Secure Data Collection in Wireless Sensor Networks Using Randomized Dispersive Routes

M.Madhusudhan[1], N.SubbaReddy[2], G Kiran Kumar[3]
[1] *M.Tech, CSE Dept MLRIT, Hyderabad*
[2] *M. Tech(CSE),Asst. Professor, MLRIT, Hyderabad*
[3] *M. Tech(CSE),*Associate Professor & HOD-CSE, MLR Institute of technology

***Abstract*—Compromised node and denial of service are two key attacks in wireless sensor networks (WSNs). In this paper, we study data delivery mechanisms that can with high probability circumvent black holes formed by these attacks. We argue that classic multipath routing approaches are vulnerable to such attacks, mainly due to their deterministic nature. So once the adversary acquires the routing algorithm, it can compute the same routes known to the source, hence, making all information sent over these routes vulnerable to its attacks. In this paper, we develop mechanisms that generate randomized multipath routes. Under our designs, 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 generated routes are also highly dispersive and energy efficient, making them quite capable of circumventing black holes. We analytically investigate the security and energy performance of the proposed schemes. We also formulate an optimization problem to minimize the end-to-end energy consumption under given security constraints. Extensive simulations are conducted to verify the validity of our mechanisms.**

Index Terms—Randomized multipath routing, wireless sensor network, secure data delivery.

1 INTRODUCTION

1.1 Motivations

the various possible security threats encountered in a wireless sensor network (WSN), in this paper, we are specifically interested in combating two types of attacks: compromised node (CN) and denial of service (DOS) [22]. In the CN attack, 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 network by actively disrupting, changing, or even paralyzing the functionality of a subset of nodes. 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. Likewise, an adversary can always perform DOS attacks (e.g., jamming) even if it have any knowledge of the underlying cryptosystem.

One remedial solution to these attacks is to exploit the network's routing functionality. Specifically, if the locations of the black holes are known a priori, then data 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. First, the packet is broken into M shares (i.e., components of a packet that carry partial information) using a ðT; MÞ-threshold secret sharing mechanism such as the Shamir's algorithm [20]. 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. Second, multiple routes from the source to the destination are computed according to some multipath routing algorithm (e.g., [5], [8], [13], [24]). These routes are node-disjoint or maximally node-disjoint subject to certain constraints (e.g., min-hop routes). The M shares are then distributed over these routes and delivered to the destination. As long as at least M _ T þ 1 (or T) shares bypass the compromised (or jammed) nodes, the adversary cannot acquire (or deny the delivery of) the original 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 multipath routing algorithms is deterministic in the sense that for a given topology and given source and destination nodes, the same set of routes are always computed by the routing algorithm. As a result, once the routing algorithm becomes known to the adversary (this can be done, e.g., through memory interrogation of the compromised node), the adversary can 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 [24], actually very few node-disjoint routes can be found when the node density is moderate and the 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 30 percent probability that no node-disjoint paths can be found between the source and the destination [24]. The lack of enough routes significantly undermines the security performance of this multipath approach. Last, because the set of routes is computed under certain constraints, the routes may not be spatially dispersive enough to circum-vent a moderate-size black hole.

In this paper, we propose a randomized multipath routing algorithm that can overcome the above problems. In this algorithm, multiple paths are computed 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 not possible.

Because routes are now randomly generated, they may no longer be node-disjoint. However, the algorithm ensures that the randomly generated routes are as dispersive as possible, i.e., the routes are geographically separated as far as possible such that they have high likelihood of not simultaneously passing through a black hole. Considering the stringent constraint on energy consumption in WSNs, the main 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 lots of 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 the 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 commu-nication overhead.

1.2 Contributions and Organization

The key contributions of this work are as follows:

1. 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 informa-tion “shares”: purely random propagation (PRP), directed random propagation (DRP), nonrepetitive random propagation (NRRP), and multicast tree-assisted random propagation (MTRP). PRP utilizes only one-hop neighborhood information and pro-vides baseline performance. DRP utilizes two-hop neighborhood information to improve the propaga-tion efficiency, leading to a smaller packet intercep-tion probability. The NRRP scheme achieves a similar 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 delivery process more energy efficient.

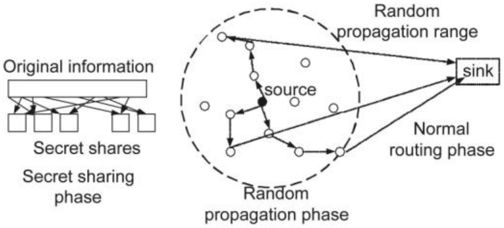


Fig. 1. Randomized dispersive routing in a WSN.

2. We theoretically evaluate the goodness of these dispersive routes in terms of avoiding black holes. We conduct asymptotic analysis (i.e., assuming an infinite number of nodes) for the worst-case packet interception probability and energy efficiency under the baseline PRP scheme. Our results can be interpreted as the performance limit of PRP, and a lower-bound on the performance of the more advanced DRP, NRRP, and MTRP schemes. Our analysis helps us better to understand how security is achieved under dispersive routing. Based on this analysis, we investigate the trade-off between the random propagation parameter and the secret sharing parameter. We further optimize these para-meters to minimize the end-to-end energy consump-tion under a given security constraint.

3. We conduct extensive simulations to study the performance of the proposed schemes under more realistic settings. Our simulation results are used to verify the effectiveness of our design. When the parameters are appropriately set, all four rando-mized schemes are shown to provide better security performance at a reasonable energy cost than their deterministic counterparts. At the same time, they do not suffer from the type of attacks faced by deterministic multipath routing.

The remainder of this paper is organized as follows: In Section 2, we elaborate on the design of the randomized multipath routing mechanism. In Section 3, we analyze the performance of the baseline PRP scheme. Section 4 evaluates the performance of all four schemes using simulations. We overview related work in Section 5 and conclude the paper in Section 6.

2 RANDOMIZED MULTIPATH DELIVERY

2.1 Overview

As illustrated in Fig. 1, we consider a three-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 ðT; MÞ-threshold secret sharing algorithm, e.g., Shamir's algorithm

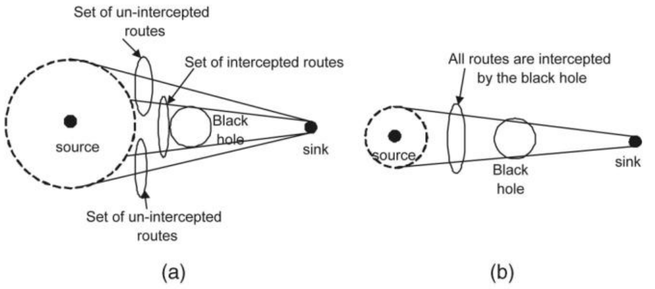

Set of un-intercepted routes
Set of intercepted routes
Black hole
source
sink
Set of un-intercepted routes
(a)
All routes are intercepted by the black hole
source
Black hole
sink
(b)

Fig. 2. Implication of route dispersiveness on bypassing the black hole. (a) Routes of higher dispersiveness. (b) Routes of lower dispersiveness.

other randomly selected neighbors, and so on. In each share, there is a TTL field, whose initial value is set by the source node to control the total number of random relays. After each relay, the TTL field is reduced by 1. When the TTL value reaches 0, the last node to receive this share begins to route it toward the sink using min-hop routing. Once the sink collects at least T shares, it can reconstruct the original packet. No information can be recovered from less than T shares.

The effect of route dispersiveness on bypassing black holes is illustrated in Fig. 2, where the dotted circles represent the ranges the secret shares can be propagated to in the random propagation phase. A larger dotted circle implies that the resulting routes are geographically more dispersive. Comparing the two cases in Fig. 2, it is clear that the routes of higher dispersiveness are more capable of avoiding the black hole. Clearly, the random propagation phase is the key component that dictates the security and energy performance of the entire mechanism.

2.2 Random Propagation of Information Shares

To diversify routes, an ideal random propagation algorithm would propagate shares as dispersively as possible. Typically, this means propagating the shares farther from their 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 propaga-tion: 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 a security standpoint. To tackle this issue, some control needs to be imposed on the random propagation process.

2.2.1 Purely Random Propagation (Baseline Scheme)

In PRP, shares are propagated based on one-hop neighbor-hood information. More specifically, a sensor node maintains a neighbor list, which contains the ids of all nodes within its transmission range. When a source node wants to send shares to the sink, it includes a TTL of initial value N in each share. It then

[20]. Each share is then transmitted to some randomly selected neighbor. That neighbor will continue to relay the share it has received to

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 it toward the sink using normal min-hop routing. The WANDERER scheme [2] is a special case of PRP with N ¼ 1.

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 analysis and simulations in subsequent sections, increasing the TTL value 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.

2.2.2 Nonrepetitive Random Propagation

NRRP is based on PRP, but it improves the propagation efficiency by recording the nodes traversed so far. Specifi-cally, 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 upstream node is appended to the NIR field. Nodes included in NIR are excluded from the random pick at the next hop. This nonrepetitive propagation guarantees that the share will be relayed to a different node in each step of random propagation, leading to better propagation efficiency.

2.2.3 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 updates the LHNL field with 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 selected, just as in the case of the PRP scheme. According to this propagation method, DRP reduces the chance of propagat-ing a share back and forth by eliminating this type of propagation within any two 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.

2.2.4 Multicast Tree-Assisted Random Propagation

MTRP 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 Fig. 1: Among the three different routes taken by shares, the route on the bottom right is the most energy efficient because it is the shortest end-to-end path. So, in order to improve energy efficiency, shares should be best propa-gated in the direction of the sink. In other words, their propagation should be restricted to the right half of the circle in Fig. 1.

Conventionally, directional routing requires location information of both the source and the destination nodes, and sometimes of intermediate nodes. Examples of location-based routing are the Greedy Perimeter Stateless Routing (GPSR) and Location-Aided Routing (LAR). 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, it requires the sink to construct a multicast tree from itself to every node in the network. Such tree construction is not unusual in existing protocols, and is typically conducted by flooding a "hello" message from the sink to every node. Once the multicast tree is constructed, a node knows its distance (in hops) to the sink and the id of its parent node on the tree. We assume that each entry in the neighbor list maintained by a node has a field that records 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 parameters are set by the source to max_{hop} ¼ n_S þ $\_{1}$ and min_{hop} ¼ n_S _ $\_{2}$, where n_S is the hop count from the source to the sink, and $\_{1}$ and $\_{2}$ are nonnegative integers with $\_{1}$ _ $\_{2}$. The para-meter $\_{1}$ controls the scope of propagation away from the sink, i.e., to the left half of the circle in Fig. 1. The parameter $\_{2}$ controls the propagation area toward the sink, i.e., the right half of the circle. A small $\_{2}$ pushes the propagation of a share 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 select the next relaying node, it first filters out neighbors that are in the LHNL, just as in DRP. Next, it filters out nodes whose hop count to the sink is 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.

3 Asymptotic Analysis of the PRP Scheme

The random routes generated by the four algorithms in Section 2 are not necessarily node-disjoint. So, a natural question is how good these routes are in avoiding black holes. We answer this question by conducting asymptotic analysis of the PRP scheme. Theoretically, such analysis can be interpreted as an approximation of the performance when the node density is sufficiently large. It also serves as a lower bound on the performance of the NRRP, DRP, and MTRP schemes. Note that the security analysis for the CN and DOS attacks are similar because both of them involve calculating the packet interception probability. For brevity, we only focus on the CN attack model. The same treatment can be applied to the DOS attack with a straightforward modification.

3.1 Network and Attack Models

We consider an area S that is uniformly covered by sensors with density _. We assume a unit-disk model for the sensor communication, i.e., the transmitted signal from a sensor can be successfully received by any sensor that is at most R_h meters away. Multihop relay is used if the intended destination is more than R_h away from the source.

We assume that link-level security has been established through a conventional cryptography-based bootstrapping algorithm, i.e., consecutive links along an end-to-end path are encrypted by symmetric link keys. So, when a node A wants to send a share to its neighbor B, it first encrypts the plaintext using link key K_{AB} and then sends the ciphertext to B. When B wants to forward the received share to its neighbor C, it decrypts the ciphertext using key K_{AB}, reencrypts the plaintext using key K_{BC} , then sends it to C, and so on. In this way, the openness of the wireless media is eliminated: a node cannot decrypt a ciphertext overheard over the wireless channel if it is not the intended receiver. We also assume that a link key is safe unless the adversary physically compromises either side of the link.

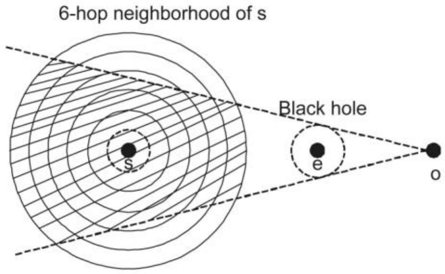


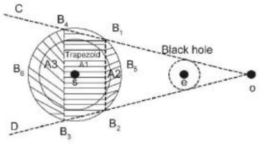
Fig. 3. Packet interception area, a six-hop random propagation example.

The adversary has the ability to compromise multiple nodes. However, we assume that the adversary cannot compromise the sink and its immediate surrounding nodes. This assumption is reasonable because the sink's neighbor-hood is usually a small area, and can be easily physically secured by the network operator, e.g., by deploying guards or installing video surveillance/monitoring equipment. Such an assumption is also widely adopted in the literature, e.g., see [18], [23].

We assume that the black hole formed by the compro-mised nodes can be approximated by its circumcircle, i.e., the smallest circle that encompasses the shape of the black hole. Note that the schemes' operation does not depend on the shape of the black hole. The analysis of the security performance is conservative (i.e., the system is more secure than what it shows by analysis) under this assumption. We denote the circle, its center, and its radius by E, e, and R_e, respectively. During the WSN's operation, any end-to-end path that traverses through this circle is considered vulner-able to eavesdropping, i.e., information shares delivered over this path are all acquired by the adversary. In addition, we also assume that the area S is sufficiently large such that the boundary effect of S can be ignored in our analysis. We will consider the boundary effect in our simulations.

3.2 Security Definition

For a given source sensor node, the security provided by the protocol is defined as the worst-case (maximum) prob-ability that for the M shares of an information packet sent from the source, at least T of them are intercepted by the black hole. Mathematically, this is defined as follows: Let the distance between the source s and the sink o be d_s. As shown in Fig. 3, we define a series of N þ 1 circles cocentered at s. For the ith circle, $1 \_ i \_ N$, the radius is iR_h. For circle 0, its radius is 0. These N þ 1 circles will be referred to as the N-hop neighborhood of s. More specifically



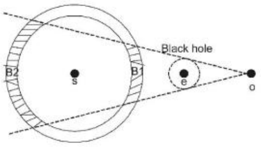


Fig. 4. Packet interception area: Case 2.

Fig. 5. Packet interception area: Case 3

we say that a node is i hops away from s if it is located within the intersection between circles $i \_ 1$ and i. We refer to this intersection as ring i. For an arbitrary share, after the random propagation phase, the id of the ring in which the last receiving node, say w, is located is a discrete random variable _ with state space f1; . . . ; Ng. The actual path from w to the sink is decided by the specific routing protocol employed by the network. Accordingly, different packet interception rates are obtained under different routing protocols. However, the route given by min-hop routing, which under high node density can be approximated by the line between w and the sink, gives an upper bound on the packet interception rates under all other routing protocols. This can be justified by noting that min-hop routing tends not to distribute traffic over various intermediate nodes and only selects those nodes that are closest to the sink. As illustrated in Fig. 3, this path-concentration effect makes min-hop routing have a smaller traversing area of the paths, and thus is more prone to packet interception, especially when compared to power-balancing routing protocols that build dispersive routes.

The worst-case scenario for packet interception happens when the points s, e, and o, in Fig. 3, are collinear (the shaded region denotes the locations of w for which the transmission from w to o using min-hop routing will be intercepted by E). Denote the distance between e and o by d_e. Given d_s and d_e, when s, e, and o are collinear, the shaded region attains its maximum area, and thus gives the maximum packet interception probability. For ring i, denote the area of its shaded portion by S_i. The interception probability for an arbitrary share of information is given by

$$P_I \text{ ¼ } \overset{N}{\underset{i¼1}{\chi}} \text{ Prf\_ ¼ ig } \frac{S_i}{\text{Area of ring i}} \qquad \text{ð 1 Þ}$$

$$\text{¼ } \overset{N}{\chi} \text{ Prf\_ ¼ ig } \frac{S_i}{\_i^2R^2 \quad \_ \ i \ \ 1 \quad ^2R^2} \text{ :}$$
$$\_ \ \text{ð} \ \_ \ \text{Þ}$$

Accordingly, the worst-case probability that at least T out of M shares are intercepted by E is given by

$$P_{S\text{ðmaxÞ}} \text{ ¼ } \overset{M}{\underset{k}{\chi}} \ {}_{T}^{-} \ M_k \, {}^{-}P_I^{\,k}\text{ð1 \_ } P_I \text{ Þ}^{M\_k}\text{:} \qquad \text{ð2Þ}$$

To proceed with the security analysis, we need to calculate the shaded area in each ring S_i and the distribution of _.

3.3 Derivation of the Packet Interception Area

The derivation of S_i falls into one of the following three cases:

Case 1: When iR_h _ d_e $\underline{R_e d_s}$ (e.g., rings 1 to 3 in Fig. 3), ring i is completely covered by the shaded region. Therefore,

$$S_i^{ð\,case\,1\,Þ} ¼ \_½i^2 \_ ði \_ 1Þ^2 \; \&R_h^2; 1\_ i \_ \frac{R_e d_s}{R_h d_e} \_: \quad ð3Þ$$

Case 2: When ði _ 1ÞR^h < $\frac{R_e d_s}{d_e}$ < iR^h, as shown in Fig. 4,

ring i is partially shaded. The shaded area of ring i is the intersection of circle i and the cone CoD minus the area of circle i _ 1. The area of this intersection is composed of three components: the trapezoid A_1 ($B_1B_2B_3B_4$), two circle segments A_2 (surrounded by arch $B_1B_5B_2$ and chord B_1B_2), and A_3 (surrounded by arch $B_3B_6B_4$ and chord B_3B_4). It can be shown that A_1 has a height h_{A1} ¼ x_1 _ x_2, where

$$x_1 \overset{def}{¼} \frac{R_e^2 d_s \; þ \sqrt{R_e^4 d_s^2 \_ d_e^2 R_2^2 d_s^2 \; þ \; d_e^4 i^2 R_h^2 \_ i^2 d_e^2 R_h^2 R_e^2}}{d_e} ; \quad ð4Þ$$

$$x_2 \overset{def}{¼} \frac{R_e^2 d_s \; \_ \sqrt{R_e^4 d_s^2 \_ d_e^2 R_2^2 d_s^2 \; þ \; d_e^4 i^2 R_h^2 \_ i^2 d_e^2 R_h^2 R_e^2}}{d_e} : \quad ð5Þ$$

The lengths of the two parallel edges of A_1 are given by

$$l_1 ¼ 2 \_ \frac{R}{d_e^2 \_ R_e^2} x_1 þ \frac{R_e d_s}{d_e^2 \_ R_e^2} ! ; \quad ð6Þ$$

$$l_2 ¼ 2 \_ \frac{R_e}{\sqrt{\;}} x_2 þ \frac{R_e d_s}{\sqrt{\;}} : \quad 7$$

Therefore, the area of A_1 is given by

$$S_i^{ðA_1Þ} ¼ \frac{ðl_1 þ l_2Þh_{A1}}{2} : \quad ð8Þ$$

The area of A_2 and A_3 are given by

$$S_i^{ðA_2Þ} ¼ ðiR_hÞ^2 \arctan\_ \frac{0:5l_1}{x_1} \_ \_ 0:5x_1l_1; \quad ð9Þ$$

$$S_i^{ðA_3Þ} ¼ ðiR_hÞ^2 \arctan\_ \frac{0:5l_2}{x_2} \_ þ 0:5x_2l_2: \quad ð10Þ$$

So the total shaded area in ring i, d_{Rhde}e _ i _ b_{Rhde} $\frac{R_e d_s}{}$ $\frac{R_e d_s}{}$ þ 1c, is given by

$$S_i^{ðcase\,2Þ} ¼ S_i^{ðA_1Þ} þ S_i^{ðA_2Þ} þ S_i^{ðA_3Þ} \_\_ ði\_ 1Þ^2 R_h^2: \quad ð11Þ$$

Case 3: When ði _ 1ÞR$_h$ _ $\frac{R_e d_s}{d_e}$, as shown in Fig. 5, the shaded area in ring i is the sum of the areas of two ring

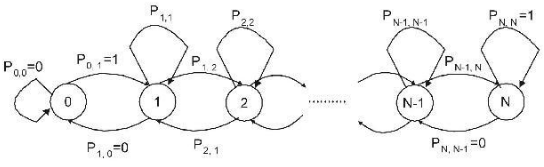


Fig. 6. ðN þ 1Þ-state Markov chain for the random walk.

segments B_1 and B_2. Following a similar approach to Case 2, the areas of B_1 and B_2 are approximated by

B

0 :5l₁

_ _ ði _

Sᵢ^ð ₚ ½i² 1Þ²&Rₕ² arctan⁻ x₁ –; ð12Þ

B 0:5l₂

_ ½i² _ ði _ 1Þ²&Rₕ²

Sᵢ^ð ₂Þ arctan⁻ _ x₂ –; ð13Þ

where x_1, x_2, l_1, and l_2 are given by (4) through (7), with i referring to the ring being calculated. So the total shaded area in ring i is

_Rₕdₑ

Sᵢ^ðcase 3Þ ¼ Sᵢ^ðB₁Þ þ Sᵢ^ðB₂Þ; i – Rₑds þ 1 : ð14Þ

3.4 Derivation of Packet Interception Probability

We derive the distribution of _ in this section. For a given share of information, its random propagation process can be modeled as a random walk, which is described by the ðN þ 1Þ-state discrete-time Markov chain in Fig. 6. A state in this Markov chain denotes the id of the ring that the share is at during the random propagation process. We first notice that the transition probability P10 ¼ 0. This is because in the random propagation process, the information share will never be relayed back to the source node. In addition, we note that the state N is an absorbing state ($P_{N;N}$ ¼ 1 and

N;N_1

¼ 0). This is because an information share takes totally N hops in the random propagation phase, and thus it can reach as the farthest ring N. Furthermore, it is trivial to see that $P_{0;1}$ ¼ 1 and $P_{0;0}$ ¼ 0. The transition probability at other states can be calculated as follows:

Suppose that after the current hop, the share of informa-tion reaches at ring i, where 2 _ i _ N _ 1. Let the location of the node that receives this share of information be w, and denote the one-hop neighborhood of w as circle O_w (this is the circle centered at w and with a radius of R_h). As shown in Fig. 7, the next hop from w has three possibilities:

Case 1: Node w picks a node in region R_1 as the next hop to relay the share. Region R_1 is defined as R_1 ¼ O_w n Circle i,

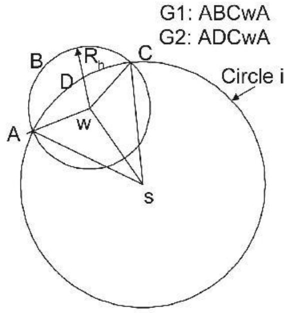


Fig. 8. Calculation of $P_{i;iþ1}$.

where the operation A n B denotes A _ A \ B. This case corresponds to the transition from state i to i þ 1 in the random walk. We use Fig. 8 to illustrate the calculation of the area of R_1. Given the distance from w to o be d, where ði _ 1ÞRₕ < d < iRₕ, the area of R_1 is the difference between the pies G_1 (the area surrounded by the arch ABC and the edges wA and wC) and G_2 (surrounded by arch ADC and the edges wA and wC). The area of G_1 is given by

ᵓG₁ ¼ Rₕ² arcsin o i²Rₕ² _ y² / R ; ð 15 Þ

@ h A

where

y ¼ d² þ ði² _ 1ÞRₕ² / 2d : ð 16 Þ

The area of G_2 is given by

ᵓG₂ ¼ i² 2 Rₕ arcsin o i²Rₕ² _ y² / iR ¹_ ᶻᵓ4Aws' ð 17 Þ

@ h A

where S_{4Aws} is the area of the triangle Aws and can be calculated according to Heron's Formula:

S₄Aws ¼ pðp _ iRₕÞðp _ dÞðp _ RₕÞ; ð18Þ

ðiþ1ÞRₕþd

where p ¼ 2 is half of the perimeter of the triangle.

Given that ði _ 1ÞRₕ _ d _ iRₕ, the conditional prob-ability density function (pdf) of d is given by

f_dðdjði _ 1ÞRₕ _ d _ iRₕÞ

2d / ð2i_1ÞRₕ / 2 forði _ 1ÞRₕ _ d _ iRₕ; ð19Þ

¼ (0; otherwise:

Therefore, the transition probability $P_{i;iþ1}$ can be calculated according to the probability theorem:

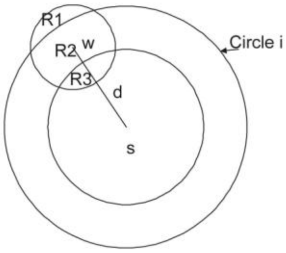


Fig. 7. Possibilities of the next hop.

$$\ulcorner_{i;i\quad 1}\quad \frac{1}{\frac{1}{4}}\ \text{Z}_{\text{ði\_1ÞR}_h}^{iR_h}\quad S_{G_1}\ d\quad S_{G_2}\ d\quad \frac{2d}{\text{ð ÞÞ}}\quad dd;\quad 20$$

$$\text{ð}2i\_ \qquad þ\quad \_R_h^{\,2}\ \text{ð} \qquad \text{ð Þ} \_ \quad 1\text{Þ}R_h^2 \qquad \text{ð Þ}$$

where S_{G1} and S_{G2} are written as functions of d.

Case 2: Node w picks a node in region R_3 as the next hop to relay the share. The region R_3 is defined as R_3 ¼ $O_w \backslash$ Circle i _ 1. This case corresponds to the transi-tion from state i to i _ 1 in the random-walk process. As shown in Fig. 9, given the distance from w to o is ði _ 1ÞR_h < d < iR_h, the area of R_3 is the sum of the areas

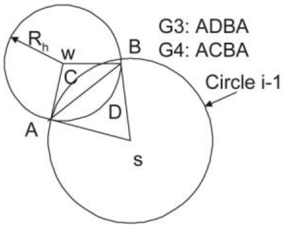


Fig. 9. Calculation of $P_{i;i\_1}$.

G_3 (surrounded by the arch ADB and the chord AB) and G_4 (surrounded by the arch ACB and the chord AB).

The area of G_3 is given by

ði _ 1Þ ^{2}R$_h$^{2} _ y^{02}
^{2} R$_h$ arcsin
$_{G_3}$ ¼ o R$_h$ 1 21
@ A ð Þ
2 ði _ 1Þ 2 R$_h$ 2 _ y^{u};
_2
2 ð$_2$ _ Þ

where y^{u} ¼ ði _2iÞR$_h$þd 2d . The area of G is given by
$_{G_4}$ i 1 ^{2}R$_h$^{2} arcsin i 4 ^{2}R$_h$^{2} v^{02}

¼ ð _ Þ o i 1 R$_h$ 1
y^0 i 1 ^{2}@ R^2 _ ð _ Þ y^2: A ð22Þ
_

Following a similar argument in Case 1, the transition probability $P_{i;i\_1}$ is calculated as

P S S 2d
i;i 1 ¼ G3 d G4 d dd: 23
Z
_ _R$_h$^{2} ði_1ÞR$_h$ ð Þ þ ð Þ ð2i _ 1ÞR$_h$^{2} ð Þ

Case 3: Node w picks a node in region R_2 as the next hop to relay the share, where R_2 ¼ O$_w$ n ðR1 [R3Þ. This corresponds to the situation that the information share will stay in ring i after the next hop relay. Obviously, the transition probability $P_{i;i}$ ¼ 1 _ P$_{i;iþ1}$ _ P$_{i;i\_1}$.

When i ¼ 1, the calculation of $P_{1:2}$ follows exactly the same analysis as in Case 1, i.e., using (20). There will not be Case 3 when i ¼ 1 ($P_{1:0}$ ¼ 0). Therefore, the transition probability $P_{1;1}$ ¼ 1 _ $P_{1;2}$.

Denote the transition probability matrix of the Markov chain in Fig. 6 by P. The element of P can be numerically calculated, according to above analysis. To calculate the distribution of _, we compute the N-step transition probability matrix by conducting the matrix power operation P^{ðNÞ}. The first row of the matrix P^{ðNÞ} gives the probability mass vector of _. Substituting (3), (11), (14), and the distribution of _ into (1), the worst-case packet interception probability is obtained.

3.5 Energy Efficiency of the Random Propagation

We assume that the energy consumption for delivering one

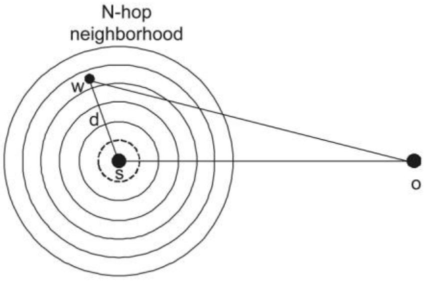


Fig. 10. The total transmission distance after random propagation.

asymptotic assumption, when min-hop routing is used, the ratio between the number of hops from w ! o and from s ! o can be approximated by the ratio of the lengths of these two paths. This ratio can be calculated as follows.

Suppose w is located in the ith ring (see Fig. 10). Let the distance between w and s be ð i _ 1 Þ R$_h$ _ d _ iR$_h$. Given that the angle between sw and so be _, the distance between w and o is given by

d$_{wo}$^{ðiÞ} ðd; _Þ ¼ d^2 þ d$_s$^{2} _ 2dd$_s$ cos _: ð24Þ

Due to the symmetry of the random propagation on all direction, _ uniformly distributed between 0 and 2_. Therefore, the average distance while taking all directions into consideration is given by

i 2_ 1 d^2 þ d$_s$^{2} _ 2dd$_s$ cos _d_: ð25Þ
—

The average distance between w and o given that ði _ 1ÞR$_h$ _ d _ iR$_h$ is given by

d$_{wo}$^{ðiÞ} ¼ ^{Z}ði_1ÞR$_h$ ^{iR_h} ^{Z}0$^{2\_}$ d / 2i 1 _R$_h$^{2} d^2 þ d$_s$^{2} _ 2dd$_s$ cos _d_dd:
ð _ Þ ð26Þ

Therefore, the unconditionally average distance between w and o is given by the weighted sum of d$_{wo}$^{ðiÞ}s with weights f ¼ g, i.e.,

d$^–$$_{wo}$ ¼ $\sum_{i ¼ 1}^{N}$ d^{ðiÞ}$_{wo}$ Pr f _¼ i g; ð27Þ

where the distribution of _ has been obtained in Section 3.4. When min-hop routing is used in the third phase, the number of hops from s to o can be approximated by d_s=R_h. Let the lengths of an information packet and a share generated by the secret sharing algorithm be L_p and L_s, respectively. Note that, in general, L_s _ $\frac{L_p}{M}$, because a share contains a header and other redundant information of its original packet. To account for this segmentation overhead, let the extra bits of a bit over one hop is a constant q. Then, the average energy consumption for delivering one packet from source s to sink o depends on the average length (in hops) of the route. Note that each random route consists of two components. The first is a fixed N-hop component attributed to the random propagation operation. The second component involves sending the share from the last random relay node, i.e., w, to

the sink o using a normal single path routing. Under the

share be a fraction, say _, of the length of the original packet, i.e., L_S ¼ $\frac{L_p}{M}$ þ _L_p. Under this notation, the average energy consumptions for delivering one information packet using PRP can be calculated as follows:

PRP ¯$_{wo}^{u}$ $^{u}_{wo}$¯ ¯q:

$Q^{ð}$ $_Þ$¼ ML_S¯N þ R $_h$¯q ¼ ð1 þ M_ÞL_p¯N þ R $_h$ ð28Þ

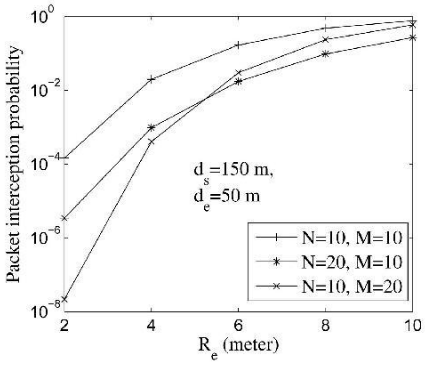


Fig. 11. Packet interception probability versus black hole size.

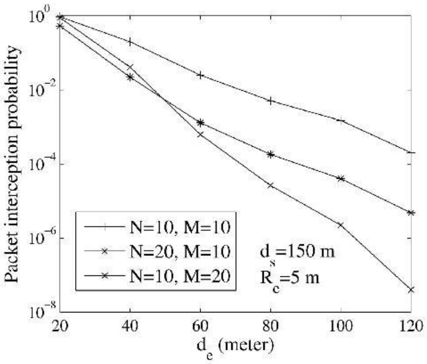


Fig. 12. Packet interception probability versus black hole location.

3.6 Energy-Optimal Secret Sharing and Random Propagation

In this section, we consider the problem of deciding the parameters for secret sharing (M) and random propagation (N) to achieve a desired security performance. To obtain the maximum protection of the information, the threshold parameter should be set as T ¼ M. Then, increasing the number of propagation steps (N) and increasing the number of shares a packet is broken into (M) has a similar effect on reducing the message interception probability. Specifically, to achieve a given $P_S^{ðmaxÞ}$ for a packet, we could either break the packet into more shares but restrict the random propagation of these shares within a smaller range, or break the packet into fewer shares but randomly propagate these shares into a larger range. Therefore, when the security performance is concerned, a trade-off relation-ship exists between the parameters M and N. On the other hand, although different combinations of M and N may contribute to the same P_S^{max}, their energy cost may be different, depending on the parameters L_s, L_p, and q. This motivates us to include their energy consumption into consideration when deciding the secret sharing and random propagation parameters: We can formulate an optimization problem to solve for the most energy-efficient combination of M and N subject to a given security constraint. Formally, this is given as follows:

minimize $Q^{ðPRPÞ}$ðM; NÞ
s:t: $P_s^{ðmaxÞ}$ $_ð$M; N Þ _ $P_S^{ðreqÞ}$; ð 29 $_Þ$
1 _ M _ M_{max};
1 _ N _ N_{max};

where M and N are variables and $P_S^{ðreqÞ}$ is the given security requirement. The upper bounds, M_{max} and N_{max}, are dictated by practical considerations such as the hardware or energy constraints. Because the range of M and N that are of practical interest is not large, e.g., at most few of tens, the space of feasible ðM; N Þ is moderate. Thus, the optimal ðM^o; N^oÞ can be solved by the exhaustive search algorithm.

3.7 Numerical Examples

In this section, we use numerical examples to give a sense of the typical security performance of the PRP scheme. In all calculations, we assume that the secret sharing threshold T ¼ M and the hop range R_h ¼ 10 meters.

3.7.1 Impact of Geometric Parameters

Fig. 11 plots the packet interception probability as a function of the black hole size under various combinations of N and M. It is clear that the message is more likely to be intercepted when the black hole becomes larger, but increasing either N or M helps to reduce the interception rate. It is also noted from the crossing between curves that there is no absolute winner between increasing N and increasing M to reduce the interception probability. In the low-interception-probability regime, increasing M gives better performance, while in the high-interception regime, increasing N becomes better. This can be explained as follows: An increase in N helps to propagate information shares more dispersively, thus reducing the interception probability of each share (P_I). Increasing M does not affect the interception probability of the share, but the black hole needs to collect more shares to recover a packet. From (2), it is clear the latter takes effect as the exponent while the former is on the base. When P_I _ 1, a larger exponent provides faster decay of the probability than reducing the base, and vice versa.

We plot the packet interception probability as a function of the black hole location in Fig. 12. It is clear that the closer the black hole to the sink, the larger the interception probability. This is in line with the many-to-one data collection paradigm in WSNs. For example, if the sink is compromised, then all packets will be intercepted by the adversary (no effective counter-attack measure exists in this case).

3.7.2 Optimization of N and M

In Fig. 13, for a target message interception probability of 10^{-3}, we show the impact of the segmentation overhead _

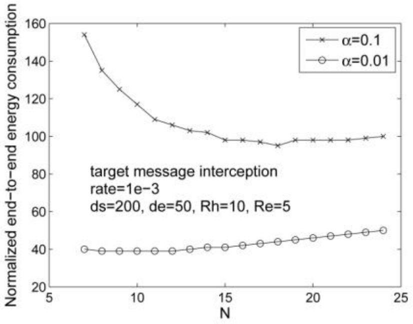


Fig. 13. Energy consumption under different (N, M).

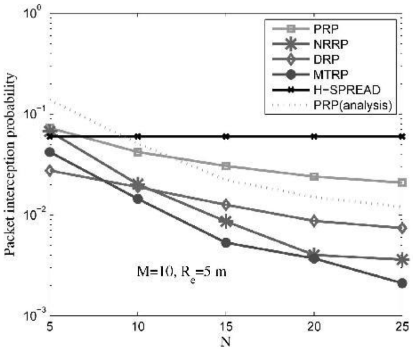


Fig. 14. Packet interception probability versus N.

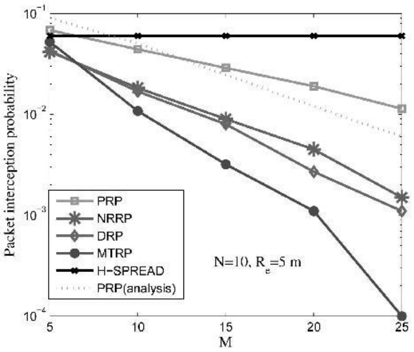


Fig. 15. Packet interception probability versus M.

on the energy-optimal value of ðN; M Þ. We fix the distance between the source node and the sink, but vary the segmentation overhead between two levels: high (_ ¼ 10 percent) and low (_ ¼ 1 percent). The energy consump-tion is normalized by L_pq, i.e., the amount of energy for transmitting the packet for one hop. Note that in this figure, only N is shown and the corresponding M satisfying the security requirement is omitted. Every point on both curves achieves the same target message interception probability. It shows that when the segmentation overhead becomes high, the optimization will favor a larger range of random propagation and a smaller breaking of the information (ðN^0 ¼$_0$ 10; M^0 ¼ 28Þ when _ ¼ 1 percent, and ðN^0 ¼ 17; M ¼ 16Þ when _ ¼ 10 percent). This can be explained by noting that under higher segmentation overhead, the delivery of each piece becomes more expensive in terms of energy consumption. Therefore, the best way is to partition the message into less number of piece, but propagate each piece into further distance. These results can serve as a guide in determining suitable M and N in the protocol's operation. We will evaluate the actual effect of this optimization using simulation in the next section.

4 Simulation Studies

4.1 Simulation Setup

In this section, we use simulation to evaluate the performance of PRP, NRRP, DRP, and MTRP under more realistic settings. To better understand the capability of these randomized multipath routing algorithms in bypassing black holes, we also compare their performance against a deterministic counterpart, H-SPREAD [10], which generates node-disjoint multipath routes to combat CN attack in WSNs.

We consider a 200 m _ 200 m 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 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 ¼ 10 m. For MTRP, we set the parameters _$_1$ ¼ 0 and _$_2$ ¼ 5. In all simulations, after the random propaga-tion 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, 1,000 information packets are sent from the source node to the sink. Our simulation results indicate that the nodes' locations have a significant impact on the absolute value of the packet interception probability of a given scheme. As a result, we emphasize that when reading the simulation results presented below, the absolute value of the mean performance is not as useful as the relative performance ranking between various schemes, and also not as useful as the general trend in performance. Because all comparisons made in our simulations are based on 50 common topologies, this common ground for compar-ison ensures that our results preserve the actual relative performance between various schemes.

4.2 Simulation Results

4.2.1 Single-Source Case

We first fix the location of the source node at ð_50; 0Þ. In Figs. 14 and 15, 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. The packet interception probability calculated according to our asymp-totic analytical model for PRP is also plotted in the same figure for comparison. 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 ¼ 20 in Fig. 14, the interception probability will 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. Furthermore, it can be observed that the simulated performance for PRP is reasonably close to its theoretical performance, especially in the medium packet-interception-probability regime (i.e., 0:01 _ P_S _ 0:1). This clearly demonstrates that the sample topologies used in our simulations are representative and sufficient, and the

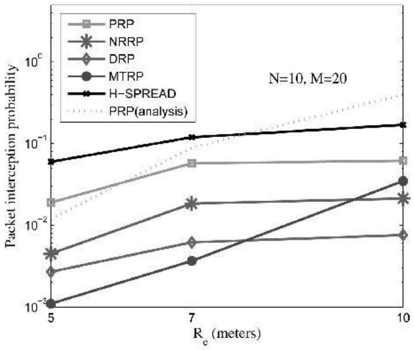


Fig. 16. Packet interception probability versus R_e.

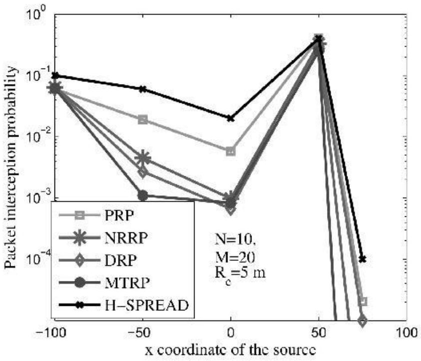


Fig. 18. Packet interception probability at different source location.

simulation results do represent the general performance trend. When the packet interception probability is high (P_S > 0:1) or low (P_S < 0:01), the gap between the theore-tical and simulated results becomes more significant. The overly-optimistic behavior of the analytical model in the low P_S regime is due to ignoring the boundary effect when modeling random propagation. The overly-pessimistic behavior in the high P_S regime is due to the asymptotic assumption made in the analytical model, which under-states the spatial separation between routes when node density is not high enough.

We plot the packet interception probability as a function of the size of the black hole in Fig. 16. It is clear that the interception probability increases with R_e. This trend is in line with our analytical results shown in Fig. 11.

In Fig. 17, we study the impact of node connectivity. The number of nodes is changed from 1,000 to 3,000, corre-sponding 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. From Fig. 11, we can find that even for the asymptotic case, for which the average node degree is infinite, the theoretical interception probability of the PRP scheme is about 1 _ 10^{2}, which is slightly smaller than the simulation results. Such insensitivity to the node connectivity/density is because the packet interception probability is mainly decided by how dispersive the shares can be geographically after random propagation, i.e., how large the cocentered circles in Fig. 3 can be and how the shares are distributed over these circles. As long as the nodes are uniformly distributed, the change of node density does not impact the size of these circles, nor the distribution of the shares over these circles. In contrast, the packet interception probability of H-SPREAD decreases significantly with the increase in node density, because more node-disjoint routes can be found.

In Fig. 18, we slide the x-coordinate of the source node along the line y ¼ 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 _ x _ 0), the closer the source is to the black hole, the smaller the packet interception probability will be. This is in line with our analytical result in Fig. 12. Note that when x ¼ _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 _ 0, the trend in interception probability is reversed. This is because more and more shares are intercepted during the propagation phase. When x ¼ 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 Fig. 19. This hop-count metric can be considered as an indirect measurement of the energy efficiency of the routes generated by the

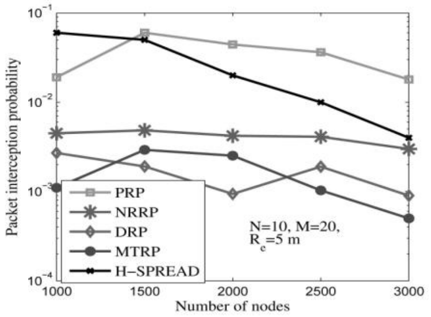


Fig. 17. Packet interception probability versus number of nodes.

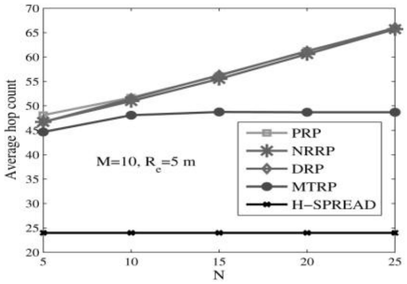


Fig. 19. Hop count of route versus N.

Fig. 20. Energy consumption under different ðN; MÞ.

Fig. 21. Packet interception probability under different ðN; MÞ.

routing schemes. 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 N does not play a role in the H-SPREAD scheme. Under large N, e.g., when N ¼ 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. The relatively large hop count in the randomized algorithms is the cost for stronger capability of bypassing black holes.

4.2.2 Effect of the Optimization of N and M

In Section 3.6, we have formulated an optimization problem for PRP, which finds the most energy-efficient parameter setting ðN^{o}; M^{o} Þ among all feasible combinations of N and M that satisfy a given security requirement. Due to the asymptotic nature of the analytical model, the solution provided by our optimization is only optimal for PRP in an idealized setting. In this section, we use the outcome of our optimization to drive the simulation under PRP, DRP, NRRP, and MTRP, and then measure the resulting packet interception probability and the end-to-end energy consumption. The results will help us better to understand the practical effect of our optimization. Our simulations are conducted as follows:

In the simulations, we assume the same target intercep-tion probability ($P_S^{\text{ðreqÞ}}$ ¼ 10^{-3}) and segmentation overhead (_ ¼ 0:1 and _ ¼ 0:01, respectively) as in the numerical example of Section 3.7.2. Under this setting, in Section 3.7.2, we obtained all feasible combinations of N and M. For each of these feasible ðN; MÞs, we run our simulation under PRP, DRP, NRRP, and MTRP schemes, respectively. Similar to the treatment in Section 3.7.2, the energy consumption measured in our simulations is also normalized by L_pq. The results are plotted in Fig. 20 as a function of N (the corresponding M is omitted in the x-axis of the figure for brevity). The actual packet interception probability ob-served in our simulation is plotted in Fig. 21. Two observations can be made. First, the actual energy con-sumption of PRP measured in the simulations presents a similar trend in ðN; M Þ to that calculated according to our analysis (see Fig. 13, for comparison). As a result, the outcome of our optimization, i.e., ðN^{o} ¼ 10; M^{o} ¼ 28Þ when _ ¼ 0:01 and ðN^{o} ¼ 17; M ^{o} ¼ 16Þ when _ ¼ 0:1, also achieves close-to-minimum energy consumption in the simulation. The discrepancy in the absolute value of the analytical results and the simulation outcome is not surprising, because when node density is not high enough, it takes more hops for a share to reach the destination than what is calculated under the asymptotic assumption. Second, Fig. 21 shows that in the case of PRP, although the use of ðN ^{o}; M^{o}Þ still achieves good energy performance, the resulting packet interception probability violates the original requirement (10^{-3}). However, under DRP, NRRP, and MTRP schemes, the ðN^{o}; M^{o}Þ achieves good energy performance and in most of the time satisfies or is close to satisfying the required interception probability. This phenomenon is attributed to two reasons. First, Figs. 19 and 20 have shown that DRP and NRRP present similar energy performance to PRP. This explains why using ðN^{o}; M^{o}Þ also results in good energy performance for the more advanced designs. Second, due to the overly-optimistic nature of the analysis, it is not surprising that under PRP, the actual packet interception probability under ðN^{o}; M^{o}Þ is higher than the constraint imposed on the optimization. However, because DRP, NRRP, and MTRP have better security performance than PRP, the actual packet interception probability under these algorithms is more likely to satisfy the constraint.

4.2.3 Multisource Case

In Fig. 22, we study the average packet interception probability of the proposed algorithms when there are multiple source nodes that are sending packets simulta-neously in the system. The locations of the five source nodes

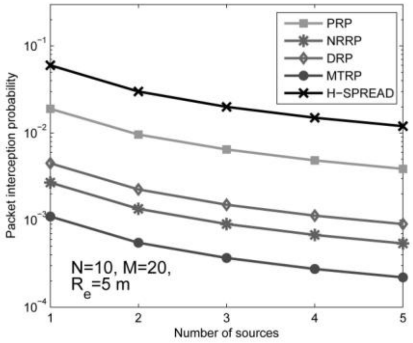


Fig. 22. Average packet interception probability versus number of sources.

used in our simulations are fixed at ð_50; 0Þ, ð_50; 25Þ, ð_50; _25Þ, ð_50; 50Þ, and ð_50; _50Þ, respectively. These nodes are added to the simulation sequentially, yielding the average packet interception probability as a function of number of source nodes. For a given number of source nodes, the average packet interception probability is defined as the total number of packets intercepted by the eavesdropper divided by the total number of packets sent by various sources. This is in contrast to the worst-case interception probability that could happen to any single source node.

From Fig. 22, it is clear that in terms of the average packet interception probability, the relative performance of various algorithms (i.e., the ranking) is similar to that under the worst-case packet interception probability. In particular, all proposed randomized multipath routing algorithms per-form better than their deterministic counterparts. This phenomenon can be explained by noting that an eaves-dropper has to collect at least T shares of the same information packet in order to intercept that packet. In the case of multiple source nodes, it is true that the total number of shares intercepted by the compromised node is greater than that in the single-source case. However, these intercepted shares come from various sources. For a particular source node, the number of shares that originate from that source and are collected by the eavesdropper may not reach threshold T . Therefore, the larger number of intercepted shares does not necessarily indicate a higher average packet interception probability. This is especially true when the source node does not lie on the line connecting the eavesdropper and the sink, whereby most of the shares that belong to the same packet circumvent the black hole. The collinear (or near-collinear) situation contributes the most to the average packet interception probability. This is why the performance ranking in terms of the average packet interception probability under multi-ple sources is similar to that under the worst-case scenario under a single source.

5 Related Work

The concept of multipath routing dates back to 1970s, when it was initially proposed to spread the traffic for the purpose of load balancing and throughput enhancement [15]. Later on, one of its subclasses, path-disjoint multipath routing, has attracted a lot of attention in wireless networks due to its robustness in combating security issues. The related work can be classified into three categories. The first category studies the classical problem of finding node-disjoint or edge-disjoint paths. Some examples include the Split Multi-ple Routing (SMR) protocol [8], multipath DSR [5], and the AOMDV [13] and AODMV [24] algorithms that modify the AODV for multipath functionality. As pointed out in [24], actually very limited number of node-disjoint paths can be found when node density is moderate and the source is far away from the destination. Furthermore, the security issue is not accounted for explicitly in this category of work.

The second category includes recent work that explicitly takes security metrics into account in constructing routes. Specifically, the SPREAD algorithm in [11], [12] 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 [10] improves upon SPREAD by simultaneously accounting for both security and reliability requirements. The work in [6], [7] presents distributed Bound-Control and Lex-Control algorithms, which com-pute the multiple paths in such a way that the maximum performance degradation (e.g., throughput loss) is mini-mized when a single-link attack or a multilink attack happens, respectively. The work in [23] considers the report fabrication attacks launched by compromised nodes. The work in [19] further considers selective forwarding attacks, whereby a compromised node selectively drops packets to jeopardize data availability. Both works are based on a similar cryptographic method: the secret keys used by sensor nodes are specific to their geographic locations, which limits the impact of a compromised node. Instead of relying on a cryptographic method for resolving the issue, our work mainly exploits the routing functionality of the network to reduce the chance that a packet can be acquired by the adversary in the first place. Other secure multipath routing algorithms include SRP [16], SecMR [14], Burmes-ter’s approach [3], and AODV-MAP [21]. Among them, SRP uses end-to-end symmetric cryptography to protect the integrity of the route discovery; SecMR protects against the denial-of-service attack from a bounded number of colla-borating malicious nodes; Burmester’s method is based on the digital signatures of the intermediate nodes; AODV-MAP is another modification of AODV, which can provide local bypass of the attacked nodes.

Given a set of paths that have been constructed, the third type of work studies the optimal way of using these paths to maximize security. For example, the Secure Message Transmission (SMT) mechanism proposed in [17] continu-ously updates the rating of the routes: For each successful (failed) share, the rating of the corresponding route is increased (decreased). The delivery of subsequent shares will be in favor of those routes with high ratings. The work in [4] studies two different ways of spreading an informa-tion packet into shares: secret sharing multipath aggrega-tion (SMA) and dispersed (message-splitting) multipath aggregation (DMA). It shows SMA achieves better security at the cost of higher overhead, while the performance of DMA is exactly the complementary of SMA. In all above work, the multipath routing algorithms are deterministic in the sense that the same set of routes is always computed under the same topology. This weakness opens the door for a pinpointed node-compromise or jamming attack, once the routing algorithm is acquired by the adversary.

Existing randomized multipath 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 multipath routing. Specifically, flooding is the most common randomized multipath routing mechanism. In flooding, every node in the network receives the packet and retransmits it once. To reduce unnecessary retransmissions and improve energy efficiency, the Gossiping algorithm [9] was proposed as a form of controlled flooding, whereby a node retransmits packets according to a preassigned 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 the packet to one randomly picked neighbor. When used to counter compro-mised-node attacks, flooding, Gossiping, and parametric Gossiping algorithms actually help the adversary intercept the packet, because multiple copies of the same 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 secret shares.

6 Conclusions

Our analysis and simulation results have shown the effectiveness of the randomized dispersive routing in combating CN and DOS attacks. By appropriately setting the secret sharing and propagation parameters, the packet interception probability can be easily 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 multipath routing. At the same time, we have also verified that this improved security performance comes at a reasonable cost of energy. Specifically, the energy consumption of the proposed randomized multipath routing algorithms is only one to two times higher than that of their deterministic counter-parts. The proposed algorithms can be applied to selective packets in WSNs to provide additional security levels against adversaries attempting to acquire these packets. By adjusting the random propagation and secret sharing parameters (N and M), different security levels can be provided by our algorithms at different energy costs. Considering that the percentage of packets in a WSN that require a high security level is small, we believe that the selective use of the proposed algorithms does not signifi-cantly impact the energy efficiency of the entire system.

Our current work is based on the assumption that there is only a small number of black holes in the WSN. In reality, a stronger attack could be formed, whereby the adversary selectively compromises a large number of sensors that are several hops away from the sink to form clusters of black holes around the sink. Collaborating with each other, these black holes can form a cut around the sink and can block every path between the source and the sink. Under this cut-around-sink attack, no secret share from the source can escape from being intercepted by the adversary. Our current work does not address this attack. Its resolution requires us to extend our mechanisms to handle multiple collaborating black holes, which will be studied in our future work.

Acknowledgments

A preliminary version of this paper was presented at the IEEE INFOCOM 2009 Mini-Conference. Part of this work was conducted while M. Krunz was a visiting researcher at the University of Carlos III, Madrid, and IMDEA Networks, Spain. This research was supported in part by the US National Science Foundation (under Grants CNS-0721935, CNS-0904681, and IIP-0832238), Raytheon, and the "Con-nection One" center. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the US National Science Foundation.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," IEEE Comm. Magazine, vol. 40, no. 8, pp. 102-114, Aug. 2002.

[2] C.L. Barrett, S.J. Eidenbenz, L. Kroc, M. Marathe, and J.P. Smith, "Parametric Probabilistic Sensor Network Routing," Proc. ACM Int'l Conf. Wireless Sensor Networks and Applications (WSNA), pp. 122-131, 2003.

[3] M. Burmester and T.V. Le, "Secure Multipath Communication in Mobile Ad Hoc Networks," Proc. Int'l Conf. Information Technology: Coding and Computing, pp. 405-409, 2004.

[4] T. Claveirole, M.D. de Amorim, M. Abdalla, and Y. Viniotis, "Securing Wireless Sensor Networks Against Aggregator Com-promises," IEEE Comm. Magazine, vol. 46, no. 4, pp. 134-141, Apr. 2008.

[5] D.B. Johnson, D.A. Maltz, and J. Broch, "DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Net-works," Ad Hoc Networking, C.E. Perkins, ed., pp. 139-172, Addison-Wesley, 2001.

[6] P.C. Lee, V. Misra, and D. Rubenstein, "Distributed Algorithms for Secure Multipath Routing," Proc. IEEE INFOCOM, pp. 1952-1963, Mar. 2005.

[7] P.C. Lee, V. Misra, and D. Rubenstein, "Distributed Algorithms for Secure Multipath Routing in Attack-Resistant Networks," IEEE/ ACM Trans. Networking, vol. 15, no. 6, pp. 1490-1501, Dec. 2007.

[8] S.J. Lee and M. Gerla, "Split Multipath Routing with Maximally Disjoint Paths in Ad Hoc Networks," Proc. IEEE Int'l Conf. Comm. (ICC), pp. 3201-3205, 2001.

[9] X.Y. Li, K. Moaveninejad, and O. Frieder, "Regional Gossip Routing Wireless Ad Hoc Networks," ACM J. Mobile Networks and Applications, vol. 10, nos. 1-2, pp. 61-77, Feb. 2005.

[10] W. Lou and Y. Kwon, "H-Spread: A Hybrid Multipath Scheme for Secure and Reliable Data Collection in Wireless Sensor Net-works," IEEE Trans. Vehicular Technology, vol. 55, no. 4, pp. 1320-1330, July 2006.

[11] W. Lou, W. Liu, and Y. Fang, "Spread: Enhancing Data Confidentiality in Mobile Ad Hoc Networks," Proc. IEEE INFOCOM, vol. 4, pp. 2404-2413, Mar. 2004.

[12] W. Lou, W. Liu, and Y. Zhang, "Performance Optimization Using Multipath Routing in Mobile Ad Hoc and Wireless Sensor Networks," Proc. Combinatorial Optimization in Comm. Networks, pp. 117-146, 2006.

[13] M.K. Marina and S.R. Das, "On-Demand Multipath Distance Vector Routing in Ad Hoc Networks," Proc. IEEE Int'l Conf. Network Protocols (ICNP), pp. 14-23, Nov. 2001.

[14] R. Mavropodi, P. Kotzanikolaou, and C. Douligeris, "SecMR—a Secure Multipath Routing Protocol for Ad Hoc Networks," Ad Hoc Networks, vol. 5, no. 1, pp. 87-99, Jan. 2007.

[15] N.F. Maxemchuk, "Dispersity Routing," Proc. IEEE Int'l Conf. Comm. (ICC), pp. 41.10-41.13, 1975.

[16] P. Papadimitratos and Z.J. Haas, "Secure Routing for Mobile Ad Hoc Networks," Proc. SCS Comm. Networks and Distributed Systems Modeling and Simulation Conf. (CNDS), 2002.

[17] P. Papadimitratos and Z.J. Haas, "Secure Data Communication in Mobile Ad Hoc Networks," IEEE J. Selected Areas in Comm., vol. 24, no. 2, pp. 343-356, Feb. 2006.

[18] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and D. Tygar, "SPINS: Security Protocols for Sensor Networks," Proc. ACM MobiCom, 2001.

[19] K. Ren, W. Lou, and Y. Zhang, “LEDS: Providing Location-Aware End-to-End Data Security in Wireless Sensor Networks,” Proc. IEEE INFOCOM, 2006.

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

[21] B. Vaidya, J.Y. Pyun, J.A. Park, and S.J. Han, “Secure Multipath Routing Scheme for Mobile Ad Hoc Network,” Proc. IEEE Int'l Symp. Dependable, Autonomic and Secure Computing, pp. 163-171, 2007.

[22] A.D. Wood and J.A. Stankovic, “Denial of Service in Sensor Networks,” Computer, vol. 35, no. 10, pp. 54-62, Oct. 2002.

[23] H. Yang, F. Ye, Y. Yuan, S. Lu, and W. Arbaugh, “Toward Resilient Security in Wireless Sensor Networks,” Proc. ACM MobiHoc, 2005.

[24] Z. Ye, V. Krishnamurthy, and S.K. Tripathi, “A Framework for networks.

[25] Routing in Mobile Ad Hoc Networks,” Proc. IEEE Labs, INFOCOM, vol. 1, pp. 270-280, Mar. 2003.

First Author: M.Madhusudhan received B.Tech Degree in Computer Science and Engineering from S.S. Institute of Technology in the year 2012. He is currently M.Tech student in Computer Science and Engineering from MLR Institute of Technology. And his research interested areas are in the field of Mobile Computing and Networking.

Second Author: N Subba Reddy working as an Asst. Professor in MLR Institute of Technology, Dundigal, Ranga Reddy. He has completed his M.Tech CSE and he has 5 years of teaching experience. His research interested areas are Computer Networks and Mobile Computing.

Third Author: G Kiran Kumar is working as Associate Professor & HOD-CSE in MLR Institute of technology. He did M.Tech from Osmania University, Hyderabad, and submitted Ph.D from Nagarjuna University. His research areas include Data Mining, Spatial data mining, Software Engineering.