energies

MDPI

Article

EGCIR: Energy-Aware Graph Clustering and Intelligent Routing Using Supervised System in Wireless Sensor Networks

Tanzila Saba [1], Khalid Haseeb [2], Ikram Ud Din [3,*], Ahmad Almogren [4,*], Ayman Altameem [5] and Suliman Mohamed Fati [1]

[1] Artificial Intelligence & Data Analytics Lab (AIDA), College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh 11586, Saudi Arabia; drstanzila@gmail.com (T.S.); smfati@yahoo.com (S.M.F.)

[2] Department of Computer Science, Islamia College University, Peshawar 25000, Pakistan; khalid.haseeb@icp.edu.pk

[3] Department of Information Technology, The University of Haripur, Haripur 22620, Pakistan

[4] Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia

[5] Department of Natural and Engineering Sciences, College of Applied Studies and Community Services, King Saud University, Riyadh 11543, Saudi Arabia; aaltameem@ksu.edu.sa

* Correspondence: ikramuddin205@yahoo.com (I.U.D.); ahalmogren@ksu.edu.sa (A.A.)

Received: 25 June 2020; Accepted: 4 August 2020; Published: 6 August 2020

check for updates

Abstract: In recent times, the field of wireless sensor networks (WSNs) has attained a growing popularity in observing the environment due to its dynamic factors. Sensor data are gathered and forwarded to the base station (BS) through a wireless transmission medium. The data from the BS is further distributed to end-users using the Internet for their post analysis and operations. However, all sensors except the BS have limited constraints in terms of memory, energy and computational resources that degrade the network performance concerning the network lifetime and trustworthy routing. Therefore, improving energy efficiency with reliable and secure transmissions is a valuable debate among researchers for critical applications based on low-powered sensor nodes. In addition, security plays a significant cause to achieve responsible communications among sensors due to their unfixed and variable infrastructures. Keeping in view the above-mentioned issues, this paper presents an energy-aware graph clustering and intelligent routing (EGCIR) using a supervised system for WSNs to balance the energy consumption and load distribution. Moreover, a secure and efficient key distribution in a hierarchy-based mechanism is adopted by the proposed solution to improve the network efficacy in terms of routes and links integrity. The experimental results demonstrated that the EGCIR protocol enhances the network throughput by an average of 14%, packet drop ratio by an average of 50%, energy consumption by an average of 13%, data latency by an average of 30.2% and data breaches by an average of 37.5% than other state-of-the-art protocols.

Keywords: energy efficiency; graph clustering; key distribution; link security; wireless sensor networks

1. Introduction

In many critical applications, tiny wireless objects, known as sensor nodes, have been exploited to improve the network coverage and communications [1–4]. Sensor nodes are operated in dynamic, ad-hoc and self-configured mode to gather environmental data. Due to limited energy, transmission and memory constraints, the field of WSN is to make them different from traditional networks. Among all the constraints, the energy consumption gained much research interest in improving the performance of networks. Sensor nodes transmit their data towards the BS using an appropriate proxy node, also called a gateway or cluster head [5–8]. Due to its extra burden, the cluster heads consume energy resources more rapidly than other nodes. Such a problem mostly creates a routing hole in the network regions and results in packet drop ratios with network latency. Normally, the data routing of static sensors in the comparison of mobile sensors is most suitable for the small-sized region due to predefined data forwarding paths. While in the scenario of a large region, mobile nodes are preferred due to their mobility, scalability and coverage characteristics [9–12]. In recent years, the technology of machine learning improved the monitoring of targeting fields and learned the data routing due to their intelligent and lightweight processes. The area of machine learning makes the environment for low-powered sensor nodes more robust, cost-effective and energy-efficient.

The hierarchical schemes divide similar nodes based on certain conditions into various clusters [13–16]. A cluster is a particular region and nodes are bounded to communications with BS through cluster heads. The clusters can be achieved using either one of the two approaches, i.e., top–down and bottom–up. Within each cluster, one node is assigned a primary duty for data aggregation and forwarding, such node is called cluster head [17,18]. However, the energy of cluster heads is consumed most swiftly due to their computational, processing and transmission powers in the data routing. Thus, clustering nodes and uniformly distribute the load among cluster heads are major research problems for network efficacy [19,20]. Along with securing the sensors' data against attackers, the efficient generation and distribution of keys for transmission sessions are other important concerns [21–24].

This paper presents an energy-aware graph clustering and intelligent routing protocol using a supervised system to address the restrictions of existing energy-efficient routing solutions. The EGCIR protocol exploits a centralized cluster formation with the selection of energy-efficient and robust cluster heads, which results in minor network overhead and communication cost. The EGCIR protocol makes use of a supervised machine learning algorithm to share the sensors into various clusters and make the network smarter. Furthermore, it avoids the consumption of additional energy of sensors due to forwarding of control and route request messages. Moreover, it introduces fault-tolerant routing paths between cluster heads using a graph-based technique and achieves an optimized network lifetime. In addition, it offers a secure mechanism to obtain trustworthiness and authentic communications between the intracluster and intercluster transmission. The encryption keys are generated and distributed in the hierarchy from the BS to cluster heads and from cluster heads to sensor nodes. The malicious entities are adversaries and they aim to prevent sensor nodes to attain network services reliably. The proposed solution also excludes such malicious nodes from the existing routes and makes the transmission links more intellectual using security methods. The simulation experiments demonstrate improved outcomes of the EGCIR protocol in the comparison of existing solutions. The main contributions of this paper can be summarized as follows:

- A review of the literature on various secure and efficient routing protocols for low-power WSNs;
- Proposal of an energy-aware graph clustering with fault-tolerant routing using a supervised system for improving the network lifetime with secure data transmission;
- The improvement of network performance by the EGCIR protocol for different metrics in the comparison of other state-of-the-art solutions.

The rest of the paper is structured as follows: The related work is presented in Section 2; Section 3 highlights the problem background; Section 4 explains the methodology of the EGCIR protocol; Simulation setup and network parameters are presented in Section 5; Section 6 discusses the experimental results and their analysis in a detailed manner; and Section 7 presents the conclusion with suggestions for the future.

2. Related Work

The next-generation sensor network [25,26] is one of the most promising research topics because of its cost, size and robustness for real-time applications, i.e., healthcare, military, agriculture, etc. In such a network, smart sensors are dispersed either in a predefined location or in a random manner. The main aim of such sensors is observing the targeting object and sending the information toward network users for necessary actions. However, smart sensors have very limited capabilities and with time, they degrade the network performance especially in terms of energy and computational power [27–29]. Furthermore, due to the unpredictable behavior of sensors and some external sources, the conditions of WSNs are changed promptly. Therefore, since the last two decades, many researchers have been trying to propose different solutions for optimizing network performance, authenticity and balancing the network load between large numbers of low-powered sensor nodes.

In [30], the authors proposed a topology adaptive spatial clustering (TASC) approach, which aims to assign a weight within a locality of the node. The weighted value is based on some factors, i.e., distance, connectivity and density information. Initially, all nodes determine their weighted value and flood the information to their 2-hop-away neighbors. Based on the highest weighted value, appropriate nodes are selected as cluster heads. The selected cluster heads announce their status in the 2-hop-neighbor and the nodes are grouped towards a particular cluster head.

Authors in [31] proposed clustering communication based on a number of neighbors (CNN) for WSNs. The proposed solution is based on determining the number of neighbors, cluster head selection, clusters formation and TDMA scheduling phases. The cluster heads are selected using a random number and each cluster head adjusts their transmission range based on its number of neighbors. In [32], the authors proposed a solution to improve the localization scheme by using k-means and fuzzy-c means algorithms. By these two algorithms, the sink node trains the overall system. All network fields are divided into various clusters based on the Received Signal Strength Interference (RSSI) value. Moreover, all formulated clusters are trained individually to determine the coordinates of sensor nodes.

The zone-based energy-efficient routing protocol (ZEEP) is presented in [33] by exploiting a fuzzy system for mobile WSNs. The proposed solution makes use of a fuzzy inference system based on four factors, i.e., energy, distance, density and mobility, to improve data routing. In addition, the proposed solution balances the energy consumption in the network field by selecting an optimal number of cluster heads. The authors in [34] proposed advanced caching for distributing sensor data through programmable nodes, which aims to improve the download latency than the legacy and Internet-based solutions. The proposed solution is based on the virtualization of the deployed network nodes and service modularization using the environment of NetServ.

In [35], the authors studied different routing protocols for WSNs that aim to increase energy efficiency and network performance in terms of data latency. Most of the proposed solutions offered minimum energy consumption for WSN applications, however, such a solution unnoticed the optimal routing decisions and lead to degrading data delivery performance. In addition, it is observed that most of the solutions are not appropriate for secure data transmission under the existence of network attacks and are therefore negotiated with network operations. Thus, the knowledge of significant parameters for the formation of network structure and routing protocols is an important factor. A distributed adaptive cooperative routing protocol (DACR) is proposed in [36] using a reinforcement learning mechanism. The proposed solution significantly improves the quality of service (QoS) for low-power sensor networks with improved network lifetime. Moreover, the proposed solution makes use of the reinforcement learning algorithm and chooses optimal data forwarders and achieves network reliability. In [37], the authors proposed an ant-based QoS-aware routing protocol for heterogeneous WSNs, which aims to incorporate multimedia and scalar nodes. Furthermore, the routing decision of the proposed solution is based on control, scalar and multimedia traffic and achieves better outcomes in terms of delivery ratio and network latency. The authors in [38] proposed an energy-efficient and QoS-aware routing algorithm for industrial WSNs that aims to deliver the data in a timely and reliable manner. First, the proposed algorithm offers a link reliability estimation method and increases the delivery performance in data routing. Second, this solution guarantees that the data of different types are transmitted using various routing strategies.

3. Problem Background

Based on related studies, it can be seen that sensor nodes are self-configured, autonomous, unpredictable and have limited constraints. The main task of such tiny devices is to capture information from critical fields, process it, and transfer towards the sink node via wireless communication channels. As sensor nodes are restricted in terms of memory, processing and battery power, such limitations degrade the performance of the network in terms of well-timed and efficient data delivery. Moreover, the cluster heads—especially around the BS—consume their battery power more frequently than the far nodes, which results in network disconnectivity and extra cost in route reconstruction. Although, a machine learning-based algorithm has been proposed to improve the network lifetime and delivery ratio of WSNs. However, most of the proposed solutions offer only energy-efficient routing, and they leave the most significant aspects if any malicious node becomes a part of the network. In addition, data transmission failure may occur due to the selection of unreliable and low-quality links; such reasons may be harmful to network and route stability. Although some solutions have been proposed to cope with data security, key distribution is not dealt with reliably and there are no randomness in the key generation and data encryption methods. Moreover, malicious nodes may be able to generate false packets for accessing encryption keys, thus compromising network security. Most of the proposed secure solutions overlook intracluster secure communications, which leads to malicious nodes becoming the part of a cluster and being able to steal information. Furthermore, some solutions avoid transient link failure due to high congestion and time latency, and as a result, the fraction of route failure and packet loss increases. In addition, some solutions have incorporated link evaluation in their routing decision, but they incur the additional cost in terms of energy consumption and computational overheads.

This article aims to propose a graph-based clustering protocol with trusted routing using a supervised system for WSNs. The EGCIR protocol significantly improves the energy consumption between the sensor nodes and balances the forwarding load on cluster heads. While keeping in mind the limited constraints of low-powered sensors, the EGCIR protocol exploits the intelligent techniques to learn energy-efficient decisions with lower computational power. Furthermore, a trusted mechanism for data transmission with secure key distribution to increase the level of privacy and data integrity is also provided. Normally, keys are the combination of random bits and are used for the aim of data privacy. In the proposed solution, encryption keys are generated and distributed hierarchically to nodes to cope with large network sizes. On trusted routes, this improves the ratio of successful packet delivery and reduces the chances of faulty nodes to disclose the network data. The EGCIR protocol also produces some randomness in the key request and data encryption packets, thus, malicious nodes will not be able to resend the request packet for getting keys and encrypted data. Moreover, it maintains the routing path more intellectually based on an undirected graph model and leads to avoiding link disturbing and data breaches.

4. EGCIR Protocol

In this section, the energy-aware graph clustering and intelligent routing using a supervised system are discussed in detail. Before explaining the design of the protocol, we highlight the network assumptions as given in the following subsection.

4.1. Assumptions

The following assumptions are considered while designing and developing the EGCIR protocol:

i. The sensor nodes are distributed randomly;

ii. Sensor nodes remain static after dispersal with restricted constraints;
iii. After network deployment, nodes cannot be added or removed;
iv. Neither the source nor receiver is faulty or malicious;
v. The BS has unlimited resources in terms of processing, storage, transmission and energy;
vi. Each node has equipped with a global positioning system (GPS) to identify its position.

4.2. Design of the EGCIR Protocol

The design of the EGCIR protocol was based on supervised clustering, routing setup and route management phases. In the first phase, the deployed sensor nodes were arranged into particular boundaries only once at the time of network initialization using a K-nearest neighbor (K-NN) algorithm. Unlike frequent network formulation at a particular time, it leads to a decrease in the fraction of energy exhaustion and node overheads. Within each boundary, a specific node was selected as a cluster head to transmit cluster data towards the BS. Initially, all nodes store their neighbor information in their local tables, which comprise residual energy, distance to the BS and distance to the source node. In addition, the BS creates a generalized table to store information about all sensors, i.e., position, energy, time edge and public and secret keys for future decisions. The stored information is also updated frequently to know the exact state of the network field. In the routing setup, the trusted and efficient routes in terms of energy, time edge, and communication distance were determined. The EGCIR protocol improves the route lifetime with the least transmission cost and selected cluster heads are mapped in an undirected graph model. Furthermore, the routing path in the undirected graph model is re-adjusted using routing updates and it increases the network performance for delivery ratio and energy efficiency. In the end, routing paths are protected against faulty or misbehaving nodes and retain data security with integrity. Along with the BS that secures intercluster communications, the cluster heads also act as a local node controller and protect the intracluster communications against threats. Moreover, based on routing requirements, an appropriate next-forwarder was selected to avoid unnecessary retransmissions and information damages. The design of the EGCIR protocol is illustrated in Figure 1.

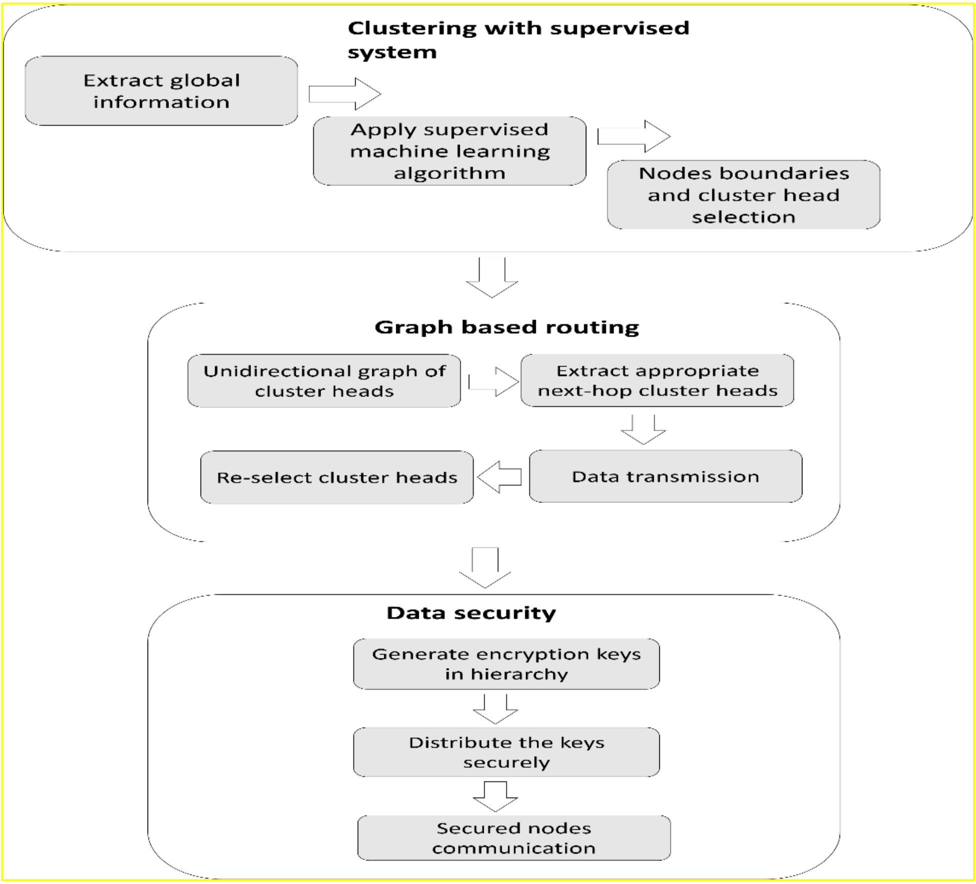
Clustering with supervised system
Extract global information
Apply supervised machine learning algorithm
Nodes boundaries and cluster head selection
Graph based routing
Unidirectional graph of cluster heads
Extract appropriate next-hop cluster heads
Data transmission
Re-select cluster heads
Data security
Generate encryption keys in hierarchy
Distribute the keys securely
Secured nodes communication

Figure 1. Research design of energy-aware graph clustering and intelligent routing using a supervised system.

4.3. Graph-Based Clustering with Supervised System

This section presents the algorithm that was developed to classify the network nodes in different boundaries, i.e., clusters. In the EGCIR protocol, the BS supervises the whole cluster formation mechanism and also creates a generalized table to hold the network information. Table 1 contains all information about sensor nodes regarding their identifier ID, energy, position, time edge, status and public and secret keys. Sensor nodes send their information to neighbors and BS to create tables both at the node and BS levels. In the EGCIR protocol, the cluster formation is based on a centralized manner and the BS acts as an organizer for such activities. The centralized mechanism offers a balance sized cluster in an efficient and orderly manner with uniform energy consumption between sensor nodes. Upon receiving the position information of the nodes, the BS determines various centroids. Afterward, the BS exploits a supervised machine learning, a K-NN algorithm [39,40] to part the dense neighbors into a particular cluster. Such a mechanism guarantees that the cluster members are closely interconnected to each other with nominal transmission distance. Furthermore, a single sensor node is belonged to only a particular cluster at a time based on the least Received Signal Strength Interference (RSSI). In the EGCIR protocol, the Manhattan distance function [41,42] is used to determine the distance between sensor nodes and centroid. This function utilizes the absolute value and provides robust results. Let us consider that x_i, y_i are the positioning coordinates of the sensor node i and (x_n, y_n) are the positioning coordinates of the centroid, then Manhattan distance $f(d)$ can be determined using Equation (1).

$$f(d) = |x_i - x_n| + |y_i - y_n| \tag{1}$$

Table 1. Structure of the generalized table.

Id	Position (x_j, y_j)	Energy e_i	Time Edge E_{ti}	Keys Information PU_{ki} and s_k	Status
—	———	———	———	———	———

Subsequently computing the distance of sensor nodes for centroid points using Equation (1), the BS formulates the k number of nearest nodes in a specific cluster, whereas the value of k depends on the external input. The BS slightly increases its transmission power to $t_r(p)$ by particular length l_o to locate the other centroid point as given in Equation (2), and accordingly, the k number of nearest neighbors is divided into another cluster.

$$t_r(p)' = t_r(p) + l_o \tag{2}$$

Afterward, the bounded nodes determine the time edge that is taken in transmitting the packets to neighbors over the transmission links. The time edge is the ratio of distance measurement using equation 1 from n_1 to n_2 and velocity of data transmission for a particular link. Let us consider that node n_1 floods the probe packets to neighbor n_2 on a transmission link at a prefixed time interval. Accordingly, the source node n_1 computes the time edge E_{t0}, $E_{t1} \ldots$, E_{tn} for probe packets P_0, P_1, $\ldots$, P_n and determines the weighted value as $E_{t0} + E_{t1} + \ldots, +E_{tn}$. Accordingly, the bounded nodes n_1, $n_2, \ldots$, n_k within the clusters compute their time edges towards neighbors and forwards the values to the BS. Based on the weighted value of time edge and residual energy, the BS selects the most suitable node as a starting cluster head and announces its status among the particular cluster members.

All the cluster members mark the entry of selected cluster head ID in their node level table. On the other hand, the BS also updates its generalized table and makes an entry in front of each sensor node regarding its status, i.e., the node is either cluster head or member. In this way, the BS keeps track of all selected cluster heads and exactly only one cluster head will be allocated to its nearest neighbors. The structure of the generalized table for node i, which is created at the BS, is given in Table 1.

When the clustering phase is over, the proposed protocol initiates the steps for intercluster routing. During intercluster routing, all cluster heads are arranged in the form of an undirected graph. In the EGCIR protocol, we assume that the cluster heads are modeled as undirected graph G. Let us consider that the cluster head is set as undirected graph G and all the edges are bidirectional, i.e., $G = (N, \epsilon)$, where N is a set of selected cluster heads and ϵ is a set of undirected edges. Whenever the source cluster head needs to send its cluster data towards the BS, it extracts the information for the next-hop cluster head from graph G and forwards the data towards it. The EGCIR protocol makes use of the following rules to identify the optimal routes:

i. If the source cluster head in graph G is only one hop away from the BS, then the route is marked as optimal and data packets are forwarded directly to the BS;
ii. If the source cluster head identifies multiple upstream cluster heads in the graph, the cluster head which meets energy, time edge and least transmission power requirements has given a higher priority;
iii. If the upstream cluster head is already busy in data transmission, then it simply drops the route request (RREQ) packet and the source cluster head extracts the information for the next upstream cluster head in the graph. This leads to distribute the data traffic evenly and avoid the congestion on nodes.

The BS continues to update the information of selected upstream cluster heads in its generalized table. It may be a case that during data forwarding, the cluster head in Graph G drops its energy power and time edge to a preset value, then the BS reselects the new cluster head, which meets the energy, time edge and distance requirements.

4.4. Data Security with Hierarchical Key Distribution

This section presents the details of proposed data security with hierarchical key distribution algorithm on the formulated clusters, as discussed in Section 4.3. In the EGCIR protocol, the process of key generation and distribution is performed in the hierarchy. The BS acts as a key distribution center (KDC), which generates and shares the symmetric key among cluster heads for data encryption and authentication. In addition, the cluster heads act as a local BS to ensure secure communications within a cluster. Symmetric keys are used for encoding and decoding sensors' data between the sending and receiving nodes. To initiate the routing session, the source sensor nodes send a request packet R_{req} towards cluster heads for issuing symmetric keys s_k. The routing session aims to identify the next-hop for data forwarding over the wireless channels. Upon receiving the R_{req}, the cluster head transmits the s_k towards sensor nodes for the particular session, also the generated s_k is encrypted with the private key PR_k of the cluster head. Similarly, on receiving the s_k from the cluster head, sensor nodes decrypt it with the public key PU_k of the cluster head, which ensures the authenticity of it. Accordingly, the cluster data $D(C_i)$ from all member nodes n_1, n_2,..., n_n is securely forwarded to the associated cluster head, as given in Equation (3).

$$D(C_i) = n_1(d) \oplus s_{k1} + n_2(d) \oplus s_{k2} + \ldots + n_n(d) \oplus s_{kn} \tag{3}$$

In addition—instead of assigning encryption keys to cluster heads before intercluster routing—the proposed security algorithm needs the cluster heads to negotiate with the BS. Both the source and upstream cluster heads i and j search the BS in their local tables, if they are found, then the cluster heads i and j send R_{req} to the BS directly. It may be a case that there is no entry of the BS in the local tables of cluster heads, then R_{req} is sent to the BS through an intermediate node. Upon receiving the request packets, the BS generates and sends back the s_k key to the cluster heads either directly or using the intermediate nodes. Likewise, the BS stores the generated s_k for the particular cluster heads in its generalized table. To secure the symmetric keys against intermediate nodes, the BS extracts the public keys of source and upstream cluster heads from the generalized table. Afterward, the BS encrypts the symmetric keys using public keys of source and upstream cluster heads. Accordingly, the symmetric keys can only be decrypted using the associated private keys and no other intermediate node can misuse the symmetric keys for data transmission. The WSN can perform different functions on open-space and there are many possibilities for malicious nodes to capture the requested packets. Therefore, the EGCIR protocol incorporates nonce in the R_{req} packet as {(ID, time) + N_i)} to produce randomness and avoid the replay attack. The nonce is a cryptographic value and is used only once in the entire communication session. The presence of nonce is a R_{req} packet, which guarantees that the same packet cannot be regenerated by any malicious node for obtaining the symmetric key. In the EGCIR protocol, the BS performs the role of main server, which is consulted before data routing takes place. Such a mechanism improves the level of security against malicious threats because each time a unique symmetric key is generated by the BS to initiate the intercluster routing. After receiving the s_k from the BS, both cluster heads i and j can exchange the secret information to authenticate with each other. Furthermore, the EGCIR protocol makes the symmetric key more secure by encryption and decryption mechanisms based on private–public keys. It exploits the RSA [43,44], which is a standard security algorithm to generate private–public keys. Let us consider that (KU_i, KR_i) and $\left(KU_j, KR_j\right)$ are the pairs of private–public keys for cluster heads i and j. The cluster head i generates a secure packet E_k by encrypting the symmetric key using the public key of cluster head j as $E(KU_j(s_k))$. The cluster head i also incorporates the nonce N_i at time t_o in the encryption process, while sharing the symmetric key with a cluster head j, such a method increases the security level of encryption in the presence of mischievous nodes and avoiding the chances of a replay attack. Then the encrypted packet $E_k(i,j)$ from cluster head i to cluster head j can be denoted as given in Equation (4).

$$E_k(i,j) = \left[E\left(KU_j(s_k)\right) + N_i(t_o)\right] \tag{4}$$

On receiving the encrypted packet from cluster head i, cluster head j applies the decryption function using its private key to recover the symmetric key, which was generated by cluster head i at time t_o. Now both cluster heads i and j have the symmetric key that was generated by the BS for a particular session. Using the shared symmetric key, both cluster heads can encrypt and decrypt the data packets securely and authentically. Accordingly, cluster heads are provided symmetric keys by the BS and afterward, when the cluster head i send data D_i to the cluster head j, the XOR $\oplus$ operation is performed along with the nonce $N_{i,j}$ to produce the encryption and randomness features, as given in Equation (5).

$$E_{i,j}(D_i) = (D_i \oplus Sk_i) + N_{i,j} \tag{5}$$

The XoR is an additive cipher that operates on bits level between sensor data and encryption keys. The encrypted data $E_{i,j}(D_i)$ is further communicated with cluster head j, which can be decrypted by applying XOR operation with the same symmetric key Sk_i, as given in Equation (6). Accordingly, the encryption and decryption process based on symmetric keys between the set of selected cluster heads are achieved in the chaining mode.

$$D_j(m_i) = E_{i,j}(D_i) \oplus Sk_i \tag{6}$$

In the end, the encrypted clusters' data $D(C_i)$, i.e., $D(C_1) + D(C_2) \ldots . + D(C_n)$, is delivered and stored at the BS. After verifying the identity and symmetric keys of cluster heads from the constructed generalized table, the cluster data are decrypted.

5. Simulation Setup

A well-known and open source network simulation tool NS-3 [45,46] was used for experiments. The EGCIR protocol was simulated against the existing work as discussed in [37,38]. The various parameters used in the simulation are highlighted in Table 2. The sensor nodes were dispersed in random order within the range of 300 × 300 m^2 over the squared sized area. The simulation time was set to 1000 s. The amount of sensors and attacker nodes was arranged from 100 to 500 and 10, respectively, which were deployed randomly. In simulation experiments, the value of k varies in different odd series to check the accuracy and avoid the ties. The size of the key was set to 64 bits and all nodes transmit data based on constant bit rate (CBR). After deployment, the energy resource of all nodes was set to 5 J. The BS had no limitations in terms of constraints.

Table 2. Simulation parameters.

Parameter	Value
Simulation area	300 × 300 m^2
Deployment	Random
Sensor nodes	100 to 500
Simulation time	1000 s
Malicious nodes	10
Packet size, k	64 bits
Energy level	5 J
Control message	25 bits
Transmission range	20 m
Traffic type	CBR

6. Simulated Results

This section exhibits the different simulated results along with their discussions in terms of network throughput, packet drop ratio, energy consumption, data latency and data breaches in the comparison between the EGCIR protocol and the existing solution. The simulation experiments are done based on a varying number of sensors.

6.1. Network Throughput

Figure 2 illustrates the simulation experiment between the EGCIR protocol and existing solutions in terms of network throughput under a varying number of nodes. The network throughput can be defined as the transferring of sensors data towards the BS at a particular period. It is observed that the EGCIR protocol outperforms the network by an average of 9% and 19% than the existing work. This is due to the fact that it identifies the attackers in the observing area at the earliest position and avoids them in data routing. Furthermore, the cluster heads are arranged in the graph-based interconnectivity to achieve a reliable multi-hop transmission paradigm, which significantly increases data productivity. Unlike other solutions that impose extra overheads to achieve data security and drop more data packets, the EGCIR protocol using machine-learning techniques reduces the additional overheads on the part of sensors and increases the delivery rate between cluster heads and BS. Furthermore, overlooking constraint resources of sensors and frequent reselection of cluster heads increase the network burden and ultimately the existing solutions decreases the network throughput.

6.2. Packets Drop Ratio

Figure 3 shows the experimental results of the EGCIR protocol in the comparison of existing solutions. Based on the results, it is revealed that the EGCIR protocol reduces the fraction of packet drop ratio by an average of 34% and 70% than other solutions under varying number of nodes and attackers. The packet drop ratio is the fraction of lost packets during data transmission. In the existing solutions, the slow security convergence provides open access to malicious nodes to simply alter or drop data packets. Moreover, the frequent exchange of control messages for construction routing paths explicitly decreases the energy resources of the sensors and leads to an increased packet drop ratio. In addition, the EGCIR protocol excludes the energy inefficient sensors from data routing and significantly contributes to a decrease in the number of packets drop. Unlike other solutions that perform frequent re-election over the entire network boundaries, the EGCIR protocol uses energy threshold and analysis of time frame. Whenever any cluster head during data routing drops its energy level, the source cluster head reselects alternative cluster heads among its members for data routing, which particularly reduces the chances of link failure and packets retransmission.

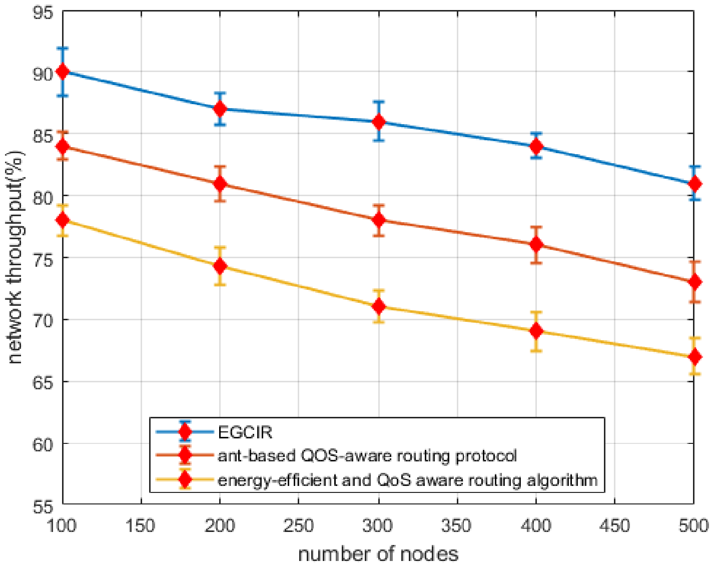

Figure 2. Impact of the nodes and network throughput.

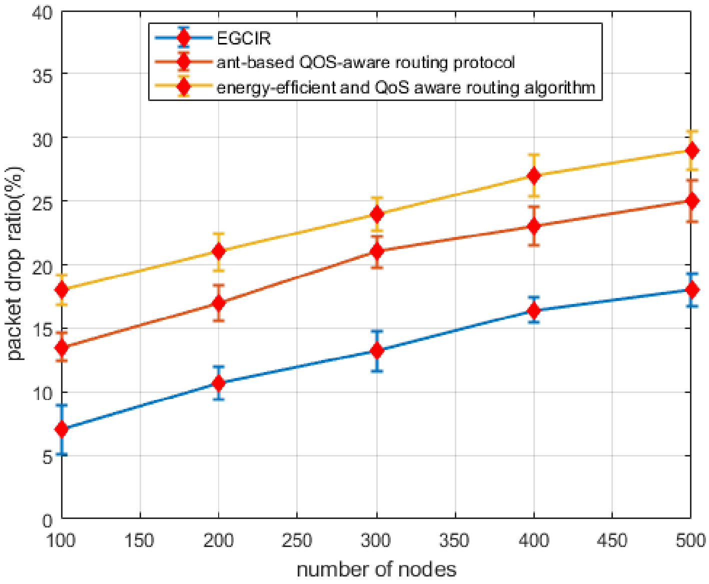

Figure 3. Impact of the nodes and packet drop ratio.

6.3. Data Latency

In Figure 4, the simulation experiments have been done for the evaluation of data latency between EGCIR protocol and existing works. The data latency is defined as the time taken from source nodes to the destination during sending sensors' information. It can be seen that the EGCIR protocol reduces the network delay in the comparison of other solutions by an average of 28% and 32.4%. This is due the fact that it reduces the length of transmission path over multi-hop points and all the points are reliable in terms of network conditions. In addition, explicit data security and authentication have been imposed by the EGCIR protocol to reduce the chances of malicious activities for data altering and dropping. Unlike other solutions that take a longer time to reselect the alternative route in data forwarding, the EGCIR protocol exploits intelligent methods to evaluate the robust paths with the reduction of route rediscoveries and ultimately it decreases the network end-to-end delay. Furthermore, the cluster members are closer to selected cluster heads, which results in reducing the communication distance among them and leads to improve the fraction of network delay.

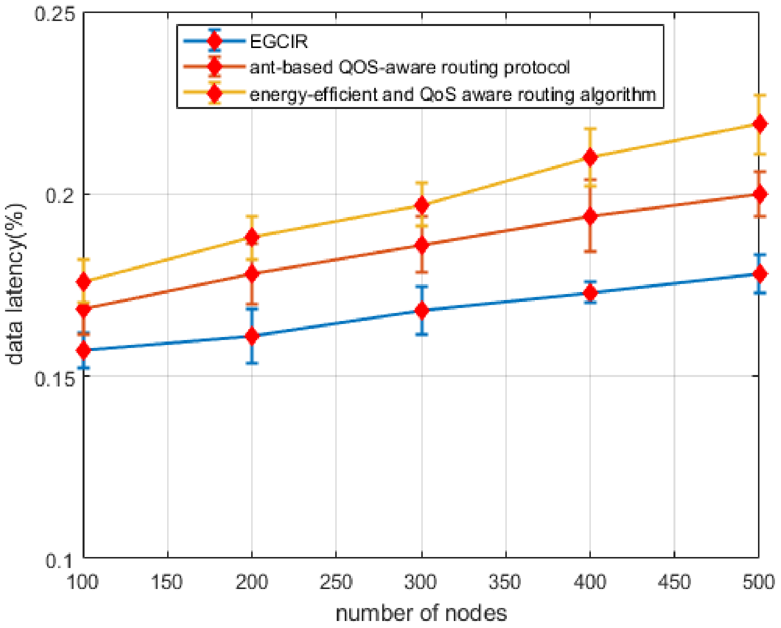

Figure 4. Impact of the nodes and data latency.

6.4. Energy Consumption

Figure 5 demonstrates the evaluation of energy consumption of the EGCIR protocol in the comparison of other solutions under a varying number of nodes. The energy consumption is defined as the fraction of depleted energy resources among sensor nodes while transmitting, aggregating and receiving the network data. It is observed that the EGCIR protocol significantly reduces a load of energy consumption over the network field by an average of 9% and 17%. This is because it exploits a centralized cluster formation and cluster head selection process. All information of the entire network field is placed on a centralized location at the BS, which is based on the updated node conditions. Furthermore, cluster heads only change their positions when there is a demand of member nodes instead of the regular epoch. Unlike other solutions that impose extra energy load on the sensors to protect data packets from malicious nodes, the EGCIR protocol securely forwards the communication keys between both the source and next-hop cluster heads using a light computational XoR mathematical function. Furthermore, the existing solutions deplete additional energy constraints on transmitted more control messages for the formation of routing paths. On the other hand, the EGCIR protocol using graph-based approach for the formation of cluster heads linkage consumes minor energy in the process of route management.

6.5. Data Breaches

In Figure 6, the simulation results demonstrate the analysis of the ratio of data breaches of the EGCIR protocol in the comparison of the existing solutions. The data breach is the planned or unplanned release of confidential information to untrusted nodes and is therefore compromised to network performance. It is noted that the EGCIR protocol reduces the fraction of data breaches by an average of 30% and 45%. The reason behind this reduction is such that the EGCIR protocol is extremely energy-efficient and reliable in terms of clustering and data security processes. Moreover, unlike other solutions that manage data transmission in a distributed manner, the EGCIR protocol uses the BS as a KDC, which guarantees secure key generation and distribution among sensors. In addition, the integration of nonce in key sharing, data encryption and decryption reduces the probabilities of malicious nodes to perform misbehaving actions in the observing area. Moreover, unlike other solutions where routing paths are more prone to failure due to heavy load network setup, the EGCIR protocol uses a graph learning-based intelligent routing decision to extract the shortest, energy-efficient and most trustworthy nodes from the undirected graph—which results in reducing data breaches and increases the level of security for sensor data.

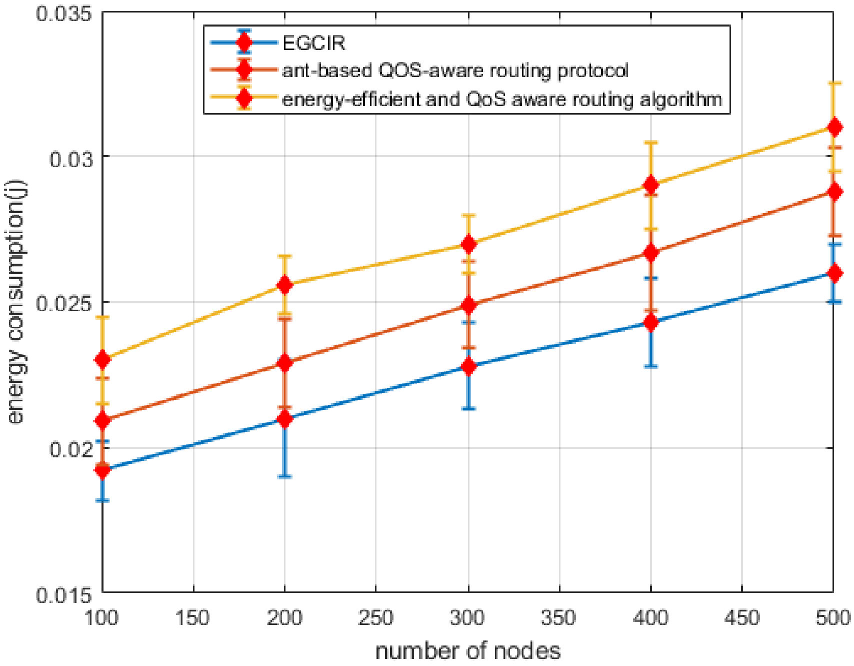

Figure 5. Impact of the nodes and energy consumption.

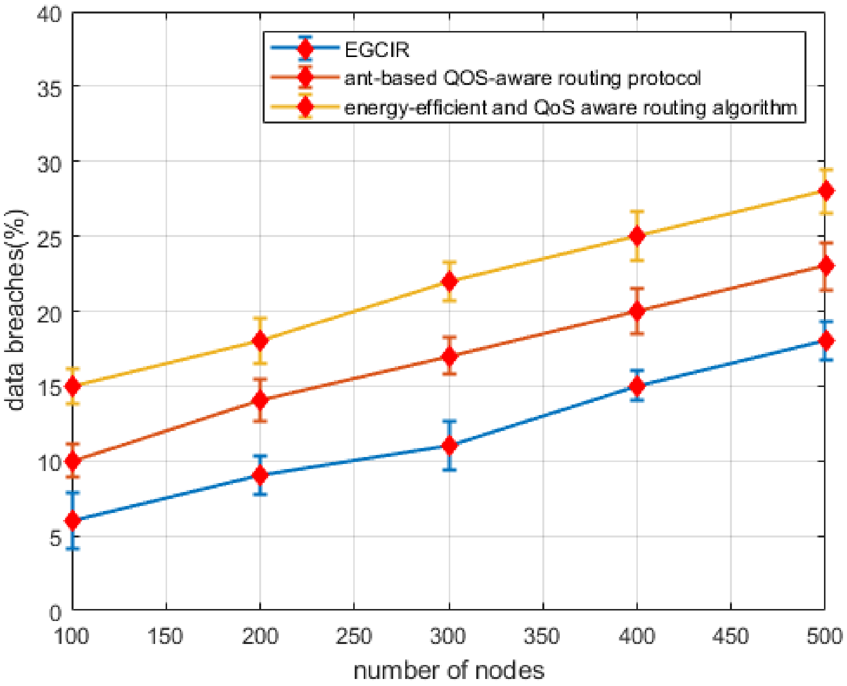

Figure 6. Impact of the nodes and data breaches.

7. Conclusions

Low-powered sensor nodes are deployed in various applications due to their cost-effective and easy management. However, resource restriction on the part of sensors harms the performance of real-time applications. Such impacts degrade network delivery and lifetime in the large-scale applications. In addition, due to ad-hoc and insecure communication media, sensor-based networks are open to various attackers by disclosing node data and impacting network openness. This paper proposes energy-aware graph clustering and intelligent routing using a supervised system for WSNs. The EGCIR protocol makes the use of graph clustering, which is a machine-learning technique that learns the decisions based on diverse conditions. Moreover, the analysis of the residual energy and time edges to compute the transmission latency improves the route lifetime and energy consumption in data transmission. Furthermore, the secure distribution of keys in the hierarchy, i.e., from the BS to cluster heads and from cluster heads to sensor nodes, leads to avoiding the chances for malicious threats with nominal overheads for large scale networks. The simulation experiments demonstrate that the EGCIR protocol outperforms the existing solutions in terms of various routing and security measurements. However, the EGCIR protocol needs to train the sensor nodes with a neural network-based algorithm to further decrease the processing overheads and improve the energy efficiency.

Author Contributions: Conceptualization, K.H. and S.M.F.; methodology, K.H.; software, A.A. (Ayman Altameem); validation, I.U.D.; T.S. and A.A. (Ahmad Almogren); formal analysis, A.A. (Ayman Altameem); investigation, T.S.; resources, K.H.; data curation, S.M.F.; writing—original draft preparation, T.S.; writing—review and editing, I.U.D.; visualization, T.S.; supervision, I.U.D.; project administration, A.A. (Ahmad Almogren); funding acquisition, A.A. (Ahmad Almogren). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by King Saud University, Riyadh, Saudi Arabia, through Researchers Supporting Project number RSP-2020/184.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rawat, P.; Singh, K.D.; Chaouchi, H.; Bonnin, J.M. Wireless sensor networks: A survey on recent developments and potential synergies. *J. Supercomput.* **2014**, *68*, 1–48. [CrossRef]
2. Yang, M.; Li, Y.; Jin, D.; Zeng, L.; Wu, X.; Vasilakos, A.V. Software-defined and virtualized future mobile and wireless networks: A survey. *Mob. Netw. Appl.* **2015**, *20*, 4–18. [CrossRef]
3. Yetgin, H.; Cheung, K.T.K.; El-Hajjar, M.; Hanzo, L.H. A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 828–854. [CrossRef]
4. Din, I.U.; Asmat, H.; Guizani, M. A review of information centric network-based internet of things: Communication architectures, design issues, and research opportunities. *Multimed. Tools Appl.* **2019**, *78*, 30241–30256. [CrossRef]
5. Ahmed, G.; Zou, J.; Zhao, X.; Fareed, M.M.S. Markov chain model-based optimal cluster heads selection for wireless sensor networks. *Sensors* **2017**, *17*, 440. [CrossRef]
6. Hu, S.; Han, J.; Wei, X.; Chen, Z. A multi-hop heterogeneous cluster-based optimization algorithm for wireless sensor networks. *Wirel. Netw.* **2015**, *21*, 57–65. [CrossRef]
7. Ucar, S.; Ergen, S.C.; Ozkasap, O. Multihop-cluster-based IEEE 802.11 p and LTE hybrid architecture for VANET safety message dissemination. *IEEE Trans. Veh. Technol.* **2015**, *65*, 2621–2636. [CrossRef]
8. Awan, K.A.; Din, I.U.; Almogren, A.; Guizani, M.; Khan, S. StabTrust—A stable and centralized trust-based clustering mechanism for IoT enabled vehicular ad-hoc networks. *IEEE Access* **2020**, *8*, 21159–21177. [CrossRef]
9. Yuan, J.; Zhang, J.; Ding, S.; Dong, X. Cooperative localization for disconnected sensor networks and a mobile robot in friendly environments. *Inf. Fusion* **2017**, *37*, 22–36. [CrossRef]
10. Bouaziz, M.; Rachedi, A.; Belghith, A. EC-MRPL: An energy-efficient and mobility support routing protocol for Internet of Mobile Things. In Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2017.
11. Chen, R.; Guo, J. Hierarchical trust management of community of interest groups in mobile ad hoc networks. *Ad Hoc Netw.* **2015**, *33*, 154–167. [CrossRef]
12. Ullah, U.; Khan, A.; Zareei, M.; Ali, I.; Khattak, H.A.; Din, I.U. Energy-effective cooperative and reliable delivery routing protocols for underwater wireless sensor networks. *Energies* **2019**, *12*, 2630. [CrossRef]
13. Guleria, K.; Verma, A.K. Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks. *Wirel. Netw.* **2019**, *25*, 1159–1183. [CrossRef]
14. Naureen, A.; Zhang, N.; Furber, S. Identifying energy holes in randomly deployed hierarchical wireless sensor networks. *IEEE Access* **2017**, *5*, 21395–21418. [CrossRef]

15. Habib, M.A.; Moh, S. Game theory-based routing for wireless sensor networks: A comparative survey. *Appl. Sci.* **2019**, *9*, 2896. [CrossRef]
16. Haseeb, K.; Islam, N.; Almogren, A.; Din, I.U. Intrusion prevention framework for secure routing in WSN-based mobile Internet of Things. *IEEE Access* **2019**, *7*, 185496–185505. [CrossRef]
17. Pan, J.-S.; Dao, T.-K. A compact bat algorithm for unequal clustering in wireless sensor networks. *Appl. Sci.* **2019**, *9*, 1973.
18. Jadoon, R.N.; Zhou, W.; Jadoon, W.; Khan, I.A. RARZ: Ring-zone based routing protocol for wireless sensor networks. *Appl. Sci.* **2018**, *8*, 1023. [CrossRef]
19. Tang, L.; Chen, Z.; Cai, J.; Guo, H.; Wu, R.; Guo, J. Adaptive Energy Balanced Routing Strategy for Wireless Rechargeable Sensor Networks. *Appl. Sci.* **2019**, *9*, 2133. [CrossRef]
20. Abuarqoub, A.; Hammoudeh, M.; Adebisi, B.; Jabbar, S.; Bounceur, A.; Al-Bashar, H. Dynamic clustering and management of mobile wireless sensor networks. *Comput. Netw.* **2017**, *117*, 62–75. [CrossRef]
21. Granjal, J.; Monteiro, E.; Silva, J.S. Security in the integration of low-power Wireless Sensor Networks with the Internet: A survey. *Ad Hoc Netw.* **2015**, *24*, 264–287. [CrossRef]
22. Farash, M.S.; Turkanović, M.; Kumari, S.; Hölbl, M. An efficient user authentication and key agreement scheme for heterogeneous wireless sensor network tailored for the Internet of Things environment. *Ad Hoc Netw.* **2016**, *36*, 152–176. [CrossRef]
23. Ashibani, Y.; Mahmoud, Q.H. Cyber physical systems security: Analysis, challenges and solutions. *Comput. Secur.* **2017**, *68*, 81–97. [CrossRef]
24. Haseeb, K.; Islam, N.; Almogren, A.; Din, I.U.; Almajed, H.N.; Guizani, N. Secret sharing-based energy-aware and multi-hop routing protocol for IoT based WSNs. *IEEE Access* **2019**, *7*, 79980–79988. [CrossRef]
25. Erol-Kantarci, M.; Mouftah, H.T. Wireless multimedia sensor and actor networks for the next generation power grid. *Ad Hoc Netw.* **2011**, *9*, 542–551. [CrossRef]
26. Minoli, D.; Sohraby, K.; Occhiogrosso, B. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet Things J.* **2017**, *4*, 269–283. [CrossRef]
27. García Villalba, L.J.; Sandoval Orozco, A.L.; Cabrera, A.T.; Abbas, C.J.B. Routing protocols in wireless sensor networks. *Sensors* **2009**, *9*, 8399–8421. [CrossRef]
28. Bapu, B.T.; Gowd, L.S. Link quality based opportunistic routing algorithm for QOS: Aware wireless sensor networks security. *Wirel. Pers. Commun.* **2017**, *97*, 1563–1578. [CrossRef]
29. Din, I.U.; Guizani, M.; Hassan, S.; Kim, B.-S.; Khan, M.K.; Atiquzzaman, M.; Ahmed, S.H. The Internet of Things: A review of enabled technologies and future challenges. *IEEE Access* **2018**, *7*, 7606–7640. [CrossRef]
30. Virrankoski, R.; Savvidees, A. TASC: Topology adaptive spatial clustering for sensor networks. In Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, Washington, DC, USA, 7 November 2005.
31. Shigei, N.; Morishita, H.; Miyajima, H. Energy efficient clustering communication based on number of neighbors for wireless sensor networks. In Proceedings of the International multi-conference on engineers and computer scientists (IMECS), Hong Kong, China, 17–19 March 2010.
32. Bernas, M.; Płaczek, B. Fully connected neural networks ensemble with signal strength clustering for indoor localization in wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 403242. [CrossRef]
33. Srivastava, J.R.; Sudarshan, T. A genetic fuzzy system based optimized zone based energy efficient routing protocol for mobile sensor networks (OZEEP). *Appl. Soft Comput.* **2015**, *37*, 863–886. [CrossRef]
34. Femminella, M.; Reali, G.; Valocchi, D.; Francescangeli, R.; Schulzrinne, H. Advanced caching for distributing sensor data through programmable nodes. In Proceedings of the 2013 19th IEEE Workshop on Local & Metropolitan Area Networks (LANMAN), Bussels, Belgium, 10–12 April 2013.
35. Shabbir, N.; Hassan, S.R. Routing protocols for wireless sensor networks (WSNs). *Wirel. Sens. Netw. Insights Innov.* **2017**. [CrossRef]
36. Razzaque, M.A.; Ahmed, M.H.U.; Hong, C.S.; Lee, S. QoS-aware distributed adaptive cooperative routing in wireless sensor networks. *Ad Hoc Netw.* **2014**, *19*, 28–42. [CrossRef]
37. Malik, S.K.; Dave, M.; Dhurandher, S.K.; Woungang, I.; Barolli, L. An ant-based QoS-aware routing protocol for heterogeneous wireless sensor networks. *Soft Comput.* **2017**, *21*, 6225–6236. [CrossRef]
38. Zhang, W.; Liu, Y.; Han, G.; Feng, Y.; Zhao, Y. An energy efficient and QoS aware routing algorithm based on data classification for industrial wireless sensor networks. *IEEE Access* **2018**, *6*, 46495–46504. [CrossRef]

39. Soucy, P.; Mineau, G.W. A simple KNN algorithm for text categorization. In Proceedings of the 2001 IEEE international Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001.
40. Pan, L.; Li, J. K-nearest neighbor based missing data estimation algorithm in wireless sensor networks. *Wirel. Sens. Netw.* **2010**, *2*, 115. [CrossRef]
41. Singla, A.; Karambir, M. Comparative analysis & evaluation of euclidean distance function and manhattan distance function using k-means algorithm. *Int. J.* **2012**, *2*, 298–300.
42. Mohibullah, M.; Hossain, M.Z.; Hasan, M. Comparison of euclidean distance function and manhattan distance function using k-mediods. *Int. J. Comput. Sci. Inf. Secur.* **2015**, *13*, 61.
43. Zhou, X.; Tang, X. Research and implementation of RSA algorithm for encryption and decryption. In Proceedings of the 2011 6th International Forum on Strategic Technology, Harbin, Heilongjiang, China, 22–24 August 2011.
44. Milanov, E. The RSA algorithm. *RSA Labs.* 2009, pp. 1–11. Available online: https://sites.math.washington.edu/~{}morrow/336_09/papers/Yevgeny.pdf (accessed on 4 August 2020).
45. Kumar, A.A.; Rao, S.; Goswami, D. Ns3 simulator for a study of data center networks. In Proceedings of the 2013 IEEE 12th International Symposium on Parallel and Distributed Computing, Bucharest, Romania, 27–30 June 2013.
46. Katkar, P.S.; D.Ghorpade, V.R. Comparative study of network simulator: NS2 and NS3. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2016**, *6*, 608–612.