Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018)

Routing for Efficient Alarm Aggregation in Smart Grids: A Genetic Algorithm Approach

Sérgio Sabino[a,b], António Grilo[a,b]

[a] *Instituto Superior Técnico-Universidade de Lisboa, Avenida Rovisco Pais, Lisboa 1049-001, Portugal*
[b] *INESC-ID, Rua Alves Redol, Lisboa 1000-029, Portugal*

Abstract

We propose a Genetic Algorithm based routing scheme to maximize the number of alarm messages received at the sink node in a Wireless Sensor Network embedded in an electric power grid upon a fault/outage detection. In this scenario, sensor nodes rely on their scarce energy storage to operate and report alarm messages to the sink. Sensed data in neighboring nodes tend to be correlated. Therefore, nodes will send redundant data to the sink while wasting energy and congesting the network. In order to mitigate the energy waste and increase the message delivery rate, a data aggregation scheme is proposed.

Keywords: Wireless Sensor Networks, Smart Grid, Data Aggregation, Routing, Genetic Algorithms.

1. Introduction

A Smart Grid corresponds to the modern concept of an electric power grid infrastructure, with improved efficiency, reliability, and safety, achieved through intelligent management control, enabled by the seamless integration of Information and Communication Technologies (ICT)[1,2]. Unlike the traditional power grid, which generates and distributes electricity following a rigid hierarchical architecture comprising three isolated subsystems (generation, transmission and distribution) and an unidirectional flow of electricity, Smart Grids enable bi-directional energy and information flows to provide automated and more efficient energy delivery[3]. Smart Grids demand secure and accurate information exchange for the grid management and control to prevent possible disruption in the electrical system due to unexpected failures. Therefore, a highly reliable, scalable, robust and cost-effective communication network between substations and the remote control center is vital[2].

* Corresponding author. Tel.: +351-969-356-209 ; fax: +351-213-145-843.
E-mail address: sergio.sabino@tecnico.ulisboa.pt

1877-0509

Recently, Wireless Sensor Networks (WSN)s have been widely recognized as a technology promising to improve various aspects in Smart Grids, namely those that deal with power generation, bidirectional delivery and seamless monitoring, providing an energy efficient, reliable and low-cost solution for control and management[4].

This study proposes a Genetic Algorithm (GA) based routing scheme capable of maximizing the number of alarm notifications received at the sink node. These alarms are trigged upon detection of a fault (e.g., short circuit followed by a grid shutdown) in the grid, being used by monitoring systems for fault detection and localization. Usually, sensed data of neighboring nodes tend to be correlated. Therefore, nodes will send redundant data to the sink while wasting energy and potentially congesting the network. To tackle this issue, we propose an aggregation scheme, which operates simultaneously with the GA based routing. The aggregation scheme takes into account the spatial correlation of the alarm notifications in the electric power grid.

The remainder of this article is structured as follows: In Section 2, related work on routing protocols in WSNs based on GA are presented. Section 3 presents the network model. Section 4 presents the proposed data aggregation scheme. Section 5 presents the proposed GA based routing scheme and operators. Section 6 presents the simulation settings and results. Finally, Section 7 concludes the article.

2. Related work

Traditionally, routing protocols in WSNs are classified based on the network architecture (data-centric, hierarchical and location-based) or depending on the protocol operation (multipath-based, query-based, negotiation-based, QoS-based, or coherent-based)[5]. In this work we focus on routing protocols which use meta-heuristic algorithms such as GA and data aggregation mechanisms in WSNs.

Bari et al.[6], propose a GA scheme in a two-tiered sensor network architecture, where higher powered special nodes, named *relay* nodes, act as cluster heads, while sensor nodes transmit their data directly to their respective cluster heads. The objective function aims to find a scheduling for data gathering of relay nodes, which can significantly extend (maximize) the network lifetime in each data gathering round. The relay nodes can be used as cluster heads in hierarchical sensor networks and can be provisioned with higher energy as compared to the sensor nodes.

Gupta et al.[7] proposed a centralized GA based routing protocol scheme named GAR (Genetic Algorithm-based Routing) which considers the energy efficiency of the relay nodes. The GAR finds out the route from all the relay nodes to the base station (BS) by minimizing their overall distances with the view that the consumed energy to transmit the diffused data is proportional to the square of the distance between sender and receiver. Recently, Gupta et al.[8] proposed a GA scheme for multipath routing in WSNs which is also energy efficient. The objective function takes into account the distance between the sender and receiver, the distance between next hop node to the BS and on the number of hops to send the data from the next node to the BS. What mainly distinguishes our work from the previous ones is that we have considered a flat network of homogeneous nodes in the field, i.e., there are no cluster heads, and the aggregation is performed by each intermediate node according to defined aggregation policies. Our model considers a fixed link length between the sender and the receiver, though the path length from each source to the sink may vary.

Shiobara et al.[9] proposed an aggregation scheme which uses two methods to aggregate data. The first is the *combining method*, which aggregates messages and combines all of them without modifying any data, then sends them to the upstream node as one large message. The intermediate nodes remove all individual headers and includes a single header for the large aggregated message. The second is the *manipulating method*, which reveals the calculation results such as average, maximum, minimum, and summation values. This reduces the total size of the messages. However, the completeness of all messages is lost because they are compressed by the calculation. Our proposed aggregation scheme is based on those aforementioned methods, though with some differences on how they are performed as will be described along this work.

3. Network model

The physical topology of the WSN is modeled as an undirected connected graph $G(\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of wireless nodes, and $\mathbf{E}$ is the set of edges representing a wireless connectivity between two nodes considering that they are within communication range of each other. The logical topology is overlaid on the physical topology, representing the data routing path in the communication network. It is modelled as a directed acyclic tree subgraph $G'(\mathbf{V}, \mathbf{E}')$, where $\mathbf{E}' \subseteq \mathbf{E}$ represents a set of $N-1$ edges of the tree subgraph G'. The following assumptions apply:

- All nodes are static, and placed at known locations in the branches of the electric power grid;
- There is only one centralized sink node acting as the root, which is represented by node $v_0 \in \mathbf{V}$ with unlimited energy capacity E_s, and all other nodes in the communication network $\{v_1, v_2, v_3 ... v_N\} \subseteq \mathbf{V}$, have equal initial energy capacity E_i, $i = 1, 2, ...N$, when operating autonomously (i.e., relying on their supercapacitors);
- There is no clustering, i.e., flat routing is assumed. Nodes rely on each other to deliver the messages to the sink in a multihop fashion.
- The radio model for transmission and reception of b-bit data messages between a sender and receiver node separated by distance r is given by Equations 1 and 2 [10], respectively:

$$E_T(b, r) = (E_{T_x} + E_{amp} * r^{\beta}) * b \tag{1}$$

$$E_R(b) = E_{R_x} * b \tag{2}$$

  Where, $E_T(b, r)$ and $E_R(b)$ are the total energy dissipated to transmit and receive b-bit of the data message, respectively. The parameters E_{T_x} and E_{R_x} in Equations 1 and 2, are the amount of energy dissipated to run the radio electronics upon transmission and reception, respectively. E_{amp} is the energy required by the transmitter amplifier to maintain an acceptable signal-to-noise ratio (SNR) in order to transfer b-bit of the data messages reliably. β is called path-loss exponent, which may assume a value of 2 or 4, depending on the distance to the receiver.
- In WSNs nodes share a common channel to report their collected data to the sink. Nodes at distance r from each other interfere in their communication, therefore, collisions and data loss are prone to occur. The packet error probability (P_r) for b-bit data message is given by Equation 3:

$$P_r = 1 - (1 - BER)^b \tag{3}$$

  Where BER is the Bit Error Rate. Based on Equation 3 the corresponding average number of transmission attempts including successful transmission is computed as follows:

$$\tau = \frac{1}{1 - P_r} \tag{4}$$

Regarding the routing layer, it is assumed that routing paths are calculated and disseminated during normal operation of the grid. Upon the occurrence of a fault in the grid, during which power flow is interrupted, the nodes already know the next hops toward the sink, to which the fault reports will be issued.

4. Data aggregation

Data aggregation in WSNs is mostly used to reduce redundant data transmission across the network. In this work, the proposed aggregation scheme assumes that event detections in an electrical grid are spatially correlated along a distribution feeder. The feeder topology is considered to form a tree connecting all nodes in $\mathbf{V}$, whose model was described in Section 3. In our proposed scheme, aggregation takes place only between nodes deployed sequentially along the feeder. Hence, the feeder coincides with the aggregation tree, where the sink node acts as the root and intermediate nodes are responsible for aggregating data based on topological correlation of measurement data. If the electric power grid topology is known, it can be used to establish an ordered relationship between the positions of nodes along the feeder branches, with each node being either upstream or downstream from each of its neighbor. We can also represent a group of cascading node ranges in a feeder branch, from the downmost (most distant from the sink) to the uppermost (closest to the sink) node. In this case, the latter identifiers are enough to represent all the range, since the ordered relationship is well known by the sink node. This avoids explicitly including the identifiers of all the detecting nodes in data messages when they have detected the same event. In practice, this model could be used in Smart Grids monitoring, with nodes reporting voltage and current level measurements or issuing fault detection and location reports [11].

4.1. Data message structure

We propose a data message structure as depicted in Figure 1. It is composed by a header with two fields of 2 octets each, and a data field of variable length. The message fields are defined as follows:

- **T**: Identifies the type of event that has been trigged (e.g., warnings, fault/outage alarm notifications, etc.).
- **L**: Is the total number of **Node ranges** (see Figure 1) which compose the content of the message;
- **Node range**: Represents the range of succeeding nodes in each branch. It is composed by two fields of 2 octets each. The first one, **Unode**, stores the identifier of the uppermost node in the branch, while the other, **Dnode**, stores the downmost node identifier. Note that the Data field is composed by **L Node range** items.

We use Equation 5 to compute the data message size S_v in octets. S_v is the analytical representation of the data message size depending on the size of **L** at each node in the field. The sink node is expected to receive $\lambda * S_v$, where, λ is the number of children directly connected to it.

$$S_v = 4 * (1 + L) \tag{5}$$

4.2. Aggregation model

Essentially, an aggregation tree is built during the aggregation process, where correlated data is compressed to avoid network congestion and unnecessary transmission. We use the previously defined data message structure and Equation 5 to represent the data at each node, assuming that the routes have been previously defined. We designate the nodes with the highest number of hops from the sink and which have detected a common event, as *aggregation leaves*. We have modeled the aggregation process considering the following rules:

- R_1: Nodes which have data to transmit should fill the **Node range** with their unique node id: **Unode**=**Dnode**=localID.
- R_2: We consider that $\lambda_{v_i} \geq 0$ at each node. Leaf nodes ($\lambda_{v_i} = 0$) can only perform R_1 and forward the message. Intermediate nodes, upon a reception of a message from their children, should perform the aggregation using the combining method, if $\lambda_{v_i} > 1$, otherwise use manipulation method and forward the aggregated result upstream. Differently from the usual aggregation methods[9], intermediate nodes do not compute the average, maximum, minimum and summation. Instead, they fill the **Node range** fields as follows: **Dnode** will contain the downmost node identifier along the feeder branch and **Unode** will contain its own identifier as the uppermost node in the branch. Figure 2 illustrates the aggregated messages at each node.

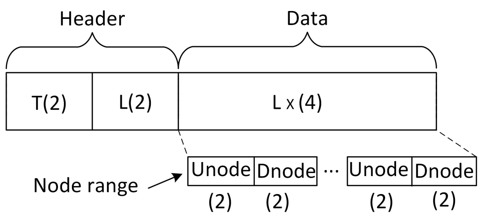

Fig. 1. Data message structure.

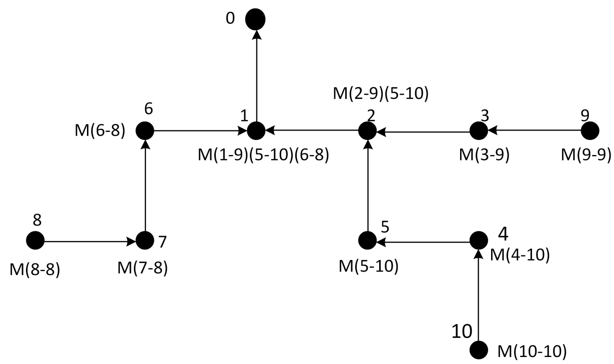

Fig. 2. Network topology of WSN, showing data message aggregation process.

4.3. Waiting Time (WT)

Waiting time (WT) is the time frame when intermediate nodes should wait for their children to hand over their data message to be aggregated. The WT for an intermediate node is computed based on the number of hops towards the sink and the number of predecessor nodes in the aggregation tree. Nodes with different number of hops towards the sink should have different WT. The difference is called minimum cascading time φ[12]. We denote by Δt the fixed amount of time to aggregate data at each intermediate node. Variables h_{max} and h_v are the maximum hop count in the network and the number of hops from the current node to the sink, respectively. We compute the minimum cascading time φ_v at a node using Equation 6.

$$\varphi_v = (h_{max} - h_v)\Delta t \tag{6}$$

The estimated time to transmit a data message is calculated as follows:

$$Tx_v = \frac{8\tau S_v}{Bw} \tag{7}$$

Where Bw is the bit rate of the transmission medium in bit/s. Let us assume that in Figure 2 all nodes have detected the same event. Node v_{10} is an *aggregation leaf*, and defines the h_{max}=5. Table 1 summarizes the expected WT and transmission time for the network depicted in Figure 2.

Table 1. Waiting Time and expected transmission duration for a given node according to the Figure 2.

h_{max}= 5					
Level	Node v	$\varphi_v[s]$	L_v	WT_v $[s]$	Tx $[s]$
5	10	0	1	0	$8\tau(4 + L_{10})/Bw$
4	4	Δt	1	Tx_{10}+φ_4	$8\tau(4 + L_4)/Bw$
	8	Δt	1	φ_8	$8\tau(4 + L_8)/Bw$
	9	Δt	1	φ_9	$8\tau(4 + L_9)/Bw$
3	3	2Δt	1	Tx_9+φ_3	$8\tau(4 + L_3)/Bw$
	5	2Δt	1	Tx_4+φ_5	$8\tau(4 + L_5)/Bw$
	7	2Δt	1	Tx_8+φ_7	$8\tau(4 + L_7)/Bw$
2	6	3Δt	1	Tx_7+φ_6	$8\tau(4 + L_6)/Bw$
	2	3Δt	2	Tx_3+Tx_5+φ_2	$8\tau(4 + L_2)/Bw$
1	1	4Δt	3	Tx_2+Tx_6+φ_1	$8\tau(4 + L_1)/Bw$

5. GA based routing approach and operators

It is assumed that the routes are centrally computed at the sink node using a GA, and disseminated to the sensor nodes during normal grid operation, using an appropriate protocol (the details of such protocol are out of scope of this paper). The routing algorithm seeks to find an optimal balance between using the default aggregation tree and the quality of the wireless links. In order to maximize the chances of reducing the message size, the message data should follow as much as possible the hierarchy of nodes defined by the aggregation tree. Whenever the next hop of an upstream data message does not correspond to the preassigned parent node in the aggregation tree topology, the new parent simply uses the combining method, otherwise performs aggregation according to the aggregation rules from Section 4.2.

5.1. Chromosome and Fitness function

The chromosome is represented as a string of size N, where, N is equal to the total number of nodes in the WSN, including the sink. Each index position of the string represents the unique ID of a node in the network. The value stored at each position (index) corresponds to the ID of the subsequent node (parent) towards the sink. We use this chromosome to represent the network traffic flow. For instance, the network in Figure 2 could be represented as depicted in Figure 3.

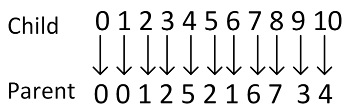

Fig. 3. Chromosome representing WSN.

We compute the fitness function as the total number of reported data messages at the sink node as follows:

$$Maximize\ Z = \sum_{k=1}^{\lambda} \sum_{j=1}^{L_k} \alpha_j \tag{8}$$

where λ is the number of children directly connected to the sink, L_k is the number of **Node ranges** at child k, and α_j is the number of combined or manipulated data messages at each **Node range** j. Note that the higher the value, the fitter is the chromosome.

$$\alpha_j = \begin{cases} |h_{j_{Unode}} - (h_{jDnode} + 1)| \ , & j_{Unode} \neq j_{Dnode} \\ 1, & \text{Otherwise} \end{cases} \tag{9}$$

6. Simulation setup and simulation results

In this Section, we present the results of simulations performed with the proposed GA. The algorithm is implemented in **Python** 2.7.x programming language. We have used Eclipse as the development environment. A successful transmission is modeled probabilistically using Equation 4. The simulated deployment topologies consist of square matrices of different sizes, ranging from 4X4 up to 10X10 . For each of these matrix sizes, random tree feeder topologies were overlaid spanning all the nodes with the sink as root. The sink is placed in different locations in the matrix. We have applied the GA scheme working simultaneously with the aggregation model in a Smart Grid scenario, where each node is placed at known location along the distribution feeders for fault detection and location. In this case, the squared communication grid topology maps the node placement in the electric power grid. Nodes inside the communication grid are only able to communicate with their vertical and horizontal immediate neighbors and have the ability of issuing alarm notifications upon anomaly detection. In the considered scenario, the worst case will be considered, whereby an alarm situation will be detected by all nodes in the network. During the normal operation of the electric power grid, nodes are energized by the electric power grid itself. At the moment of the fault, all nodes are assumed to have a fixed and residual amount of initial energy in their energy storage (i.e., supercapacitor), except the sink, which is considered not to be power constrained. This residual energy, allows them to operate autonomously for a short time.

As already mentioned, nodes rely on each other to deliver the alarm notifications in a multihop fashion. Some alarm notifications may not reach the sink, due to node failure (some nodes may be drained) along the path. We expect that the proposed GA based routing scheme operating simultaneously with the aggregation scheme reduces the energy waste and improves the alarm notifications delivery ratio.

6.1. Simulation setup

The performance of the GA is affected by parameters such as the population size, the rate of crossover and mutation, and the method of replacement[6]. A small population size may lead to premature convergence before reaching an acceptable solution. On the other hand, if the population size is too large, it leads to unnecessary computations. The rates of crossover and mutation are closely related: if set too low, the GA will converge earlier before the optimal solution is found, and if too high, it may cause stability problem in the population. Therefore, exhaustive simulations where performed to determine a suitable GA parameter set to run the GA based routing approach. In this study we use

Table 2. GA parameters

Number of generations	100
Population size	80
Crossover probability (p_c)	0.7
Mutation probability (p_m)	0.4

tournament selection to pick the best chromosomes to the next generation. Single point crossover is performed with the probability (chance) p_c, and during the mutation, a node is randomly chosen and assigned a new parent based on a probability p_m. Table 2 presents the resulting parameter set used in the simulations.

The communication and interference range between two different nodes is fixed at $r = 50m$. The radio model parameters are set as follows [10]: The path loss exponent is constant, β=2, $E_{T_x} = E_{R_x} = 50\ nJ/bit$, $E_{amp} = 100\ pJ/bit/m^2$. For the intermediate nodes, the energy during data aggregation and the processing duration are set as 5 $nJ/bit/signal$ and Δt=0.001s respectively. For the choice of initial energy E_i, we performed simulations to find a suitable energy amount where nodes could struggle to deliver data messages. Obviously, if nodes are supplied with infinite energy storage, all messages will reach the sink at some point. In the end, E_i was chosen to be $0,2mJ$.

6.2. Results

The performance of the network was evaluated in five different configurations: static routing along the distribution feeder, GA routing with aggregation, static routing along the distribution feeder without aggregation, GA routing without aggregation and Dijkstra algorithm with aggregation. In our implementation, the Dijkstra algorithm minimizes the hop lengths of the routing paths. The network size is defined by those aforementioned different matrix sizes in Section 6. Essentially, we are interested in improving the message delivery ratio, as mentioned in the previous Section. Figure 4 and 5, show the results of the simulations. These results correspond to the mean number of received alarm notifications at the sink node with a confidence interval of 95% when it is positioned at the corner and center, respectively, for the different matrix sizes. In order to facilitate the evaluation of the performance for different models, we have normalized the number of received alarm notifications based on the total expected alarm notifications for each matrix size. Since we assumed that all nodes have detected a fault (worst case), the total expected number of alarm notifications is equal to the number of nodes in the field, excluding the sink. From these results, we observe that the GA based routing approach operating with the aggregation scheme outperforms the other models regardless of the sink location in the communication topology. Moreover, the performance tend to decreases when the network size increases. We can also observe that, in some instances, when the GA based routing scheme operates without aggregation, it shows better results than the static aggregation tree. In this case, although there is no aggregation, the GA routing scheme was able to find routes with better link quality than those found by the static aggregation tree for the same matrix size, which means that it is not worth to maximize aggregation if the benefits are destroyed by the overhead caused by constant retransmissions. Considering both scenarios, the improvement achieved by GA with aggregation scheme compared to Dijkstra algorithm with aggregation lies between 0% and 28% along different matrix sizes. All models show an improved message delivery ratio when the sink is positioned at the center of the grid. The performance improvement is expected, since the mean distance (i.e., the number of hops) to the sink node is reduced and the number of intermediate nodes directly connected to the sink may increase, allowing their predecessors to have more alternative routes (for instance, non-congested routes) to forward the traffic. On the other hand, in all the schemes with aggregation, there is elimination of redundant or spatially correlated alarm notifications.

7. Conclusion

In this work, we have proposed an aggregation and a GA based routing scheme in WSN to maximize the number of messages received by the sink. We ran simulations with the sink placed in different locations in the communication grid. From the results, we observed that the GA routing scheme operating simultaneously with the aggregation scheme significantly improved the message delivery. It outperformed other models regardless of the sink location in the communication grid topology.

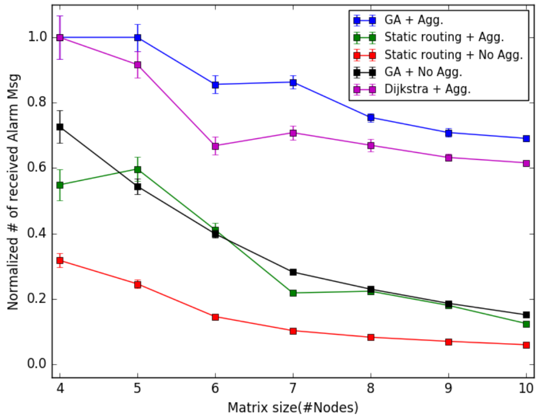

Fig. 4. Message delivery ratio, with the sink at the corner of the grid.

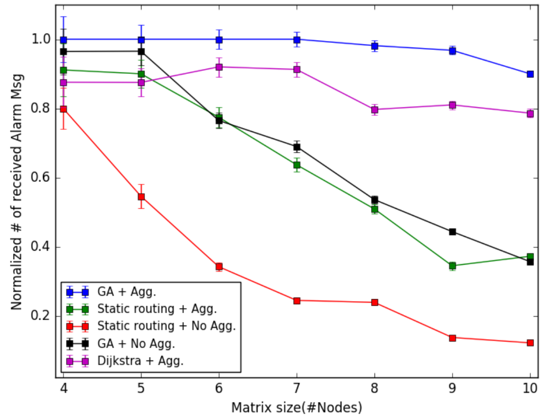

Fig. 5. Message delivery ratio, with the sink at the center of the grid.

This study does not take into account the delivery delay, which is an important aspect for time critical applications in Smart Grid. During the aggregation process, intermediate nodes should wait for a predefined time, WT, to collect data from their children. In a time critical application, data may reach the sink when is no longer needed (outdated), therefore, the optimization of the WT should also be studied. For the future work, we consider working on multi-objective problem. We intend to find a compromise between maximizing the number of messages reported at the sink, while minimizing the delivery delay time and the energy waste in the network, addressing scenarios other than Smart Grids, where sensing correlation also applies, e.g., pipeline monitoring.

Acknowledgements

This work was partially supported by Fundação Calouste Gulbenkian and by Portuguese national funds through Fundação para a Ciência e Tecnologia (FCT) with reference UID/CEC/50021/2013.

References

1. Amin SM, Wollenberg BF. Toward a smart grid: power delivery for the 21st century, *IEEE power and energy magazine*, vol. 3, no. 5, pp. 34–41, 2005.
2. Gungor VC, Lu B, Hancke GP. Opportunities and challenges of wireless sensor networks in smart grid, *IEEE transactions on industrial electronics*, vol. 57, no. 10, pp. 3557–3564, 2010.
3. Fang X, Misra S, Xue G, Yang D. Smart gridthe new and improved power grid: A survey, *IEEE communications surveys & tutorials*, vol. 14, no. 4, pp. 944–980, 2012.
4. Fadel E, Gungor VC, Nassef L, Akkari N, Malik MA, Almasri S, Akyildiz IF. A survey on wireless sensor networks for smart grid, *Computer Communications*, vol. 71, pp. 22–33, 2015.
5. Al-Karaki JN, Kamal AE. Routing techniques in wireless sensor networks: a survey, *IEEE wireless communications*, vol. 11, no. 6, pp. 6–28, 2004.
6. Bari A, Wazed S, Jaekel A, Bandyopadhyay S. A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks, *Ad Hoc Networks*, vol. 7, no. 4, pp. 665–676, 2009.
7. Gupta SK, Kuila P, Jana PK. Gar: An energy efficient ga-based routing for wireless sensor networks. in *ICDCIT*. Springer, 2013, pp. 267–277.
8. Gupta SK, Kuila P, Jana PK. Energy efficient multipath routing for wireless sensor networks: A genetic algorithm approach, in *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*. IEEE, 2016, pp. 1735–1740.
9. Shiobara T, Palensky P, Nishi H. Effective metering data aggregation for smart grid communication infrastructure, in *Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE*. IEEE, 2015, pp. 002 136–002 141.
10. Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on wireless communications*, vol. 1, no. 4, pp. 660–670, 2002.
11. Silva N, Basadre F, Rodrigues P, Nunes MS, Grilo A, Casaca A, Melo F, Gaspar L. Fault detection and location in low voltage grids based on distributed monitoring, in *Energy Conference (ENERGYCON), 2016 IEEE International*. IEEE, 2016, pp. 1–6.
12. Bagaa M, Challal Y, Ksentini A, Derhab A, Badache N. Data aggregation scheduling algorithms in wireless sensor networks: Solutions and challenges, *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1339–1368, 2014.