Review Article


Aly Mohamed El-Semary\textsuperscript{1,2} and Mohamed Mostafa Abdel-Azim\textsuperscript{1,3}

\textsuperscript{1} College of Computer Science and Engineering, Taibah University, Al-Madina Al-Munawwarah, Saudi Arabia
\textsuperscript{2} Faculty of Engineering, Al-Azhar University, Cairo, Egypt
\textsuperscript{3} Faculty of Industrial Education, Beni-Suef University, Beni-Suef, Egypt

Correspondence should be addressed to Mohamed Mostafa A. Azim; mmazim@ieee.org

Received 21 February 2013; Accepted 15 April 2013

Academic Editor: S. Khan

Copyright © 2013 A. M. El-Semary and M. M. A. Azim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, Wireless Sensor Networks (WSNs) have been deployed into a variety of applications including homeland security, military systems, and health care. Sensor nodes deployed in such networks are subject to several attacks such as sinkhole and select forwarding, wormhole, Hello flood, and replication attacks. Therefore, developing secure and energy-efficient routing protocols to protect WSNs against these attacks while efficiently utilizing the energy of the deployed nodes has become imperative. Several routing protocols have been proposed in the literature for WSNs. Most of these protocols assume static nodes and sinks to collect data from network fields. However, they may be highly movable, and recent advances show that mobile sensors in WSNs have a promising performance. Therefore, this paper surveys the state of the art on routing protocols related to WSNs and presents the security issues or problems associated with the current protocols as well as discusses the future trends and open research issues on secure routing protocols of WSNs.

1. Introduction

Wireless Sensor Networks (WSNs) are emerging networking technologies for low-cost, unattended monitoring of a wide range of environments. WSNs received a great importance in the current decade due to the recent development of the IEEE 802.15.4 standard \cite{IEEE} for the physical and MAC layers and the forthcoming ZigBee standard \cite{ZigBee} for the network and application layers. Moreover, the latest delivery of the IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) standard \cite{6LoWPAN} enables us to communicate with devices across the Internet directly without having to go through a translation step such as a ZigBee-to-IP translation.

A WSN typically consists of tens, hundreds, or even thousands of low-cost, low-power, and multifunctional sensor nodes that are deployed in a region of interest \cite{WSN} as shown in Figure 1. These tiny sensor nodes are equipped with sensing functionalities, embedded microprocessors, radio transceivers, and small memory. They communicate over a short distance wirelessly and cooperate to accomplish a common task, such as environmental monitoring.

In this paper, we survey the state of the art on the routing protocols used in WSNs. We demonstrate different security issues to be taken into consideration when deploying WSNs indicating their security requirements for such networks and demonstrating different types of attacks that may face. We also summarize the state-of the art secure routing protocols proposed for WSNs. Finally, we present the future directions and open research issues or problems in the field of routing protocols of WSNs.

The remainder of this paper is organized as follows. Section 2 gives an overview of the applications of wireless sensor networks. Section 3 presents a summary of the state of the art routing protocols used in wireless sensor networks. Section 4 demonstrates security issues in wireless sensor networks. Section 5 presents the up to date secure routing protocols employed in WSNs. Next, Section 6 discusses the recent research issues related to routing protocols of...
WSNs. Finally, the paper concludes with the future research directions in Section 7.

2. Applications of Wireless Sensor Networks

WSNs are currently being employed in a variety of domains ranging from commercial, industrial, environmental, and healthcare to military applications to monitor data that would be difficult or expensive to capture using wired sensors [5, 6]. Based on these fields, a variety of applications have been presented in the literature including aircraft monitoring [7], ecological habit monitoring [8], and geological monitoring [9]. Wireless biomedical sensor networks (WBSNs) [10] are another application domain for WSNs that is characterized by the necessary low error rates compared to traditional WSNs.


3. Wireless Sensor Networks

Routing Protocols

Due to the natural limitations imposed on sensor nodes, several network layer protocols have been proposed to utilize sensor’s energy to prolong the life time of deployed wireless sensor networks (WSNs). These protocols can be generally classified into four main categories: data-centric, hierarchical, location-based, and Quality of Service- (QoS-) based routing protocols. Extensive surveys on routing protocols can be found in [12–18].

3.1. Data-Centric Protocols. Data-centric routing protocols are negotiation-based and application-specific protocols. In data-centric routing, the sink node forwards queries to selected regions and waits for data replayed for the sensors in the specified regions. While data is being requested through queries, attribute-based naming should be utilized to identify the characteristics of required data. Several data-centric protocols have been proposed in the literature including Flooding and Gossiping [19], SPIN [20], DD [21], and Rumor [22].

Flooding and Gossiping are two classical algorithms to convey data in sensor networks without any routing protocols and topology maintenance. In Flooding algorithm, when a sensor receives a data packet, it broadcasts the data packets to all neighbors within its transmission range. Each sensor receiving the packet repeats the same process until the packet reaches its either destination or time to live. Even though Flooding algorithm is straightforward to implement, it has several disadvantages including implosion, overlap, and resource blindness. Implosion occurs when a node receives the same packet from more than one node. Gossiping algorithm overcomes the shortcoming of implosion by just making each receiving node forwards the data packet to a randomly chosen neighbor which in turn selects another random neighbor to forward the packet to and so on.

Sensor Protocols for Information via Negotiation (SPIN) is a family of adaptive routing protocols. SPIN is designed to improve classical mechanisms to overcome the problems of implosion and overlap. It gives a high-level name (called metadata) to completely describe the collected data, and also it uses three types of messages: advertisement (ADV), request (REQ), and DATA to convey data messages throughout the network. In SPIN protocols, when a node has a data to convey, it first sends an ADV message attached with a metadata to its neighbors for negotiation. Next, each node receiving the ADV message replays with a REQ message if it is interested in the data. Then, when the node that has the data receives the REQ messages, it starts to send the DATA message to each node sending a REQ. Finally, once a node receives the DATA message, it repeats the same sequence of steps to convey the DATA messages to all interested nodes all over the network. SPIN has the advantage that it provides more energy savings over the classical protocols. This is achieved through using metadata negotiation which almost reduces the redundant data halves. On the other hand, the data advertisement mechanism of SPIN cannot guarantee delivery of data. The implosion problem still exists during the REQ phase and the solution for the collision is not considered.

Directed Diffusion (DD) is a routing protocol for distributed activities where a large number of limited resource nodes locally coordinate to achieve a sensing task. The DD protocol has two main features: Interests and Gradients. The Interest is a named data or task description to define the sensor events that an originator is interested in, and it has a list of several attribute-value pairs including type, region, rate, duration, and time stamp. The type is used to impose what type of data should be sensed, and the region defines the network part from which the data should be drawn. The rate attribute sets how often the data should be forwarded, while the duration sets how long the interest should be active, and the time stamp is used to refresh the interest. Each
interest entry has several gradients, one to each neighbor. Each gradient also has a set of attributes including a node identity to which the data will be forwarded, data rate that tells how often the data will be forwarded, and the duration field which sets how long the gradient will be active. The DD protocol is completed by diffusing the interest, setting up gradients, and sending data and path reinforcement.

Rumor routing algorithm is a variation of DD, and it is characterized by events, queries, and agents. An event is a phenomenon occurring in a fixed region of the network. A query is a request for information or an order to collect data, while an agent is a long-lived packet used to create paths leading to events. Each node on the path contains both a list of its neighbors and an event table in which a row has an event name, number of hops to the event, and the next node in the path toward the event. In the rumor routing, any node can generate a query and then send the query in a random walk to find the path. The query keeps going on until it finds the path or its time to live (TTL) expires. If the node originated, the query discovers that the query is dead; without finding a path to the event, it can retransmit, stop, or flood the query.

3.2 Hierarchical Routing Protocols. Hierarchical routing protocols often group sensor nodes into clusters that form a hierarchy. Several hierarchical protocols have been introduced in the literature including LEACH [23], TEEN [24], and PEGASIS [25].

Low Energy Adaptive Clustering Hierarchy (LEACH) is a self-organizing, adaptive clustering protocol that utilizes randomization to balance the energy load among the deployed nodes in the network. LEACH assumes that all nodes start with uniform energy distribution, and all nodes can communicate directly with the base station. It works in two rounds, set-up and steady-state. The set-up round is responsible for forming clusters and cluster heads. During this round, sensors vote for themselves to be cluster heads at any given time with certain probability based on their energy. Next, the cluster head advertises their status to other sensors in the network. Accordingly, each sensor node decides the cluster that it wants to belong to by selecting the cluster head that needs the least amount of communication energy. Finally, once all the nodes are arranged into clusters, each cluster head creates a schedule for the nodes in its cluster to avoid collision. This results in the network structure that consists of a base station or sink, cluster heads that communicate with the sink, and sensor nodes each is recognized by a cluster and communicated with its cluster head that is a single hop to the sink.

The steady-state phase is concerned with transferring the data from the sensors in the network to the base station or sink node. During this phase, the cluster head collects the data from the sensors in its cluster. Once the data is collected from all nodes in the cluster, the cluster head locally aggregates the data in some way based on the application to remove the unreliable data. Next, the cluster head transmits the collected data to the base station. LEACH uses the local processing to reduce global communication and also randomizes the rotation of cluster heads. Therefore, it prolongs the network’s life time. On the other hand, it is not applicable to large networks and time-critical applications. In addition, the idea of dynamic clustering brings extra overhead. Furthermore, the cluster heads send data to the sink through high power link which make these clusters consume their energy faster.

Threshold-Sensitive Energy-Efficient Protocols (TEEN) are designed for time-critical applications in which a sudden change in the measured attribute (e.g., temperature) may happen. The TEEN networks employ multilevel clustering mechanism to avoid the transmission over high power link implemented in LEACH. In TEEN, once the clusters and cluster heads are formed, each cluster head broadcasts two main parameters: hard threshold (HT) and soft threshold (ST) to the sensor nodes in its clusters. Each sensor in the network has a variable called SV to store the sensed value. Also, each sensor node continuously measures the environment and if the measured value is above its HT for the first time, the node stores the value into SV, turns on its transmitter, and sends the measured value to its cluster head which in turn forwards it to another cluster head until it reaches the sink node. The node will next transmit the measured value in the current cluster period if two conditions are met: (1) the current measured value is greater than HT and (2) the current measured value varied from the SV by the value of ST. Using this technique, TEEN reduces more network traffic and extends the life time of the LEACH networks into two things: (1) it reduces the large energy consumed by transmission over high power link through utilizing multilevel clustering algorithm and (2) it reduces the network traffic through implementing the HT and ST.

Power Efficient Gathering in Sensor Information Systems (PEGASIS) implements a chain-based algorithm to eliminate the overhead resulted from dynamic formation of cluster in LEACH and TEEN. PEGASIS assumes that all nodes know the overall topology and implements greedy algorithm to form data chain. To form the data chain, a downstream node chooses the closest node as an upstream node from the unchained nodes. Once the data chain is formed, the last node of the chain is called the chain leader which begins to send data to the base station. Any node on the chain receives the data from its downstream, aggregates the data with its own in some way, and then sends the aggregated data to its upstream node. This process is repeated until the data reaches the chain leader which in turn sends it to the base station. The data chain is constructed once and followed by several rounds of data communication, but if a node on the chain is dead, the chain is reconstructed again in the same manner to avoid the dead node. Even though PEGASIS overcomes the overhead resulted from dynamic cluster formulation by constricting data chains, it has several drawbacks: (1) it assumes that each node can communicate directly with the BS, (2) the chain leader can become a bottleneck, and (3) it uses greedy algorithm which is locally efficient but globally inefficient.

3.3 Location-Based Routing Protocols. Location-based routing protocols are a family of routing protocols in which each deployed sensor node should know its local location information by some means, for instance, GPS [26]. In addition, it also may require that each sensor node is aware
of its remaining amount of energy. It uses this information as forwarding metrics during the forwarding of data packets from a source node to a distention node. This family of protocols include GEAR [27], Min-Hop [28], MAP [29], and PEW [30].

Geographic and Energy-Aware Routing (GEAR) is designed for routing queries to specific regions. It assumes that all sensor nodes are aware of their remaining energy and their locations by some means such as GPS. In addition, each node should know the location and remaining energy of its neighboring nodes simply by Hello flood messages. GEAR utilizes this information to construct a heuristic function that avoids energy holes and chooses sensors to route a packet toward the target region. This forwarding process is repeated until the packet is disseminated inside the target region.

Minimum hop (Min-Hop) routing protocol forwards packets from a source node to the sink node through an optimal path. The optimal path is the route that has the shortest path to the sink. The path is represented as the number of hops to the sink. Therefore, the source node chooses the node of the next hop as the one that has the shortest path or the minimum number of hops to the sink. If several nodes have the same number of hops to the sink, the one with the maximum energy is considered. This forwarding process is continued until the packet reaches either the sink node or its predefined maximum of hops to avoid packet loop. The Min-Hop works in two phases: initialization phase and routing phase. The initialization phase is concerned with building the routing tables of the network nodes, while the routing phase starts after initialization phase, and it is responsible for forwarding data packets through the network until they reach the sink node. Although the Min-Hop improves the average energy consumption in the network by using the shortest paths, it overutilizes the nodes along the shortest paths. This results in increasing energy gaps or holes and decreasing network life time.

Maximum Available Power (MAP) is an enhanced version of the Min-Hop to lengthen the network life time. The MAP distributes network load by choosing the neighbor node with the maximum energy, but this leads to significantly longer paths. Both Min-Hop and MAP forward data packets based on local view of neighbors and thus may result in energy holes that lead to decreasing the network life time. In other words, selected paths may include one or more nodes with very low energy.

Path Energy Weight (PEW) overcomes this shortcoming by deploying a global view mechanism. Its main idea is to globally map energy levels of all nodes along the communication path into a single parameter. This parameter reflects how uniform the energy distribution along the path is. It gives a better weight to a path with balanced energy level over a path with unbalanced energy level.

3.4. QoS-Based Routing Protocols. Wireless Multimedia Sensor Networks (WMSNs) [31, 32] are a new powerful class of sensor-based distributed intelligent systems capable of ubiquitously retrieving multimedia information derived by the recent technological advancements in microelectromechanical systems. WMSNs consist of sensor nodes equipped with audio and visual information collection modules that have the ability to retrieve multimedia data, store or process data in real time, correlate and fuse multimedia data originated from heterogeneous sources, and wirelessly transmit collected data to the desired destinations.

In the QoS-based routing protocols, routing process of selecting the next forwarding node is performed based on Quality of Service (QoS) parameters such as bounded latency or delay, bandwidth, jitter, and reliability. However, satisfying the stringent QoS requirements of multimedia transmission in a resource-constrained sensor network environment places new challenges to routing. Hence, optimal energy and application-specific QoS-aware routing for WMSNs have gained considerable research attention recently. The authors in [33] claim that classic multipath routing approaches are vulnerable to black holes attacks, mainly due to their deterministic nature. They also proposed mechanisms that generate randomized multipath routes. Besides routes randomization, the generated routes are also highly dispersive and energy efficient, making them quite capable of circumventing black holes.

The authors in [34] focus on the selective jamming attacks in WSNs. To face these attacks, they developed three schemes that prevent real-time packet classification by combining cryptographic primitives with physical layer attributes.

In [35], the authors propose a secure node disjoint multipath routing protocol in which the data packets are transmitted in a secure manner by using the digital signature cryptosystem. Their proposed protocol can improve the packet delivery and reduce the end-to-end delay.

The authors in [36] formulated the secret-sharing-based multipath routing problem as an optimization problem aiming at maximizing both network security and life time, subject to the energy constraints. They also propose a three-phase disjoint routing scheme called the Security and Energy-efficient Disjoint Route (SEDR). Based on the secret-sharing algorithm, the SEDR scheme can depressively and randomly deliver shares all over the network in the first two phases and then transmits these shares to the sink node. The proposed scheme improves the network security under both scenarios of single and multiple black holes without reducing the network's life time.


Due to the limited resources of the sensor nodes in wireless sensor networks, most of the traditional security mechanisms employed in traditional wireless networks are not applicable for wireless sensor networks (WSNs). Thus, applying a security mechanism into a WSN is a challenging task. The rest of this section introduces the main security issues associated with WSNs. These main issues are security requirements, attacks in WSNs, and key managements [37, 38].

4.1. Security Requirement in WSNs. To accomplish security in WSNs, the basic security requirements or objectives should be applied to face passive attacks, active attacks, and Denial-of-service (DoS) attacks [39, 40]. Passive attacks are silent in nature and do not affect the normal network operations;
Therefore they are very difficult to detect. However, passive attacks lay down a foundation for later launching an active attack. Integrity and availability are compromised by active attacks, while the confidentiality of end-users’ traffic is compromised by passive attacks. The confidentiality is the process of hiding messages so that unauthorized entity cannot reveal them. The integrity is the process of ensuring that messages are not altered by unauthorized entity. The availability is the process of ensuring that network services are available for authorized entities when they are needed. In wireless networks, these requirements are independent of application due to the data aggregation at the intermediate nodes. However, some systems may satisfy some of the requirements based on the application [41]. The DoS attack will be discussed in details in the following subsection.

4.2. Attacks on Routing Protocols of WSNs. Wireless sensor networks are subject to several types of attacks which can be categorized into attacks on information in transit, node replication attack, Denial-of-service, and routing attacks. Since the routing protocols are the scope of this paper, the first three categories are discussed briefly hereafter, and the rest of section is devoted for routing attacks.

Attacks on information in transit are the most common attacks against WSNs in which the information in transit is vulnerable to eavesdropping, modification, injection, interruption, and traffic analysis. Most of these attacks can be thwarted by implementing confidentiality, authentication, and integrity. In a node replication attack, an attacker can insert a new node into a network which has been cloned from an existing node. This new node can act exactly like the old node, or it can have some extra behavior, such as transmitting interested information directly to the attacker. The Denial-of-Service (DoS) attacks can take place at different layers. At the physical layer, one or more attackers continuously transmit a radio signal that interferes with the radio frequency used by the sensors. This jamming can render the network sensors to be ineffective. Also, DoS attack can occur at the data link layer by violating the communication protocol. For example, a malicious node continuously transmits messages to generate collision or to deplete the power of the target nodes [42]. DoS attacks at network layer involve attacking the routing protocols which is discussed in the rest of this section.

A large number of the routing protocols of wireless sensor networks are designed without security in mind. Consequently, these protocols are subject to several attacks associated with the network layer of ad hoc networks. Most of these attacks fall into one of the following categories: acknowledgment spoofing attack [42], false routing information attack [43], wormhole attack [44], selective forwarding attack [45], sinkhole attack [46], Hello flood attack [46], and Sybil attack [47].

In the acknowledgement spoofing attack, an adversary may forge on acknowledgement aiming to influence that a weak link is strong or a dead node is alive. As a result, a weak link may be chosen for routing, and hence packets delivered through that link may be lost or corrupted. An attacker using acknowledgment spoofing attack can effectively mount a selective forwarding attack by encouraging the target node to send out packets through those weak links.

In the false routing information attack, an attacker can reshape the whole network by conveying false routing information to the other nodes in the network. As a result, the attacker can create routing loops, attract or repel network traffic from specific nodes, extend or shorten source routes, generate fake error messages, cause network partitions, increase end-to-end latency, and reduce the network life time.

In the wormhole attack, two adversaries cooperated to tunnel messages received from sensors at a network location over a low latency link and then replay them to sensors at a different location. The wormhole attack is dangerous for WSN routing protocols since attackers can achieve it without compromising any sensor node in the network; even all of the sensor nodes in the network utilize effective authentication and confidentiality mechanism.

In selective forwarding, malicious nodes may not forward specific messages and simply drop them. A simple aspect of the selective forwarding called sometimes grey-hole attack is when a malicious node acts like a black hole by refusing to forward every observable packet. The selective forwarding attack is most effective when the attacker’s node is able to include himself on the forwarding path of target messages.

In the sinkhole attack, one or more malicious nodes prevent the base station from obtaining complete and correct sensing data, thus forming a serious threat to higher-layer applications. The sinkhole attack is achieved by making a compromised node look attractive to its neighbor nodes with respect to the routing metrics. Consequently, the attacker manages to draw as much traffic as possible that is designated to the base station. By involving himself in the routing process, the attacker is then able to launch more severe attacks such as selective forwarding, modifying or dropping the received packets.

In the Hello flood attack, an adversary exploits the behavior of most routing protocols requiring each node to broadcast a Hello message to discover its neighbor and automatically create a network. Nodes receiving this message mark the sending node as their parent. As a result, each of these nodes will forward the packets to its parent. In the Hello flood attack, an attacker can use a wireless device with a large enough transmission power for compromising all nodes in the network that this device is its parent. For example, if an attacker broadcasts a Hello message with a large enough transmission power, the malicious node will convince each node in the network that the adversary is its neighbor and marks the adversary as its parent. This makes all nodes send their messages to the adversary, but because they are far away from this adversary, they send them to oblivion.

In the Sybil attack, a malicious node unlawfully claims several identities to represent a large number of other sensor nodes in the network. In this attack, each of these identities is called a Sybil node which can be obtained be an attacker in one of two ways. In the first way, the attacker can simply generate a new Sybil identity from the address space to each Sybil node. For example, if a node is identified by a 16-bit integer, the attacker can simply assign a random 16-bit integer for each Sybil node. In the second way, the
attacker can steal or spoof a legitimate identity for each Sybil node. The way is most likely used when the address space is intentionally limited to prevent attackers from injecting any new identity. Once each of the Sybil nodes is assigned an identity, it can communicate with legitimate nodes directly or indirectly through malicious nodes. Consequently, the Sybil attack imposes a significant impact on the routing process of the underlying network.

4.3. Key Management in WSNs. The wireless sensor networks (WSNs) have a set of security requirements that must be achieved to protect the networks against most of the associated attacks. These security requirements are confidentiality, integrity, availability, authentication, and refreshment. To provide these requirements, a key management mechanism suitable for WSN must be implemented. The key management in WSN is a set of key distribution mechanisms; each mechanism is responsible for establishing cryptographic key or key material among all sensors nodes in the network. In addition, it is concerned with revoking and refreshing keys. A good key distribution mechanism should have the following features: scalability, efficiency, connectivity, and resilience.

The scalability means that the key distribution mechanism should support large networks and be flexible against substantial increase in network size even after deployment. Efficient mechanism should consider sensors limitations such as storage, processing, and communication. Key connectivity is the probability that two or more nodes share the same key. Enough key connectivity must be provided for a WSN in order to perform its task. Resilience means that the algorithm is resisting against node capture. For example, compromising the security credentials stored on a sensor node should not reveal information about the secrecy of any other links in the WSN. These features conflict with each other, and thus a tradeoff should be considered. Accordingly, researchers developed several key distribution mechanisms including [48–52].

5. Secure Routing Protocols

Even though the above mentioned routing protocols utilize the limited capabilities of sensor nodes, they have not been designed with a security goal in mind. Consequently, they are not applicable into adversarial environment, such as military systems and disaster relief, due to their susceptibility to a great number of attacks against routing protocols. These attacks include selective forwarding, sinkhole, wormhole, and Sybil described in Section 4. To protect WSNs against these routing attacks, a number of secure routing protocols have been proposed in the literature including SPINS [53], SSPIN [54], ITSRP [55], COOL [56], EESRP [57], DSSRP [58], TTSS [59], SRPMND [60], and STAPLE [61].

SPINS enabled security features through two security protocols: SNEP and μTESLA. SNEP provides confidentiality, integrity, authentication, and freshness, while μTESLA [62] provides authenticated broadcast. SPINS gives more attention on key management.

Secure SPIN (SSPIN) is a secure version of SPIN family discussed in Section 3. It uses Message Authentication Code (MAC) to provide integrity and packet correctness, but it did not clearly specify how the confidentiality is achieved. Like SPIN, SSPIN works in three stages: ADV, REQ, and DATA. When a node has new data, it creates ADV message with its MAC and then broadcasts the ADV to its neighbors. Once a node obtains the ADV, it verifies the ADV through its MAC. If it is a valid ADV, the receiver sends the REQ message to a source which in turn will verify the REQ through the associated MAC. If the REQ is valid, the DATA stage starts. The source generates the MAC, attaches it to the data packet, and then sends the packet to the destination. The authors prove that SSPIN is secure if MAC scheme is secure against existential forgery attack.

ITSRP (Intrusion-Tolerant Secure Routing Protocol) implements mechanisms for authenticated key exchange and energy factor as well as resistance to some types of attacks such as sinkhole and wormhole attacks. The authors of ITSRP developed Distributed key management scheme running at the sink. It is responsible for distributed and session key initialization. ITSRP uses cryptography to provide authentication and data secrecy.

COmpromised nOde Locator (COOL) is an authentication scheme proposed by Zhang et al. [56]; the main idea of this protocol is to detect and locate compromised nodes once they misbehave in the network based on the observation of well-behaved sensor nodes.

Energy-Efficient Secure Routing Protocol (EESRP) is designed to provide both security and energy efficiency through developing two protocols: Roulette-Wheel Routing Protocol (RWRP) and Secure Routing Protocol (SRP). The RWRP is developed to forward data packets from a source to the sink node. The forwarding decision of each node is independent from other nodes (i.e., the node does not collect information from other nodes to make the decision). Thus, the node decision cannot be deceived by other nodes. SRP is interested in securing data during its traveling from a source to the sink. It implements μTESLA to authenticate packets from the sink. In addition, it used shared keys, Message Authentication Code (MAC), and time stamp to provide confidentiality, authentication and integrity, and refreshment of packets, respectively.

Dual Sink Secure Routing Protocol (DSSRP) is an enhanced version from the EESRP to prolong the network life time by using two sinks. Like the EESRP, DSSRP provides its functionality through two protocols: Next Node Selection Protocol (NNSP) and Network Protection Protocol (NPP). NNSP and NPP are a modification version of RWRP and SRP to adapt the operation for two sinks, respectively.

TTSS (The Three-Tier Security Scheme) is a secure routing protocol designed to eliminate mobile sink replication attacks associated with sensor networks that have mobile sinks. It differentiates between three types of nodes: mobile sinks, stationary access nodes, and sensor nodes. The stationary access nodes are a small number of nodes chosen randomly from the network nodes, and they are used as access points to mobile sinks in order to collect data from sensor nodes. Also, TTSS uses two separate pools of polynomial keys: the mobile polynomial pool MP of size M and the static polynomial pool SP of size S. TTSS works in two phases: static
and mobile polynomial predistribution and key discovery between mobile node and stationary node. In the first phase which was executed before deployment, all mobile sinks and stationary access nodes are randomly given as $k_m$ and one polynomial ($k_m > 1$) from MP such that the number of polynomials in each mobile sink is greater than the number of polynomial in stationary access nodes. This is to assure that a mobile sink shares with high probability a mobile polynomial with a stationary access node and to reduce the number of compromised mobile polynomials in case a stationary access node is compromised. Also, all stationary access nodes and sensor nodes randomly pick a subset of $k_l$ from SP. In the second phase, when a sensor node establishes a direct pairwise key with a mobile sink, it has to find a stationary access node in its neighborhood that share a pairwise polynomial with the mobile sink. Using two polynomials makes the authentication of mobile sinks independent of the key distribution of sensor nodes.

SRPMDND stands for Secure Routing Protocol with Malicious Node Detecting and Diagnosing for Wireless Sensor Networks. It uses $\mu$TESLA authentication protocol to protect packets sink node against the modification, forging, and replay. Also, the protocol implements an acknowledgment mechanism to detect the malicious nodes. For example, if the node did not hear an acknowledgment within a specific period of time, this means that the packet is forwarded to a malicious node. Therefore, a secure route can be created when each node on the path forwards its packet and waits for an acknowledgement until the packet reaches the sink.

STAPLE is a secure routing and aggregation protocol with low energy cost for sensor networks. It applies one-way hash chain and multipath routing to gain the security and uses hash-based message authenticated code to achieve authentication and data integrity. In addition, it uses symmetric cryptography to provide data secrecy. STAPLE works in three stages: initialization, transmission, and source authentication. The first stage is responsible for both assigning the nodes in different levels according to the minimum hop from the sink node and distributing keys. The second stage is concerned with forwarding packets from child node to parent and so on until they reach the sink. During this stage, STAPLE achieves child authentication and data integrity authentication as well as detection of false packets. The last stage authenticates the source and verifies integrity.

In addition, the authors in [63] propose a security approach that uses secret key cryptography and key management along with rekeying support. The proposed protocol provides backward secrecy, privacy, data integrity, and secure management to a wireless sensor network. It is also capable of identifying different attacks such as replay attack, DoS attack, and Sybil attack with low-resource requirements. Khan et al. [64] proposed a robust and secure routing protocol for an infrastructure-based wireless mesh (SRPM) networks. The proposed protocol performs well against a variety of multi hop threats tested over a range of networks scenarios. The authors in [65] propose a cross-layer design to improve the overall performance of multi hop wireless networks. The proposed architecture provides different parameters at different layers that can be utilized in protocol design phase. The authors in [66] presented cross-layer secure and resource-aware on demand routing (CSROR) protocols for hybrid WMN which is designed to ensure routing security and provide different applications specific requirements for multimedia delivery and real-time transmissions. CSROR selects an optimum route on the basis of route security taking into consideration the different cross-layer parameters.

6. Recent Research Issues in WSN Routing Protocols

Due to the continual proliferation of WSN applications, especially, QoS crucial applications, the routing protocols for these networks have introduced a lot of challenges that need more study and exploration. These research challenges or issues include energy efficiency and QoS guarantee, network dynamics, multiple sources and sinks, dynamic holes bypassing, and secure routing [16]. Due to the importance of these issues, the rest of this section is devoted to discuss them for future study.

6.1. Energy Efficiency and QoS Guarantee. Utilizing nodes’ energy is an essential factor in designing routing protocols of WSNs to prolong the network life time of traditional WSNs. However, employing such energy-efficient routing protocols in WMSN may result in energy holes due to the large amount of data transferred in such networks. Figure 2 describes this issue in which Figure 2(a) explores the energy-efficient path from node A to the sink node. However, overutilizing this efficient path may result in energy holes along the path as shown in Figure 2(b). A possible solution to overcome the energy hole problem is to explore multipath techniques that satisfy both energy efficiency and QoS requirements between the node A and the sink node as depicted in Figure 2(c).

Multipath routing can be interpreted in two different ways. First, it can be envisioned as a multipath exploration while employing a single path randomly at a time for data transmission. The objective of the random selection of a path is to evenly distribute the energy consumption among the explored paths. In this case, multipath routing can improve the reliability and strengthen the security by avoiding the failed or compromised paths. Second, multipath routing can also be achieved by exploring multipaths and employing them for carrying the distributed data along the explored paths simultaneously. In addition to the advantages of the first scenario, the second one will reduce the end-to-end delay and facilitate the detection and prevention of selective forwarding and sinkhole attacks. However, data segmentation at the source node and aggregation at the sink node will introduce an overhead. Multipath routing providing QoS guarantee requires further studies for investigating the tradeoff between the energy efficiency and the QoS parameters.

6.2. Network Dynamics. Most of the current routing algorithms assume that network sensor nodes are static. However, they may be highly movable, and latest advances show that mobile sensors in WSNs have a promising performance [16]. Also, the recent research on data collection reveals that
reporting data through implementing and leveraging sink mobility is promising for energy-efficient data gathering than reporting data through long, multi-hop, and error-prone paths to a static sink in either tree or cluster network structure [67]. Accordingly, Liu et al. [68] proposed a data reporting protocol that implements a mobile sink. Their protocol is called SinkTrail in which a mobile sink moves around the network field with relatively low velocity and keeps listening to data report packets. The mobile sink stops at some places called "trail points" for a short amount of time. Next, it broadcasts a message called "trail message" to the whole network and simultaneously listens to data report packets. Finally, it moves on to another trail point as shown in Figure 3.

The distance between any two consecutive trail points is assumed to be the same by the authors in order to facilitate the tracking of a mobile sink. The authors showed that the networks which implemented sink mobility have a great performance over networks that deployed static sink. The authors also revealed that the movement of the sink in a circular pattern outperforms each of random, angular, and linear patterns.

Even though the recent research revealed that leveraging mobility into the WSNs has a great or promising performance, the sink mobility introduced several issues that require more consideration study. The first of these issues is that the sink mobility presented the sink replication attack in which a malicious sink claims that it is the intended one. This may reveal confidential data for the public or prevent it from reaching the authorized sink. Therefore, the researcher should consider this attack along with the other associated attacks when designing secure routing protocols implementing sink mobility. The second issue is the minimization of the...
overhead resulting from the frequent updates of new location information. The third issue is what is the optimal number of mobile sinks as a function of sensor nodes deployed into the network field? The last issue, but not least, is the optimization of moving trajectory for a mobile sink since it is an NP-hard problem [69]. This issue also includes finding out the proper moving trajectory pattern such as random, circular, or triangular as seen in Figures 4(a), 4(b), and 4(c), respectively.

In addition, we need to study the performance when dynamic nodes together with a static sink or mobile sink are deployed in a network field as depicted in Figures 4(d) and 4(e), respectively. Furthermore, we need to explore the performance related to a hybrid combination of static and mobile sinks in environments with static and dynamic regular nodes as shown in Figure 4(f). Consequently, supporting mobility into routing protocols can be shown as an interesting area for future studies.

6.3. Multiple Sources and Sinks. Most of the current routing protocols forward data from a single source to a single sink. When a source node has data to share with the sink, it forwards the data to the sink node over an explored path between them. This may have several impacts on the deployed network. The network life time may be reduced due to consuming a lot of nodes’ energy over long and multi-hop routes. Next, the collision may be increased when several sources send their data to the sink due to the fact that many paths from different sources to the sink have overlapping links. In addition, the nodes in the neighborhood of the sink node consume their energy faster because they are responsible for forwarding their own data along with the received data from other nodes. Consuming the energy of the neighboring nodes faster may result in a network partition in which the sink node is in one partition and the other nodes are in another partition as visualized in Figure 5 by the authors of [30].

The network partition renders the network ineffective and reduces the network life time. To clarify the point, Figure 5 visualizes the energy of the network nodes deploying either the Min-Hop or the PEW routing protocols. The network model used by any of these two protocols is depicted in Figure 5(a). It has a single static sink at (0, 0) in xy-coordinates, and it has also 300 nodes that are distributed randomly all over the network field. Figure 5(b) surrounds the neighbor nodes of the sink by a dashed circle. The energy of these nodes is consumed faster than the other nodes in the network as visualized in Figures 5(c) and 5(d) for the Min-Hop and PEW protocols, respectively.

Fortunately, the aforementioned problems resulted from deploying only a single sink seen in Figure 6(a) which can
be elevated by implementing multisinks. For example, the DSSRP protocol [58] implements two static sinks and reveals a promising performance over a single static sink. Also, networks supporting several sinks may be entailed to accept several events simultaneously.

In addition, these network configurations increase the reliability and improve the network life time. However, leveraging the multisinks into the networks introduces new issues including the number of sinks, deploying mobile or static sinks, and hybrid sinks. Because the number of sinks has a great influence on the network performance, it should be studied carefully to find out the optimal number of sinks as a function of the network size. It is also required to study whether to (1) implement only static sinks as in Figure 6(b) which has four static sinks, as an example, distributed at the corners of the network field, (2) deploy only mobile nodes as depicted in Figure 6(c) which has four mobile sinks that move in a circular patterns as an example, or (3) leverage or combine both static and mobile sinks into the network field as shown in Figure 6(d) which has four static sinks and one mobile sink. As an example in this aspect of mobility, Suganthy et al. [70] deployed the multisinks technique in which they use three mobile sinks along with a static sink. They confine one mobile sink to circular pattern mobility, the second one to rectangular pattern mobility, and the other one to constant speed mobility. They measure the remaining energy of the deployed nodes without sink mobility and with sink mobility. The authors revealed that the remaining energy of the nodes with sink mobility is better than the remaining energy of nodes without sink mobility. This extends the life time of the deployed network. However, forwarding metrics for moving the mobile sinks and optimal number of sinks need more explorations and studies. Furthermore, the securely underlying scenarios should be considered. Thus, implementing multisinks into network fields can be considered as a new area of exploration.

6.4. Dynamic Holes Bypassing. Dynamic holes occur due to overutilizing optimal paths. Figure 7(a) reveals that when a source node A always sends its data over an optimal path to the sink, it may result in energy holes as depicted in Figure 7(b). An energy hole or energy gap is a region in the network field that is not covered with radio signals. The increasing number of energy gaps reduces network life time and may render network ineffective. Therefore, the rest of this section presents possible bypassing solutions that elevate the problems resulted from energy holes.

Most of the current routing protocols such as [19, 20, 23–25, 28–30, 58] bypass energy holes by exploring new path to the sink node as shown in Figure 8(a), but it may result in an extra delay.

This solution works fine when a source node (e.g., node A) has a connection with the sink node but it fails when a source node is disconnected from the sink node. For example, the nodes surrounded by a dash,ed circle form a network partition, and they cannot deliver their data to the sink node.
To overcome this shortcoming, the network dynamic should be implemented into networks. The network dynamic can be introduced into the network by implementing one of the following three aspects:

1. Small set of mobile nodes with a static sink,
2. Static nodes with mobile sink,
3. Small set of mobile nodes, mobile sink, and static sink.

Implementing a small set of mobile nodes into a network field can cover energy hole regions with radio signals by moving one or more mobile nodes towards these regions as depicted in Figure 8(b). This solution introduces new issues that need a further consideration and study. These issues are summarized as follows: which mobile nodes should be moved towards energy hole regions, when they should be moved, and where exactly they should be moved. For example, which of the nodes B, C, or D in Figure 8(b) will be moved towards the indicated region? One solution to these issues can be achieved by giving a weight to each mobile node. The weight should be a function of the distance between the mobile node and the center of the region to be covered. Also, the weight should consider the priority of each energy gap region. In addition, it should have a concern about the current area covered by the mobile node.

The second aspect is to implement a mobile sink with static nodes. The mobile sink will move around to collect data from all sensors in the network field as seen in Figure 8(c) in which the sink goes into the region to collect the data. This will bypass the energy hole but introduces an extra overhead due to the updates concerning new location information. In addition, the mobile sink movement pattern should be defined as discussed in the network dynamics section.

The last aspect is to combine the static sink, mobile sink, mobile nodes, and static nodes into network field as shown in Figure 8(d) as an example. In this aspect, a source node has more choices to bypass an energy hole varying from exploring new route to adjust topology changes required to cover the holes. However, the performance- and security-related issues resulted from combining these different technologies need extra studies and exploration. In addition, more studies are needed to know how to adjust topology changes resulted from holes.

6.5. Secure Routing. WSNs are used in several applications such as homeland security, military, and healthcare in which the security is essential. Such networks are subject to several routing attacks which include sinkhole attack, selective forwarding attacks, wormhole attack, Hello flood attack, Sybil
A Sensor node
Base station
Low bandwidth links
(a) Overutilization of energy efficiency

Sensor node
Base station
Low bandwidth links
(b) Energy hole

Figure 7: Energy efficiency and energy holes.

Sensor node
Dynamic sink
Low bandwidth links
(c) Energy hole bypassing through deploying a mobile sink

Sensor node
Dynamic sink
Low bandwidth links
(d) Energy hole bypassing through hybrid mobility of both sink and regular sensors

Figure 8: Different techniques for energy holes bypassing.
attack, Denial-of-service (DoS) attack, and a sink replication attack associated with mobile sinks. Figure 9 depicts different types of routing attacks.

Figure 9(a) shows a sinkhole attack scenario that is launched when an adversary puts himself on the forwarding path. Once the adversary is on the forwarding path, he will refuse to forward any received packets. Figure 9(b) depicts the selective forwarding attack in which the attacker puts himself on the forwarding path and then drops certain packets and prevents them from reaching the sink node. Figure 9(c) visualizes the wormhole attack launched with two adversaries that share the low latency link denoted by a dashed line. One adversary attracts nodes in its neighborhood to send their data to the adversary. When the adversary receives data, he will forward the received data to the other adversary which in turn conveys it to the nodes in its region. Figure 9(d) depicts the Hello flood attack which is launched when an attacker, equipped with a laptop that has a high-quality radio signal, sends Hello messages to the network field. Once a node receives a Hello message, it marks the attacker as its parent. When the node has data that it wants to share, it sends it to the attacker, but because the attacker is far away, the node sends its data into oblivion [46, 71].

Another type of WSN attack on routing protocols is Sybil attack which reduces the fault tolerance, topology maintenance, and resource utilization and weakens the routing mechanism. In this type of attack, a node steals the identities of many nodes to pretend to be as them to degrade the data integrity and security of the network. But in reality, it is only a single malicious node injected cleverly into the specific sensor network to alter valuable information [47, 71].

The DoS attack in WSNs can occur intentionally by attackers or unintentionally due to unexpected node failure. The unexpected node failure may result from various software bugs, exhausted power supply system, environmental disaster, and complication in data transmission and communication, or even intentional intruder attack may execute DoS attack. Often, attackers try to weaken or destroy a network or cause an interruption in secure data communication by sending loads of unnecessary data packets to the victim nodes and therefore exhibit DoS attack. Different types of DoS attacks may take place at different network layers. At physical layer, it may cause jamming and tampering, at data link layer, it causes exhaustion and data collision, at network layer, it causes misdirection and negligence of data, and at transport layer, it could perform data flooding and malicious attack [71].
Finally, the sink replication attack is associated with sink mobility. It is achieved when a malicious sink attracts sensor nodes to forward their data to it. The aforementioned attacks may render the network inefficient. Therefore, several secure routing protocols such as [53–59, 72] have been proposed to defend against these attacks. However, new applications and new advance in the technology may need new security algorithms to protect networks against the current and new attacks that may be introduced. Thus, the security is an endless area of research, and integrating security into routing techniques of WSNs is an important area of current and future research.

7. Future Research Direction

Based on the recent trends discussed in Section 6, routing protocols that implement network dynamics initially contribute in solving energy holes bypassing problems and have a promising performance over routing protocols that implement static networks. Also, the routing protocols that use multisink outperform the routing protocols that deploy only a single sink. However, implementing these technologies into routing protocols introduces security holes. Therefore, the future research directions on WSNs routing protocols should focus on implementing multisink and nodes mobility along with satisfying security and QoS requirements into the new routing protocols. These routing protocols will help to protect their networks and prolong their life time.

Acknowledgment

This work is supported by the Deanship of Scientific Research, Taibah University, KSA under the Grant no. 434/4301.

References


