundary), the gap between the proposed schemes are small, because all shares can only be propagated to the

right, making the random propagation process of PRP, DRP, and NRRP similar to that of MTRP. However, when the source is close to the black hole, i.e., $x \geq 0$, the trend in the interception probability is reversed. This is because more and more shares are intercepted during the propagation phase. When $x = 50$, which corresponds to the scenario where the source is placed right at the center of the black hole, the interception probabilities reach their maximum value. After that, they decrease quickly as the source gets farther away from the black hole. In all segments, the packet interception probabilities of the DRP, NRRP, and MTRP schemes are smaller than that of H-SPREAD.

We evaluate the average number of hops of the end-to-end route as a function of the TTL value in Figure 7. It can be observed that the hop-count under PRP, DRP, and NRRP increases linearly with N, while the hop-count under MTRP only increases slowly with N. The TTL value does not play a role in the H-SPREAD scheme. Under large N, e.g., when $N = 25$, the randomized algorithm achieves better security performance than H-SPREAD. However, the hop-count of H-SPREAD is about $1/3$ of that of PRP, DRP, and NRRP, and about $1/2$ of that of MTRP scheme. The relatively large hop-count in the randomized algorithms is the cost for stronger capability of bypassing black holes.

IV. RELATED WORK

Recently, several works have taken security metrics into account when constructing (deterministic) multi-path routes. Specifically, the SPREAD algorithm in [7] attempts to find multiple most-secure and node-disjoint paths. The security of a path is defined as the likelihood of node compromise along that path, and is labeled as the weight in path selection. A modified Dijkstra algorithm is used to iteratively find the top-K most secure node-disjoint paths. The H-SPREAD algorithm [6] improves upon SPREAD by simultaneously accounting for both security and reliability requirements. The work in [4] presents distributed Bound-Control and Lex-Control algorithms, which compute multiple paths, respectively, in such a way that the performance degradation (e.g., throughput loss) is minimized when a single-link attack or a multi-link attack happens, respectively. Other examples of secure deterministic multi-path routing algorithms include SRP [9], SecMR [8], Burmester's approach [3], and AODV-MAP [11].

Existing randomized multi-path routing algorithms in WSNs have not been designed with security considerations in mind, largely due to their low energy efficiency. To the best of our knowledge, the work presented in this paper fills a void in the area of secure randomized multi-path routing. Specifically, flooding is the most common randomized multi-path routing mechanism. As a result, every node in the network receives the packet and retransmits it once. To reduce unnecessary retransmissions and improve energy efficiency, the Gossiping [5] algorithm was proposed as a form of controlled flooding, whereby a node retransmits packets according to a pre-assigned probability. It is well known that the Gossiping algorithm has a percolation behavior, in that for a given retransmission probability, either very few nodes receive the packet, or almost all nodes receive it. Parametric Gossiping was proposed in [2] to overcome the percolation behavior by relating a node's retransmission probability to its hop count from either the destination or the source. A special form of Gossiping is the Wanderer algorithm [2], whereby a node retransmits
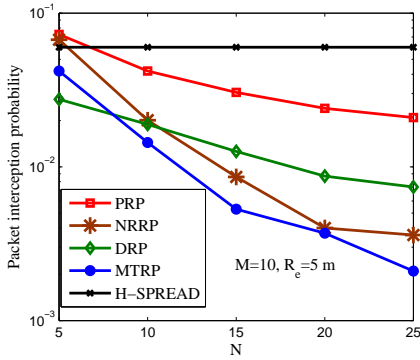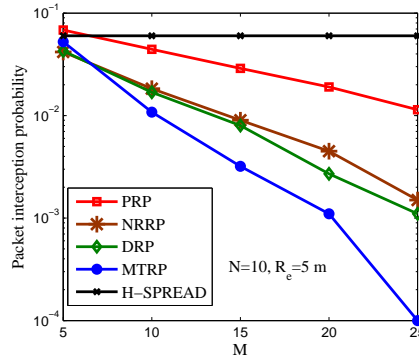
Fig. 2. Packet interception prob. vs. N.

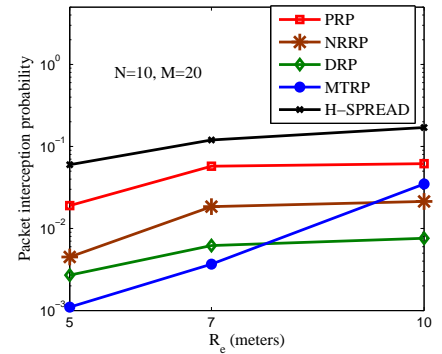

Fig. 3. Packet interception prob. vs. M.

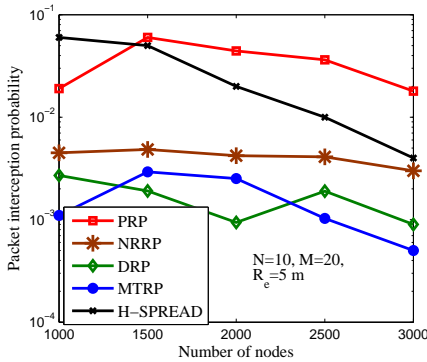Fig. 4. Packet interception prob. vs. R_e.

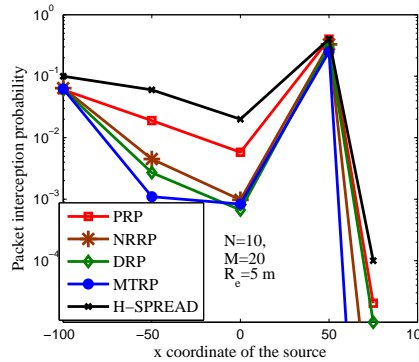Fig. 5.   Packet interception prob. vs. number of nodes.

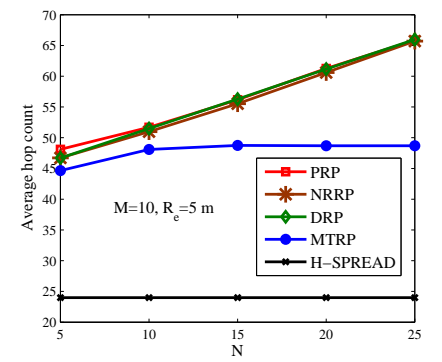Fig. 6.   Packet interception prob. at different source locations.

Fig. 7.   Hop count vs. N.

the packet to one randomly picked neighbor. When used to counter compromised-node attacks, flooding, Gossiping, and parametric Gossiping actually help the adversary intercept the packet, because multiple copies of a secret share are dispersed to many nodes. The Wanderer algorithm has poor energy performance, because it results in long paths. In contrast, the NRRP, DRP, and MTRP schemes proposed in this paper are specifically tailored to security considerations in energy-constrained WSNs. They provide highly dispersive random routes at low energy cost without generating extra copies of secrete shares.

V. CONCLUSIONS

Our simulation results have shown the effectiveness of randomized dispersive routing in combating CN and DOS attacks. By appropriately setting the secret sharing and propagation parameters, the packet interception probability can easily be reduced by the proposed algorithms to as low as 10^{-3}, which is at least one order of magnitude smaller than approaches that use deterministic node-disjoint multi-path routing. At the same time, we have also verified that this improved security performance comes at a reasonable cost of energy.

ACKNOWLEDGEMENTS

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114, Aug. 2002.

[2] C. L. Barrett, S. J. Eidenbenz, L. Kroc, M. Marathe, and J. P. Smith. Parametric probabilistic sensor network routing. In *Proceedings of the ACM International Conference on Wireless Sensor Networks and Applications (WSNA)*, pages 122–131, 2003.

[3] M. Burmester and T. V. Le. Secure multipath communication in mobile ad hoc networks. In *Proceedings of the International Conference on Information Technology: Coding and Computing*, pages 405–409, 2004.

[4] P. C. Lee, V. Misra, and D. Rubenstein. Distributed algorithms for secure multipath routing in attack-resistant networks. *IEEE/ACM Transactions on Networking*, 15(6):1490–1501, Dec. 2007.

[5] X. Y. Li, K. Moaveninejad, and O. Frieder. Regional gossip routing wireless ad hoc networks. *ACM Journal of Mobile Networks and Applications*, 10(1-2):61–77, Feb. 2005.

[6] W. Lou and Y. Kwon. H-spread: a hybrid multipath scheme for secure and reliable data collection in wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 55(4):1320–1330, July 2006.

[7] W. Lou, W. Liu, and Y. Fang. Spread: enhancing data confidentiality in mobile ad hoc networks. In *Proceedings of the IEEE INFOCOM Conference*, volume 4, pages 2404–2413, Mar. 2004.

[8] R. Mavropodi, P. Kotzanikolaou, and C. Douligeris. Secmr- a secure multipath routing protocol for ad hoc networks. *Elsevier Journal of Ad Hoc Networks*, 5(1):87–99, Jan. 2007.

[9] P. Papadimitratos and Z. J. Haas. Secure routing for mobile ad hoc networks. In *Proceedings of SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS)*, 2002.

[10] D. R. Stinson. *Cryptography, Theory and Practice*. CRC Press, 2006.

[11] B. Vaidya, J. Y. Pyun, J. A. Park, and S. J. Han. Secure multipath routing scheme for mobile ad hoc network. In *Proceedings of IEEE International Symposium on Dependable, Autonomic and Secure Computing*, pages 163–171, 2007.

[12] A. D. Wood and J. A. Stankovic. Denial of service in sensor networks. *IEEE Computer Magazine*, 35(10):54–62, Oct. 2002.

[13] Z. Ye, V. Krishnamurthy, and S. K. Tripathi. A framework for reliable routing in mobile ad hoc networks. In *Proceedings of the IEEE INFOCOM Conference*, volume 1, pages 270–280, Mar. 2003.