Data Aggregate Point Placement for Smart Grid with Joint Consideration of Communication and Power Networks

Amany Mahdy[†], Peng-Yong Kong[†], Bashar Zahawi[†], and George K. Karagiannidis[‡]
[†]Khalifa University of Science, Technology and Research, Abu Dhabi, United Arab Emirates.
[‡]Aristotle University of Thessaloniki, Thessaloniki, Greece.

Abstract—**Data aggregate point (DAP) placement is one of the great challenges in implementing smart grid. The placement problem involves determining the optimal number and locations of DAPs, while maintaining a desired level of quality for each communication link that connects a smart meter to its associated DAP. Having smart meters on different feeders in the power network, connecting to the same DAP in the communication network may lead to inefficient data aggregation. Therefore, the placement problem must not consider the communication network in isolation from the power network. In this paper, a DAP placement scheme is proposed to jointly consider the topologies of both power and communication networks, in achieving a higher communication efficiency, and fulfilling a desired level of transmission reliability. In the proposed scheme, a mixed-integer non-linear optimization is formulated to determine the number and locations of DAPs that would reduce the delay in the network, so that the amount of data traffic and the system cost are minimized. Numerical results show that having the number of DAPs equals to the number of feeders is the optimal one. In addition, it is concluded that having less number of feeders and lower data arrival rate would lead to a smaller delay.**

Index Terms—**Smart grid, data aggregate point, smart meter, node placement.**

I. INTRODUCTION

In the new smart grid paradigm, the move from automated meter reading (AMR) to the new advanced metering infrastructure (AMI) requires the integration of information and communication technology (ICT) into the power network [1], [2]. A substantial amount of data will need to be exchanged and transmitted between smart meters (SM), data aggregate points (DAP) and the utility center (UC) through the communication links if the benefits promised by this new model are to be realized. For proper operation of various smart grid applications, the data transmissions must satisfy some stringent quality of service (QoS) requirements [3]–[7]. For efficiency, these QoS requirements must be met at the minimum QoS. This paper studies the problem of finding the optimal number of DAPs, and their respective locations, to meet the QoS requirements at the minimum cost.

A umber of such optimization studies are available in the literature. A common approach for finding the optimal DAPs locations is to use heuristic algorithms such as the K-mean and K-median [8]. The cost of data transmissions was minimized in [9] by proposing a distributed minimum packet forwarding algorithm that would choose the best location for a fixed number (1, 3, 7, and 15) of DAPs. In [10], the cost is reduced by controlling the distance between the nodes and the DAPs while [11] suggests that the total installation and transmission cost could be decreased by using the overhead power line poles whenever available, to mount the DAPs. The delay between the smart meters and the DAPs is minimized in [12] by following the procedure presented in [11]. The K-mean algorithm is used to cluster the smart meters which are centered by one DAP. Binary programing is then used to choose the shortest path between the DAP and the poles located 50m away from it. Finally, the algorithm estimates the average delay using queuing theory with a finite buffer size. If the delay does not meet the target, the number of clusters increases by 1, and the loop is repeated until a satisfactory solution is obtained.

Another research area where the placement of DAPs is an important and well-studied subject is wireless sensor networks (WSN), where data is transmitted between the sensors and the relay nodes for further processing, noticing that relay node in WSN is equivalent to DAP in smart grid. The author in [13] is considered among the first to deal with the problem of constrained relay node placement, where relay nodes need to be placed in optimal locations to meet connectivity and survivability requirements. In [14], a WSN is used as an infrastructure in a smart building, where no line-of-sight exists. The algorithm uses mixed integer programming in two different kinds of relay placement (deterministic and robust placement) with the aim of minimizing cost. However, all of the algorithms mentioned above were concerned solely and exclusively with the communication network topology when determining the required number of DAPs and their locations. No previous study had ever considered the power network topology alongside that of the communications network. This is a serious and significant omission in smart grid applications where the topology of the power distribution network is very likely to have a significant impact on the system's cost and QoS.

In this paper, an optimization algorithm is proposed for the optimum placement of DAPs within a given geographical area by considering power distribution network topology along with the communication network topology to minimize the total latency and cost without compromising QoS. This is achieved

978-1-5090-5454-1/17/$31.00

by focusing on the idea of aggregating and compressing data associated with the same power feeder at the appropriate DAP before this is sent to the UC. This means a lower number of packets to be transmitted from the DAP to the UC, resulting in less data traffic and a decrease in the total delay in the system. Another benefit of aggregating the data in the DAP and the resulting lower number of packets is a significant reduction in the monetary cost of data transmissions. This is because data transmission from the smart meters to the DAP can be performed using the license-free radio frequency band, while transmission from the DAP to the UC incurs a real cost that increases with the number of packets being transmitted.

The rest of the paper is organized as follows: section II describes the proposed communications system model and the proposed representation of the power distribution network in the given geographical area. Four components of the communication models are discussed: media access control, retransmission delay, queuing delay and the propagation model. In section III, the method of solution and the proposed algorithm is considered. Evaluation results of the proposed algorithm are presented and discussed in section IV. The paper ends with the conclusion in section V.

II. SYSTEM MODEL

In this section, the geographical layout of the power and communication networks and the consideration are described first, followed by discussion of the communication network used in the study.

A. power and communication network

In this study, it is assumed that the geographical area under examination contains 33 kV/11 kV substations each supplying a number of 11 kV feeders which provide electricity to a variety of load types within the area including residential, commercial and industrial loads (with N_k being the total number of 11k feeders). It is also assumed that each house or load center will be equipped with a smart meter capable of transmitting relevant data to anyone of N_c DAPs distributed around the geographical area (with N_s being the total number of smart meters).

Each smart meter has a fixed location coordinate, i.e., (sm_x, sm_y), while the DAPs are placed at locations ($conc_x$, $conc_y$) to be identified by the proposed optimization algorithm. The distribution network operator (DNO) would be most interested in information related to each separate feeder (e.g. current-power flows, reactive power flows, etc.). Therefore, the optimization strategy proposed in this study is to optimize the number and locations of the DAPs taking into account the the location of the smart meters (loads) on the 11kV feeders and not simply the Euclidean distance between the smart meters and the DAPs as previous studies have tended to do. This means that two smart meters that are in close physical proximity may be connected to different DAPs depending on their locations on the electric power network.

B. Communication Network

For the proposed algorithm, it is assumed that there is only one hop between each SM and the DAP. The communication channel between them is IEEE 802.15.4g because this standard supports the TDMA that is needed in the algorithm and it targets the smart utility. IEEE 802.15.4g covers a large number of unlicensed frequency bands in order to span most regions. Worldwide, the frequency band 2400 to 2483.5 MHz, can offer up to sixteen communication channels [15]. In addition, the communication channel between the DAPs and the control center is a wire with high speed communication link. It is important to note that there is no communication between the smart meters.

The frame duration that is supported by the mentioned IEEE standard is assumed to be the same for all the smart meters within the same cluster where a cluster constitutes all smart meter on a feeder. Then, for a cluster j, the frame duration is described as follows:

$$T_{f_j} = \sum_{i=1}^{Ns} \frac{\chi_{kj} L}{\nu_{ij}}, \tag{1}$$

where χ_{kj} is a binary number with value of 1 when the the smart meter in the kth cluster is connected to the j-th DAP and zero other wise. χ_{kj} comes from the equation $\chi_{kj} = \frac{\alpha_k}{j}$, where α_k is a vector with N_k elements, each element is an integer number that could take a value from 1 to N_c. When $\alpha_k = j$, it means the k-th cluster is connected to the j-th DAP. The fraction $\frac{\alpha_k}{j}$ is equal to 1 when $\alpha_k = j$ and zero otherwise, resulting in a binary value for χ_{kj}.

In the equation above, L is the number of the packets that are produced by the the smart meter in a time frame. The last term ν_{ij} is the transmission rate for smart meter i to the j-th DAP in packets/seconds. Its derived equation is shown later in the propagation model. Furthermore, the characteristics of the frames is shown in the media access control (MAC) subsection.

1) MAC: In each cluster, every smart meter transmits the packet without facing interference from other smart meters in the same cluster, using the Time Division Multiple Access (TDMA) technique. In each time slot within the same cluster, only one smart meter is allowed to transmit. However, the same time slot may be used by smart meters from an adjacent cluster, producing interference.

By using the TDMA technique, each smart meter should wait for a period of time until it can send the packets to its time slot in each frame for the transmission. This requires a packet to be saved in a buffer causing an access delay equal to half the frame duration. The buffer collects the packets from the devices in the house and wait for the transmission.

It is assumed that the frame duration T_{f_j} is different from one cluster to another depending on the number of packets produced by the connected smart meters and their transmission rate. In addition, The duration of a time slot, $T_{s,i}$ for smart meter i, depends on its data rate $\nu_{i,j}$, and is the same in all time frames. The number of slots in each frame varies depending on how many smart meters the DAP *j* is connected to, where the maximum number *w* is equal to the total number of smart meters N_s. The number of packets in one frame produced by the i-th smart meter is equal to *L*. The packet size, M for all packets produced by the smart meters is the same.

2) Delay: The end-to-end delay between smart meter and DAP is consist of the queuing delay, the access delay, the transmission and retransmission delay and the propagation delay. In this proposed system, the propagation delay is negligible, resulting in the following delay equation.

$$D_{ij} = \chi_{kj}(D_{TDMA_j} + D_{sy_{ij}}), \tag{2}$$

where D_{TDMA_j} is the access delay for each smart meter in the cluster *j*. It is the average time needed for the head-of-queue packet to wait for its time slot in the frame. It is equal to half the frame duration $D_{TDMA_j} = T_{f_j}/2$. $D_{sy_{ij}}$ is the system delay in node *i* that is connected to the DAP *j*. The system delay is the summation of the queuing delay and the service time). The queuing model that is used is M/G/1 model, with an infinite buffer size. The packet arrival time follows a Poisson process with arrival rate λ and a general service delay distribution equal to 1/μ.

The system delay $D_{sy_{ij}}$ can be written as:

$$D_{sy_{ij}} = D_{s_{ij}} + \frac{\lambda E[D_{sij}^2]}{2(1-\rho)}. \tag{3}$$

Equation (3) shows the system delay, where it includes the delay occurs due to the queuing and the service time. From the equation, ρ is the utilization factor that is equal to $\frac{\lambda}{\mu}$. $D_{s_{ij}}$ is the service delay and $E[D_{sij}^2]$ is its second moment.

Our service time, μ^{-1} depends on the transmission and retransmission delays for each packet. In total, the average service time for each packet D_{sij} can be viewed as:

$$D_{s_{ij}} = \sum_{n=1}^{\inf} P_{s_{ij}}(1-P_{s_{ij}})^{n-1}[n\frac{1}{\nu_{ij}} + (n-1)T_{f_j}]. \tag{4}$$

In this equation, the first term shows the successful transmission probability, where the packet success probability $P_{s_{ij}} = (1 - P_{b_{ij}})^M$, where M being the packet size. P_b is the bit error probability and equal to $P_{b_{ij}} = 0.5(1 - \sqrt{SNIR_{ij}/(SNIR_{ij}+1)}$.

After multiplying it by the second term with the summation, it shows the expected time needed to transmit the packets successfully to the UC. The third term shows the time duration between each two packets in the same frame.

By using the geometric series, D_{sij} in a closed form equation can be written as:

$$D_{sij} = \frac{P_{s_{ij}}}{(1-P_{s_{ij}})^2}\left(\frac{1}{\nu_{ij}} + T_{f_j}P_{s_{ij}}\right) \tag{5}$$

However, the second moment is descried as:

$$E[D_{sij}^2] = \sum_{n=1}^{\inf} P_{s_{ij}}(1-P_{s_{ij}})^{n-1}[n\frac{1}{\nu_{ij}} + (n-1)T_{f_j}]^2 \tag{6}$$

By considering the service rate $\mu_{ij} = \frac{1}{D_{sij}}$ in (5), and substituting that in (3), where $\rho = \frac{\lambda}{\mu}$, the delay could be written as:

$$D_{sy_{ij}} = D_{s_{ij}} + \frac{\lambda E[D_{sij}^2]}{2(1-\lambda D_{s_{ij}})} \tag{7}$$

3) Propagation model: Bit rate of a link can be determined depending on the link's radio propagation model. The radio channel between the smart meter and the DAP suffers a Rayleigh fading. The amplitude of the signal fades as the signal hits some obstacles such as buildings, other houses, etc. As it is mentioned earlier, in one cluster, when the smart meter sends the data to the DAP in a specific time slot, that DAP receives signals from smart meters of neighboring clusters. Those alien signals are considered as interferences to the transmitted data in a time slot. The signal to interference and noise, SINR is determined.

$$SINR_{ij} = \frac{P_{tx,n} d_{ij}^{-\sigma} \left|h^2\right|}{(\sum_{k=1}^{N_k, k\neq K} \frac{\sum_{f=1}^{Ns, f\neq i} d_{fk}^{-\sigma}}{N})(P_{tx,m}\left|h^2\right|)) + \zeta} \tag{8}$$

where K is the *i*'th smart meter's cluster, $P_{tx,n}$ and $P_{tx,m}$ are the transmission power in the nth smart meter and the transmission power for the interference mth smart meters. d_{ij} is the distance between the smart meter *i* and the DAP *j*, however, d_{fk} is the average interference distance coming from smart meters in other clusters. In general, the distance d_{ij} is calculated as $d_{ij} = \sqrt{(conc_x - sm_x)^2 + (conc_y - sm_y)^2}$, where $conc_x, conc_y$ are the coordinates for the DAP position in x-axes and y-axes respectively and sm_x, sm_y are the coordinates for the smart meters position. The channel gain is represented by h and ζ is the noise power caused by the environment temperature. For further simplification, assuming that transmission powers are the same for all the smart meters

$$SINR_{ij} = \frac{d_{ij}^{-\sigma}}{\sum_{k=1}^{N_k, k\neq K} \frac{\sum_{f=1}^{Ns, f\neq i}(d_{fk})^{-\sigma}}{N} + \Upsilon} \tag{9}$$

Using the SNIR equation, the bit rate ν_{ij} for smart meter *i* to the DAP *j* is equal to $\nu_{ij} = \frac{C_{ij}}{M}$, so the final equation for the transmission rate is:

$$\nu_{ij} = \frac{Blog_2(1 + SNIR_{ij})}{M} \tag{10}$$

III. PROBLEM FORMULATION

The main objective of the optimization algorithm proposed in this paper is to find two variables, namely the coordinates of the DAP's ($conc_x, conc_y$) and α_k, that would minimize the total average delay in the system. The total average delay is the delay occurs while transmitting the data from all the smart meters to their DAPs and from the DAPs to the utility center, under some constraints.

For that, the DAP placement problem can be viewed as:

$$\min_{\alpha_k, conc_y, conc_x} \left(\sum_{j=1}^{N_c} \sum_{i=1}^{N_s} D_{ij} + \frac{M}{r_{j\theta}} \right)$$

subject to

$$\sum_{j=1}^{N_c} (\chi_{kj}) = 1 \qquad \forall k \in N_k \qquad (11)$$

$$\sum_{k=1}^{N_k} (\chi_{kj}) \leq Amax \qquad \forall j \in N_c \qquad (12)$$

The optimization is done over the coordinates of the DAPs and α_k. The first term of the objective function shows the maximum delay caused by transmitting the packets from the smart meters that are connected to the DAP *j*. The second term shows the delay needed for the aggregated and compressed packets to convey from the DAP to the utility center in the wired channel.

The constraint (11) ensures that each cluster is connected to only one DAP. Concerning about DAP capacity, the constraint (11) is used to limit the number of the smart meters that are connected to one DAP to $Amax$. The only scenario allowed in the proposed algorithm is that all the smart meters in the same cluster should send data to the same DAP. Overall, the objective function could be described as a non-linear function, with linear constraints. In addition, it should be optimized over integer and continuous variables. Hence, this is indeed a mixed integer non-linear optimization problem. This optimization problem in NP-hard and cannot be solved quickly for a practical problem size. In view of the challenge, we have solved the problem using genetic algorithm.

It is worth mentioning that genetic algorithm is the most used technique in the computation research [16]. It starts by randomly generating number of chromosomes or individuals and then the genetic algorithm cycle begins. Each chromosome goes through some genetic operations, commonly used ones are crossover and mutation. In the crossover, two chromosomes exchange their genes to produce a new chromosome. However, in the mutation, some random gens are changing within the same chromosome. After that, the selection process is applied, where the chromosomes with the optimal fitness function values go to the second generation, waiting for the crossover and mutation operations.However, the chromosomes that have not being selected as fit chromosomes are removed. The cycle keeps going until the fitness function value is in its optimal status [17].

IV. RESULTS

Table1 below lists the typical values for various variables used in the proposed schemes. These values should be assumed while reading the numerical results presented in this section. There are many factors that could affect the total transmission delay occurs in the system, however, three of them are going to be studied which are: the number of DAPs deployed, the total number of smart meters that are distributed in the area, and the packet arrival rate.

TABLE I
PARAMETERS FOR COMMUNICATION MODEL

parameter	representation	values
N	Number of smart meters per cluster	12
N_k	Total number of feeders	4,10,15
N_s	Total number of smart meters	48,120,180
σ	Path loss exponent	3.7
L	packets produced per smart meters	4
M	packets size	128 bytes
B	Channel bandwidth	250 kHz
λ	Number of packets arrived per seconds	0.06,0.5,0.9
ν	Number of packets transmitted per seconds	4
ζ	The noise power	-174 dBm
A	Max. No. of SMs to be connected to DAP	36
$r_{j,\Theta}$	link capacity between DAP and UC	250 Mbps

First, increasing the number of DAPs in the examined area would exponentially decrease the overall delay in the system, as seen in Fig. 1 for $\lambda = 0.5$ and $N_k = 10$. The figure shows the total transmission delay for different number of DAPs. Increasing the number of DAPs allows more transmission operations to be done in parallel, resulting in decreasing total transmission delay until it reaches the minimum. The delay reaches the minimum when the number of DAP is the same as the number of clusters. That happens because in this study, it is assumed that all the smart meters that are placed in the same cluster should send the data to the same DAP When the number of DAPs are less compared to number of clusters, it means higher delay. However, more DAPs lead to a higher cost. In the case above, the total number of DAPs need to reach minimum total delay, i.e 0.296s is 10 DAPs. which is the same as the number of DAPs.

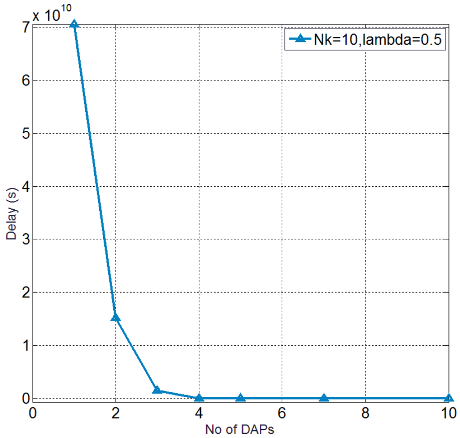

Fig. 1. The total transmission delay for different No. of DAPs

Secondly, different number of clusters would lead to different total transmission delay. Fig. 2 shows three scenarios; when number of clusters is 4, 10 and 15 respectively for the same $\lambda = 0.5$. Increasing the number of clusters tends to have more smart meters in the examined area (since each cluster contains

12 smart meters) and hence, a higher delay. For that, the line is shifted toward the up-right-side of the figure with increasing the number of clusters. For $N_k = 4$, it takes 4 DAPs 0.0395s to transmit the data from the houses to the UC, however, these four DAPs need 10 times that value to transmit the data when number of clusters is 10 (0.3675s). Nevertheless, 4 DAPs are not enough to transmit the data when number of clusters is 15 as one DAP can handle a total of 3 clusters.

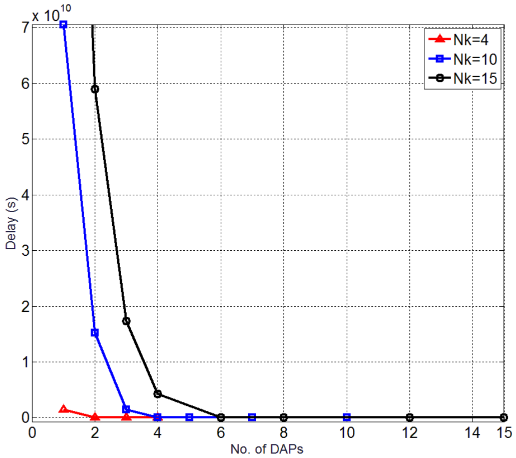

Fig. 2. The total transmission delay for different No. of clusters

Finally, as the arrival rate λ increases, a higher delay is observed. Fig. 3 shows the delay with different arrival rate, when total number of clusters are fixed to be 4. It needs 2 DAPs to achieve a delay of 0.1463s when the arrival rate is 0.9 packets/second. For a lower arrival rate, , which is 0.5 packets/second, these two DAPs need only 0.1421s, which is 1.0296 times less. Eventually, when the arrival rate is further lower to 0.06 packets/second, the delay recorded is 0.1346s.

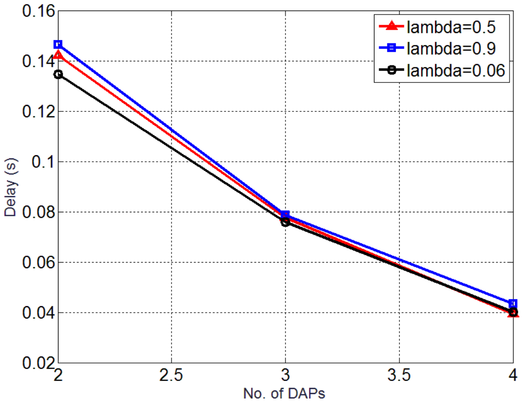

Fig. 3. The total transmission delay for different packet arrival rates

V. CONCLUSION

As a conclusion, aggregating and compressing the packets coming from the same cluster in the DAP before transmitting them to the utility center would decrease the overall delay. That happened because aggregation method would decrease both the size and number of packets needed to be transmitted to the utility center, and hence decrease the total transmission delay and cost. However, the number of DAPs deployed will affect the delay. In the algorithm we proposed, each smart meter in the same cluster should send data to the same DAPs, for that, the adequate number of DAPs placed in the examined area should be the same as number of clusters. In addition, having more clusters in the system or increasing the arrival rate would negatively affect the delay.

REFERENCES

[1] N. Myoung, Y. Kim, and S. Lee, “The design of communication infrastructures for smart das and ami,” in *2010 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2010, pp. 461–462.

[2] P.-Y. Kong and G. K. Karagiannidis, “Charging schemes for plug-in hybrid electric vehicles in smart grid: A survey,” *IEEE Access*, vol. PP, no. 99, pp. 1–1, Oct. 2016.

[3] P.-Y. Kong, “Wireless neighborhood area networks with qos support for demand response in smart grid,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1913–1923, Jul. 2016.

[4] P. D. Diamantoulakis, K. N. Pappi, P.-Y. Kong, and G. K. Karagiannidis, “Game theoretic approach to demand side management in smart grid with user-dependent acceptance prices,” in *IEEE Vehicular Technology Conference (VTC) Fall*, Sep. 2016.

[5] P.-Y. Kong, “Effects of communication network performance on dynamic pricing in smart power grid,” *IEEE Systems Journal*, vol. 8, no. 2, pp. 533–541, Jun. 2014.

[6] P.-Y. Kong and Y. Han, “Cooperative wireless transmissions of dynamic power price and supply information for smart grid,” in *IEEE WCNC*, Apr. 2013.

[7] P.-Y. Kong, C.-W. Liu, and J.-A. Jiang, “Cost efcient placement of communication connections for transmission line monitoring,” *IEEE Tans. Industrial Electronics*, 2017.

[8] D. B. Shmoys, É. Tardos, and K. Aardal, “Approximation algorithms for facility location problems,” in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*. ACM, 1997, pp. 265–274.

[9] Z. Lu and Y. Wen, “Distributed algorithm for tree-structured data aggregation service placement in smart grid,” *Systems Journal, IEEE*, vol. 8, no. 2, pp. 553–561, 2014.

[10] R. Peng, X. Rui-hua, and Q. Jin, “Bi-level simulated annealing algorithm for facility location problem,” in *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*, vol. 3. IEEE, Dec. 2008, pp. 17–22.

[11] F. Aalamifar, G. N. Shirazi, M. Noori, and L. Lampe, “Cost-efficient data aggregation point placement for advanced metering infrastructure,” in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*. IEEE, Nov. 2014, pp. 344–349.

[12] G. B. de Castro Souza, F. H. T. Vieira, C. R. Lima, G. A. de Deus, M. S. d. C. Júnior, S. G. de Araújo, and T. L. Vasques, “Developing smart grids based on gprs and zigbee technologies using a queueing modeling based optimization algorithm,” *ETRI Journal*, vol. 7, no. 4, pp. 1913 – 1923, 2016.

[13] S. Misra, S. D. Hong, G. Xue, and J. Tang, “Constrained relay node placement in wireless sensor networks: Formulation and approximations,” *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 2, pp. 434–447, Apr. 2010.

[14] Z. Zhou, X. Zhao, and P. H. J. Chong, “Optimal relay node placement in wireless sensor network for smart buildings metering and control,” in *Communication Technology (ICCT), 2013 15th IEEE International Conference on*. IEEE, Nov. 2013, pp. 456–461.

[15] O. Hersent, D. Boswarthick, and O. Elloumi, *The internet of things: applications to the smart grid and building automation*. Wiley, 2012.

[16] S. Sivanandam and S. Deepa, *Introduction to genetic algorithms*. Springer Science & Business Media, 2007.

[17] P. U. Kadam and M. Deshmukh, “Real-time intrusion detection with genetic, fuzzy, pattern matching algorithm,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, March 2016, pp. 753–758.