Received: 23 April 2018 | Revised: 14 July 2018 | Accepted: 18 July 2018

DOI: 10.1002/ett.3503

SPECIAL ISSUE ARTICLE

WILEY

Bio-inspired routing protocol for WSN-based smart grid applications in the context of Industry 4.0

M. Faheem[1,2] | R. A. Butt[3] | B. Raza[4] | M. W. Ashraf[5] | Seema Begum[6] | Md. A. Ngadi[1] | V. C. Gungor[2]

[1]Department of Computer Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

[2]Department of Computer Engineering, Abdullah Gul University, Kayseri, Turkey

[3]Department of Telecom Engineering, NED University of Engineering and Technology, Karachi, Pakistan

[4]Department of Computer Science, COMSATS University, Islamabad, Pakistan

[5]Department of Computer Engineering, Bahauddin Zakariya University, Multan, Pakistan

[6]School of Computer Science and Technology, Dalian University of Technology, Dalian, China

Correspondence

Md. A. Ngadi, Department of Computer Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
Email: dr.asri@utm.my

Funding information

International Doctoral Fellowship (IDF) program from UTM; Research Grant from UTM; TUBA-GEBIP, Grant/Award Number: V.G./TUBA-GEBIP/2013-14

Abstract

Recently, the advences of Industry 4.0 have paved the way for a systematical deployment of the smart grid (SG) to manage continuously growing energy demand of the 21st century. This even allows the fourth stage of the industrial revolution in the power sector, which is known as the smart grid industry (SGI) 4.0. In SGI 4.0, the industrial wireless sensor networks (WSNs) and the Internet of Things are envisioned as key promising communication technologies for monitoring various SG applications due to their large-scale coverage, fault tolerance characteristics, and cost reduction. However, highly dynamic nature of the SG environments brings several unique challenges caused by systems and operating devices. This results in hampering the quality-of-service communication requirements for WSNs-based SG applications. In SGI 4.0, the routing infrastructure not only requires a reliable but also fulfills the communication requirements of diverse SG applications. Thus, a sophisticated, reliable and QoS-aware multi-hop communications network architecture enabling a real-time exchange of data for various WSNs-based SG applications is essential in SGI 4.0. Hence, this paper proposes a novel bio-inspired self-optimized butterfly mating optimization-based data gathering routing scheme called Self-optimized Intelligent routing protocol (SIRP) for WSNs-based SG applications. The extensive simulations reveal that the proposed scheme achieves its defined goals compared to existing routing schemes designed for WSNs-based applications.

1 | INTRODUCTION

The recent advances in information and communication technologies (ICTs) have permitted the fourth stage of industrial revolution known as Industry 4.0 (I 4.0). The main ideas of the term I 4.0 were first time discussed in 2011 at the Hannover Fair as a high-tech strategy for 2020 in Germany.[1] Similar ideas have also been discussed in other industrial countries, such as “Internet +” in China, “the fourth industrial revolution” in Korea, “Factories of the Future” in Europe, and “Industrial Internet” in the USA.[2,3] The key idea of I 4.0 is to make existing industries intelligent enough to identify the need for a significant reduction in faults and short-cycle adaptation for higher productivity to increase economic benefits. Industry 4.0 will serve as a conceptual base for intelligent realizing and controlling physical production processes locally as well as globally by using advance ICTs.[4] Therefore, in I 4.0, ICTs play a dominant role in strengthening the overall flexibility of manufacturing resources for increasing productivity. Industry 4.0 will allow customers to choose a variety of desired quality level products at a constant price with a high degree of freedom anytime, anywhere in the world. This will definitely increase the number of competitors and sellers to the buyer's market as well as the life of the people in the living world.

In I 4.0 pattern, all manufacturers, because of the growing interconnectivity of machines, parts, products, and operators, demand a highly stable connectivity and autonomous interaction among different systems and subsystems in their factories. To this end, the key aim of Internet of Things (IoT) is to connect different types of cyber-physical Systems (CPSs) via the Internet of Services for generating and sharing information with each other and with humans to cooperate closely.[5] This enables real-time data collection and sharing of industry systems on the cyber layers to carry out monitoring and control logic intelligently from anywhere worldwide. The efficient and effective use of this information and knowledge changes the relationship among suppliers, consumers, and stakeholders in the industry. Moreover, it reduces or scales the costs by offering the opportunity to use services and storage in the cloud such as web shops, big data analytics, and online portals. In addition, the use of IoT in I 4.0 allows making intelligent systems, management decision support systems, and predictive diagnostic systems to increase the production capacity and thus result in significant financial benefits.[6] This real-time exchange of meaningful information from the entire production and consumption floor will allow manufacturing companies to monitor, control, and manage factory operations in a more efficient, reliable, and flexible manner. Therefore, IoT has various applications in the technical market such as agriculture, traffic monitoring, parking systems, logistics, smart grids, smart vehicles, smart homes, smart cities, smart security, and several others.[7,8] The advancements in IoT will form the interface between the physical and the virtual worlds for exchanging instant industrial information is seen as a real game-changer for industrial networking in the I 4.0.

Generally, IoT is incorporating two types of communication technologies (CTs), namely wired and wireless. The key aim of the both CTs at the communication layer is to provide highly stable networking for automated exchange of information between different production systems in the smart factories to empower I 4.0. However, the design and implementation of these CTs in a two-way manner is extremely challenging because of the security and diverse QoS requirements of various factory applications. Therefore, it is extremely important that utilities should understand the various QoS requirements of their industrial applications before deploying the communication infrastructure for the I 4.0.[9] Presently, many devices to streamline management operations are connected using wired networks working over industrial protocols in traditional factories. However, compared with a wired network, the wireless networking solution plays a complementary role to empower control and management competencies of the factory elements in I 4.0. To this end, industrial wireless sensor networks (IWSNs) are envisioned as a valuable technology to realize the vision of I 4.0 because of their low-cost sensing, identifying, processing, and networking capabilities. The IWSNs provide physical industry world information to a user located in a remote location to realize distinct objectives.[10,11] The fundamental characteristics of IWSN-based networking solution in I 4.0 is to provide least maintenance cost by avoiding excessive breakdowns in a defined time interval for enabling smart production reliability and increasing economic benefits. Therefore, IWSN, with these aims, has revealed its prominence in various industrial applications, including modern agriculture, transportation, steel mills, healthcare systems, area surveillance, gas and oil industry, and many others.[12,13]

The quality-of-service (QoS)–aware cooperative multihop communication in IoT-enabled smart grid (SG) is highly dependent on a number of performance metrics like coverage, scalability, stability, latency, lifetime, and complexity of the network. In IoT-enabled SG, IWSNs through IoT connect various industrial components of the power grid to the information world, which results in high-quality power generation and distribution and innovative services.[14] Consequently, IWSNs in the SG can considerably diminish global infrastructure cost by enhancing quality of the product, accelerating production, raising the flexibility, and making installation easier to empower SGI 4.0. This makes them to be a part of the strategic decision-makers and flexible problem-solvers in the technical complexity that might even create entirely new manufacturing concepts.[15] In smart grid industry (SGI) 4.0, a wide range of existing and anticipated wireless sensor networks (WSNs)–based smart grid applications are given in the work of Bukhari et al.[16] In the fourth-generation global marketplace, all these SG applications provide a vying edge, leading to new products and services for improving the ordinary citizens day-to-day lives. However, the realization of these applications leans on efficient and reliable communication architecture proficiencies deployed in the SG.[17] On the other hand, recent studies reveal that in SG, the wireless channels because of fading, multipath effects, equipment noise, heat, electromagnetic interference, and dusty environments have several distinct challenges, which generally do not exist in other environments. As a result, the wireless link quality between sensor nodes deployed for SG applications varies in different time and locations of the network. This makes QoS-aware multihop data transmissions challenging for WSN-based for SG applications.[18] Hence, under adverse wireless communications conditions, providing the quality-aware data delivery to accomplish QoS requirements like network coverage, scalability, stability, reliability, throughput, latency, lifetime, and complexity are key design issues in SG.[19]

To tackle aforementioned challenges, this paper proposes an innovative bio-inspired self-optimized butterfly mating optimization-based data gathering routing scheme called SIRP for WSNs-based event monitoring SG applications to empower SGI 4.0. The proposed bio-inspired routing protocol finds near-optimal shortest loop-free routing solutions in both sparse and densely deployed networks by considering the natural behavior of butterfly and evolution of genetic alteration in the mating process. The designed scheme selects highly stable links among sensors from the source toward the destination by considering the SG environments in the network. This results in high network throughput, packet delivery ratio (PDR) with low data packet error rate, delay, congestion, and energy consumption at extremely low network complexity in the SG. Moreover, the use of dual sinks and the key idea of packet forwarding in a greedy manner along a narrow routing path in each subregion significantly distribute the data traffic load and minimize the sensor nodes memory overflow and hotspot issues in the network. In addition, the proposed routing protocol avoids excessive route failures and employs a self-learning–based intelligent mechanism to tackle route failure problems in a bounded time interval in the network. The proposed routing protocol is appropriate for both sparse and densely deployed large and small-scale WSNs-based SG applications.

The remainder of this article is structured as below. Section 2 gives an extensive literature review in the SG. Section 3 presents problem statement and motivations. Section 4 explains the network model and our proposed scheme. Section 5 reveals simulation setup and path-loss-model parameters. Section 6 provides a detailed performance analysis, and finally, Section 7 summarizes the article, including future research work.

2 | LITERATURE REVIEW

The current research on QoS-aware and energy-efficient routing solutions designed for WSN-based SG applications is limited, though some data-gathering solutions exist for SG applications. For example, Mu[20] proposes a minimum physical distance-based information-gathering scheme for SG applications, in which data packets are transmitted along the trajectory of the devices with the minimum payload to optimize the transmission of the monitoring in advanced metering infrastructure (AMI). The designed scheme improves the end-to-end (e2e) delay by providing short paths in the network. Yaacoub and Abu-Dayya[21] propose a contention-based channel-aware reservation approach to investigate the real-time transmission of smart metering data. The suggested scheme reduces data packet collision in the network. Xiang et al[22] present an innovative geographical routing protocol for reliable e2e data transmission for SG applications. The proposed scheme performs well in terms of achieving a high level of reliable data transmission and energy efficiency in the network. The study by Lin et al[23] offers an energy-efficient clustering routing protocol to handle traffic variability for monitoring transmission lines in the SG. The proposed scheme takes advantages of low energy consumption and real-time network reliability performance in the network. The research by Zhang et al[24] also presents a self-healing routing protocol for monitoring long-range power transmission lines in the SG. The designed scheme provides data confidentiality, data integrity, mutual authentication, and energy efficiency for e2e packet transmissions in the network.

The work by Yang et al[25] proposes a novel ant colony optimization-based routing technique to improve the QoS in cognitive radio-enabled AMI networks in SG. The proposed scheme selects the best route from the source toward the destination based on global optimization mechanism in the network to reduce harmful interference impact of transmission reliability in the SG. Deng et al[26] is offering a novel multigateway backup routing scheme to improve the route decision with effective link quality, queue optimization, throughput, and overhead in the network. The scope of the work of Faheem and Gungor[27] mainly focuses on a multihop spectrum and capacity-aware routing protocol for WSNs-based SG application. The proposed routing protocol considerably decreases the data packet collision and enhances the packet delivery ratio, network throughput, energy utilization, and e2e delay in the network. The work of Fadel et al[28] presented a bio-inspired multichannel routing scheme in cognitive radio sensor network to optimize e2e reliable data packet transmission for SG applications. The suggested scheme performs well in terms of network delay, packet delivery ratio, and network throughput in a specific SG environment. Hui et al[29] proposed a top-k query-based clustering routing protocol for efficient and reliable data transmission in SG. The proposed scheme, by employing a top-k query mechanism, optimizes the routing process for SG applications. The suggested scheme significantly improves the residual energy by minimizing high transmission cost in the network.

The work of Erol-Kantarci and Mouftah[30] provides extensive home energy management performance by using a novel data collection mechanism in the SG. The designed scheme efficiently improves the packet delivery ratio and decreases the total cost of monitoring for real-time pricing in the SG. A novel link quality-aware routing protocol has been discussed in the work of Farooq and Tang Jung[31] for AMI in the SG. The designed scheme, by employing shortest path routing architecture, successfully improves the packet delivery ratio and minimizes e2e delay, energy consumption, and normalized routing overload in the network. The research by Hou et al[32] deals with a multicast routing tree for minimizing e2e delay and interference impact on packet transmission for SG applications. Moreover, Kim and Jin[33] presents a branch-based data collection routing mechanism. In the designed scheme, the entire routing network is divided into several multiple tree branches for efficient data collection in the SG. The offered scheme successfully minimizes data packet collision and information processing time for SG monitoring. A multilayer data collection scheme for SG systems is presented in the work of Islam and Koo.[34] The designed model fuses routing information at each protocol layer to minimize packet error rates in the entire network. The proposed scheme takes into account the advantages of real-time robust, secure, and energy-efficient packet transmission for SG applications. Lastly, in the work of Li et al,[35] a maximum likelihood communication protocol has been proposed for reliable and energy efficient data delivery in the SG. The designed scheme, by employing novel level-based routing mechanism, takes the advantages of high packet delivery, energy efficiency, and network lifetime.

3 | PROBLEM STATEMENT AND RESEARCH MOTIVATION

The SGI 4.0 paradigm offers a platform, vision, and architecture of high-quality electricity generation and distribution in the SG. The key objective of the SGI 4.0 is to provide intelligent electricity by using advanced IoT envisioned to offer a variety of advantages in the following areas: emerging economies, renewable energy sources, environmental, efficiency, reliability, security, and safety.[36] In SGI 4.0, the SG systems will be capable of collecting data, communicate with computers for analyzing it, and give an advice on it to the operator for necessary actions. This self-cognition, self-optimization, and self-customization of the SGI 4.0 will enable different systems of the power grid to operate independently or with less human interventions for high-quality electricity.[37] Thus, in SGI 4.0, the systems, workers, customers, and other stakeholders will be interconnected and cooperate closely. This real-time interaction between systems, workers, and customers through the continuous and the autonomous exchange of information will build a highly flexible power generation model that will greatly improve power efficiency and competitiveness in the power industries. This will not only change the traditional methods of electricity generation but also guide for future improvements in terms of internal complexity and external unpredictable events. However, the multifunction and autonomous equipment's in a large-scale SG generate a huge amount of ambiguous data in the network. In addition, the communication requirements in the SGI 4.0 are increased regarding data rates, bandwidth, latency, efficiency, scalability, flexibility, and reliability. The existing information exchange approaches are not well suited to handle this massive data, which dramatically increases the internal complexity of the systems, leading to more uncertainties because of atypical events. Some major drawbacks of existing routing schemes designed for SG applications are summarized as below.

Generally, the existing approaches (see Section 2 for detail) forming an ad hoc network among the sensor nodes and involve multihop communication for relaying data from the sources toward the single static sink in a network deployed for WSNs-based SG monitoring applications. In most existing schemes, a major problem is observed when sensor nodes near the static sink forward a huge amount of data from farther sensors in the deployed network. Thus, these sensors carrying a heavy traffic load result in the many-to-one traffic pattern in the network. Therefore, they are more prone to energy exhaustion and causing the network partition problem when consuming their energy. This results in the premature lifetime ending of a WSN since no more data packet can be transmitted to the sink for the SG applications. Moreover, in order to balance the total network lifetime, the existing routing schemes are forwarding data packets over excessive hops from data sources toward the sink. This could be efficient in terms of balancing the network lifetime; however, it rapidly consumes a notable amount of energy at each node in both sparse and dense deployment networks. Thus, it decreases the overall network lifetime for WSNs-based SG monitoring applications. In addition, the increasing number of relaying sensor nodes in a data forwarding process may increase the delay and security concerns, and thus, data reliability becomes one of the main challenges in these routing schemes designed for SG applications. Furthermore, in most existing works, the sensors two-hops or one-hop away to the sink are facing congestion problem, resulting in excessive loss of data packets because of memory overflow issues in a large-scale network. Thus, the buffer of sufficient size should be implemented at each sensor node for WSN-based SG applications. To this end, a dual sink-based routing mechanism can efficiently solve most of the aforementioned QoS-aware reliable data collection problems for WSNs-based SG applications.

Additionally, most existing schemes do not consider the appropriate dynamic nature of the SG environment while multihop packet transmissions and thus resulting in an excessive corrupted data packets in the deployed network. Moreover, the topology of the network changes frequently because of excessive route failures in the network. This results in unacceptable control message overhead since nodes need to send out query packets periodically for updating neighboring profiles in a new topology. Although some cognitive radio schemes provide a high packet delivery ratio, they are facing high network complexity, overhead, energy consumption, and implantation issues for various SG applications. In addition, the traditional routing schemes require enormous computational efforts that result in an increasing problem size. Thus, they cannot provide optimal or a set of near-optimal values for reliable multihop data transmission and thus fails to find appropriate alternative data paths in a bounded time for conveying data in the network. Lastly, the existing schemes without modification in the protocol stack cannot be directly implemented for the IoT-based SG applications in the context of SGI 4.0. In summary, the existing routing schemes designed for various SG applications are not fully smart enough to provide QoS-aware data gathering in harsh nature SG environments. This motivates us to designing an innovative bio-inspired, self-optimized intelligent butterfly mating optimization data gathering routing scheme (called SIRP) for WSN-based event monitoring SG application to empower SGI 4.0. The designed scheme is appropriate and performs well in terms of network delay, lifetime, residual energy, throughput, packet delivery ratio, congestion management, packet error rates, and balancing the overall network lifetime at extremely low network complexity in both sparse and densely deployed networks for various SG applications.

4 | PROPOSED SELF-OPTIMIZED QOS-AWARE DATA GATHERING SCHEME

The following sections explain the working principle of our designed scheme in the SG.

4.1 | Network model

Recently, many renewable energy generators like solar, tidal, biomass, and wind generators, which produce a few kilowatts up to some megawatts, are becoming extensively diffused around the world.[36] These large-scale distributed generation systems are transformation of the present power systems containing several hundreds of generators, characterized by power, current, voltage, and different technologies and topologies. The whole electrical power system is the integration of these large numbers of geographically distributed systems and subsystems. The digital integration complexity of these geographically distributed components in the existing power grid makes it very vulnerable and thus requiring an intelligent, two-way, and secure communication mechanism for monitoring events in the electric grid. To this end, the network model of the designed routing protocol is depicted in Figure 1. Herein, we assumed a 550 kV outdoor grid station consists of various types of systems, subsystems, and poles for electricity transmission and distribution. In this harsh nature SG environment providing QoS-aware data gathering using WSNs for real-time monitoring applications is extremely challenging. This is because of fading, multipath effects, equipment noise, heat, electromagnetic interference, and dusty environments. Without loss of generality, the whole network model has been partitioned into main three layers, namely sensor layer, sink layer, and user interaction layer usually known as application layer or Internet layer.

The sensor layer consists of homogenous sensor nodes deployed randomly in an ad hoc manner in the network. This entire sensor network is divided into equally distributed regions and subregion to shift the data traffic burden to both sinks. Each sensor node in the network is equipped with equally limited energy, buffer size, data processing capability, and omnidirectional communication capability. The key responsibilities of the tiny sensor nodes at sensor layer is to monitor surrounding, collaborate with each other, and report the sensed data to closer associated sink. At sink layer, two sinks having the same communication and data processing capabilities are deployed at distinct positions in the network. Both static sinks are connected to each other via line-of-sight highly stable microwave communication link. Thus, they can directly communicate with each other without creating interference in the network. The sink nearer to the BS directly communicates, whereas the farthest sink can communicate with the BS via a closer neighboring sink. Consequently, the farthest sink, after collecting the packets from the associated sensors, sends this data to neighboring sink, which is directly connected to the BS. At the application layer, the user/s may connect to the BS via three different types of network, namely wide area network (WAN), local area network (LAN), and neighborhood area network (NAN). The user/s can directly monitor, control, and configure any deployed sensor node through the BS and sink in the network.

In proposed scheme, we assume the following. First, all the wireless or wired communication links between sinks, sink-to-BS, and BS-to-user/s are bidirectional and highly stable in LAN, NAN, and WAN. Second, we assume that the user/s may connect to the BS via one of the following communication technologies like Cellular, Satellite, or Ethernet in the SG. Third, we also assume that the communication link between sinks and BS could be Cellular, Microwave, or Ethernet. Fourth, the location of each sensor node as well as sinks is known, which can be obtained using a

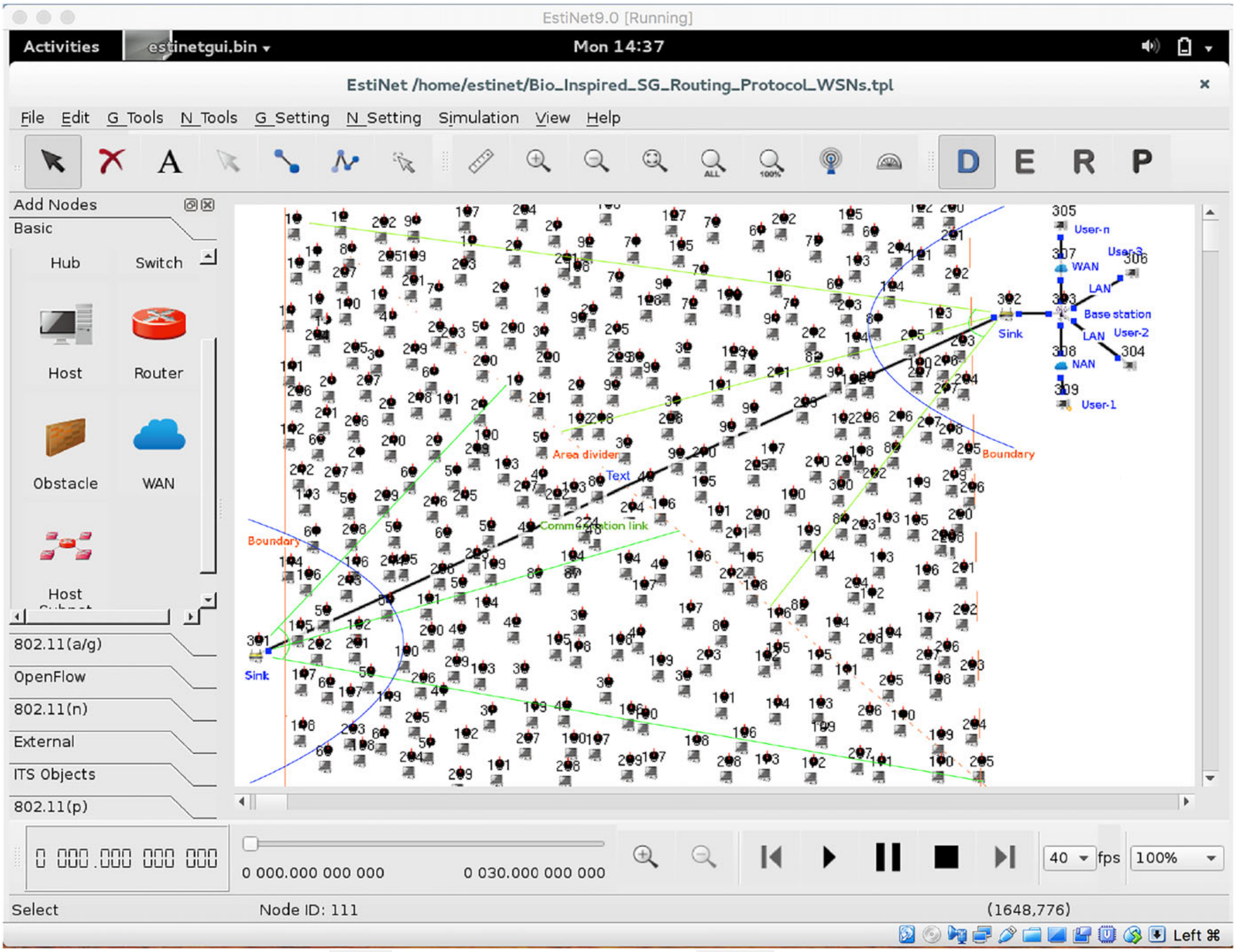

FIGURE 1 Shows the snapshot of our network model taken from EstiNet 9.0.[38] The black color icons are the wireless sensor nodes. The unique number on the right side of each sensor node shows the identity in the network. The devices equipped with dual antennas on the left and right sides of the deployed network are the sinks while the pole like icon is the BS. The orange color lines on the left and right side defined the network boundary while the dotted orange color line in the network divides the entire network into two regions. The green color lines divide each region into subregion. Moreover, green color circular lines show the angle formation due to subregion at each sink in the network. On the other hand, the blue color circular lines show the sink range for message transmission and reception in the network. The black line between sinks, sink-to-BS, and BS-to-user/s shows the highly stable bi-directional communication links. The cloud like icon indicates the network is either a local area network, neighborhood area network, or wide area network

location mechanism.[39] Fifth, the data transmission delay between sink-to-sink, sink-to-BS, and BS-to-user/s is negligible and has no loss of data packets. Sixth, it is assumed that the sinks and BS are rich in data rates, information processing, and energy compared with sensor nodes. Finally, to evade a data packet collision we assume a carrier sense multiple access (CSMA) technique in the network.

4.2 | Bio-inspired computing (BIC)

The BIC is a great and immense source of inspiration to solve real complex problems by developing new algorithms, methods, and techniques in various scientific and engineering domains.[40] The are several bio-inspired techniques, including Genetic Algorithm, Culture Algorithm, Memetic Algorithm, Bird-mating Optimization, and many others.[41] To solve complex problems, a BIC algorithm finds an optimal or a range of near-optimal solutions by employing a perfect balance among its components due to extremely dynamic, diverse, complex, robust, and fascinating phenomenon.[40] This is the thrust behind BIC exploring the new era of computation for solving optimization problems in diverse SG applications.

4.3 | Butterfly mating optimization (BMO)

In BIC, the behavior of butterflies has inspired researchers in the pursuit of solutions for solving routing problems in highly dynamic, complex, and evolvable SG environments.[42] Generally, because of their short lifespan, the communication in butterflies is mainly focusing on mating to produce a new fittest generation in a habitat. Butterflies mate-locating system is important to promote the population of butterflies because the expected lifetime of the butterfly is very short, mostly about 1 to 2 weeks. Males usually fly around the fields searching for females and mate with several females during their lives, but in the most species, the females mate more than once based on some constraints. There are several methods or approaches that have been used by the males of the butterfly for searching and determining whether he has found the right female of his own species. Usually, the mate location strategy varies significantly between butterfly species. Generally, two methods of mate location that males butterfly usually use to find females have been categorized into two main strategies: patrolling and perching behavior.[43] In patrolling, the males butterfly fly above a region where the females butterfly are present and when males found the insect with the same size and color; males fly closer to examine it. However, in patrolling species, the male usually finds the female; moreover, males use color more, and some females are found by the use of female pheromone when a patrolling male approaches within a few meters of the female. In addition, throughout the habitat patrolling species mate at any time of day, even in cold habitat where flight can attend as a heat gain approach. The behavior of patrolling is an active search for receptive females; however, it is not obvious and hard to distinguish from the usual flight over a broader habitat.

On the other hand, in perching approach, the male can survey and intercept passing females by spending a long time sitting on a prominent leaf or on a particular patch. In perching species, perching males use movement in the initial approach to a potential female. Perching species usually mate in limited areas of the habitat, often during only part of the day. Perching behavior may adopt within species that mate in the slight area of the habitat. This is because, investigating a passing female, a perching male often returns to a site nearby the preceding place in the habitat. A perching male, to intercept passing females, adopts a sit-and-wait approach at a specific site. In the field, this behavior is very conspicuous. Territorial behavior is categorized as this behavior. Thus, in patrolling species, males usually are mobile, whereas males in perching species are more immobile. Most butterfly species are able to see all the colors of the visible spectrum.[44] However, males of many butterfly species can switch between territorial perching and no territorial patrolling behavior, and that depends on the density of conspecific males at the encounter site. Moreover, male butterflies may switch between perching and patrolling behavior depending on the fundamental environmental conditions. The switching of male behavior has been taken in terms of the economics of territory defense, with increasing defense cost more than the reproductive benefits of this activity. In addition, males in perching try to exclude other males from their perching site since the males are habitually faithful to their sites. In sum, mostly butterfly species are perchers and the others certain are limited patrollers, whereas in some other species, both behavioral strategies occur. In many species of butterfly, a male butterfly spends most of his active time for finding an appropriate mate/s in a suitable area, and distinctive mate site strategies may have importance for their distribution in their habitat and daily activities. When males encounter females, they approach them in the initial stage of mating.

The butterfly male will look for female butterflies with wings that have the correct color, flight ability, size act, and ventral surface of the female wings, which are important functions that play important role in promoting a mating reaction from males in the early stages. Moreover, ultraviolet patterns on the butterfly wings often enable males to recognize their own species. The females receive ultraviolet patterns reflected by the males based on the distances between them in a habitat. When compared with the farthest one, the nearest one receives more ultraviolet, which can be numerically shown as

$$\mathcal{UV}_{i\rightarrow j} = \mathcal{UV}_i \times \frac{\mathcal{D}_{ij}^{-1}}{\sum_k \mathcal{D}_{ik}^{-1}} \qquad (1)$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n.$$

In addition, when the butterflies get closer to each other, they use an additional method that is pheromones. The butterflies find the females of the same species by releasing their pheromones. In some butterflies species, the pheromone stimulation for a pair of females is not enough to start the copulation, so butterflies start the second stage of mating, which is courtship before copulation. In the courtship pattern, the male releases his pheromone and dances around the female by moving his wings to spread his pheromone to the female's antennae. The courtship can take a few seconds or several minutes, which depend on the species of butterflies, and the female can accept or rejects the male in courtship pattern.

During mating process, the crossover and mutation operator are incorporated into the optimization process to generate a set of fittest butterfly in the habitat. A self-adaptive modified version of the binomial crossover ($\mathcal{C}_r$) and mutation ($\mathcal{M}_u$) operators used in this study can be numerically shown as

$$\mathcal{BF}_i^{\mathcal{G}_k+1} = \mathcal{BF}_j^{\mathcal{G}_k+1} \times (1 - \mathcal{C}_r) + \mathcal{BF}_j^{\mathcal{G}_k} \times \mathcal{C}_r \tag{2}$$

$$\mathcal{C}_r = 0.9 + 0.1 \times \frac{f\left(\mathcal{BF}_j^{\mathcal{G}_k}\right) - f\left(\mathcal{BF}_{i,\text{Best}}\right)}{f\left(\mathcal{BF}_{k,\text{Worst}}\right) - f\left(\mathcal{BF}_{i,\text{Best}}\right)} \tag{3}$$

$$\mathcal{M}_u = \mathcal{BF}(\overline{x})_i^{\mathcal{G}_k+1} = \begin{cases} x_i^{\mathcal{G}_k} + \Delta\left(\mathcal{G}_k x_i^{\alpha} - x_i^{\mathcal{G}_k}\right) & \text{if } r \leq 0.3 \\ x_i^{\mathcal{G}_k} - \Delta\left(\mathcal{G}_k, x_i^{\mathcal{G}_k} - x_i^{\beta}\right) & \text{otherwise} \end{cases} \tag{4}$$

$$\Delta(\mathcal{G}_k, y) = y\left(1 - \alpha^{\left(1 - \frac{\mathcal{G}_k}{\mathcal{G}_{k\,\text{max.}}}\right)^{\gamma}}\right) \tag{5}$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n.$$

Note that the value of $\mathcal{C}_r$ is in [0.1, 0.9] and $\mathcal{M}_u$ is in [0, 0.3], which tends to search the space locally and uniformly and maintain diversity in the generated population. Mating with virgin or nonvirgin females or males can have an effect on the mating behavior of butterfly such as the individual preference and mating costs. Consequently, both the male and female butterflies in the habitat memorize each mating history, eg, distance traveled to find a partner, habitat environment, number of mating, etc. This helps to increase the mating probability, which may result in higher fecundity, longevity, and egg weights. After completing the genomic representation of crossover and mutation, the female butterfly lays eggs onto the gentian plants, which results in a population of caterpillars. These caterpillars grow and suffer from predators and environmental extremes, leading to density problem in the first few generations of a new habitat. Finally, at least one fittest caterpillar is surviving per bud in the new generated population in a habitat numerically can be shown as

$$\mathcal{CP}_{i,\text{Best}}^{\mathcal{G}_k+1} = \begin{cases} \mathcal{CP}_i^{\mathcal{G}_k+1}, & f\left(\mathcal{CP}_i^{\mathcal{G}_k+1}\right) < f\left(\mathcal{CP}_j^{\mathcal{G}_k}\right) \\ \mathcal{CP}_j^{\mathcal{G}_k}, & \text{otherwise} \end{cases} \tag{6}$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n.$$

To ensure at least one caterpillar is surviving per bud via contest competition, we employ a competition function, which numerically can be shown as

$$\mathcal{G}_k(\mathcal{N}_t, \mathcal{P}_i) = \frac{\mathcal{FBF}_i(\mathcal{E}_j)\mathcal{N}_t \cdot \mathcal{P}_i(\mathcal{FB}_j)\mathcal{S}_{\mathcal{E}_j,\mathcal{CP}_i}}{\mathcal{S}_{\mathcal{CP}_i}(\mathcal{FB}_j)} \tag{7}$$

such that

$$\mathcal{S}_{\mathcal{CP}_i}(\mathcal{FB}_j) = \mathcal{FBF}_i(\mathcal{E}_j)\mathcal{S}_{\mathcal{E}_j,\mathcal{CP}_i} \cdot \mathcal{N}_t / \mathcal{P}_i\,(\mathcal{CP}_i) \tag{8}$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n.$$

Lastly, the young caterpillars grow and survive from generalist predators and environmental extremes, and the remaining caterpillars convert to mature butterflies. In each iteration, the fittest butterfly is selected, which can further undergo the mating process after a specific time interval in a habitat, which can be numerically shown as

$$\mathcal{BF}_{i,\text{Best}}^{\mathcal{G}_k+1} = \begin{cases} \mathcal{BF}_i^{\mathcal{G}_k+1}, & f\left(\mathcal{BF}_i^{\mathcal{G}_k+1}\right) < f\left(\mathcal{BF}_j^{\mathcal{G}_k}\right) \\ \mathcal{BF}_j^{\mathcal{G}_k}, & \text{otherwise} \end{cases} \tag{9}$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n.$$

This entire process of mating to produce eggs that turn into larva later into a pupa and finally a beautiful fittest matured butterfly repeats until the mating season is ended. The entire terms used above are explained in the Table 1.

4.3.1 | Terms and concepts

The aforementioned entire butterfly mating optimization phenomenon is implemented to sensor nodes deployed for events monitoring in the SG. Herein, the butterflies are the sensor nodes. The habitat means a network, in which sensor nodes are deployed in an SG environment. The male butterflies represent a set of sensor nodes, which need some sort of information about neighboring nodes or a sink/s in the network. On the other hand, the female butterflies are a set of sensor nodes with the information required by the male butterflies. The behavior of a male or female butterfly is iteratively changing based on above two types of information characteristics in the network. A node in patrolling behavior communicates with specific sensor nodes to share initial information in the network. On the other hand, a sensor node in perching behavior directly communicates with the associated sensor nodes in the network. The search space or prominent leaf represents a specific region where a specific node/s is located in the network. These distinct regions are further divided into subregion to specify the butterflies' activities in the network. The butterfly dancing means that the nearest sensor nodes are free for sharing information, which may be identified using a CSMA mechanism in the network. The pheromones releasing means that a neighboring sensor node with closed distance to a specific node that may have some required information for sharing in the network. In addition, the females/males receive ultraviolet patterns reflected by the males/females based on the Euclidean distance between each pair of sensor nodes in the network. The sensor nodes with minimum distance information have a high probability to share precise information over a highly reliable link compared with the farthest nodes in the network.

The nonvirgin or virgin female or male related to the concept of a sensor node that may or may not has exchanged information with a specific sensor node in the network. If a sensor node already has shared information with a specific sensor node stored in the routing table, then it is called nonvirgin node; otherwise it is virgin. The mating number means that the number of times information is exchanged between sensor nodes. Consequently, the several mating means, a number of times the information is exchanged between sensor nodes are in a specific subregion in the network. Each time, during mating process, the genetic value is represented by 0 and 1 for a sensor node in a chromosome of a female butterfly. The value 0 means a sensor node with no energy or inactive, whereas 1 means a sensor node is active in the network. The courtship can take a few seconds or several minutes, which depend on the species of butterflies and the female means that information exchange between each pair of nodes in the network. Herein, the communication between each pair of sensor nodes is computed in milliseconds in the network. Consequently, each time the shared information between each pair of the sensor node is memorized in the routing table. The newly generated butterflies mean an optimal or a set of near-optimal information available to reach a specific node or a sink in the network. The fittest butterfly means the best next hop node in the routing path from the source toward the sink or a specific node in the network. The predators' attack or ending a lifetime of butterfly means the sensor nodes without energy, which may be because of battery power exhaustion or inaccessible by neighboring sensor nodes due to void problems in the network. This butterfly phenomenon has been modeled as a QoS-aware multihop routing for WSN-based real-time event monitoring applications in the SG. The entire working mechanism of butterfly in term of WSNs for SG applications has been explained in the sections given below.

4.4 | Network initialization

Initially, the deployed sensor nodes do not have any prior knowledge about their neighboring nodes in the network. Therefore, an accurate neighbor discovery is extremely important and challenging for sensor nodes due to varying link quality in complex and harsh nature SG environments. During the network initialization phase, a user via BS initiates a neighbor discovery process in the network. After receiving a user instruction, the base station is responsible to broadcast a predefined number of neighbor discovery messages ($\mathcal{nd\_msg}$) to static sinks in the network. After receiving the neighbor discovery message, each sink starts to broadcast a predefined number of neighbor discovery messages to sensor nodes in its range in the network. Then, after receiving neighbor discovery message from the sinks, each sensor node broadcasts a predefined number of $\mathcal{nd\_msg}$ messages in its range to neighboring sensors in the network. This whole process repeats until the messages are reached to the entire network. The neighbor discovery message contains information about the node identity number, sink identity, position information, remaining energy, and its distance to neighboring sensors as well as to the sender sensor node. After receiving a message, each sensor node fetches the identity of the transmitter node

TABLE 1 Notations used in SIRP

Notation	Description
$\mathcal{UV}_i$	is the ultra-violet pattern emitted by ith male butterfly
$\mathcal{UV}_{i\rightarrow j}$	is the ultra-violet pattern absorbed by jth female butterfly from ith male butterfly
$\mathcal{D}_{ij}^{-1}$	is the Euclidean distance between ith male and jth female butterfly
$\mathcal{D}_{ik}^{-1}$	is the Euclidean distance between ith male and kth female butterfly
$\mathcal{BF}_{i,\text{Best}}$	is the best butterfly individual in butterfly population
$\mathcal{BF}_{k,\text{Worst}}$	is the worst butterfly individual in butterfly population
$\mathcal{BF}_j^{\mathcal{G}_k}$	is the fitness of butterfly j in a subpopulation
$\mathcal{G}_k$	is current generation number
r	is a uniformly distributed random number in $[0, 1]$
$x_i^{\beta}, x_i^{\alpha}$	are lower and upper bounds of the ith component of the decision vector, respectively
α	is a uniformly distributed random number in $[0, 1]$
$\mathcal{G}_{k\,\text{max.}}$	is the maximum number of generations in a habitat
γ	is a parameter, determining the strength of the mutation operator
x, y	are the genomic values of a pair of female and male butterfly, respectively
$\overline{x}$	is the mutated genomic value in a chromosome of a female butterfly
$\mathcal{CP}_{i,\text{Best}}^{\mathcal{G}_k+1}$	is the caterpillar survives for the next generation $\mathcal{G}_k + 1$
$\mathcal{CP}_i^{\mathcal{G}_k+1}$	is the fitness of $i$thcaterpillar
$\mathcal{CP}_j^{\mathcal{G}_k}$	is the fitness of jth caterpillar
$\mathcal{BF}$, $\mathcal{FBF}$	is the butterfly and female butterfly, respectively
$\mathcal{MBF}$	is the male butterfly
$\mathcal{N}_t$	is the adult butterfly population in a specific time t.
$\mathcal{P}_i$	is the number of plants
$\mathcal{P}_i(\mathcal{FB}_j)$	is the number of flower buds in a plant $\mathcal{P}_i$
$\mathcal{CP}_i$	is the number of caterpillars that can survive in a plant $\mathcal{P}_i$
$\mathcal{G}_k$	is the number of generations in a habitat
$\mathcal{FBF}_i(\mathcal{E}_j)$	is the total number of eggs laid by a female butterfly divided by 2 by assuming the 1 : 1 sex ratio
$\mathcal{S}_{\mathcal{E}_j,\mathcal{CP}_i}$	is the proportion of eggs and young caterpillars surviving from density-independent causes of mortality
$\mathcal{S}_{\mathcal{CP}_i}(\mathcal{FB}_j)$	is the number of competing caterpillars per resource unit, ie, flowering bud
$\mathcal{BF}_{i,\text{Best}}^{\mathcal{G}_k+1}$	is the butterfly survives for the next generation $\mathcal{G}_k + 1$
$\mathcal{BF}_i^{\mathcal{G}_k+1}$	is the fitness of ith butterfly
$\mathcal{BF}_j^{\mathcal{G}_k}$	is the fitness of jth butterfly
$\mathcal{X}$	is a binary variable that sets the assignment of the sensor node $\mathcal{S}_i \in \mathcal{L}_i$ to the sink $\mathcal{S}_i \in \mathbb{S}_\kappa$
$\mathcal{Y}$	is a binary variable for the sink such that $\mathcal{S}_i \in \mathbb{S}_\kappa$ in the network
$\mathcal{Z}$	is an integer variable that represents the number of sensor nodes that are assigned to a sink
$\mathcal{B}$	is a binary variable that identifies if $\mathcal{S}_i \in \mathcal{L}_i$ in the route $\mathcal{R}_{\mathcal{P}(k)}$ of a sink $\mathbb{S}_k \in \mathbb{B}\mathbb{S}$ in the network
$\mathcal{F}$	is a b binary variable that identifies the data flow between the intermediate sensor nodes $\mathcal{S}_i$ and $\mathcal{S}_j$ over highly reliable links $\ell_i\left(\mathcal{S}_i, \mathcal{S}_j\right)$ or a sink $\mathbb{S}_\kappa$ uses link $\ell_i\left(\mathcal{S}_i, \mathcal{S}_j\right)$ in the network. It takes the value 1 if a link is active and 0 otherwise
ρ_r, $\mathcal{R}e$	is the probability and the relay node, respectively.
$\mathcal{W}$	denotes the integer variables that represents the number of sensor nodes data packets that are picked up/dropped off each hop sensor node in a subregion in the network
$\mathcal{A}$	shows the array with the number of sensor nodes data packets that are picked up/dropped off at relay sensor node $\mathcal{S}_i$ at location $\mathcal{L}_i$ in the network
$\mathcal{T}_{\max}$	is the maximum defined time
$\mathcal{E}_0$	is the initial energy of the sensor nodes in the network
$\mathcal{E}_1(r)$	is the anticipated amount of energy spent by the one-hop transmission of a sensor node at a distance κ away from the sink in the network

(Continues)

TABLE 1 (Continued)

Notation	Description
$\hbar(\varkappa, r)$	is the anticipated number of hops that sensor's packet need to make to reach the sink in the network
$\mathfrak{R}$	is the radius of a sensor node in the network
$\mathcal{D}_{\mathcal{T}}$	is the total amount of events data collected in each round from the entire subregion in the network
$\mathcal{X}, \mathcal{Y}$	are the integer variables
$\mathbb{B}\mathbb{S}$	is the base station
$\mathcal{D}is_{\mathbb{S}_\kappa}^{\mathbb{S}_i}$	is the distance from a sensor node i to the sink in the network
$\delta_{\mathcal{R}_{\mathcal{P}(i)}}$	is the size of routing path i
$\mathcal{LR}_{\mathcal{P}(i)}$	is the length of routing path i
$\mathcal{R}_{\mathcal{P}(k)}$	is the set of routing paths k from the source and destination in the network.
$\delta_{\mathcal{R}_{\mathcal{P}_{\ell 1(j)}}}$	is the link-path incidence dummy, ie, one if data path j uses lthe ink ℓ_1 and zero otherwise

and updates its routing table. Then, after computing the information, each sensor node replies to the sender node via a reply message (rp_msg) with the information about its identity number, residual energy, location information, and its distance to neighboring sensor nodes that can be estimated using a Euclidean distance formulation in the network.

During the message exchange process, the packet collision may take place when two or more sensor nodes concurrently exchange packets in the same time slot since all the nodes intensively participate in the neighbor discovery process. Therefore, we employ a CSMA mechanism during message exchange in the network. Thus, the collision-free messages are exchanged among sensor nodes since all sensor nodes perform carrier sense before sending a message to neighboring nodes in the entire network. Then, based on the sensor nodes location information and distance to the sink entire network is partitioned into two main regions by the user. The network partition occurred in such a way that the total numbers of nodes are divided equally for each sink to share data traffic burden in the network. Then, each sink sends a sink association message (ass_sink) to its associated sensors in a specific partition region in the network. Note that all the sensors are linked to as a minimum one sink in the network. In the entire message exchange process, each sender sensor node receives an acknowledge message (ack_msg) from the receiver sensors, which guarantees the successfully delivery of the message in the network. Hence, the link between each pair of sensor nodes is bidirectional tested in the network. Finally, the entire network information is forwarded to the user/s by sinks and the BS.

4.5 | Routing mechanism

The harsh SG environments due to power systems' noise, interference, fading, and obstacles make it hard to convey real-time QoS-aware data in a large-scale network. This is because of the time- and space-dependent varying link quality, which hampered the data reliability for WSNs in the SG. Therefore, designing a large-scale WSN-based QOS-aware data gathering scheme for SG applications is challenging. In addition, usually, the density of deployed sensor nodes in the SG is expected to be higher as compared with traditional WSN-based applications. Therefore, the data generation rate is higher and highly depends on the event occurrence frequency in the smart grid. As a result, the sensor nodes buffer overflow time may not be same at various sites in the SG. The sensor nodes generating data at higher rates may be immediately forwarded to the sink to avoid buffer overflow compared with a sensor node generating data at lower rates in the network. Moreover, the excessive delay in data transmission may also cause packet loss because of buffer overflow or invalid data packets due to not reaching to the user in defined time intervals. Hence, the frequent packet forwarding in an efficient and timely fashion is essential from the sensors located in higher data rates regions in the SG. Therefore, the data gathering from the field by using a single sink takes so long time and faces issues like congestion, hotspot, high energy consumption, and invalid data packets due to not reaching at the destination in a bounded time interval for WSN-based smart grid applications. Thus, it might not be a proficient solution for various large-scale WSN-based applications in the SG. To this end, a dual-sink–based data collection mechanism can provide QoS-aware energy-efficient data gathering in both dense and sparse sensor network deployments for various SG applications.

In the proposed scheme, after network initialization, the entire sensor network is partitioned into two main groups based on geographical locations of the sensor nodes to minimize the sensor buffer overflow, data delivery latency, and data traffic load to both sinks. Then, each group is further divided into several subgroups so that the sensor nodes in the same group have the same deadline of packet forwarding to the sink. The subgroups are divided based on the angle information computed by each sink in the network. The minimum and maximum value of the angle is set to 15 and 25 degrees left and right sides of a semicircle. However, the value of each sink angle periodically changes in an individual round of the data collection phase of the proposed scheme. The key aim to future divide the main two groups into subgroups is to minimize the routing path length and unnecessary multihop packet forwarding, ie, data path loops in the network. This feature avoids interference issues and node buffer overflow because of narrow routing packets over a set of sensor nodes with highly stable link quality in the network. In the second phase, the route construction process starts after receiving route discovery message *rd_msg* from the remote user. The sensor nodes that precisely decode the route discovery message start to broadcast a predefined number of route discovery messages in their communication range to the neighboring nodes in the field. The route discovery message contains information about the sender node identity, distance, residual energy, location information, and six single-digit bits information in the network. The key aim of sending these six single-digit bits is to ensure the link quality between each sensor's pair in the network.

The link is considered as a highly stable link when no alteration is found in six single-digit bits computed using a checksum mechanism in the network. On the contrary, the link stability decreases as the number of alterations in bits increases from zero to five. The receiver sensor nodes that decode the route discovery message accurately reply to the sender node via reply message (re_msg). The reply message contains the information about receiving node identity, remaining energy, position, distance, and the number of errors found during receiving six single-digit bits in the network. Then, each sensor node, upon receiving the reply message, marks the sender current location and updates the information in its routing table time if already exists. Otherwise, it creates a new entry for storing the information of each new sender sensor node in the routing table. Note that information are stored in the routing table in decreasing priority, which means that a node with higher link stability, residual energy, and lower distance toward the sink has high priority. If two or more sensor nodes have the same priority, then they are stored in the routing table, opportunistically. In addition, herein, it is possible that a sensor node may receive multiple copies of the same message from the same sender node in the smart grid. In this case, the receiver node simply ignores the rest of the messages. On the other hand, if it receives multiple messages from different sender sensor nodes, then it simply replies to each sender node based on First-Come-First-Serve (FCFS) policy by employing collision avoidance CSMA mechanism. Consequently, during the route discovery process, the routing table information is updated periodically with recent information and a ack_msg is sent to the sender sensor node in the network. This process continues until the nodes closer to the sink are found in the network. At this stage, each sensor node in the network has information to reach the sink through various routing paths established over a set of multi-hops sensors in the network.

In data collection phase, the observed information from the sensor nodes is gathered in both active and passive manners in the smart grid. Initially, a user is responsible to initiate data collection process by sending a predefined member of data collection messages (dc_msg) periodically to the sensor nodes via the base station and sinks in the network. The sensor nodes that correctly decode the data collection messages explore the smart grid environment. Then each sensor node before transmitting the packets first sends a beacon message (bc_msg) to neighboring sensor node to acquire the channel using CSMA mechanism. If a sensor node successful acquires a free channel then it transmits the data over the channel. Otherwise, it continuously monitors the surroundings and store monitoring data in its cache and waits for a predefined amount of time to acquire the channels. This cached data is then immediately transmitted to the associated sink as it acquires the medium. Herein, it is important to note that the sensor nodes send their gathered information directly to associated sink if only it is in the transmission range. The sensor nodes away from the sink start to communicate with neighboring nodes in order to find an appropriate next hop relay sensor nearer to the sink. Herein, it is important to notice that initially a next hop relay sensor node in a routing path is appointed based on its low transmission distance, higher residual energy and link quality in the network. However, after a predefined set of iterations, a data forwarding relay sensor is appointed by considering its low transmission distance, high link quality, buffer overflow time and residual energy in the network. This prior history based self-learning ability of the designed scheme helps to reduce a significant amount of data packets loss due to buffer overflow and corrupted data packets in the network. Thus, during multi-hop packet transmission, each relay node after receiving the information to the associated sensors aggregates the events data based on FCFS policy and forwards this data to next hop relay sensor that is closer to the sink. Consequently, the receiver sensor node repeats the same procedure and forwards it again to the relay sensor in greedy manner that is further nearer to the sink.

Lastly, the aggregated events information is directly uploaded to the sink for user inspections through short distance communication via a relay node in the network. After transmitting the events related information, the sent data is immediately removed from the sensor node cache to allow memory for new events recordings. The next data transfer deadline of the sensor node is immediately updated periodically by considering the previous history stored in the network. Herein, the data transfer deadline means the probability of sending packets in the next time intervals. Then, the updated deadline message ($up\_deadline$) is transmitted to the next hop neighboring nodes in the network. This mechanism avoids the excessive data packets loss that occurs due to sensor buffer overflow and minimizes the data packets forwarding delay since the receiver is aware of receiving data packets in the network. In addition, the path counter at each relay sensor node from the source toward the sink is decremented by 1 until it reaches to 0 for the relay node, which is a single hop away to the sink. On the contrary, at each relay node, the path counter from the sink towards the source is incremented by 1, until its maximum defined value for the relay node, which is a single hop away to the source in the network. Thus, the entire observed information from the sensing field is relayed in a greedy manner over a shortest multihop routing path to data collector sink in the network. Finally, the data collection schedule of each subregion is concatenated to form the whole schedule in a region associated to a sink. The designed scheme due to its self-learning based routing mechanism significantly reduces the data packet collision, data path loops, and delay due to appointing an optimal or near the optimal narrow path routing architecture from the source toward the destination in the network. This further reduces the interference effects on the routing path due to reducing the number of multi-hop sensor links in the network. Moreover, the low distance data transmission improves the overall lifetime of the network.

In designed scheme, the entire routing phenomenon is modeled as mixed integer linear programming in the smart grid. Let a set of the sensor node $\mathcal{S}_i = \{\mathcal{S}_1, \mathcal{S}_2, + \cdots +, \mathcal{S}_n\}$ deployed randomly with their known position $\mathcal{L}_i = \{\mathcal{L}_1, \mathcal{L}_2, + \cdots +, \mathcal{L}_n\}$ in different subregion $\mathcal{R}_i = \{\mathcal{R}_1, \mathcal{R}_2, + \cdots +, \mathcal{R}_n\}$ for monitoring events in the SG. The link and distance between each pair of sensor node are shown by $\ell_i = \left\{\ell_{1(\mathcal{S}_1,\mathcal{S}_2)} + \ell_{2(\mathcal{S}_1,\mathcal{S}_3)} +, \ldots, \ell_{n(\mathcal{S}_n,\mathcal{S}_n)}\right\}$ and $\mathcal{D}_{\mathcal{S}_i,\mathcal{S}_j} = \left\{\mathcal{D}_{\mathcal{S}_1,\mathcal{S}_2} + \mathcal{D}_{\mathcal{S}_1,\mathcal{S}_3} +, \ldots, \mathcal{D}_{\mathcal{S}_n,\mathcal{S}_n}\right\}$, respectively. The routing path from the source toward the destination is indicated by $\mathcal{R}_{\mathcal{P}(k)} = \{\mathcal{R}_{\mathcal{P}1} + \mathcal{R}_{\mathcal{P}2} +, \ldots, \mathcal{P}_{\mathcal{P}n}\}$. In addition, the sum of data collected from the sensors in various times is shown by $\mathcal{D}_i = \{\mathcal{D}_1, \mathcal{D}_2, + \ldots +, \mathcal{D}_n\}$ and $\mathcal{T}_i = \{\mathcal{T}_1, \mathcal{T}_2, + \ldots +, \mathcal{T}_n\}$ for $\forall i = 1, 2, \ldots, n; \forall j = 1, 2, \ldots, n; \forall k = 1, 2, \ldots, n$, respectively, in the network. Consequently, the objective function (ψ) of the proposed scheme is numerically written as

$$\psi = \text{Min} \sum_{\mathcal{S}_i}^{\mathcal{S}_j} \sum_{\mathcal{S}_j}^{\mathbb{S}_\kappa} \left(\mathbb{D}_e + \mathbb{P}_{er} + \mathbb{C}_{on} + \mathbb{D}_a\right)^{i,j,k} + \text{Max} \sum_{\mathcal{S}_i}^{\mathcal{S}_j} \sum_{\mathcal{S}_j}^{\mathbb{S}_\kappa} \left(\mathbb{R}_e + \mathbb{P}_{dr} + \mathbb{T}_p\right)^{i,j,k} \tag{10}$$

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n; \mathbb{S}_\kappa = 2.$$

The key aim of the objective function is to minimize total data gathering cost for monitoring smart grid events in the network. In Equation (10), $\mathbb{D}_e$ is the delay, $\mathbb{P}_{er}$ is the packet error rate, $\mathbb{C}_{on}$ is the network congestion, $\mathbb{D}_a$ is the data traffic load in the network, $\mathbb{R}_e$ is the residual energy, $\mathbb{P}_{dr}$ is the packet delivery ratio, $\mathbb{T}_p$ is the throughput, and $\mathbb{S}_\kappa$ is the sink in the network.

Subject to:

$$\forall i = 1, 2, \ldots, \ell; \forall j = 1, 2, \ldots, m; \forall k = 1, 2, \ldots, n \sum_{\mathcal{S}_i \in \mathcal{S}_n} \mathcal{X}_{\mathcal{S}_i(\mathcal{L}_j)}(\mathbb{S}_\kappa) \geq 1. \tag{11}$$

Constraint (11) guarantees that the sensors deployed in a smart grid environment are associated to as a minimum one sink in the network.

$$\mathcal{R}_i(\mathcal{S}_j) \in \mathbb{S}_\kappa = 1. \tag{12}$$

Constraint (12) states that each sensor belongs to a region is assigned to a sink in the network.

$$\mathcal{X}_{\mathcal{S}_i(\mathcal{L}_i)} \leq \left(\mathcal{Y}_{\mathcal{S}_i(\mathcal{L}_i)}(\mathbb{S}_\kappa)\right)_{\mathcal{R}_j} = 1. \tag{13}$$

Constraint (13) assures that each sensor in a subregion is assigned to exactly one sink in the network.

$$\mathcal{X}_{\mathcal{S}_i(\mathcal{L}_i)}(\mathbb{S}_\kappa) \leq \mathcal{Y}_{\mathcal{S}_j(\mathcal{L}_j)}. \tag{14}$$

Constraint (14) states that the sensor associated to a single sink are less than the total number of sensor nodes in the network.

$$\rho_r(\mathbb{S}_\kappa)_{\mathcal{R}_i} = 1 - \prod_{i=1}^{n} \left(1 - \rho_r(\mathcal{S}_i)\right). \tag{15}$$

Equation 15 indicates the coverage probability of a sink to its member sensors located in a region in the network.

$$\mathcal{D}_T = \sum_{k=1}^{2} \mathbb{S}_\kappa(\mathcal{D}(\mathcal{S}_i))^j. \tag{16}$$

Constraint in Equation 16 specifies that a sink gathers the data from its member sensors in each round in the network.

$$\mathcal{Z}_{\mathcal{S}_i} \leq \sum_{\mathcal{S}_i \in \mathcal{R}_1} \mathcal{X}_{\mathcal{R}e(\mathcal{S}_i)} \left(\mathcal{R}_{\mathcal{P}(\mathscr{k})}\right), \quad \mathcal{Z}_{\mathcal{S}_i} \leq \mathcal{Y}_{\mathcal{R}_{\mathcal{P}(\mathscr{k})}} \times \mathfrak{max}\left(\mathcal{R}e\left(\mathcal{S}_i\right)\right), \mathfrak{max}\left(\mathcal{R}e\left(\mathcal{S}_i\right)\right) < \mathcal{S}_j. \tag{17}$$

Constraint (17) assures that the number of relay sensors in a routing path must not greater than the maximum defined value in a region in the network.

$$\mathcal{Y}_{(\mathcal{S}_i,\mathcal{S}_j)}\ell_{(\mathcal{S}_i,\mathcal{S}_j)} \leq \left|\mathcal{R}_{\mathcal{P}(\mathscr{k})}\right| \mathcal{X}_{(\mathcal{S}_i,\mathcal{S}_j)}\ell_{(\mathcal{S}_i,\mathcal{S}_j)}, \forall i, j \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}. \tag{18}$$

Constraint (18) states that information flow can only occurs if a highly stable link between each pair of sensor is on the routing path.

$$\sum_{\mathcal{S}_i \in \mathcal{R}\mathcal{P}_{\mathscr{k}}, \mathcal{S}_i \neq \mathcal{S}_j} \left(\mathcal{X}_{(\mathcal{S}_i,\mathcal{S}_j)}\ell_{(\mathcal{S}_i,\mathcal{S}_j)}\right)^{\mathcal{R}_i} \geq 1, \forall j \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}. \tag{19}$$

$$\sum_{\mathcal{S}_j \in \mathcal{R}\mathcal{P}_{\mathscr{k}}, \mathcal{S}_j \neq \mathcal{S}_i} \left(\mathcal{X}_{(\mathcal{S}_i,\mathcal{S}_j)}\ell_{(\mathcal{S}_i,\mathcal{S}_j)}\right)^{\mathcal{R}_i} \geq 1, \forall i \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}. \tag{20}$$

Both constraints (19) and (20) ensure that each sensor in the routing path must have at least one link pointing away from it and pointing toward it in the network.

$$\sum_{\mathcal{S}_j \in \mathbb{S}_\kappa} \mathcal{F}^{\mathbb{S}_\kappa}_{\mathcal{S}_j \in (\mathcal{L}_j)} = \sum_{\mathcal{S}_i \in \mathbb{S}_\kappa} \mathcal{F}^{\mathbb{S}_\kappa}_{\mathcal{S}_i \in (\mathcal{L}_i)} = 1 \tag{21}$$

$$\sum_{\mathcal{S}_j \in \mathbb{S}_\kappa} \mathcal{F}^{\mathbb{S}_\kappa}_{\ell_1(\mathcal{S}_i,\mathcal{S}_j)} - \sum_{\mathcal{S}_i \in \mathbb{S}_\kappa} \mathcal{F}^{\mathbb{S}_\kappa}_{\ell_1(\mathcal{S}_j,\mathcal{S}_i)} = 0. \tag{22}$$

Both constraints (21) and (22) assure that one unit of data must be forwarded to the sink over a set of highly stable links between sensors in the network.

$$\sum_{\mathcal{S}_j \in \mathcal{R}\mathcal{P}_{\mathscr{k}},\ \mathcal{S}_j \neq \mathcal{S}_i} \mathcal{F}\ell_{1(\mathcal{S}_i,\mathcal{S}_j)} = \sum_{\mathcal{S}_j \in \mathcal{R}\mathcal{P}_{\mathscr{k}},\ \mathcal{S}_i \neq \mathcal{S}_j} \mathcal{F}\ell_{1(\mathcal{S}_j,\mathcal{S}_i)}. \tag{23}$$

Constraint (23) confirms that the information must flow continuously among the intermediate sensors in the network.

$$\sum_{\mathcal{S}_j \in (\mathcal{L}_j)} \mathcal{B}_{\mathcal{S}_j(\mathcal{D})_{\mathscr{k}}} \cdot \mathbb{S}_\kappa\left(\mathcal{L}_\kappa\right) = \sum_{\mathcal{S}_i \in (\mathcal{L}_i)} \mathcal{B}_{\mathcal{S}_i(\mathcal{D})_{\mathscr{k}}} \cdot \mathbb{S}_\kappa\left(\mathcal{L}_\kappa\right). \tag{24}$$

Constraint (24) ensures the data flow continuity towards the sink from the sensors in the network.

$$\rho_{\mathrm{r}}\left(\mathcal{R}_{\mathcal{P}(i)}\right) = \mathcal{R}_{\mathcal{P}(j)} + \ln\ \mathcal{D}_n\left(i\right) \Big/ \sum_{j \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}} \mathcal{R}_{\mathcal{P}(j)} + \ln \mathcal{D}_n\left(j\right). \tag{25}$$

Equation (25) provides the probability of data packets from the source toward the destination traveled through a data path i in the network. The value of $\mathcal{D}_n(i) = 1$ and $\mathcal{D}_n(i) = 0$ indicate that a data route i exist and the sensor nodes are unaware of a data path or path is unavailable, respectively. That is, the log-of-zero limit is negative infinity, whereas for an unavailable alternative, the exponent of utility is zero in the network.

$$\delta_{\mathcal{R}_{\mathcal{P}(i)}} = \sum_{j \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}} \frac{\ell_{1(\mathcal{S}_i,\mathcal{S}_j)}}{\mathcal{L}\mathcal{R}_{\mathcal{P}(i)}} \frac{1}{\sum_{j \in \mathcal{R}_{\mathcal{P}(\mathscr{k})}} \left(\frac{\mathcal{R}_{\mathcal{P}(i)}}{\mathcal{R}_{\mathcal{P}(j)}}\right)^{\gamma} \delta_{\mathcal{R}_{\mathcal{P}\ell 1(j)}}}. \tag{26}$$

Equation (26) indicates that the $\ell_{1(\mathcal{S}_i,\mathcal{S}_j)}$ is the length of link 1 and γ is a constant with value 1. This ensures that the data paths with different length are generated in the network.

$$\sum_{\mathcal{S}_j,\mathcal{S}_i \in \mathbb{S}_\kappa} \mathcal{F}^{\mathbb{S}_\kappa}_{\ell_1(\mathcal{S}_i,\mathcal{S}_j)} \mathcal{R}_{\mathcal{P}(\mathscr{k})}(\mathcal{D})_{\mathscr{k}} = \sum_{\kappa=2} \mathbb{S}_\kappa^{(\mathcal{D})_{\mathscr{k}}} = \sum_{\mathbb{B}_{\mathbb{S}}=1} \mathbb{B}_{\mathbb{S}}(\mathcal{D})_{\mathscr{k}}. \tag{27}$$

Constraint (27) guarantees that the sum of data packets received by the sink and the base station are identical to the data packets forwarded by a source sensor over a set of sensor nodes along a routing path in the network.

$$\sum_{\mathcal{S}_i,\mathcal{S}_j \in (\mathcal{R}_{\mathcal{P}})} \mathcal{R}_{\mathcal{P}(\mathscr{k})} \left(\ell_{1(\mathcal{S}_i,\mathcal{S}_j)} + \ell_{1(\mathcal{S}_j,\mathcal{S}_i)}\right) \leq 1. \tag{28}$$

Constraint (28) assures that the intermediate sensors avoid data path loops during forwarding information in the network.

$$\sum_{\mathcal{S}_j \in (\mathcal{L}_j)} \mathcal{R}e_{\mathcal{S}_j}(\mathcal{D})_{\mathscr{k}} > \sum_{\mathcal{S}_i \in (\mathcal{L}_i)} \mathcal{S}_i\left(\mathcal{L}_i\right) \cdot (\mathcal{D})_{\mathscr{k}}. \tag{29}$$

Constraint (29) assures that each relay sensor ($\mathcal{R}e$) carries more data packets than a normal sensor in the network.

$$\mathcal{W}_{\mathbb{S}_j}(\mathcal{L}_i) = \sum_{\mathbb{S}_j \in (\mathcal{L}_j)} \mathcal{A}_{\mathbb{S}_j(\mathcal{D})_k} \cdot \mathcal{B}_i\left(\mathbb{S}_j\right) \leq 1. \tag{30}$$

Constraint (30) assures that the sum of data packets received by each sensor from a set of associated sensors is not higher than its maximum buffer size in the network.

$$\sum_{\mathbb{S}_i \in (\mathcal{L}_i)} \mathcal{E}_{\mathbb{S}_i}(\mathcal{R}_j) \leq \mathcal{E}_0. \tag{31}$$

Constraint (31) assures that the sensor node energy consumption location in a region must not exceed than its initial energy in the network.

$$\mathbb{B}_{\mathbb{S}}(\mathcal{T}_k) - \mathbb{S}_\kappa\left(\mathcal{T}_j\right) - \mathcal{R}e_{\ell_i(\mathbb{S}_i,\mathbb{S}_j)}(\mathcal{T}_i) \leq \mathcal{T}_{\max}. \tag{32}$$

Constraint (32) ensures that the maximum data collection time of the BS from the sink and relay sensors located in a subregion should not exceed the maximum defined time in the network.

$$\mathcal{D}is_{\mathbb{S}_\kappa}^{\mathbb{S}_i} = \min \frac{2}{\mathfrak{R}^2} \int_0^{\mathfrak{R}} \mathcal{E}_1(\mathrm{r})\, \hbar(\varkappa, \mathrm{r})\, \varkappa \mathrm{d}\varkappa. \tag{33}$$

Equation (33) shows the minimum distance and energy consumption to the sink from the associated sensors in the network. The terms used above are defined in Table 1.

4.6 | Route recovery and new node joining

In designed scheme, each sensor node is capable of finding a new or inactive neighboring node in the SG. Initially, after deployment, a new sensor node starts to broadcast a predefined number of neighbor discovery messages in its range in a subregion in the SG. This neighbor discovery message contains the information such as node position information, remaining energy, and identity in the network. After receiving the message accurately, a sensor node creates a new entry for storing the information of the new sender node based on decreasing priority in the routing table. Then, it sends a reply message to new sensor node with the information such as node identity, distance, residual energy, location, and six single-digit bits to the sender node in the network. On receiving the reply message, the new sensor node creates a routing table and stores the information of neighboring sensor nodes with decreasing priority based on FCFS mechanism. Then, an acknowledgment message is sent by new sensor node equipped with the six single-digit bits information to the neighboring sensors in the SG environment. Herein, the acknowledgment message has twofold purposes. First, it guarantees successfully delivery of the messages, and second, it estimates the link quality between each pair sensors in the harsh nature SG environment. Consequently, each receiver sensor node, by using a checksum mechanism, verifies each bit of the message and updates information in the routing table with decreasing priority and sends a joining message (jo_msg) to the new node in the network. Thus, at this stage, the new sensor node located in a subregion is the part of the network. This entire information is forwarded to user/s located in a remote area by the sensors located in a subregion in the network. In a subregion, the route construction process starts again in a case if the new sensor behaves as a data packet forwarding node (ie, relay node) for its associated sensor node/s in the network. The entire route construction process is same as discussed above in detail (see Section 4.5).

In designed scheme, each sensor node in a subregion periodically sends alive messages (al_msg) to its neighboring sensor nodes to discover an inactive or dead sensor node in the network. If a neighboring sensor node in a predefined amount of time does not reply to the sender node during monitoring cycles, then it is assumed to be dead in the network. This updated information is then forwarded to the remote user. After receiving information, a user sends a neighbor discovery message only to the neighboring sensor nodes belonging to a subregion to confirm the sensor; either it is active or inactive in the network. After receiving the messages accurately, the neighboring sensors start to communicate with the inactive sensor node based on its location information by sending a predefined number of neighbor discovery messages in a subregion in the network. The location information of the assumed inactive sensor node could be found in the routing table since it is updated periodically in each round in the network. After a predefined amount of time, if at least one neighboring node does not get any reply message from the inactive sensor node, then it is assumed to be dead, and updated information is sent to the user. Then, based on the information received from the neighboring sensor nodes, the user limited broadcasts a dead node message (de_msg) only to the associated sensors in a subregion in the network. In a subregion, the route construction process starts again in a case if the inactive sensor was serving as a packet forwarding node to its associated node in previous rounds in the network. The data path construction procedure is same as discussed above in detail (see Section 4.5).

5 | PATH LOSS MODEL AND SIMULATION PARAMETERS

The path loss model given in the work of Gungor et al[45] is used because it provides more precise channel estimation results compared with Nakagami model. The path loss model numerical can be illustrated as

$$\gamma(d)_{dB} = P_t - PL(d_0) - 10\eta\log_{10}\frac{d}{d_0} - X_\sigma - P_\eta. \quad (34)$$

In addition, the performance evaluation of our designed scheme is measured by using a discrete-event network simulator, namely EstiNet9.0,[38] with the recent published bio-inspired routing scheme known as Energy and Quality Aware Spectral Honey bee inspired Clustering (EQSHC)[28] for SG applications. In fact, for real-time event monitoring applications, we compute the performance of our designed scheme in terms of PDR, which is measured as the ratio of generated packets in the network to the packets received successfully at the sink. The data packet error rate is computed as the number of corrupted packets to the sum of packets received by the user/s in the network. The memory overflow data loss occurs when a sensor node receives data packets from the neighboring sensor nodes higher than its maximum defined capacity in the network. The congestion is the ability of sensor nodes to handle data traffic efficiently to avoid the bottleneck in the network. The efficient memory usage is the characteristic of the sensors to manage and utilize their limited cache for managing data packets to provide robust routing in the network. Throughput is measured as the sum of bits processed in a defined time.[46] Usually, it is computed as bits per second (bits/sec). From the source toward the destination, the remaining energy after a successful packet transmission is known as residual energy, whereas the control message overhead is the sum of messages required to control the entire routing process in the network. Herein, it is important to note that the sum of energy consumed during data aggregation, idle listening, to keep sensor radios on, and data and control packets reception and transmission is known as the energy consumption in the network. Finally, the time difference when a data packet is sent by the source node and received at the sink along the shortest path is known as delay. In designed scheme, in order to provide consistent results, we employ batch mean method[47] for each data point with 96% confidence intervals in the network. In our simulation study, a total of 500 sensor nodes were used for real-time SG event monitoring applications. The value for total number of male butterfly and female butterfly sensor nodes were set to 200 and 300 from a set of 500 sensor nodes in the network. In addition, a set of near-optimal fittest broods associated to a butterfly after mating process were set to 5 in the network. The detail of simulation and path loss model parameters value used in this study is given in Table 2.

TABLE 2 Simulation and path loss model parameters value and description

Simulation Model Parameters	Values	Path loss Model Parameters	Values
Initial sensor node energy	5 J	Path loss at a reference distance d_o	PL (d_O)
Maximum hop distance	80 meters (m)	Signal to noise ratio	γ(d)
High transmission power	0.93Watts (W)	Transmit power in dB	P_t
Low transmission power	0.81 W	Zero-mean Gaussian variables with standard deviation	X_σ
Packet receiving power	0.05 W	Path loss exponent	η
Ideal listening	0.023 W	Non–line of sight	NLOS
Sleeping power	3×10^{-6}W	Noise power measured in dBm	P_η
Data aggregation	0.015 W	Path loss exponent (n) for both line of sight (LOS), NLOS	2.4, 3.5
Packet length	43 bytes	Noise floor for both LOS, NLOS	−87, −91
Topology	Random	Shadowing deviation (σ) for both LOS, NLOS	3.12, 2.92
Sensor data rate	250 kbit/second (s)	Number systems/subsystems in a grid	100
Physical layer standard	802.11 g	Number of electric grid towers	320
Set of simulations	53	The maximum distance between sinks	2 km
Area: 2D (length × width)	(1600 × 800) m	The maximum distance between nearest sink and BS	500 m

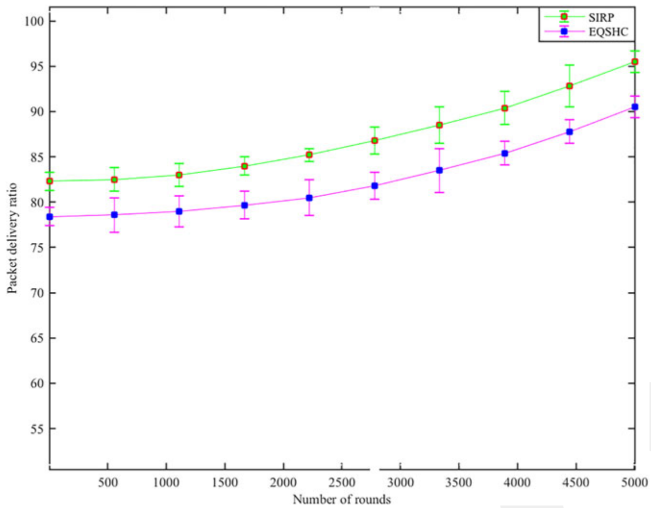

FIGURE 2 Shows the packet delivery ratio vs number of round between 1 and 5000

6 | PERFORMANCE ANALYSIS

Figure 2 shows that the total number of data packets delivered by each routing scheme to a user for real-time monitoring events in the SG. It clearly shows that the PDR in both routing schemes increases with the increase in network size between round numbers 1 and 5000. The PDR decreases rapidly when sensor nodes start to die in the ending rounds between 4000 and 5000 in the SG. The rate of decreasing PDR is observed more in EQSHC than SIRP scheme in the SG. The PDR of SIRP is reliant on offered events data traffic load and delivers information around 96.08% of the total initiated data packets in the SG. On the contrary, the EQSHC delivers events data up to 91.12% of the total initiated data packets to a user in the SG. The simulation studies reveal that the PDR in both routing schemes depends on the path selection that is relatively stable, least congested, and in which the intermediate sensor nodes hold massive residual energy in the SG. Moreover, it also depends on the number of intermediate nodes along a routing path since the excessive number of sensor nodes increases the probability of memory-overran issues and corrupted data packets as revealed in Figures 3 and 4, respectively. This memory-overrun brings congestion management issues for relaying sensors nodes near to the sink. Figure 5 clearly shows that the proposed scheme performs well in terms of congestion management compared with EQSHC routing scheme in the SG. In addition, the PDR in both routing schemes is also affected by the excessive delay, which directly depends on the network congestion along a data path from the source toward the destination in the SG. Based on the simulation facts, we reveal a number of possible reasons that might cause in decreasing the PDR in EQSHC

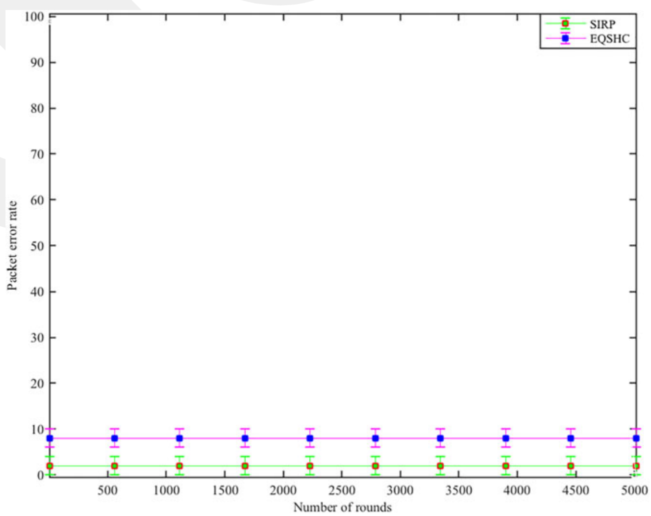

FIGURE 3 Indicates the packet error rate vs number of rounds between 1 and 5000

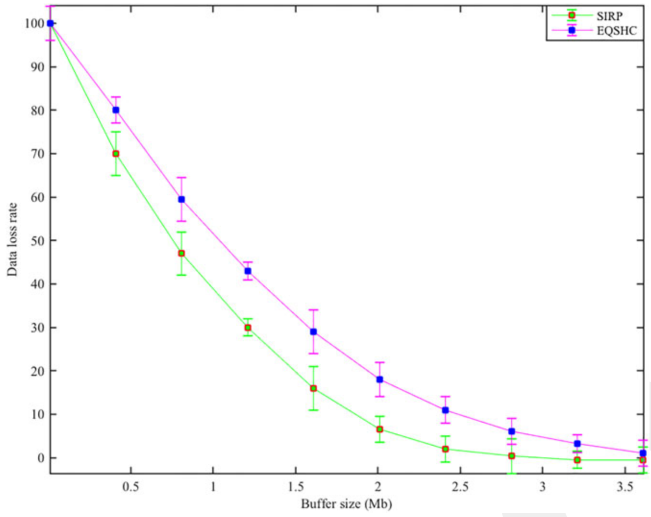

FIGURE 4 Depicts data loss rate vs node buffer size from 1 to 3.6 Mb

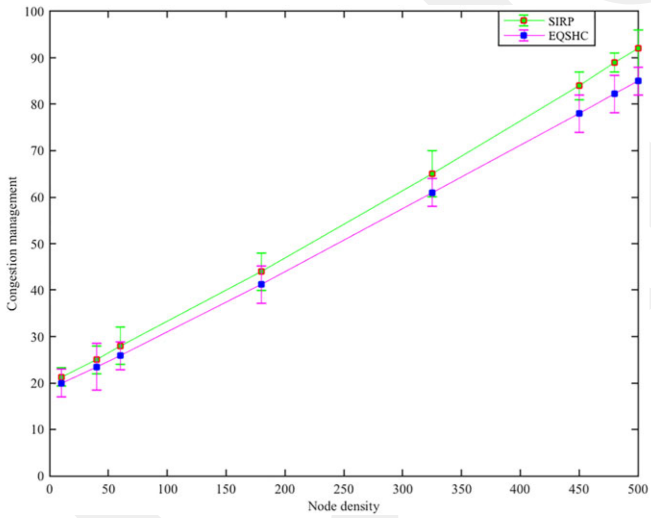

FIGURE 5 Expresses the congestion management vs number of sensor nodes between 1 and 500

routing scheme for real-time monitoring in the SG. First, we observe that the increase in sensing time decreases the false alarm and misdetection probability with the expense of higher probability of false alarm, which directly affects PDR in EQSHC in the SG. Second, most of the time, route failures in EQSHC result in low PDR due to frequent rerouting in the network. This may possibly increase the time spent by witnessing the data packets in the cache of the originated sensor node because the packets need to wait until a routing path is established in the network. Third, due to the lack of using limited sensor node buffer size efficiently in EQSHC increases the probability of data packet loss in the network. Figure 4 clearly shows that as compared with EQSHC, the SIRP routing scheme performs better and significantly decreases the packet loss rate in the network. Moreover, it also clearly shows that no packet loss is observed when the cache size is about 3.24 Mb in the network. At the same time, the packet loss rate was observed nearly 8.7% in EQSHC of the originated data traffic in the network. Figure 6 clearly shows that EQSHC most of the time fails to utilizes limited sensor node memory efficiently and thus results in data packet collisions, which directly affect the total packet delivery ratio.

Fourth, the superfluous multihop data transmission in EQSHC may help increase the coverage range of sensor network; however, it causes excessive interference which increases the corrupted data packets in EQSHC compared with SIRP during monitoring an event as shown in Figure 3. This reduces the packet delivery ratio and results in low network throughput in EQSHC compared with SIRP in the SG. The network throughput profile of both EQSHC and SIRP routing schemes is depicted in Figure 7. In EQSHC, the network throughput directly depends on the arrival rate of primary and secondary user's activity in the network. The rate of secondary user's arrival to take the control channel for the desired

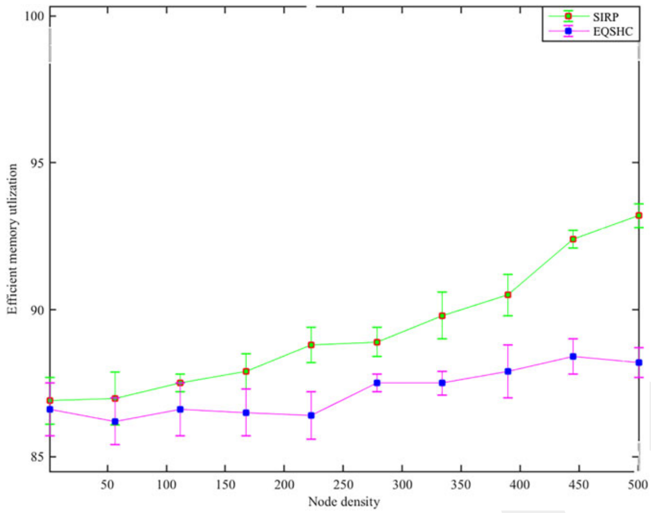

FIGURE 6 Illustrates the efficient memory utilization vs number of sensor nodes between 1 and 500

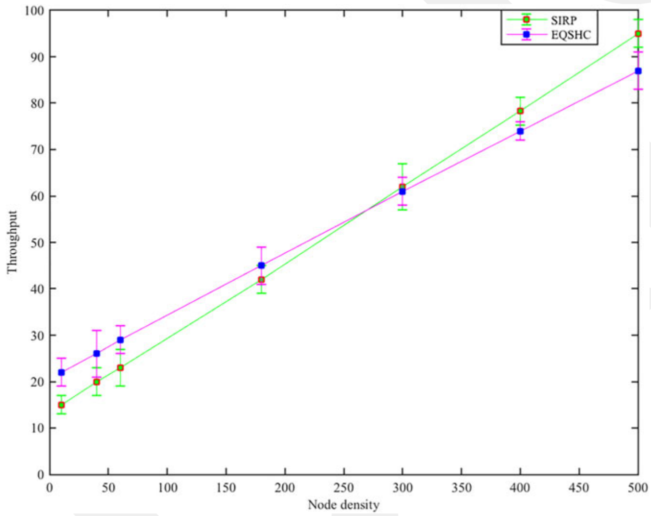

FIGURE 7 Shows the network throughput vs number of sensor nodes between 1 and 500

throughput is found enormously slower in EQSHC routing scheme in the network. We observed a couple of reasons, which may possibly be caused to decrease network throughput in high EQSHC during monitoring events rivaled to SIRP in the SG. First, in EQSHC, the secondary connection data delivery time decreases due to not considering an appropriate ideal probability of the available channel in the network. Second, in EQSHC, the data delivery time of the primary connections is higher compared with the secondary connections, which lead to lower throughput, and increases the chance of tainted data packets. Third, in EQSHC, with the increasing primary users, the probability of arrivals of primary users increases in the transmission period of the secondary user, in turn, increasing the probability of interference, which leads to a decrease in secondary user throughput more during monitoring events in the SG. Unlike EQSHC in SIRP, the data delivery time increases because of taking into account the highly stable link between each pair of sensor nodes and thus results in high network throughput than EQSHC routing scheme in the SG.

A detailed view of the network delay profile is shown in Figure 8. It clearly shows that the SIRP routing scheme achieves lower network delay compared with EQSHC during forwarding packets from the source towards the destination in the SG. In EQSHC, most of the time, due to the excessive intermediate hop counts, successful transmission becomes more improbable on the selected routing path for during monitoring events in the SG. Herein, we expect that a certain number of packets become invalid because of not reaching a defined threshold time from the source toward the sink in the network. During simulation studies, we noticed that this is due to increasing the route length caused by a large number of intermediate hops compared with the SIRP in the SG. Studying the simulation facts, it is revealed that our proposed

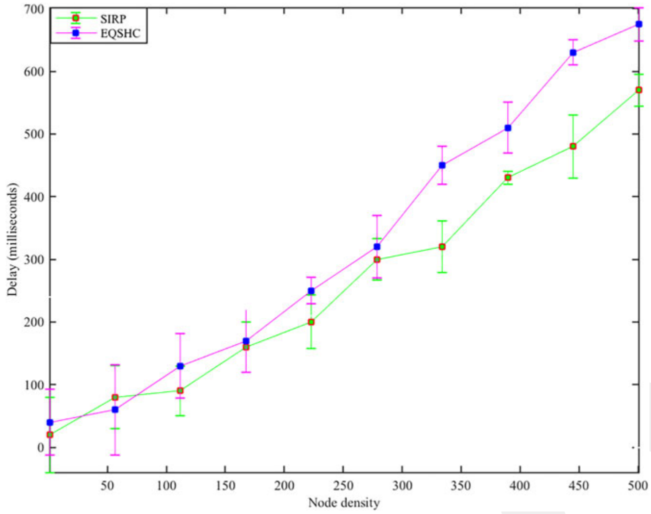

FIGURE 8 Indicates the network delay vs number of sensor nodes between 1 and 500

scheme uses near-optimal shortest paths from a source toward the sink in the network. Therefore, the designed scheme leads to low network delay when compared with EQSHC for monitoring events in the SG as revealed in Figure 8. Note that the delay in EQSHC during monitoring events is relatively superior when network size is small around 50 sensor nodes in the SG. In addition, this lower delay behavior of SIRP is found more robust in large-size networks due to its unique delay constraint requirements. One of the main reasons for the excessive delay in EQSHC is due to its cumulative hand-off delay during monitoring events in the SG. Due to cumulative handoff delay in EQSHC, secondary users need to spend excessive waiting time while changing its operating channel due to the high busy probability of the selected target channel. This handoff latency increases with the increase in time in EQSHC in the SG. Generally, when the distance between each pair of sets of nodes is too low, it means that the receiver and transmitter nodes are very near to each other, which might cause more intermediate hops count as in EQSHC. Sometimes, it may also lead to data paths looping in EQSHC, which degrades the advantage of shortest-path routing in the network.

A detailed view of the residual energy profile of SIRP and EQSHC routing schemes is revealed in Figure 10. It clearly shows that the residual energy profile of SIRP is superior compared with EQSHC routing scheme in the SG. This low energy consumption leading to high network lifetime in SIRP for long-lasting events monitoring compared to EQSHC in the SG. In the designed scheme, the entire sensor network is partitioned into two main groups based on geographical locations of the sensor nodes to minimize the sensor buffer overflow, data delivery latency, and data traffic load to both sinks. Then, each group is further divided into several subgroups so that the sensor nodes in the same group have the same deadline of packet forwarding to the sink. These subgroups are divided based on the angle information computed by each sink in the network. This minimizes the routing path length and unnecessary multihop packet forwarding, ie, data path loops in the network. Moreover, this feature avoids interference issues and node buffer overflow due to narrow routing packets over a set of sensor nodes with stable link quality in the network. In addition, a highly stable link between sensors is computed by exchanging six single-digit bits in the SG environment. Thus, the cached data are immediately transmitted to the associated sink as it acquires the medium in the SG. During the data forwarding phase, a next-hop relay sensor, node along a routing path, is appointed based on its low transmission distance, higher residual energy, and link quality in the network. However, after predefined iterations, a next-hop data forwarding sensor node is appointed by considering its low transmission distance, high link quality, residual energy, and buffer overflow time in the network. This self-learning ability of the designed scheme that is based on the prior history helps reduce a significant amount of data packets loss and corrupted data packets in the network. Thus, during multihop packet transmission, each relay node, after receiving the information from associated sensors, aggregates the information based on FCFS policy and forwards this data to next-hop data forwarding sensor node nearer to the sink. This multihop data forwarding in a narrow routing path toward the sink extremely minimize nodes energy consumption and balance the overall sensors remaining energy in the network.

In addition, the highly stable and reliable communication links between each pair of sensor node minimize the frequent route failures in SIRP that result in low control message overheads compared with EQSHC as shown in Figure 9. Thus, a notable sum of node energy is saved for extending the lifespan of the sensors in SIRP as sown in Figure 10. It clearly shows that the control message overheads in SIRP are significantly less compared with EQSHC routing scheme in the SG.

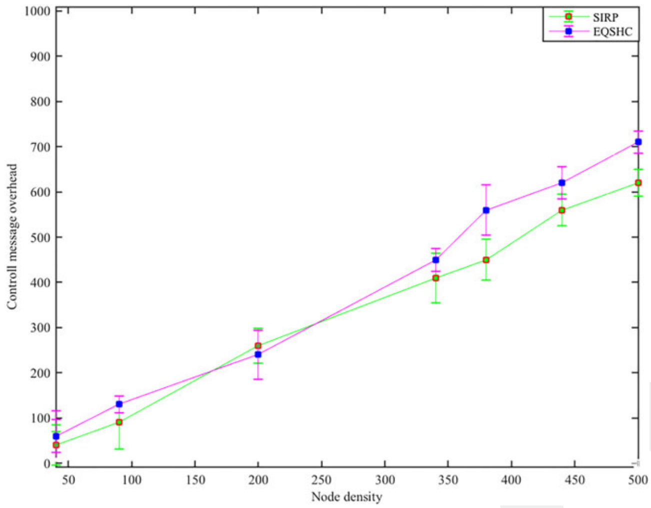

FIGURE 9 Illustrates the control message overhead vs number of sensor nodes between 1 and 500

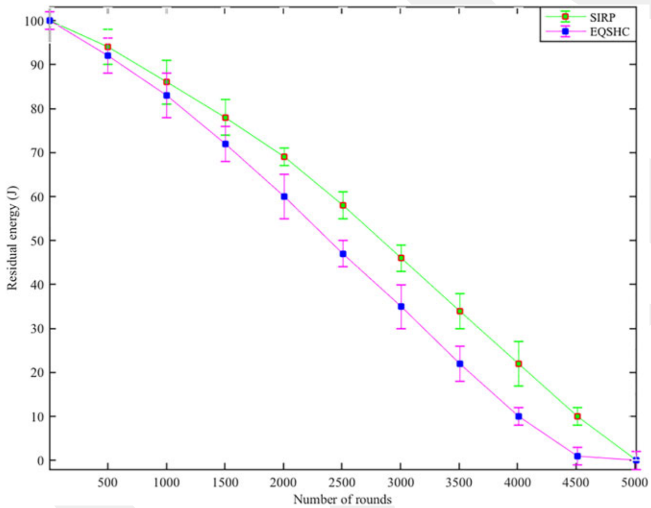

FIGURE 10 Depicts the sensor node residual energy vs number of rounds between 1 and 5000 in the network

Therefore, it is extremely appropriate for sparse and highly dense sensor networks for the SG. In SIRP, the data collection mechanism has a very small communication overhead since prior knowledge about neighbor existence is required with extremely low control message retransmission in the network. On the other hand, in EQSHC, the need for more data collection control messages leading to more packets collisions and consequently more retransmissions. Consequently, the event-related aggregated data are directly uploaded to the sink for user inspection through short-distance communication by a relay sensor in the network. After transmitting information, the sent data are immediately removed from the sensor node cache to free memory for new event recordings. The next data transfer deadline of the sensor node is immediately updated periodically by considering the previous history in the network. This mechanism avoids the excessive data packets loss that occurs due to sensor buffer overflow and minimizes the data packets forwarding delay since the receiver is aware of receiving data packets in the network.

The designed scheme also takes the benefit of shortest path multihop routing and appoints the near-optimal number of hops by taking into account the minimum and maximum coverage range values from the source toward the destination in the SG. This avoids excessive multihop and data path looping in SIRP even in highly dense and sparse deployment in the SG. The designed scheme, due to its self-learning–based routing mechanism, significantly reduces the data packet collision, data path loops, and delay due to appointing an optimal or near the optimal narrow path routing architecture between a source and destination in the network. Moreover, the low distance data transmission over a set of relay nodes in greedy manner improves the overall network lifetime in the smart grid. Herein, the $O(n)$ and $O(\mu \cdot n)$ indicate the overall computed time complexity of the routing table and designed scheme, respectively. On the other hand, the overall computed time complexity of the EQSHC routing scheme is found as $O\left(\mu \cdot n^2\right)$ in the network. The routing algorithm computational time complexity is linear to μ. Herein, n and μ are the number of sensors information stored in the routing table and number of initializations, respectively. In the case of a route breakdown due to a single or multiple nodes failure, a next-hop data forwarding relay sensor with a secondary high value in the routing table is selected to convey data packets than EQSHC in the network. Thus, it significantly saves sensor node energy by reducing the impact of communication overheads to increase network lifespan than all other routing mechanism in the SG.

7 | CONCLUSION

Inspired by the fourth industrial revolution and growing day-to-day energy demand, the conventional power grids are currently being transformed into SGs. To this end, IoT and IWSNs are envisioned as key promising communication technologies for numerous SG applications to empower the SGI 4.0. However, despite several benefits, they have several unique challenges caused by fading, multipath effects, equipment noise, and electromagnetic interference at different protocol stack layers when deploying for SG monitoring and control applications in SGI 4.0. This results in hampering the two-way QoS communication requirements for wireless sensor networks-based SG applications. To tackle these challenges, this article proposed an innovative self-optimized intelligent butterfly mating optimization data gathering multihop routing scheme for wireless sensor networks-based smart grid applications. The designed bio-inspired routing scheme offers highly reliable multihop communication network architecture for QoS-aware information gathering for WSN-based SG applications in SGI 4.0. The simulations extensively performed through a network simulator called EstiNet 9.0 have revealed that the proposed scheme accomplishes its defined goals compared with existing routing scheme designed for WSN-based applications with the expense of data redundancy in the network. As a future work, the researchers may introduce a novel communication framework to guarantee diverse QoS-aware data gathering with least data redundancy for various WSN-based SG applications. In addition, the experimental evaluation of the proposed protocols in WSN testbeds is an important future research direction.

ACKNOWLEDGMENT

The work of Muhammad Faheem was supported by the International Doctoral Fellowship (IDF) program from UTM, the work of Md. Asri Ngadi was supported by the Research Grant from UTM, and the work of V.C. Gungor was supported by the TUBA-GEBIP under grant V.G./TUBA-GEBIP/2013-14.

ORCID

M. Faheem http://orcid.org/0000-0003-4628-4486
M. W. Ashraf http://orcid.org/0000-0003-1591-7041

REFERENCES

1. Mrugalska B, Wyrwicka MK. Towards lean production in Industry 4.0. *Procedia Eng*. 2017;182:466-473.
2. Oliff H, Liu Y. Towards Industry 4.0 utilizing data-mining techniques: a case study on quality improvement. *Procedia CIRP*. 2017;63:167-172.
3. Sung TK. Industry 4.0: a Korea perspective. *Technol Forecast Soc Change*. 2018;132:40-45.
4. Majeed AA, Rupasinghe TD. Internet of things (IoT) embedded future supply chains for Industry 4.0: an assessment from an ERP-based fashion apparel and footwear industry. *Int J Supply Chain Manag*. 2017;6:25-40.
5. Saleem Y, Crespi N, Rehmani MH, Copeland R, Hussein D, Bertin E. Exploitation of social IoT for recommendation services. Paper presented at: 2016 IEEE 3rd World Forum on Internet of Things; 2016; Reston, VA.
6. Arshad S, Azam MA, Rehmani MH, Loo J. Information-centric networking based caching and naming Schemes for Internet of Things: A survey and future research directions. 2017. arXiv preprint arXiv:1710.03473.
7. Zikria YB, Afzal MK, Ishmanov F, Kim SW, Yu H. A survey on routing protocols supported by the Contiki internet of things operating system. *Future Gener Comput Syst*. 2018;82:200-219.
8. Saleem Y, Crespi N, Rehmani MH, Copeland R. Internet of things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions. 2017. arXiv preprint arXiv:1704.08977.
9. Guerreiro BV, Lins RG, Sun J, Schmitt R. Definition of Smart Retrofitting: First steps for a company to deploy aspects of Industry 4.0. In: Hamrol A, Ciszak O, Legutko S, Legutko M, eds. *Advances in Manufacturing*. Cham, Switzerland: Springer; 2018:161-170.

10. Ishmanov F, Bin Zikria Y. Trust mechanisms to secure routing in wireless sensor networks: current state of the research and open research issues. *J Sens*. 2017;2017:1-16.
11. Ghafoor S, Rehmani MH, Cho S, Park S-H. An efficient trajectory design for mobile sink in a wireless sensor network. *Comput Electr Eng*. 2014;40:2089-2100.
12. Amjad M, Sharif M, Afzal MK, Kim SW. TinyOS-new trends, comparative views, and supported sensing applications: a review. *IEEE Sensors Journal*. 2016;16:2865-2889.
13. Umer T, Amjad M, Afzal MK, Aslam M. Hybrid rapid response routing approach for delay-sensitive data in hospital body area sensor network. Paper presented at: 7th International Conference on Computing Communication and Networking Technologies; 2016; Dallas, TX.
14. Faheem M, Abbas MZ, Tuna G, Gungor VC. EDHRP: energy efficient event driven hybrid routing protocol for densely deployed wireless sensor networks. *J Netw Comput Appl*. 2015;58:309-326.
15. Faheem M, Gungor VC. Energy efficient and QoS-aware routing protocol for wireless sensor network-based smart grid applications in the context of Industry 4.0. *Appl Soft Comput*. 2018;68:910-922.
16. Bukhari SHR, Rehmani MH, Siraj S. A survey of channel bonding for wireless networks and guidelines of channel bonding for futuristic cognitive radio sensor networks. *IEEE Commun Surv Tutor*. 2016;18:924-948.
17. Rehmani MH, Reisslein M, Rachedi A, Erol-Kantarci M, Radenkovic M. Integrating renewable energy resources into the smart grid: recent developments in information and communication technologies. *IEEE Trans Ind Inform*. 2018.
18. Gungor VC, Sahin D, Kocak T, et al. A survey on smart grid potential applications and communication requirements. *IEEE Trans Ind Inform*. 2013;9:28-42.
19. Shah GA, Gungor VC, Akan OB. A cross-layer QoS-aware communication framework in cognitive radio sensor networks for smart grid applications. *IEEE Trans Ind Inform*. 2013;9:1477-1485.
20. Mu J. A minimum physical distance delivery protocol based on ZigBee in smart grid. *EURASIP J Wirel Commun Netw*. 2014;2014:108.
21. Yaacoub E, Abu-Dayya A. Automatic meter reading in the smart grid using contention based random access over the free cellular spectrum. *Comput Netw*. 2014;59:171-183.
22. Xiang M, Bai Q, Liu W. Trust-based adaptive routing for smart grid systems. *J Inf Process*. 2014;22:210-218.
23. Lin J, Zhu B, Zeng P, Liang W, Yu H, Xiao Y. Monitoring power transmission lines using a wireless sensor network. *Wirel Commun Mob Comput*. 2015;15:1799-1821.
24. Zhang X, Ye F, Fan S, Guo J, Xu G, Qian Y. An adaptive security protocol for a wireless sensor-based monitoring network in smart grid transmission lines. *Secur Commun Netw*. 2016;9:60-71.
25. Yang Z, Ping S, Aijaz A, Aghvami A-H. A global optimization-based routing protocol for cognitive-radio-enabled smart grid AMI networks. *IEEE Syst J*. 2018;12(1):1015-1023.
26. Deng X, He L, Li X, Liu Q, Cai L, Chen Z. A reliable QoS-aware routing scheme for neighbor area network in smart grid. *Peer-To-Peer Netw Appl*. 2016;9:616-627.
27. Faheem M, Gungor VC. Capacity and spectrum-aware communication framework for wireless sensor network-based smart grid applications. *Comput Stand Interfaces*. 2017;53:48-58.
28. Fadel E, Faheem M, Gungor VC, et al. Spectrum-aware bio-inspired routing in cognitive radio sensor networks for smart grid applications. *Comput Commun*. 2017;101:106-120.
29. Hui W, Zhitao G, Tingting Y, Yue X. Top-k query framework in wireless sensor networks for smart grid. *China Commun*. 2014;11:89-98.
30. Erol-Kantarci M, Mouftah HT. Wireless sensor networks for cost-efficient residential energy management in the smart grid. *IEEE Trans Smart Grid*. 2011;2:314-325.
31. Farooq H, Tang Jung L. Energy, traffic load, and link quality aware ad hoc routing protocol for wireless sensor network based smart metering infrastructure. *Int J Distributed Sens Netw*. 2013;9:597582.
32. Hou R, Wang C, Zhu Q, Li J. Interference-aware QoS multicast routing for smart grid. *Ad Hoc Netw*. 2014;22:13-26.
33. Kim K, Jin S-Ii. Branch-based centralized data collection for smart grids using wireless sensor networks. *Sensors*. 2015;15:11854-11872.
34. Islam MT, Koo I. Compressed sensing-based multi-layer data communication in smart grid systems. *KSII Trans Internet Inf Syst*. 2013;7(9):2213-2231.
35. Li X, Liang Q, Lau FC. A maximum likelihood routing algorithm for smart grid wireless network. *EURASIP J Wirel Commun Netw*. 2014;2014:75.
36. Faheem M, Gungor VC. MQRP: mobile sinks-based QoS-aware data gathering protocol for wireless sensor networks-based smart grid applications in the context of Industry 4.0-based on internet of things. *Future Gener Comput Syst*. 2018;82:358-374.
37. Junaid M, Ahmad S, Ali QW. Smart grid (WAMS) for transmission line through GSM. *Int J Electr Comput Sci*. 2013;13:42-47.
38. EstiNet Technologies Inc. 2018. http://www.estinet.com
39. Pandey S, Varma S. A range based localization system in multihop wireless sensor networks: a distributed cooperative approach. *Wirel Pers Commun*. 2016;86:615-634.
40. Dressler F, Akan OB. A survey on bio-inspired networking. *Comput Netw*. 2010;54:881-900.
41. Farooq M, Di Caro GA. Routing protocols for next-generation networks inspired by collective behaviors of insect societies: An overview. In: Blum C, Merkle D, eds. *Swarm intelligence*. Berlin, Germany: Springer; 2008:101-160.
42. Jada C, Vadathya AK, Shaik A, Charugundla S, Ravula PR, Rachavarapu KK. Butterfly mating optimization. In: Berretti S, Thampi S, Srivastava PR, eds. *Intelligent Systems Technologies and Applications*. Cham, Switzerland: Springer, 2016:3-15.

43. Arora S, Singh S. Butterfly optimization algorithm: a novel approach for global optimization. *Soft Comput*. 2018:1-20.
44. Wu H, Liu C, Li B, Xie X. A novel binary butterfly mating optimization algorithm with subarray strategy for thinning of large antenna array. *Prog Electromagn Res*. 2017;60:101-110.
45. Gungor VC, Lu B, Hancke GP. Opportunities and challenges of wireless sensor networks in smart grid. *IEEE Trans Ind Electron*. 2010;57:3557-3564.
46. Afzal MK, Kim BS, Kim SW. Throughput enhancement in cooperative wireless ad hoc networks. Paper presented at: 2014 IEEE International Conference on High Performance Computing and Communications (HPCC), 2014 IEEE 6th International Symposium on Cyberspace Safety and Security (CSS) and 2014 IEEE 11th International Conference on Embedded Software and Systems (ICESS); 2014; Zhangjiajie, China.
47. Fishman GS, Yarberry LS. An implementation of the batch means method. *INFORMS J Comput*. 1997;9:296-310.

How to cite this article: Faheem M, Butt RA, Raza B, et al. Bio-inspired routing protocol for WSN-based smart grid applications in the context of Industry 4.0. *Trans Emerging Tel Tech*. 2019;30:e3503. https://doi.org/10.1002/ett.3503