UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Faculdade de Engenharia
Câmpus de Ilha Solteira

WILSON ENRIQUE CHUMBI QUITO

DIGITALIZATION AND DECARBONIZATION OF ENERGY SYSTEMS: A GEOSPATIAL AND EDGE COMPUTING APPROACH

Ilha Solteira
2026

WILSON ENRIQUE CHUMBI QUITO

DIGITALIZATION AND DECARBONIZATION OF ENERGY SYSTEMS: A GEOSPATIAL AND EDGE COMPUTING APPROACH

Postdoctoral activity report submitted to the Faculty of Engineering of the Ilha Solteira Campus - UNESP, in fulfillment of the requirements for the Postdoctoral Research Program in Electrical Engineering.

Prof. Dr. Jônatas Boás Leite

Advisor

Ilha Solteira

2026

ABSTRACT

The modernization of power systems into smart grids (SGs) is fundamental to achieving a decarbonized economy through the integration of distributed energy resources (DERs), such as solar PV, electric vehicles and emerging energy carriers such as green hydrogen. This transition requires a paradigm shift from traditional reactive planning toward a proactive framework based on the principles of predict, prevent, and optimize. Advanced data-driven techniques and decentralized decision-making provide a promising framework for the optimization of decentralized resources, provided that the operational constraints of the physical grid are respected. Accurate grid state estimation requires a reliable digital representation of the physical network. To support this requirement, synthetic distribution networks (SDNs) are constructed using street map layers and geospatial load data through a GIS-based graph methodology. By employing two-dimensional KD-trees along with a graph reduction process, this approach achieves a 2.6-fold reduction in node count compared to traditional geometric approximations. The effective management of distributed resources further depends on a robust communication and efficient data-processing infrastructure. The increasing volume of internet of things (IoT) data, combined with the strict real-time requirements of grid automation, motivates the developing of a multi-tier edge computing architecture. This decentralized infrastructure distributes computational tasks across edge, fog, and cloud layers using the lightweight MQTT communication protocol. By strategically placing fog brokers through spatial-electrical clustering (K-means and DBSCAN), transmission latency is reduced by 81% relative to centralized cloud models. Each of these fronts contribute to moving into the scalable infrastructure necessary for sustainable and intelligent grid operations.

The activities, analyses, and scientific outcomes summarized in this report were developed during the postdoctoral research stage conducted from March 15, 2025 to March 15, 2026.

Keywords: Distribution network planning; geographic information systems; internet of things data; multi-tier edge computing architecture; smart grids.

RESUMO

A modernização dos sistemas de energia em redes inteligentes (RIs) é fundamental para alcançar uma economia descarbonizada por meio da integração de recursos energéticos distribuídos (REDs), como energia solar fotovoltaica, veículos elétricos e novos vetores energéticos, como o hidrogênio verde. Essa transição exige uma mudança de paradigma, passando do planejamento reativo tradicional para uma proativa baseada nos princípios de prever, prevenir e otimizar. Modelos baseados em dados e a tomada de decisão descentralizada oferecem uma estrutura promissora para a otimização de recursos descentralizados, desde que as restrições operacionais da rede física sejam respeitadas. A estimativa do estado da rede requer uma representação digital confiável da rede física. Para atender a esse requisito, redes de distribuição sintéticas (RDSs) são construídas utilizando camadas de mapas de ruas e carga, e uma metodologia de grafos baseada em SIG (Sistema de Informação Geográfica). Ao empregar árvores KD bidimensionais juntamente com um processo de redução de grafos, essa abordagem monstra uma redução de 2,6 vezes na contagem de nós em comparação com as aproximações geométricas tradicionais. O gerenciamento de recursos distribuídos depende ainda de uma infraestrutura robusta de comunicação e processamento de dados. O crescente volume de dados da internet das coisas (IdC), aliado às limitações de latência da automação da rede elétrica, motiva o desenvolvimento de uma arquitetura de computação de borda em múltiplas camadas. Essa infraestrutura descentralizada distribui tarefas computacionais entre as camadas de borda, neblina e nuvem, utilizando o protocolo de comunicação MQTT. Ao posicionar estrategicamente os brokers de névoa por meio de agrupamento espacial-elétrico (K-means e DBSCAN), a latência de transmissão é reduzida em 81% em comparação com modelos de nuvem centralizados. Cada uma dessas frentes contribui para a transição para a infraestrutura escalável necessária para operações de RIs.

As atividades, análises e resultados científicos resumidos neste relatório foram desenvolvidos no período coberto de 15/03/2025 a 15/03/2026.

Palavras-chave: arquitetura de computação de borda em múltiplas camadas; dados da internet das coisas; planejamento de redes de distribuição; sistemas de informações geográficas; smart grids.

LIST OF FIGURES

Figure 1 — Electrical network modeling using (a) Geometric and (b) GIS-driven approaches 14
Figure 2 — Methodology overview for SDN construction 16
Figure 3 — Topology correction stage to achieve undirected street graphs 17
Figure 4 — KD-Tree and load association stage 18
Figure 5 — Case study overview: (a) street network of the study area; (b) distribution street segment lengths. 20
Figure 6 — Length distribution of service drops: (a) synthetic; (b) real. 20
Figure 7 — (a) Real low-voltage distribution network, (b) the proposed method, (c) Delaunay, and (d) KNN. 21
Figure 8 — Three-layer edge-fog-cloud architecture for IoT networks. 26
Figure 9 — MQTT-based publish–subscribe communication architecture. 28
Figure 10 — IoT node device integration architecture for SG data harmonization. 29
Figure 11 — Clustering process for the k-means algorithm. 31
Figure 12 — Clustering process for DBSCAN algorithm. 32
Figure 13 — 136-bus test system and distribution of clustered buses 35
Figure 14 — DBSCAN and K-means algorithm comparison results 37

LIST OF TABLES

Table 1 — SDN construction methods comparison 22
Table 2 — Graph size evolution according the methodology proposed..... 24
Table 3 — Comparative limitations of cloud, fog, and edge computing for SGs 27
Table 4 — Comparison of latency and total transmission distances between two architectures 34
Table 5 — DBSCAN parameter tuning with different ε and $MinPts$..... 35
Table 6 — Performance metrics for k-means algorithm under k clusters 37

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviations:

AI	Artificial Intelligence
CIM	Common Information Model
CSA	Case Study Area
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DBI	Davies-Bouldin Index
CHI	Calinski-Harabasz Index
DER	Distributed Energy Resource
DMS	Distribution Management System
EV	Electric Vehicle
FLISR	Fault Location, Isolation, and Service Restoration
GIS	Geographic Information System
IED	Intelligent Electronic Device
IoT	Internet of Things
KNN	K-Nearest Neighbors
MST	Minimum Spanning Tree
MQTT	Message Queuing Telemetry Transport
OMS	Outage Management System
OPTICS	Ordering Points To Identify the Clustering Structure
PV	Photovoltaic
RTU	Remote Terminal Unit
SCADA	Supervisory Control and Data Acquisition
SDN	Synthetic Distribution Network
SG	Smart Grid
UP	Urban Parish

Units:

GWh	Gigawatt hour
km	kilometer
kV	Kilovolts
m	Meter
ms	Millisecond
MinPts	Minimum number of points
μs	Microsecond
ε	Neighborhood radius threshold

CONTENTS

1	**INTRODUCTION**	**9**
1.1	OBJECTIVES	10
1.2	CONTRIBUTIONS	11
1.3	STRUCTURE	12
2	**GIS-DRIVEN GRAPHS FOR SDNs CONSTRUCTION**	**13**
2.1	BACKGROUD	13
2.2	GIS-DRIVEN GRAPHS METHODOLOGY	15
2.3	NETWORK CONSTRUCTION ASSESSMENT	19
2.4	DISCUSSIONS	23
3	**INTERNET OF THINGS AND MULTI-TIER COMPUTING**	**25**
3.1	THE LATENCY BOTTLENECK	25
3.2	COMMUNICATION PROTOCOL: MQQT	27
3.3	HARMONIZATION OF THE IOT NODE IN THE DISTRIBUTION SYSTEM	28
3.4	CLUSTERING IN POWER NETWORKS	30
3.5	TRANSMISSION LATENCY IN MULTI-TIER ARCHITECTURES	33
3.6	CLUSTERING PERFORMANCE	34
4	**CONCLUDING REMARKS**	**38**
4.1	CHALLENGES AND FUTURE DIRECTIONS	39
BIBLIOGRAPHY		**40**
APPENDIX A – Scientific Production		**45**

1 INTRODUCTION

The global energy sector is undergoing an unprecedented paradigm shift driven by the need to achieve a decarbonized economy and the massive integration of edge technologies (Sioshansi, 2017). This transition relies on the large-scale integration of distributed energy resources (DERs), including solar photovoltaic generation, electric vehicles (EVs), and emerging clean energy vectors like green hydrogen. At the same time, traditional power grids are evolving into smart grids (SGs), requiring a fundamental structural transformation from reactive planning toward proactive and data-driven operation (Mahmood et al., 2024). This new approach is grounded in the principles of predicting, preventing, and optimizing system behavior to improve grid stability, resilience, and efficiency (Pilo; Silvestro, 2026).

However, the modernization of large-scale distribution networks faces two major challenges: accurate network representation and efficient data-processing infrastructure. First, the evaluation of advanced smart-grid functionalities and future expansion scenarios requires high-fidelity network models (Palmintier et al., 2021). Traditional geometric approaches often ignore urban layouts and physical constraints, resulting in unrealistic and overly complex electrical topologies. In this context, synthetic distribution networks (SDNs) have emerged as strategic tools for enabling scalable, open, and reproducible research environments (Mateo Domingo et al., 2011). Although SDNs do not fully reproduce all operational characteristics of real systems, they provide sufficient spatial, topological, and statistical fidelity to support planning studies, DER integration, resilience analysis, and smart-grid applications (Li et al., 2020). Since electrical infrastructure deployment is strongly constrained by streets, buildings, and geographical barriers, the quality and realism of SDNs depend directly on the fidelity of their spatial representation.

Second, beyond topology modeling, the digital transformation of power systems also demands efficient communication and computational infrastructures capable of supporting near real-time monitoring, automation, and control functions. The increasing deployment of internet of things (IoT) devices throughout electrical networks generates massive volumes of data that must be processed with low latency and high reliability (Al-Ali et al., 2024). In conventional cloud-centric architectures, centralized data processing often introduces excessive transmission delays, bandwidth congestion, and limited responsiveness, making these solutions unsuitable for time-critical applications. To address these limitations, the multi-tier computing

architectures integrating edge, fog, and cloud layers have emerged as promising alternatives for decentralized processing (Veeramanikandan; Sankaranarayanan, 2019). In this framework, edge devices perform local sensing and data acquisition with ultra-low latency, fog nodes execute regional aggregation and fast local decision-making, and cloud platforms provide global coordination, long-term analytics, and large-scale data storage. Furthermore, communication protocols, interoperability standards, and the optimal allocation of fog brokers become critical factors for enabling efficient distributed computing paradigms (Leite; Kezunovic, 2023).

To address these challenges, this postdoctoral work guides two researches that tackle both network modeling and decentralized computing infrastructures. On one side, a GIS-driven graph methodology is proposed to construct SDNs constrained by urban street maps. On the other hand, a multi-tier edge-fog-cloud computing architecture is studied by using lightweight message queuing telemetry transport (MQTT) communication and spatial-electrical clustering algorithms for fog-broker allocation and latency reduction. Together, these contributions support the development of scalable planning and operational tools for the next generation of digital and low-carbon power systems.

1.1 OBJECTIVES

The main question of this work is: “How can geospatial information and decentralized multi-tier computing be integrated to support the digitalization and decarbonization of modern electrical distribution systems?” To address this question, two data-driven frameworks based on geospatial information systems (GIS), the IoT, and edge-fog-cloud computing architectures are investigated.

To achieve this goal, the research focuses on the following objectives:

i. *GIS-Driven SDN Construction:* Develop a GIS-driven graph methodology for constructing realistic utility infrastructure and service-drop connections constrained by urban street layouts;
ii. *Network Scalability:* Apply graph reduction techniques to generate compact, scalable, and computationally efficient distribution topologies suitable for DER integration studies;
iii. *SDN methodology Validation:* Validate the methodology using a real distribution-system and compare its performance against geometric approaches, such as delaunay triangulation and K-nearest neighbors (KNN);

iv. *Multi-Tier IoT Communication Framework:* Design a lightweight and interoperable IoT communication architecture based on MQTT protocols and common information model (CIM) integration for decentralized applications;
v. *Fog-Broker Allocation and Clustering:* Evaluate multidimensional clustering techniques, including K-means and DBSCAN, to determine the optimal allocation of fog brokers considering spatial and electrical characteristics;
vi. *Latency Reduction and SG Automation:* Assess transmission-latency models in edge-fog-cloud architectures to support time-critical SG applications, such as Fault Location, Isolation, and Service Restoration (FLISR).

1.2 CONTRIBUTIONS

The main contributions of this postdoctoral research to the fields of network modeling and decentralized power system computing are as follows:

i. *A GIS-Driven Graph Methodology for SDNs:* Developed a graph-based methodology that transforms raw GIS data into topologically validated and street-constrained SDN;
ii. *Reduction in Graph Computational Complexity:* Developed a network abstraction methodology capable of reducing the number of nodes and edges by approximately 2.6 times compared with conventional geometric approaches;
iii. *High-Accuracy Load Association Method:* Implemented a space-partitioning 2D KD-tree algorithm to associate extensive load points to utility poles at ultra-high speeds. This approach reproduced realistic utility connection patterns with a mean error of only 2.42% in service-drop length estimation compared with real network data;
iv. *Multi-Dimension Clustering for Fog Allocation:* Formulated a multidimensional clustering approach that integrates spatial coordinates, network impedance, and active power demand to determine strategic locations for fog-computing brokers within distribution systems;
v. *Transmission Delays Validation for Real-Time Automation:* Validated, through a geographically referenced 136-bus distribution test system, that the strategic deployment of fog brokers reduces round-trip transmission latency by 81% compared to centralized cloud models;
vi. *Harmonization of Heterogeneous Communication Standards:* Established a dual-path communication framework based on the topic-based MQTT protocol to enable interoperability between legacy utility control platforms and edge-level IoT data streams.

1.3 STRUCTURE

The remaining sections of this document are organized to present the theoretical foundations, implementation methodologies, and experimental validations of the research fronts as follows:

Section 2, entitled *GIS-Driven Graphs for SDNs Construction*, addresses the topological and spatial modeling. This section introduces the importance of high-fidelity spatial representations and presents the proposed GIS-driven graph methodology for constructing SDNs. The framework includes topological construction, GIS data correction, KD-tree-based load association, and MST reduction procedures. The methodology is validated using real-world data from the CENTROSUR utility and compared against conventional geometric approaches.

Section 3, *Internet of Things and Multi-Tier Computing*, addresses the decentralized communication and processing infrastructure required for modern SG applications. This section discusses the limitations of centralized cloud architectures and presents a hierarchical edge-fog-cloud framework integrated with MQTT-based communication. In addition, interoperability between utility standards, including IEC 61850 and IEC 61968/61970, is presented through a unified data model. The section also evaluates the strategic allocation of fog brokers using spatial-electrical clustering techniques, comparing the performance of K-means and DBSCAN algorithms under latency-sensitive scenarios.

Finally, Section 4 summarizes the main findings and contributions of this postdoctoral research, highlighting the importance between geospatial network modeling and decentralized computing paradigms. The report concludes with Appendix A, which presents the scientific publications and technical outputs derived from this research program.

2 GIS-DRIVEN GRAPHS FOR SDNs CONSTRUCTION

2.1 BACKGROUD

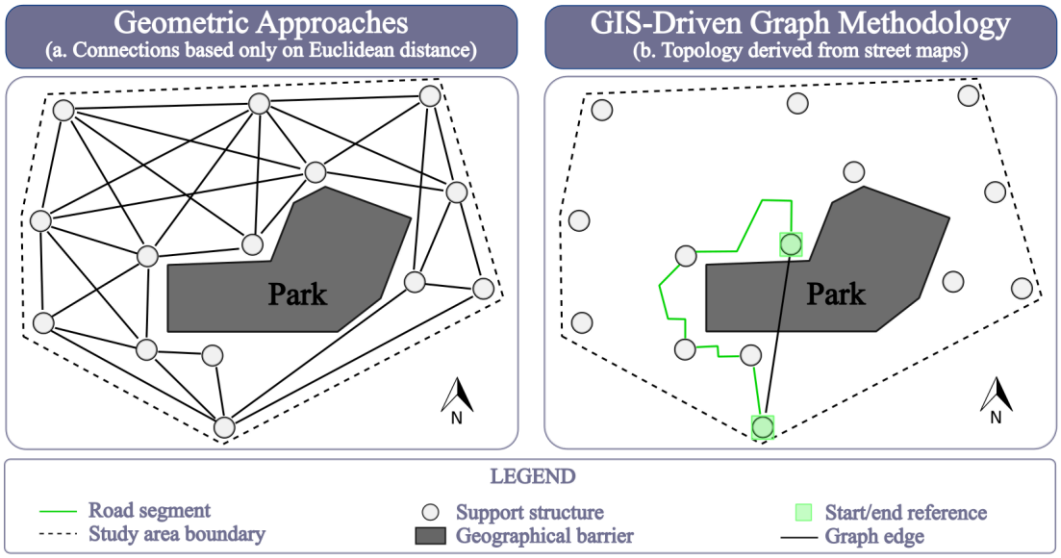
The rapid integration of DERs, the large-scale transport electrification, and the proliferation of active microgrids are reshaping the global energy landscape (Löbbe; Sioshansi; Robinson, 2022). These highly dynamic, bi-directional edge technologies introduce severe operational challenges, including localized voltage violations, bidirectional power flows, and altered protection requirements. Addressing these challenges requires advanced data-driven techniques capable of optimizing decentralized resources while supporting decision-making processes and ensuring compliance with the operational constraints of the physical grid (Kays et al., 2017).

The planning and operation of increasingly complex distribution systems require realistic modeling environments capable of supporting proactive analysis and decision-making (Ali et al., 2023). In this context, SDNs have emerged as essential tools for modeling network topologies and evaluating grid performance under the integration of emerging technologies (Tayab et al., 2024). While they do offer a pathway toward open and reproducible datasets, their primary value lies in providing a high-fidelity environment to simulate real-world operational behaviors and the validation of advanced planning and operational algorithms (Caetano et al., 2024). They are constructed using real geographical data and localized utility statistics to emulate the exact structural, spatial, and operational behaviors of real grids (Mateo et al., 2020). While SDNs do not fully replicate all dynamic and operational characteristics of real networks, they provide sufficient topological and spatial fidelity to support the analysis of future scenarios, grid functionalities, and infrastructure investment strategies.

Historically, SDN has been constructed using unconstrained geometric approaches, such as Delaunay Triangulation and KNN, which establish connectivity solely based on Euclidean proximity (Li et al., 2020). Although these methods quickly produce connected networks, they neglect physical and geographical constraints, often generating unrealistic connections that cross buildings, parks, rivers, or other inaccessible areas. To overcome these limitations, recent studies have combined GIS with graph theory to incorporate real-world spatial constraints into network construction (Oneto et al., 2025). In this paradigm, street networks define feasible routing corridors, while graph structures model the connectivity between infrastructure elements. Customers are then mapped onto these paths through spatial indexing algorithms.

The fundamental divergence between geometric and GIS-driven approaches is illustrated in the Figure 1. In the geometric model (Figure 1a), load points are interconnected according to spatial proximity, resulting in a dense topology with numerous connections that disregard urban infrastructure constraints. In contrast, the GIS-driven approach (Figure 1b) derives network topology from validated street maps, where intersections are represented as nodes and street segments as edges. The resulting graph follows the urban layout, producing physically plausible routing paths and a more realistic foundation for synthetic distribution network modeling.

Figure 1 — Electrical network modeling using (a) Geometric and (b) GIS-driven approaches



Source: Elaborated by the author.

SDNs are commonly represented using graph-theoretical models because graph theory provides a rigorous mathematical framework for describing interconnected systems. In these models, physical components such as intersections, utility poles, transformers, and loads are represented as nodes, while conductors and service connections are represented as edges (Zambrano-Asanza et al., 2021). Consequently, graph theory is particularly suitable for modeling and analyzing network topology, connectivity, and structural properties. Therefore, it is important to introduce some fundamental graph-theoretical concepts.

A graph is a mathematical structure composed of a set of vertices (or nodes) and a set of edges that connect these vertices. Each edge represents a relationship or connection between two vertices in the graph. Graphs can be directed, where the edges have a specific direction, or undirected, where the edges have no orientation. A graph is denoted by $G = (N, L)$, where $N = \{n_1, n_2, \dots\}$ represents the set of nodes and $L = \{l_1, l_2, \dots\}$ represents the set of edges. In an

undirected graph, each edge $l_k \in L$ is associated with an unordered pair of nodes (n_i, n_j). These nodes linked by the edge l_k are called the endpoints of l_k (Deo, 2016).

2.2 GIS-DRIVEN GRAPHS METHODOLOGY

The methodology to construct a simplified yet representative SDN employs a four-stage process, as illustrated in Figure 2. First, in the *topology construction stage*, geospatial data from street maps are transformed into a primal graph. The intersections are modeled as nodes and road segments as edges, establishing the physical backbone of the network. This is followed by the *correction stage*, where geometric inconsistencies, such as disconnected segments, undershoots, and overshoots, are repaired to ensure graph integrity and connectivity. In the *load association stage*, edges are first subdivided into segments to represent realistic span limits for pole placement modeling; then, customer load points are linked to these newly created nearest support structures using a 2D KD-tree algorithm. This enables the integration of demand data into the topology. Finally, the *graph reduction stage* eliminates non-essential nodes and redundant elements by concentrating demand at the support nodes while preserving topological and statistical fidelity. Each of the steps is detailed below.

In the *topology construction stage*, the topological construction begins with a multipart explosion to separate geometrically disconnected segments that share identical attributes and identifiers. This procedure ensures that each edge corresponds to a single continuous segment, making any implicit spatial disconnections explicit within the network topology. This avoids false connectivity assumptions when converting raw GIS spatial features into a graph structure. To guarantee the topological integrity of the resulting base graph G, the following four conditions must be satisfied, (1) to (4).

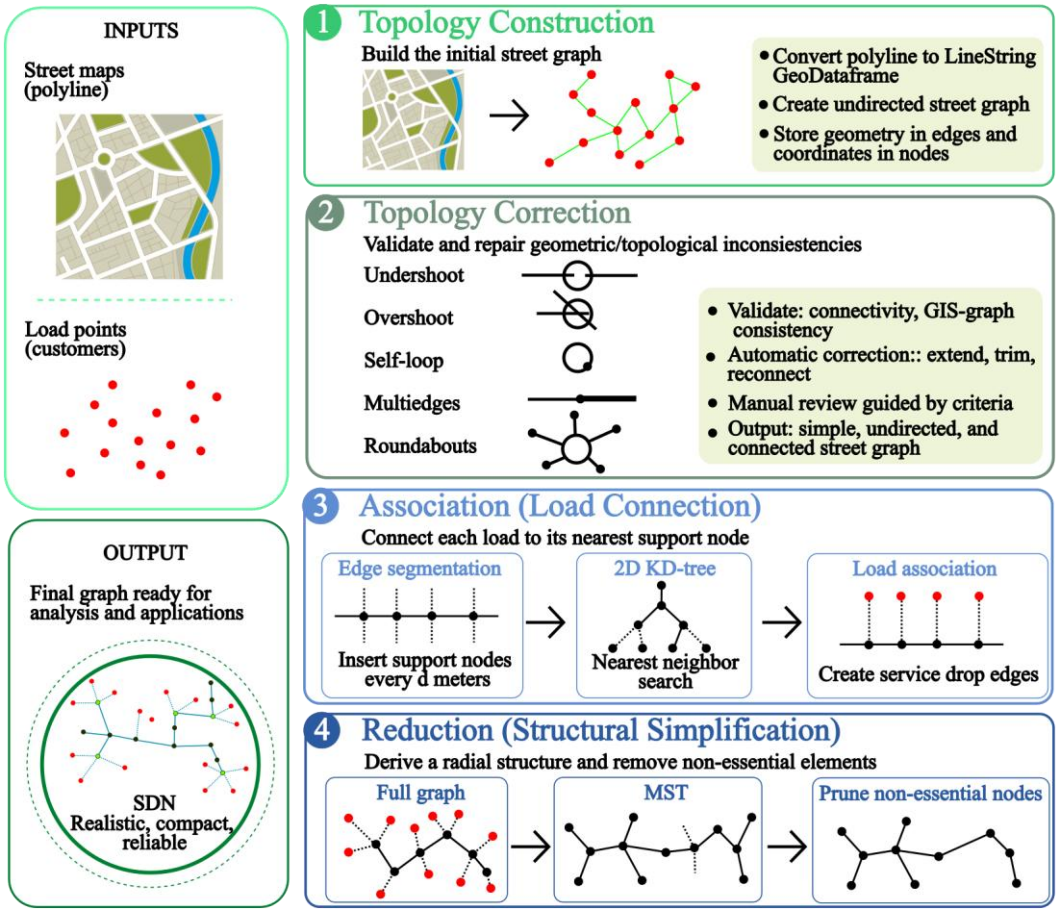
$$|Comp(G)| = 1 \tag{1}$$

$$|\{(n,m) \in L(G)\}| = 0 \tag{2}$$

$$|\{|(n,m) \in N \times N||\{(n,m),(n,m)\} \cap L(G)| > 1\}|=0 \tag{3}$$

$$Num(L_{GDF}) = |L(G)| \tag{4}$$

Figure 2 — Methodology overview for SDN construction



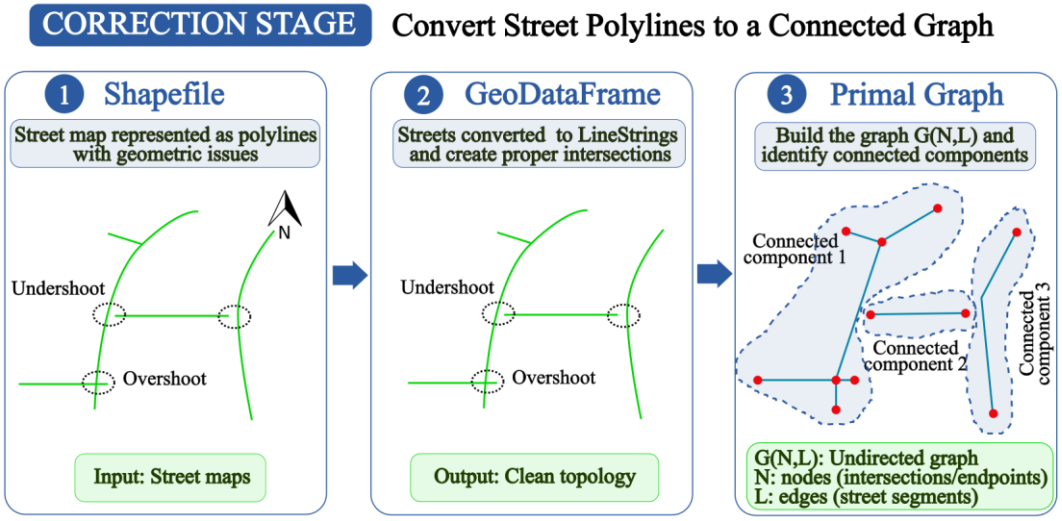
The first condition, connectedness, given in equation (1), requires that the graph consists of exactly one connected component. That is, every node must be reachable from any other node, ensuring there are no isolated subgraphs or topological islands. This constraint is critical because a disconnected graph would imply electrically isolated regions, which poses a severe problem for the unified modeling of a distribution feeder or service area.

A self-loop is defined as an edge that connects a node to itself. The second condition ensures the absence of self-loops, which would otherwise introduce mathematical inconsistencies in graph algorithms, such as shortest-path routing and power flow analysis. The third condition, the absence of multiedges, guarantees that no duplicate edges exist between the same pair of nodes. This constraint avoids duplicated GIS features.

The fourth condition, GIS–Graph consistency, ensures that the set of edges L_{GDF} in the GeoDataFrame (the GIS layer) equals the set of edges $L(G)$ in the graph representation. This condition enforces a strict, one-to-one correspondence between the spatial domain and the topological domain, thereby avoiding data loss (such as edges lost during conversion) and topological distortion (such as artificially created edges).

When the street network is represented as a raw polyline dataset, topological inconsistencies frequently arise from the digitization process. In the *correction stage*, common digitizing errors such as undershoots and overshoots, highlighted in the Figure 3, are resolved using distance-based thresholds. These thresholds prevent unintended connections, while ensuring that all valid road crossings correspond to explicit intersection points. Without these corrections, geometric errors propagate directly into the graph domain, leading to a fragmented graph $G(N, L)$ with multiple disconnected components. Beyond closing spatial gaps, this stage also systematically resolves advanced topological anomalies, including self-loops, multiedges, and complex intersections like roundabouts.

Figure 3 — Topology correction stage to achieve undirected street graphs



For edge segmentation, edges are subdivided into realistic spans (e.g., $d = 40m$) to create candidate support nodes that mimic utility pole placement. The number of required divisions n and the interval spacing Δs are calculated by (5) and (6).

$$n = {}^{L}/_{d} \tag{5}$$

$$\Delta s = {}^{L}/_{n} \tag{6}$$

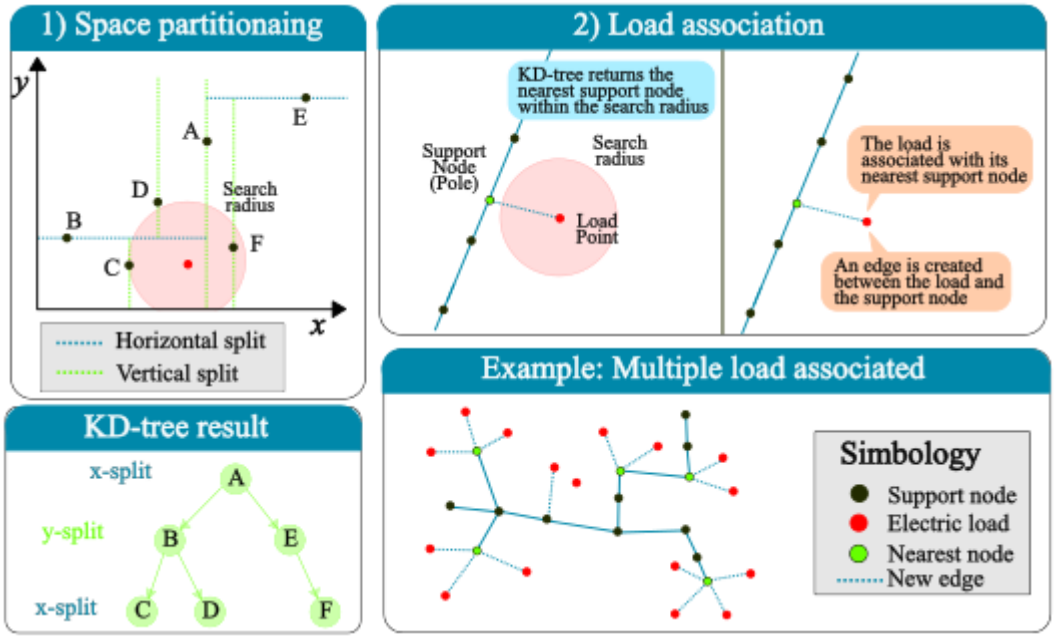
where L denotes the total length of the original edge, and n represents the smallest integer greater than or equal to the ${}^{L}/_{d}$ ratio. This process ensures that the synthetic network adheres to engineering span limits rather than arbitrary mathematical spacing.

In the *load association stage*, the links between load points and the electrical infrastructure are optimized using a two-dimensional (2D) KD-tree algorithm. By partitioning spatial coordinates into a balanced binary tree structure, the framework executes high-speed proximity queries to link each customer load to its nearest support node. This spatial-indexing

approach ensures that the resulting service drop length distributions correlate closely with the minimum-distance patterns observed in real-world utility networks.

A KD-tree is a space-partitioning data structure used for organizing points in a multi-dimensional space. The 2D KD-tree is built using the spatial coordinates (x, y) of all support nodes generated during the preceding edge-segmentation phase. The data structure recursively divides the 2D geometric plane into orthogonal regions, enabling rapid nearest-neighbor search queries. For every customer load node, the search algorithm performs a query configured with a neighborhood parameter of $k = 1$, retrieving the single closest infrastructure support structure based on Euclidean distance. Once the nearest support node is identified, a new edge representing the physical service drop line is created to connect the load to the grid backbone. This is illustrated in Figure 4.

Figure 4 — KD-Tree and load association stage



The final stage of the SDN construction is *graph reduction*. While the preceding stages focus on establishing a topologically and spatially faithful model, the resulting graph often contains a high density of nodes that are non-essential for power flow analysis or optimization algorithms. To streamline the network, the reduction stage first transfers the electrical demand from each customer load node to its assigned infrastructure support node. Once this demand is concentrated at the support structures, the original load nodes and their corresponding service drop edges are removed from the network backbone. Subsequently, a MST is generated using Kruskal's algorithm to ensure connectivity and establish a radial low-voltage layout, with edge weights defined by the geometric length of the line segments. Finally, any remaining non-essential nodes (characterized by a node degree of one, zero electrical demand, and non-critical

junction status) are recursively pruned alongside their incident edges to produce an updated, simple layout.

The methodology ensures that the synthetic grid is not only topologically sound but also statistically congruent with real-world distribution systems. The four-stage process establishes a robust baseline infrastructure generation framework upon which advanced power system planning, operational optimization, and large-scale DER integration models may be developed.

2.3 NETWORK CONSTRUCTION ASSESSMENT

The proposed methodology for constructing SDNs was validated through a case study conducted on the electrical system of "Empresa Eléctrica Regional Centro Sur C.A." (CENTROSUR) in Cuenca, Ecuador. The study utilized 43,689 load points and 582.88 km of street data. The case study area (CSA) comprises 46% of the urban parishes (UPs), as illustrated in Figure 5(a). The final street map dataset includes 9,449 polylines. The segment length distribution in Figure 5(b) shows a heavy concentration of segments shorter than 3 m, which are attributed to digitization errors; this 3-meter threshold was explicitly adopted for correcting undershoot and overshoot anomalies. The histogram of street segment lengths uses a logarithmic scale, where the X-axis represents the length of street segments in meters and the Y-axis represents how many segments exist for each length range. The vertical red dashed line marks the 3-meter boundary separating these digitization errors from the valid street geometries represented in the blue region.

The proposed methodology was executed on a computing system equipped with an Intel Core i5-12650H processor and 16 GB of RAM. To evaluate the load association accuracy, Figure 6(a) shows the length distribution of service drops extracted from the real distribution network, while Figure 6(b) presents the corresponding distribution generated synthetically. Both distributions exhibit a similar and correlated pattern. The real network presents a mean service drop length of 12.7 m compared to the synthetic network's mean of 12.4 m, yielding a difference of 2.42% between their mean values.

Figure 5 — Case study overview: (a) street network of the study area; (b) distribution street segment lengths.

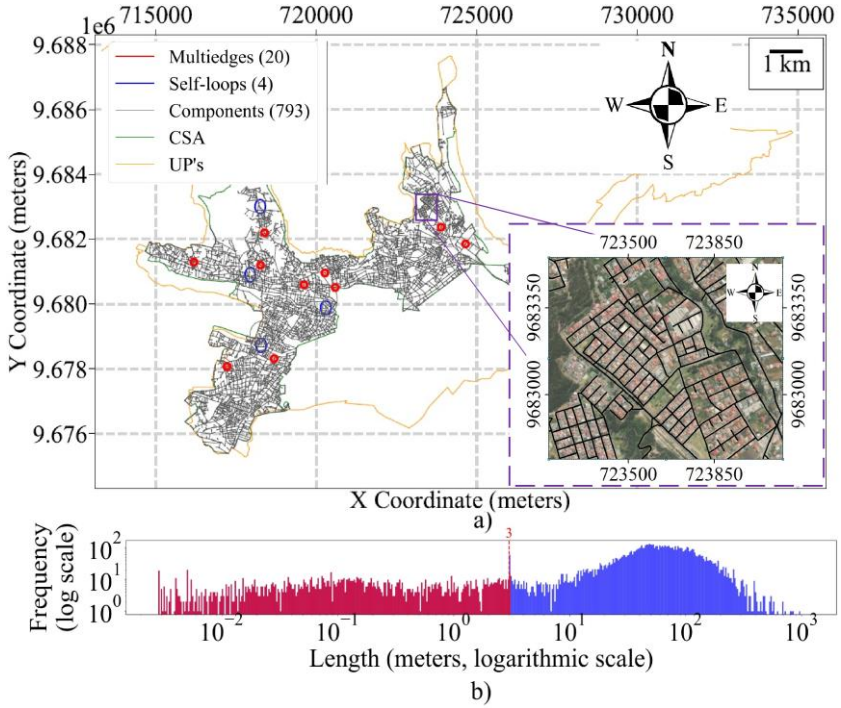
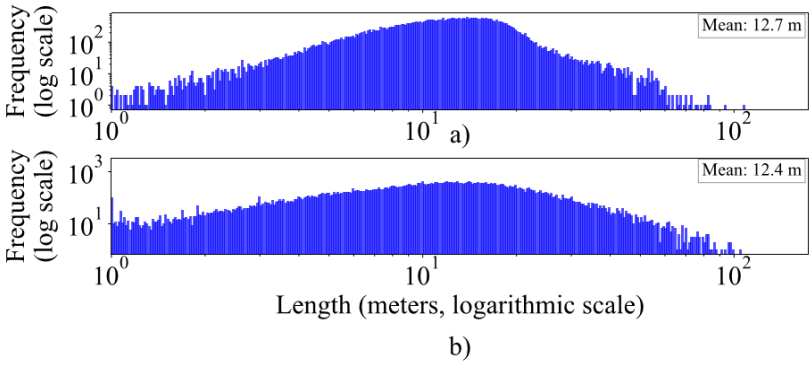


Figure 6 — Length distribution of service drops: (a) synthetic; (b) real.



The study compared the proposed GIS-driven method against two standard geometric approaches: Delaunay Triangulation and K-Nearest Neighbors (KNN). These geometric baseline approaches have been used in previous literature for preliminary distribution and transmission network design (Li et al., 2020), synthetic street network generation (Mateo Domingo et al., 2011), and graph partition-based clustering (Zambrano-Asanza et al., 2021). Both geometric methods construct network topologies solely based on the spatial locations of the load points. For the KNN, the neighbor parameter was set to k = 11 to guarantee that the resulting graph forms a fully connected component. A visual comparison of the real network alongside the SDNs generated by each methodology after applying the MST is presented in Figure 7, displaying the real layout (a) against the proposed (b), Delaunay (c), and KNN (d) configurations.

Figure 7 — (a) Real low-voltage distribution network, (b) the proposed method, (c) Delaunay, and (d) KNN.



The real low-voltage distribution network, shown in Figure 7(a), exhibits a structural layout strongly constrained by the street topology, where multiple consumers are connected directly to common support structures distributed along the road network. The synthetic network produced by the proposed methodology, illustrated in Figure 7(b), preserves these spatial and topological characteristics. By integrating the 2D KD-tree with the street graph, the resulting layout naturally follows the street geometry and generates service-drop connections that are highly consistent with real-world distribution practices and urban infrastructure constraints.

By contrast, the Delaunay-based network shown in Figure 7(c) creates direct geometric connections between load points without considering street constraints. Consequently, the layout does not preserve realistic service-drop organization. A similar behavior is observed for the KNN-based network in Figure 7(d). As a consequence of bypassing the street map corridors, both geometric approaches produce artificial interconnections crossing city blocks and urban regions without alignment to the urban infrastructure.

To quantify these structural differences, Table 1 and Table *2* details the topological performance of the GIS-graph algorithms alongside a comparative analysis against Delaunay triangulation and KNN across key performance and modeling metrics. These metrics track the evolution of total node and edge counts as the networks are built, segmented, and ultimately reduced into finalized low-voltage layouts.

Table 1 — SDN construction methods comparison

Metric	**Proposed Method**	**Delaunay Triangulation**	**K-Nearest Neighbors (K=11)**
Model Compactness (Final Size)	2.6x smaller in total nodes and edges	Baseline size; significantly larger network	Baseline size; significantly larger network
Edge-to-Node Composition (Final MST Structure)	16,788 nodes and 16,787 edges	43,689 nodes and 43,688 edges	43,689 nodes and 43,688 edges
Service Drop Accuracy	2.42% relative difference in mean service length	N/A (Direct geometric links only)	N/A (Direct geometric links only)
Infrastructure Fidelity	Adheres to street paths and urban layout	Ignores road geometry; produces infeasible overlaps	Ignores road geometry; produces infeasible overlaps

Quantitative metrics extracted from the case study confirm the clear advantages of the GIS-driven graph methodology over unconstrained geometric configurations. As shown in Table 1, the proposed method achieves superior model compactness upon final reduction. The resulting graph structure is approximately 2.6 times smaller in both node and edge counts than the geometric baselines. This reduction occurs because geometric counterparts fail to consolidate demands at shared physical structures, leading to larger graph sizes. In terms of edge-to-node relationships, the GIS-based approach significantly reduces structural complexity, achieving between 2.1× and 4.8× fewer edges, whereas Delaunay triangulation suffers from high edge redundancy and KNN requires a relatively large neighborhood parameter (e.g., $k = 11$) to maintain connectivity. Furthermore, the proposed framework introduces the capacity to synthesize realistic service drop lines. It matches the statistical distribution of the physical utility network with a minor 2.42% relative difference between their mean values (12.4 m synthetic versus 12.7 m real). The GIS-graph approach ensures infrastructure fidelity by constraining network topology to street paths, in contrast to geometric models that disregard road geometry and may produce physically infeasible connections.

2.4 DISCUSSIONS

The proposed GIS-driven framework progressively transforms the initial street map into an optimized SDN through successive topological refinement and reduction stages. Following the initial validation and topological correction phase, the framework successfully eliminates digitizing errors to yield a simple, fully connected street graph containing 6,592 nodes and 8,312 edges. During the subsequent edge segmentation process, the grid structure expands to accommodate infrastructure constraints; enforcing a regular spacing threshold between support structures expands the node count to 17,876 support nodes and increases the graph to 19,596 routing edges. Finally, during the load-to-support node association stage, the framework integrates the 43,689 raw customer coordinates into the layout. By linking each load vertex to its nearest spatial support structure via the 2D KD-tree, the network reaches its peak topological density, containing 61,565 nodes and 63,285 edges. This deliberate expansion captures every physical asset and connection corridor before the graph reduction phase eliminates non-essential elements to compress the network's final layout.

Following this peak topological expansion during load integration, the subsequent MST consolidation process at the support nodes drastically compact the network size. Although the load association stage temporarily increases the overall graph size due to the integration of customer nodes, the subsequent reduction phase systemically transfers the system's electrical demand directly onto the support nodes. After recursively removing non-essential, zero-demand leaf nodes and their incident edges, the final reduced network is compressed down to just 16,788 nodes and 16,787 edges. The overview process is presented in Table 2. This structural optimization produces a highly compact topological footprint, successfully minimizing network complexity while fully preserving realistic spatial routing and service-drop distribution characteristics along the street paths.

In comparison, the unconstrained geometric baselines generate massive structural redundancy prior to the tree extraction phase because they build layouts purely from the spatial locations of load points without considering street map corridors. For the same underlying 43,689 customer load points, the initial Delaunay triangulation graph swells to 131,027 edges, while the KNN graph swells to 306,603 edges to maintain connectivity. Even after applying identical MST extraction rules, both geometric approaches remain inflated graph dimensions, maintaining networks with 43,689 nodes and 43,688 edges that scale linearly and rigidly with the number of customer loads. This contrast highlights the proposed method's ability to collapse consumer nodes onto shared physical infrastructure assets, achieving a final model layout that

is approximately 2.6 times smaller in both nodes and edges while ensuring physical infrastructure fidelity.

Table 2 — Graph size evolution according the methodology proposed

Stage	**Nodes**	**Edges**	**Description**
Topological correction	6,592	8,312	Repairs undershoots, overshoots, and loops to ensure a simple street graph
Graph segmentation	17,876	19,596	Inserts support nodes along segments to emulate realistic pole spacing constraints.
Load association	61,565	63,285	Connects customer loads to nearest support nodes via 2D KD-tree for service-drop modeling
MST generation	17,876	17,875	Generates a radial low-voltage layout using Kruskal's algorithm.
Reduced MST	16,788	16,787	Removes non-essential nodes with zero demand and degree one.

3 INTERNET OF THINGS AND MULTI-TIER COMPUTING

The "Three Vs" of big data (Volume, Velocity, and Variety) render centralized cloud computing obsolete for time-critical SG applications. This section shows how to achieve self-healing grid capabilities through a multi-tier hierarchy (Edge, Fog, Cloud) architecture. This decentralized architecture decouples latency-critical restoration protocols from long-term market analytics. By leveraging the MQTT protocol, we strategically deploy communication brokers across these distributed operational layers. Data routing is optimized using a data-driven methodology based on geographic clustering to determine the optimal placement of intermediate fog brokers. Aligning these decentralized computational resources directly with the topology of the electrical network minimizes overall transmission delays. This alignment ensures that grid automation tools, such as Fault Location, Isolation, and Service Restoration (FLISR), operate in near real-time.

3.1 THE LATENCY BOTTLENECK

While cloud computing architectures offers high-level data management, its centralized nature introduces significant bottlenecks for time-sensitive grid applications (Singh; Yassine, 2019). The physical distances between end devices and distant cloud servers cause delays that can compromise grid stability and operational safety. These delays can critically compromise active grid stability, automation effectiveness, and operational safety during transient events (Hong; Varghese, 2020).

Transmitting all raw IoT data to a central server leads to high network traffic (bandwidth congestion) and increased communication costs. Centralized models also establish vulnerable single points of failure and depend heavily on stable, high-speed wide-area internet connections, thereby introducing substantial reliability and connectivity issues into core grid monitoring and control loops (Al-Ali et al., 2024). To address the limitations of centralized data processing, a hierarchical multi-tier architecture is adopted, as illustrated in Figure 8, where computational resources are distributed across edge, fog, and cloud layers to support real-time SG operation.

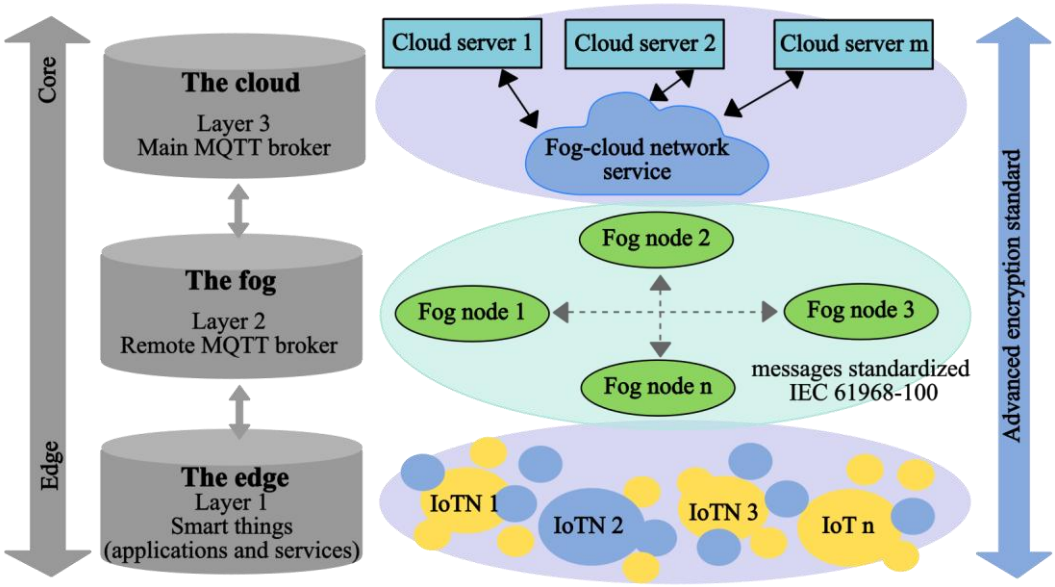
At the edge layer (Layer 1), a large number of IoT devices, including sensors and actuators, are deployed across the network to perform ultra-low-latency tasks such as local data

acquisition, event detection, and preliminary filtering (Bhupathi; Pandraka; Malladi, 2024). These devices generate high-frequency data streams that are transmitted to the fog layer.

The fog layer (Layer 2) comprises multiple distributed nodes, such as gateways and local servers, interconnected within a fog-cloud network service. Fog nodes may communicate laterally and vertically, enabling localized data aggregation, preprocessing, and time-sensitive decision-making (Singh; Yassine, 2019). This layer also incorporates remote MQTT brokers and supports standardized communication protocols (e.g., IEC 61968), ensuring interoperability and reduced latency compared to fully centralized architectures.

The cloud layer (Layer 3), which includes many cloud servers, is part of a centralized platform that handles global system coordination, advanced analytics, and storing large amounts of data (Bhattarai et al., 2019). The cloud hosts the main MQTT broker and manages the core infrastructure of the utility, providing computational resources for non-real-time tasks. Additionally, the architecture supports secure data exchange through advanced encryption mechanisms across all layers, ensuring data integrity and system reliability (Alomar, 2023).

Figure 8 — Three-layer edge-fog-cloud architecture for IoT networks.



The limitations of cloud, fog, and edge computing architectures are primarily determined by the location of data processing (Yang et al., 2019). Cloud computing, while suitable for large-scale analytics, suffers from high latency, dependence on stable connectivity, and centralized storage, which introduces reliability and security vulnerabilities. Fog computing mitigates these issues by enabling intermediate processing and regional coordination; however, it requires additional infrastructure, increasing system complexity and deployment costs. Edge computing provides ultra-low latency and enhanced privacy through local data processing, but

its performance is constrained by limited computational resources, memory, and energy availability at the device level. Thus, edge devices are best suited for lightweight, real-time tasks, while more complex analytics must be offloaded to fog or cloud layers. Table 3 summarizes these trade-offs, highlighting their implications for SG operation.

Table 3 — Comparative limitations of cloud, fog, and edge computing for SGs

Dimension	Cloud	Fog	Edge	Implication for SG	References
Latency	High (remote processing)	Moderate (near-device)	Ultra-low (on-site)	Delays hinder real-time control	(Bhattarai et al., 2019)
Bandwidth	High (raw data transfer)	Medium (partial preprocessing)	Low (local filtering)	Network congestion and higher costs	(Yang et al., 2019)
Scalability	Limited by central bottlenecks	Improved, requires infrastructure	Device-dependent scaling	Constrains large-scale IoT integration	(Singh; Yassine, 2019)
Reliability	Dependent on connectivity	Improved via distribution	High for local operation	Failures affect monitoring/control	(Ahmad; Zhang, 2021)
Security & Privacy	Centralized, single point of failure	Reduced exposure	Strongest (local data retention)	Risk of breaches and data loss	(Alomar, 2023)
Energy	High (data centers, transmission)	Moderate (shared resources)	Low–moderate (device-side)	Impacts operational cost and sustainability	(Hong; Varghese, 2020)

3.2 COMMUNICATION PROTOCOL: MQQT

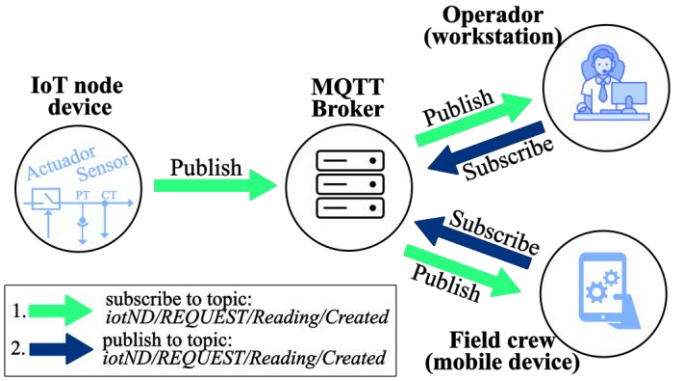
The MQTT protocol is a lightweight communication standard widely adopted in IoT-based smart grid architectures. This enables efficient, reliable, and scalable data exchange between distributed field devices and central utility systems (Jia et al., 2018). It defines the communication rules and message exchange mechanisms within multi-tier computational networks.

The MQTT protocol operates under a publish-subscribe architecture. In this model, a central broker acts as an intermediary that decouples message producers (publishers) from consumers (subscribers), enabling asynchronous data exchange. Communication is organized through a topic-based structure. Client devices subscribe to specific topics of interest, and the broker automatically forwards any messages published to those topics to all relevant subscribers. This mechanism eliminates the need for direct device-to-device communication (Veeramanikandan; Sankaranarayanan, 2019). It significantly enhances system flexibility and

scalability, as nodes can be dynamically added or removed without requiring network reconfiguration. Furthermore, MQTT incorporates reliability features such as persistent messaging. The broker retains the most recent message for each topic, allowing newly connected subscribers to immediately access the latest available data, even if the publishing device has experienced temporary disconnections.

In practical implementations, IoT nodes, databases, and field devices act as publishers, while operator workstations and mobile units function as subscribers, all interacting through the broker as a communication hub. The overall message flow and interaction among system components are illustrated in Figure 9.

Figure 9 — MQTT-based publish–subscribe communication architecture.



3.3 HARMONIZATION OF THE IOT NODE IN THE DISTRIBUTION SYSTEM

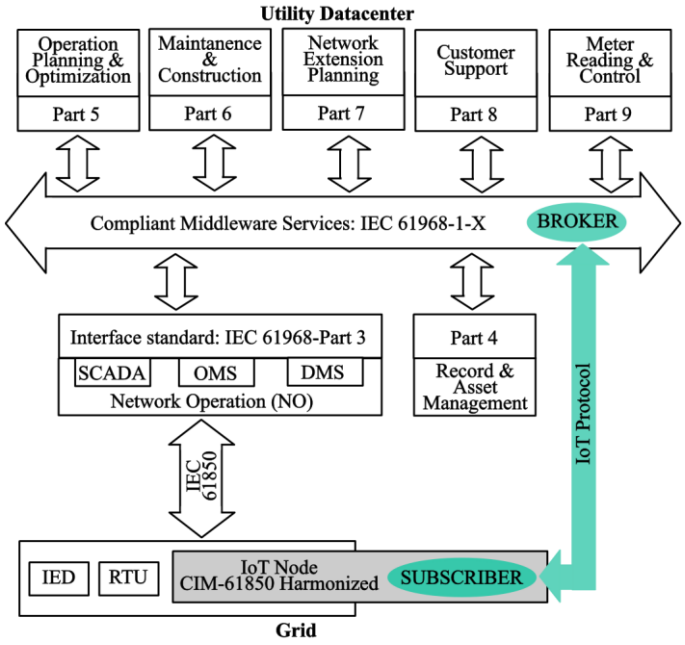
Harmonizing IoT nodes within the distribution system relies on establishing interoperability between heterogeneous field devices and utility platforms (De Barros; Leite, 2021). To ensure that these devices can communicate seamlessly across the grid, communication and data standards must be harmonized into a unified architecture. The harmonization relies on establishing a common information model that integrates the IEC 61970/61968 standards (typically used in utility control centers) with the IEC 61850 standard (used for substation and feeder automation) (Leite; Kezunovic, 2023). By harmonizing these standards, modern IoT nodes operating at the grid's edge can seamlessly connect and exchange information with the centralized distribution management system (DMS).

The harmonization is structured across a three-tier hierarchical integration architecture, as illustrated in Figure 10. At the top layer, the utility datacenter functions as the central hub for high-level operations, including supervisory control, system planning, and maintenance scheduling. At the bottom layer, the physical grid comprises field-level equipment such as

intelligent electronic devices (IEDs) and remote terminal units (RTUs), complemented by IoT nodes that enhance sensing capabilities and enable fine-grained monitoring across the network. Bridging these two layers, the middleware provides the necessary coordination framework, hosting critical services such as Supervisory Control and Data Acquisition (SCADA), the Outage Management System (OMS), and the Distribution Management System (DMS), thereby enabling seamless integration between operational technologies and information systems.

To support efficient and reliable communication across this architecture, a dual-path communication framework is employed. The traditional communication path facilitates interoperability between legacy devices, such as IEDs and RTUs, and centralized systems (e.g., SCADA and DMS) using standardized protocols like IEC 61850. In parallel, a modern IoT communication path is introduced to accommodate IoT nodes, leveraging lightweight protocols optimized for high-frequency data transmission and low-latency requirements. In the new path (green path), each IoT node acts as a publisher or subscriber, transmitting data to a central middleware broker. This broker translates and manages topic-based message distributions, interfacing directly with the utility's main communication backbone (standardized under IEC 61968).

Figure 10 — IoT node device integration architecture for SG data harmonization.



This architecture ensures that high-velocity IoT data is natively understood by central utility systems alongside traditional grid data. This allows the grid to combine local edge computing intelligence with centralized coordination, enabling rapid, automated responses essential for advanced self-optimizing and self-healing grid operations, such as Fault Location, Isolation, and Service Restoration (FLISR).

3.4 CLUSTERING IN POWER NETWORKS

MQTT-based communication and strategically allocated fog brokers could minimize transmission latency, enabling grid automation tools. The strategic positioning of computational resources is a critical engineering challenge within an expansive and often irregular physical grid topology. A distribution network consists of numerous buses and branches spatially distributed with respect to a reference node (root bus), which typically serves as the central communication point. As the network expands, the average distance between buses and the root increases, leading to longer communication paths and higher latency. To mitigate this effect, spatial clustering techniques are applied to group buses based on their geographical proximity, thereby reducing communication distances and enabling efficient decentralized processing.

The K-means algorithm aims to partition the dataset $X = \{x_i \mid i = 1, \dots, n\}$ into k non-overlapping clusters by minimizing the sum of squared Euclidean distances between each data sample x_i and its corresponding cluster centroid m_j. The dataset X of n samples is partitioned in C_j clusters as shown in (7). The algorithm follows an iterative process that begins by randomly selecting the centroid of each cluster m_j and assigning each sample x_i to the nearest centroid by (8).

$$C = \bigcup_{j=1}^{k} C_j \tag{7}$$

$$C_0 = \min_{C} \sum_{j \in k} \sum_{m \in C_j} d^2(x_m, m_j) \tag{8}$$

The algorithm, at each iteration t, assigns each sample to the nearest centroid according to (9). Subsequently, centroid positions are updated as the mean of the points within each cluster according to (10).

$$C_j^t = \left\{ x_i | d_2\left(x_i, m_j^t\right) \leq d_2\left(x_i, m_{j^*}^t\right), \forall j^* = 1,2, \dots, k \right\} \tag{9}$$

$$m_j^t = \frac{1}{\left|C_j^t\right|} \sum_{x_j \epsilon C_j^t} x_j \tag{10}$$

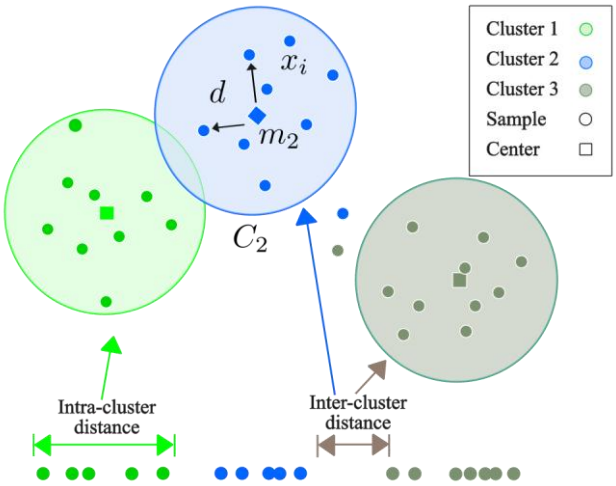
The initialization of centroids plays a critical role in the performance of the algorithm. It can be based on the spatial boundaries of the dataset, defined by the minimum and maximum coordinates along the $X$and Yaxes according to equations (11) and (12). Alternatively, centroids may be initialized randomly using existing bus locations within the distribution network.

$$m_{X,j}^{0} = X_{min} + \frac{X_{max} - X_{min}}{k+1} * j, for\ j = 1,2, \dots, k \quad (11)$$

$$m_{Y,j}^{0} = X_{min} + \frac{Y_{max} - Y_{min}}{k+1} * j, for\ j = 1,2, \dots, k \quad (12)$$

The clustering process is illustrated in Figure 11, where buses are grouped into distinct clusters C_1, C_2, C_3, each associated with a centroid m_j. Within each cluster, intra-cluster distances are minimized, ensuring compactness, while inter-cluster separation is preserved (VanderPlas, 2016).

Figure 11 — Clustering process for the k-means algorithm.



The centroid location can be further refined by incorporating additional components that account for network characteristics beyond purely geometric distance. In this context, the distance metric is extended as presented in (13).

$$d_2(x_m, m_j)^2 = \sqrt{(X_{xm} - X_{mj})^2 + (Y_{xm} - Y_{mj})^2 + (Z_{xm} - Z_{mj})^2 + (P_{xm} - P_{mj})^2} \quad (13)$$

where X and Y represent spatial coordinates, Z denotes network impedance, and P corresponds to load power. This formulation captures not only geometric separation but also topological and operational proximity, leading to a more representative clustering of electrical network elements (Chumbi Wilson E et al., 2022).

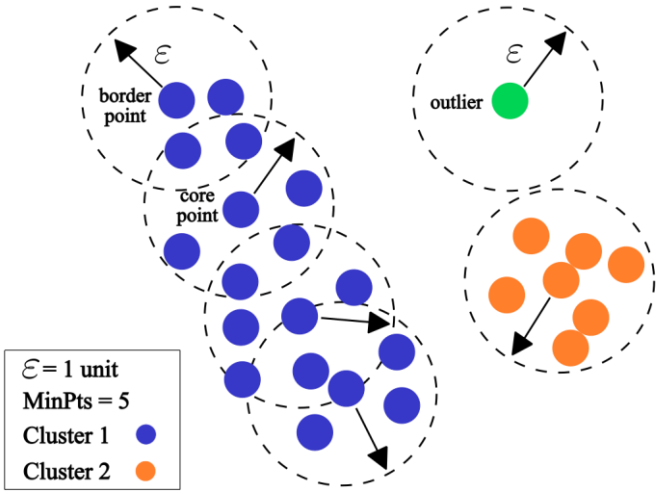
While the K-means algorithm effectively partitions data into compact and well-separated clusters by minimizing distance, it assumes clusters to be spherical and of similar size. These assumptions may not hold in real-world distribution systems, where bus locations and load patterns are often irregular. To address this limitation, density-based clustering methods are introduced as a complementary approach.

Algorithms such as DBSCAN, OPTICS, and DENCLUE identify clusters based on the spatial density of data points rather than distance alone. Among these, DBSCAN is selected due

to its ability to detect clusters of arbitrary shape and handle noise effectively. A key advantage of DBSCAN is that it does not require prior specification of the number of clusters, instead determining them dynamically based on data density (Chumbi Wilson E et al., 2022). Additionally, it provides robustness to outliers and allows parameter tuning based on domain-specific knowledge.

DBSCAN operates using two parameters: the neighborhood radius (ε) and the minimum number of points ($MinPts$) required to define a dense region. A point is classified as a core point if its ε-neighborhood contains at least $MinPts$ points. A border point lies within the neighborhood of a core point but does not satisfy the density condition itself, while outliers are points that do not belong to any cluster. Figure 12 illustrates this classification, where dashed circles represent ε-neighborhoods with $\varepsilon = 1$ and MinPts = 5.

Figure 12 — Clustering process for DBSCAN algorithm.



Each centroid obtained from the clustering process represents a candidate location for fog broker deployment. However, practical implementation must consider physical and operational constraints, including the availability of communication infrastructure, power supply, environmental conditions, and maintenance accessibility (Meehan, 2013). In practice, fog nodes are typically deployed in three types of locations: (i) substations, which offer secure and well-connected environments; (ii) pole-mounted gateways, providing cost-effective and decentralized solutions for feeder-level applications; and (iii) regional data centers, which act as intermediaries between fog and cloud layers.

To enhance deployment feasibility, the centroid-based allocation can be integrated with a site suitability analysis (Chumbi et al., 2024). This approach enables the selection of locations that balance spatial proximity, infrastructure availability, and network accessibility, recognizing that optimal deployment sites may not coincide with purely geometric centroids.

3.5 TRANSMISSION LATENCY IN MULTI-TIER ARCHITECTURES

Latency is defined as the total time required for a signal or data packet to travel from a source to a destination and return to its origin. The transmission latency, denoted as λ_t, is a key performance indicator for evaluating the responsiveness of communication architectures in smart grids. In multi-tier IoT frameworks, latency is influenced by the number of communication stages and the distances between interacting components across the edge, fog, and cloud layers.

For each event generated by an IoT node δ, typically triggered by a sensor or actuator, a request is transmitted to a remote broker φ located in the fog layer. The broker processes the request locally and, if required, forwards it to the main broker Φ in the cloud layer for further analysis. The response message follows the reverse communication path, forming a complete round-trip transmission. The total latency for the fog-based architecture is expressed in (14).

$$\overline{\lambda_{t_f}} = \sum \left(l_{\delta}^{\varphi} + l_{\varphi}^{\Phi} + l_{\Phi}^{\varphi} + l_{\varphi}^{\delta} \right) \tag{14}$$

where l represents the communication distance between two endpoints.

In contrast, the cloud-based architecture involves direct communication between each IoT node and the central broker, bypassing intermediate fog processing. The corresponding latency is given by (15).

$$\overline{\lambda_{t_c}} = \sum \left(l_{\delta}^{\Phi} + l_{\Phi}^{\delta} \right) \tag{15}$$

Due to long transmission distances and potential environmental interference, data packets may require retransmission to ensure successful delivery. This effect is represented by a retransmission factor τ, assumed to correspond to 1.5% of the total transmission time (Veeramanikandan; Sankaranarayanan, 2019). Incorporating this factor, the average transmission latency per IoT node is defined in (16).

$$\lambda_t = \frac{\overline{\lambda_t} + \tau}{n} \tag{16}$$

where n denotes the total number of IoT nodes generating communication requests. This formulation accounts for both the cumulative transmission delay and the additional overhead introduced by retransmissions, normalized over the number of nodes.

3.6 CLUSTERING PERFORMANCE

When applying the K-means algorithm to cluster network buses, determining the optimal number of clusters k is a critical step. This problem is addressed using the bisection method, a numerical technique based on the intermediate value theorem, which iteratively refines the search interval $[a, b]$ until a convergence criterion is satisfied.

In this study, the interval is defined as $a = 2$, representing the minimum feasible network partition, and $b = n$, corresponding to the extreme case where each bus forms an individual cluster. The optimization objective is the minimization of the total communication distance, defined as the sum of distances between each bus and its assigned centroid, plus the distance between each centroid and the root bus.

By integrating K-means clustering with the bisection method, it is possible to simultaneously determine the optimal number of clusters and the corresponding distance configuration, as explained in (Fontolan; Quito; Leite, 2024). Once the clustering structure is established, latency values for both cloud-based and fog-based architectures are computed using (15) and (14), respectively, while the average latency is obtained from (16).

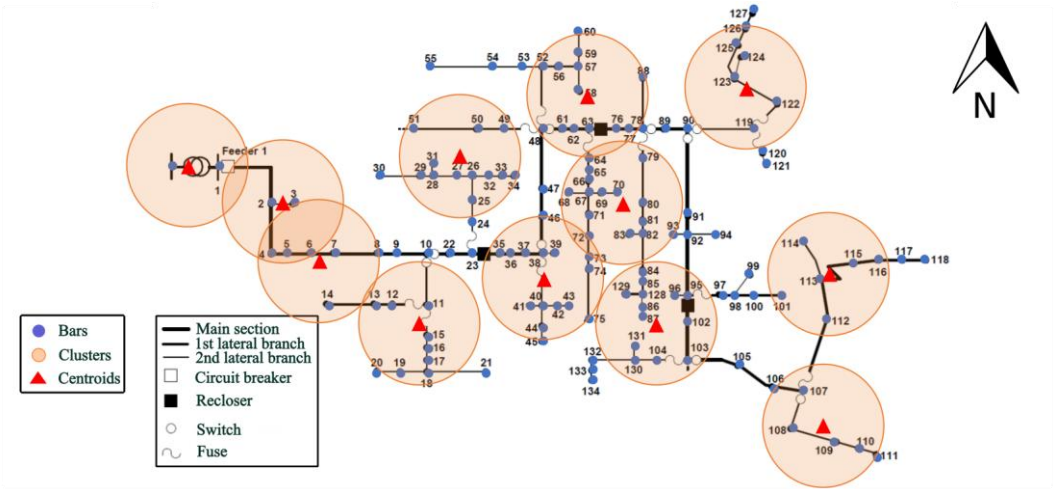
The methodology is validated using a 3.8 kV distribution network based on the IEEE 136-bus test system. The network is geographically referenced to enable spatial clustering and distance calculations between key components, including buses, remote brokers, and the main broker located at the root bus. A propagation delay of 4 μs/km is adopted, consistent with terrestrial communication standards (ITU-T G.114, 2003). The resulting transmission distances and latency values are summarized in Table 4.

Latency in the fog-based architecture is calculated considering round-trip communication paths and the same propagation delay. Applying the bisection method yields an optimal number of clusters $k = 12$, where centroids represent candidate locations for fog brokers. As reported in Table *4*, the fog-based approach achieves a significant reduction in communication overhead, with an approximate 81% decrease in latency compared to the cloud-based model. The resulting clustering configuration is depicted in Figure 13**.**

Table 4 — Comparison of latency and total transmission distances between two architectures

Model	**Cloud-based model**	**Fog-based model**
Sum of routes (km)	71.21	13.56
Sum of Latencies (ms)	0.57	0.11
Average latency (μs)	4.25	0.81

Figure 13 — 136-bus test system and distribution of clustered buses



Unlike K-means, the DBSCAN algorithm does not require a predefined number of clusters. However, its performance depends on two key parameters: the neighborhood radius ε and the minimum number of points MinPts. A common practice sets MinPts $\in [4, 2 \cdot \text{dim}]$, where *dim* represents the dimensionality of the dataset.

The parameter ε is typically determined using a k-distance graph, where distances to the k-th nearest neighbor are sorted in ascending order. The point of maximum curvature ("elbow") indicates an appropriate threshold, separating dense regions from outliers.

A sensitivity analysis was conducted using a dataset of 251 geographically referenced electrical loads, represented in a three-dimensional space (x, y, power). The results, shown in Table 5, indicate that increasing MinPts reduces the number of clusters while increasing the number of outliers. Similarly, higher values of ε lead to fewer clusters and fewer outliers due to the expansion of neighborhood regions. The Silhouette Score metric improves as the number of clusters decreases, indicating better-defined cluster structures.

Table 5 — DBSCAN parameter tuning with different ε and $MinPts$

Scenario	ε	$MinPts$	Number of clusters	Outliers	Silhouette score	Execution time
1	0.30	4	10	21	0.217	0.19
	0.35	4	9	12	0.288	0.18
	0.40	4	6	9	0.485	0.16
	0.45	4	5	9	0.538	0.15
2	0.30	5	7	33	0.328	0.25
	0.35	5	6	25	0.277	0.24
	0.40	5	4	19	0.490	0.22
	0.45	5	4	14	0.524	0.22

3	0.30	6	7	38	0.359	0.16
	0.35	6	6	29	0.272	0.16
	0.40	6	4	23	0.482	0.14
	0.45	6	4	14	0.524	0.14

A visual comparison of clustering results is presented in Figure 14, highlighting the differences between K-means and DBSCAN. DBSCAN identifies clusters with irregular shapes and explicitly detects outliers, adapting naturally to the spatial distribution of data. In contrast, K-means produces compact, convex clusters with a predefined number of groups, assigning all data points to a cluster. K-means benefits from flexibility in defining distance metrics. In this work, a multidimensional distance function incorporates spatial coordinates, electrical distance, and load power, enabling a more comprehensive spatial-electrical clustering. This approach extends traditional geographic clustering to a multi-attribute analysis, allowing simultaneous grouping of loads and partitioning of the electrical network.

The quality of clustering results is assessed using three standard metrics: the Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI), and Silhouette Score. Lower DBI values indicate better separation and compactness, while higher CHI and Silhouette values reflect improved cluster cohesion and distinction. As shown in Table 6, the configuration with $k = 4$ provides the best overall performance, achieving the lowest DBI (0.5129) and the highest Silhouette Score (0.531) with moderate computational cost.

Figure 14 — DBSCAN and K-means algorithm comparison results

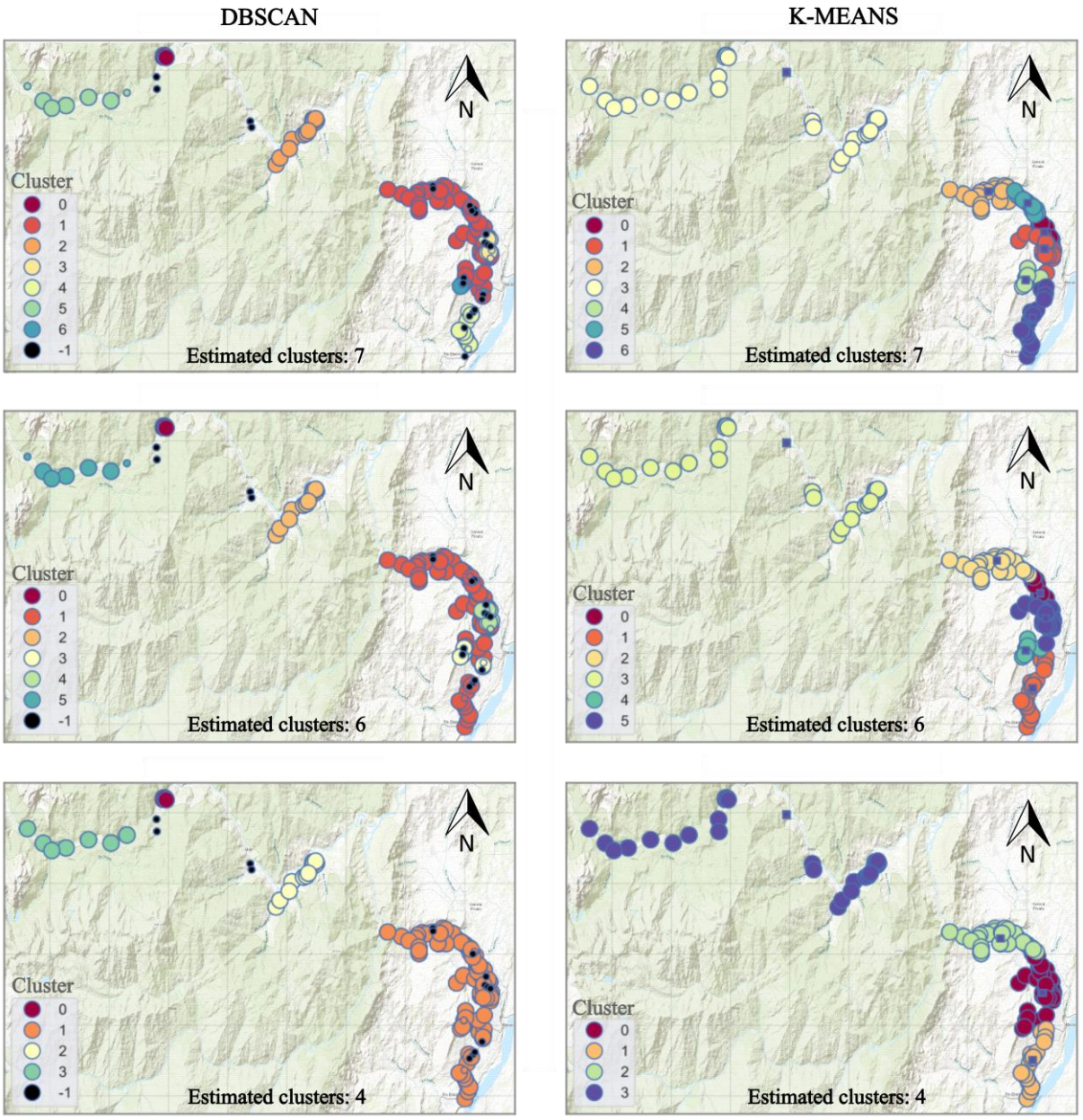


Table 6 — Performance metrics for k-means algorithm under k clusters

Metric	$k = 7$	$k = 6$	$k = 5$	$k = 4$	$k = 3$
Davies–Bouldin Index	0.7099	0.5792	0.5427	0.5129	0.5707
Calinski–Harabasz Index	327.3976	388.1579	453.623	562.927	596.445
Silhouette Score	0.417	0.486	0.477	0.531	0.506
Execution Time (s)	92.508	86.501	63.368	74.054	32.989

4 CONCLUDING REMARKS

This postdoctoral research addressed two fundamental pillars required for the digitalization and decarbonization of modern power distribution systems: the development of realistic digital representations of electrical infrastructure and the design of decentralized computational architectures capable of supporting data-intensive smart-grid applications. The results demonstrate that geospatial information systems, graph theory, internet of things technologies, and edge computing paradigms can be mixed to provide scalable solutions.

Regarding the first research front, the proposed GIS-driven graph methodology positively transformed raw street-map information into synthetic distribution networks constrained by real urban infrastructure. The validation carried out on the electrical infrastructure of the utility CENTROSUR demonstrated that integrating spatial constraints from street maps eliminates the unrealistic, non-feasible geometric crossings. Furthermore, by incorporating a space-partitioning 2D KD-tree algorithm, customer loads were associated with support structures at high speeds, achieving a minor 2.42% relative difference in mean service-drop length estimation compared to real physical networks. Finally, the consecutive graph reduction phase compressed the finalized low-voltage topologies by approximately 2.6 times in both node and edge counts. This reduction drastically minimizes graph computational complexity while fully preserving topological, spatial, and statistical fidelity, creating computationally efficient environments suitable for large-scale DER integration studies.

On the second research front, the research presented the benefits of decentralized edge-fog-cloud architectures for smart-grid applications. By leveraging the lightweight MQTT publish-subscribe communication protocol and establishing a unified semantic data model, interoperability between legacy substation standards (IEC 61850) and edge IoT data streams can become successfully harmonized. To optimize data routing, a multidimensional clustering framework incorporating spatial coordinates, network impedance, and active power demand was evaluated. The K-means algorithm coupled with the numerical bisection method enables get an optimal configuration of clusters within the electrical network. The results showed that properly positioned fog brokers can significantly reduce communication distances and transmission latency, achieving reductions of approximately 81% compared with centralized cloud-based architectures. Such improvements are particularly relevant for time-critical applications, such as FLISR management.

4.1 CHALLENGES AND FUTURE DIRECTIONS

Based on the findings of this research, several opportunities for future work can be identified:

- *Enhanced Electrical Modeling of SDNs:* Although the proposed GIS-driven methodology provides realistic spatial and topological representations, future developments should incorporate additional electrical attributes, including conductor parameters, transformer allocation, protection devices, voltage regulation equipment, and time-varying load profiles. Integrating these characteristics would improve the applicability of SDNs for operational analysis, resilience assessment, and planning studies under high penetrations of DERs;
- *Intelligent and Secure Multi-Tier Computing Architectures:* The increasing complexity of multi-tier edge-fog-cloud architectures creates opportunities for the application of artificial intelligence (AI) to support distributed decision-making. Future research may investigate AI-based approaches for autonomous resource management, DER coordination, and communication optimization, while integrating secure communication frameworks and distributed transaction mechanisms.
- *Extension to Multi-Energy Systems:* Recent studies have highlighted the growing interaction between electricity networks and other energy vectors, including hydrogen, thermal systems, and industrial energy processes. Future work may extend the proposed geospatial modeling and edge-computing frameworks to support integrated multi-energy infrastructures, enabling coordinated planning and operation across coupled energy systems;
- *Grid-aware decentralized energy markets:* The increasing deployment of DERs motivates the development of decentralized market mechanisms and the adoption of energy communities. Future research may investigate peer-to-peer energy trading frameworks based on optimization AI techniques, allowing local energy transactions constrained by realistic operational constraints of the underlying distribution network;
- *Adaptive Fog Resource Allocation for Resilient Grid Operation:* Although the present work considers static fog-broker allocation, future smart-grid applications will require adaptive communication infrastructures capable of responding to changing operating conditions. Future studies may develop dynamic clustering and fog-resource allocation strategies that continuously reconfigure communication architectures according to network conditions, contingencies, and extreme events.

BIBLIOGRAPHY

AHMAD, T.; ZHANG, D. Using the internet of things in smart energy systems and networks. **Sustainable Cities and Society**, v. 68, maio 2021. ISSN: 22106707.

AL-ALI, A. R.; GUPTA, R.; ZUALKERNAN, I.; DAS, S. K. Role of IoT technologies in big data management systems: A review and smart grid case study. **Pervasive and Mobile Computing**, v. 100, p. 101905, maio 2024. ISSN: 15741192.

ALI, M.; PRAKASH, K.; MACANA, C.; RAZA, M. Q.; BASHIR, A. K.; POTA, H. Modeling synthetic power distribution network and datasets with industrial validation. **Journal of Industrial Information Integration**, v. 31, p. 100407, fev. 2023. ISSN: 2452414X.

ALOMAR, M. A. An IOT based smart grid system for advanced cooperative transmission and communication. **Physical Communication**, v. 58, p. 102069, jun. 2023. ISSN: 18744907.

BHATTARAI, B. P.; PAUDYAL, S.; LUO, Y.; MOHANPURKAR, M.; CHEUNG, K.; TONKOSKI, R.; HOVSAPIAN, R.; MYERS, K. S.; ZHANG, R.; ZHAO, P.; MANIC, M.; ZHANG, S.; ZHANG, X. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. **IET Smart Grid**, v. 2, n. 2, p. 141–154, jun. 2019. ISSN: 2515-2947.

BHUPATHI, H. P.; PANDRAKA, V. K.; MALLADI, L. S. Internet of Things (IoT) in smart grid technology and its applications. Em: 2024 IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS, SMART AND GREEN TECHNOLOGIES (ICISSGT) 2024, Visakhapatnam, India. **Proceedings**, Visakhapatnam, India: IEEE, nov. 2024. p. 91–96.

CAETANO, H. O.; DESUÓ N., L.; FOGLIATTO, M. de S. S.; RIBEIRO, V. P.; BALESTIERI, J. A. P.; MACIEL, C. D. A Bayesian hierarchical model to create synthetic power distribution systems. **Electric Power Systems Research**, v. 235, p. 110706, out. 2024. ISSN: 03787796.

CHUMBI, W. E.; MARTÍNEZ-MINGA, R.; ZAMBRANO-ASANZA, S.; LEITE, J. B.; FRANCO, J. F. Suitable site selection of public charging stations: A Fuzzy TOPSIS MCDA framework on capacity substation assessment. **Energies**, v. 17, n. 14, p. 3452, jul. 2024. ISSN: 1996-1073.

CHUMBI WILSON E; C. JOSÉ EDUARDO; JARAMILLO-LEON BRIAN; B. LEITE JÔNATAS. Análise de métodos para agrupamento de cargas elétricas em redes de distribuição de grande porte usando sistema de informação geográfica. Em: XXIV CONGRESSO BRASILEIRO DE AUTOMÁTICA - CBA 2022, Fortaleza, Brazil. **Proceedings**, Fortaleza, Brazil: Sociedade Brasileira de Automática (SBA), out. 2022. p. 686–693.

DE BARROS, J. V.; LEITE, J. B. Development of a relational database oriented on the common information model for power distribution networks. Em: 2021 IEEE URUCON 2021, Montevideo, Uruguay. **Proceedings**, Montevideo, Uruguay: IEEE, nov. 2021. p. 63–66.

DEO, Narsingh. **Graph Theory with Applications to Engineering & Computer Science**. First Edition ed. New Jersey: Dover Publications, ago. 2016. ISBN: 0486807932.

FONTOLAN, L.; QUITO, W. E. C.; LEITE, J. B. Determinação da latência de transmissão em duas arquiteturas de internet das coisas na digitalização das redes de distribuição. Em: LVI SBPO 2024, Fortaleza, Brazil. **Proceedings**, Fortaleza, Brazil: Simposio Brasileiro de Pesquisa Operacinal, nov. 2024.

HONG, C.-H.; VARGHESE, B. Resource management in fog/edge computing. **ACM Computing Surveys**, v. 52, n. 5, p. 1–37, set. 2020. ISSN: 0360-0300.

JIA, K.; XIAO, J.; FAN, S.; HE, G. A MQTT/MQTT-SN-based user energy management system for automated residential demand response: Formal verification and cyber-physical performance evaluation. **Applied Sciences**, v. 8, n. 7, p. 1035, jun. 2018. ISSN: 2076-3417.

KAYS, J.; SEACK, A.; SMIREK, T.; WESTKAMP, F.; REHTANZ, C. The generation of distribution grid models on the basis of public available data. **IEEE Transactions on Power Systems**, [s. l.], v. 32, p. 2346–2353, jun. 2017.

LEITE, J. B.; KEZUNOVIC, M. Integrating IoT devices with distribution energy management system by harmonizing their logical models using IEC standards 61970/61968 and 61850. Em: IEEE POWER AND ENERGY SOCIETY GENERAL MEETING 2023, Orlando, USA. **Proceedings**, Orlando, USA: IEEE Computer Society, jul. 2023.

LI, H.; WERT, J. L.; BIRCHFIELD, A. B.; OVERBYE, T. J.; ROMAN, T. G. S.; DOMINGO, C. M.; MARCOS, F. E. P.; MARTINEZ, P. D.; ELGINDY, T.; PALMINTIER, B. Building highly detailed synthetic electric grid data sets for combined transmission and distribution Systems. **IEEE Open Access Journal of Power and Energy**, v. 7, p. 478–488, out. 2020. ISSN: 2687-7910.

LÖBBE, Sabine.; SIOSHANSI, F. P. .; ROBINSON, David. **Energy communities: customer-centered, market-driven, welfare-enhancing?** Lonndon, UK: Academic Press, 2022. ISBN: 9780323911399.

MAHMOOD, M.; CHOWDHURY, P.; YEASSIN, R.; HASAN, M.; AHMAD, T.; CHOWDHURY, N.-U.-R. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. **Energy Conversion and Management: X**, v. 24, p. 100790, out. 2024. ISSN: 25901745.

MATEO, C.; POSTIGO, F.; DE CUADRA, F.; ROMAN, T. G. S.; ELGINDY, T.; DUEÑAS, P.; HODGE, B.-M.; KRISHNAN, V.; PALMINTIER, B. Building large-scale U.S. synthetic electric distribution system models. **IEEE Transactions on Smart Grid**, v. 11, n. 6, p. 5301–5313, nov. 2020. ISSN: 1949-3053.

MATEO DOMINGO, C.; GOMEZ SAN ROMAN, T.; SANCHEZ-MIRALLES, A.; PECO GONZALEZ, J. P.; CANDELA MARTINEZ, A. A reference network model for large-scale distribution planning with automatic street map generation. **IEEE Transactions on Power Systems**, v. 26, n. 1, p. 190–197, fev. 2011. ISSN: 0885-8950.

MEEHAN, B. **GIS for Enhanced Electric Utility Performance**. Boston: House Artech, 2013. ISBN: 9781608075591.

ONETO, A.; GJORGIEV, B.; TETTAMANTI, F.; SANSAVINI, G. Large-scale generation of geo-referenced power distribution grids from open data with load clustering. **Sustainable Energy, Grids and Networks**, v. 42, p. 101678, jun. 2025. ISSN: 23524677.

PALMINTIER, B.; ELGINDY, T.; MATEO, C.; POSTIGO, F.; GÓMEZ, T.; DE CUADRA, F.; MARTINEZ, P. D. Experiences developing large-scale synthetic U.S.-style distribution test systems. **Electric Power Systems Research**, v. 190, p. 106665, jan. 2021. ISSN: 03787796.

PILO, F.; SILVESTRO, F. Active distribution network operation and planning. Em: **CIGRE Green Books**: Springer, Cham, abr. 2026. p. 563–608.

SINGH, S.; YASSINE, A. IoT big data analytics with fog computing for household energy management in smart grids. Em: Switzerland: Springer Nature, 2019. p. 13–22.

SIOSHANSI, F. P. . **Innovation and disruption at the grid's edge: how distributed energy resources are disrupting the utility business model**. London, UK: Academic Press Elsevier, 2017. ISBN: 9780128117583.

TAYAB, U. B.; RAHMAN, M. T.; AZIZ, S.; HASAN, K. N. Modelling and validation of geographic information system (GIS)-based synthetic distribution network. Em: 2024 IEEE POWER & ENERGY SOCIETY GENERAL MEETING (PESGM) 2024, Seattle, WA, USA. **Proceedings**, Seattle, WA, USA: IEEE, jul. 2024. p. 1–5.

VANDERPLAS, J. **Python data science handbook**. first ed. Sebastopol, USA: O'REILLY, dez. 2016. v. 53, ISBN: 9788578110796.

VEERAMANIKANDAN, M.; SANKARANARAYANAN, S. Publish/subscribe based multi-tier edge computational model in internet of things for latency reduction. **Journal of Parallel and Distributed Computing**, v. 127, p. 18–27, maio 2019. ISSN: 07437315.

YANG, Y.; LUO, X.; CHU, X.; ZHOU, M. T. **Fog-enabled intelligent IoT systems**. Switzerland AG: Springer International Publishing, out. 2019. ISBN: 9783030231859.

ZAMBRANO-ASANZA, S.; CANDO, D. J.; CHUQUI, F. H.; SANANGO, J.; FRANCO, J. F. Graph partitioning-based clustering for the planning of distribution network topology using spatial- temporal load forecasting. Em: 2021 IEEE PES INNOVATIVE SMART GRID TECHNOLOGIES CONFERENCE - LATIN AMERICA (ISGT LATIN AMERICA) 2021, Lima, Peru. **Proceedings**, Lima, Peru: IEEE, set. 2021. p. 1–5.

APPENDIX A – Scientific Production

The scientific production generated during the postdoctoral research period includes publications and collaborative contributions directly associated with the research activities reported herein. These contributions have resulted in publications in international scientific venues and book chapters under editorial publishing.

Type	Tittle	Status	Publisher	Relation to Postdoc
IEEE Conference Article	GIS-Driven Graphs and KD-Trees for Synthetic Distribution Network Construction	Published (Sep. 2025)	ISGT LA	Direct
Book Chapter	Multi-tier edge computing for IoT-based smart grids: Latency-aware data processing and resource management	Accepted (Apr. 2026)	Springer Nature	Direct

International Research Proposal Development

During the postdoctoral period, an international research proposal entitled *"Advanced AI for Decentralized Multi-Energy Systems (AI4MES)"* was submitted to the European Commission under the Marie Skłodowska-Curie Postdoctoral Fellowships (MSCA-PF 2025) programme in collaboration with the Universidade do Porto (Portugal). The proposal focused on the application of Artificial Intelligence for decentralized multi-energy systems, including peer-to-peer energy trading, deterministic optimization models, and adaptive AI-driven market management.

The proposal received a final evaluation score of 88.80/100, exceeding the funding threshold of 70/100 and obtaining an "Excellent" rating in the Excellence criterion. Although it was not selected for funding due to budget limitations and the high competitiveness of the call, the evaluation demonstrates the scientific quality and international relevance of the research developed during the postdoctoral period.

Proposal	Program	Host Institution	Score	Result
AI4MES – Advanced AI for Decentralized Multi-Energy Systems	Horizon Europe MSCA-PF 2025	Universidade do Porto	88.8/100	Evaluated above funding threshold

GIS-Driven Graphs and KD-Trees for Synthetic Distribution Network Construction

1st Diego J. Cando
Department of Electrical Engineering, Electronics and Telecommunications
Universidad de Cuenca
Cuenca, Ecuador
diego.cando@ucuenca.edu.ec
https://orcid.org/0009-0001-6540-6795

2nd Santiago P. Torres
Department of Electrical Engineering, Electronics and Telecommunications
Universidad de Cuenca
Cuenca, Ecuador
santiago.torres@ucuenca.edu.ec
https://orcid.org/0000-0002-8803-6811

3rd Wilson E. Chumbi
Department of Electrical Engineering
São Paulo State University - UNESP
Ilha Solteira, SP, Brazil
enrique.quito@unesp.br
https://orcid.org/0000-0001-8347-0208

4th Milton Patricio Agudo
Department of Electrical Engineering
São Paulo State University - UNESP
Ilha Solteira, SP, Brazil
patricio.agudo@unesp.br
https://orcid.org/0000-0003-4995-7770

5th Sergio Zambrano-Asanza
Centre for Power and Energy Systems, Institute of Systems and Computer Engineering, Technology and Science – INESC TEC
Porto, Portugal
sergiopza.24@gmail.com
https://orcid.org/0000-0003-3662-0220

2025 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT LA) | 979-8-3315-9793-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISGTLA64895.2025.11371192

***Abstract*—Distribution networks were not designed to accommodate high penetration of distributed energy resources (DERs), microgrids, and electric vehicles (EVs), resulting in operational challenges. Studying new topologies is essential, and synthetic distribution networks (SDNs) are key tools for modeling and validating present or future grid architectures. In this context, this paper presents a four-stage GIS–graph methodology that integrates graph topology with geospatial layers. A GIS-based graph is constructed from a street map and load points corresponding to customer coordinates. In the first step, the process validates street geometries and repairs disconnections. Then, it segments edges to respect realistic span limits, creating candidate support nodes aligned with typical pole placements. In the third stage, a two-dimensional KD-tree links each load point to its nearest support node. Finally, loads are grouped at support nodes and unnecessary elements are removed. The resulting graph-based network is topologically, spatially, and statistically congruent with the existing network; moreover, it is approximately 2.6 times smaller than networks produced with geometric approximations such as Delaunay triangulation and K-nearest neighbors (KNN).**

***Index Terms*—synthetic networks, large-scale systems, kd-tree, graph theory**

I. Introduction

Distribution networks now face increasing complexity due to the integration of distributed energy resources (DERs), active customer participation, local markets, microgrid connections, and protection schemes [1], [2]. Synthetic distribution networks (SDNs) provide test systems for evaluating advanced algorithms designed to address these issues [3], [4].

While not exact replicas, SDNs are built using real-world data and characteristics to emulate the statistical and operational behavior of real networks [4], [5], thus providing open-access platforms for planning and operations tool development [3], [6]–[12].

An SDN typically starts from a street map and customer coordinates, which together define an undirected, connected graph; this base-graph is then enriched with a device catalog, network statistics, and expert knowledge [3], [4], [13]. GIS–graph coupling ranges from single-use static queries [5], [6], through bidirectional schemes that repeatedly consult road geometry during network synthesis [3], to fully dynamic approaches in which GIS attributes are embedded within the evolving graph [14].

In SDN design, the street map defines the feasible corridor for network layout, with three main sources in the literature: (i) OpenStreetMap (OSM), widely used when available [8], [9], [11], [12], [15], though its detail depends on the area; (ii) synthetic or geometric maps, created from load points [7], [11] or approximations like Delaunay or K-nearest neighbors (KNN) graphs refined with detour factors [2], [5], [6]; and (iii) governmental or commercial digital maps, whose geometric inconsistencies are acknowledged but whose correction steps are rarely documented [14], [16], [17].

With the street map ready, customers can be linked using four main anchoring strategies: (i) orthogonally projecting each customer onto the nearest road segment to define the LV service drop [9]; (ii) snapping load nodes directly to street-graph vertices, preserving topology but adding one node per customer [15], [16]; (iii) interpolating uniformly spaced virtual poles along the centerline of each street and connecting each customer to its nearest pole using a nearest-feature operation [14]; and (iv) placing regularly spaced potential service drop points that act as customer anchors and Steiner nodes for later optimization [3]. However, all of these approaches have

Authorized licensed use limited to: UNIV ESTADUAL PAULISTA JULIO DE MESQUITA FILHO. Downloaded on March 09,2026 at 20:20:27 UTC from IEEE Xplore. Restrictions apply.

unesp

Wilson Enrique Chumbi Quito <enrique.quito@unesp.br>

Decision on Chapter for the book: "Edge Computing for Smart Grid"

3 mensagens

Amitkumar V. Jha <amit.jhafet@kiit.ac.in> 10 de abril de 2026 às 03:28
Para: Wilson Enrique Chumbi Quito <enrique.quito@unesp.br>
Cc: nicu bizon <nicubizon@yahoo.com>, "bhargav.appasanifet" <bhargav.appasanifet@kiit.ac.in>

Dear Dr. Wilson,
We are pleased to inform you that your chapter titled "**Multi-tier edge computing for IoT-based smart grids: Latency-aware data processing and resource management**" with **Chapter No: 3**, has been **ACCEPTED**, for publication in the book.

Comments:

- Alt text is required for each figure

As per the Springer Guidelines, we need **Contributor Agreements form**. **Each corresponding author** should print the form, enter the title of the chapter as well as the names of the authors at the top of the form and sign it. It is important to note that the details of the chapter authors (sequence, names, name, and affiliation of corresponding author) given in the Contributor Agreement should be complete, final, and correspond to the details given in the manuscript. As this is a legal document, we cannot accept changes at a later stage and ask for your understanding in this matter.

You need to enter the following details in the contributor agreement form.

- Author and corresponding author (on page 1)
- Chapter title (on page 2)
- Sign, name and date of corresponding author (on page 9)

Please provide the following files by **12 April 2026.** For the submission, you can send the files as a reply to this email:

- A zipped folder containing all figures (High resolution)
- Contributor Agreement Form
- ALT text for each figure

Files attached for reference:

- ALT text template
- Contributor Agreement Form

Regards,
Dr. Amitkumar V. Jha
Associate Professor
School of Electronics Engineering
KIIT-DU (IoE)
Bhubaneswar-751024
Mob. 7978822792/766300813
Facebook || Twitter || Google Scholar || Website

Associated with document Ref. Ares(2026)692015 - 21/01/2026

Proposal Evaluation Form

EUROPEAN COMMISSION

Horizon Europe Framework Programme (HORIZON)

Evaluation Summary Report

Call:	HORIZON-MSCA-2025-PF
Type of action:	HORIZON-TMA-MSCA-PF-EF
Proposal number:	101279831
Proposal acronym:	AI4MES
Duration (months):	24
Proposal title:	Advanced AI for Decentralized Multi-Energy Systems
Activity:	EF-ENG

N.	Proposer name	Country	Total eligible costs	%	Grant Requested	%
1	UNIVERSIDADE DO PORTO	PT	191,343.12	100.00%	191,343.12	100.00%
	Total:		191,343.12		191,343.12	

Abstract:

The transition to a decarbonized economy requires innovative management of multi-energy systems (MES) and the empowerment of energy communities through peer-to-peer (P2P) trading. However, aligning local economic transactions with physical grid constraints at scale remains a major challenge, demanding computationally efficient and adaptive tools. The AI4MES project addresses this gap by developing an AI-enhanced optimization framework for Multi-Energy Distribution Grid Management (MEDGM), integrating electricity, heating, cooling, gas, and hydrogen systems.The framework combines three modules: (i) a P2P market mechanism for decentralized energy trading, (ii) a deterministic multi-energy grid optimization model to ensure system security and efficiency, and (iii) an adaptive AI-driven module based on deep reinforcement learning to handle uncertainty and scale. Its core innovation is an AI tool capable of internalizing network physics, such as congestion and voltage constraints, within the operation of local energy markets.Practical validation will be achieved through a 3-month secondment at CEVE, an industrial partner specializing in distribution networks, ensuring technology transfer and real-world relevance. Expected outcomes include high-impact publications, open-source software, and a scalable decision-support tool for Distribution System Operators (DSOs) and energy communities. Ultimately, AI4MES will strengthen Europe's leadership in energy digitalization and position the fellow as an independent research leader at the intersection of innovation and sustainable energy systems.

Evaluation Summary Report

Evaluation Result

Total score: 88.80 % (Threshold: 70 /100.00)

Criterion 1 - Excellence

Score: **4.60** (Threshold: 0 / 5.00 , Weight: 50.00%)

- **Quality and pertinence of the project's research and innovation objectives (and the extent to which they are ambitious, and go beyond the state of the art)**
- **Soundness of the proposed methodology (including interdisciplinary approaches, consideration of the gender dimension and other diversity aspects if relevant for the research project, and the quality of open science practices)**
- **Quality of the supervision, training and of the two-way transfer of knowledge between the researcher and the host**
- **Quality and appropriateness of the researcher's professional experience, competences and skills**

Quality and pertinence of the project's research and innovation objectives:

Excellent

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

Ilha Solteira/SP, 17 de junho de 2026.

À Seção Técnica Acadêmica – STA/FEIS/UNESP

Eu, Jonatas Boas Leite, professor do Departamento de Engenharia Elétrica da Universidade Estadual Paulista, UNESP, do câmpus de Ilha Solteira, SP, em uma avaliação final, declaro que o estagiário de pós-doutorado, Wilson E. C. Quito, realizou todas as atividades previstas no projeto de pesquisa. O estagiário, também, participou de bancas de defesa no curso de graduação em h elétrica e no programa de pós-graduação em engenharia elétrica. No período de vigência do pós-doutoramento, foram publicados um artigo de congresso e um capítulo de livro.

Portanto, o estagiário, Wilson E. C. Quito, cumpriu com êxito as atividades previstas para o estágio de pós-doutorado.

Fico a disposição para esclarecimentos adicionais.

Atenciosamente,

gov.br Documento assinado digitalmente
JONATAS BOAS LEITE
Data: 17/06/2026 17:10:10-0300
Verifique em https://validar.iti.gov.br

Professor J. B. Leite
Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP
Faculdade de Engenharia de Ilha Solteira
Departamento de Engenharia Elétrica
Fone: +55 (18) 3742-1965
E-mail: jb.leite@unesp.br

Chapter 3: Multi-tier edge computing for IoT-based smart grids: Latency-aware data processing and resource management

Wilson Enrique Chumbi[1], Jonatas B. Leite[2].

[1]Electrical Engineering Department, Sao Paulo State University (UNESP), Ilha Solteira, SP, Brazil, enrique.quito@unesp.br, https://orcid.org/0000-0001-8347-0208.

[2]Electrical Engineering Department, Sao Paulo State University (UNESP), Ilha Solteira, SP, Brazil, jb.leite@unesp.br, https://orcid.org/0000-0002-1204-7178.

Abstract:

The modernization of power systems into Smart Grids (SGs) is driven by the need to manage increasing electrical loads and effectively integrate a multitude of smart devices. SGs leverage real-time data from widespread Internet of Things (IoT) devices to shift from traditional reactive planning to a proactive, "predict, prevent, and optimize" model. However, conventional centralized cloud computing introduces significant latency, a critical bottleneck for time-sensitive applications like grid automation and self-healing. This chapter addresses this challenge by presenting a latency-aware, multi-tier computational architecture tailored for SGs. We detail the transition from a centralized to a decentralized paradigm, deploying computational resources closer to the data source. The multi-tier architecture highlights the use of the Message Queuing Telemetry Transport (MQTT) protocol, with brokers deployed across cloud, fog, and edge layers. Resource management is achieved by grouping geographically the IoT devices and strategically placing these fog brokers. This placing of fog brokers optimizes the data routing, resulting in significant performance improvements for MQTT-based communication. A data-driven methodology based on geographic clustering is assessed to determine the effective number and location of fog brokers, aligning the network's computational resources with the physical topology of the grid. This chapter presents how a multi-tier MQTT architecture achieves a substantial latency reduction compared to the cloud model. Ultimately, this chapter provides a practical implementation strategy for designing scalable, low-latency infrastructures, bridging theoretical concepts with concrete implementation strategy for intelligent grid operations.

Keywords: Distribution network operation, internet of things, multi-tier edge computing architecture, self-healing, smart grid, transmission latency.

List of abbreviations

AI	Artificial intelligence.
CAGR	Compound annual growth rate.
DBSCAN	Density-based spatial clustering of applications with noise.
DER	Distributed energy resource.
DMS	Distribution management system.
EaaS	Energy-as-a-Service.
EV	Electric vehicle.
FLISR	Fault location, isolation, and service restoration.
ICT	Information and communication technology.

IED	Intelligent electronic device.
IoT	Internet of Things.
MQTT	Message queuing telemetry transport.
O&M	Operation and maintenance.
OMS	Outage management system.
PMU	Phasor measurement unit.
RTU	Remote terminal unit.
SCADA	Supervisory control and data acquisition.
SG	Smart grid.
VPP	Virtual power plant.

1. Introduction

The IoT is an information and communication technologies (ICTs) paradigm that expands internet connectivity beyond traditional computers and mobile devices to a vast array of everyday objects. The fundamental concept is to provide "connectivity for anything" by embedding transceivers into these items, enabling new forms of communication between people and things, and between things themselves. As IoT technologies mature, they are having a transformative impact on traditional industries by introducing unprecedented levels of automation and data-driven decision-making[1]. This technological shift is also driving an evolution in data models, moving from current data-centric systems to future knowledge-centric systems, powered by artificial intelligence (AI) and advanced computational resources, such as the fog and edge computing layers explored in this chapter[2].

1.1. Data management challenges and digital transformation in smart grids

Today, power grids face growing demands for reliability and security, and the integration of new energy sources. The SG vision implies the evolution of traditional power systems into highly instrumented, data-driven infrastructures[3]. Sensors, smart meters, phasor measurement units (PMUs), distributed energy resources (DERs), and IoT devices continuously generate large volumes of heterogeneous data that must be processed and acted upon, often in near real-time[4]. This transition brings multiple data-related challenges that must be resolved for the grid to realize its promised benefits including enhanced load balancing, self-healing capabilities, and operational resilience[5].

A wide range of ICTs are deployed across power plants, transmission lines and towers, distribution substations, and consumer units. The connectivity, automation, and tracking of such a large number of devices requires distributed monitoring, analysis, and control through high-speed, ubiquitous, and bidirectional digital communication[6]. This vision is increasingly being realized through the IoT paradigm, which establishes a network capable of connecting virtually any object to the Internet via standardized protocols for data exchange and communication among smart devices[7]. By integrating electricity-consuming data and grid-operating assets, such as smart meters, switches, and reclosers, into networks, utilities can enable remote operation, automated management, and cost-efficient control, thereby advancing toward more intelligent and adaptive grid operations.

In 2008, the number of devices connected to the Internet surpassed the global human population, marking a pivotal milestone in the evolution of digital connectivity. The IoT encompasses a network of physical objects equipped with embedded technologies that enable interaction with their internal states and external

environments[8]. Trought this network, these objects can sense, analyze, control, and make decisions either autonomously or collaboratively. Such behavior is essential for SG operations, positioning the SG as one of the most significant and impactful applications of IoT [9]. The standardization of IoT communication protocols creates new opportunities for innovation in future knowledge-centric energy systems.

In recent years, IoT technologies have matured into well-defined business models, accelerating the digital transformation of electric utilities[10]. When end users turn into prosumers, agents that both produce and consume energy, they represent a central source within the SG. In SG context, IoT plays a decisive role by enabling efficient energy use through device connectivity and real-time data collection[11]. The scale and effectiveness of these impacts depend on the technological maturity of each implementation stage, ranging from pilot stage to mature and social acceptance [12]. These technologies facilitate increasingly sophisticated business models across the energy value chain, as summarized in Table 1.

Emerging markets and developing economies could significantly benefit from SG deployments, to both upgrade their electricity network infrastructure and tackle numerous business opportunities[13]. IoT-enabled services show remarkable potential for economic growth and operational efficiency. For instance, the Virtual Power Plant (VPP) market is a rapidly expanding sector, projected to grow from USD 1.6 billion in 2022 to USD 6.9 billion by 2030, reflecting a compound annual growth rate (CAGR) of 19.3%. VPPs are particularly advantageous, enabling the integration of 10–20% more DERs compared to conventional methods[14]. Demand-side management has demonstrated 10–15% peak demand reduction in pilot projects[15]. The Energy-as-a-Service (EaaS) model is forecasted to reach USD 147.5 billion by 2030, with over 60% of commercial buildings in North America anticipated to adopt EaaS contracts[16]. Decentralized energy trading, enabled by blockchain and IoT, offers significant savings and new revenue for prosumers[17]. The global microgrid market is projected at USD 84.4 billion by 2032[18]. Furthermore, IoT-based predictive maintenance is transforming operational practices, reducing unplanned downtime by 30–50% and cutting operation and maintenance (O&M) costs by up to 25%[19][20].

Table 1. Emerging business models for SGs and IoT technologies categorized in utility-focused, consumer-focused, and third-party/cross-sectoral.

Category	Business model	Features	Value proposition	Maturity level	References
Utility-focused	Grid reliability & asset management	IoT sensors monitor grid assets, enabling predictive maintenance and performance optimization.	Reduce outages, extend asset lifespan, cut O&M costs.	Growing. Widely adopted by leading utilities, scaling globally	[21]
	Grid optimization & flexibility	AI + IoT balance supply/demand, manage storage, and coordinate VPPs.	Integrate renewables effectively, enhance grid stability.	Growing. Strong adoption in regions with high renewable penetration.	[22]
	Demand-side management & response	Smart meters + IoT engage customers in demand response programs.	Lower peak demand, reduce stress on grid, lower bills.	Mature. Established programs in U.S., EU, parts of Asia.	[23],[24]

	Data-as-a-service	Utilities monetize anonymized/aggregated IoT data.	New revenue streams, support urban planning, energy forecasting, and market analysis.	Emerging. Pilots in Europe and North America.	[6],[16]
	Microgrid-as-a-service	Localized energy systems that can operate independently (islanded) or connected to the main grid.	Enable revenue from local energy sales and ancillary services to the main grid.	Growing. Adoption expanding in industrial and community settings.	[25],[26]
Consumer-focused	Peer-to-peer (P2P) energy trading	Blockchain + IoT enable prosumers to trade excess energy locally.	Empower consumers, enable local renewable energy markets.	Pilot/Emerging. Active pilots in EU, Asia-Pacific.	[27]
	EaaS	Customers pay for package of services such as guaranteed energy savings or comfortable building environments, rather than raw energy units.	Low upfront costs, reliable energy savings.	Growing. Strong adoption in commercial/industrial sectors.	[28]
	Pay-per-usage pricing	IoT meters enable dynamic billing by actual usage.	Incentivize mindful consumption, reward off-peak usage.	Mature. Implemented widely in advanced smart metering markets.	[29],[30]
	Data-driven home energy management	IoT + external data (weather, prices) automate appliances.	Personalized efficiency, cost savings, carbon reduction.	Growing. Common in developed economies with smart home adoption.	[31],[32],
Third-party / Cross-sectoral	IoT expertise-as-a-service	Tech firms provide IoT deployment/management for utilities.	Fill skill gaps, accelerate IoT adoption.	Emerging. Early-stage adoption among utilities with limited IoT expertise.	[33],
	Predictive maintenance as a service (PMaaS)	AI + IoT analytics predict equipment failures for clients.	Prevent downtime, shift to proactive maintenance.	Emerging/Growing. Increasing traction in industrial and utility sectors.	[19],[20]

	VPP aggregation	Aggregators bundle DERs and sell flexibility into markets.	Monetize DERs, support wholesale markets.	Growing. Scaling rapidly, particularly in Europe and Australia.	[34],[14]
	E-mobility charging services	Smart technologies + IoT provides EV charging demand management.	Balance grid load, enable variable pricing, and support sustainable charging options.	Growing/Mature. Strong adoption in Europe and China; scaling in U.S.	[35],[36]
	Decentralized finance (DeFi) for energy	Blockchain, smart contracts, and IoT, platforms can create new financial instruments for energy trading and investment.	Increase transparency, automate transactions, streamline energy financing.	Pilot/Emerging. Limited pilots, early research and regulatory exploration.	[37],[38]

Beyond the SG domain, IoT drives innovation and across the energy sector. The market for IoT in energy data itself was valued at USD 24.6 billion in 2023 and is projected to reach USD 78.6 billion by 2030, highlighting the accelerating adoption of data-centric energy management solutions[39]. The home automation market is expected to grow to USD 21.6 billion by 2030, reducing household energy consumption by 10–20% through intelligent monitoring and control[40]. The rise of e-mobility, with an estimated 230 million electric vehicles (EVs) projected by 2030 requires IoT-based charging coordination to optimize grid stability and charging infrastructure utilization. IoT-enabled asset monitoring enhances grid reliability by 20–30% and enables fault detection processes that are up to 40% faster. Complementing these advancements, the global energy analytics market is expected to reach USD 12.3 billion by 2030, reflecting the increasing reliance on data analytics for predictive and adaptive energy management. The wide range of IoT-driven services and business opportunities for both utilities and private sector stakeholders is illustrated in Figure 1.

Digitalization via IoT presents opportunities to enhance energy system efficiency, reliability, and resilience by transforming vast data streams into actionable insights. However, this transformation is constrained by the “three V’s” of big data (volume, velocity, and variety): Volume (exponential data growth from PMUs and smart meters), velocity (continuous streams demanding instantaneous processing), and variety (heterogeneous formats). Beyond these dimensions, latency, the delay between data transmission and response, emerges as a critical factor. Many key SG functions, including fault detection, microgrid control, demand response, and stability monitoring, require real-time or near-real-time operation. When all raw data are transmitted to a centralized cloud, significant delays can arise from transmission distances and network congestion, potentially compromising automation effectiveness, grid stability and operational safety.

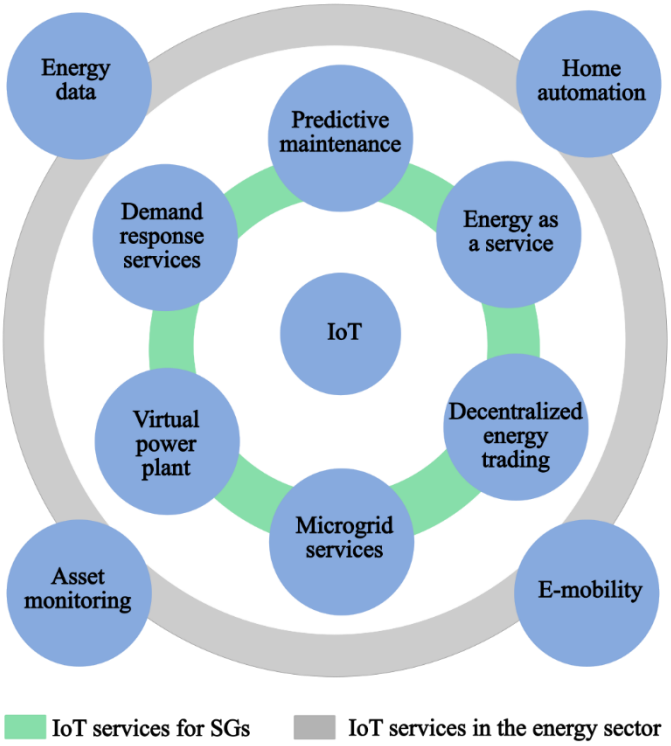


Figure 1. IoT applications serving as business strategies for utilities and the private sector in emerging energy markets.

1.2. The synergistic role of IoT nodes and edge computing in grid automation

Edge computing and the IoT are inherently complementary technologies. While IoT enables pervasive sensing, and data acquisition from millions of distributed devices, edge computing provides the computational infrastructure required to process this massive data flow with minimal latency. By relocating data processing and analytics to the network edge, closer to where IoT data are generated, edge computing helps address the limitations of traditional cloud-centric models. This synergy allows IoT nodes to not only collect and transmit information but also to execute local analytics, control decisions, and event responses autonomously. Consequently, the integration of IoT with edge computing transforms the grid into a responsive, self-optimizing system capable of supporting advanced functionalities like fault location, isolation, and service restoration (FLISR).

A distributed architecture is achieved through a multi-tier computational model, establishing a hierarchical structure for data processing across the edge, fog, and cloud layers. In this configuration, data generated by IoT edge devices first passes through an intermediate fog layer, composed of distributed nodes including gateways, routers, and local servers that perform urgent, geo-distributed processing. Meanwhile, non-urgent tasks, long-term analytics, and large-scale data management are delegated to the cloud layer, where high-performance computing resources provide centralized coordination and storage. By introducing the intermediate fog layer, the architecture tries to meet real-time operational demands by addressing latency constraints, intermittent connectivity, and bandwidth restrictions. This hierarchical structure forms the backbone of IoT–edge integration in SGs, ensuring that localized intelligence and global coordination operate seamlessly within a unified, adaptive computing framework.

Through local preprocessing and filtering, edge nodes can reduce redundant data before transmitting it upstream to alleviate bandwidth congestion. The distributed processing facilitates the integration of DERs, such as rooftop photovoltaic systems, battery storage units, and electric vehicle chargers, by allowing local optimization and control strategies at the grid's periphery. A major example of this capability is the FLISR system, which autonomously detects faults, isolates the affected segments, and reroutes power flows to restore service with minimal latency. Successful functioning depends on interoperability: heterogeneous devices and platforms trying to communicate across the grid. Achieving this requires the harmonization of communication and data standards, such as integrating the IEC 61970/61968 standards used in control centers with the IEC 61850 standard applied in substation and feeder automation [41]. Establishing a common information model, IoT nodes at the edge can seamlessly integrate and exchange information with the centralized distribution management system (DMS), as illustrated in Figure 2.

The IoT device integration architecture illustrated in Figure 2 is structured across three hierarchical layers, each serving a distinct functional role within the smart grid ecosystem. At the top layer, the utility datacenter functions as the centralized control center, responsible for high-level operations such as system planning, maintenance scheduling, and supervisory decision-making. At the bottom layer, the physical grid layer encompasses field-level equipment, including traditional IEDs and Remote Terminal Units (RTUs), complemented by modern IoT Nodes that extend connectivity and sensing capabilities throughout the network. Bridging these two layers is the middleware layer, which provides the essential communication and coordination infrastructure. This layer hosts network operation services such as Supervisory Control and Data Acquisition (SCADA), the Outage Management System (OMS), and the DMS, ensuring seamless information flow between field devices and the control center.

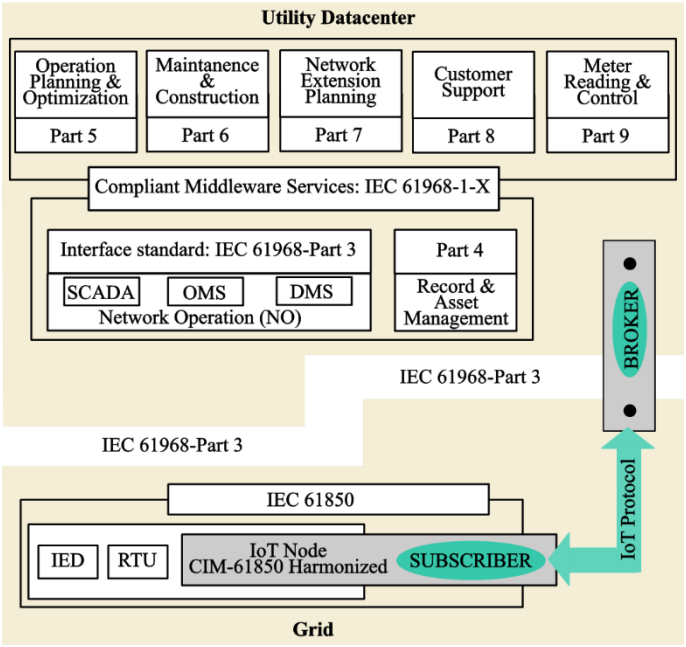


Figure 2. Harmonized IoT node device in the distribution system.

Within this three-tier framework, two parallel communication pathways coexist. The traditional communication path relies on the **IEC 61850** standard, enabling IEDs and RTUs to exchange data with SCADA and DMS systems. In contrast, the modern communication path serves IoT nodes, which operate

using a lightweight "IoT protocol" optimized for high-frequency, low-latency data exchange. In this configuration, each IoT node acts as a publisher/subscriber, transmitting data to a central broker. The broker interfaces with the utility's main communication backbone, standardized under **IEC 61968**, which governs information exchange for distribution management applications. This dual-path architecture allows IoT data to be natively understood by central systems, enabling faster, automated grid responses.

2. Transition from centralized to distributed computing

2.1. Limitations of cloud-based models for smart grids

The main limitations of cloud, fog, and edge computing are primarily determined by the location of data processing[42]. Its reliance on a stable Internet connection introduces reliability issues, and its centralized storage creates a single point of failure for security. Fog computing mitigates some of these issues by introducing intermediate processing nodes. However, it also requires the deployment of additional infrastructure, increasing system complexity and cost. Fog computing can aggregate and coordinate data from multiple edge nodes performing regional analytics, event correlation, and resource orchestration[43]. Edge computing limitations are related to local resource constraints. While it provides ultra-low latency, its performance is limited by the processing power, memory, and energy consumption of individual devices. This makes it ideal for immediate, lightweight tasks like event detection and data filtering, rather than computationally intensive analytics, which must be sent to the fog or cloud. Table 2 summarizes the limitations and capabilities of cloud, fog, and edge computing architectures, highlighting the implications of distributed computing models in SG applications.

Table 2. Limitations of cloud, fog, and edge systems and the implications for the SG.

Dimension	Cloud Computing	Fog Computing	Edge Computing	Implications for IoT-enabled SG
Latency	High latency, servers are often distant from end devices.	Moderate latency, processing closer to devices via fog nodes.	Ultra-low latency, processing at or near data source.	Delays in data transmission hinder real-time control and automation.
Bandwidth Usage	High: all raw IoT data transmitted to central server.	Medium: partial preprocessing reduces data sent to the cloud.	Low: local processing minimizes network traffic.	Network congestion and higher communication costs, especially in large-scale deployments.
Scalability	Performance bottlenecks arise as billions of IoT devices connect to the cloud simultaneously.	More scalable in distributed deployments but requires additional infrastructure.	Performance may be constrained by device capacity, but it scales with the number of local devices.	Limits the ability to support massive IoT integration.
Reliability	Dependence on stable, high-speed internet connections.	More reliable, as processing is distributed across regional fog nodes.	Highly reliable for critical local operations, even if disconnected from network.	Outages or disruptions in connectivity can interrupt grid monitoring and control.

Security & Privacy	Centralized storage creates single points of failure and vulnerability to cyberattacks.	Improved: data partially processed locally, reducing exposure.	Strongest: sensitive data can remain within the device or local network.	Increased risk of data breaches, unauthorized access, and loss of consumer trust.
Energy Consumption	Data transmission and large data center operations are energy intensive.	Moderate: shared fog resources reduce transmission and cloud load.	Low: less transmission, though device-side processing can increase local energy use.	Raises operational costs and increases carbon footprint, conflicting with sustainability goals.

2.2. MQTT-based multi-tier architecture

The MQTT protocol is a lightweight communication standard widely adopted in IoT-based smart grid architectures to enable efficient, reliable, and scalable data exchange between distributed devices and central utility systems. It relies on a central intermediary, the broker, which is a server acting as a communications hub that decouples message senders (publishers) from receivers (subscribers) [44]. The data flow begins when client devices, such as operator workstations or field crew mobile units, subscribe to specific message topics. An IoT node (publisher) then transmits messages to the broker on those topics. The broker immediately forwards the data to all subscribed clients, ensuring that both the operator and field crew receive the necessary information without direct communication with the IoT node itself. Publishers and subscribers don't need to know each other's network addresses, which makes the system incredibly flexible; devices can be added or removed without reconfiguring the entire network. In addition to scalability, MQTT offers robust reliability through its persistent messaging mechanism. This feature allows the broker to retain the last message for each topic, ensuring that any newly connected subscriber immediately receives the most recent data update, even if the publisher is temporarily offline [45]. The overall message flow and interaction between publishers, subscribers, and the broker are illustrated in Figure 3.

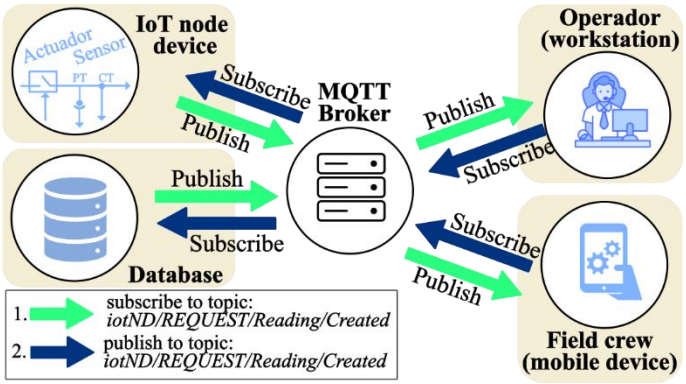


Figure 3. Message flow architecture of the MQTT publish–subscribe protocol for SG communication.

In this framework, the MQTT protocol defines the communication rules and message exchange mechanisms, while the multi-tier architecture provides a better structure for that communication to take place [41]. Within multi-tier architecture, the components of the MQTT protocol are strategically

distributed across three layers. Layer 1 (edge layer), where IoT node devices execute MQTT client instances to publish and subscribe to operational data in real time. Layer 2 (fog layer), comprises remote brokers that assist the main broker by performing localized data aggregation, preprocessing, and filtering for time-sensitive applications. Layer 3 (cloud layer), hosts the central MQTT broker, which functions as the primary communication hub, coordinating data management, analytics, and long-term storage within the utility's core infrastructure. This three-tier MQTT-based architecture is illustrated in Figure 4.

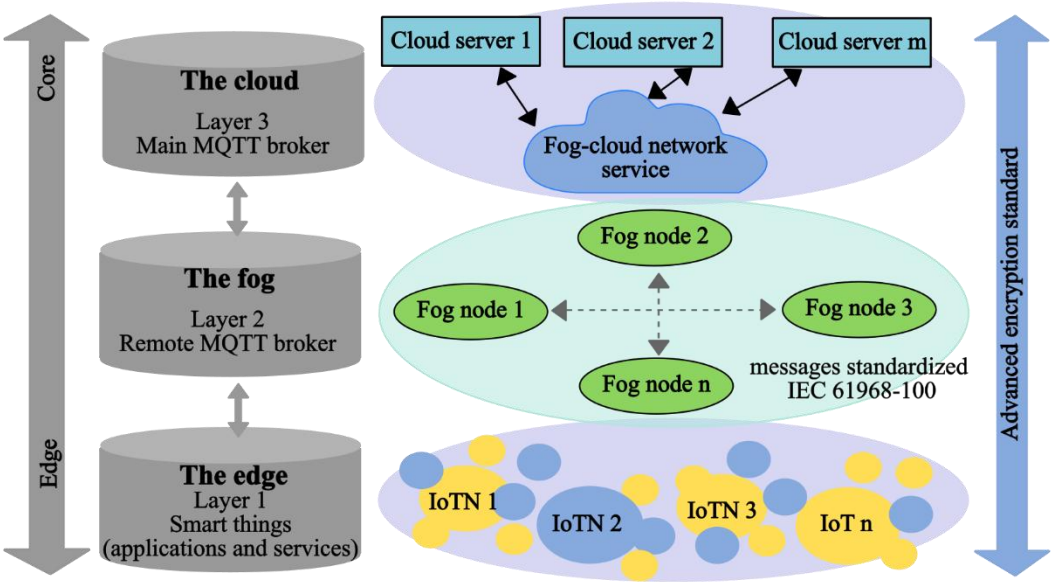


Figure 4. Three-layer architecture integrating cloud, fog and edge computing technologies.

The architecture is composed with end-to-end encryption applied to the data payload between the end nodes and the application server, thereby ensuring security across the entire communication chain. Cloud is typically left to focus on aggregating data and turning the data into knowledge. Note that the further from the edge to the core, the greater the knowledge that can be obtained. The fog computing architecture is flexible, allowing its structure to be tailored to the specific requirements of each application domain [46]. In most implementations, fog deployments consist of N tiers of nodes, each responsible for managing data processing and coordination within distinct operational layers. Each fog element can represent a cluster of interconnected fog nodes that collaboratively execute common functions, enhancing resilience, scalability, and fault tolerance. In simpler configurations, certain fog and cloud elements may be consolidated into a single physical deployment, optimizing system design for smaller-scale or resource-constrained environments. Conversely, in more dynamic and distributed scenarios, such as vehicle-to-vehicle (V2V) communication for traffic coordination or EV networks managing charging schedules, each fog element may form a mesh of peer-to-peer fog nodes [47].

2.3. Transmission latency in multi-tier architectures

Latency refers to the total time required for a signal or data packet to travel from one point to another and return to its origin. Transmission latency, denoted as λ_t, serves as a key performance indicator for assessing system responsiveness. Within a multi-tier IoT architecture, several transmission stages occur as data propagate between different endpoints across the edge, fog, and cloud layers. For each event generated by an IoT node δ, triggered by a sensor and/or actuator, the node initiates a notification process and transmits a request to the remote broker φ located in the fog layer. The broker φ processes the request and, when

necessary, forwards it to the main broker Φ in the cloud layer for further analysis or decision-making. The corresponding response message then follows the reverse route. The total transmission latency is shown in (1), where the communication distance between two endpoints is denoted by l.

$$\overline{\lambda_{t_f}} = \sum \left(l_{\delta}^{\varphi} + l_{\varphi}^{\Phi} + l_{\Phi}^{\varphi} + l_{\varphi}^{\delta} \right) \tag{1}$$

In the cloud architecture, the transmission path is simplified, with data packets being sent directly from each IoT node δ to the main broker Φ. The latency expression is presented in (2).

$$\overline{\lambda_{t_c}} = \sum \left(l_{\delta}^{\Phi} + l_{\Phi}^{\delta} \right) \tag{2}$$

Considering the potential delays introduced by long transmission distances and environmental losses, certain data packets may require retransmission to ensure successful delivery. This retransmission factor is denoted by τ. Assuming that τ represents 1.5% of the total transmission time, the average transmission latency per IoT node can be expressed as shown in (3).

$$\lambda_t = \frac{\overline{\lambda_t} + \tau}{n} \tag{3}$$

Let n denote the total number of IoT nodes (or buses) generating transmission requests. The equation (3) calculates the overall transmission latency, adjusted by the retransmission factor τ, and then normalizes it by dividing the result by the number of requests n.

3. Broker deployment analysis using clustering and latency evaluation

This section details the methodology used to compare the transmission latency of a cloud-based model against the multi-tier IoT architecture. The core of this approach lies in the strategic deployment of fog brokers, guided by geographical clustering of network buses to optimize data flow. The analysis begins with a geographic clustering approach using the K-means algorithm, combined with the bisection method, to determine the appropriate number of clusters. Subsequently, a density-based spatial clustering of applications with noise technique (DBSCAN), is introduced as an alternative for identifying non-spherical clusters and effectively managing outliers in irregular network topologies. Finally, the K-means algorithm is extended into a multi-dimensional spatial-electrical clustering, incorporating network impedance and load power.

3.1. Irregular network data clustering for fog broker placement

A distribution network is composed of numerous buses and branches, each positioned at a defined distance from a reference point, or root bus, which serves as the central communication node. As the network expands, the average distance between buses and the reference point increases, leading to longer communication paths. To mitigate this effect, bus clustering is employed to group buses according to their spatial proximity, thereby reducing average communication distances[48]. Clustering algorithms partition georeferenced buses into clusters centered around centroids. Each centroid can then be associated with a fog broker, minimizing intra-cluster communication distances.

The K-means clustering algorithm aims to determine an optimal configuration C, as defined in (4), by partitioning a dataset into k non-overlapping clusters. Let $X = \{x_i \mid i = 1,2,\dots,n\}$ denote an n-dimensional dataset, where each x_i represents a data sample, and k is the predefined number of clusters. Each cluster C_j contains a subset of samples such that the sum of the Euclidean distances between each

sample and its corresponding cluster centroid is minimized. The algorithm follows an iterative process that begins by randomly selecting the centroid of each cluster, denoted as m_j. Each sample x_i is then assigned to the nearest centroid by computing the squared Euclidean distance, as expressed in (5).

$$C = \bigcup_{j=1}^{k} C_j \tag{4}$$

$$C^0 = \min_{C} \sum_{j \in k} \sum_{m \in C_j} d_2(x_m, m_j)^2 \tag{5}$$

In each iteration t, the reassignment of samples to their respective centroids is performed according to equation (6), and the centroid positions are updated based on (7) until the convergence criterion is met. At each iteration, the algorithm aims to improve the quality of the clustering.

$$C_j^t = \{x_i | d_2(x_i, m_j^t) \leq d_2(x_i, m_{j^*}^t), \forall j^* = 1,2, \dots, k\} \tag{6}$$

$$m_j^t = \frac{1}{|C_j^t|} \sum_{x_j \epsilon C_j^t} x_j \tag{7}$$

When applying the K-means algorithm, it is necessary to determine the minimum and maximum values along the X and Y axes ($X_{\max}, X_{\min}, Y_{\max}, Y_{\min}$) to establish the spatial boundaries of the dataset. These parameters are used to define the initial positions of the k centroids according to equations (8) and (9). Alternatively, the centroids can be randomly initialized at existing bus locations within the distribution network.

$$m_{X,j}^0 = X_{min} + \frac{X_{max} - X_{min}}{k+1} * j, for\ j = 1,2, \dots, k \tag{8}$$

$$m_{Y,j}^0 = X_{min} + \frac{Y_{max} - Y_{min}}{k+1} * j, for\ j = 1,2, \dots, k \tag{9}$$

The data clustering is illustrated in Figure 5, which shows three clusters (C_1, C_2, C_3) each represented by a distinct color and associated with its corresponding centroid (m_j). Within each cluster, the samples (x_i) are grouped according to their spatial proximity to the centroid. The intra-cluster distance quantifies the compactness of data points within a cluster, while the inter-cluster distance measures the separation between clusters.

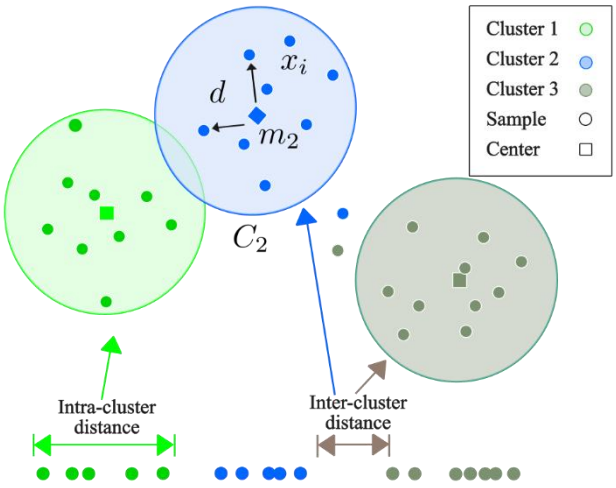


Figure 5. Three clusters (C_1, C_2, C_3) and their centroids (m_j) based on k-means algorithm.

The centroid location can be further refined by incorporating two additional components related to network distance and load density, as shown in (10).

$$d_2(x_m, m_j)^2 = \sqrt{(X_{xm} - X_{mj})^2 + (Y_{xm} - Y_{mj})^2 + (Z_{xm} - Z_{mj})^2 + (P_{xm} - P_{mj})^2} \quad (10)$$

where Z represents the network impedance and P denotes the load power. The expression accounts not only for geometric separation but also for both topological and operational proximity[49].

While the K-means algorithm partitions data into compact and well-separated clusters based on distance minimization, it assumes that clusters are spherical and of similar size. However, in real-world applications, where data distribution may be irregular, these assumptions may not always hold. To address this limitation, density-based clustering methods can be employed as a complementary approach.

Density-based clustering algorithms, such as DBSCAN, OPTICS (Ordering Points to Identify Clustering Structure), and DENCLUE (DENsity CLUstEring), are designed to identify clusters by analyzing the spatial density distribution of data points. Among these methods, the DBSCAN algorithm was chosen to perform spatial grouping based on its ability to detect both high- and low-density regions with noise[49]. The main characteristics of DBSCAN include:

- No need to predefine the number of clusters, as the algorithm determines the grouping dynamically based on the specified density threshold.
- Ability to detect clusters of arbitrary shapes, which is particularly valuable for geographically referenced data affected by network topology.
- Robustness to outliers, as noise points are easily identified and excluded from clusters.
- Parameter flexibility, allowing users with domain knowledge of network characteristics to fine-tune the algorithm and improve clustering accuracy.

The DBSCAN algorithm operates based on two key parameters: the neighborhood radius (ε) and the minimum number of points (MinPts) required for a region to be considered dense. A point is defined as a core point if its ε-neighborhood contains at least MinPts data points. Conversely, a border point has fewer than MinPts neighbors but lies within the ε-neighborhood of a core point, thereby remaining part of a cluster. Points that are neither core nor border are classified as outliers, representing isolated or noise elements within the dataset. Figure 6 illustrates the DBSCAN clustering classification of data points according to their neighborhood density. In the figure, each dashed circle represents the ε-neighborhood of a sample point, with $\varepsilon = 1$ and a minimum number of points MinPts $= 5$.

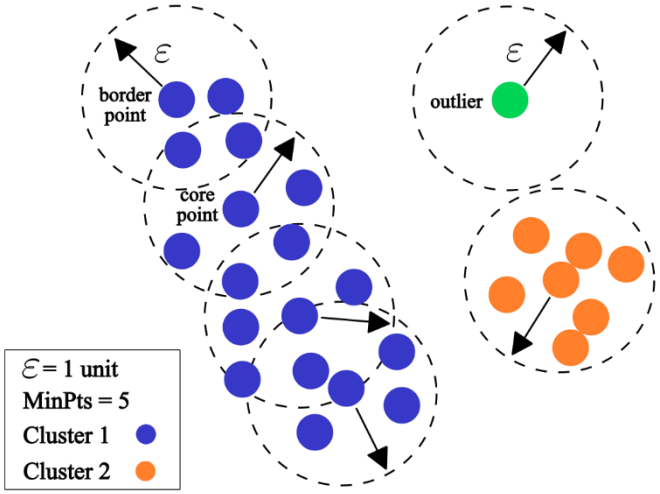


Figure 6. DBSCAN clustering with neighborhood radius ($\varepsilon = 1$) and minimum points (MinPts $= 5$) to form a cluster.

Each centroid represents the ideal position for a broker. However, practical deployment sites within the electrical network must consider several physical and operational constraints, including the availability of existing communication assets, power supply, environmental conditions, and maintenance accessibility. The deployment of fog computing nodes typically aligns with three distinct types of infrastructure. (1) substations, which provide secure and well-connected environments for hosting brokers; (2) pole-mounted gateways, offering cost-effective and decentralized solutions for rural or feeder-level applications; and (3) regional data centers, serving as high-capacity intermediaries between local fog nodes and the central cloud. Thus, the mapping process can be improved by correlating the centroid location with a site suitability analysis. This correlation enables the selection of deployment sites that minimize the distance between infrastructure availability, and topological accessibility, which may differ from the linear geometric distance[50].

3.2. Clustering evaluation and latency outcomes

When applying the K-means algorithm to cluster the network buses, it is essential to determine the optimal number of clusters (k). This optimization problem can be solved using the bisection method, a numerical technique grounded in the intermediate value theorem. The method iteratively partitions a defined interval $[a, b]$ into subintervals $[a, c]$ and $[c, b]$, successively refining the search until a stopping criterion is satisfied.

The bisection method is applied to evaluate possible values for the number of centroids within the interval $[a, b]$, where $a = 2$ represents the minimum possible subdivision of the distribution network, and $b = n$ represents the situation in which each bus acts as its own centroid. The stopping criterion is defined as the minimization of the total distance, calculated as the sum of distances between each bus and its cluster centroid, plus the distance between each centroid and the root bus.

By combining the K-means clustering algorithm with the bisection method, it is possible to determine both the ideal number of clusters and the set of distances between buses, centroids, and the root bus, as summarized in Algorithm 1. Once the optimal clustering configuration is defined, latency values are computed for both cloud and fog-based architectures, according to (1) and (2), respectively. The average

latency for each architecture is calculated using (3), to enable a comparative assessment of their performance.

Algorithm 1: Clustering k-means and bisection algorithm

Set $a = 2$ and $b = n$;
Solve k-means(a) and k-means(b); *# split the network into a and b clusters*
Compute $D_t(a)$, $D_t(b)$ and $D = D_t(a) - D_t(b)$;
While $D > 0$;
    $c \leftarrow (a + b)/2$; *# new midpoint*
    Solve k-means(c); *# split the network into c clusters*
    If $D_t(a) < D_t(b)$ **then**:
        Compute $D_t(c)$ and $D = D_t(a) - D_t(c)$; *# total distances*
        $k \leftarrow a$;
        $b \leftarrow c$;
    Else:
        Compute $D_t(c)$ and $D = D_t(b) - D_t(c)$; *# total distances*
        $k \leftarrow b$;
        $a \leftarrow c$;
    End if
End while
Return k; *# optimal number of clusters*

In the comparative analysis of latency between the cloud and fog-based architectures, a three-phase unbalanced 13.8 kV distribution network comprising 136 buses was employed. The topology of the distribution network, is illustrated in Figure 7 (a). The system must be geographically referenced to clustering bars and determine the distances between critical components, including the remote brokers, the main broker, and the sensors or client devices. Additionally, the position of the root bus was established on the bus one where corresponding to the location of the system's main broker.

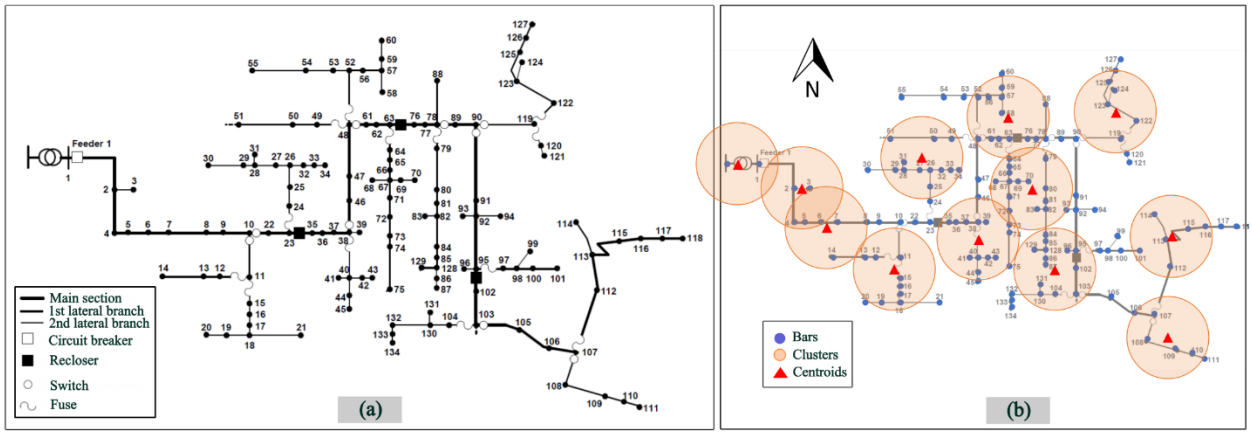


Figure 7. Multi-tier IoT architecture applied to the 136-bus test system: (a) network topology; (b) clustered bus distribution.

Latency for the cloud-based architecture was computed using the root bus position and network bus coordinates. The corresponding total transmission distances are summarized in Table 3. For latency computation, a propagation delay of 4 μs/km was adopted, consistent with typical values for terrestrial radio communication systems (ITU-T G.114, 2003).

For fog-based architecture case using Algorithm 1, the optimal number of clusters was determined as $k = 12$. The Figure 7 (b) illustrates the spatial distribution of the buses and the centroids, where remote brokers are assumed to allocated at the centroid positions. From these coordinates, the distances between each bus and its respective cluster centroid, as well as between each centroid and the root bus, were calculated and the corresponding distance sums are shown in Figure 8.



Cluster	Bars to centroid	Centroid to root bar	Total
0	59	19.67	78.67
1	28	143.89	171.89
2	266.43	214.94	481.37
3	797.33	359.63	1156.96
4	988.17	355.32	1343.49
5	872.19	478.18	1350.37
6	965.8	518.8	1484.6
7	1094.73	556.99	1651.72
8	1175.79	627.16	1802.95
9	640.16	713.39	1353.55
10	683.8	820.1	1503.9
11	316.53	862.26	1178.79

Figure 8. Clustering of the 136-bar test system and summation of distances.

The fog-based latency was computed considering the round-trip path and the same propagation delay. Subsequently, the average fog-based latency is presented in Table 3. A comparison between the latency values reveals that the fog-based architecture achieves an approximately 81% reduction in latency relative to the cloud-based architecture.

Table 3. Total transmission distances and latency comparison between cloud-based and fog-based architectures.

Model	Cloud-based model	Fog-based model
Sum of routes (km)	71.21	13.56
Sum of Latencies (ms)	0.57	0.11
Average latency (µs)	4.25	0.81

Unlike K-means, DBSCAN algorithm does not require the number of clusters to be specified in advance. However, its performance is highly sensitive to its two key parameters: the neighborhood radius (ε) and the minimum number of points ($MinPts$) required to form a dense region. The $MinPts$ parameter can often be specified using domain knowledge. A common practice is to set this value between four and two times the number of dimensions in the dataset $MinPts \in [4, 2 \cdot dim]$. The number of dimensions refers to the features or attributes used to describe each individual data point for the clustering analysis. The neighborhood radius is more critical and can be determined methodically using a k-distance graph. This is generated by calculating the distance of each point to its k-th nearest neighbor (where $k = MinPts - 1$), and then plotting these distances in ascending order. The "elbow" or point of maximum curvature in this plot suggests an optimal ε value. Points with a k-distance greater than ε are likely to be considered outliers.

A sensitivity analysis was performed to examine the interaction between the DBSCAN parameters and the clustering performance. The analysis employed a test system comprising 251 electrical loads, with an

average energy demand of 217.8 MWh, distributed along a three-phase feeder rated at 22 kV and extending over 102.90 km. The dataset is geographically referenced, allowing each data point used in the clustering analysis to be represented in a three-dimensional space, defined by its x-coordinate, y-coordinate, and power electric load. The sensitive analysis is presented in the Table 4. The results indicate that increasing $MinPts$ reduces the number of identified clusters while increasing the number of points classified as outliers. Similarly, as ε increases, the number of clusters generally decreases, but the number of outliers also declines due to the formation of larger neighborhoods. The silhouette score improves as the number of clusters decreases, indicating stronger intra-cluster cohesion and greater inter-cluster separation. A score approaching +1 denotes compact and well-separated clusters, reflecting better clustering quality.

Table 4. DBSCAN parameter tuning outcomes with different ε and $MinPts$.

Scenario	ε	$MinPts$	Number of clusters	Outliers	Silhouette Score	Execution Time (s)
1	0.30	4	10	21	0.217	0.19
	0.35	4	9	12	0.288	0.18
	0.40	4	6	9	0.485	0.16
	0.45	4	5	9	0.538	0.15
2	0.30	5	7	33	0.328	0.25
	0.35	5	6	25	0.277	0.24
	0.40	5	4	19	0.490	0.22
	0.45	5	4	14	0.524	0.22
3	0.30	6	7	38	0.359	0.16
	0.35	6	6	29	0.272	0.16
	0.40	6	4	23	0.482	0.14
	0.45	6	4	14	0.524	0.14

A comparative visualization reinforcing the findings of the sensitivity analysis under varying parameter settings is presented in Figure 9. The figure highlights the influence of parameter selection on the number and spatial distribution of clusters. It provides a direct visual comparison of the clustering results, corresponding to 7, 6, and 4 clusters, obtained from the DBSCAN and K-means algorithms when applied to the same dataset of electrical loads. In the DBSCAN results, the clusters follow the natural, irregular, and elongated shapes of the spatial data. The algorithm effectively identifies points that do not belong to any dense cluster (outliers), marking them as black dots labeled "Cluster −1." The three clustering configurations (7, 6, and 4 clusters) were not preassigned but emerged from the tuning of the parameters ε and $MinPts$, which directly affect cluster density and the number of outliers. In contrast, the K-means algorithm produces compact, convex partitions with a fixed number of clusters determined a priori by the parameter k. For comparative purposes, the number of clusters (7, 6, and 4) was explicitly set to match the DBSCAN results. K-means visibly divides the dataset into spherical groups. This behavior is particularly evident in the bottom plot, where the algorithm splits a large, dense group on the right into two separate clusters (green and red) while merging two naturally separate clusters into a single one (purple). Every data point is assigned to the nearest cluster centroid, so no outliers are identified. The small squares in the K-means plots indicate the cluster centroids.

The K-means algorithm's strength lies in its use of distance metrics to identify how data points are interrelated. It leverages a capability known as vector quantization to calculate the similarity between data points, even in multidimensional spaces. It is achieved by defining an appropriate distance function tailored to the specific characteristics of the problem under study. Equation (10) exemplifies this flexibility. Getting

back to the case study depicted in Figure 9, the electrical loads were grouped using a custom multidimensional distance function that incorporates the *x*- and *y*-coordinates, load power, and electrical distance between the loads. Consequently, the clustering process scales from a simple geographic clustering to a more advanced spatial-electrical analysis, achieved by integrating four distinct dimensions into a single distance calculation. This multidimensional approach enables the simultaneous clustering of electrical loads and partitioning of the physical network, as well as the cluster centroid's position on the electrical grid.

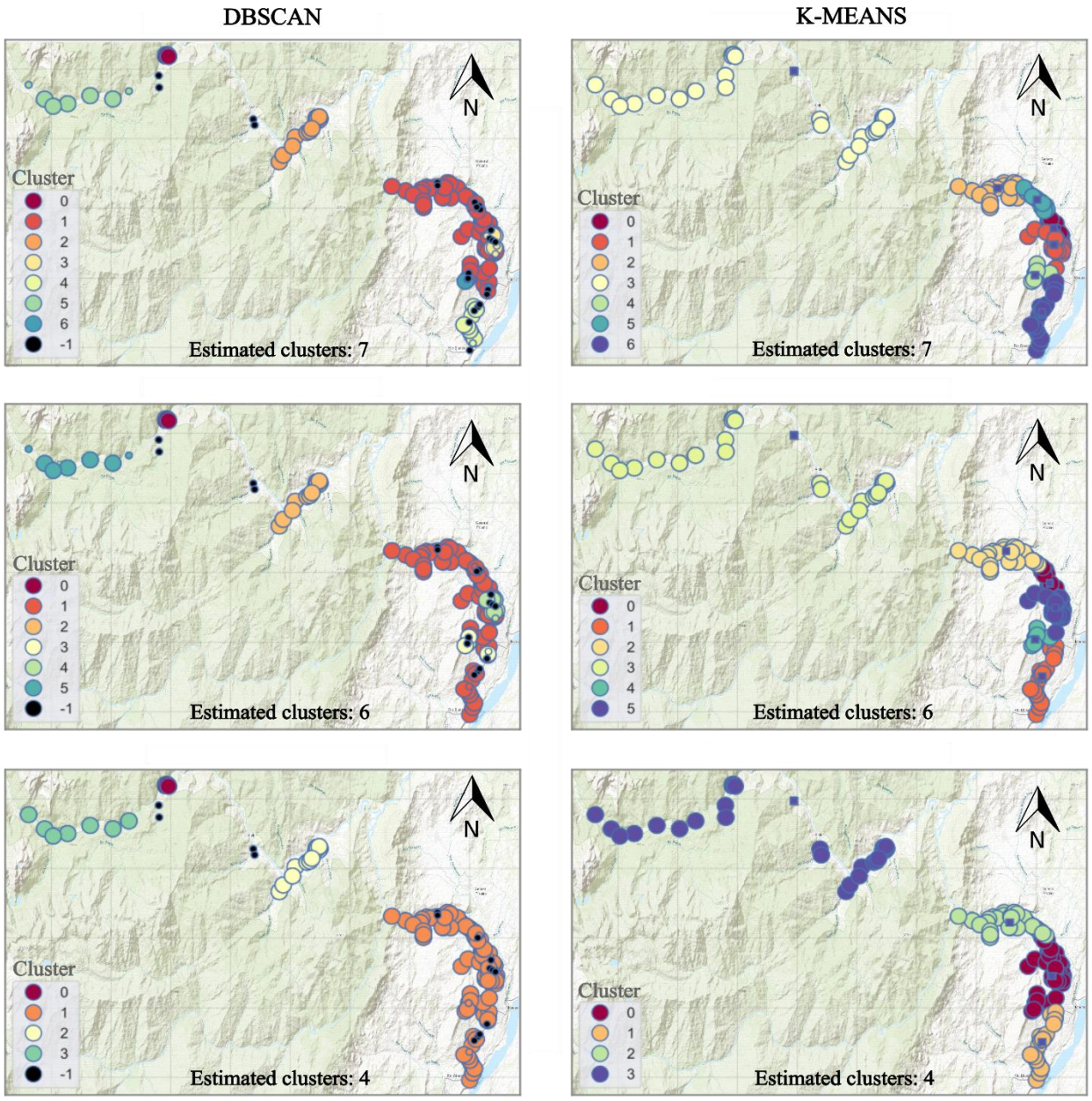


Figure 9. Comparison of the clustering results from the DBSCAN and K-means algorithms.

The Davies–Bouldin Index, Calinski–Harabasz Index, and Silhouette Score are all metrics used to evaluate the quality of a clustering algorithm's results. These metrics collectively address a fundamental question: *Are the resulting clusters dense, compact, and well-separated?* Lower Davies–Bouldin values indicate

better cluster separation and compactness, whereas higher Calinski–Harabasz and Silhouette scores signify improved cluster cohesion and distinctiveness. Table 5 presents the sensitivity analysis of K-means algorithm using these three-performance metrics, together with the execution time, for different values of *k,* but now extended to the four-dimensional case. The results show that *k = 4* provides the most favorable clustering configuration, characterized by the lowest Davies–Bouldin Index (0.5129) and a high Silhouette Score (0.531) at moderate computational time.

Table 5. Performance metrics for K-means clustering under varying numbers of clusters (k).

Metric	$\boldsymbol{k = 7}$	$\boldsymbol{k = 6}$	$\boldsymbol{k = 5}$	$\boldsymbol{k = 4}$	$\boldsymbol{k = 3}$
Davies–Bouldin Index	0.7099	0.5792	0.5427	0.5129	0.5707
Calinski–Harabasz Index	327.3976	388.1579	453.623	562.927	596.445
Silhouette Score	0.417	0.486	0.477	0.531	0.506
Execution Time (s)	92.508	86.501	63.368	74.054	32.989

4. Conclusions and future directions

This section summarizes the research findings and outlines the roadmap for future implementations.

4.1. General conclusions

When evaluating K-means and DBSCAN for smart grid applications, several trade-offs become evident. DBSCAN is particularly effective in identifying clusters that follow the natural spatial distribution of network assets, regardless of their shape. It is also robust in identifying outliers, which can represent isolated loads. In contrast, K-means partitions data into spherical clusters, which may not always align with the irregular geographic topology of a distribution network. However, K-means can be easily scaled to incorporate not only geographic coordinates but also electrical distance (impedance) and load characteristics. This capability allows the simultaneous clustering of loads and partitioning of the electrical network, useful feature for microgrid planning that DBSCAN does not inherently offer. While DBSCAN is sensitive to parameter selection, K-means can be computationally more expensive, with its runtime complexity related to the number of iterations, clusters, points, and dimensions. The performance of K-means is best when evaluated using multiple metrics to ensure a compact clustering solution.

The strategic resource management methodology, utilizing K-means clustering optimized via the bisection method, proved highly effective in aligning computational assets with the grid's physical topology. This systematic approach to fog broker placement yielded a definitive performance success: the fog-based architecture achieved a substantial 81% reduction in average transmission latency (from 4.25 μs to 0.81 μs) when tested on the 136-bus distribution network. This quantifiable performance gain validates the potential of edge and fog computing to facilitate the next generation of automated, self-healing grid operations.

A critical advancement of this methodology is its move beyond simple geometric clustering. The adoption of a multi-dimensional distance metric, which integrates geographic coordinates (X, Y) with operational parameters like electrical impedance (Z) and load characteristics (P), established a robust spatial-electrical correlation for resource placement. This nuanced algorithmic approach is fundamental for developing advanced distribution automation functions, as it enables the simultaneous optimization of communication efficiency and electrical network partitioning, supporting critical functions such as microgrid formation and coordinated local control. The complementary use of DBSCAN further enhances analysis by identifying

irregular cluster shapes and crucial outliers that K-means overlooks, providing richer context for network planning.

Building upon these methodological insights, this research validates that the modernization of power systems into resilient, high-performance SGs mandates a paradigm shift away from centralized cloud processing toward a distributed, multi-tier computational framework. This work presented and quantified the benefits of a latency-aware architecture spanning the edge, fog, and cloud layers, integrated via the scalable and decoupled MQTT protocol. The deployment of computational resources closer to the data source is a necessity for achieving the velocity requirements of time-critical grid applications.

4.2. Future research challenges

To transition the validated model from assessment to sustained operational deployment, future research must focus on addressing dynamic challenges and practical engineering constraints.

- Dynamic resource management: The current static optimal placement determined by clustering must evolve into a dynamic system. Future efforts should focus on developing algorithms that enable real-time adaptation of fog broker capacity, computational allocation, and cluster boundaries in response to shifting grid conditions, volatile DER integration, and changes in data traffic load, ensuring low-latency performance is maintained across all operational scenarios;
- Integration of knowledge-centric AI: The low-latency, multi-tier infrastructure is the necessary prerequisite for deploying advanced AI and machine learning models. Future research must leverage this responsiveness to deploy AI directly at the edge and fog layers, enabling proactive decision-making for future knowledge-centric SGs, particularly for tasks essential to the "predict, prevent, and optimize" mandate, such as advanced self-healing and predictive asset maintenance;
- Cybersecurity and resilience hardening: The distributed nature of architecture necessitates robust security solutions. A critical future direction is the detailed investigation and implementation of distributed trust mechanisms and enhanced security protocols, including end-to-end encryption, to protect the network of MQTT brokers and IoT nodes against cyber threats targeting essential grid infrastructure;
- Legacy system integration and cost analysis: Practical utility deployment requires solving the engineering challenge of seamlessly integrating the modern IoT/MQTT data path with existing, established SCADA and DMS. Future work must provide detailed practical solutions for harmonizing the IoT data models with industry standards (IEC 61850/61968) and deliver comprehensive economic analyses detailing operational costs, energy consumption, and capital expenditure to guide utilities in adopting this resilient computational framework.

5. References

[1] N. 691405 Grant agreement, “The SmartNet Project,” 2020. Accessed: Oct. 30, 2025. [Online]. Available: https://smartnet-project.eu/the-project/index.html

[2] M. M. Alam, M. J. Hossain, M. A. Habib, M. Y. Arafat, and M. A. Hannan, “Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions,” *Renewable and Sustainable Energy Reviews*, vol. 211, p. 115251, Apr. 2025, doi: 10.1016/j.rser.2024.115251.

[3] S. Hafez *et al.*, “Advancements in grid resilience: Recent innovations in AI-driven solutions,” *Results in Engineering*, vol. 26, p. 105042, Jun. 2025, doi: 10.1016/j.rineng.2025.105042.

[4] A. R. Al-Ali, R. Gupta, I. Zualkernan, and S. K. Das, “Role of IoT technologies in big data management systems: A review and Smart Grid case study,” *Pervasive Mob Comput*, vol. 100, p. 101905, May 2024, doi: 10.1016/j.pmcj.2024.101905.

[5] H. P. Bhupathi, V. K. Pandraka, and L. S. Malladi, “Internet of Things (IoT) in Smart Grid Technology and its Applications,” in *2024 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT)*, IEEE, Nov. 2024, pp. 91–96. doi: 10.1109/ICISSGT58904.2024.00028.

[6] B. P. Bhattarai *et al.*, “Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions,” *IET Smart Grid*, vol. 2, no. 2, pp. 141–154, Jun. 2019, doi: 10.1049/iet-stg.2018.0261.

[7] M. A. Alomar, “An IOT based smart grid system for advanced cooperative transmission and communication,” *Physical Communication*, vol. 58, p. 102069, Jun. 2023, doi: 10.1016/j.phycom.2023.102069.

[8] M. Mahmood, P. Chowdhury, R. Yeassin, M. Hasan, T. Ahmad, and N.-U.-R. Chowdhury, “Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications,” *Energy Conversion and Management: X*, vol. 24, p. 100790, Oct. 2024, doi: 10.1016/j.ecmx.2024.100790.

[9] Y. M. Rind, M. H. Raza, M. Zubair, M. Q. Mehmood, and Y. Massoud, “Smart Energy Meters for Smart Grids, an Internet of Things Perspective,” *Energies (Basel)*, vol. 16, no. 4, p. 1974, Feb. 2023, doi: 10.3390/en16041974.

[10] K. Malewska, S. Cyfert, A. Chwiłkowska-Kubala, K. Mierzejewska, and W. Szumowski, “The missing link between digital transformation and business model innovation in energy SMEs: The role of digital organisational culture,” *Energy Policy*, vol. 192, p. 114254, Sep. 2024, doi: 10.1016/j.enpol.2024.114254.

[11] C. Stephen Ball and D. Degischer, “IoT implementation for energy system sustainability: The role of actors and related challenges,” *Util Policy*, vol. 90, p. 101769, Oct. 2024, doi: 10.1016/j.jup.2024.101769.

[12] T. Ahmad and D. Zhang, “Using the internet of things in smart energy systems and networks,” May 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.scs.2021.102783.

[13] M. Singh, J. Jiao, M. Klobasa, and R. Frietsch, “Servitization of Energy Sector: Emerging Service Business Models and Startup’s Participation,” *Energies (Basel)*, vol. 15, no. 7, p. 2705, Apr. 2022, doi: 10.3390/en15072705.

[14] S. Abdelkader, J. Amissah, and O. Abdel-Rahim, “Virtual power plants: an in-depth analysis of their advancements and importance as crucial players in modern power systems,” *Energy Sustain Soc*, vol. 14, no. 1, p. 52, Aug. 2024, doi: 10.1186/s13705-024-00483-y.

[15] M. S. Bakare, A. Abdulkarim, M. Zeeshan, and A. N. Shuaibu, “A comprehensive overview on demand side energy management towards smart grids: challenges, solutions, and future direction,” *Energy Informatics*, vol. 6, no. 1, p. 4, Mar. 2023, doi: 10.1186/s42162-023-00262-7.

[16] F. P. . Sioshansi, *Consumer, prosumer, prosumager : how service innovations will disrupt the utility business model*. Academic Press, an imprint of Elsevier, 2019.

[17] M. Poyyamozhi, B. Murugesan, N. Rajamanickam, M. Shorfuzzaman, and Y. Aboelmagd, “IoT—A Promising Solution to Energy Management in Smart Buildings: A Systematic Review, Applications, Barriers, and Future Scope,” *Buildings*, vol. 14, no. 11, p. 3446, Oct. 2024, doi: 10.3390/buildings14113446.

[18] G. Ruan, D. Qiu, S. Sivaranjani, A. S. A. Awad, and G. Strbac, “Data-driven energy management of virtual power plants: A review,” *Advances in Applied Energy*, vol. 14, p. 100170, Jul. 2024, doi: 10.1016/j.adapen.2024.100170.

[19] “Predictive Maintenance: how IoT helps reduce O&M costs Barbara IoT.” Accessed: Nov. 01, 2025. [Online]. Available: https://www.barbara.tech/blog/predictive-maintenance

[20] “Why Smart Factories Switch to Predictive Maintenance in 2025.” Accessed: Nov. 01, 2025. [Online]. Available: https://carbonminus.com/why-smart-factories-switch-to-predictive-maintenance-in-2025/

[21] A. P. Azodo, F. C. T. Mezue, O. S. Onyekwere, and O. Idama, “Leveraging AI for predictive maintenance in energy systems: enhancing reliability and efficiency,” *Australian Journal of Multi-Disciplinary Engineering*, pp. 1–19, Sep. 2025, doi: 10.1080/14488388.2025.2557137.

[22] I. Rojek, D. Mikołajewski, A. Mroziński, M. Macko, T. Bednarek, and K. Tyburek, “Internet of Things Applications for Energy Management in Buildings Using Artificial Intelligence—A Case Study,” *Energies (Basel)*, vol. 18, no. 7, p. 1706, Mar. 2025, doi: 10.3390/en18071706.

[23] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, “Recent advances in the analysis of residential electricity consumption and applications of smart meter data,” *Appl Energy*, vol. 208, pp. 402–427, Dec. 2017, doi: 10.1016/j.apenergy.2017.10.014.

[24] M. Afzalan and F. Jazizadeh, “Residential loads flexibility potential for demand response using energy consumption patterns and user segments,” *Appl Energy*, vol. 254, p. 113693, Nov. 2019, doi: 10.1016/j.apenergy.2019.113693.

[25] T. Sachs, A. Gründler, M. Rusic, and G. Fridgen, “Framing Microgrid Design from a Business and Information Systems Engineering Perspective,” *Business & Information Systems Engineering*, vol. 61, no. 6, pp. 729–744, Dec. 2019, doi: 10.1007/s12599-018-00573-0.

[26] Y. Lopez Alzate, E. Gómez-Luna, and J. C. Vasquez, “Innovative Microgrid Services and Applications in Electric Grids: Enhancing Energy Management and Grid Integration,” *Energies (Basel)*, vol. 17, no. 22, p. 5567, Nov. 2024, doi: 10.3390/en17225567.

[27] W. Tushar *et al.*, “Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges,” *Appl Energy*, vol. 282, p. 116131, Jan. 2021, doi: 10.1016/j.apenergy.2020.116131.

[28] Sabine. Löbbe, F. P. . Sioshansi, and David. Robinson, *Energy communities : customer-centered, market-driven, welfare-enhancing?* Academic Press, 2022.

[29] T. Alquthami, A. H. Milyani, M. Awais, and M. B. Rasheed, “An Incentive Based Dynamic Pricing in Smart Grid: A Customer’s Perspective,” *Sustainability*, vol. 13, no. 11, p. 6066, May 2021, doi: 10.3390/su13116066.

[30] L. Wan, Z. Zhang, and J. Wang, “Demonstrability of Narrowband Internet of Things technology in advanced metering infrastructure,” *EURASIP J Wirel Commun Netw*, vol. 2019, no. 1, p. 2, Dec. 2019, doi: 10.1186/s13638-018-1323-y.

[31] L. Che, X. Zhang, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, “Optimal Interconnection Planning of Community Microgrids with Renewable Energy Sources,” *IEEE Trans Smart Grid*, vol. 8, no. 3, pp. 1054–1063, 2017, doi: 10.1109/TSG.2015.2456834.

[32] A. Srivastava, S. Van Passel, and E. Laes, “Dissecting demand response: A quantile analysis of flexibility, household attitudes, and demographics,” *Energy Res Soc Sci*, vol. 52, pp. 169–180, Jun. 2019, doi: 10.1016/j.erss.2019.02.011.

[33] Z. Ma *et al.*, “CELSIUS: an international project providing integrated, systematic, cost-effective large-scale IoT solutions for improving energy efficiency of medium- and large-sized buildings,” *Energy Informatics*, vol. 5, no. S4, p. 41, Dec. 2022, doi: 10.1186/s42162-022-00221-8.

[34] Y. Dai, Z. Huang, Y. Li, and R. Lv, “Robust Optimal Scheduling of Multi-Energy Virtual Power Plants with Incentive Demand Response and Ladder Carbon Trading: A Hybrid Intelligence-Inspired Approach,” *Energies (Basel)*, vol. 18, no. 18, p. 4844, Sep. 2025, doi: 10.3390/en18184844.

[35] J. Hildermeier, J. Burger, A. Jahn, and J. Rosenow, “A Review of Tariffs and Services for Smart Charging of Electric Vehicles in Europe,” *Energies (Basel)*, vol. 16, no. 1, p. 88, Dec. 2022, doi: 10.3390/en16010088.

[36] T. Kunj and K. Pal, “Role,application and challenges of IoT in smart EV charging management:a review,” *Future Batteries*, vol. 7, p. 100102, Sep. 2025, doi: 10.1016/j.fub.2025.100102.

[37] J. Dafda, S. Patil, and K. Parmar, “Decentralized Energy Trading Model with Smart Grids and Blockchain,” in *2022 IEEE 19th India Council International Conference (INDICON)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/INDICON56171.2022.10039962.

[38] S. S. Reka *et al.*, “Decentralized energy trading in smart grid using secured post quantum encryption,” *Results in Engineering*, vol. 27, p. 105767, Sep. 2025, doi: 10.1016/j.rineng.2025.105767.

[39] “IoT In Utilities Market Size, Share And Growth Report, 2030.” Accessed: Nov. 01, 2025. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/iot-utilities-market-report

[40] “Global Smart Home Automation Market Size Report, 2030.” Accessed: Nov. 01, 2025. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/smart-home-automation-market

[41] J. B. Leite and M. Kezunovic, “Integrating IoT Devices with Distribution Energy Management System by Harmonizing Their Logical Models Using IEC Standards 61970/61968 and 61850,” in *IEEE Power and Energy Society General Meeting*, IEEE Computer Society, 2023. doi: 10.1109/PESGM52003.2023.10252950.

[42] S. Singh and A. Yassine, “IoT Big Data Analytics with Fog Computing for Household Energy Management in Smart Grids,” 2019, pp. 13–22. doi: 10.1007/978-3-030-05928-6_2.

[43] R. Das and M. M. Inuwa, “A review on fog computing: Issues, characteristics, challenges, and potential applications,” *Telematics and Informatics Reports*, vol. 10, p. 100049, Jun. 2023, doi: 10.1016/j.teler.2023.100049.

[44] Y. Monges, M. Lopez, E. Flecha, and E. Chaparro, “Performance Evaluation of MQTT and CoAP Communication Protocols in IIoT for Electrical Substations Based on the IEC 61850 Standard,” in *2023 15th IEEE International Conference on Industry Applications (INDUSCON)*, IEEE, Nov. 2023, pp. 774–779. doi: 10.1109/INDUSCON58041.2023.10374829.

[45] K. Jia, J. Xiao, S. Fan, and G. He, “A MQTT/MQTT-SN-Based User Energy Management System for Automated Residential Demand Response: Formal Verification and Cyber-Physical Performance Evaluation,” *Applied Sciences*, vol. 8, no. 7, p. 1035, Jun. 2018, doi: 10.3390/app8071035.

[46] Y. Yang, X. Chen, R. Tan, and Y. Xiao, *Intelligent IoT for the Digital World*. Wiley, 2021. doi: 10.1002/9781119593584.

[47] Y. Yang, X. Luo, X. Chu, and M. T. Zhou, *Fog-enabled intelligent IoT systems*. Springer International Publishing, 2019. doi: 10.1007/978-3-030-23185-9.

[48] J. VanderPlas, *Python Data Science Handbook*, vol. 53, no. 9. O’REILLY, 2017.

[49] Chumbi Wilson E, C. José Eduardo, Jaramillo-Leon Brian, and B. Leite Jônatas, “Análise de Métodos para Agrupamento de Cargas Elétricas em Redes de Distribuição de Grande Porte usando Sistema de Informação Geográfica,” Fortaleza, Brazil, Oct. 2022. doi: 10.20906/CBA2022/3269.

[50] B. Meehan, *GIS for Enhanced Electric Utility Performance*. Boston: House Artech, 2013.

[51] L. Fontolan, W. E. C. Quito, and J. B. Leite, “Determinação da latência de transmissão em duas arquiteturas de internet das coisas na digitalização das redes de distribuição,” Fortaleza, Brazil, 2024. doi: 10.59254/sbpo-2024-193642.