Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2015, Article ID 737269, 11 pages
http://dx.doi.org/10.1155/2015/737269

Research Article

Network Coding versus Replication Based Resilient Techniques to Mitigate Insider Attacks for Smart Metering

Pierre Brunisholz,[1] Ochirkhand Erdene-Ochir,[2] Mohamed Abdallah,[3,4] Khalid Qaraqe,[3] Marine Minier,[5] and Fabrice Valois[5]

[1]*INP, LIG, CNRS UMR, 5217 Grenoble, France*
[2]*Department of Electrical Engineering, Qatar University, Doha, Qatar*
[3]*Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar*
[4]*Department of Electrical and Communications Engineering, Faculty of Engineering, Cairo University, Giza, Egypt*
[5]*INRIA, INSA-Lyon, CITI, Université de Lyon, 69621 Lyon, France*

Correspondence should be addressed to Marine Minier; marine.minier@insa-lyon.fr

Received 18 December 2014; Accepted 7 June 2015

Academic Editor: Chao Song

The main focus of this paper is the resilience of communication protocols for data gathering in distributed, large scale, and dense networks. In our previous work, we have proposed the resilient methods based on random behavior and data replications to improve route diversification, thus to take advantage of redundant network structure. Following these previous methods, we propose in this paper a new resilient method based on network coding techniques to improve resilience in Wireless Sensor Networks (WSNs) for smart metering applications. More precisely, using our resilience metric based on a performance surface, we compare several variants of a well-known gradient based routing protocol with the previous methods (random routing and packet replications) and the new proposed methods (two network coding techniques). The proposed methods outperformed the previous methods in terms of data delivery success even in the presence of high attack intensity.

1. Introduction

Recent advances in wireless communications and electronics have developed a next generation of distributed, large scale, and dense networks. In particular, Wireless Sensor Networks (WSNs) have become popular for smart metering to gather data from multifunctional sensor nodes communicating at short distance to collect and transmit data to one or more data collectors. WSNs have well-known features such as low-power consumption, changing topology awareness, open noisy environment, and unreliable radio links. This leads to possible collisions and interferences which makes data gathering a real challenge. In addition, the nature of wireless communication medium of the smart devices combined with their deployment in an open environment makes them vulnerable to malicious attacks.

An outsider attacker could eavesdrop on communications and alter transmitted messages. Smart electronic devices deployed in an unattended and possibly hostile environment for resource monitoring enable physical attacks and their resource limitations (computation, energy, and communication) could ease node tampering. Due to node compromise, numerous malicious insider attacks are possible such as injecting bad data to the network to manipulate control actions and provide multiple Denial-of-Service (DoS) attacks to disrupt data gathering process for monitoring.

The main focus of this paper is the resilience of such constrained networks for data gathering. Resilience study encompasses a wide range of multidisciplinary research topics and is still a relatively new concept in networking. Recently, in works on the resilience of Internet, several concomitant domains (fault tolerance, security, survivability, etc.) were jointly considered in [1–4], where the lack of a metric and a valid definition of the resilience in networking is underlined. Resilience of routing protocols has been defined

as their ability to absorb the performance degradation under some failure pattern (random or intentional) and to continue delivering messages with an increasing number of k compromised nodes in [5]. This definition introduces an analogy related to the original definition of resilience in mechanics which characterizes the properties of the materials to resist a shock.

In our previous work, a quantitative metric was proposed in [5] and several resilient routing techniques such as Random Gradient Based Routing (RS-GBR) and Random Gradient Based Routing with Replication (RM-GBR) were studied in [6]. They consist of three main features: (i) introduction of random behavior, (ii) limitation of route length, and (iii) data replication. Random behaviors increase uncertainty for an adversary, making the protocols unpredictable. Limiting the route length is necessary to reduce the probability of a data packet meeting a malicious insider along the route. Data replication allows route diversification between the smart devices and the data collector, thus improving the delivery success rate and the fairness of the network. Such techniques are particularly interesting for smart metering to take advantage of the redundant topology created through wireless medium.

The originality and novelty of this paper is introducing a new resilient routing technique based on network coding mechanisms to exploit data redundancy, thus to take advantage of the route diversity inherently present in wireless networks. Network coding [7] provides an interlaced multiple packets generation process, where each packet possesses a part of the data information a source node wants to route to the data collector. The network coding techniques are well investigated [7]; however, to our knowledge, this is the first study applying those techniques to resilience context to improve routing protocols against insider attacks. The objective is thus to compare the proposed network coding techniques with the data replication mechanisms studied previously [8] in terms of our resilience metric [5].

The rest of the paper is organized as follows. Section 2 gives an overview of the attack categories in smart metering. In Section 3, the previous resilient methods are presented including random behavior and packet replications. In Section 4, we propose new resilient methods based on network coding techniques. Section 5 defines the resilience metric, assumptions, and simulation parameters followed by the results and comparative analysis. Finally, Section 6 concludes this paper and outlines future work directions.

2. Attack Categories

This section introduces the main attack categories in smart metering communication network according to the ontology introduced in [9].

According to its *capabilities* an attacker can be characterized as *laptop-class* and *mote-class*. A *laptop-class* attacker may have access to powerful devices with more computational resources. A single laptop-class attacker might be able to eavesdrop on and/or jam the entire network. In smart metering infrastructure, *mote-class* attacker with no resource advantages over legitimate nodes is also possible because an ordinary smart device can be captured and compromised.

According to its *intent* an attacker can be characterized as *passive* and *active*. A *passive* attacker attempts to learn or make use of information from a system but does not affect system resources. For example, passive eavesdropping that simply gathers information can compromise privacy and confidentiality. An *active* attacker attempts to alter system resources or affect system operations. Compared to the passive attacker, here the goal is to produce DoS attacks to disrupt communication by destroying links or exhaust available resources such as bandwidth or energy. Such attacks are challenging for smart metering, which relies on reliable data gathering from smart devices for resource monitoring.

According to *point of initiation* an attacker can be characterized as *outsider* and *insider*. An *outsider* attack is initiated from the outside of the security perimeter by an unauthorized or illegitimate user of the system. For instance, jamming, eavesdropping, and injecting replayed or fabricated messages are examples of such attacks. An *insider* attack is one that is initiated by an entity inside the security perimeter, that is, an entity that is authorized to access system resources but uses them in a way not approved by the party that granted the authorization. This is possible for smart metering because an attacker could tamper low cost network devices like smart meters deployed in an open and unattended environment. Such attacks are particularly critical for smart metering because an *insider* attacker could inject false data to manipulate control actions and/or provide numerous DoS attacks such as selective forwarding, Sinkhole, Sybil, node replication, and Wormhole to disrupt data gathering process [10].

In this paper, we deal with an attacker that has no resource advantages over legitimate devices because ordinary network nodes could be captured and compromised by an adversary. We deal with an *insider attacker* (physical attacks on electronic devices deployed in an open environment), who is authorized to access network resources but uses them in an inappropriate way. An adversary is *active*: he/she attempts to alter network resources and/or affect network operations especially at routing layer.

3. Previous Resilient Methods

Most of the routing protocols for data gathering in smart metering are deterministic, based on the "best" route selection criterion to be efficient. For instance, the RPL routing protocol standard adopted by IETF in March 2012 favors the most "stable" routes [11]. As a result, the same route is used to deliver all data packets of a source to a destination. Its sensitivity to faults and attacks has been shown in [12, 13]. Note that the packet delivery success and failure are not fairly distributed among the network nodes; some nodes will have good delivery ratio and others will be completely disconnected. This is a limitation of the protocol, since the redundant topology created by wireless communications is not exploited to benefit from physically existing alternative routes.

In previous work, we have shown through simulations in [6] and analytically in [14] that the random behavior improves the resilience of communication protocols in the presence of insider attacks. It increases uncertainty for an adversary, making the protocols unpredictable. In addition, it allows route diversity, as each data packet takes potentially different routes thanks to the random selection of the next hop. This enhances the connectivity between a source and a destination.

When the random route selection is combined with data replication, the delivery success of each data packet is improved thanks to the route diversity and data redundancy. Based on these observations, to take advantage of data redundancy and path diversity, thus to improve resilience, we introduce a new resilient technique based on network coding mechanisms. The main goal of this paper is to study both data replication and network coding based techniques and to compare them in terms of resilience against insider attacks.

3.1. Random Behavior. We have proposed a theoretical framework of the resilience based on the biased random walks [14]. The theory of random walks is widely used in various fields such as mathematics, physics, and telecommunications. In networking, random walk is related mostly to a data packet generated at a node, traveling randomly across the network to reach a destination. Given a graph and a starting node, a neighbor node is selected at random and a data packet is sent to this neighbor; then again a neighbor of the current node is selected at random, and the data packet is sent to it, and so forth. The random sequence of nodes selected in this way to route a data packet from a source to a destination is a random walk on a graph [15].

In the context of networking, the traditional unbiased random walks are not relevant due to their low performance in average number of hops required to reach a destination [16]. In unbiased random walk, no state information on the direction is available and the probability of selecting the next hop is uniform. Thus, a data packet could travel a long time across the network before reaching a destination. A bias could be introduced to random walks to reduce the route length. Biased random walks study the influence of bias on the stochastic process (random behavior) to determine its influence on the performance of random walks. The most biased random walk is based on the shortest path principle allowing reaching the destination faster, while keeping a random behavior.

Previously, we have introduced the random behavior to the well-known gradient based routing (GBR) protocol [6] considering different parameters for the route length. Simulation results showed that introducing random selection of the next hop to the GBR routing protocol based on shortest paths improved the resilience against insider attacks without bringing about a significant extra cost. However, despite this improvement, the general shape of the average delivery success curve was still concave down, reflecting a sensitivity to insider attacks. Therefore, data replication mechanisms are necessary to provide data redundancy, thus, to increase the delivery success of each data packet.

3.2. Data Replication. To improve the delivery success of each message, the packet replication is introduced and combined with random selection of the next hop. If the original packet is lost, some replicated copies could reach the destination successfully. The classical deterministic protocols cannot take advantage of data replication as a source uses the same route for all messages, whereas the randomized variants may increase the delivery success thanks to multiple routes.

In a previous study, we have studied several replication methods such as replicating packets at the sources only, both at the source and at each intermediate node along the route [6], depending on the distance of a source to the destination (distant nodes replicate more than closest nodes) [8], and so forth. The best method that allows a good trade-off between the data delivery success and the overhead was the replication at the sources only [8].

Even though the data replication combined with random routing improves the resilience, it brings about an overhead. Note that providing redundant data has an important cost in terms of energy consumption. Now, the main goal is to study other mechanisms to provide data redundancy and take advantage of route diversity in the network. It seems to us that the most intuitive method to provide data redundancy is to introduce network coding mechanisms to improve the resilience of routing protocols and compare this method to the data replication methods studied previously.

4. The Proposed Resilient Methods

Still aiming at a delivery success improvement, we introduce a new resilient method based on network coding instead of packets replication. This version is still combined with random next hop selection. The main idea is to generate multiple coded packets from the data a node wants to send. Each coded packet will contain a part of information from the original one. Due to the random next hop selection, each coded packet will take a different route, increasing the overall delivery success rate. As soon as the data collector receives enough coded packets, it decodes them in order to retrieve the original data.

In order to understand how this version works, it is important to give several definitions of network coding [7]. Basically, we can say that *coding at a node in a network is network coding*, where coding means a causal mapping from inputs to outputs. This definition has the inconvenience of not distinguishing the network coding we are going to speak about from the channel coding used in noisy networks.

We will then define the network coding as *coding at a node in a network with error-free links*. Moreover, this definition helps us to make a difference between network coding and source coding.

But this definition can be more specific, and if we are considering that we are in packets networks, we can define network coding as *coding content of packets inside a node*. If we had a little generalization by saying that we apply the coding above the physical layer, we can distinguish the network coding function from the information theory. Then

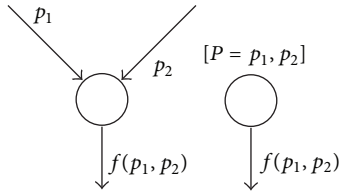

Figure 1: Network coding illustrations.

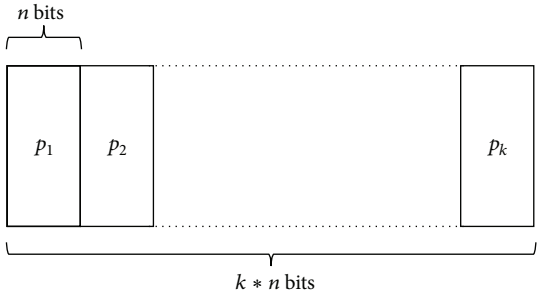

Figure 2: Subdivision illustration.

we base our work on the previous definition of network coding.

There are two major versions of network coding: the interflow network coding and the intraflow network coding as shown in Figure 1. Interflow network coding consists in coding different packets from various origins together to create a packet containing information from all of them, while intraflow network coding consists in creating multiple coded packets from an original one. In this work, we focus on intraflow network coding. More precisely, we will use intraflow network coding based on random linear network coding.

In packets networks, intraflow (intrasession) network coding consists in dividing a message (a data packet) into multiple submessages of the same size and then creating a linear dependency between them before transmitting [7]. When the data collector receives enough packets, it can recreate the initial message, by resolving the linear system created by the linearly independent subpackets. More precisely, random linear network coding works as follows [7].

Step 1 (the packet subdivision). As we have seen before, intrasession network coding relies on the division of a data packet into a predefined number of the same sized packets. Here, we consider a message as a chain of bits.

The initial source node has to split a data packet p into k packets $p_1, p_2, \ldots, p_k$ of n bits as illustrated in Figure 2. This implies that the original message has to be a multiple of n. Usually, taking $n = 8$ means that the required coefficients are chosen in a Galois Field of size 2^8. This field size allows the created packets to be linearly independent with a probability $P = 0.996$, and every coefficient has the size of a *byte*, which is a good compromise between practicability and linear independence [17].

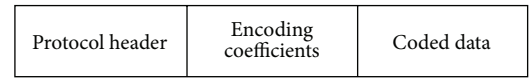

Figure 3: Encoded packet.

Step 2 (the coding coefficients choice). For each packet p_i, the node has to randomly choose $k' \geq k$ coefficients $\langle c_{p_i}^1, c_{p_i}^2, \ldots, c_{p_i}^{k'} \rangle$ from a Galois Field GF(2^8) to form the coefficients vector. To enhance performances, the coefficients are randomly picked in a precomputed Galois Field table.

Step 3 (the coding). We put the previous vectors in a $k * k'$ matrix to obtain the coefficients matrix:

$$k' \geq k \overbrace{\begin{pmatrix} c_1^1 & c_2^1 & \cdots & c_k^1 \\ c_1^2 & c_2^2 & \cdots & c_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{k'} & c_2^{k'} & \cdots & c_k^{k'} \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}}^{k}. \quad (1)$$

Then we create the encoded data Y_j using the formula $Y_j = \sum_{i=1}^{k} c_i^j p_i$, with $j = 1, \ldots, k'$, and obtained the following matrices:

$$\begin{pmatrix} c_1^1 & c_2^1 & \cdots & c_k^1 \\ c_1^2 & c_2^2 & \cdots & c_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{k'} & c_2^{k'} & \cdots & c_k^{k'} \end{pmatrix} \longrightarrow \begin{bmatrix} Y_1 = \sum_{i=1}^{k} c_i^1 p_i \\ Y_2 = \sum_{i=1}^{k} c_i^2 p_i \\ \vdots \\ Y_{k'} = \sum_{i=1}^{k} c_i^{k'} p_i \end{bmatrix}. \quad (2)$$

Step 4 (the dissemination). Each encoded data Y_j is then encapsulated with its coefficients vector $\langle c_1^j, c_2^j, \ldots, c_k^j \rangle$ in a packet (Figure 3) to be broadcast in the network, using any routing protocol.

Moreover, because we are in a Galois Field, each intermediate node receiving b encoded packets $Y_1, Y_2, \ldots, Y_b$ with their $c_1^i, c_2^i, \ldots, c_k^i$ ($i = 1, \ldots, b$) coefficients can pick new encoding coefficients from GF(2^8) and can create new packets repeating previous steps by linearly combining new coefficients.

Step 5 (the decoding). Whenever the data collector receives m packets, it puts the received coefficients vectors into a matrix. If these coefficients satisfy the full rank matrix condition [18], which means that they are all linearly independent, the data collector can retrieve the original

subdivided messages and thus the original data packet p computing

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{bmatrix} = \begin{pmatrix} c_1^1 & c_2^1 & \cdots & c_k^1 \\ c_1^2 & c_2^2 & \cdots & c_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^m & c_2^m & \cdots & c_k^m \end{pmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}. \quad (3)$$

We have to notice that the vector $\langle p_1, p_2, \ldots, p_m \rangle$ is equal to the original one ($\langle p_1, p_2, \ldots, p_k \rangle$) and is ordered regardless of the Y_i reception order (i.e., there is no particular order for the Y_i vector's elements in order to decode). The original data p is then obtained by assembling p_i together.

Thus, the idea is to use network coding in addition to random behavior to create resilient routing protocols as it will be detailed in Section 5.1.

5. Resilience Evaluation

Resilience evaluation of the proposed resilient methods based on network coding is presented in this section and compared to the previous resilient methods based on random behavior and packet replications. The focus of our simulations is comparing four GBR versions in terms of the resilience metric presented in Section 5.2. GBR has been preferred to other routing protocols as the classical version has obtained the best results in terms of resilience metric according to [8]. The four versions of the GBR routing protocol under study described in Section 5.1 are (i) random variant without data redundancy (RS-GBR), (ii) random behavior with data replication at the sources (RM-GBR), (iii) random variant with network coding without ACK mechanism (RS-GBR-NC), and (iv) random variant with network coding with ACK mechanism (RS-GBR-NC-ACK). The first two variants introduce the previous methods and the last two variants provide the new proposed methods.

5.1. Routing Protocols under Study. *GBR* [19] is a classical flooding based routing protocol well adapted for constraint environment such as WSNs and smart metering. It constructs the routes incrementally using gradient information. The data collector floods an INIT packet in order to set up a gradient. The INIT packet records the number of hops taken from the data collector. Then a node can discover its minimum number of hops from the data collector, called the node "height." The height difference between a node and one of its neighbors is the gradient on that link. Then source nodes send their DATA packets to one of their minimum gradient neighbors and their neighbors do the same until the data collector is reached constructing a single route that guarantees the shortest path.

5.1.1. Previous Resilient Methods. Instead of fixed routes with minimal gradients, *Random Gradient Based Routing (RS-GBR)* avoids deterministic routing by introducing some random choices in the next hop selection. Before forwarding, a node randomly chooses a next hop among its neighbors with a certain probability depending on the distance to the data collector [8]. More precisely, a node divides its next hop node possibilities into two groups: the subset of neighbors closer to the data collector, that is, the set of nodes with *next hop gradient = node's gradient − 1*, and the subset of neighbors with the same height as itself, that is, the set of nodes with *next hop gradient = node's gradient*. The node will randomly choose a next hop in the first group with a probability $P(next\ hop)$ equal to 0.8 and will randomly choose a next hop in the second group with probability $P(next\ hop)$ equal to 0.2. Thus, longer routes are generated but the route diversity for each DATA packet is guaranteed; thus several messages may take potentially different routes and routes are never the same. In this way, malicious nodes are not as powerful as if they were on a static route.

The classical shortest path routing protocols cannot take advantage of data replication as each source uses the same route for all messages, while the randomized versions combined with data replication may increase the delivery success thanks to multiple paths. *Random Gradient Based Routing with Replication (RM-GBR)* replicates a packet a chosen number of times and sends it this number of times over different randomly chosen paths using the RS-GBR random mechanism. For our simulation, RM-GBR replicates two times each DATA packet at the source. Thus, if the original packet is lost, some replicated copies could reach the data collector successfully increasing the probability of the transmission success for each message.

5.1.2. The Proposed Resilient Methods. To show the benefits of network coding in terms of resilience, we designed an *Intrasession Network Coding Random Gradient Based Routing (RS-GBR-NC)* protocol. In this alternative, the DATA packets are coded at the source. The number of generated encoded submessages is fixed and is equal to 16. The original message size is equal to 32 bits and is divided into four 8-bit submessages to use coefficients from GF(2^8). Then, when a node wants to send a DATA message, it generates 16 encoded messages with coefficients taken over GF(2^8). Those encoded messages will take different randomly chosen paths using the RS-GBR mechanism. When the data collector receives enough packets (to retrieve the original message, the data collector needs at least four coded packets), it decodes the message and drops any other related incoming packets. Each packet generated from the same original message has a unique ID identifying the original message, the coded data, and the coding vector used to code the submessages.

We also propose *Random Gradient Based Routing with Network Coding and Acknowledgment (RS-GBR-NC-ACK)* which is a slight improvement of RS-GBR-NC. In this version, the number of coded packets is not fixed. Instead, an acknowledgment mechanism is introduced to dynamically generate coded packets. As soon as the data collector achieves a successful decoding, it sends back to the node an acknowledgment message to stop the coded packets generation. To ensure that each coded packet possesses enough time to reach the data collector and to avoid useless packets generation, a delay between each coded packet generation is introduced.

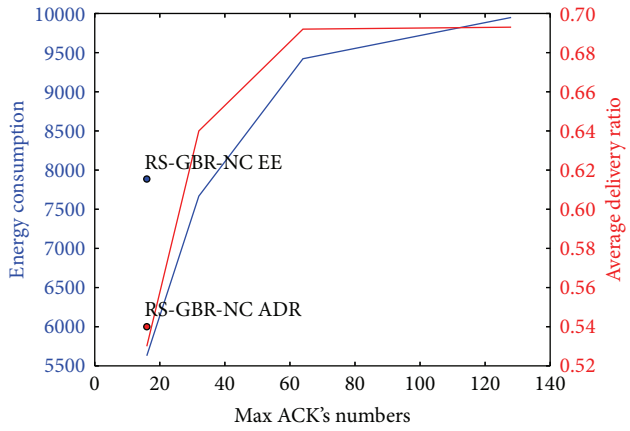

FIGURE 4: Average delivery ratio and energy consumption depending on the maximum number of generated coded packets while waiting for an ACK with 50% malicious nodes.

Moreover, the introduction of an ACK mechanism with unlimited retransmission (i.e., the sender continues to generate new coded packets while it has not retrieved any ACK from the data collector) highlights the fact that some nodes will never be able to communicate with the data collector. Indeed, some devices could be completely disconnected from the network due to a massive presence of malicious nodes in their neighborhood. In this case, these nodes unnecessarily create and send an unlimited amount of messages leading to consumption of a lot of energy in the network for nothing.

This possibility leads to bounding the number of coded packets generated while the node is waiting for the data collector ACK to limit the energy exhaustion of disconnected nodes. Figure 4 shows the results of the increase of the maximum number of generated packets in terms of energy consumption and average delivery ratio while being in an environment with 50% of malicious nodes. As a reference, the values of energy consumption and average delivery ratio of the RS-GBR-NC in the same attacker context are also added to Figure 4. As shown in Figure 4, a good compromise between the average delivery ratio and the energy consumption seems to be for 32 generated coded packets. Indeed, having a number of retransmissions greater than 64 does not improve the average delivery ratio while increasing the energy consumption; and choosing a number of retransmissions equal to 16 leads to lowering the average delivery ratio when compared with RS-GBR-NC even if it greatly decreases the energy consumption. Thus, if the number of generated coded packets is equal to 32, the average delivery ratio is increased by 20% with a slightly lower energy consumption when compared with RS-GBR-NC.

Thus, in the rest of this paper, we choose a number of generated coded packets equal to 32 in the case of RS-GBR-NC-ACK.

5.2. Resilience Metric. A new metric to measure and thus quantify resilience has been proposed in [5]. This metric introduces a new method to aggregate meaningfully several performance metrics using a two-dimensional graphical representation, because of the numerous manifestations of resilient behavior with respect to various performance metrics. The average packet delivery alone is not enough to represent the resilience and we should consider it in combination with several other performance parameters such as fairness, protocol overhead, delay, and average throughput. To obtain a comprehensive measure of resilience, all these metrics need to be somehow meaningfully aggregated. The new resilience metric provides an equiangular polygon surface for each protocol to represent its resilience, where each performance metric is presented by an axis (see Figure 5). Not only does this method allow discerning visually various trade-offs, but also a quantitative value is obtained by computing the polygon surface.

For more comprehensive description, let us consider n routing protocols to compare according to m performance metrics. A routing protocol i ($i \geq n$) is presented by the *resilience surface* of its performance vector $p'_i(k)$, where k is the attack intensity ($k = \{0, 10\%, 20\%, 30\%, 40\%, 50\%\}$, the percentage of compromised nodes among all network nodes).

$p'_i(k) = \{p'_{i,j}(k)\}$ is obtained by rescaling each value $p_{i,j}(k)$, expressed in its own range and unit using the following formula:

$$p'_{i,j}(k) = \frac{p_{i,j}(k) - \min\left(p_{1,j}(k), \ldots, p_{n,j}(k)\right)}{\max\left(p_{1,j}(k), \ldots, p_{n,j}(k)\right) - \min\left(p_{1,j}(k), \ldots, p_{n,j}(k)\right)}, \quad (4)$$

with $p'_{i,j}(k) \in [0;1]$. We need to normalize all values to obtain a polygon surface from several performance metrics with different units. This representation allows zooming in on the protocol differences for an efficient comparison.

Following the work of [5], the resilience could also be directly computed as the area of the polygon using the following formula:

$$R_i(k) = \left(\sum_{j=1}^{n-1}\left(p_{i,j}(k)\, p_{i,j+1}(k)\right) + p_{i,n}p_{i,1}\right)\frac{1}{2}\sin\left(\frac{2\pi}{n}\right). \quad (5)$$

A routing protocol is considered as resilient if this computation stays constant when the percentage k of attackers increases.

5.3. Resilience Parameters. In the context of smart metering, data gathering should be firstly successful, secondly efficient, and finally fairly distributed among all network nodes. The following five performance parameters were selected for the resilience metric evaluation:

(1) *Average Delivery Ratio (ADR).* It is defined as the fraction of the number of received packets by a data collector and the number of sent packets by all sources. It represents the main goal of a WSN in terms of data delivery.

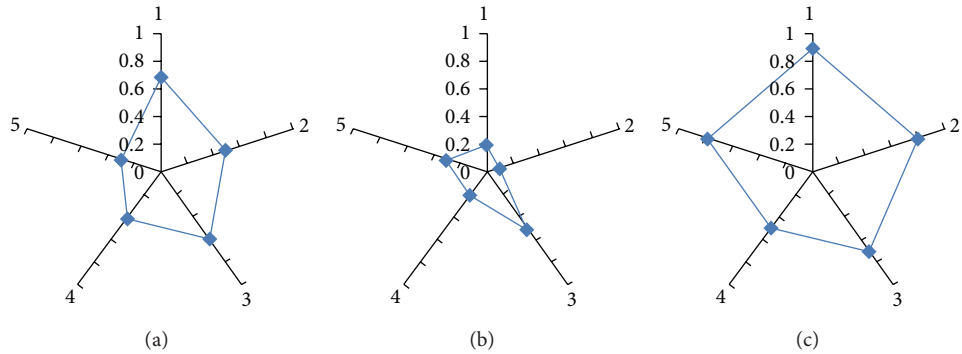

Figure 5: Resilience surface $m = 5$: (a) ordinary case, (b) bad resilience, and (c) good resilience.

(2) *Energy Efficiency (EE)*. It is defined as the efficiency of the overall energy expenditure for all routing generated and forwarded (CONTROL and DATA) packets by all network nodes; this is the protocol overhead in terms of power consumption.

(3) *Delay Efficiency (DE)*. It is defined as the efficiency of average time delay (including queuing, retransmissions, and propagation delays) required for a packet to go from the source to the data collector; this is a function of the average path length (hop count). It characterizes the network efficiency in terms of speed.

(4) *Average Throughput (AT)*. It is defined as the average of the maximum amount of data flows received by a data collector per source per unit of time; this in a sense represents the network capacity (the overall throughput when the traffic is saturated).

(5) *Delivery Fairness (DF)*. It is defined as the standard deviation of the delivery ratio of each source from the average. It characterizes both the fact that sources eventually reach the data collector and the data delivery success distribution among the sources which should be as uniform as possible for good geographic coverage required in smart metering.

5.4. Assumptions and Simulation Parameters. Simulations have been performed using the WSNet simulator [20] and averaged over 100 trials for each case with a 95% confidence interval. For each simulation 300 nodes were deployed randomly through uniform distribution over a 100 m × 100 m square area with a single data collector at the center. Smart meters have a transmission range of 20 m, leading to an average node degree about 31. Table 1 sums up the simulation parameters.

Selective forwarding attack [6, 10] is considered, where forwarding malicious nodes (k% of randomly designated compromised nodes among all network nodes) drop all data packets instead of retransmitting. This attack is relevant in the context of smart metering because the data reliability characterizes the delivery success of routing protocols; the given attack is easy to launch for an adversary and it is common to all routing protocols under study.

Table 1: Summary of the simulation parameters.

Parameter	Value
Number of nodes	300
Area size	100 × 100 meters
Transmission range	20 meters
Topology	Uniformly distributed
Traffic generation	Poisson distribution $\lambda = 1$ packet per second
Simulation time	100 seconds
Number of packets	30000
Number of runs	100

5.5. Results and Analysis. The focus of our simulations in this subsection is the comparison of the four variants of GBR routing protocol with the previous methods (RS-GB and RM-GBR) and the new proposed methods (RS-GBR-NC and RS-GBR-NC-ACK) as described in Section 5.1 in terms of resilience polygon diagram and of resilience surface.

The resilience quantitative evaluation of the four protocols along the 5 axes (ADR, DE, EE, AT, and DF) defined in Section 5.3 is shown in Figure 6. The different protocols exhibit different behaviors. Firstly, as shown in Figure 7, there is a gap between RS-GBR on the one hand and RM-GBR, RS-GBR-NC, and RS-GBR-NC-ACK on the other hand in terms of energy efficiency (EE). Indeed, RM-GBR, RS-GBR-NC, and RS-GBR-NC-ACK transmit more packets than RS-GBR to achieve a successful delivery. As the normalization is computed according to the best result (here RS-GBR), their energy efficiencies become close to zero. However, RS-GBR-NC-ACK has a better energy behavior than RS-GBR-NC when few malicious nodes are present in the network.

Secondly and as shown in Figure 8, we observe the same behavior with the average throughput (AT) metric. As RS-GBR-NC generates 16 times more packets, and because the network capacity is never saturated, this protocol presents a huge throughput compared to RS-GBR, RM-GBR, and RS-GBR-NC-ACK. This leads to the fact that their metrics are close to zero due to normalization and that the throughput variations are not significant.

Finally and as shown in Figure 9, we observe quite the same thing concerning the delay efficiency (DE), but the

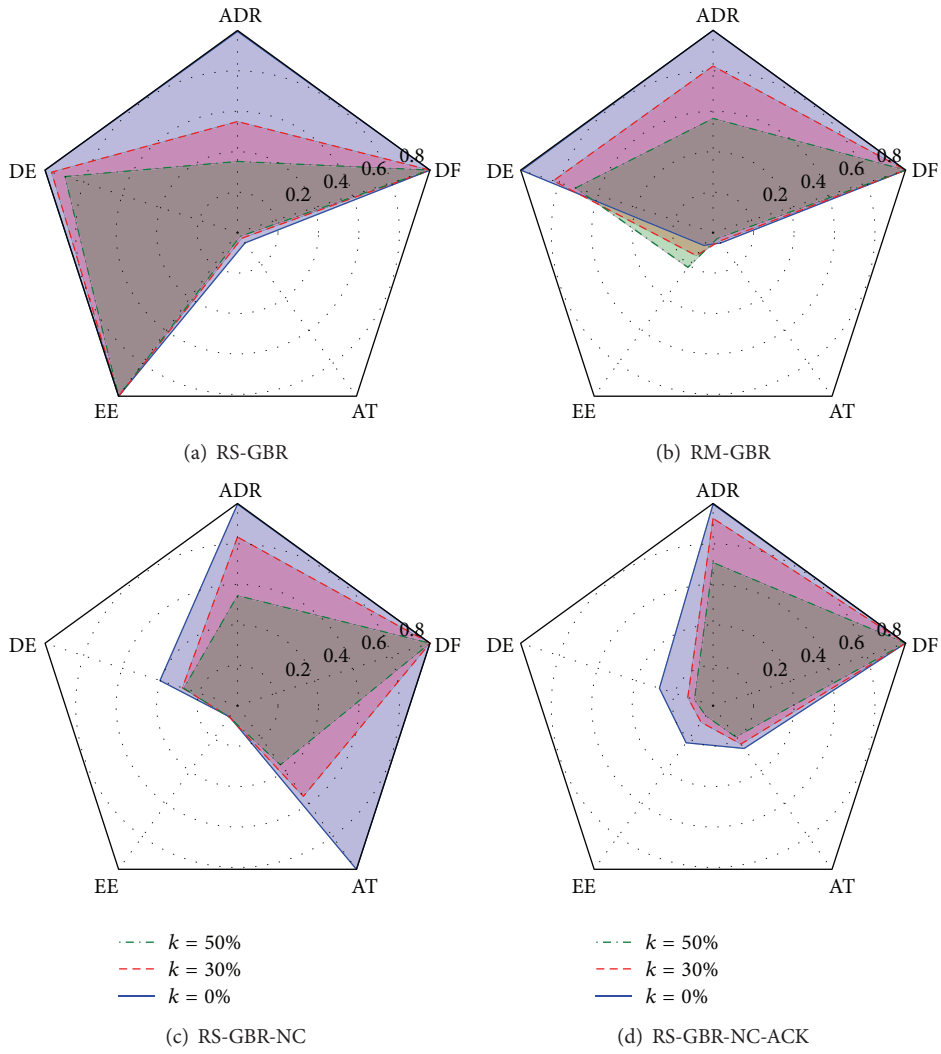

Figure 6: Resilience area depending on the percentage of malicious nodes according to the 5 metrics defined in Section 5.3.

result is not as sharply contrasted as the previous ones. RS-GBR-NC is less efficient due to the fact that the decoding to the data collector is time consuming and that it has to wait until there are enough packets in order to decode. Due to the delay introduction in RS-GBR-NC-ACK to mitigate the number of generated coded packets, this version is slightly below RS-GBR-NC in terms of delay efficiency.

The resilience surface evolution over the percentage of malicious nodes is presented in Figure 10.

As shown in Figure 10, RS-GBR has the best resilience surface over the other protocols. This is mostly due to the fact that RS-GBR has greater energy efficiency than the others. But the resilience surface evolution of RS-GBR is not as good as the others. More precisely, the RS-GBR resilience falls quicker than the others when the malicious nodes increase from 0% to 20%. This means that this protocol is less resilient than the others to the malicious nodes increase. In the same way, RS-GBR-NC has about the same constant decrease as RS-GBR: its resilience to malicious nodes increase has always the same intensity.

Oppositely, RM-GBR resilience falls less quicker during the same malicious nodes increase, which means that it is more resilient to the first increases. The introduction of redundancy when RM-GBR is in use means that even if there is a malicious node on one path, a duplicated packet takes another path which could be malicious-nodes-free. So, redundancy introduction could be seen as a good compromise when the number of malicious nodes present in the network does not exceed 30%.

Concerning RS-GBR-NC-ACK, when the percentage of malicious nodes is below 20%, we could see that RS-GBR-NC-ACK has very low variations and that its resilience surface stays constant. This flat curve means that RS-GBR-NC-ACK is very resilient when the number of malicious nodes begins to increase. However, this protocol has a low resilience value because the flat evolution has a strong cost in terms of delay efficiency and energy efficiency.

Finally, considering only the good delivery of information, regardless of the delay and the energy consumption, Figure 11 presents the results obtained for average delivery

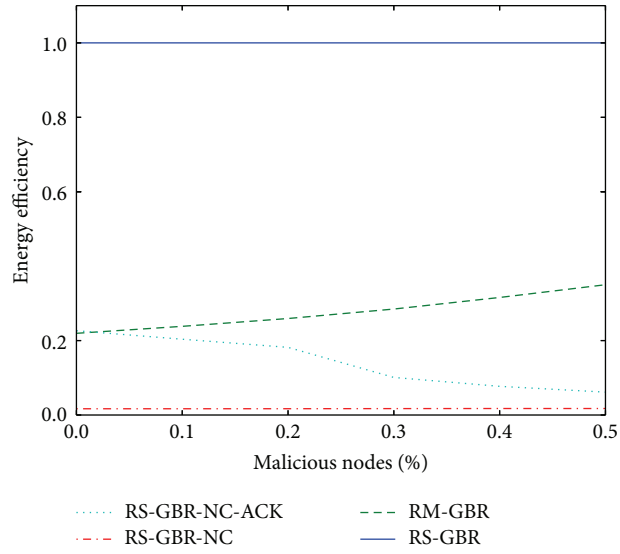

Figure 7: Evolution of energy efficiency in the presence of malicious nodes.

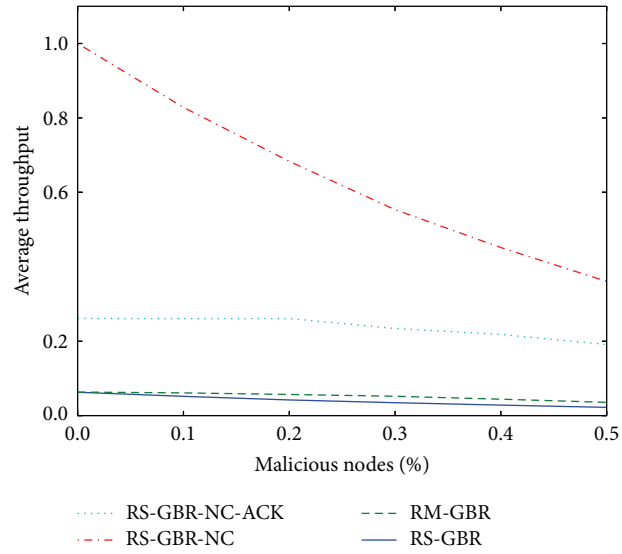

Figure 8: Evolution of average throughput in the presence of malicious nodes.

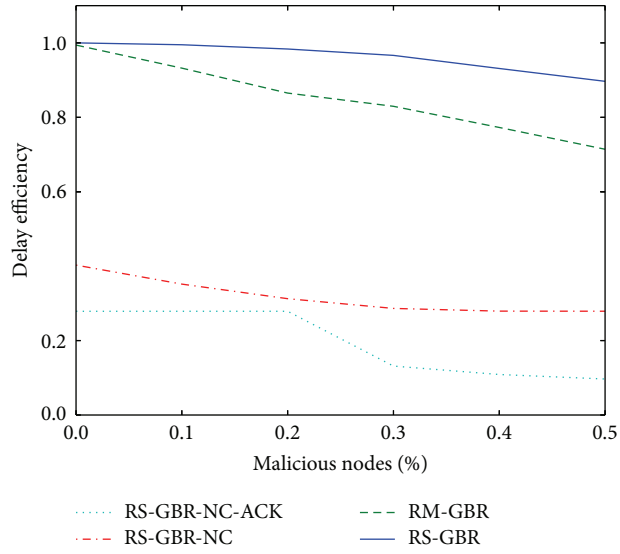

Figure 9: Evolution of average throughput in the presence of malicious nodes.

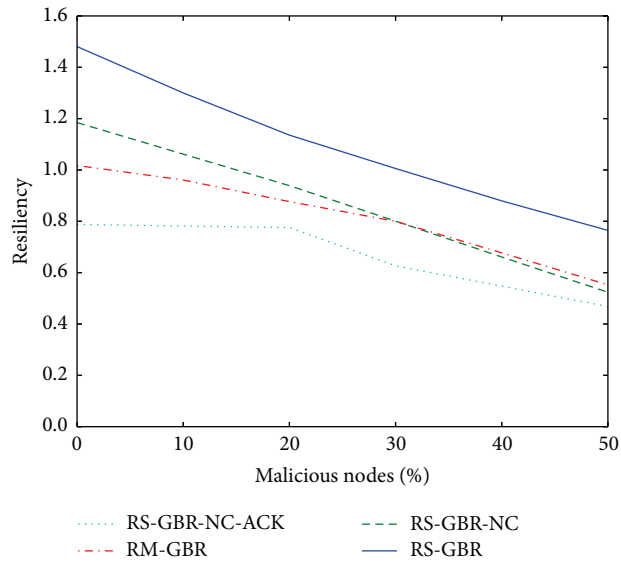

Figure 10: Resilience surface evaluation in the presence of malicious nodes.

ratio alone. The first noticeable thing is that there is a clear gap in terms of delivery ratio between solutions with multiple packets generation (RM-GBR, RS-GBR-NC, and RS-GBR-NC-ACK) and the other (RS-GBR). For solutions with multiple packets, clearly and as shown in Figure 11, the overall network keeps a good average delivery ratio longer before the fall than RS-GBR whereas RS-GBR sees its average delivery ratio falls as soon as there are some malicious nodes in the network. Among the solutions with multiple packets generation the best one is clearly RS-GBR-NC-ACK which has a graceful degradation and thus allows a large number of malicious nodes before having an average delivery ratio below 90% even if this good behavior has a cost as seen before.

We also should notice that because of the nature of the resilience computation, the order of the parameters matters as noticed in [8]. They are acting as a weighting of each other. Using the same experimental protocol as in [21], delay efficiency weights the average delivery ratio, and energy efficiency weights both delay efficiency and average throughput. Even if the importance of the axes position was clearly justified in [21], as we deal here with protocols based on redundancy to achieve good deliveries and because of the sharp contrasts between delay efficiency, energy consumption, and average throughput, maybe another parameters order could be more relevant in this particular case.

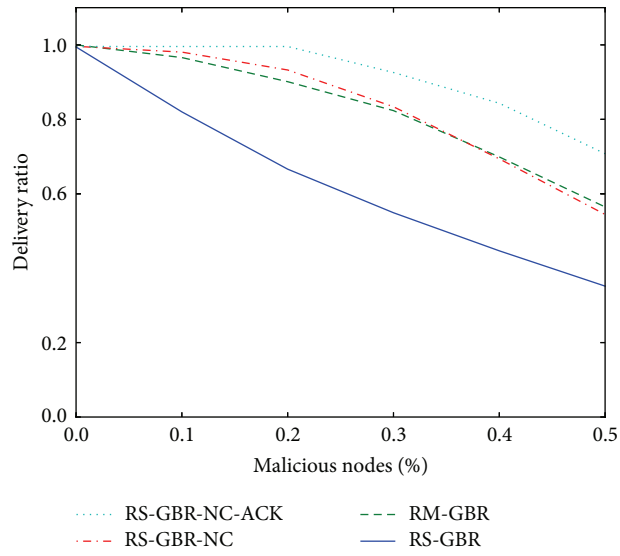

Figure 11: Evolution of average delivery ratio in the presence of malicious nodes.

6. Conclusion

In this paper, we have compared using the metric introduced in [5] the resilience of several routing protocols that exploit redundancy under maliciously packet dropping attacks. We essentially compared redundant methods which are data replication and network coding. Simulation results show that even if RS-GBR has the best resilience surface, the versions that include network coding especially RS-GBR-NC-ACK keep the same resilience surface when the number of attackers increases leading us to think that there is no performance degradation. Moreover, we show that, in terms of average delivery ratio, the routing protocols with multiple packets strategies clearly behave better than RS-GBR.

So, in summary, we proposed some methods that are able to maintain a high average delivery ratio even in the presence of many attackers and even if they have of course an energy cost to pay.

In near future, we plan to modify the resilience metric itself to better take into account, in the context of multiple packet protocols, the gain in terms of average delivery ratio. In the same research direction, the average throughput metric may introduce a bias because, in all cases, we do not achieve the network capacity. We think of using a better parameter like "Goodput" to characterize the quantity of useful data received by time unit.

Disclaimer

The statements made herein are solely the responsibility of the authors.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was made possible by NPRP Grant no. 6-149-02-058 from the Qatar National Research Fund (a member of Qatar Foundation).

References

[1] E. K. Çetinkaya, M. J. F. Alenazi, A. M. Peck, J. P. Rohrer, and J. P. Sterbenz, "Multilevel resilience analysis of transportation and communication networks," *Telecommunication Systems*, 2015.

[2] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya et al., "Resilience and survivability in communication networks: strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, 2010.

[3] P. Smith, D. Hutchison, J. P. G. Sterbenz et al., "Network resilience: a systematic approach," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 88–97, 2011.

[4] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner, and A. Jajszczyk, "A survey of resilience differentiation frameworks in communication networks," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 4, pp. 32–55, 2007.

[5] O. Erdene-Ochir, A. Kountouris, M. Minier, and F. Valois, "A new metric to quantify resiliency in networking," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1699–1702, 2012.

[6] O. Erdene-Ochir, M. Minier, F. Valois, and A. Kountouris, "Toward resilient routing in wireless sensor networks: gradient-based routing in focus," in *Proceedings of the 4th International Conference on Sensor Technologies and Applications (SENSORCOMM' 10)*, pp. 478–483, IEEE, Venice, Italy, July 2010.

[7] T. Ho and D. S. Lun, *Network Coding: An Introduction*, Cambridge University Press, New York, NY, USA, 2008.

[8] O. Erdene-Ochir, M. Minier, F. Valois, and A. Kountouris, "Resiliency taxonomy of routing protocols in wireless sensor networks," in *Proceedings of the 37th Annual IEEE Conference on Local Computer Networks (LCN '12)*, pp. 324–327, IEEE, Clearwater, Fla, USA, October 2012.

[9] W. Znaidi, M. Minier, and J.-P. Babau, "An ontology for attacks in wireless sensor networks," Inria Research Report RR-6704, 2008, https://hal.inria.fr/inria-00333591.

[10] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 293–315, 2003.

[11] T. Winter and P. Thubert, "Rpl: IPv6 routing protocol for low-power and lossy networks," RFC 6550, Internet Engineering Task Force, 2012.

[12] K. Heurtefeux, H. Menouar, and N. Abuali, "Experimental evaluation of a Routing Protocol for WSNs: RPL robustness under study," in *Proceedings of the IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '13)*, pp. 491–498, IEEE, Lyon, France, October 2013.

[13] K. Heurtefeux, O. Erdene-Ochir, N. Mohsin, and H. Menouar, "Enhancing RPL resilience against routing layer insider attacks," in *Proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications (AINA '15)*, pp. 802–807, Gwangiu, Republic of Korea, March 2015.

[14] O. Erdene-Ochir, M. Abdallah, K. Qaraqe, M. Minier, and F. Valois, "A theoretical framework of resilience: biased random walk routing against insider attacks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '15)*, New Orleans, Lo, USA, March 2015.

[15] L. Lovász, "Random walks on graphs: a survey," in *Combinatorics, Paul Erdős is Eighty*, vol. 2, pp. 1–46, Bolyai Society Mathematical Studies, 1993.

[16] I. Mabrouki, G. Froc, and X. Lagrange, "Biased random walk model to estimate routing performance in sensor networks," in *9ème Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications*, pp. 73–76, 2007.

[17] L. Wang, Y. Yang, and W. Zhao, "Network coding-based multipath routing for energy efficiency in wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 115, 2012.

[18] Y. Yang, C. Zhong, Y. Sun, and J. Yang, "Network coding based reliable disjoint and braided multipath routing for sensor networks," *Journal of Network and Computer Applications*, vol. 33, no. 4, pp. 422–432, 2010.

[19] C. Schurgers and M. B. Srivastava, "Energy efficient routing in wireless sensor networks," in *Proceedings of the Military Communications Conference (MILCOM '01). Communications for Network-Centric Operations: Creating the Information Force*, vol. 1, pp. 357–361, IEEE, McLean, Va, USA, October 2001.

[20] E. B. Hamida, G. Chelius, and J.-M. Gorce, "Scalable versus accurate physical layer modeling in wireless network simulations," in *Proceedings of the 22nd Workshop on Principles of Advanced and Distributed Simulation (PADS '08)*, pp. 127–134, IEEE, Roma, Italy, June 2008.

[21] O. Erdene-Ochir, A. Kountouris, M. Minier, and F. Valois, "Enhancing resiliency against routing layer attacks in wireless sensor networks: gradient-based routing in focus," *International Journal on Advances in Networks and Services*, vol. 4, pp. 38–54, 2011.