Computer Communications 101 (2017) 106–120

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Spectrum-aware bio-inspired routing in cognitive radio sensor networks for smart grid applications

E. Fadel[a,*], M. Faheem[b], V.C. Gungor[b], L. Nassef[a,e], N. Akkari[a], M.G.A. Malik[a,d], S. Almasri[a,d], I.F. Akyildiz[a,c]

[a] Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia
[b] Department of Computer Engineering, Abdullah Gül University, Kayseri, 38039, Turkey
[c] Broadband Wireless Networking Lab, School of Electrical and Computer Eng., Georgia Institute of Technology, Atlanta, GA 30332, USA
[d] Faculty of Computing and IT, University of Jeddah, Saudi Arabia
[e] Department of Computer Sciences and Information, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt

ARTICLE INFO

Article history:
Received 5 December 2015
Revised 22 November 2016
Accepted 29 December 2016
Available online 30 December 2016

Keywords:
Smart grid
Wireless sensor networks
Honey bee mating
Bio-inspired
Clustering

ABSTRACT

Cognitive radio sensor networks (CRSNs) have been proposed to serve as a reliable, robust, and efficient communications infrastructure that can address both the existing and future energy management requirements of the smart grid. The existing and envisioned applications of CRSN-based smart grid include substation automation, overhead transmission line monitoring, home energy management, advanced metering infrastructure, wide-area situational awareness, demand response, outage management, distribution automation, asset management. To realize these applications, in this paper, honey bee mating optimization-based routing and cooperative channel assignment algorithms have been proposed. The developed framework significantly decreases the probability of packet loss and preserves high link quality among sensor nodes in harsh smart grid spectrum environments. The proposed approach performance has been evaluated in terms of packet delivery ratio, delay, and energy consumption demonstrating that it has successfully addressed the QoS requirements of most of the SG applications presented.

1. Introduction

The existing power grid is built a century ago, which means its outdated and does not meet the new power demands. Wireless sensor networks (WSNs) have been proposed to effectively monitor, control and operate the power grid [1]. The existing and envisioned applications of WSN-based smart grid include substation automation, overhead transmission line monitoring, home energy management, advanced metering infrastructure, wide-area situational awareness, demand response, outage management, distribution automation, asset management. It is evident that these applications have different quality of service (QoS) requirements in terms of bandwidth, reliability and latency as shown in Table 1 [1]. However, recent field tests have shown that wireless links in smart grid environments have high packet loss rates and varying link qualities because of power grid equipment noise, obstructions, electromagnetic interference, multipath effects, and fading [3]. In these field tests, it has been observed that the average noise level varies between −89 dBm and − 93 dBm in outdoor 500 kV substation environments. Please note that this noise level is much higher than that of outdoor noise levels, which has been observed as − 105 dBm. Thus, the main design challenge of WSN-based smart grid applications is to enable reliable and energy-efficient data delivery under adverse wireless communication conditions of smart grid propagation environments [4].

To this end, cognitive radio sensor networks (CRSNs) can resolve the bandwidth scarcity issues in smart grid by permitting a sensor node to transmit in unused spectrum holes without causing harmful interference to neighboring nodes. In CRSN, if a secondary user (SU) encounters the high noise and/or primary user (PU), it changes its spectrum band or stays in the same band without creating interference with the licensed-user by adapting its radio parameters. In unlicensed spectrum bands, all users have the same right to access the unlicensed spectrum bands by avoiding noisy channels [5]. Hence, CRSNs can enhance the spectrum utilization efficiency to address communication challenges of WSN-based smart grid applications, which are time and location dependent link-quality variations, harsh environmental conditions, resource, energy, and memory limitations of sensor nodes.

* Corresponding author.
E-mail addresses: eafadel@kau.edu.sa (E. Fadel), muhammad.faheem@agu.edu.tr (M. Faheem), cagri.gungor@agu.edu.tr (V.C. Gungor), lmohamed@kau.edu.sa (L. Nassef), nakkari@kau.edu.sa (N. Akkari), mgmalik@kau.edu.sa, mgmalik@uj.edu.sa (M.G.A. Malik), smalmasri@kau.edu.sa, smalmasri@uj.edu.sa (S. Almasri), ian@ece.gatech.edu (I.F. Akyildiz).

http://dx.doi.org/10.1016/j.comcom.2016.12.020
0140-3664/

Table 1
QoS requirements of WSN-based smart grid applications [1,5].

Applications	Bandwidth	Reliability	Latency
Price signalling	9.6–56 kbit/s	99%	2000 ms
Substation Automation (SA)	9.6–56k kbit/s	99.0–99.99%	15–200ms
Overhead Transmission Line Monitoring	9.6–56 kbit/s	99.0–99.99%	15–200ms
Automated feeder switching	9.6–56 kbit/s	99%	1000–2000ms
Home energy management	9.6–56 kbit/s	99.0–99.99%	300–2000ms
Distribution Automation (DA)	9.6–56 kbit/s	99.0–99.99%	20–200ms
Distribution Management (DM)	9.6–100 kbit/s	99.0–99.99%	100 ms-2 second
Wide-Area Situational Awareness	600–1500 kbit/s	99.0–99.99%	15–200ms
Advanced metering infrastructure	10–100 kbit/s per node, 500 kbit/s for backhaul	99.0–99.99%	2000ms
Demand Response	14–100 kbit/s per node	99%	500-ms several minutes
Outage Management	56 kbit/s	99.0%	2000ms
Residual energy management	9.6–56 kbit/s	99%	1000ms
Asset Management	56 kbit/s	99.0%	2000ms
Vehicle to Grid (VG)	9.6–56 kbit/s	99.0–99.99%	2sc-5 min

To address these challenges, in this paper, a honey bee mating optimization-based routing and cooperative channel assignment algorithms for WSN-based smart grid applications is proposed. The proposed communication framework deals with the channel impairments by opportunistically switching among different spectrum bands and employs clustering-based routing for organizing CRSN into a connected hierarchy for load balancing and prolonging lifetime of CRSNs in the smart grid. To this end, the dynamic spectrum access capabilities of the proposed system are benefited to mitigate the noisy and congested spectrum bands, yielding reliable and high capacity links for low-power wireless communications in smart grid. The developed approach significantly decreases the probability of packet loss and preserves high link quality among sensor nodes in various smart grid applications. Overall, the main contribution of this paper is the design of spectrum-aware and bio-inspired routing algorithm maximizing spectrum usage efficiency in harsh smart grid spectrum environments. Comparative performance evaluations demonstrate that the proposed approach is efficient in terms of packet delivery ratio, delay and energy consumption.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 explains the proposed clustering, channel assignment and routing algorithms. Section 4 presents the network model, channel model and simulation parameters. Section 5 shows comparative performance results. Finally, Section 6 concludes the paper.

2. Related work

Due to late recognition of smart grid applications, the research on smart grid communications protocols design is found to be limited [2]. Although some QoS-aware routing protocols [5–13] for smart grid exist, but their scope is limited to certain smart grid applications, such as price signaling, automated metering, and emergency handling. The majority of these existing solutions have been designed to meet application-specific design objectives and requirements in a particular scenario. Though these studies guide design decisions for WSN-based smart grid applications, most of the existing routing schemes achieve one or two of these design objectives at the expense of others. For example, energy saving during data collection introduces excessive delay in the network. Moreover, they generally ignore the impact of external interference and noise on transmission reliability in harsh smart grid (SG) environments [2]. Furthermore, they do not have the ability of channel adaptation to alleviate the interference induced at a certain channel due to the power grid equipment.

Recently, cognitive radio sensor networks (CRSNs) have been proposed to serve as a reliable, robust and efficient data aware communications infrastructure that can address both the existing and future energy management requirements of the SG [5]. The study in [5] proposes a QoS-aware communication framework in CRSNs for smart grid applications. The designed framework provides energy efficient data delivery with low end-to-end delay in the network. However, it faces the problem of data redundancy and node memory overflow during routing, which leads to high data packet collisions in the network. In addition, it does not consider the issue of unnecessary multi-hop data packet transmission and data path looping in the network.

Furthermore, in [15] a spectrum-aware cluster based routing scheme for CRSNs is proposed. However, it consumes a significant amount of energy while finding an appropriate next hope relay node in the network. Therefore, the routing in CRSNs in the context of collision free cooperative spectrum sensing is still an important problem for SG applications, yielding bursty traffic depending on the event characteristics. Recent field tests show that the time varying nature of the smart grid environment due to the variations of spectrum availabilities needs frequent re-channel assignment [1,4]. However, the existing network solutions due to their fixed or inefficient channel allocation strategies fail to handle dynamic spectrum access challenges and thus, they are not resilient or efficient enough to provide desired reliable data delivery for WSN-based smart grid applications [2].

Although these studies [5–15] provide valuable insights for WSN-based smart grid applications, they generally ignore the impact of fading, external interference and noise on transmission reliability in smart grid. Moreover, they do not incorporate dynamic channel adaptation to mitigate the interference induced at a certain channel due to the power grid equipment, and hence, the delay and energy efficiency performance is substantially hampered for the sake of achieving communication reliability. Therefore, in this study particular importance is given to design an energy-efficient bio-inspired routing scheme improving spectrum usage efficiency in harsh smart grid environments. The comparison table showing the differences between the proposed approach and the existing studies is presented in Table 2. In this table, the reliability is a metric of how reliable a communication system can perform data transfers. Some of the SG applications, such as substation automation, overhead transmission line monitoring, automated feeder switching, expect highly reliable data communication, and some of them can tolerate some packet losses in communications. In addition, throughput is the rate of successful message delivery over a communication channel for SG applications presented in Table 1. The energy efficiency is the usage of less energy to provide the same services for long time for various SG applications. Generally, the design and implementation of WSNs is mainly constrained by limited battery energy supply. For this reason, communication protocols for WSNs-based SG applications are mainly tailored to provide high energy efficiency.

Table 2
Comparison of routing solutions in smart grid.

Routing solution	Architecture	Spectrum aware	Energy efficient	Reliability Aware	Throughput aware
Proposed solution	Clustering	√	√	√	√
DCA [5]	Flat	√		√	
HRL-AODV [6]	Flat		√	√	
Lin et al., [7]	Flat			√	
Ronghui et al., [8]	Tree				
Li et al., [9]	Flat		√		
HWMP [10]	Flat			√	√
Kwangsoo Kim and Seong-il Jin [11]	Tree			√	
Sungwook Kim [12]	Flat	√		√	√
SCR [15]	Clustering	√		√	

3. Proposed clustering, channel assignment and routing algorithms

Over the last few decades, nature-inspired intelligence has become increasingly popular in advanced communication networks and information systems design. Several meta-heuristics have been developed to create a robust search tool to handle optimization and local optima problems in various research domains. In general, these meta-heuristics are designed to attack complex optimization problems and have been found to be effective in dealing with several NP-hard problems, where classical optimization methods have failed to be effective and efficient in various scientific and engineering domains [16].

Recently, the evolutionary algorithm, a form of meta-heuristic has become one of the most popular techniques in both theoretical analysis and industrial applications and is used as an optimization tool in different research fields [17], to solve real-world problems. These problems are usually very challenging for the conventional computing techniques, the main purpose of the nature-inspired evolutionary algorithms is to identify the mechanism of such complex computational systems and to develop a flexible, proficient and robust technique with a realistic execution time near optimal results. In evolutionary algorithms, the fitness function is responsible for measuring the quality of an individual. To this end, the individual with superior quality has more chance to be elected as a solution of the initial population. In the mating pool, several available variation operators, such as crossover, mutation and selection, also can be used to alter recent information. Generally, the evolutionary algorithm follows user-defined instructions under which recombination and mutation are made.

The Honey-bee Mating Optimization (HBMO) a typical nature-inspired approach in which the search algorithm is inspired by the process of real honey bees mating to address highly nonlinear combinatorial optimization problems in various science and engineering fields [18]. The queen and drone bees are the most important members of the family in a bee colony. A queen bee which is the only member can lay eggs and has the responsibility of reproducing individuals. A drone can be considered as a father of the colony which only acts as an amplifier for its queen genome without altering its genetic inheritances, except for mutations. During modelling the mating process, the queen bee keeps producing new members where the genome is attached to one solution (to one bee) for the family with multiple male (drone) bees for the studied optimization problem [19]. These above mentioned advantages of the evolutionary approaches make them appropriate for the complex problems associated with the highly unpredictable smart grid communication environments.

Furthermore, in recent years spectral clustering has gained an increasing interest due to its simplicity and ease of implementation in a wide-range set of applications, such as image processing, data mining and biological etc., to solve modern clustering problems [20]. For a given set of data points $\{x_i\}_{i=1}^{n}$, $x_i \epsilon\ R^{1\times d}$ (x is a data point ranges from $i = 1$ *to* n in region R having dimensions $l \times d$), the goal of the spectral clustering is to distribute the data points into groups based on their similarities between all pairs of data points. In spectral clustering algorithms, the point of departure is a weighted similarity graph $G(V, E)$, where the weights correspond to the pairwise similarities and vertices correspond to data points [21]. The Evolutionary Spectral Clustering techniques based on these weights form the Laplacian graph and calculate the Eigen decomposition of this Laplacian. To perform bi-partitioning of the given data, where large numbers of clusters are created, most of the developed algorithms use the thresholding notion for the second eigenvector of the Laplacian mean. Hence, the eigenvector with the second least eigenvalue is nominated. However, to find a k-way clustering directly the designed algorithm uses multiple eigenvectors. Further, since the queen in the family of a bee colony is the fittest individual, it is expected that this will evolve better solutions, if the new generated solution does not satisfy the individual best solution. This leads to local minima problem to find the robust optimal clustering and routing solution in highly dynamic smart grid environments. Therefore, instead of relying on the mutation operator for the new generation, we extract the advantages of evolutionary algorithms and apply cross over, mutation and selection operators to find the best existing optimal or near optimal solution for the new generated bees (broods). Hence, inspired by the above mentioned advantages of the evolutionary spectral clustering approaches, we developed a hybrid Energy-efficient Spectral Honey bee Mating Optimization-based Clustering (ESHC) algorithm to solve complex clustering problems in energy efficient manner for CRSNs-based smart grid monitoring applications.

3.1. Energy-efficient spectral honey bee mating optimization-based clustering (ESHC)

In our proposed hybrid clustering algorithm, one of the main objectives is to find an energy-efficient clustering solution at low complexity in the network. In the proposed approach, the function of minimum Euclidian distance information plays a key role to minimize high transmission energy consumption cost by providing stable links among sensor nodes in the network. The proposed honey bee mating optimization-based clustering algorithm is inspired by the process of mating in a real honey bee colony where the queen, drones and brood bees have their own genome composed of genes. This genome is attached to each individual solution (one bee) of the studied optimization problem while modelling the mating process. A list of numerical values is used to describe one genome where each value is attached to a decision variable (gene) that indicates an unknown of the problem solutions, including the best solution, candidate solution and the trivial solution. The fitness function of the problem from such a list depends on the values of unknowns (greater or smaller) and presents the genome associated (bee) solution in a stronger or weaker manner. This genotype representation presents a solution (bee) that has several unknowns (genes) equivalent to the total number of clusters. This solution forms the essential part of the designed algorithm, which studies clustering at the lowest network cost (the best performance). Here, the aim of generating high quality is to find the fittest clustering solutions, which maximizes the fitness function for each individual cluster generated in the network. In the proposed distributed scheme, a sink node is responsible to initiate clustering process for partitioning the set of nodes into multiple groups based on their similarity values where each group corresponds to a set of clusters. The entire network goes through five different phases, as discussed below:

Phase 1: Generating the drone set and initial queen

At the beginning, an initial population of individual bees (sensor nodes) is generated (of size $n=200$) for the potential solution of the problem by using a random number generator and each individual is evaluated by a predefined fitness function. Each genome randomly builds within admissible ranges of the variables by considering its genome list of similarity of the unknowns. Then, based on the fitness function values, the initial population of bees is ranked decreasingly and the one with the best quality solution is selected as an initial queen bee ($\mathcal{Q}_i$) in the existing population for the current round in region R_i. After employing a queen bee, the remaining solution forms a list of drones (trivial solutions) ranked in decreasing order to mate with the queen during the first mating flight. In addition to its genome, the queen is characterized by her spermatheca capacity (Sc_n) as well as her speed $v(i)$ and/or energy $E(i)$. The spermatheca capacity is equal to the maximum number of drones that may mate with the queen. Herein, it is set constant during all mating flights. An initial population (P_{ini}) of drones and queens in the entire search space to study the clustering problem can be numerically indicated as:

$$P_{ini} = \sum_{i=1}^{n} \mathcal{B}_i \tag{1}$$

$$\mathcal{Q}_k = \sum_{i=1}^{k=n-l} \mathcal{Q}_i \tag{2}$$

$$\mathcal{D}_l = \sum_{i=1}^{l} \mathcal{D}_i \tag{3}$$

$$Sc_i(\mathcal{Q}_i)^{\mathcal{D}_{best}} = \sum_{i=1}^{n} (Sc_i)^{\mathcal{D}_j} \tag{4}$$

where $\mathcal{B}_i$ represents the individual bees to initialize for $n = 200$, $\mathcal{Q}_i$ shows the best individual bees (queen bees) of size $k = 23$ in different regions $R_{1,2,\dots,n}$, $\mathcal{D}_i$ indicates the drones of size $l = 177$. In $Sc_i(\mathcal{Q}_i)^{\mathcal{D}_{best}}$, Sc_i, $i \in \{1, 2, \dots, n\}$ is the spermatheca capacity of a queen $\mathcal{Q}_i$ containing best drones $\mathcal{D}_{best}$ for mating of maximum size $\mathcal{D}_j$, $j = 10$.

Phase 2: Mating flight

In mating flight, the current queen bee $\mathcal{Q}_i$ (i.e., the best clustering solution leading to the lowest clustering network cost in region R_i) randomly selects some drones from the drone's population list $\mathcal{D}_n$ and mates with each individual drone and stores resultant genome in her spermatheca. The queen $\mathcal{Q}_i$ before leaving the hive is initialized with some energy and/or speed parameter values randomly generated between 0.1 and 1. The probability of mating among the drones $\mathcal{D}_n$ and queen $\mathcal{Q}_i$ can be numerically computed as follows

$$\begin{aligned} P_b(\mathcal{Q}_i, \mathcal{D}_n) &= exp\left(-\Delta f / \mathcal{Q}^i_{\nu(i)}(t_i)\right) \\ &= exp(-|f(\mathcal{Q}_i) - f(\mathcal{D}_i)|) / \mathcal{Q}^i_{\nu(i)}(t_i) \end{aligned} \tag{5}$$

where $P_b(\mathcal{Q}_i, \mathcal{D}_n)$ is the mating probability of a queen $\mathcal{Q}_i$ to a set of drones $\mathcal{D}_n$, $\mathcal{Q}^i_{\nu(i)}(t_i)$ is the queen speed at time t_i, $f(\mathcal{Q}_i)$ and $f(\mathcal{D}_i)$ are the performance functions of $\mathcal{Q}_i$ and $\mathcal{D}_i$, respectively.

The above function is highly associated with the queen speed and offers superior values when the queen speed is greater or when the drone fitness function is closed to the queen fitness function. Herein, mating probability can also be measured by generating a random number $P_b(\mathcal{Q}_i, \mathcal{D}_i) > r$, where random number $r \in [0, 1]$. If the generating probability is greater than 0 then drone $\mathcal{D}_i$ can successfully mate with the queen and his genome is stored in her spermatheca. The queen speed and energy decay iteratively depending on the time t_i even if the mating process succeed or not can be numerically indicated as follows

$$\mathcal{Q}^i_{\nu(i)} t_{\{i+1\}} = \partial \mathcal{Q}^i_{E(i)} t_{\{i+1\}} = \left(\mathcal{Q}^i_{E(i)} t_{\{i+1\}} - \delta \mathcal{Q}^i_{\nu(i)} t_{\{i+1\}}\right) \tag{6}$$

where $\partial \in [0,1]$ indicates the decay coefficient and δ shows the energy loss after each iteration. The above process repeats either until the queen's speed/energy decays down to a minimum level as $\mathcal{Q}^i_{\min[\nu(i)]}$ and $\mathcal{Q}^i_{\min[E(i)]}$ or until queen spermatheca is full.

Phase 3: Breeding Process

In the breeding process, a drone $\mathcal{D}_i$ genome is randomly selected from the queen spermatheca and combine with the queen own genome by applying crossover operator which leads to genome of a new bee (B_i). The Genetic algorithm is used here with a single heuristic crossover operator to generate new genome can be numerically indicated as:

$$B_i = Q_i + \gamma (r(\mathcal{Q}_i + \mathcal{D}_i)) \tag{7}$$

where B_i is the new generated bee, $\mathcal{D}_i$ indicates the drone genome randomly selected from the queen $\mathcal{Q}_i$ spermatheca, r in a random number $\in [0, 1]$ and γ helps to round solution towards the nearest integer.

Phase 4: Brood fitness improvement by applying Evolutionary operators

After generating an initial population of the new bees an extracted looping feature of the EA algorithm, including selection, recombination/crossover and mutation operators are applied on each individual broods, until termination criterion is satisfied for the complete clustering solution as $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n$. Herein, it is assumed that the applied evolutionary operators are equal to the number of worker bees. By applying selection mechanism, proposed algorithm produces new individuals of the new generation where each individual solution (clustering solution as an individual string) is measured based on its fitness value. The selection operator is responsible to assign each set of pair of bees (sensor nodes) a positive weight w_{ij} in the current population to find a partition graph (G) such that $G \in w_{ij}$. The value of this assigned weight is based on Euclidean distance function which is proportional to the fitness function (Fit) of each individual bee in the string S_i. Thus, the probability of a particular string is being selected can be formulated as follows

$$P_{(S_i)} = \frac{Fit(S_i)}{\sum_{j=1}^{n} Fit\left(S_j\right)} \tag{8}$$

The strings with better fitness values, i.e. lower cost have greater probability of being generated. The probability of an individual bee B_i selected in string S_i with the highest weight is given as follows

$$P_{(B_i)} = \frac{Fit(B_i)}{\sum_{j=1}^{n} Fit(B_i)} \tag{9}$$

In generation k, we denoted the bee population members as $x^k_1, x^k_2, \dots, x^k_n$, where each population member represents an individual sensor node. Assume that an individual having n genes and each of them can be denoted as 0 for the dead or inactive bee (node) with insufficient energy and 1 for the active node having sufficient amount of energy. Then, n individual solutions of the population $\mathcal{Q}_n$ can be numerically described as follows

$$\forall_i \in \left\{x_1^K, x_2^K, \ldots, \ x_n^K\right\} \text{and} \, \forall_j \in \left\{x_1^k, x_2^k, \ldots, \ x_n^k\right\} \tag{10}$$

$$\mathcal{Q}_i^j = \begin{cases} 1 & \text{if } x_n^K > 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

To form an initial population for the required percentage of the strings in the network each individual bee is initialized randomly through the probability P_b by choosing either 1 or 0 can be indicated as follows

$$\forall_i \in \left\{x_1^1, x_2^0, \ldots, \ x_n^1\right\} \text{and} \, \forall_j \in \left\{x_1^0, x_2^1, \ldots, \ x_n^1\right\} \tag{12}$$

$$\mathcal{Q}_i^j = \begin{cases} 1 & \text{if } x_n^0 \parallel \ x_n^1 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

The minimum Euclidean distance (d) between each set of pair of individuals i and j in term of weight w_{ij}, can be calculated as follows

$$E_{min(d)} = \sum_{i=1}^{n} \sum_{j=1}^{S} w_{ij} \parallel x_i - y_j \parallel^2 \tag{14}$$

The delay between each set of pair of individuals i and j in term of weight w_{ij}, can be defined as:

$$E_{min(t)}(w) = \sum_{i=1}^{n} \sum_{j=1}^{S} w_{ij} \parallel diff_{ij} \parallel^2 \tag{15}$$

where $w_{ij} \in diff_{ij}$ is the difference between individuals i and j based on the packet time of arrival and time to departure. The residual energy (Re) of each individual in term of weight w_{ij} can be numerically written as:

$$Re_{maxi}(w) = \sum_{i=1}^{n} \sum_{j=1}^{S} w_{ij} \parallel E_{i,j,\ldots,n} \parallel^2 \tag{16}$$

The probability of a particular queen bee $\mathcal{Q}_i$ (sensor node) being selected as a cluster head (CH) in each individual cluster $\mathcal{C}_i$ in region R_i is given as follows

$$P_b(CH)_{\mathcal{C}_i} = \frac{Fit(\mathcal{Q}_i)}{\sum_{j=1}^{n} Fit\left(\mathcal{Q}_j\right)} \tag{17}$$

The associated weight pattern of a set of neighboring brood bees $B_i, \ i \in \{1, 2, \ldots, n\}$ associated to the jth individual best queen bee Q_j belongs to a cluster center can be numerically defined as:

$$Q_j = \frac{1}{n} \sum_{i=1}^{n} w_{ij}(B_i) \tag{18}$$

subject to

$$w_{ij} = \begin{cases} 1 & \text{if } B_i \in \mathcal{C}_i \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

In entire search space (k), the cost function (C_f) based on minimum Euclidean distance of a set of individual brood bees $B_i, \ i \in \{1, 2, \ldots, n\}$ associated to the jth individual best queen bee Q_j belongs to a cluster center can be numerically defined as:

$$C_f\left(Q_j\right) = \frac{1}{n} \sum_{i=1}^{n} d\left(Q_j, B_i\right)^k \tag{20}$$

The average Euclidean distance of all the brood bees (the best solution for sensor nodes) to their associated cluster centers can be numerically defined as:

$$d_{B_i \to \mathcal{C}_n} = \frac{\sum_{j=1}^{\mathcal{C}_n} \left[\sum_{\forall B_i \in \ \mathcal{C}_i} d\left(B_i, \ \mathcal{C}_{ij}\right)\right] / \ \left|\mathcal{C}_{ij}\right|}{N_c} \tag{21}$$

where

$$d\left(B_i, \ \mathcal{C}_{ij}\right) = \sqrt{\sum_{i=1}^{n} \left(B_i - \ \mathcal{C}_{ij}\right)^2} \tag{22}$$

where $\mathcal{C}_n$ shows the total number of clusters in the entire search space, B_i are the number of brood bees (cluster member sensor nodes) belongs to a cluster $\mathcal{C}_i$, $d(B_i, \ \mathcal{C}_{ij})$ shows the minimum distance between a set of brood bees to associated cluster center, $|\mathcal{C}_{ij}|$ is the number of sensor nodes in a group of a set $\mathcal{C}_{ij}$. The maximum intra-distance (χ) associated to brood bees sensor nodes and cluster centers can be mathematically defined as:

$$\chi_{max}\left(B_i, \ \mathcal{C}_{ij}\right) \ = maxi_{j=1,\ldots,n}\left(\sum_{\forall B_i \in \ \mathcal{C}_{ij}} d\left(\mathcal{C}_{ij}, B_i\right) / \left|\mathcal{C}_{ij}\right|\right) \tag{23}$$

The minimum inter-distance ($\mathcal{Y}$) between any pair of clusters can be numerically written as:

$$\mathcal{Y}_{min}\left(\mathcal{Q}_{i,J}\right) \ = min_{\forall j_1, j_2, \ j_1 \neq j_2}\left\{d\left(\mathcal{C}_{ij}^1, \ \mathcal{C}_{ij}^2\right)\right\} \tag{24}$$

To get better clustering qualities, the idea of partitioning the small clustering sets that are compact and well separated by maximum inter-distance can be calculated with the following formula

$$d_{\min}(\mathcal{C}_{i,\ldots,n}) = \min_{1 \le i \le n} \left\{ \min_{\substack{1 \le i \le \mathcal{C}_n \\ i \neq j}} \left\{ \frac{d\left(\mathcal{C}_i, \ \mathcal{C}_j\right)}{\max_{1 \le k \le \mathcal{C}_n} \{d^*(\mathcal{C}_k)\}} \right\} \right\} \tag{25}$$

where $d(\mathcal{C}_i, \ \mathcal{C}_j)$ is the distance between clusters $\mathcal{C}_i$ and $\mathcal{C}_j$ and $d^*(\mathcal{C}_k)$ is the intra-cluster distance of cluster C_k. The total energy consumes during appointing cluster heads (the best queens) in the network can also be measured as the sum of the energy dissipated during communication. Thus, for CH^i, $i \in \{1, 2 \ldots, n\}$ the total number of cluster heads generated in the network can also be written in the form as:

$$E_{EHSC} = \sum_{j=1}^{\mathcal{C}_n} \sum_{Sn \in k_i} \sum_{(T_x + R_x)_{\mathcal{C}_i}} CH_i \tag{26}$$

where $\mathcal{C}_n$ is the total number of cluster heads (best selected queens), $Sn \in k_i$ are non-cluster head sensor nodes associated with the ith cluster head. The energy consumption of each individual cluster head can be numerically written as:

$$E_{EHSC} = \sum_{j=1}^{Sn} \sum_{(T_x + R_x)_{\mathcal{C}_i}} CH_i \tag{27}$$

where Sn are the member sensor nodes in a cluster $\mathcal{C}_i$ and $(T_x + R_x)_i$ is the transmission and receiving power consumption associated to a cluster head CH_i. Thus, the fitness function in term of each individual cluster during appointing a normal node as a cluster head can be written as:

$$Fit(EHSC) = \frac{1}{\sum_{j=1}^{Sn} \sum_{Sn \in k_i} \sum_{(T_x + R_x)_{\mathcal{C}_i}} CH_i} \tag{28}$$

Thus, the fitness value for cluster heads selection cost in entire network can be formally indicated as:

$$Fit(EHSC) = \frac{1}{\sum_{j=1}^{\mathcal{C}_n} \sum_{Sn \in k_i} \sum_{(T_x + R_x)_{\mathcal{C}_i}} CH_i} \tag{29}$$

To form a mating pool of n parents, selection operator randomly chooses best individuals from the current population set and repeat this process n times until the defined criteria satisfy. In formal way, selection operator $SO : \mathcal{Q}_i \Leftrightarrow \mathcal{Q}_j$ can be numerically defined as: let $\mathcal{Q}_{i,S1}, \forall_i \in \{x_1^K, x_2^K, \ldots, x_n^K\}$ be the set of individuals in string 1 and $\mathcal{Q}_{j,\ S2} \forall_j \in \{x_1^k, x_2^k, \ldots, x_n^k\}$ be the set of individuals in string 2, which are uniformly distributed from the set $k \in \{x_1^k, x_2^k, \ldots, x_n^k\}$, can be described as:

$$\mathcal{Q}_{i,S1},\ \ \mathcal{Q}_{j,\ S2} \forall_i \in \left\{x_1^K, x_2^K, \ldots,\ x_n^K\right\} \text{ and } \forall_j \in \left\{x_1^k, x_2^k, \ldots,\ x_n^k\right\} \quad (30)$$

$$\mathcal{Q}_{i,S1} \cup\ \ \mathcal{Q}_{j,\ S2} \forall_i \in \left\{x_1^K, x_2^K, \ldots,\ x_n^K\right\} \cup \forall_j \in \left\{x_1^k, x_2^k, \ldots,\ x_n^k\right\} \quad (31)$$

$$\mathcal{Q}_{i,S1} \cup \mathcal{Q}_{j,\ S2} = \left\{x_1^k, x_2^K, \ldots,\ x_n^{K\|k}\right\} \quad (32)$$

$$\mathcal{Q}_{i,S1} \parallel \mathcal{Q}_{j,\ S2}\ = \begin{cases} \mathcal{Q}_{i,S1} & \text{if } Fit\left(x_i^K\right) > Fit\left(x_j^k\right) \\ \mathcal{Q}_{j,\ S2} & \text{otherwise} \end{cases} \quad (33)$$

To produce a pair of offspring strings, reproduction operator is responsible for partial swapping of the bits between two individual strings. For a given crossover probability ($P_c = 0.90$), each pair of strings from the new brood bee population is randomly picked out by proposed algorithm. Then, a uniformly distributed random number is generated in the predefined range between 0 and 1. If the generated random number is less than the predefined value then the developed algorithm is responsible to apply the crossover operator for producing swapping new bee generation, otherwise it does not apply crossover operator on these two particular strings. Through applying the crossover operator along with the strings a crossover point is also randomly generated. The crossover operator plays an important role for swapping individual bee values from one string to another pairwise string in the population. For a given crossover probability ($P_c =\ 0.90$), on average crossover in each generation are given as $P_c = 100\ x\ 0.9 = 90$. In formal way, crossover $CO : \mathcal{Q}_i\mathcal{Q}_j \Leftrightarrow \mathcal{Q}_i^{\sim}\mathcal{Q}_j^{\sim}$ can be numerically demarcated as: let $\mathcal{Q}_{i,S1}, \mathcal{Q}_{j,\ S2}, \forall_j \in \{x_1^K, x_2^K, \ldots,\ x_n^K\}$ be the individuals and $\mathcal{Q}_{i,S1}^{\sim}, \mathcal{Q}_{j,S2}^{\sim}\ \in \{x_1^k, x_2^k, \ldots,\ x_n^k\}$ are the individuals uniformly distributed from the set $\{x_1^k, x_2^k, \ldots,\ x_n^k\}$. This genomic type depiction silently facilitates the creation of dynamic strings in the whole network by randomly selecting a single cut point c_1 from the predefined range $\{1, \ldots,\ n-1\}$, for the each set of parents. Then, these two points contribute to parent individuals $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are swapped as follows

$$\left(\mathcal{Q}_{i,S1}\mathcal{Q}_{j,\ S2}\right)\forall_{i,j} \in \left\{x_2^k x_2^K, x_3^k x_5^K, x_8^k x_{12}^K, \ldots,\ x_n^{K\|k}\right\} \quad (34)$$

$$\left(\mathcal{Q}_{i,S1}^{\sim}, \mathcal{Q}_{j,S2}^{\sim}\right)\forall_i^{\sim} \in \left\{x_1^K, x_6^k, \ldots,\ x_n^K\right\} \text{ and } \forall_j^{\sim} \in \left\{x_2^k, x_4^K, \ldots,\ x_n^k\right\} \quad (35)$$

For more clear understanding, we can write the above equation as:

$$\mathcal{Q}^{\sim} = \left(\mathcal{Q}_{i,S1}^{\sim}, \mathcal{Q}_{j,S2}^{\sim}\right)\forall_i^{\sim} \in \left\{x_1^K, x_6^k, \ldots,\ x_n^K\right\}\ \text{ and } \forall_j^{\sim} \in \left\{x_2^k, x_4^K, \ldots,\ x_n^k\right\} \quad (36)$$

$$\mathcal{Q}_{new}^{\sim} = \begin{cases} \mathcal{Q}_{i,S1}^{\sim} & \text{if } Fit\left(n_{i,j}\right) \geq\ Fit\left(n_{i.j}\right) \\ \mathcal{Q}_{j,S2}^{\sim} & \text{otherwise} \end{cases} \quad (37)$$

We consider a non-uniform mutation operator with the probability of $P_m =\ 0.02$ and $P_m =\ 0.01$, to perform mutation operation with the assertion that no important genetic material is lost. The new generation, which is the result of reproduction and crossover, the designed algorithm takes into account each string bit by bit (i.e. node by node), by generating a uniformly distributed random number between 0 and 1. If the random number value is less than

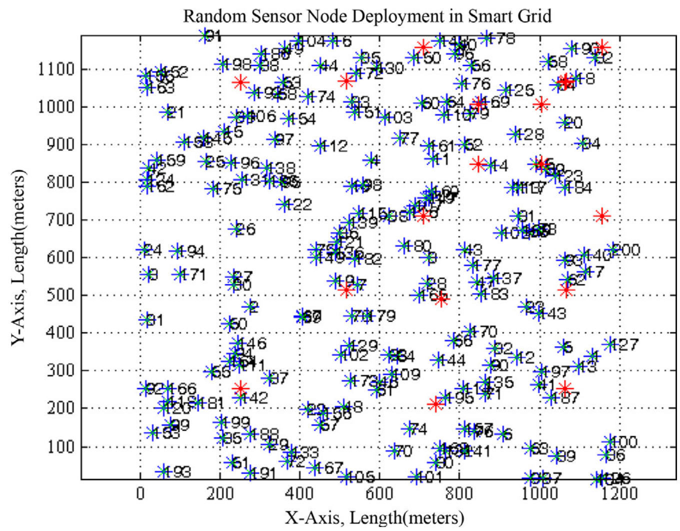

Fig. 1. Sensor node deployment in the smart grid of 2D area (M×M) meter. Here, red color icons indicate the network coordinators while the blue color icons are the sensor nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

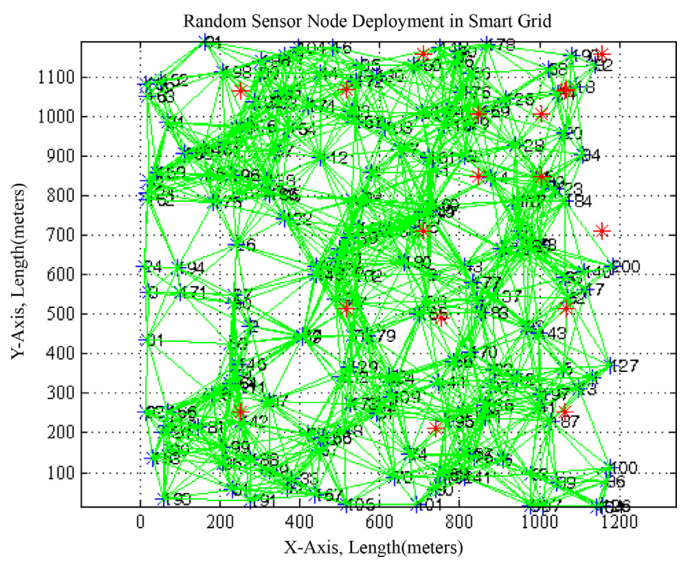

Fig. 2. Clustering architecture generated by ESHC algorithm in smart grid for both dense and sparse regions.

the predefined value 0.01 or 0.02, then the developed algorithm is responsible to apply the mutation operator to the new individual else it does not ponder mutation operator to that particular node. In a non-uniform mutation operator where the value of gene j is selected for mutation is modified to

$$\mathcal{Q}_{ij_{(new)}} = \text{round}\left(\mathcal{Q}_{ij} + P_{m1}\left(\mathcal{Q}_{ij_{(maxi)}} -\ \mathcal{Q}_{ij}\right)\right) \text{ if } P_{m1} = 0.01 \quad (38)$$

and

$$\mathcal{Q}_{ij_{(new)}} = \text{round}\left(\mathcal{Q}_{ij} + P_{m2}\left(\mathcal{Q}_{ij} - \mathcal{Q}_{ij_{(\min)}}\right)\right),\ P_{m2} = 0.02 \quad (39)$$

where P_{m1} and P_{m2} are random numbers, and $\mathcal{Q}_{ij_{(maxi)}}$ and $\mathcal{Q}_{ij_{(\min)}}$ are the limits of gene values.

Phase 5: Termination criterion

In designed algorithm, one of the basic objectives is to find a best individual solution for a given population (the best solution with greater fitness functions) as shown in Fig. 1 and Fig. 2. If the improved solution belongs to an individual (new generated brood bee) cannot be found for a given time interval over a predefined number of cycles then that individual is supposed to be abandoned. Then, the current solution (current best queen) will be considered as a best solution otherwise phase 2 repeats iteratively. In the offered algorithm, the value of the predefined number of cycles is an essential control factor which is called 'limit' for the abandonment. Let's assume that an individual i in $\mathcal{Q}_i$ is abandoned and $j \in \{1, 2, \ldots, n\}$, then the individual better solution discovered to be replaced with $\mathcal{Q}_i$. This process can be defined as follows

$$\mathcal{Q}_i^j = \mathcal{Q}^j_{\min[\nu(i)]} + r\left(\mathcal{Q}^j_{\mathrm{maxi}[\nu(i)]} - \mathcal{Q}^j_{\min[\nu(i)]}\right) \tag{40}$$

where $\mathcal{Q}^j_{\mathrm{maxi}[\nu(i)]}$ and $\mathcal{Q}^j_{\min[\nu(i)]}$ indicates the maximum and minimum speed of the queen, respectively.

In the last, by applying these mutation operators the probability (P_m) value of each individual changes from 0 to 1 or 1 to 0 in the new generated population in the form of

$$M_{(P_m)} : \mathcal{Q}_i^{j\sim} \to \mathcal{Q}_i^{j'} \tag{41}$$

$$\mathcal{Q}_{new} = \begin{cases} \mathcal{Q}_i^{j'} & \text{if } \mathcal{Q}_i^{j'} > 0 \\ \mathcal{Q}_i^{j\sim} & \text{otherwise} \end{cases} \tag{42}$$

This entire process continues until the termination criterion of the proposed algorithm is satisfied. Thus, a set of best queens $best\_\mathcal{Q}_i{}'$ from the fittest queens $Fit(\mathcal{Q}_j{}^{\sim})$ are selected in the entire search space as the feasible clustering solution for the current population formally can be described as:

$$\forall \mathcal{Q}_{new}^{i,\ldots,n} : \ Fit\left(\mathcal{Q}_j{}^{\sim}\right) \le Fit\ \left(best\_\left(\mathcal{Q}_i{}'\right)\right) \tag{43}$$

3.2. Quality aware channel assignment algorithm (QCA)

In CRSNs, one of the key design issues is spectrum sensing. During spectrum sensing each secondary user sensor node periodically monitors primary user activities for a vacant frequency band. If available, then immediately transmits its data over that channel without creating interference to neighboring nodes. Thus, in smart grid, one of the main challenging issues is to estimate available channel quality for efficient and reliable transmission in the network. A channel facing high interference issues must be replaced immediately with the appropriate one to improve the sensing or data transmission reliability for smart grid applications. Generally, sensor nodes are usually battery-operated devices and spectrum sensing is a repetitive process, which consumes a significant amount of energy for implementing spectrum sensing in all nodes may not be efficient in term of energy consumption. Therefore, the spectrum sensing task can be performed in a distributing manner between primary users and some intelligent devices. In general, spectrum sensing algorithms are classified into cooperative and a non-cooperative approaches. In cooperative algorithm approach, secondary users perform spectrum sensing and forward the measurements to the central fusion specialized device that fuses the spectrum results and makes decision on the use of spectrum. On the other hand, secondary user makes local decision based on its own spectrum measurement in non-cooperative algorithm approach. It can be envisioned that cooperative spectrum sensing can significantly increase the systems aptitude in recognizing and avoiding primary users compared to non-cooperative approach. Practically, seems cooperative approach is an efficient and cost-effective solution, and can provides more robust results to employ in smart grid. Therefore, in this study, we focus on cooperative spectrum sensing and a number of cooperative intelligent devices are deployed in an ad-hoc manner with their known location and communication range ψ_r meters for monitoring smart grid systems. The deployed network also includes cognitive radio nodes where each of them is equipped with a single radio transceiver that can be tuned to any channel opportunistically in the radio spectrum. Herein, it is assumed that there are total number of ϕ_n channels which are equal to the primary users transmitters φ_{pu} with their know position and the maximum coverage range ψ_r meters. There also exist φ_{su} secondary users which can occupy channel ϕ_i from ϕ_n opportunistically when φ_{pu} goes silent for τ_{off} seconds. In designed channel assignment algorithm, during spectrum sensing each cooperative device employs energy based signal detection mechanism presented in [22] for primary signal detection with a predefined threshold (σ) to identify either i^{th} channel is occupied (τ_{on}) or not (τ_{off}). It can be numerically denoted as:

$$\text{Spectrum Sensing Decision}(\lambda) = \begin{cases} S_{pu}^i & \text{if } E \ge \sigma \\ S_{ideal}^i & \text{if } E < \sigma \end{cases} \tag{44}$$

where S_{pu}^i indicates the i^{th} channel is busy due to the active primary user with probability ρ_{pu}^i and S_{ideal}^i shows the i^{th} channel is idle with probability ρ_{ideal}^i such that $\rho_{pu}^i + \rho_{ideal}^i = 1$.

In proposed channel assignment algorithm, one of the main purposes of the multiple specialized devices called network coordinator (NC) is to allocate empty available spectrum efficiently by avoiding excessive message exchange in the network. These network coordinators can communicate and periodically update each other with spectrum information, as they are fully aware of the available spectrum τ_{on} and τ_{off} of i^{th} channel at any given time t_i. To avoid data packet collision during message exchange between secondary users and network coordinators, the designed algorithm considers a Carrier Sense Multiple Access (CSMA) mechanism in the network. In addition, an individual cluster head in each cluster plays an important role to communicate with its member nodes for measuring assigned channel quality for reliable data transmission. After cluster formation, each active cognitive radio equipped sensor node such that $N_i \in N_a$ sense the environment interference in the smart grid by taking into account the parameters given in following equation using default assigned channel.

$$y_j(t) = \sum_{a=1}^{N_a^*} \sum_{i=1}^{B_n} \xi_a^i(t) * \xi_{a,j}^i(t) + \xi_{mn}(t) + \xi_{nn}(t) + \xi_{eme}(t) \tag{45}$$

where $\xi_a^i(t)$ indicates that the communication already active between communication source and reference cognitive node $\sum_{a=1}^{N_a^*} \sum_{i=1}^{B_n} \xi_a^i(t) * \xi_{a,j}^i$ on the i^{th}channel (ϕ_i), $\xi_{mn}(t)$ shows environmental conditions (i.e., temperature, moisture, etc., based on the deployment field), $\xi_{nn}(t)$ shows human noise, $\xi_{eme}(t)$ indicates the machine or system noise, and N_a^* indicates the active nodes and B_n is the bandwidth.

To occupy high quality channel, after sensing, a secondary user (SU_i) interested to participate in communication process sends an initiation message (*init_msg*) to associated cluster head (CH_i). The cluster head CH_i replies with its initiation acceptance message (init_accpt) in time t_i and allocate a partial set of available spectrum channels in a timely manner after negotiations with NCs. Note that CH_i follows the first come first serve (FCFS) based policy which means that a CH_i will starts communication with the SU_i to whom it receives the message first and rest of the requests send by others SU_j will be maintained in channel assignment list in decreasing order. To measure the channel link quality of i^{th} channel for reliable data transmission, SU_i begins rapid message exchange where in the first $n1 = (n \times \rho_b)$ rounds rapidly exchange sequential challenging bits while in the rest of $n2 = n - n1 = n - (n \times \rho_b)$ rounds random challenging bits are sent $(0 \le \rho_b \le 1)$ to CH_i. The CH_i also sends and computes the bits response to sequential and random challenging bits by taking into account the current and maximum l previous challenging bits $(l \ge 0)$. For quality channel assignment, the maximum total numbers of errors (corrupted bits) that are acceptable by the SU_i and CH_i are ε_1 and ε_2 over i^{th}channel respectively, occur during n iterations for which the algorithm successfully ends (ε_1 and $\varepsilon_2 \geq 0$).

$$\text{Channel Quality Decision } (\omega) = \begin{cases} C_i & \text{if } \varepsilon_1 \parallel \varepsilon_2 \geq \varepsilon_o \\ 0 & \text{otherwise} \end{cases} \tag{46}$$

where ε_o is the error reported in ith channel C_i. Eq. (46) sets the decision criterion with respect to the received signal quality in Eq. (45). Specifically, if the received quality is higher than or equal to the reported error of the channel, the decision will indicate that the communication has occurred; if the quality is lower than the error, the decision will indicate no communication occurs. Moreover, even if ε_1 incorrect or corrupted bits are received to CH_i sent by SU_i due to i^{th} channel errors, the NC_i identifies the SU_i validity over i^{th} channel. Similarly, even if ε_2 incorrectly bits are received to SU_i sent by the CH_i due to incorrect or corrupted bits, the SU_i identifies the CH_i validity over the i^{th} channel. Note that ε_1 and ε_2 are two significant parameters for implementing the algorithm over the noisy channel. The probability (P_b) of i^{th} channel C_i that p errors are identified by CH_i and SU_i at ith section and also the P_b of the maximum incorrect or corrupted bits $\varepsilon_1 \| \varepsilon_2$ errors are detected by NC_i and SU_i due to environmental challenges, including ξ_a^i, ξ_{mn}, ξ_{nn} and ξ_{eme} factors in ith round at time t_i can be numerically obtained as follows:

Sequentially challenging bits sent by SU_i at time t_i can be written as:

$$SU_i(t)_i = \xi_a^i(t)_i^{1,\ldots,n} + \xi_{mn}(t)_i^{1,\ldots,n} + \xi_{nn}(t)_i^{1,\ldots,n} + \xi_{eme}\ (t)_i^{1,\ldots,n} \tag{47}$$

The sum of the sequential challenging bits sent by SU_i can be estimated as:

$$SU_i\ (t)_i = \sum_{k=1}^{n} \Big(\xi_a^i\ (t)_i{}^{1,\ldots,n} + \xi_{mn}\ (t)_i{}^{1,\ldots,n} + \xi_{nn}(t)_i{}^{1,\ldots,n} + \ \xi_{eme}\ (t)_i{}^{1,\ldots,n}\Big)^k \tag{48}$$

Random challenging bits sent by SU_j at time t_j can be denoted as:

$$SU_j(t)_j = \xi_a^i\ (t)_j^{1,3,\ldots,n} + \xi_{mn}\ (t)_j^{1,5,\ldots,n} + \xi_{nn}(t)_j^{1,6,\ldots,n} + \xi_{eme}\ (t)_j^{1,4\ldots,n} \tag{49}$$

The sum of the random challenging bits sent by SU_i can be estimated as:

$$SU_j\ (t)_j = \sum_{k=1}^{n} \Big(\xi_a^i\ (t)_j{}^{1,3,\ldots,n} + \xi_{mn}\ (t)_j{}^{1,5,\ldots,n} + \xi_{nn}(t)_j{}^{1,6,\ldots,n} + \ \xi_{eme}\ (t)_j{}^{1,4\ldots,n}\Big)^k \tag{50}$$

The sum of the both sequential and random challenging bits (CBs) sends by SU_i numerically indicated as:

$$CBs^k = \sum_{k=1}^{n} (SU_i\ (t)_i)^k + (SU_j\ (t)_j)^k \text{ where } k = 1, 2, \ldots, n \tag{51}$$

$$CBs^k = \sum_{k=1}^{n} (SU_i\ (t)_i + SU_j\ (t)_j)^k \tag{52}$$

The probability of sequentially and random challenging bits received by CH_i at time $(t)_{i+1,\ j+1}$ are shown as:

$$CH_i(t)^1_{i+1,j+1} = \xi_a^i(t)^{\alpha}_{i,j} + \xi_{mn}\ (t)^{\alpha}_{i,j} + \xi_{nn}(t)^{\alpha}_{i,j} + \xi_{eme}\ (t)^{\varepsilon}_{i,j} \tag{53}$$

$$CH_i(t)^2_{i+1,\ j+1} = \xi_a^i(t)^{\varepsilon}_{i,j} + \xi_{mn}\ (t)^{\varepsilon}_{i,j} + \xi_{nn}(t)^{\varepsilon}_{i,j} + \xi_{eme}\ (t)^{\varepsilon}_{i,j} \tag{54}$$

$$CH_i(t)^3_{i+1,j+1} = \xi_a^i(t)^{\alpha}_{i,j} + \xi_{mn}\ (t)^{\varepsilon}_{i,j} + \xi_{nn}(t)^{\varepsilon}_{i,j} + \xi_{eme}\ (t)^{\varepsilon}_{i,j} \tag{55}$$

$$CH_i(t)^k_{i+1,j+1} = \xi_a^i(t)^{\varepsilon}_{i,j} + \xi_{mn}(t)^{\varepsilon}_{i,j} + \xi_{nn}(t)^{\varepsilon}_{i,j} + \xi_{eme}\ (t)^{\varepsilon}_{i,j} \tag{56}$$

The sum of the challenging bits response generated by CH_i to SU_i at time t_l can be numerically revealed as:

$$rCH_i(t)_l{}^k = \sum_{k=1}^{n} \Big(r\xi_a^i\ (t)_{i,j}{}^{\varepsilon} + r\xi_{mn}\ (t)_{i,j}{}^{\varepsilon} + r\xi_{nn}(t)_{i,j}{}^{\varepsilon} + \ r\xi_{eme}\ (t)_{i,j}{}^{\varepsilon}\Big)^k \tag{57}$$

Thus, the sum of the sequentially and random challenging bits sent by CH_i to SU_i can be represented as follows:

$$CH_i\ (t)_i = \sum_{k=1}^{n} \Big(\xi_a^i\ (t)_i{}^{1,\ldots,n} + \xi_{mn}\ (t)_i{}^{1,\ldots,n} + \xi_{nn}(t)_i{}^{1,\ldots,n} + \ \xi_{eme}\ (t)_i{}^{1,\ldots,n}\Big)^{k'} \tag{58}$$

$$CH_j\ (t)_j = \sum_{k=1}^{n} \Big(\xi_a^i\ (t)_j{}^{1,3,\ldots,n} + \xi_{mn}\ (t)_j{}^{1,5,\ldots,n} + \xi_{nn}(t)_j{}^{1,6,\ldots,n} + \ \xi_{eme}\ (t)_j{}^{1,4\ldots,n}\Big)^{k'} \tag{59}$$

The sum of the challenging bits response probability generated by SU_i to CH_i at time t_l can also be numerically obtained as:

$$rCH_i(t)_l{}^k = \sum_{k=1}^{n} \Big(r\xi_a^i\ (t)_{i,j}{}^{\varepsilon} + r\xi_{mn}\ (t)_{i,j}{}^{\varepsilon} + r\xi_{nn}(t)_{i,j}{}^{\varepsilon} + \ r\xi_{eme}\ (t)_{i,j}{}^{\varepsilon}\Big)^{k'} \tag{60}$$

which may be written as:

$$rCH_i(t)_i = r\xi_a^i(t)_i^o + r\xi_{mn}\ (t)_i^o + r\xi_{nn}(t)_i^o + r\xi_{eme}\ (t)_i^o \tag{61}$$

The probability of the both SU_i and CH_i that ε errors are detected correctly for the half (1/2) of the time at ith section over i^{th} channel by $SU_i \| CH_i$ due to ξ_a^i, ξ_{mn}, ξ_{nn} and ξ_{eme} factors are obtained as follows

$$P[SU_i \parallel CH_i] = \sum_{k=1}^{n} P\big[\xi_a^i + \xi_{mn} + \ \xi_{nn} + \ \xi_{eme}\big]^k \tag{62}$$

Since the $SU_i \| CH_i$ guesses the challenging bits at each round independently, thus

$$P_b\big[\xi_a^i + \xi_{mn} + \ \xi_{nn} + \ \xi_{eme}\ \big] = \prod_{i=n1+k(j-1)+1}^{n+kj} \left(\frac{1}{2}\right)^k \tag{63}$$

if $\xi_a^i + \xi_{mn} + \ \xi_{nn} + \ \xi_{eme} \ = \xi_{amne}$ then Eq. (63) becomes

$$P_b[\xi_{amne}] = \prod_{i=n1+k(j-1)+1}^{n+kj} \left(\frac{1}{2}\right)^k \tag{64}$$

The probability of maximum ε errors are detected in $n_s = \frac{n_2}{k}$ sections can be computed by using the probability of the exactly p errors are detected by $SU_i \| CH_i$ in each section by the following multinomial distributions for every particular k.

$$\text{For } K = 0,\ P_b\big[\xi_a^i + \xi_{mn} + \ \xi_{nn} + \ \xi_{eme}\ \big] \text{ is } \sum_{i_{1=0}}^{\varepsilon} P_b[\xi_{amne0}]^{i_1} \times \frac{(n_s)!}{(i_1)!(n_s - i_1)!} \tag{65}$$

where ns ∈ round numbers and for $K=1$ P_b $[\xi_a^i+\xi_{mn}+\ \xi_{nn}+\xi_{eme}\ ]$ is

$$\sum_{i_2=0}^{\frac{\varepsilon}{2}}\sum_{i_1=0}^{\varepsilon-2i_2} P[\xi_{amne1}]^{n_s-i_1-i_2} P[\xi_{amne0}]^{i_1}\times P_b[\xi_{amne1}]^{i_2} \times \frac{(n_s)!}{(i_1)!(i_2)!(n_s-i_1-i_2)!} \tag{66}$$

For $K=2\,P_b\left[\xi_a^i+\xi_{mn}+\xi_{nn}+\xi_{eme}\ \right]$ is

$$\times\sum_{i_3=0}^{\lfloor\frac{\varepsilon}{3}\rfloor}\sum_{i_2=0}^{\lfloor\frac{\varepsilon}{2}\rfloor-3i_3}\sum_{i_1=0}^{\varepsilon-3i_3-2i_2} P_b[\xi_{amne_2}]^{n_s-i_1-i_2-i_3} \times P_b[\xi_{amne_0}]^{i_1}P_b[\xi_{amne_1}]^{i_2}P_b[\xi_{amne_2}]^{i_3} \times\frac{(n_s)!}{(i_1)!(i_2)!(i_3)!(n_s-i_1-i_2-i_3)!} \tag{67}$$

Thus, the probability of $(\xi_a^i+\xi_{mn}+\ \xi_{nn}+\ \xi_{eme}\ )_{\varepsilon,\ n_2-1,\ n_s}$ can be summarized as:

$$P_b[(\xi_{amne_0},i)]+P_b[(\xi_{amne_1},i)]+P_b[(\xi_{amne_2},i)]+\ldots+P_b[(\xi_{amne_\varepsilon},i)] \tag{68}$$

Thus, an ith channel C_i with least sum of errors $\varepsilon_1 \parallel \varepsilon_2=\xi_{amne_\varepsilon}\geq\varepsilon_o$ will be occupied by the SU_i for the quality communication for specific time intervals t_k in the ith round which can be numerically expressed as:

$$\omega=\begin{cases}C_i & \text{if } \xi_{amne_\varepsilon}\geq\varepsilon_o\\ \varepsilon_1\parallel\varepsilon_2<\varepsilon_o & \text{otherwise}\end{cases} \tag{69}$$

while the channel with errors $\varepsilon_1\|\varepsilon_2\ <\ \varepsilon_o$ will be discarded. This entire process repeats n time until a channel with an appropriate quality compatible to environment is selected for efficient and reliable data transmission in the network.

3.3. Energy efficient honey bee mating optimization routing algorithm (EHRA)

One of the main objectives of proposed routing algorithm is to route data packets over loop free spanning tree from source towards destination in the network. In the proposed routing algorithm, observed information is routed over a set of cluster heads by considering minimum distance and over-assignment cost to reduce high transmission energy consumption and packet loss. After dividing the entire sensor nodes into multiple clusters an initial population of individual bees is generated (of say size CHs = 23), using a random number generator and each individual is evaluated by a predefined fitness function. The individual in the initial population of bees (CHs) with greater fitness value is chosen as the queen of the hive while all other remaining individuals of the population are the drones and ranked decreasingly in the list. The current selected queen $\mathcal{Q}_i$ (the best next hop CH leading to the lowest energy and over-assignment cost for robust data delivery) in the mating flight randomly selects some drones (CHs) from the individual population list and mates with each individual drone. During each mating individual drone's genome is stored in her spermatheca (set to maximum 3 neighboring CHs). The probability of mating the drones and a queen $\mathcal{Q}_i$ can be numerically computed by taking into account the Eq. (5). In breeding process, a drone genome is randomly selected from the queen spermatheca and combine with the queen own genome by applying cross over operator which leads to genome of a new bee (the best next hop relay CH). To improve the quality of generating broods, a loop of selection, recombination/crossover and mutation operators is applied to each individual through the predefined probabilities as defined in ESHC algorithm. This entire process repeats iteratively until the termination criterion is satisfied for the complete cluster based routing solution over a set of cluster heads in greedy manner from the source towards sink. Let's consider a fully connected undirected network graph, selection operator assigns each cluster head in the population a set of weight $G(V, E)$, where a finite set of vertices $V=\ \{1,2,\ldots,N\}$ and edge $E\ =\ \{(i,j)|i,j\ \in\ V\}$ are connected between each cluster head node in the system. Every edge in the fully connected graph assigned a positive real number value denoting distance among cluster heads can be represented as $w=\ \{w_1,\ w_2,\ldots,w_{(N-1)N}\}$, where decision variable can be expressed as:

$$CH_{ij}=\begin{cases}1 & \text{if edge}(i,j)\text{is slelected}\\ 0 & \text{otherwise}\end{cases}\quad \text{for}\ \forall\, edges(i,j)\in E \tag{70}$$

Then, according to prime algorithm if V_1- Sink, E_o- null, and $V_2=\ V-V_1$ then by selecting a neighboring cluster head edge E_1 with minimum distance with edge weight w_i toward the sink. Let assume that E_i is directly connected to the sink then set $E_1=\{Sink,\ E_i\}$, $E_o=\{Sink,\ E_i\}$, $V_2=\ V_2-V_1$. Thus, for each individual cluster head CH_i in V_1 selects least weight $w_{i,j}$ where $CH_i\in V_1$, $V_j\in V_2$ and $e_o=(V_i,V_j)$ but is not E_o then $V_1=V_1\cup\ V_j, E_o=\{V_i,\ V_j\}\cup e_o$, $V_2=V_2-V_j$ as minimum spanning tree from the CH_i toward the sink. Consequently, if there are N total numbers of clusters in the minimum spanning tree network and there are n numbers of sensor nodes in each cluster where a cluster head can be represented as CH_i in each individual cluster. Then the average shortest path (ASP) for N number of cluster heads is the length of a shortest path from a cluster head node i to sensor node k towards the sink given as:

$$ASP=\frac{1}{N(N-1)}\sum_{i=1}^{k}[CH_1+,\ldots,CH_k\ ] \tag{71}$$

In EHRA, for measuring CH_i over-assignment cost, the entire network is divided into internal and external traffic by considering M/M/1 and M/M/N queuing models as discussed in [23], respectively. The sensor nodes in the same cluster generate internal traffic while the external data traffic cost is from the other cluster heads can be numerically indicated as:

$$\mathcal{C}_i=\ \sum_{r=1}^{n}\mathcal{C}_j\cdot\lambda_r(1-P_b) \tag{72}$$

where λ_r is the packets arrival rate from a neighboring cluster $\mathcal{C}_j$ with probability P_b. The packets arrival rate from n number of clusters to a cluster $\mathcal{C}_i$ can be indicated as

$$\mathcal{C}_i=\ \sum_{k=1}^{n}\sum_{r=1}^{n}\mathcal{C}_k\cdot\lambda_r(1-P_b) \tag{73}$$

Jobs blocking probability (BP_b) of a cluster head $\mathcal{C}_i$ when queue is full can be indicated as

$$BP_b(\mathcal{C}_i)=\ \sum_{k=1}^{n}\sum_{r=1}^{n}L_k(\mathcal{C}_i)\cdot\ \mathcal{C}_r\lambda_r(1-P_b) \tag{74}$$

where L_k denotes the queue length of a cluster head for storeddata packets.

To represent the connection between a cluster head to its neighboring cluster head, we simply draw a $n\times n$ topological matrix as m_1 and $n\times m$ topological matrix as m_2. To show the connectivity among a set of neighboring clusters and also to cluster member sensor nodes (Sn_i), we use $m_1i\to j$ and $m_2i\to j$ in the entire network where the spanning tree matrix $[m_1m_2]$ define as $n\times(n+m)$. Then, we can formulate a topological problem as:

$$m_{1ij}=\begin{cases}1 & \text{if } Sn_i\in CH_j\\ 0 & \text{otherwise}\end{cases}\quad \text{for}\ i,j=1,2,\ldots,n \tag{75}$$

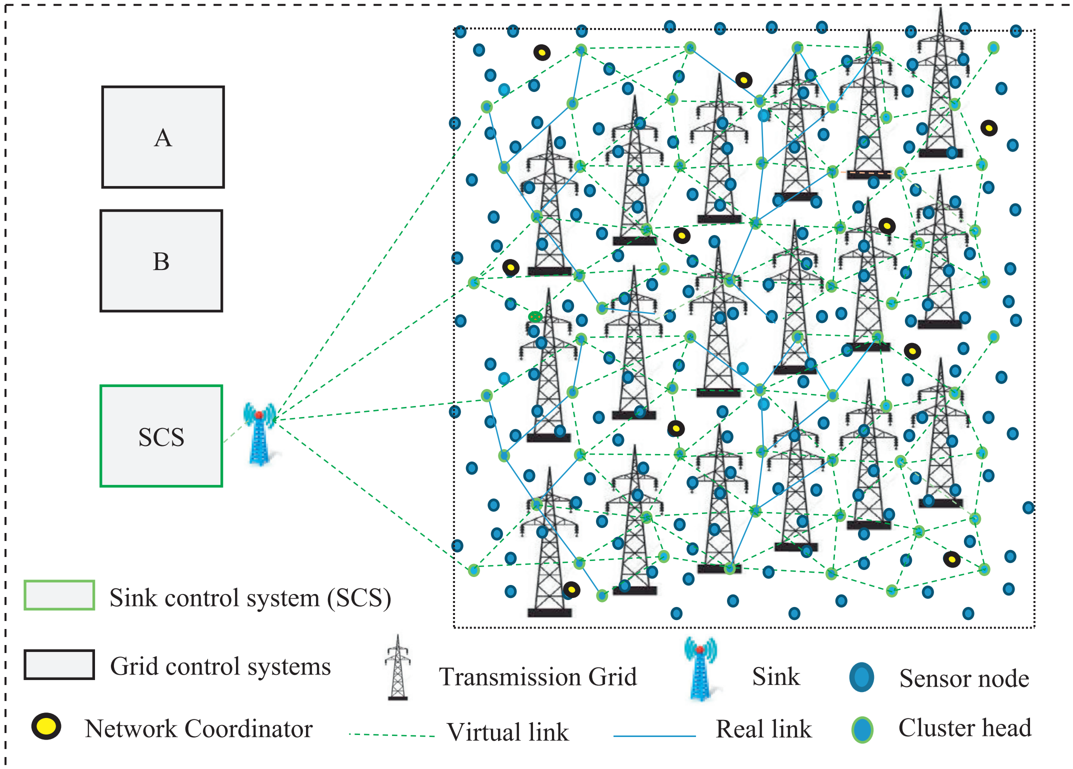

Fig. 3. Network model for proposed EQSHC scheme.

$$m_{2ij} = \begin{cases} 1 & \text{if } CH_i \in CH_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i, j = 1, 2, \ldots, n \tag{76}$$

$$\min \frac{1}{\delta_t} \left[\sum_{i=1}^{n} \frac{T_{CH(i)}}{DTC_i - T_{CH(i)}} + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{CH(i,j)} g T_{CH(i,j)} \right] + \min \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{lij} g m_{1ij} + \sum_{i=1}^{n} \sum_{j=1}^{m} w_{2ij} g m_{2ij} \tag{77}$$

where δ_i denotes the entire offered data traffic, $T_{CH(i)}$ shows the maximum data traffic at cluster head *i*, $T_{CH(i,\,j)}$ denotes total data traffic through the link (*k*, *l*), $DTC_{(i)}$ represents the data traffic capacity of cluster head *i*, $d_{CH(i,\,j)}$ indicates the distance between cluster head node *i* and cluster head node *j*, *g* is a constant set to 0.1 and, w_{1ij} and w_{2ij} show the weight of the link between sensor node *i* to cluster head *j*, and cluster head *i* to cluster head *j*, respectively.

4. Network model

In this study, we consider a smart grid environment where a power grid station under 500 *KV* is considered as the local power distribution area (LPDA) as shown in Fig. 3. The primary substation including a base station and primary data management center has an important role for regional power management and controls substations. A power distribution network in which 350 transmission grids are involved is considered to test the designed protocol effeteness. The distance between each individual transmission grid is assumed to be 15 to 20 m in local power distribution area. In our performance evaluations, we consider some assumptions. First, after deployment, all sensor nodes and multiple network coordinators are static and only can communicate with each other if they are in range. Second, all randomly deployed sensor nodes have the same capability in terms of initial energy and transmission ranges. Third, omnidirectional antenna model is considered. Fourth, we assumed asymmetric communication model. In addition, unlike most of the existing work to test routing protocols in a more realistic smart grid scenario, we assumed that the arrival time follows Poisson distribution, while the service time is considered exponentially distributed.

4.1. Path loss model

In this study, we used log-normal shadowing path-loss model presented in [4]. Recent studies show that this model provides more accurate multipath channel estimation results compared to the Nakagami and Rayleigh models in a smart grid environment. The mathematical form of the log-normal shadowing model is numerically expressed as:

$$\gamma(d)_{dB} = P_t - PL(d_0) - 10\eta \log_{10} \frac{d}{d_0} - X_\sigma - P_\eta \tag{78}$$

where $\gamma(d)$ represents signal to noise ratio, P_t is the transmit power in dBm, $PL(d_0)$ shows the path loss at a reference distance d_0, η indicates the path-loss exponent, X_σ denotes a zero mean Gaussian random variable with standard deviation σ, and P_η express the noise power measured in dBm. Here, path loss exponent for line of sight (LOS) and non-line of sight (NLOS) substation environments are set to 2.42 and 3.51, respectively. While the shadowing deviations for both LOS and NLOS substation environments are set to 3.12 and 2.95, respectively. Noise floor for LOS and NLOS are set to −93 dBm, respectively [4]. Please note that this noise level is much higher than that of outdoor noise levels, which is measured as – 105 dBm.

4.2. Simulation model

In our simulation study, we consider that 200 sensor nodes having maximum data rate of 250 kbit/s are deployed randomly for monitoring smart grid systems. These sensor nodes are deployed in two-dimensional smart grid area 2D(M × M) of length and width with values (1200 × 1200) meters. In smart grid, the responsibility of each node is to obtain a certain view of the employed environment. However, it can only cover a limited physical area of the environment due to its limited communication and sensing range. The maximum communication range of each sensor node is set to 100 m. Total number of network coordinators and resource-rich nodes, is 20. The initial energy, data aggregation, ideal listening, sleeping and receiving power is set to 5 joules, 0.043 W, 0.13 W, 3×10^{-6} W and 0.35 W, respectively. These energy consumption values are typical values of the sensor nodes used in smart grid applications [2-4]. Moreover, the high and low transmission power were set to 0.97 W and 0.87 W, respectively. In addition, we set data packet length to 40 bytes. The performance evaluations consist of 50 sets of simulations. All sensor nodes in a region r_i sense the data periodically and forwards sensed data to cluster head in the network. The performance results discussed in Section 5 are generated using MATLAB.

5. Performance results and discussions

To evaluate the performance of the proposed scheme against the recent routing schemes, we take into account three basic parameters explained below:

5.1. Packet delivery ratio (PDR)

is regarded as the ratio between the number of packets successfully received by the sink and those sent by the source node and can be numerically defined as:

$$PDR = \left(\sum_{i=1}^{n} D_{sink}(i) \Big/ \sum_{i=1}^{k} D_{sni}(i) \right) \times 100\% \tag{79}$$

where $D_{sni}(i)$ and $D_{sink}(i)$ are the number of data packets send by the source node and data packets received successfully at the sink, respectively.

5.2. Delay (De)

is the time when the first copy of a data packet sent along the shortest path from the source node i to sink (s). The average delay of all n nodes of a network having n number of sensor nodes can be formally indicated as:

$$De_{avg} = \frac{1}{n}\sum_{i=1}^{n} De(i) = \frac{1}{n}\sum_{i=1}^{n}\left(t_i^{sink} - t_i^{sni}\right) \tag{80}$$

where t_i^{sni} and t_i^{sink} denote the packet sending time at source i and the packet receiving time at the sink, respectively.

5.3. Energy consumption (Ec)

can be numerically defined as:

$$Ec = \sum_{i=1}^{n}\left(\sum\nolimits^{CA} + \sum\nolimits^{BB} + \sum\nolimits^{RC} + \sum\nolimits^{C} + \sum\nolimits^{ID} + \sum\nolimits^{DA}\right) \tag{81}$$

where total energy consumption is the sum of channel assignment cost $\sum^{CA}$, backbone construction cost $\sum^{BB}$, is rib node joining cost $\sum^{RC}$, communication cost $\sum^{C}$, idle energy consumption $\sum^{ID}$ and data aggregation cost $\sum^{DA}$ for each individual round $i = 1$ to n.

The PDR of each routing scheme vs. number of channels for SG applications is shown in Fig. 4. The PDR of EQSHC, SCR and HRL-AODV is dependent on offered traffic load and deliver data up to 94%, 84% and 81% of the originated data packets using default channel in the deployed network, respectively. However, the PDR of EQSHC rapidly increases and delivers data up to 100% of the originated data packets when the number of channels is assumed to be 19 as shown in Fig. 4. Similarly, the PDR of SCR also increases and delivers data up to 93% of the originated data packets when the numbers of channels is 22. On the other hand, the PDR of HRL-AODV using default static channel remains maximum 81%, and does not increase. This is because the HRL_AODV is not considering the dynamic channel adaptation in SG. On the other hand, the increase in size of the channel history significantly increases the channel selection reliability and is a tradeoff between routing table management complexity and processing delay for channel assignment in both EQSHC and SCR routing schemes. Herein, it is observed that in all routing schemes the PDR depends on the path selection that are relatively stable, less congested and in which the intermediate sensor nodes hold high capacity data channel. It also depends on the intermediate nodes in the routing path because as the number of intermediate hops increases the probability of route failures also increases. Moreover, it is noticed that high PDR in all routing schemes is also affected by the excessive end-to-end packet delay and PU activity to gain channel control instantaneously. When a PU appears in the channel occupied by a SU, the SU has to switch to another free channel with least switching delay. The SU will be dropped if there is no other free channel. Also, when a new SU wants to initiate a session, it will be blocked if all the primary channels and secondary channels are occupied, otherwise, it selects a channel with high data capacity and least error probability to fulfill the applications requirement. These high data capacity channels significantly improve PDR in EQSHC compared to SCR routing scheme for SG applications. Considering the Poisson distribution, we assume the initial transition intensity of a SU as λ. After appearing PU, the new transition intensity is given by:

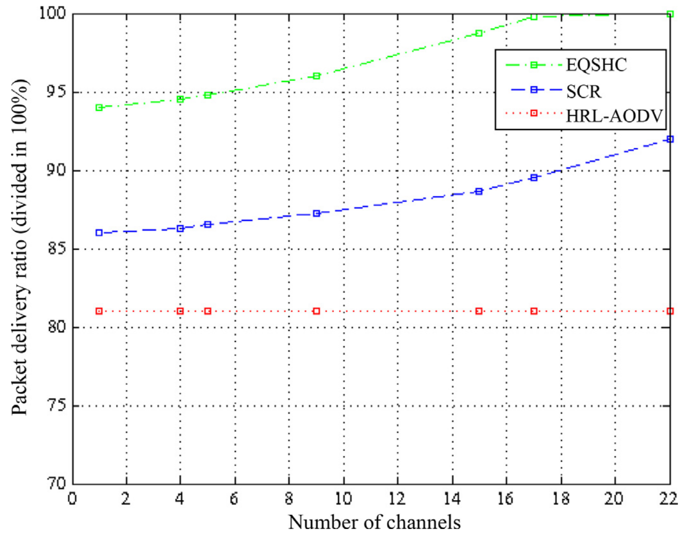

Fig. 4. Packet delivery ratio vs number of channels Fig. 5. SU throughput vs PU activity.between 1 and 22.

$$\lambda_{switch}^{(1-i)} = \frac{1}{\left(\frac{1}{\lambda} + 1\right)} + t_{switch}^{(1-i)} \tag{82}$$

where $t_{switch}^{(1-i)}$ $(i = 1, 2)$ is the delay for a SU to switch from a channel to a primary channel $(i = 1)$ or from a primary channel to a secondary channel $(i = 2)$. During channel handover process, we can get the blocking probability of SUs as:

$$P_{block}^{(SU)} = \sum_{i=0}^{c_1} P_{i,\ c_i - i, c_2} \tag{83}$$

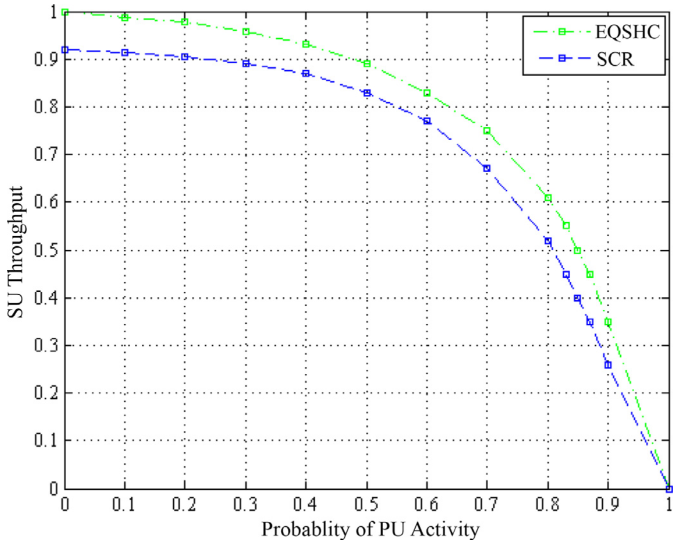

Fig. 5. SU throughput vs PU activity.

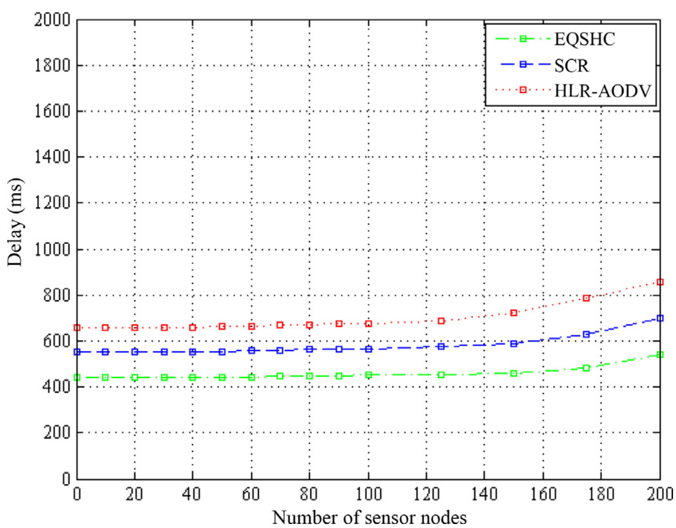

Fig. 6. Delay vs node density between 1 and 200.

The dropping probability of SUs is given by:

$$P^{(SU)}_{block} = \lambda_1 \sum_{i=0}^{c_1} P_{i,\ c_i-i,c_2} / \left[\lambda_2\left(1 - P^{(SU)}_{block}\right)\right] \tag{84}$$

The throughput of SUs is given by:

$$\text{Throughput}(SU) = \ \lambda_2\left(1 - P^{(SU)}_{block}\right)\left(1 - P^{(SU)}_{block}\right) \tag{85}$$

where Poisson with a rate λ_1 and λ_2 are the call arrival process of PUs and SUs, respectively. The SUs throughput is highly pretentious by the PUs activity and decreases with the increase in PUs activity to gain channel control periodically for monitoring events in the smart grid as shown in Fig. 5. Herein, it is observed that the proposed scheme leads to high throughput compared to SCR routing scheme.

The network delay in all routing schemes depends on the data traffic congestion level and intermediate hops count in a specific data path from the source to destination in the network. It also indirectly depends on the sensor nodes residual energy in the selected path and path reliability, which are the two main possible reasons for route failure in the network. The delay of all routing schemes for SG applications is also shown in Fig. 6. It clearly shows that our proposed scheme achieves less network delay compared to other routing scheme during forwarding information from the source towards the destination. Routing failures in both HRL-AODV and SCR routing schemes are due to frequent rerouting, which increases the time spent by witnessed data packets in the sending cache of the originated sensor node, as packets need to wait for new paths to be established. Moreover, the node's buffer length in our proposed scheme adversely affects and plays a major role to achieve low end-to-end delay and increased packet delivery in the network. The increase in node buffer length means to reduce the probability of congestion when some new sensor nodes join new data paths and vice versa. On the other hand, small buffer size as in HRL-AODV and SCR routing schemes would increase the probability of data packet loss due to node buffer overflow, which may reduce high PDR in the network. The other reason of decreasing high packet delivery ratio in HRL-AODV is data packet collisions due to high RF interference among the sensor nodes. A slightly better PDR of the SCR is due to its consideration of better link stability among nodes during forwarding information in an appropriate manner. In addition, in HRL-AODV observed information is conveyed indirectly via multi-hops in the network. Here, we expect if the route length increases, then number of intermediate unnecessary hops also increases, which makes a certain amount of data packets invalid due to not delivering at a predefined threshold time to the sink node. On the other hand, the HRL-AODV has a significant tail for some packets, taking up to 6 or more hops longer than optimal number of hops, which decreases a significant amount of PDR in the network. Overall, high network delay in both HRL-AODV and SCR routing schemes causes a certain amount of data packets dropped due to the creation of path loops from source to destination in the network.

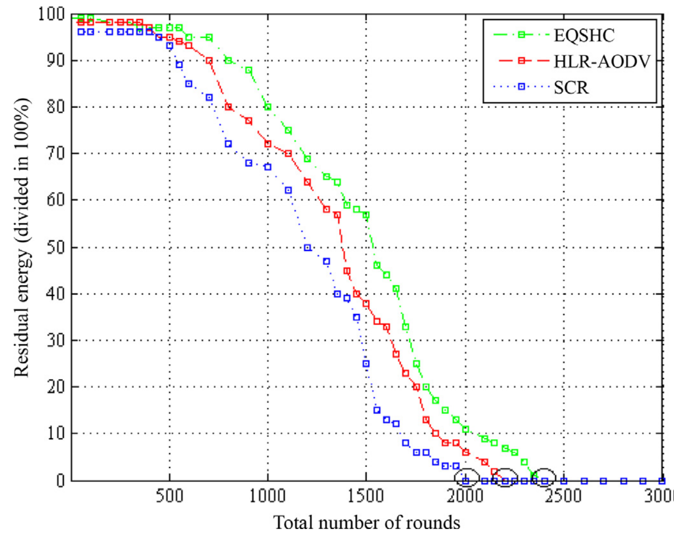

Fig. 7. Residual energy vs round numbers between 1 and 3000.

The residual energy profile of all routing schemes in smart grid deployment is shown in Fig. 7. In all routing schemes, the nodes battery lifespan depends on the most recent values of residual energy during the time of path election and aggregation of the observed information waiting to be conveyed on the selected path in the smart grid deployment. The low residual energy profile in HRL-AODV and SCR routing schemes is due to their frequent route failures from source to destination. To discover effective routes, excessive message retransmissions are required, which consumes a significant amount of nodes energy leading to low residual energy in the network. In HRL-AODV, routing paths with high number of hops lead to significant amount of energy consumption in the network. In addition, high interference is another main challenging issue in SCR and increases the data packet corruption rate leading to high energy consumption in the network. Residual energy profile is found better in HRL-AODV compared to SCR routing scheme. To achieve high residual energy profile in our proposed scheme, the proposed clustering algorithm (ESHC) plays an important role by providing a robust and reliable energy efficient clustering architecture at low complexity in the network. To this end, the ESHC algorithm is responsible to distribute energy load evenly by appointing a set of normal nodes as cluster heads based on their residual energy and minimum Euclidean distance information in the entire cluster network.

Furthermore, in our proposed EQSHC scheme, the channel assignment algorithm assigns high quality channels among sensor nodes by taking into account the highly noisy smart grid spectrum environments. Thus, it saves network lifetime by decreasing data packets retransmission and overhead in the network. In addition, to distribute energy load evenly in the entire clustering network, the proposed routing algorithm plays an important role by providing highly stable links among cluster heads based on distance information from the source towards the destination in the smart grid. In the proposed routing algorithm, the entire network is divided into multiple loop-free minimum spanning trees. In the proposed algorithm, if an appropriate relay cluster head node is not found in the range, then a sender cluster head can switch to its high transmission power. This process helps to find an appropriate next hop relay node in case of sparse network deployment in smart grid. Moreover, during data forwarding phase, if a next hop CH does not reply in defined time interval t_i, then a next hop node with the second highest priority value is selected to convey information in a greedy manner from the source towards the sink. This entire process repeats until the observed information is reached from the source node to the sink. Thus, it reduces the impact of control message overheads and saves a significant amount of network energy compared to other routing schemes. Overall, performance results show that the proposed scheme successfully addresses the QoS requirements of most of the SG applications presented in Table 1. However, some applications, such as substation automation, overhead transmission line monitoring, wide-area situational awareness, are still challenging due to their specific strict latency requirements. Adaptive transmission power control algorithms and cross-layer communication protocols could be promising for these applications to reduce communication delay and the development of such protocols is left as a future work.

6. Conclusion

Recently, WSNs have been recognized as a promising technology for smart grids, next-generation power systems, due to their collaborative nature. The main deployment challenge of WSN-based smart grids is to establish reliable, energy-efficient communications in regard to harsh smart-grid spectrum environments. This paper proposes a spectrum-aware bio-inspired routing algorithm in CRSNs, which maximizes spectrum efficiency with minimum consumed energy. Specifically, honey-bee mating optimization-based routing and cooperative channel assignment algorithms have been jointly proposed that opportunistically switch among spectrum bands and employ clustering-based routing to organize CRSNs into a connected hierarchy. The developed solution balances the loads of energy consumption and prolonging the lifetime of CRSNs in smart grids. Performance evaluations show that our solution outperforms existing designs in terms of packet delivery ratio, transmission delay, and residual energy, thus facilitating various smart-grid applications. The future work will be the implementation of an adaptive, cross-layer communication framework for WSN-based smart-grid applications.

Acknowledgements

This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH) – King Abdulaziz City for Science and Technology - the Kingdom of Saudi Arabia – award number (12-INF2731-03). The authors also, acknowledge with thanks Science and Technology Unit, King Abdulaziz University for technical support.

References

[1] V.C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, et al., A survey on smart grid potential applications and communication requirements, Ind. Inf., IEEE Trans. 9 (2013) 28–42.

[2] E. Fadel, V.C. Gungor, L. Nassef, N. Akkari, M.G.A. Malik, S. Almasri, et al., A survey on wireless sensor networks for smart grid, Comput. Commun. 71 (2015) 22–33 11/1/.

[3] C. Eris, M. Saimler, V. Gungor, E. Fadel, I. Akyildiz, Lifetime analysis of wireless sensor nodes in different smart grid environments, Wireless Netw. 20 (2014) 2053–2062 2014/10/01.

[4] V.C. Gungor, B. Lu, G.P. Hancke, Opportunities and challenges of wireless sensor networks in smart grid, Ind. Electron. IEEE Trans. 57 (2010) 3557–3564.

[5] G.A. Shah, V.C. Gungor, O.B. Akan, A cross-layer QoS-aware communication framework in cognitive radio sensor networks for smart grid applications, Ind. Inform. IEEE Trans. 9 (2013) 1477–1485.

[6] H. Farooq, L.T. Jung, Health, link quality and reputation aware routing protocol (HLR-AODV) for wireless sensor network in smart power grid, in: Computer & Information Science (ICCIS), 2012 International Conference on, 2012, pp. 664–669.

[7] J. Lin, W. Yu, D. Griffith, X. Yang, G. Xu, C. Lu, On distributed energy routing protocols in the smart grid, in: R. Lee (Ed.), Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 492, Springer International Publishing, 2013, pp. 143–159.

[8] R. Hou, C. Wang, Q. Zhu, J. Li, Interference-aware QoS multicast routing for smart grid, Ad Hoc Netw. 22 (2014) 13–26.

[9] X. Li, Q. Liang, F.C. Lau, A maximum likelihood routing algorithm for smart grid wireless network, EURASIP J. Wireless Commun. Networking 2014 (2014) 1–7.

[10] X. Deng, L. He, X. Li, Q. Liu, L. Cai, Z. Chen, A reliable QoS-aware routing scheme for neighbor area network in smart grid, Peer-to-Peer Networking Appl. (2015) 1–12.

[11] K. Kim, S.-i. Jin, Branch-based centralized data collection for smart grids using wireless sensor networks, Sensors 15 (2015) 11854–11872.

[12] S. Kim, Biform game based cognitive radio scheme for smart grid communications, J. Commun. Netw. 6 (2012) 614–618.

[13] B. Erman Bilgin, V. Cagri Gungor, Adaptive error control in wireless sensor networks under harsh smart grid environments, Sens. Rev. 32 (2012) 203–211.

[14] S.C. Lin, K.C. Chen, Improving spectrum efficiency via in-network computations in cognitive radio sensor networks, Wireless Commun. IEEE Trans. 13 (3) (2014) 1222–1234.

[15] G.A. Shah, O.B. Akan, Spectrum-aware cluster-based routing for cognitive radio sensor networks, in: Communications (ICC), 2013 IEEE International Conference on, 2013, pp. 2885–2889.

[16] B. Alik, M. Teguar, A. Mekhaldi, Minimization of grounding system cost using PSO, GAO, and HPSGAO techniques, Power Delivery IEEE Trans. 30 (2015) 2561–2569.

[17] M. Niu, Z. Xu, Efficiency ranking-based evolutionary algorithm for power system planning and operation, Power Syst. IEEE Trans. 29 (2014) 1437–1438.

[18] Y. Marinakis, M. Marinaki, G. Dounias, Honey Bees Mating Optimization algorithm for large scale vehicle routing problems, Nat.Comput. 9 (2010) 5–27 2010/03/01.

[19] J. Senthilkumar, M. Chandrasekaran, Y. Suresh, S. Arumugam, V. Mohanraj, Advertisement timeout driven bee's mating approach to maintain fair energy level in sensor networks, Appl. Soft Comput. 11 (2011) 4029–4035.

[20] M.C. Nascimento, A.C. De Carvalho, Spectral methods for graph clustering–a survey, Eur. J. Oper. Res. 211 (2011) 221–231.

[21] S. Ding, L. Zhang, Y. Zhang, Research on spectral clustering algorithms and prospects, Computer Engineering and Technology (ICCET), 2010 2nd International Conference on, 2010 V6-149-V6-153.

[22] H. Urkowitz, Energy detection of unknown deterministic signals, Proc. IEEE 55 (1967) 523–531.

[23] A.O. Allen, Probability, Statistics, and Queueing Theory, Academic Press, 2014.

Dr. Etimad Fadel Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia. Received her Bachelor's degree in Computer Science at King Abdul Aziz University with Senior Project title ATARES: Arabic Character Analysis and Recognition in 1994. She was awarded the Mphil/Ph.D. degree in computer science at De Montfort University (DMU), Liecester, UK. Currently she is working as an Assistant Professor at the Computer Science Department at KAU. This was between 2008 – 2010. Her main research interest is Distributed Systems, which are developed based on middleware technology. Currently she is looking into and working on Wireless Networks, Internet of Things and Internet of Nano-Things. In addition, she is working on smart grids and HetNets.

Muhammad Faheem received the B.*Sc.* Computer Engineering degree in 2010 from the Department of Computer Engineering at the University College of Engineering & Technology, Bahauddin Zakariya University Multan, Pakistan. In 2012, he received an M.S. degree in Computer Science from the Faculty of Computer Science and Information System at Universiti Teknologi Malaysia. Currently, he is a Ph.D. student at Abdullah Gul University, Kayseri, Turkey. His research interest includes wireless sensor networks and smart grid.

Vehbi Cagri Gungor received his B.S. and M.S. degrees in Electrical and Electronics Engineering from Middle East Technical University, Ankara, Turkey, in 2001 and 2003, respectively. He received his Ph.D. degree in electrical and computer engineering from the Broadband and Wireless Networking Laboratory, Georgia Institute of Technology, Atlanta, GA, USA, in 2007. Currently, he is an Associate Professor and Chair of Computer Engineering Department, Abdullah Gul University (AGU), Kayseri, Turkey. His current research interests are in smart grid communications, machine-to-machine communications, next-generation wireless networks, wireless sensor networks. Dr. Gungor has authored several papers in refereed journals and international conference proceedings, and has been serving as an editor, reviewer and program committee member to numerous journals and conferences in these areas. He is also the recipient of the IEEE Trans. on Industrial Informatics Best Paper Award in 2012, the European Union FP7 Marie Curie RG Award in 2009, Turk Telekom Research Grant Awards in 2010 and 2012, and the San-Tez Project Awards supported by Alcatel-Lucent, and the Turkish Ministry of Science, Industry and Technology in 2010.

Laila Nassef received the Bachelor of Science Degree in electronics and communications from Ain Shams University, Cairo, Egypt in 1984 and Master of Science Degree in electronics and computers from Ain Shams University, Cairo, Egypt in 1989. She received her Ph.D. degree in computer engineering from Anglia Polytechnic University, England in 1997. She joined the National Telecommunications Institute in Cairo, Egypt in 1985. She joined the Institute of Statistical Studies and Research of Cairo University, Cairo, Egypt in 2001. Currently, she is an associate professor with department of computer science, faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, KSA. Her research interests include wireless ad hoc and sensor networks, cognitive networks, heterogeneous networks and smart grid communications. Her research has been supported by the King Abdulaziz City of Science and Technology (KACST), KSA.

Nadine Akkari received the B.S and the M.S. degrees in Computer Engineering from University of Balamand, Lebanon, in 1997 and 1999, respectively. She received the Master degree in Networks of Telecommunications from Saint Joseph University and the Faculty of Engineering of the Lebanese University, Lebanon, in 2001 and Ph.D. degree in Networking from National Superior School of Telecommunications (ENST), France, in 2006. She is currently an associate professor with the faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, Saudi Arabia. She is a member of IEEE. Her research interests are in wireless networks, mobility management protocols, nanonetworks and smart grids.

M. G. Abbas Malik Muhammad Ghulam Abbas Malik was born in Lahore, Pakistan, in 1979. He received his Bachelor of Science (B. *Sc.* - Major in Applied & Pure Mathematics and Statistics) and Master of Science in Computer Science from University of the Punjab in 1999 and 2003 respectively. In 2006, he received his Master in Linguistics and Computer Science from University of Paris 7 - Denis Didrot. He received his Ph.D. degree in Computer Science from University of Grenoble (Ex. University Joseph Fourier), Grenoble, France, in July 2010. Currently, he is an Assistant Professor with Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah. His current research interests are in electromagnetic nano-networks, graphene-enabled wireless communication, Terahertz Band communication networks and the Internet of Nano-Things.

Suleiman Almasri received the Computer Science degree from Philadelphia University, Amman, Jordan in 1998 and MS degree in computer science from Amman Arab University, Amman, Jordan in 2005, and the PhD degree in computer science (wireless and mobile networks) from Anglia Ruskin University, Chelmsford, UK in 2009. He is currently an assistant professor of computer science at King Abdulaziz University, Jeddah, KSA and a founding member of the Advanced Wireless Networking Group in the university. His research interests include wireless networks, nano networks, Terahertz band communication, optical networks, mobile computing, internet of nano-things, location based services and smart grid.

Ian F. Akyildiz received the B.S., M.S., and Ph.D. degrees in Computer Engineering from the University of Erlangen-Nrnberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is the Ken Byers Chair Professor in Telecommunications with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, the Director of the Broadband Wireless Networking Laboratory and Chair of the Telecommunication Group at Georgia Tech. Dr. Akyildiz is an honorary professor with the School of Electrical Engineering at Universita tPolitcnica de Catalunya (UPC) in Barcelona, Catalunya, Spain and the founder of N3Cat (NaNo Networking Center in Catalunya). He is a Consulting Chair Professor at the Department of Information Technology, King Abdulaziz University (KAU) in Jeddah, Saudi Arabia. Since September 2012, Dr. Akyildiz is also a FiDi Pro Professor (Finland Distinguished Professor Program (FiDiPro) supported by the Academy of Finland) at Tampere University of Technology, Department of Communications Engineering, Finland. He is the Editor-in-Chief of Computer Networks (Elsevier) Journal, and the founding Editor-in-Chief of the Ad Hoc Networks (Elsevier) Journal, the Physical Communication (Elsevier) Journal and the Nano Communication Networks (Elsevier) Journal. He is an IEEE Fellow (1996) and an ACM Fellow (1997). He received numerous awards from IEEE and ACM. His current research interests are in nano networks, Long term evolution (LTE) advanced networks, cognitive radio networks and wireless sensor networks.