Energy-Efficient Data Aggregation Point Placement in LoRaWAN Networks for Smart Metering Applications

Thiago Allisson R. da Silva[1,2], Geraldo A. Sarmento Neto[1], Luís H. O. Mendes[1], Pedro F. F. Abreu[1], Artur F. S. Veloso[1], Ricardo A. Lira Rabêlo[1], Francisco Airton Silva[1], José Valdemir dos Reis Jr.[1]

[1]Federal University of Piauí (UFPI), Teresina - PI – Brazil

[2]Federal Institute of Maranhão (IFMA), Barra do Corda – MA – Brazil

{thiago.allisson,valdemirreis}@ufpi.edu.br

Abstract. *Electric grids have been restructured with Smart Grids (SGs) and the deployment of Advanced Metering Infrastructure (AMI) systems is a fundamental part of this process. The definition of the number of Data Aggregation Points (DAPs) and their installation positions is essential for the operation of an AMI, and this work proposes a method to minimize the number of DAPs to be deployed in a Long Range Wide-Area Network (LoRaWAN) network. The method is evaluated through simulations with up to 1000 Smart Meters (SMs) and compares the performance of the clustering algorithms K-Means, K-Medoids, and Fuzzy C-Means in the placement of DAPs. Furthermore, network transmission capacity is analyzed applying the Spreading Factor (SF) allocation schemes Highest - Spreading Factor Allocation (H-SFA), Closest - Spreading Factor Allocation (C-SFA), and Capture Effect - Spreading Factor Allocation (CE-SFA). The results demonstrate that the method is promising, the Fuzzy C-Means algorithm reduces the number of DAPs by up to 33.33%, and the CE-SFA scheme increases network energy efficiency by up to 228.41 b/J.*

1. Introduction

The traditional Electric Power System (EPS) is being transformed into a SG, and one of the fundamental steps in this process is the deployment of the AMI system. The primary component of an AMI system is the SMs, devices installed in homes, businesses, and industries to collect data such as power consumption and current measurements. SMs are equipped with communication interfaces that transmit data to DAPs, which subsequently forward the data to the utility company. Consequently, it is essential to determine the quantity and locations of DAPs to meet the Quality of Service (QoS) requirements of the AMI system, as well as to define the appropriate configuration of the communication technology used between SMs and DAPs [Khan et al. 2023].

In line with the evolution of EPS, the paradigm Internet of Things (IoT) enables real-time data collection and assists the utility company in optimizing the energy distribution. The use of connected sensors and devices enables real-time monitoring of energy consumption, facilitates the integration of renewable energy, and improves load management. Furthermore, data collected by SMs is sent to databases and made available through applications, allowing consumers to monitor and manage their energy consumption. As a result, the EPS achieves a higher level of resilience and can quickly detect and resolve issues [Da Silva et al. 2024].

In this context, this paper proposes a method evaluated through simulations to determine the quantity and locations of DAPs that communicate with SMs using LoRaWAN, a technology widely used in IoT applications [Jouhari et al. 2023]. The positions of the DAPs are determined using clustering [Mahdi et al. 2021], a technique successfully applied to LoRaWAN infrastructure planning [Loh et al. 2023], and this work advances the state of the art by identifying the best-performing algorithm. In addition, the method determines the most suitable SF allocation scheme for the operation of SMs.

The main contributions of this paper can be summarized as follows:

- A method that minimizes the number of DAPs required for the operation of the AMI system, leading to cost savings in AMI deployment.
- The application of the method to determine the positions of the DAPs based on the clustering of the x and y coordinates of the SMs, along with a comparison of the clustering algorithms to identify the most suitable algorithm for the test scenarios.
- Introduction of the CE-SFA SF allocation scheme for LoRaWAN, which enhances energy efficiency and reduces latency compared to existing SF allocation schemes.

The remainder of this work is structured as follows. Section 2 presents the theoretical foundation of this work and related studies. Section 3 describes the proposed method. Section 4 outlines the applied methodology. Section 5 displays and discusses the results. Finally, Section 6 concludes the work and presents future research directions.

2. Theoretical Foundation

This section initially introduces the key concepts related to AMI and clustering. Next, it provides an overview of LoRaWAN technology and discusses related work.

2.1. Advanced Metering Infrastructure

Traditional EPS consists of the layers of generation, transmission, distribution, and consumption, as shown in Figure 1. The generation grid is made up of power plants, such as hydroelectric and photovoltaic plants. The transmission grid refers to the part of the electrical sector responsible for transferring the generated energy to the consumption points. The distribution layer serves as the point of interconnection between the transmission and the end consumers. Finally, the consumer layer consists of residential, commercial, and industrial locations that use energy in their daily operations [Ufa et al. 2022].

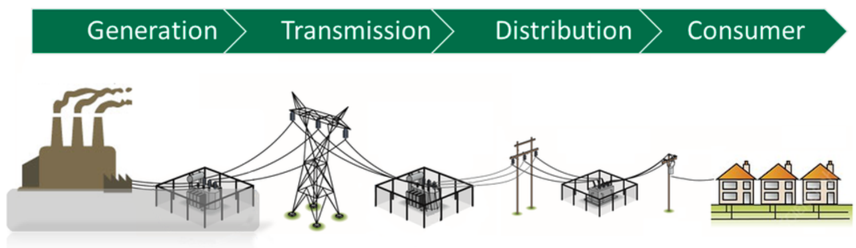


Figure 1. Description of the EPS.

The deployment of AMI allows utilities to take advantage of consumer data to improve the quality of the power supply. The AMI architecture, illustrated in Figure 2, comprises SMs distributed in Home Area Network (HAN), Building Area Network (BAN), and Industrial Area Network (IAN), which communicate with DAPs via LoRaWAN, while the DAPs use a WAN connection to forward the data to the power utility company. The HAN covers residential areas, the BAN includes commercial and institutional buildings, and the IAN handles industrial environments [Da Silva et al. 2024].

Figure 2. Architecture of the AMI System.

In the context of AMI, various applications coexist and can be classified into normal and critical traffic classes. Normal traffic includes applications such as Interval Meter Reading (IMR), which transmit data related to electricity consumption. Similarly, critical traffic consists of Remote-Control Command (RCC) and Power-Control Command (PCC). RCC encompasses messages for disconnecting and reconnecting equipment, while PCC handles messages to execute load control actions. These applications have QoS requirements primarily defined by the maximum delay, measured in milliseconds (ms), and the reliability rate in packet delivery [Khan et al. 2023], as shown in Table 1.

Table 1. QoS Requirements of AMI Applications.

Application	*Delay* (ms)	*Reliability* (%)
IMR	60000	99–99.9
RCC	1000	99
PCC	1000	99

2.2. Clustering

Clustering is an unsupervised learning technique used to identify patterns and discover knowledge by classifying unlabeled data based on their similarities. This technique has been successfully applied to data clustering problems in various domains, such as medical science, manufacturing, power grids, robotics, and more [Mahdi et al. 2021]. Consequently, numerous studies have utilized clustering in the planning of LoRaWAN networks [Matni et al. 2020, Loh et al. 2023] and in communication infrastructures for SGs [Gallardo et al. 2021, Lang et al. 2022, Piechowiak et al. 2023].

Clustering algorithms can be classified in various ways, often based on the method of cluster formation. A common distinction in the literature is between hard and soft clustering. In hard clustering, each object is assigned to a single cluster, whereas in soft clustering, an object may belong to multiple clusters with varying degrees of membership. Typically, a data point is assigned to a particular cluster if it has the highest membership value for that group. For example, in fuzzy clustering, each data point is assigned a degree of membership that quantifies its association with each cluster [Mahdi et al. 2021].

Hard clustering algorithms include K-Means, which assigns points to clusters based on proximity to centroids computed as the mean of cluster members, and K-Medoids, which selects actual data points, named medoids, as cluster centers to minimize intra-cluster dissimilarity [Mahdi et al. 2021]. Lastly, Fuzzy C-Means is a fuzzy clustering algorithm that generates fuzzy partitions. It minimizes an objective function that incorporates the weighted sum of distances between data points and centroids, with weights determined by the degree of fuzzy membership, allowing the data points to partially belong to multiple clusters [Matni et al. 2020].

2.3. LoRaWAN

LoRaWAN is an IoT technology that belongs to the Low-Power Wide-area Network (LPWAN) class and is particularly characterized by its long-range coverage, low power consumption, and low deployment cost. Additionally, it operates in license-free Industrial Scientific Medical (ISM) bands. This technology adopts a star-of-stars topology, where multiple End Devices (EDs) transmit data to Gateways (GWs), which then forward it to the Network Server (NS), and subsequently, the data is transmitted to Application Servers [Stancanelli and Filho 2024].

Communication between EDs and GWs is influenced by the following key parameters: (i) SF, ranging from 7 to 12, where higher values reduce data rate, increase energy consumption, and extend communication range; (ii) Transmission Power (TP), controlling signal strength; and (iii) bandwidth, defining the spectrum for transmission [Marini et al. 2022]. SF presents a critical role in LoRaWAN networks, and an appropriate allocation scheme is essential to balance the communication range of the device and the performance of the network. Various allocation strategies exist, ranging from simple approaches that set SFs to 7 or 12, to more sophisticated schemes that consider minimum reception sensitivity or maximum range distance [Farhad et al. 2020].

Figure 3 illustrates two sensitivity-based allocation schemes. The first scheme, referred to as H-SFA and illustrated in Figure 3(a), allocates the SF to an ED by initially computing all the reception powers P_r obtained by the GWs for a packet sent by the ED. Based on the highest P_r value obtained, the SF allocated to an ED is the one where P_r exceeds the minimum sensitivity value (S) for a given SF, as shown in Figure 3(c), which lists the sensitivity values for SF7 to SF12. For example, if P_r is greater than -124, the SF of the ED is set to 7. The second scheme, C-SFA, shown in Figure 3(b), computes P_r for the nearest GW and allocates the SF of an ED whose P_r value exceeds S for a given SF, following a similar approach to the previous scheme [Farhad et al. 2020].

2.4. Related Work

Table 2 lists the related works that have proposed methods for planning LoRaWAN networks, focusing on the placement of GWs and the design of communication infrastructures to support smart metering systems. The comparison between these methods is based on the following criteria: the applied communication technology, the techniques used for GWs or DAPs placement, and the implementation of SFs allocation schemes.

The alternative solutions presented focus on the placement of GWs or DAP, with their coordinates defined using the clustering algorithms K-Means [Lang et al. 2022, Loh et al. 2023, Da Silva et al. 2024], K-Medoids [Gallardo et al. 2021], and Fuzzy C-Means [Matni et al. 2020], which are also evaluated in this work, advancing the state of

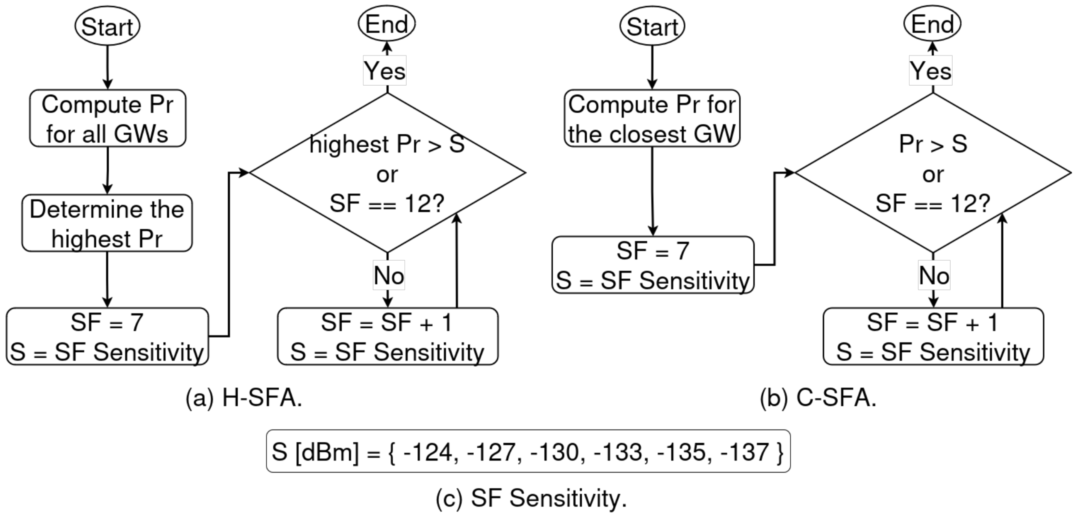


Figure 3. Flowchart of SF Allocation Schemes.

Table 2. Comparison Among Related Works.

Reference	*Technology*	*Placement*	*SF Allocation*
[Matni et al. 2020]	LoRaWAN	Fuzzy C-Means	Sensitivity
[Gallardo et al. 2021]	Multi-hop Wireless	K-Medoids	None
[Lang et al. 2022]	Multi-hop Wireless	K-Means	None
[Piechowiak et al. 2023]	LoRaWAN	K-Means	Sensitivity
[Loh et al. 2023]	LoRaWAN	K-Means	Distance
[Da Silva et al. 2024]	LoRaWAN	K-Means	Sensitivity
This Work	LoRaWAN	K-Means, K-Medoids, Fuzzy C-Means	H-SFA, C-SFA, CE-SFA

the art by determining the best-performing algorithm. The studies in [Matni et al. 2020, Piechowiak et al. 2023, Loh et al. 2023, Da Silva et al. 2024] propose placement methods for LoRaWAN networks, which is also one of the objectives of this study. Meanwhile, the research in [Gallardo et al. 2021, Lang et al. 2022] focuses on placement in multi-hop wireless networks. However, many of these studies overlook the impact of SF allocation schemes on LoRaWAN performance. For this reason, this work also conducts a comparative analysis of these schemes.

The method proposed in [Matni et al. 2020] positions LoRaWAN GWs using the Fuzzy C-Means algorithm and allocates SFs through a sensitivity-based scheme. This approach primarily aims to minimize the distances between GWs and EDs, as well as to reduce network deployment costs. The authors in [Gallardo et al. 2021] present a method for DAP placement in a generic multi-hop wireless network, aiming to improve coverage and reduce the distances between SMs and DAPs. Similarly, the solution proposed in [Lang et al. 2022] introduces a method for DAP placement in a multi-hop wireless routing scenario using the K-Means algorithm. This method applies a heuristic to minimize installation, operation, and transmission costs while ensuring signal coverage.

The authors in [Loh et al. 2023] propose a LoRaWAN GWs placement approach based on the K-Means algorithm to reduce the packet collision rate. SF allocation is performed by computing the maximum communication range of each SF. The study conducted in [Piechowiak et al. 2023] introduces a method that employs the K-Means algorithm to position DAPs and utilizes an optimization scheme to minimize the number of these devices. SF allocation in this method is based on reception sensitivity. Finally, the work presented in [Da Silva et al. 2024] proposes a method that applies the K-Means algorithm to position DAPs, considering the coordinates of SMs and the operational constraints of the LoRaWAN network. This method aims to reduce the number of DAPs while ensuring the QoS requirements of the tested application.

3. Proposed Method

This section describes the SF allocation scheme developed as a complementary part of the proposed solution, followed by a detailed explanation of the method.

3.1. CE-SFA

The CE-SFA scheme leverages the capture effect phenomenon associated with packet reception in LoRaWAN networks, which enhances the ability to correctly demodulate packets even in the presence of multiple collisions. Thus, a packet transmitted by an ED, despite partially or completely colliding with other packets transmitted on the same SF and frequency, can be processed by the receiving GW if the received power differential exceeds the signal-to-interference-plus-noise ratio [Kufakunesu et al. 2024].

The CE-SFA scheme, illustrated in Figure 4, is based on reception sensitivity per SF. Initially, as shown in Figure 4(a), this scheme determines that the SF for each ED should be the one where the P_r value exceeds the minimum sensitivity of a specific SF, similarly to the H-SFA. Once the SFs for all EDs are assigned, the procedure described in Figure 4(b) is executed. The EDs are then sorted in descending order of P_r, and 3 to 5% of the EDs with the highest P_r values have their SFs switched to a lower SF, following the approach described in [Kufakunesu et al. 2024]. This results in a decrease in Time on Air (ToA) and energy consumption at the expense of an increased collision rate, as it increases the data traffic load on lower SFs.

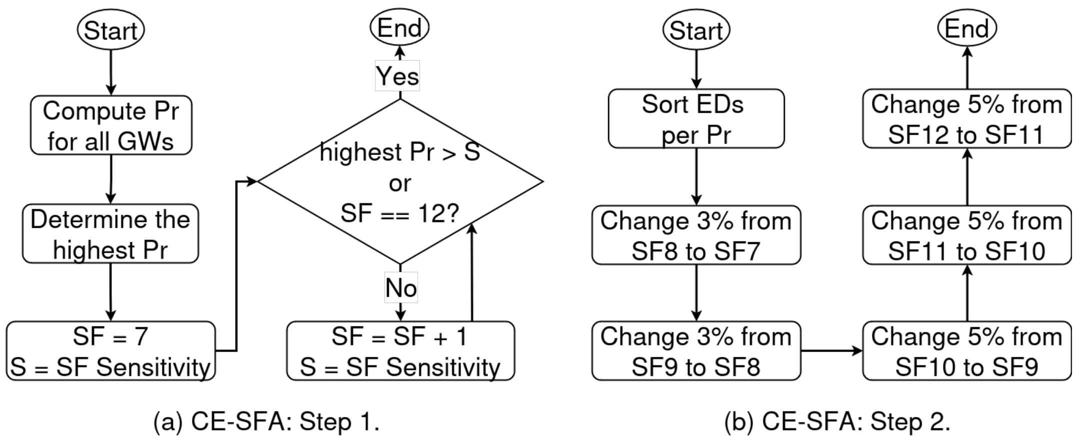


Figure 4. Flowchart of the CE-SFA Scheme.

3.2. Method Description

The method presented in this work proposes a heuristic to define a communication infrastructure that minimizes the number of DAPs while ensuring that the QoS requirements of the tested AMI applications are met. The DAPs correspond to the GWs, and their positions are determined using the K-Means, K-Medoids, and Fuzzy C-Means algorithms. The SF allocation schemes tested are H-SFA, C-SFA, and CE-SFA, which define the SFs of the SMs, implemented as EDs.

Algorithm 1 presents the method in pseudocode format. It is run for all test scenarios with varying numbers of SMs, which are randomly placed in the simulation area, as described in Section 4. The clustering algorithms are applied to determine the positions of k DAPs, and simulations with different SF allocation schemes are conducted. The objective is to identify the clustering algorithm that defines the communication infrastructure with the fewest DAPs and the allocation scheme that delivers the best performance. Furthermore, in the clustering process performed by Fuzzy C-Means, the fuzzification factor, m, is set to 2, and each SM is assigned to the cluster with which it has the highest degree of membership among all others.

Algorithm 1: AMI Planning.

```
1  for all the scenarios do
2      for all the algorithms do
3          for all the SFAs do
4              k = 2
5              while true do
6                  dapCoords = Clustering(smCoords, alg, k)
7                  metrics = SimulateLoRaWAN(smCoords,
                    dapCoords, SFA)
8                  ok = CheckQoS()
9                  if ok == true then
10                     SaveData(k, metrics, dapCoords, alg, SFA)
11                     break
12                 k = k + 1
13 bestAlg, bestSFA = AnalysisOfResults()
14 return bestAlg, bestSFA
```

The core of the proposed method for AMI planning is executed between lines 4 and 12. Initially, k is set to 2, meaning that two clusters must be generated, with their centroids representing the positions of the DAPs. The *Clustering* step receives the x and y coordinates of the SMs, $smCoords$, the clustering algorithm to be executed, alg, and the variable k. At the end of its execution, it returns the x and y coordinates of the DAPs, $dapCoords$, which are calculated based on $smCoords$. Next, $smCoords$, $dapCoords$, and the SF allocation scheme to be tested, SFA, are passed to the *SimulateLoRaWAN* step. This step configures the test scenario in the simulator, executes the simulation by applying the SF allocation schemes, transmitting packets from the IMR and PCC applications, and storing the LoRaWAN network analysis metrics in the variable $metrics$.

Subsequently, in line 8, the QoS requirements of the simulated AMI applications are checked to ensure they are all met. Specifically, 99% of the packets sent by the SMs must be correctly received by the DAPs with a delay less than or equal to the maximum delay established as a constraint. If all the constraints are satisfied, the variable ok stores the value $true$, and k, $dapCoords$, alg, and SFA are stored in log files for further analysis by the AMI planner. If ok is $false$, the variable k is incremented by one, and the entire process from line 4 is re-executed. Finally, the method returns the algorithm that minimizes the number of DAPs required for installation, $bestAlg$, and the allocation scheme with the best performance, $bestSFA$.

4. Methodology

This section outlines the methodology used in this paper, describing the test scenarios, the parameterization of the LoRaWAN technology, and the evaluation metrics.

4.1. Simulation Setup

The evaluation of the proposed method is carried out through of simulations in the Network Simulator 3 (NS-3) tool, version 3.43[1], complemented by the lorawan module, provided by SIGNET[2]. The simulation scenarios consider a square urban area of 7 km by 7 km, with 200 to 1000 SMs distributed throughout this area, as established in [Farhad et al. 2020]. Each SM runs the IMR application, sending 1 packet every 15 minutes [Da Silva et al. 2024], and the PCC application, sending 1 packet per hour [Khan et al. 2023], as described in Table 1. Both applications transmit packets with a payload of 50 B [Khan et al. 2023].

The total simulation time is 24 hours, and each simulation is replicated 33 times with different random seeds to ensure the reliability of the results, following the methodology established in [Matni et al. 2020]. These results are analyzed using a 95% Confidence Interval (CI). The LoRaWAN technology is configured to operate at a carrier frequency of 868 MHz, EU-868, with a bandwidth of 125 kHz [Marini et al. 2022]. The DAPs are configured as GWs, while the SMs are configured as EDs. The SMs transmit at a power of 14 dBm, and SFs are assigned using the schemes described in Sections 2 and 3.

Furthermore, the LoRaWAN network is parameterized using the Okumura-Hata propagation loss model [Da Silva et al. 2024] and the Correlated Shadowing model [Da Silva et al. 2024] to account for power losses in the propagation of packets transmitted by the SMs to the DAPs, consistent with the transmission of LoRaWAN in a realistic urban scenario. Finally, Table 3 summarizes the parameters used in this work and their respective values.

4.2. Evaluation Metrics

The metrics employed in this work are described in this section. The number of DAPs is used to measure the performance of the clustering algorithm, while the Packet Delivery Ratio (PDR) measures the reliability rate of the communication infrastructure [Jouhari et al. 2023] and is described by Equation 1:

[1]https://www.nsnam.org/releases/ns-3-43
[2]https://github.com/signetlabdei/lorawan

Table 3. Applied Parametrization.

Parameters	*Values*
Simulation Area	7 km x 7 km (49 km^2)
Number of SMs	{200, 400, 600, 800, 1000}
IMR Packet Rate	1 pkt/15min (96 pkts/day)
PCC Packet Rate	1 pkt/h (24 pkts/day)
Payload Size	50 B
Simulation Time	24 h (1 day)
Carrier Frequency	868 MHz (EU-868)
Bandwidth	125 kHz
Transmission Power	14 dBm
SF	7 – 12
Propagation Loss	Okumura-Hata
Shadowing	Correlated Shadowing

$$PDR = \frac{N_{rec}}{N_{sent}} \tag{1}$$

where N_{sent} is the number of packets sent by the SMs, and N_{rec} is the number of packets successfully received by the DAPs.

Latency measures the average delay, $\overline{T}$, for the reception of the N_{rec} packets, as described by Equation 2, with T_i corresponding to the reception delay of the i-th packet.

$$\overline{T} = \frac{1}{N_{Rec}} \sum_{i=1}^{N_{Rec}} T_i \tag{2}$$

Energy efficiency, EE, is measured in b/J [Jouhari et al. 2023] and described by Equation 3:

$$EE = \frac{N_{bits} \cdot N_{SMs}}{\sum\limits_{i=1}^{N_{SMs}} E_i} \tag{3}$$

where N_{bits} is the number of bits successfully received, N_{SMs} is the number of SMs, and E_i is the energy consumption of each SM.

The number of expired packets indicates the effectiveness of the communication infrastructure to ensure the time required for packet delivery [Khan et al. 2023]. Meanwhile, the distribution of SFs, $\eta(SF_i)$, indicates the percentage of SMs configured with SF7 to SF12, SF_i, and is calculated by Equation 4:

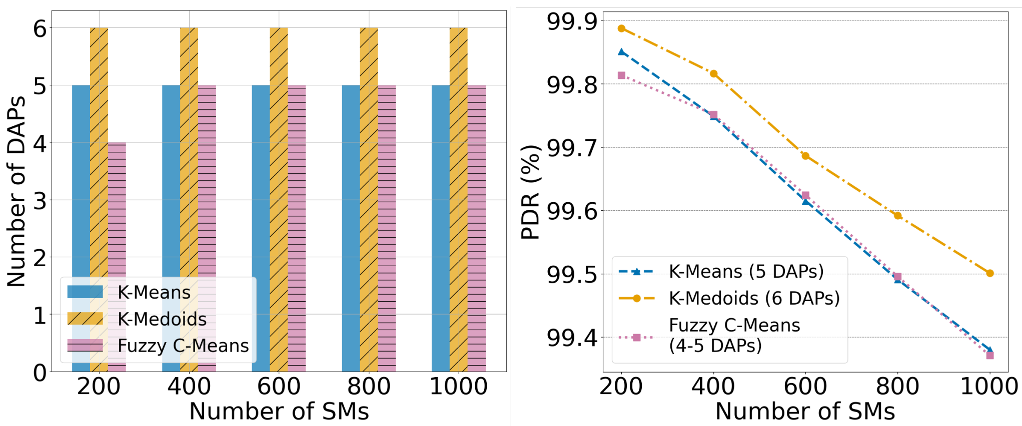
$$\eta(SF_i) = \frac{N_{SF_i}}{N_{SMs}}, \quad \forall i \in \{7, 8, 9, 10, 11, 12\} \tag{4}$$

where N_{SF_i} is the number of SMs operating with SF_i.

5. Results and Discussion

The evaluation of the clustering algorithms is presented first, and the number of DAPs is shown in Figure 5(a). In the scenario with 200 SMs, the Fuzzy C-Means algorithm defines an infrastructure with 4 DAPs to meet the QoS requirements of the applications, whereas K-Means and K-Medoids require 5 and 6 DAPs, respectively. Consequently, the infrastructure determined by Fuzzy C-Means reduces the number of DAPs by 20% and 33.33% compared to K-Means and K-Medoids, respectively. In contrast, for the other scenarios, both Fuzzy C-Means and K-Means use 5 DAPs, while K-Medoids utilizes 6 DAPs.

The PDR values obtained by the clustering algorithms are presented in Figure 5(b). Fuzzy C-Means defines an infrastructure with the fewest DAPs for 200 SMs, and therefore its PDR value is slightly lower than those of the other algorithms. The behavior of the PDRs obtained by all the algorithms is decreasing; as the number of SMs increases, the network traffic load grows, and more packets are lost. However, the number of DAPs ensures a reliability rate of 99%. The average PDR value for 1000 SMs using Fuzzy C-Means is 99.371%, with a CI ranging from 99.361% to 99.381%.



(a) Number of DAPs.

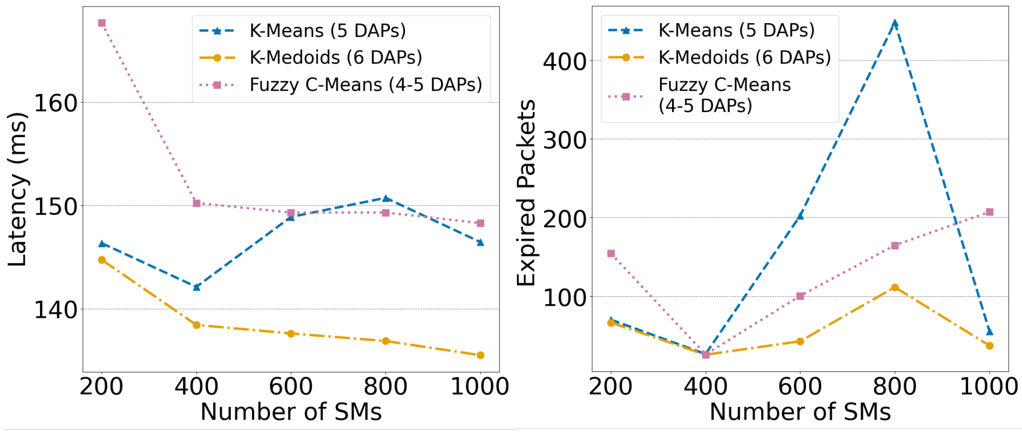
(b) PDR.

Figure 5. Clustering: Number of DAPs and PDR.

Figure 6(a) presents the latencies obtained by the clustering algorithms. The results indicate that as the number of SMs increases, latency values tend to decrease because packets from more distant SMs, which typically experience higher delays, are more frequently lost. However, an exception is observed for K-Means, where latency increases beyond 600 SMs due to the higher usage of SFs greater than 7, as shown in Figure 7(b). The latency for Fuzzy C-Means is higher than that of the other algorithms for 200 SMs due to the deployment of only 4 DAPs, with an average value of 167.648 ms and a CI ranging from 165.42 to 169.876 ms.

The number of expired packets received by DAPs for the clustering algorithms is presented in Figure 6(b). The average results obtained for 200 SMs are: 70 expired packets for K-Means, 67 for K-Medoids, and 155 for Fuzzy C-Means, which uses fewer DAPs. For 800 SMs, the number of expired packets is 448, 112, and 165 for K-Means, K-Medoids, and Fuzzy C-Means, respectively. The increase in packet loss for K-Means is due to a higher proportion of SMs using higher SFs, leading to a higher volume of expired

PCC packets.



(a) Latency. (b) Expired Packets.

Figure 6. Clustering: Latency and Expired Packets.

The results related to SFs distribution are presented in Figure 7. In the scenario with 200 SMs, Fuzzy C-Means uses fewer DAPs, which must receive packets from more distant SMs, leading to a higher number of SFs greater than 7. Conversely, K-Medoids utilizes more DAPs, assigns more SF7 transmissions, and achieves slightly better performance in the evaluated metrics. Meanwhile, K-Means maintains the same number of DAPs across scenarios with 400 to 1000 SMs, but this results in increased latency and a higher number of expired packets.

In the scenario with 800 SMs, K-Means has an average of 12 SMs using SF10, 5 using SF11, and 1 using SF12. The ToA values for SF10, SF11, and SF12 are 657.408 ms, 1232.9 ms, and 2301.95 ms, respectively. Furthermore, in a LoRaWAN network, since a packet can be received by multiple DAPs, the same packet may still be correctly demodulated by several DAPs even after exceeding the expiration time due to the high ToA of the aforementioned SFs. Thus, if all 18 SMs using SF10, SF11, and SF12 are sending 1 pkt/h and their packets are received beyond the maximum delay, an average of 432 packets will expire, which is a value close to the one shown in Figure 6(b).

The evaluation of the SF allocation schemes is presented in Figures 8-9. The number of DAPs, shown in Figure 8(a), indicates that for 200 SMs, Fuzzy C-Means defines infrastructures that satisfy the QoS requirements of the tested applications with 4, 4, and 6 DAPs when applying the H-SFA, CE-SFA, and C-SFA schemes, respectively. For scenarios with 400 to 1000 SMs using also the Fuzzy C-Means algorithm, the schemes define infrastructures with 5, 5, and 6 DAPs. Thus, it can be concluded that the H-SFA and CE-SFA schemes exhibit better performance for this metric.

The results associated with the energy efficiency metric are shown in Figure 8(b) and demonstrate that CE-SFA exhibits higher efficiency across all test scenarios compared to the H-SFA scheme. For 400 SMs, CE-SFA achieved 228.41 b/J more efficiency, with an average value of 8692.94 b/J and a CI ranging from 8640.913 to 8744.967 b/J. Meanwhile, H-SFA achieved an average value of 8464.53 b/J, with a CI of 8418.37 to 8510.70 b/J. In turn, for the scenarios with 600 to 1000 SMs, the efficiency of CE-SFA shows a slight decrease compared to the 400 SMs scenario, similar to H-SFA. This reduction is primarily

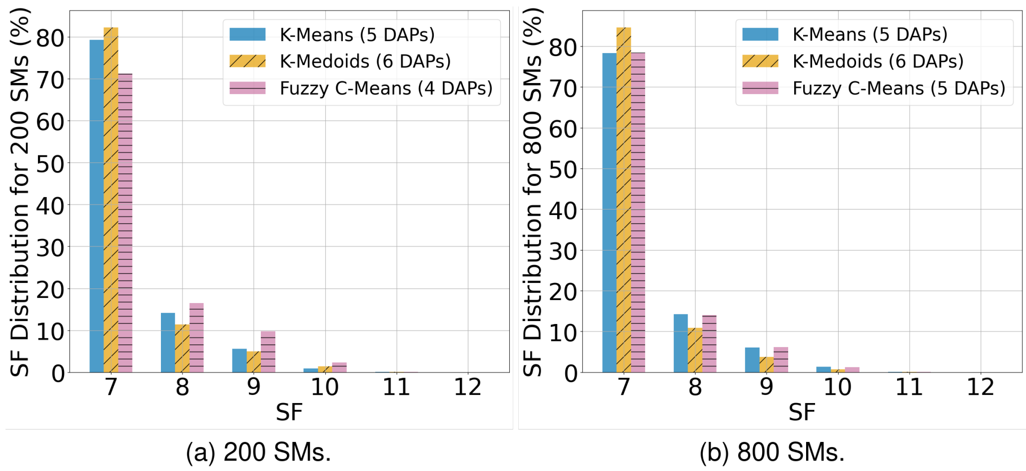


Figure 7. Clustering: SF Distribution.

due to the increase in network traffic load.

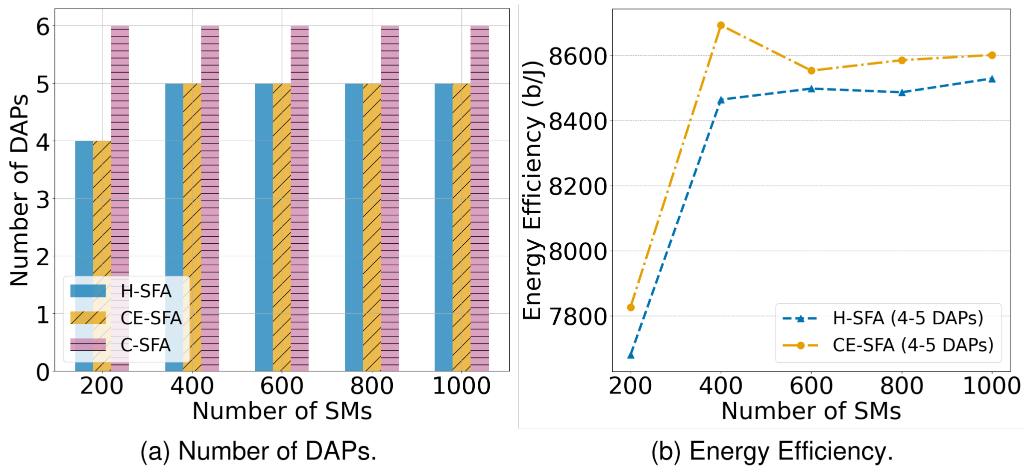
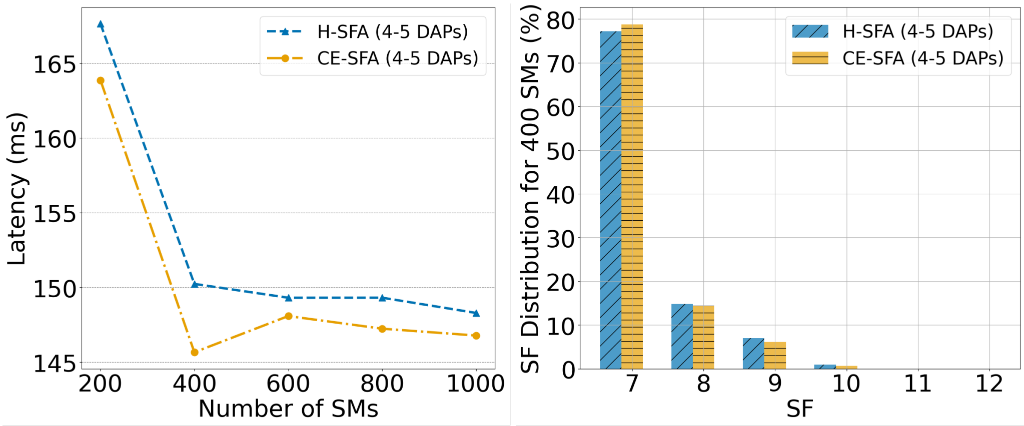


Figure 8. SF Allocation: Number of DAPs and Energy Efficiency.

The latency analysis for H-SFA and CE-SFA is shown in Figure 9(a). Similarly to energy efficiency, CE-SFA achieves better results in all test scenarios. For example, in the scenario with 400 SMs, CE-SFA has a 4.58 ms lower average latency than H-SFA, with a mean value of 145.66 ms. These results confirm that the use of lower SFs can enhance network performance, even with an increased collision rate.

Figure 9(b) presents the SF distribution for 400 SMs. The results show that CE-SFA ensures the efficient operation of applications with lower energy consumption compared to H-SFA, as it applies the capture effect in the redistribution of SFs among SMs. Specifically, CE-SFA utilizes more SF7 and does not assign SFs 11 and 12 to SMs. Consequently, CE-SFA reduces the ToA in transmissions and decreases packet delivery delays, even with the increased data traffic load on lower SFs.

The results demonstrate that the method assists in the definition of a LoRaWAN communication infrastructure that meets the QoS requirements demanded by AMI applications. The analyzed metrics also indicate that Fuzzy C-Means achieved the best perfor-



(a) Latency. (b) SF Distribution.

Figure 9. SF Allocation: Latency and SF Distribution for 400 SMs.

mance, obtaining the lowest number of DAPs for all tested scenarios. Furthermore, CE-SFA proves to be a promising scheme for SF allocation, although dynamic adjustments in the redistribution of higher SFs to lower values should be an aspect to be explored.

6. Conclusion and Future Works

The proposed method presents a heuristic to minimize the number of DAPs to be deployed in an AMI system and determines their positions through clustering. The clustering algorithms applied and compared are K-Means, K-Medoids, and Fuzzy C-Means, and the results confirm the applicability of the clustering technique, as well as the fact that Fuzzy C-Means can reduce the required number of DAPs by up to 33.33% while maintaining stable performance for the other analyzed metrics. In addition to this analysis, the evaluation of the SF allocation schemes demonstrates that CE-SFA, proposed as part of the method, is promising and can increase the network energy efficiency by up to 228.41 b/J.

Future works for the evolution of the proposed method should explore the application of other clustering algorithms, perform a sensitivity analysis of the proposed method, develop an optimal approach based on mixed-integer programming, analyze the long-term impact on energy consumption and battery lifespan of SMs, as well as propose solutions to mitigate the rate of expired packets. In line with these aspects, studies should be conducted to enhance the operation of CE-SFA through the development of a dynamic SF distribution approach and the application of techniques to reduce the number of collisions in the network. All these aspects may lead to a reduction in the number of DAPs to be deployed and a more efficient use of the LoRaWAN technology.

Acknowledgments

Acknowledgments to IFMA, as well as FAPEMA for the financial support via Grant No. 004542/2023, and CNPq for the funding via Grants No. 307967/2022-0 and No. 407274/2021-9.

References

Da Silva, T. A. R., Sarmento Neto, G. A., Abreu, P. F. F., Veloso, A. F. D. S., Mendes, L. H. d. O., and Dos Reis, J. V. (2024). A novel data aggregation point placement method for smart metering service using lorawan technology. In *2024 11th International Conference on Future Internet of Things and Cloud (FiCloud)*, pages 55–62.

Farhad, A., Kim, D.-H., Subedi, S., and Pyun, J.-Y. (2020). Enhanced LoRaWAN Adaptive Data Rate for Mobile Internet of Things Devices. *Sensors*, 20(22):6466.

Gallardo, J. L., Ahmed, M. A., and Jara, N. (2021). Clustering Algorithm-Based Network Planning for Advanced Metering Infrastructure in Smart Grid. *IEEE Access*, 9:48992–49006.

Jouhari, M., Saeed, N., Alouini, M.-S., and Amhoud, E. M. (2023). A survey on scalable lorawan for massive iot: Recent advances, potentials, and challenges. *IEEE Communications Surveys & Tutorials*, 25(3):1841–1876.

Khan, A., Shirazi, S. H., Adeel, M., Assam, M., Ghadi, Y. Y., Mohamed, H. G., and Xie, Y. (2023). A QoS-aware Data Aggregation Strategy for Resource constrained IoT-enabled AMI Network in Smart Grid. *IEEE Access*.

Kufakunesu, R., Hancke, G. P., and Abu-Mahfouz, A. M. (2024). Collision avoidance adaptive data rate algorithm for lorawan. *Future Internet*, 16(10):380.

Lang, A., Wang, Y., Feng, C., Stai, E., and Hug, G. (2022). Data Aggregation Point Placement for Smart Meters in the Smart Grid. *IEEE Transactions on Smart Grid*, 13(1):541–554.

Loh, F., Baur, C., Geißler, S., ElBakoury, H., and Hoßfeld, T. (2023). Collision and Energy Efficiency Assessment of LoRaWANs with Cluster-based Gateway Placement. In *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 391–396. IEEE.

Mahdi, M. A., Hosny, K. M., and Elhenawy, I. (2021). Scalable Clustering Algorithms for Big Data: A Review. *IEEE Access*, 9:80015–80027.

Marini, R., Mikhaylov, K., Pasolini, G., and Buratti, C. (2022). Low-Power Wide-Area Networks: Comparison of LoRaWAN and NB-IoT Performance. *IEEE Internet of Things Journal*, 9(21):21051–21063.

Matni, N., Moraes, J., Oliveira, H., Rosário, D., and Cerqueira, E. (2020). LoRaWAN gateway placement model for dynamic Internet of Things scenarios. *Sensors*, 20(15):4336.

Piechowiak, M., Zwierzykowski, P., and Musznicki, B. (2023). LoRaWAN Metering Infrastructure Planning in Smart Cities. *Applied Sciences*, 13(14):8431.

Stancanelli, E. and Filho, F. S. (2024). The impacts of chase combining-based retransmissions on lorawan performance. In *Anais do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 365–378, Niterói, RJ.

Ufa, R., Malkova, Y. Y., Rudnik, V., Andreev, M., and Borisov, V. (2022). A review on distributed generation impacts on electric power system. *International Journal of Hydrogen Energy*, 47(47):20347–20361.