Reliable Communications over Multihop Networks Under Routing Attacks

Mai Abdelhakim* Leonard Lightfoot† Jian Ren* Tongtong Li*
*Department of Electrical & Computer Engineering, Michigan State University, East Lansing, MI 48824, USA.
† Air Force Research Lab, Wright-Patterson Air Force Base, Dayton, OH 45433, USA
Email: {abdelhak, renjian, tongli}@egr.msu.edu; Leonard.Lightfoot@us.af.mi

***Abstract*—This paper considers reliable multihop transmission under routing attacks, where a malicious relay can modify or drop a packet as it is being forwarded to the destination. We propose a transmission scheme that detects malicious nodes launching routing attacks through incorporating diversity over multi-layer relays, where each relay can establish a direct connection with relays at preceding and succeeding hop levels. We prove that the proposed approach can efficiently detect malicious nodes, provided that there is at least one honest relay at each hop level. We highlight the trade-off between network efficiency and security, and show the impact of the diversity level and the number of hops on the network performance through theoretical analysis and simulation examples. Our results provide insights on general network architecture development and topology design.**

***Index Terms*—Multihop networks, malicious node detection, wireless security, routing attacks.**

I. Introduction

Multihop transmission is a solution to increase network coverage under power constraints. By dividing the communication into multiple shorter-distance transmissions, the overall performance could improve due to the reduced path loss at each hop. However, achieving reliable communications over multihop networks is challenging, especially if the network is deployed in a hostile environment. If any node along a routing path failed or become compromised, then data transfer may fail. Meanwhile, due to the absence of a direct communication with a sink, it could be difficult to identify malicious nodes along a path.

One serious threat to multihop networks is routing attack, where a compromised node along a routing path can drop or modify packets as they are being forwarded [1]. Routing attacks could not be resolved through cryptographic approaches solely, since nodes launching these attacks may have been authenticated to access the network. Utilizing multpath diversity can potentially combat routing attacks [1]–[3].

For efficient transmission in multihop networks, malicious nodes should be identified and avoided. In [4], malicious relays that perform selective forwarding are detected by exploiting a 2-hop downstream acknowledgment and end-to-end assessment. In this scheme, nodes add their opinion about their neighbors in the packets they forward, and the destination establishes an opinion about every node accordingly. A similar approach is presented in [5]. Both schemes in [4], [5] assume single-path routing, and heavily rely on the nodes to accurately report their opinion about their neighbors, which could be difficult to guarantee, especially when multiple relays along a route are malicious [6].

In this paper, first, we propose to incorporate multipath diversity and training/pilot packets to detect malicious relays in multihop networks. In our proposed model, there are multiple relays at each hop level to assist the transmission; each relay can establish a direct communication with relays at preceding and succeeding hop levels. We show that under the assumption that there is at least one honest relay at every hop level, all the malicious relays between the source and the destination can be identified. Then, we highlight the trade-off between network reliability and efficiency, which is reflected in the selection of the diversity level and hop count in multihop networks. Our results provide insights on general network topology design. It is observed that: (i) higher diversity levels would improve the network reliability, but would result in a lower throughput performance and larger complexity. (ii) for a fixed per hop distance, lower number of hops would be generally desired for higher throughput performance and lower overhead, but would limit the distance between the transmitter and the receiver; this raised the concept of hop number control, in which the number of hops in a transmission is limited through proper node deployment and network topology design.

II. Malicious Node Detection Under Routing Attacks

In this section, we present our proposed transmission model for detecting malicious nodes launching routing attacks, where a malicious node would alter the packet being forwarded or drop it.

For a given source-destination pair, we utilize multiple relays at each hop level to assist in a transmission; each intermediate relay at a hop level i from the source, can communicate with all relays at the next hop level $(i+1)$. We assume that at least one relay at every hop level is honest. Let N be the maximum number of hops from the source to the destination, and R_i be the number of relays at hop level i. Denote R_{ij} as the jth relay at the ith hop, where $i = \{1, .., N-1\},\ j = \{1, .., R_i\}$. Let $\alpha_{i,j}$ be a binary flag indicating that $R_{i,j}$ is benign. That is,

$$\alpha_{i,j} = \begin{cases} 1, & \text{if } R_{i,j} \text{ is Benign,} \\ 0, & \text{if } R_{i,j} \text{ is Malicious.} \end{cases} \quad (1)$$

The source S transmits packet M_j to $R_{1,j}, \forall j = \{1, .., R_1\}$. Each relay at the ith hop level ($R_{i,j}$) sends the packet(s) it receives to all relays at the $(i+1)$th hop level. Note that we have $\prod_{h=1}^{N-1} R_h$ possible paths between the source and the destination. The destination should receive all R_1 packets from all paths.

To ensure that packets have not been modified by intermediate relays, the destination verifies the packets' integrity through a cryptographic keyed Hash function [7]. In addition, training/pilot signals are used to enable the sink identify malicious relays that drop packets. Pilot signals are inserted in pre-defined locations that are agreed upon between the source and the destination; pilot locations should be secret to avoid having malicious relays dynamically changing their behavior in response to a pilot signal.

More specifically, we have the following result:

Proposition 1. *For a reliable multihop transmission, when diversity is incorporated by defining a set of relays at every hop to assist the transmission, such that a relay at each hop level can communicate with all relays at the subsequent hop level, then malicious relays launching routing attacks can be identified provided that there exists at least one honest relay per hop.*

Proof: Recall that the number of hops is N. Since there is at least one honest relay at each hop, then there exists at least one reliable path between the source and destination. That is, there exists α_{i,n_i} for every i, such that

$$\alpha_{1,n_1}\alpha_{2,n_2}, ...\alpha_{N-1,n_{N-1}} = 1, \quad n_i \in \{1, 2, ..., R_i\}. \quad (2)$$

In other words, we have

$$\alpha_{1,n_1} = \alpha_{2,n_2} = ... = \alpha_{N-1,n_{N-1}} = 1. \quad (3)$$

Note that the reliable path(s) can be identified using integrity check and pilot packets, as explained earlier. Then, for any relay at hop $i \in \{1, .., N-1\}$, $\alpha_{i,n}$, can be obtained by checking the path corresponding to the following entry in the reliability check matrix:

$$\underbrace{\alpha_{1,n_1}\alpha_{2,n_2}...\alpha_{i-1,n_{i-1}}}_{1}\ \alpha_{i,n}\ \underbrace{\alpha_{i+1,n_{i+1}}...\alpha_{N-1,n_{N-1}}}_{1}. \quad (4)$$

That is, since there exists a reliable path, malicious nodes along any other path can be identified.

■

A. Two-Level Relay-Assisted Transmission

In this subsection, we illustrate our proposed approach through an example of a two-level relay assisted transmission.

The malicious node detection is aided by the construction of a reliability check matrix A; each element in A corresponds to a particular path, and carries a binary value indicating the reliability of the corresponding path. Let $a_{i,j}$ be the element in A at the ith row and the jth column. Then, $a_{i,j} \in \{1, 0\}$, $\forall i, j$, where $a_{i,j} = 1$ indicates that the corresponding path is reliable. A is constructed such that the jth column contains all possible paths that pass through the jth relay at the last hop to the destination.

Upon the reception of packets and verification of their integrity, the receiver can know which paths are reliable, and accordingly sets the corresponding elements in A to 1. After a certain time duration needed to receive all packets, the receiver will be able to detect the unreliable paths and identify the malicious nodes along these paths.

Consider that we have R_1 relays at the first hop level and R_2 relays at the second hop level, then the reliability check matrix A for the two-level relay-assisted transmission can be obtained as:

$$A = \boldsymbol{\alpha}_1 \times \boldsymbol{\alpha}_2^t, \quad (5)$$

where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are $R_1 \times 1$ and $R_2 \times 1$ vectors containing the α values of relays in the first and second hop level, respectively. That is, $\boldsymbol{\alpha}_1 = [\alpha_{1,1}, ..., \alpha_{1,R_1}]^t$, and $\boldsymbol{\alpha}_2 = [\alpha_{2,1}, ..., \alpha_{2,R_2}]^t$.

Now consider the case when there are two relays at each hop, i.e., $R_j = 2$, $j = 1, 2$, as shown in Figure 1. For this network, the reliability check matrix A is:

$$A = \begin{bmatrix}\alpha_{1,1}\\ \alpha_{1,2}\end{bmatrix}\begin{bmatrix}\alpha_{2,1}\alpha_{2,2}\end{bmatrix} = \begin{bmatrix}\alpha_{1,1}\alpha_{2,1} & \alpha_{1,1}\alpha_{2,2}\\ \alpha_{1,2}\alpha_{2,1} & \alpha_{1,2}\alpha_{2,2}\end{bmatrix}. \quad (6)$$

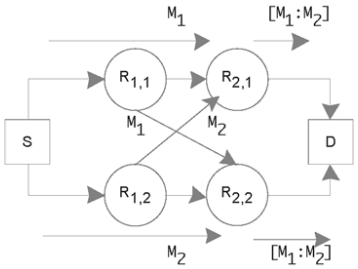

Fig. 1. Proposed model for a two-level relay assisted transmission, with $R_j = 2$, $j = 1, 2$. S is the source and D is the destination/sink.

Here, the source sends M_1 to $R_{1,1}$ and M_2 to $R_{1,2}$. Then, $R_{1,1}$/$R_{1,2}$ sends M_1/M_2 to both $R_{2,1}$ and $R_{2,2}$, respectively. The destination then receives M_1 and M_2 from both $R_{2,1}$ and $R_{2,2}$.

If $R_{1,2}$ is malicious: Consider the case when $R_{1,2}$ is malicious and all other nodes are benign. Then, assuming good channel conditions, the sink identifies the behavior of $R_{1,2}$ through the following steps:

- The sink would successfully receive M_1 from $R_{2,1}$ and $R_{2,2}$. Accordingly, it sets $a_{1,1} = \alpha_{1,1}\alpha_{2,1} = 1$ and $a_{1,2} = \alpha_{1,1}\alpha_{2,2} = 1$, which indicate that $R_{1,1}$, $R_{2,1}$ and $R_{2,2}$ are benign.
- The sink receives M_2 from $R_{2,1}$ in error (or may not receive it at all), that is because M_2 passes through $R_{1,2}$, which is malicious and may modify or drop the packet. Hence, after a time duration sufficient to receive all packets, the sink sets $a_{2,1} = \alpha_{1,2}\alpha_{2,1} = 0$. Similarly,

$a_{2,2} = 0$. Now, we have $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. From the construction of A, it can be seen that the common term in the zero elements in A, i.e., in $a_{2,1}$ and $a_{2,2}$, is $\alpha_{1,2}$. This confirms that $R_{2,1}$ is the unreliable node.

If any other node is malicious or multiple nodes are malicious at different hop levels, similar procedure is followed to detect malicious nodes.

B. Reliability Check Matrix for Three-level Relay-Assisted Transmission

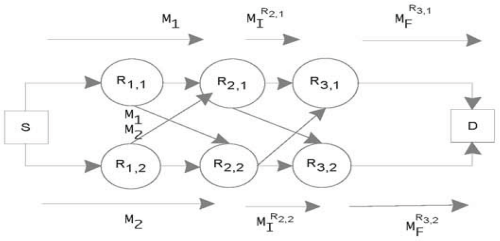

Fig. 2. Three-level relay assisted transmission with $R = 2$.

In this subsection, we illustrate the construction of the reliability check matrix of a three-level relay assisted transmission with two relays at each hop ($R = 2$), as shown in Figure 2. Here, the relays $R_{3,1}$ and $R_{3,2}$ are sending the following messages to the destination D, respectively:

$$\begin{aligned} M_F^{R_{3,1}} &= [M_I^{R_{2,1}}, M_I^{R_{2,2}}] \\ &= [M_1^{R_{1,1},R_{2,1}}\ M_2^{R_{1,2},R_{2,1}},\ M_1^{R_{1,1},R_{2,2}}\ M_2^{R_{1,2},R_{2,2}}], \end{aligned} \tag{7}$$

$$\begin{aligned} M_F^{R_{3,2}} &= [M_I^{R_{2,1}} : M_I^{R_{2,2}}] \\ &= [M_1^{R_{1,1},R_{2,1}}\ M_2^{R_{1,2},R_{2,1}},\ M_1^{R_{1,1},R_{2,2}}\ M_2^{R_{1,2},R_{2,2}}], \end{aligned} \tag{8}$$

where $M_I^{R_{i,j}}$ is the intermediate message transmitted by $R_{i,j}$, and $M_j^{R_{i,j},R_{m,n}}$ is message j that passed through $R_{i,j}$ and $R_{m,n}$. Note that when both $R_{3,1}$ and $R_{3,2}$ are benign, we will have $M_F^{R_{3,2}} = M_F^{R_{3,1}}$. Otherwise, one of them is malicious.

Recall that $\boldsymbol{\alpha}_1 = [\alpha_{1,1}, ..., \alpha_{1,R_1}]^t$, $\boldsymbol{\alpha}_2 = [\alpha_{2,1}, ..., \alpha_{2,R_2}]^t$, and $\boldsymbol{\alpha}_3 = [\alpha_{3,1}, ..., \alpha_{3,R_3}]^t$. To construct the reliability check matrix for the network in Figure 2, the following steps are performed:

1) Construct a temporary check matrix for the first two relaying levels as illustrated in the previous subsection. We get,

$$\begin{aligned} A_t &= \begin{bmatrix} \alpha_{1,1}\alpha_{2,1} & \alpha_{1,1}\alpha_{2,2} \\ \alpha_{1,2}\alpha_{2,1} & \alpha_{1,2}\alpha_{2,2} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \end{bmatrix}. \end{aligned} \tag{9}$$

2) Convert the matrix to a column vector as: $\mathbf{r_t} = \begin{bmatrix} \mathbf{a_1}^t \\ \mathbf{a_2}^t \end{bmatrix}$. That is,

$$\mathbf{r_t} = \begin{bmatrix} \alpha_{1,1}\alpha_{2,1} \\ \alpha_{1,1}\alpha_{2,2} \\ \alpha_{1,2}\alpha_{2,1} \\ \alpha_{1,2}\alpha_{2,2} \end{bmatrix}. \tag{10}$$

3) Multiply $\mathbf{r_t}$ with $\boldsymbol{\alpha}_3^t$, which is a row vector containing the α values of the last hop. That is,

$$\begin{aligned} A &= \begin{bmatrix} \alpha_{1,1}\alpha_{2,1} \\ \alpha_{1,1}\alpha_{2,2} \\ \alpha_{1,2}\alpha_{2,1} \\ \alpha_{1,2}\alpha_{2,2} \end{bmatrix} \begin{bmatrix} \alpha_{3,1} & \alpha_{3,2} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{1,1}\alpha_{2,1}\alpha_{3,1} & \alpha_{1,1}\alpha_{2,1}\alpha_{3,2} \\ \alpha_{1,1}\alpha_{2,2}\alpha_{3,1} & \alpha_{1,1}\alpha_{2,2}\alpha_{3,2} \\ \alpha_{1,2}\alpha_{2,1}\alpha_{3,1} & \alpha_{1,2}\alpha_{2,1}\alpha_{3,2} \\ \alpha_{1,2}\alpha_{2,2}\alpha_{3,1} & \alpha_{1,2}\alpha_{2,2}\alpha_{3,2} \end{bmatrix}. \end{aligned} \tag{11}$$

Based on A, all paths can be checked and malicious nodes can be identified in a similar manner as described in the previous subsection.

C. Reliability Check Matrix for R-level Relay-Assisted N-hop Transmission

For N-hop network with R relays at each hop, the check matrix can be obtained through the following steps:

1) Construct the $R \times R$ check matrix for the first two relaying levels as follows:

$$A_t = \boldsymbol{\alpha}_1 \times \boldsymbol{\alpha}_2^t. \tag{12}$$

Define a counter i, and set $i = 3$.
2) Transform A_t to a column vector $\mathbf{r_t}$, by concatenating the transpose of each row in A_t.
3) Introduce the effect of the next relaying level by multiplying $\mathbf{r_t}$ with $\boldsymbol{\alpha}_i^t$, which is a row vector containing α's of the ith relaying level. Hence, we have, $A_t = \mathbf{r_t} * \boldsymbol{\alpha}_i^t$
4) Increment i, and repeat steps 2 and 3 for each subsequent hop level, i.e., until $i = N - 1$.

Remark 1. *When TDMA protocol is used for multiple access, the relays are assigned to transmit in a sequential order based on their hop level, where nodes closer to the source are assigned to transmit first. Note that relays at higher hop levels (farther away from the source), transmit more packets, and hence are allocated larger time duration or multiple time slot.*

III. Choice of the Network Parameters

In this section, we discuss the impact of the number of hops and the diversity level on the network performance, and show the trade-offs in a multihop network topology design.

Let the number of hops in a transmission be N, and the number of relays at each hop be R. The values of N and R have a direct impact on the network performance, and should be selected to achieve a good trade-off between reliability and efficiency.

The impact of N: As N increases, the network coverage increases. However, large N could have a negative impact on the network reliability and efficiency. This is because: (i) in a single path case, assuming nodes can be compromised with an equal probability, then adding more nodes along a routing path would increase the chances of attacks/failures; (ii) in a multipath case, increasing N increases the complexity of the malicious node detection scheme. This can be seen clearly from the number of packet needed to be forwarded by each relay; (iii) even in a benign environment, under a fixed per hop distance and signal to noise ratio, as N increases, the throughput decreases. This will be illustrated in the following subsections.

The impact of R: As R increases, the diversity level increases, and theoretically the system is more reliable. However, practically, R should be limited for the following reasons: (i) increasing R would increase the complexity of the malicious node detection; (ii) increasing R would decrease the throughput and increase the delay.

In the following subsections, we first discuss the complexity of the proposed malicious node detection; next, we show the relationship between the network parameters with the percentage of malicious nodes; then, we analyze the throughput and show the impact of R and N on the attained performance.

A. Complexity of Proposed Malicious Node Detection Scheme

We measure the complexity of the proposed approach by the number of transmissions and receptions that are required at each node. Let R_i be the number of relays at hop level i. Based on the description provided in Section II, it is noted that: (i) each relay at the first level receives one packet from the source and sends one packet to all relays at the second level; (ii) each relay at the third hop level receives the R_1 packets from each of the R_2 relays at the second level and forwards them to the R_4 relays at the fourth level, and so on. That is, a relay at level $j \in \{2, .., N-1\}$ receives $\prod_{i=1}^{j-1} R_i$ packets and is required to forward them to relays at level $j+1$. Note that the complexity of the proposed approach increases as the number of relays or number of hops increases.

When malicious node detection is not employed, the conventional mulitpath routing can be used, where a relay along any path sends/receives the same number of packets transmitted by the source along the same path, regardless of the number of hops. Hence, much lower complexity would be involved, at the expense of lower reliability under malicious attacks. Therefore, a network can define two modes of operation: the detection mode and the transmission mode. In the detection mode, relays form a mesh network as described in Section II to detect malicious nodes. Then, the system goes to the transmission mode, where the more conventional multipath routing is employed with the identified malicious nodes avoided.

B. Relationship with the Percentage of Malicious Nodes

The underlying assumption of the proposed malicious node detection scheme is to have at least one benign relay at each hop level. Therefore, the diversity level has to be selected to ensure that this condition is met.

Let γ be the percentage of malicious nodes, and assume all nodes can be compromised with an equal probability. That is, the probability that a node is honest is $P_H = \Pr\{\alpha_{i,j} = 1\} = 1-\gamma,\ \forall i, j$. Let $R_i = R,\ \forall i$. Hence, the probability that there is at least one honest relays among the R selected relays is

$$\begin{aligned} P_s &= \sum_{j=1}^{R} \binom{R}{j} P_H^j (1-P_H)^{R-j} \\ &= \sum_{j=1}^{R} \binom{R}{j} (1-\gamma)^j \gamma^{R-j}. \end{aligned} \tag{13}$$

As can be seen from (13), as R increases, P_s increases. For a reliable transmission, R should be selected such that P_s is above a certain threshold, as will be further illustrated in Section IV.

For an N-hop path, the probability that there is one or more malicious nodes along the path is

$$P_x = \sum_{i=1}^{N} \binom{N}{i} \gamma^i (1-\gamma)^{N-i}. \tag{14}$$

It is clear from (14), as N increases, P_x increases.

From the discussions above, it can be concluded that: large R and small N would be preferred for a reliable transmission. However, the former would increase the system complexity and cause large overhead and delay; and the latter would result in a limited distance between the communicating ends, which has to be reflected in the network architecture and topology design. As can be seen, there is an obvious trade-off between reliability and efficiency. In the following, we will show the effect of R and N on the throughput performance.

C. Effect of R and N on Throughput Performance

The throughput T is the average number of packets per slot initiated by the source s and successfully delivered to the intended destination [8]. Recall that R_i is the number of relays at hop level i. Note that the source transmits R_1 packets to relays at the first hop level. Let the binary flag $t_s = 1$ indicates that the source and all intermediate relays are scheduled to transmit the R_1 packets of s to the destination, and $r_s = 1$ indicates that packets are successfully received at the destination. Hence, the throughput in packets per slot can be formulated as:

$$T = R_1 P(t_s = 1) P(r_s = 1 | t_s = 1). \tag{15}$$

Here, we assume that the pilot signal overhead is negligible. The throughput can be calculated before and after malicious node detection, where all paths or the reliable paths only are utilized, respectively.

1) Throughput before malicious node detection: Before malicious nodes are detected, all relays between the source s and the destination are utilized for data forwarding.

The transmission probability, $P(t_s = 1)$, depends on the adopted scheduling protocol. Here, we consider TDMA pro-

tocol and assume no bandwidth reuse. To calculate the number of time slots needed for our proposed scheme, it is noted that: (i) the source node needs R_1 time slots to transmit to the first level of relays; (ii) each relay at the first hop level needs one time slot to transmit to the second level of relays; (iii) each relay at the second level needs R_1 time slots to transmit all packets to the third level of relays, and so on. In general, a relay at level $j \in \{2, .., N-1\}$ needs $\prod_{i=1}^{j-1} R_i$ time slots, where a duration of a time slot is equal to the packet duration. Therefore, the length of the TDMA schedule, in number of slots, required to transmit all R_1 packets is $L_{TDMA} = R_1 + \sum_{j=1}^{N-1} R_j \prod_{k=1}^{j-1} R_k = R_1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} R_k$. It is noted that: as the number of hops or the number of relays per hop increases, the delay increases. Now, the transmission probability is obtained as:

$$P(t_s = 1) = \frac{1}{L_{TDMA}} = \frac{1}{R_1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} R_k}. \quad (16)$$

The probability of successful reception, $P(r_s = 1|t_s = 1)$, depends on the channel conditions as well as the percentage of malicious nodes. The information is received successfully at the destination if the first level of relays are benign and successfully receive the packets, and if any of the paths from the source to the destination is reliable and has good channel conditions, i.e., all relays along a path successfully receive the information and forward it towards the destination. Recall that γ is the percentage of malicious nodes, and that $P_H = P(\alpha_{i,j} = 1) = 1-\gamma, \ \forall i, j$. Let P be the set of paths from the source to the destination and N_p be the number of hops along path p. Here, we assume that decode-and-forward strategy is adopted at the relays. When a relay at hop level h along path p is honest, the information is received successfully at the next hop level if the corresponding signal to noise ratio $SNR_{h,p}$ is above a certain threshold SNR_{th}. That is, we have:

$$P(r_s = 1|t_s = 1) = P_1 \sum_{P} \prod_{h=1}^{N_p-1} P_H P\{SNR_{h,p} \geq SNR_{th}\}, \quad (17)$$

where P_1 is the probability that the first level of relays are benign and successfully receive the R_1 packets. Hence, we have $P_1 = \prod_{\substack{r=1 \\ r \in p}}^{R_1} P_H P\{SNR_{0,p} \geq SNR_{th}\}$, where $SNR_{0,p}$ is the signal to noise ratio from the source to a relay at the first hop level along path p.

Assume the transmit power of each node (P) is exponentially distributed with mean $\bar{P}$, i.e., $\Pr\{P = x\} = \bar{P}^{-1} \exp\{-\bar{P}^{-1}x\}$. The signal to noise ratio at hop level h of path p is $SNR_{h,p} = \frac{\bar{P} L_{h,h+1,p}^{-\beta}}{N_o}$, where $L_{h,h+1,p}$ is the distance between the relay at hop h and that at hop $h+1$ along path p, N_o is the noise power, and $\beta \geq 2$ is the path loss exponent. Note that the assignment of channels and time slots can be managed to minimize the interference. Following (17), we get:

$$P(r_s = 1|t_s = 1) = P_1 \sum_{P} \prod_{h=1}^{N_p-1} P_H \exp\left\{-SNR_{th} \frac{N_o}{\bar{P}} L_{h,h+1,p}^{\beta}\right\}. \quad (18)$$

Note that there are $\prod_{i=1}^{N-1} R_i$ possible paths from the source to the destination. Let the average per hop distance be L_{avg}, and set $L_{h,h+1,p} = L_{avg}, \ \forall h, p$. From (16)-(18), we get the following result

Proposition 2. *In a multihop-multipath transmission, assuming exponentially distributed transmit powers, if each node become compromised with probability γ, then the throughput is expressed as:*

$$T = R_1 Pr\{t_s = 1\} \sum_{P} (1-\gamma)^{N_p + R_1 - 1} \exp\left\{-(N_p + R_1 - 1) SNR_{th} \frac{N_o}{\bar{P}} L_{avg}^{\beta}\right\}, \quad (19)$$

where N_p is the number of hops along path p, R_1 is the number of relays at the first hop level, $Pr\{t_s = 1\}$ is the probability that the source and all intermediate relaying nodes are scheduled to transmit the data to the intended destination, β is the path loss exponent of the channel, L_{avg} is the average per hop distance.

Based on Proposition 2, when all paths are scheduled for transmission, if we set $R_i = R, \ \forall i, \ N_p = N, \ \forall p$, and substitute with (16), we get:

$$T = \frac{1}{1 + \sum_{j=1}^{N-1} R^{j-1}} R^{N-1} (1-\gamma)^{N+R-1} \exp\{-(N + R - 1) \ SNR_{th} \frac{N_o}{\bar{P}} L_{avg}^{\beta}\}. \quad (20)$$

Remark 2. Throughput-diversity relationship: *Assuming that there is sufficient spectrum resources, it can be seen from (20) that the throughput decreases as the diversity level R increases. Note that lower R would reduce the robustness of the network to malicious attacks, as illustrated in the previous subsection. This elucidates the trade-off between network reliability and efficiency.*

2) Throughput after malicious node detection: By utilizing the proposed approach for malicious node detection, the source can select the most reliable path for information exchange. In this case, all nodes along the path can be assumed honest. Following similar procedure, the average throughput of a multihop transmission is

$$T = \frac{1}{N} \exp\{-N \ SNR_{th} \frac{N_o}{\bar{P}} L_{avg}^{\beta}\}. \quad (21)$$

Here, the transmission probability is set to $\frac{1}{N}$, which is conservative measure assuming no bandwidth reuse along the path. Note that, in general, multipaths can be found reliable and the source can choose to transmit over all reliable paths for higher diversity. The result can be extended accordingly by averaging over all possible paths. The transmission probability is obtained by calculating the length of the TDMA schedule needed in this case.

Remark 3. Throughput and hop-number relationship: *It can be seen from (19) and (21) that under fixed SNR and average per hop distance, as N increases, the throughput decreases. That is, for high throughput performance, the network topol-*

ogy should be designed to minimize the number of hops. This enforces the concept of hop number control in large-scale networks architecture, in which the number of hops from any node to a sink is limited to a pre-specified number through node deployment and topology design [9].

IV. Simulation Results

In this section, we first show the impact of the diversity level on the network reliability, then evaluate the throughput performance highlighting the impact of number of hops and the number of relays per hop on the performance.

Example 1: Selection of diversity level R Here, we assume that the number of relays at each hop level is equal, i.e., $R_i = R,\ \forall i$. The probability that there is at least one benign relay at each hop P_s versus the the number of relays per hop R is plotted in Figure 3 for different values of the percentage of malicious nodes γ. It is shown that P_s increases as more relays are utilized per hop due to the achieved diversity. For a given γ, R can be selected to ensure that P_s is above a certain threshold. For example, to ensure that $P_s \geq 0.95$ when $\gamma = 30\%$, then three relays per hop would be sufficient. Increasing R would add more robustness to failures/errors that could be introduced by the channel, however it will also increase the complexity and delay of the system.

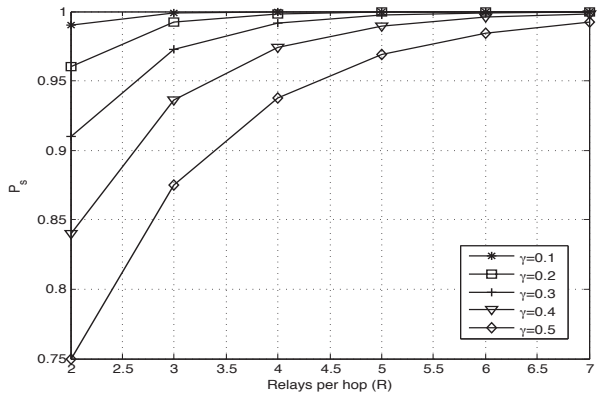

Fig. 3. Probability of having one benign relay per hop (P_s) vs. R for different percentage of malicious nodes γ.

Example 2: Throughput performance before malicious node detection In this example, we evaluate the throughput performance of a multihop transmission with different values of the number of hops (N) and diversity level (R). Here, we set $\gamma = 40\%$, $L_{avg} = 4$m, $SNR = 10dB$, $SNR_{th} = 5dB$. Figure 4 shows the throughput versus N at different values of R. As can be seen, the throughput decreases with both R and N. This echoes our theoretical analysis.

V. Conclusions

In this paper, we proposed a networking approach for detecting malicious nodes launching routing attacks. The proposed scheme incorporates diversity through exploiting multiple relays at each hop level. We proved that, if there exists at least one benign relay per hop, malicious nodes can be correctly identified. We investigated the impact of the diversity level and the number of hops on the network performance, and highlighted the trade-off between network reliability and efficiency. It is concluded that higher diversity level would improve the network reliability, but would result in a lower throughput and higher complexity. Also, for a fixed per hop distance, lower number of hops would be generally desired for higher throughput and reliability; however, it would limit the distance between the source and the destination. This trade-off raised the concept of hop number control, in which the number of hops in a transmission is limited through proper node deployment and network topology design.

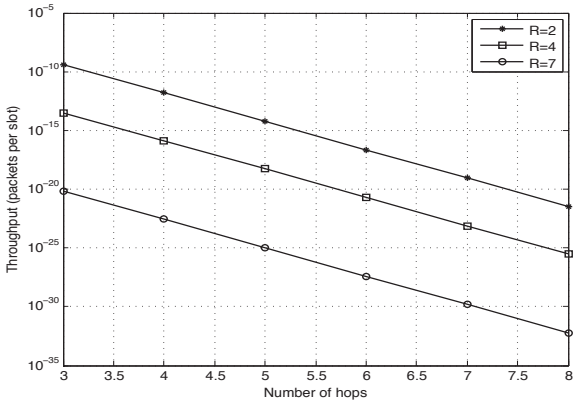

Fig. 4. Throughput vs. number of hops at different values of R. Here, $\gamma = 0.4$, $L_{avg} = 4$m, $\beta = 2$, $SNR = 10dB$, $SNR_{th} = 5dB$.

References

[1] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," *IEEE International Workshop on Sensor Network Protocols and Applications*, pp. 113 – 127, May 2003.

[2] S. Kumar and S. Jena, "SCMRP: Secure cluster based multipath routing protocol for wireless sensor networks," *2010 Sixth International Conference on Wireless Communication and Sensor Networks, WCSN'10*, pp. 1–6, Dec. 2010.

[3] M. Abdelhakim, J. Ren, and T. Li, "Throughput analysis and routing security discussions of mobile access coordinated wireless sensor networks," *IEEE Global Communications Conference, GLOBECOM'14*, Dec. 2014.

[4] Q. Liu, J. Yin, V. Leung, and Z. Cai, "Fade: Forwarding assessment based detection of collaborative grey hole attacks in wmns," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5124–5137, Oct. 2013.

[5] D. Shila, Y. Cheng, and T. Anjali, "Mitigating selective forwarding attacks with a channel-aware approach in wmns," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1661–1675, May 2010.

[6] M. Abdelhakim, L. Zhang, J. Ren, and T. Li, "Cooperative sensing in cognitive networks under malicious attack," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011*, pp. 3004–3007, May 2011.

[7] A. Alkandari, I. Al-Shaikhli, and M. Alahmad, "Cryptographic hash function: A high level view," *International Conference on Informatics and Creative Multimedia, ICICM'2013*, pp. 128–134, Sept 2013.

[8] H. Wang and T. Li, "Hybrid ALOHA: A novel MAC protocol," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5821–5832, Dec. 2007.

[9] M. Abdelhakim, J. Ren, and T. Li, "Mobile access coordinated wireless sensor networks –topology design and throughput analysis," *IEEE Global Communications Conference, GLOBECOM'13*, pp. 4627–4632, Dec. 2013.