



Edge Computing Architecture for Monitoring Distribution Networks with Distributed Photovoltaic Integration

Tran Ngoc Huy Thinh

Faculty of Engineering and Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

ARTICLE INFO

Published Online:
25 August 2026

Corresponding Author:
Tran Ngoc Huy Thinh

ABSTRACT

The increasing penetration of Distributed Energy Resources, particularly photovoltaic (PV) systems, is imposing critical requirements on low-latency Volt-Var Control (VVC) solutions in the distribution grid. Traditional Centralized Control Architectures (TCCs) are obsolete due to the transmission and calculation delays encountered when solving complex global optimization problems. This paper proposes and analyzes the Cloud-Edge Computing-Cooperative VVC Architecture (CEC-VVC) for the IEEE 33-bus power system. CEC-VVC achieves an optimal balance by utilizing Edge Servers for local optimization within the sub-zone via the MISOCP and by implementing real-time reactive power control of PV inverters using the ANN algorithm. Simultaneously, the system performance is analyzed quantitatively using the data splitting ratio. In-depth analysis shows that the CEC-VVC architecture has reduced the system's Total Response Time from 609ms (TCC) down to a minimum of 121.8ms, achieving an 80% improvement and meeting the real-time response criteria. At the same time, this data splitting ratio reduces the data capacity stored in the Cloud by 2.5 to 10 times compared to TCC. More importantly, the sensitivity analysis proves the robustness of the architecture. The optimal operating point is maintained even when the data transmission speed changes significantly. These results confirm that CEC-VVC provides a flexible balance solution between response speed and resource efficiency, making it an important proposal for stable and reliable operation of the smart grid in the future.

KEYWORDS: Cloud-Edge Computing, Volt-Var Control, Distributed Energy Resources, Low-Latency, Data Splitting Ratio.

I. INTRODUCTION

The strong development of the smart grid, coupled with the increasing integration of Distributed Energy Resources (DERs), particularly photovoltaic (PV) systems, is bringing significant benefits to the environment and energy security [1]. However, the unstable and distributed nature of PV sources poses serious challenges to power quality, notably large voltage fluctuations and local overvoltage phenomena in the distribution grid [2-4].

To maintain power quality and ensure the grid operates within safe limits, efficient Volt-Var Control (VVC) is a very important task [5]. Traditional VVC strategies are implemented in a centralized learning model, where all data is sent to a central control unit (Cloud) for global optimization [6]. Although this model achieves high economic efficiency for the entire system (minimizing loss), it faces two major limitations: high processing and communication delay (Latency) due to physical distance and a massive data

transmission burden, which severely reduces the ability to respond in real-time to rapid voltage changes [7-9].

To solve the latency problem of centralized architectures, the research community has shifted toward decentralized/local control methods [10]. These approaches are developed by directly implementing control algorithms on PV inverters or local control devices. Although achieving fast response (low latency), they usually yield locally optimal results and lack global coordination, which may lead to higher power losses and increase the risk of system instability in severe scenarios.

Recently, Edge Computing and Cloud-Edge collaborative architecture have emerged as a promising solution for smart grid applications, where there is a need to balance local speed and efficiency with global coordination and optimization [11, 12]. This architecture allows for data processing right at the Edge Servers located near the devices, reducing latency for real-time control tasks (such as local control), while still maintaining the role of the Cloud Server for global optimization tasks, thereby minimizing latency. These studies

introduce the Cloud-Edge architecture for Volt/Var Control (VVC). However, quantitative assessment of the total time latency during control data exchange, monitoring the power grid from the central control, and the volume of transmitted/stored data at the cloud data center has not been clearly demonstrated.

To fill the research gap above, this paper introduces a Cloud-Edge Computing collaborative Volt-Var control and optimization architecture designed for distributed power grids. The main contributions of this paper include:

- Propose a CEC-VVC architecture. Divide tasks for the Edge Server layer to perform high-speed local optimization and local control. Use the Cloud Server layer for long-term global optimization. This architecture is designed to overcome the limitations of latency and data throughput of the TCC.
- Provide a mathematical model and a quantitative assessment of the total system response time (T_{sys}) and the Data storage capacity ($D_{storage}$) between the Cloud and the Edge by a task-sharing mechanism based on the data splitting ratio (λ_i). This analysis establishes the basis for identifying the optimal and flexible operating points.
- Analyze the Sensitivity and Robustness of the Latency. Detailed sensitivity analysis quantitatively determines the influence of the Maximum Transmission Time ($T_{tran,max}$) on the optimal operating points. The results prove the robustness of the CEC-VVC architecture against changes in wireless communication technology.

II. OVERVIEW OF THE COOPERATIVE CLOUD – EDGE COMPUTING ARCHITECTURE

To resolve the conflict between the requirement for low latency of real-time control tasks and the need for global optimization of the system's efficiency, we propose a VVC and monitoring architecture based on the Cloud-Edge Computing collaboration model. This architecture distributes computational and control tasks based on time sensitivity, thereby leveraging the powerful processing capability near the data source of Edge Computing and the large computing capacity of Cloud Computing. A CEC-VVC architecture for a power grid integrated with distributed PV power sources, as proposed in this research, is modeled in Figure 1 [13]

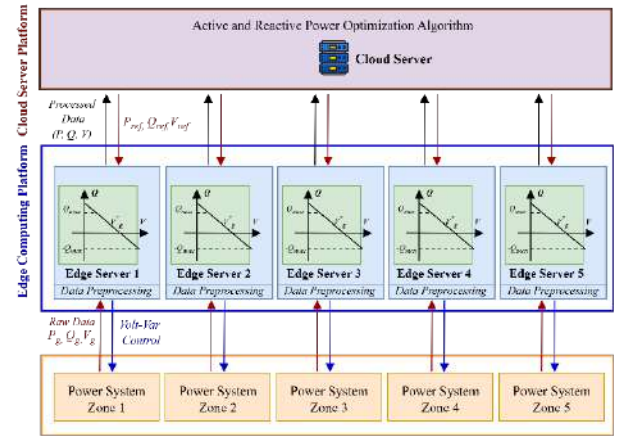


Figure 1. Cloud-Edge Collaboration Architecture for Volt-Var Control Application with Distributed Energy Resources [13]

A. Edge Computing Platform

1) Data Acquisition and Pre-processing:

Based on the description in Figure 1, the Edge Computing Platform is responsible for collecting and preliminary processing of data. The raw input data will be continuously collected and measured in real-time from the measuring devices within the corresponding ES. The crucial raw data that needs to be collected for the volt-var control requirements of the power grid presented in this paper is shown in equation (1).

$$D_i^{raw}(t) = [V_i(t), S_{i,load}(t), S_{i,pv}(t)] \quad (1)$$

The parameters used are defined as follows: (t) represents the continuous instantaneous measurement time; i is the index of the corresponding Zone and Edge Server, where i is an integer ranging from 1 to N ($i \in \{1, 2, \dots, N\}$); $V_i(t)$ denotes the voltage at the nodes belonging to $Zone_i$ at time (t); and finally, P and Q represent the active and reactive power (from both the load and PV) within $Zone_i$.

The raw data, denoted as $D_i^{raw}(t)$, undergoes pre-processing at the ES. This pre-processing step involves two primary operations: filtering noise ($\mathcal{F}$) and normalization ($\mathcal{N}$). The outcome of these procedures is the processed data ($D_i^{pro}(t)$). This initial processing of the data performed at the ES is formally described in equation (2).

$$D_i^{pro}(t) = \mathcal{N}(\mathcal{F}(D_i^{raw}(t))) \quad (2)$$

Where $D_i^{pro}(t)$ is the set of processed quantities presented in equation (3)

$$D_i^{pro}(t) = [V_i^{pro}(t), S_{i,load}^{pro}(t), S_{i,pv}^{pro}(t)] \quad (3)$$

The processed data includes active power, reactive power, and voltage at the load and PV buses of the power grid area under consideration. In [14], the authors based on the size of the processed data and the complexity of the selected algorithm to define the processing resource (RS) as:

$$RS = \frac{D_i^{proc}}{S_{in} \cdot P_{total}} \quad (4)$$

Where $D_i^{proc}(t)$ is the size of the processed data; S_{in} is the size of the input vectors/matrix; P_{total} is the total number of parameters of the algorithm. S_{in} represents the total parameters of the algorithm, which indicates the complexity of the algorithm used. Their complexity and input vector/matrix sizes are taken into account when allocating computational resources and bandwidth for balanced sharing [15]. To balance the load of multiple ESs, the authors proposed a cloud-edge selection algorithm described as follows [14]. The processing resource used by each cloud-edge within a certain time interval can be defined as:

$$RS = \{RS_1, RS_2, \dots, RS_E\} \quad (5)$$

Where E represents the total number of ESs. This means that the processing resources of all ESs are

$$RS_c = \sum_{i=1}^E RS_i \quad (6)$$

While the average processing resource is

$$RS_{average} = RS_c/E \quad (7)$$

2) Local Volt-Var Control

In recent years, the application of AI has brought many benefits in optimizing the operation of power systems. The application of AI is not limited to big data processing at the global level but is also researched and deployed at the edge of local control systems [16]. In [17], the authors examined the application of Deep Learning (DL) on Edge Computing (EC) to promote the development of edge applications. Examples include smart multimedia, smart traffic, smart cities, and smart industry. In [18, 19], the authors analyzed EC architectural designs with constraints on communication, computing capacity, and energy consumption. In this paper, the Edge Server collects data from load and PV nodes and performs local sub-region optimization based on the Mixed Integer Second Order Cone Programming (MISOCP) algorithm [11, 20]. Subsequently, the volt-var setpoints are allocated to the inverters of the PV panels to implement a smart control algorithm to either generate or consume reactive power to maintain the voltage within the safe range of 0.95 p.u. to 1.05 p.u.. The execution time of the MISOCP algorithm on the ES to issue the Volt-Var setpoints is on a time scale requiring low latency (periodically, once per minute due to the small amount of local computation). Therefore, the total system time delay T_{sys} includes the computation time at the ES. The principle of volt-var sub-region control for the power grid, which the ES is responsible for, is as follows:

ES_i uses a local optimization model $\mathcal{M}_{MISOCP}$ at $Zone_i$ to find the setpoints for P, Q, V at the PV nodes. Based on the optimal setpoints $P_{ref,local}$, $Q_{ref,local}$, $V_{ref,local}$ found, the PV inverter performs reactive power control with the pre-trained

ANN model $\mathcal{M}_{ANN}$. The reactive power output or consumption of the PV inverter is described in Figure 2 and Equation (8).

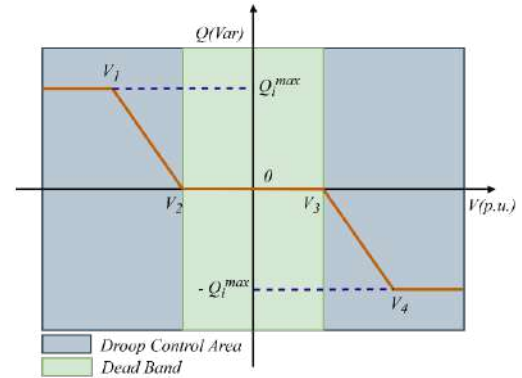


Figure 2. Conventional local voltage droop control [21]

$$Q_i^{cmd}(t) = \begin{cases} Q_i^{max} & (V(t) \leq V_1) \\ -\frac{Q_i^{max}}{V_2 - V_1}(V(t) - V_2) & (V_1 < V(t) \leq V_2) \\ 0 & (V_2 < V(t) \leq V_3) \\ -\frac{Q_i^{max}}{V_4 - V_3}(V(t) - V_3) & (V_3 < V(t) \leq V_4) \\ -Q_i^{max} & (V_4 < V(t)) \end{cases} \quad (8)$$

Control Function: The input is processed data, and the output is the instantaneous Reactive Power Control Command $Q_i^{cmd}(t)$ [22].

$$Q_i^{cmd}(t) = \mathcal{M}_{MISOCP, ANN}(D_i^{proc}(t)) \quad (9)$$

Where the value $Q_i^{cmd}(t)$ will be allocated after the ES solves the local optimization problem based on MISOCP, and the ANN algorithm will be executed at the PV inverter, maintaining the optimized setpoint $Q_i^{cmd}(t)$. The objective of VVC optimization is to simultaneously reduce power loss and voltage deviation in distribution systems as follows [23]:

$$\begin{aligned} \min_{Q_{i,t}^{PV}} \sum_{t \in T} \left\{ \sum_{(i,j) \in \mathcal{L}} P_{ij,t}^{loss} + VVR_t \right\}, \quad \forall i, \forall t \\ VVR_t = \eta \sum_{i \in K} |V_i - V_0|, \quad \forall t \\ P_{ij,t}^{loss} = r_{ij} l_{ij,t}, \quad \forall i, \forall t \\ P_{i,t}^{PV} - P_{i,t}^L = \sum_{g \in G(i)} P_{ig,t} - \sum_{j \in I(i)} (P_{ig,t} - r_{ij} l_{ij,t}) \\ Q_{i,t}^{PV} - Q_{i,t}^L = \sum_{g \in G(i)} Q_{ig,t} - \sum_{j \in I(i)} (Q_{ig,t} - x_{ij} l_{ij,t}) \\ V_{j,t}^2 = V_{i,t}^2 - 2(r_{ij} P_{ij,t} + x_{ij} Q_{ij,t}) \\ \quad + (r_{ij}^2 + x_{ij}^2) l_{ij,t}, \quad \forall i, j, t \\ l_{ij,t} = (P_{ij,t}^2 + Q_{ij,t}^2) / V_{i,t}^2, \quad \forall i, j, t \\ (Q_{i,t}^{PV})^2 + (P_{i,t}^{PV})^2 \leq (S_{i,t}^{PV})^2, \quad \forall i, t \\ V_i^{Min} \leq V_{i,t} \leq V_i^{Max}, \quad \forall i, t \end{aligned} \quad (10)$$

$$V_{min} \leq V_i(t) \leq V_{max}$$

Where V_i^{Min} and V_i^{Max} are the permissible safe voltage limits ($0.95 \text{ p.u.} \leq V_i \leq 1.05 \text{ p.u.}$)

3) Data Exchange between Edge Server and Cloud Server

After performing data processing, optimization, and local control, the ES aggregates the processed data $D_i^{proc}(t)$ and transmits it to the CS.

Data transmission to the Cloud: Data is transmitted every 15 minutes or upon event detection when a voltage exceeds the permissible limit.

$$D_{up,i} = \text{Aggregate} \left(D_i^{proc}(t) \right) \quad (11)$$

Data Transmission: This process is usually associated with the communication delay (T_{comm}) and is coordinated with the global optimization in the cloud.

$$D_{up,i} \leftarrow \{D_{up,1}, D_{up,2}, \dots, D_{up,N}\} \quad (12)$$

B. Cloud Server Platform

1) The role of Cloud Server Platform

- Global optimization (scheduled every 15 minutes) to solve VVC problems and Optimal Power Flow (OPF) problems for the entire grid. This aims to minimize power loss and improve power quality, which requires comprehensive grid data from all Zones in the distributed network.
- The CS also serves as the AI Model training center, where large datasets are aggregated from the ES to build and deploy complex deep learning models, thereby improving the accuracy of local prediction and control.
- Finally, the cloud is responsible for Big Data Storage and Non-Real-time business analytics, while maintaining the ability to self - reconfigure during severe contingency scenarios. This ensures computational capability and issues commands to restore the entire system to a safe and stable state.

2) Cloud-Edge Computing Collaboration Mechanism

The relationship between the CS and the ES is collaborative, aiming to leverage the advantages of both infrastructures, and is described in Table 1 [11]:

Table 1. Relationship between Cloud and Edge Computing

Feature	Edge Server	Cloud Server
Main Task	Real-Time Local Control, Processing Raw Data	Global Optimization, AI Training, Storage
Response Time	Low latency (a few milliseconds), Real-time	Higher latency (a few minutes), non-real-time

Scope of Operation	Local (a Zone/sub-region)	Global (the entire power grid)
Data Processed	Raw PMU/local sensor data	Processed data from all Edge Servers

Typical coordination process (Volt/VAR Control)

- Edge → Cloud (periodic, long cycle): ES collects and pre-processes data, then sends the processed parameters up to the CS.
- Cloud: The CS uses this data to perform OPF/VVC to calculate the optimal reference values ($P_{ref}, Q_{ref}, V_{ref}$ for the entire grid).
- Cloud → Edge: The CS distributes the optimal parameters (or a trained AI model) down to the ES.
- Edge: ES use the parameters received from the Cloud to fine-tune and perform real-time local Volt/VAR control, ensuring a rapid response to sudden load/PV changes within their own zone.

3) Grid Partitioning

Before deploying the Cloud-Edge Computing architecture for the IEEE 33-bus power system, the distribution power system needs to be divided into zones. This zoning is the foundation for mapping the physical zones of the power system to the corresponding ES.

Zoning Principle: The IEEE 33-bus power system is divided into N zones, based on the following criteria:

- Geographical distance
- Control boundaries
- Reactive voltage sensitivity

In this paper, an IEEE 33-bus system model is used, and parameters are set up to simulate the effectiveness of the VVC method. Figure 3 shows that the system includes 6 PV units installed at nodes 11, 17, 21, 24, 29, and 32, with a rated capacity of 1MW per unit. This IEEE 33-bus power grid is divided into sub-regions numbered zone 1, zone 2, zone 3, zone 4, and zone 5. Each region is considered a local power grid and participates in the coordinated VVC control of the power grid. Each region needs to set up a computing server to collect local data. There are several specific requirements for setting up edge computing devices as follows [11]:

- The ES should be installed at the central node of the sub-networking structure, which can be close to PV inverters.
- If a node has a large load and requires more data processing and optimal control, the edge computing device needs to be installed at this node.
- If there are many loads, additional computing servers need to be prepared to ensure sufficient computational power and data processing. In some cases, they can also be responsible for data processing in neighboring regions.

Therefore, in the IEEE 33-bus system of this paper, the ESs are installed at buses 2, 11, 17, 21, 24, and 29. Details on the zoning and the location of loads, PVs, and ESs are

described in Figure 3. Each Zone will be managed by a dedicated ES, ensuring that the server is placed as close as possible to the measuring devices (PMU/RTU) and the inverters of the DERs/PVs in that zone.

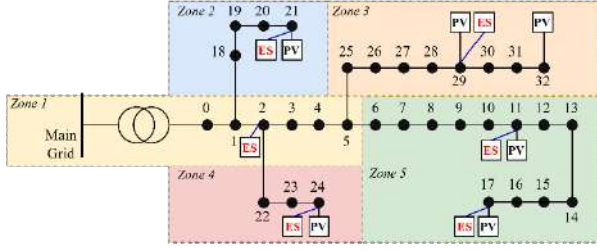


Figure 3. Schematic diagram of IEEE 33-bus

4) Response Latency and Data Splitting Ratio

At the beginning of a time interval, each ES will rely on the previous setpoints to allocate resources to the PV inverter controller. If the system voltage operates within safe limits, the ES will send aggregated data to the Cloud every 15 minutes. If the grid index exceeds the safety threshold and the ES cannot handle it, or if a fault occurs, the ES will immediately send an alert to the Cloud. Each ES can obtain RS_i of all load or PV nodes and then decide whether to forward the measurements to the remote cloud. In [14, 24], these works proposed a data splitting ratio (λ_i). This is a scaling coefficient for data processing at the edge and the cloud, aiming to optimize the computational capability at the ES and reduce the proportion of data sent to the cloud or control center. The value $\lambda_i \in [0,1]$ represents the proportion of data retained for processing at the ES. The new processing resource can be represented as $\widetilde{RS}_i = RS_i \cdot \tau_i$. Where τ_i is the time allocated for a unit of processing resource in the ES. The changing values of λ_i correspond as follows:

- $\lambda_i = 0$ (TCC Mode): In this mode, all data will be sent to the CS for processing.
- $\lambda_i = 1$ (Distributed Mode): In this mode, all data will be retained and processed at the edge (local computing mode).
- $0 < \lambda_i < 1$: Optimal hybrid/cooperative modes between the ES and the CS.

For busy ES_i the offloading rate λ_i should be chosen to minimize the processing latency T_i and the total system processing latency T_{sys} . The optimization of the offloading rate λ_i and its solution depends on the available wireless channel capacity, backhaul link capacity, computing resources at the ES, and resources in the remote cloud cluster. However, to achieve event detection and classification, there are a limited number of choices for λ_i . In this paper, the offloading rate λ_i is investigated by being chosen as 0%, 20%, 40%, 60%, 80%, or 100%. The determination of the ratio of data kept for local processing at the edge or sent to the cloud is presented in the following equations [14]:

Computational latency at the ES ($T_i^{comp,ES}$), Proportional to λ_i :

$$T_i^{comp,ES} = \lambda_i \cdot (RS_{proc,i} \cdot \tau_{ES,i}) \quad (13)$$

Data transmission latency between ES and CS and vice versa ($T_i^{comp,ES}$), directly proportional to $(1 - \lambda_i)$:

$$T_i^{tran,ES} = (1 - \lambda_i) \cdot L_i \cdot \left(\frac{1}{R_i} + \frac{1}{W_i} \right) \quad (14)$$

Where L_i is the size of the processed data, R_i is the wireless channel capacity allocated to ES and W_i is the backhaul communication capacity utilized by ES. (R_i^{-1} and W_i^{-1} can be interpreted as the time required for the wireless and backhaul links to separately transmit one bit of data [24]):

Cloud computation latency ($T_i^{comp,CS}$): Directly proportional to $(1 - \lambda_i)$:

$$T_i^{comp,CS} = (1 - \lambda_i) \cdot (RS_{proc,i} \cdot \tau_{CS,i}) \quad (15)$$

Based on the latency types above, the total system processing latency under these conditions can be described as follows:

$$T_{sys} = \max \{ T_i^{comp,ES}, T_i^{tran,ES} + T_i^{comp,CS} \} \quad (16)$$

- Case 1: Where all data is processed at the ES ($\lambda_i = 1$):

$$T_{sys} = T_i^{comp,ES} \quad (17)$$

This can be viewed as a special ES with $T_i^{tran,ES} = 0$ and $T_i^{comp,CS} = 0$. The parameters λ_i , R_i , W_i and $\tau_{c,i}$ can be carefully adjusted to minimize the total system processing latency T_{sys} . Therefore, the optimization problem can be formulated as follows:

$$\min_{\lambda_i, R_i, W_i, \tau_{c,i}} T_{sys} \quad (18)$$

$$s. t. 0 \leq \lambda_i \leq 1, \sum_i^E R_i \leq R_{all}, \sum_i^E W_i \leq W_{all} \quad (19)$$

$$\sum_i^E \tau_{c,i} \leq \tau_{c,all} \quad (20)$$

Where R_{all} and W_{all} are the constraints on the overall wireless and backhaul communication resources of the ES, and $\tau_{c,all}$ corresponds to the maximum available computation resource provided by the CS. The optimization variables include the data splitting ratio λ_i , the communication resource allocation R_i and W_i , as well as the computation resource allocation $\tau_{c,i}$.

- Case 2: Where all data is processed at the Cloud using the traditional control method ($\lambda_i = 0$):

In this case, all data from the PV nodes will simultaneously send data to the Cloud for processing and await a response.

$$T_{sys} = \{ T_i^{tran,ES} + T_i^{comp,CS} \} \text{ at } \lambda_i = 0 \quad (21)$$

- Case 3: coordination between Cloud and Edge Computing.

The problem is solved using equations (13), (14), (15), (16).

5) Analysis of the amount of data stored in the Cloud

The goal of this analysis is to quantify the difference in the volume of data transmitted and stored at the CS between the

TCC method and the CEC method over a period of one year (365 days).

- System and Data Assumptions

To carry out the calculations, we need to make appropriate assumptions for the VVC system on the PV-integrated distribution grid. Specifically, the ANN algorithm will be used to control the reactive power of the PV inverter to maintain the lowest voltage deviation and power loss (P_{loss}) in the $Zone_i$ area. In this paper, the EMTP software is used to simulate the grid to generate measurement data, including data for the PV unit. The IEEE 1547-2018 standard specifies a minimum measurement and calculation accuracy in steady state for voltage, active power, and reactive power of 10 cycles [25]. Some other assumptions are described in Table 2.

Table 2. Assumed parameters used to calculate the stored data traffic at the Cloud

Component	Symb ol	Value	Unit
Number of PV Units/Measurement Nodes	N_{PV}	6	Units
Number of Edge Servers/Zones	N_{Edge}	5	Zones
Cumulative Analysis/Calculation Period	T_{annual}	365	Days
Total Seconds over 365 Days	T_{total}	31,536,000	Seconds
Raw Measurement Data Size per Sampling Event	S_{meas}	500	Bytes
Sampling/Local Control Frequency	f_{local}	10	Hz
Size of Reported/Aggregated Data	S_{CEC}	500	Bytes
Edge → Cloud Data Transmission Period	T_{CEC}	15	Minutes

- Cloud data storage capacity of the TCC method

In the TCC method, all raw measurement data must be continuously sent to the CS for processing, calculation, and storage. Data transmission frequency to the Cloud:

$$f_{TCC} = f_{local} = 10 \text{ Hz (10 time/ second)} \quad (22)$$

Data rate from a single PV unit:

$$R_{PV,TCC} = S_{meas} \times f_{TCC}$$

$$R_{PV,TCC} = \frac{500 \text{ Bytes}}{\text{time}} \times \frac{10 \text{ time}}{\text{second}} \quad (23)$$

$$= 5,000 \frac{\text{Bytes}}{\text{second}}$$

Total data rate from all PV panels across the system:

$$R_{PV,TCC} = N_{PV} \times R_{PV,TCC} \quad (24)$$

$$R_{PV,TCC} = 6 \times 5,000 \frac{\text{Bytes}}{\text{second}} = 30,000 \frac{\text{Bytes}}{\text{second}}$$

Total stored data after 365 days:

$$D_{TCC} = R_{Total,TCC} \times T_{total}$$

$$D_{TCC} = 30,000 \frac{\text{Bytes}}{\text{second}} \times 31,536,000 \text{ second} \quad (25)$$

$$D_{TCC} = 946,080,000,000 \text{ Bytes} \approx 881 \text{ GB}$$

- Cloud data storage capacity of the CEC method

The CEC-VVC architecture is designed to optimize data transmission and storage by utilizing ES for pre-processing and aggregating data before sending it to the CS. This is completely different from the TCC architecture, where the CS must continuously receive and store raw data. The total annual data storage capacity at the CS (D_{CEC}) is calculated using the following equation:

$$D_{CEC} = S_{update,CEC} \times N_{updates} \quad (26)$$

D_{CEC} is the total data capacity required for storage at the CS over one year. $S_{update,CEC}$ is the size of the aggregated data packet sent to the Cloud after each update. $N_{updates}$ is the total number of data updates to the Cloud within one year. According to the assumptions of the CEC-VVC architecture, aggregated data is sent to the CS periodically every 15 minutes (with an update period $T_{update} = 0.25$ hours).

$$N_{updates} = \frac{365 \text{ days} \times 24 \frac{\text{hours}}{\text{day}}}{0.25 \frac{\text{hours}}{\text{update}}} = 35,040 \frac{\text{updates}}{\text{year}} \quad (27)$$

The data size $S_{update,CEC}$ must be the total size of the aggregated data packets $D_{up,i}$ sent from all ESs every 15 minutes. Based on Table 2, each ES aggregates data from the load/PV nodes within its coverage area and generates an aggregated data packet S_{CEC} to send to the Cloud. With the aggregated data size per Edge ($S_{CEC} = 500$ Bytes/Edge).

$$S_{update,CEC} = N_{EdgeServer} \times S_{CEC} \quad (28)$$

$$S_{update,CEC} = 5 \times 500 \text{ Bytes} = 2,500 \frac{\text{Bytes}}{\text{time}}$$

$$D_{CEC} = S_{update,CEC} \times N_{updates}$$

$$D_{CEC} = 2,500 \frac{\text{Bytes}}{\text{time}} \times 35,040 \frac{\text{times}}{\text{year}}$$

$$D_{CEC} = 87,600,000 \frac{\text{Bytes}}{\text{year}} \approx 0.082 \text{ GB/year}$$

III. RESULTS AND DISCUSSION

A. Analysis of system response time according to the data splitting ratio

The assumption in this paper is that the communication technology between the ES and the CS is 4G technology with

values $R_i = 100$ Mbps and $W_i = 100$ Mbps. The value L_i in equation (14) represents the total data volume (typically measured in bits or Bytes) that ES_i needs to be offloaded to the CS after performing pre-processing. This is based on the data obtained during each sampling event from Table 2. The calculated offload transmission latency is $T_i^{tran,ES} = 40$ ms. We utilize these assumed parameters to simulate resource balancing across different λ_i scenarios. The scenario assuming exclusive computation at the Edge is $T_i^{comp,ES} = 150$ ms (the physical limit of the edge server's computing capability), and the scenario assuming exclusive computation at the Cloud is $T_i^{comp,CS} = 569$ ms [26]. Based on equation (16) and the assumptions, the simplified formula for T_{sys} is rewritten as follows:

$$T_{sys}(\lambda_i) = \max\{\lambda_i \cdot 150 \text{ ms}, (1 - \lambda_i) \cdot 40 \text{ ms} + (1 - \lambda_i) \cdot 569 \text{ ms}\} \quad (29)$$

Where $T_i^{comp,CS} = 569$ ms is the MISOC/ANN processing time at the Edge. 150 ms is the physical processing limit of the ES for complex VVC tasks. In [26], the authors compared the performance of different control strategies for the IEEE 33-bus power system, focusing on voltage deviation control effectiveness and especially the execution speed. This is a crucial point for comparing real-time feasibility. In that study, the authors noted that the centralized optimization method using SOCP (Second-Order Cone Programming), although effective in controlling voltage deviation, had a long solution time (569 ms), which is slow for dealing with rapidly changing Active Distribution Networks (ADNs). Therefore, in this paper, the TCC method's resulting response time is assumed to be 569 ms [26].

- At $\lambda_i = 0$ (TCC mode): All data from all nodes will be sent to the CS. The global optimal Volt-Var models have been pre-built with operating scenarios corresponding to the input values of the collected data. Therefore, the latency is:
 $T_{sys}^{TCC} = \max\{T_i^{comp,CS}\} = 569$ ms.
- At $\lambda_i = 1$ (Maximum Latency 150 ms): This is the largest latency within the scope of the investigation. It occurs because the ES must retain and perform 100% of the complex computational load (MISOC/ANN) itself, reaching its physical limit (150 ms). This refutes the assumption that the Edge has infinite resources.
- At $\lambda_i \in [0, 1]$, the optimal point is determined.

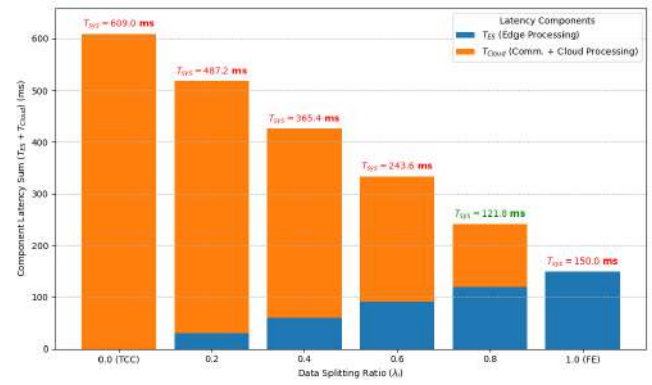


Figure 4. Total system response time results for TCC, FE modes, and varying data sharing coefficients

The results of the latency analysis in Figure 4 confirm the superiority of the CEC-VVC architecture compared to traditional control models. T_{sys} reaches a maximum of 609 ms in the TCC mode but decreases to 121.8 ms at a data splitting ratio of $\lambda_i = 0.8$, which is equivalent to an 80% reduction in latency. The $\lambda_i = 0.8$ point represents the optimal trade-off where the local processing time at the Edge is close to the transmission/processing time at the Cloud (120 ms compared to 121.8 ms). Although the purely Edge Computing mode ($\lambda_i = 1$) also achieves low latency, utilizing a purely edge mode leads to the loss of global system optimization capability. This affirms that CEC-VVC is the most effective solution for achieving ultra-low latency in smart grid operation. To quantify the degree of improvement between the novel improved methods and the TCC method (609 ms), the improvement level is calculated using the following equation:

$$\text{Improvement Level} = \left(1 - \frac{T_{sys}^{new}}{T_{sys}^{TCC}}\right) \times 100\% \quad (30)$$

Where T_{sys}^{TCC} is the response time of the TCC system, which is 569 ms (a time dominated by the complex global OPF calculation process on the Cloud). T_{sys}^{new} is the new response time achieved by the CEC-VVC method. Table 3 presents the level of improvement achieved by the CEC-VVC method with varying λ_i values compared to the TCC method.

Table 3. Improvement Level of the CEC-VVC method compared to TCC

λ_i	$T_{comp,ES}$ (ms)	$T_{tran,ES} + T_{comp,CS}$ (ms)	T_{sys} (ms)	Level of Improvement
0.0 (TCC)	0	609	609	No Improvement
0.2	30	487.2	487.2	20%
0.4	60	365.4	365.4	40%
0.6	90	243.6	243.6	60%
0.8	120	121.8	121.8	80%
1.0 (FE)	150	0	150	75%

B. Analysis of Data Storage Capacity at the Central Cloud

From formula (25), the result is $D_{\text{storage,TCC}} = 881$ GB.

Where $D_{\text{storage,CEC}} = D_{\text{periodic}} + D_{\text{offload}} \cdot D_{\text{offload}}$

$D_{\text{storage,CEC}}$ depends on $(1 - \lambda_i)$ and on the time index of events that need to be sent back to the CS for re-optimization of the entire grid $\beta_{\text{event}} = 50\%$.

$$D_{\text{storage,CEC}}(\lambda_i) = 881 \text{ GB} \cdot (1 - \lambda_i) \cdot 0.5 \quad (31)$$

Figure 5 illustrates the results comparing data storage capacity on the Cloud with varying data splitting ratio. The results show that the data capacity is reduced by 2.5 to 10 times between TCC and the CEC solution with varying λ_i .

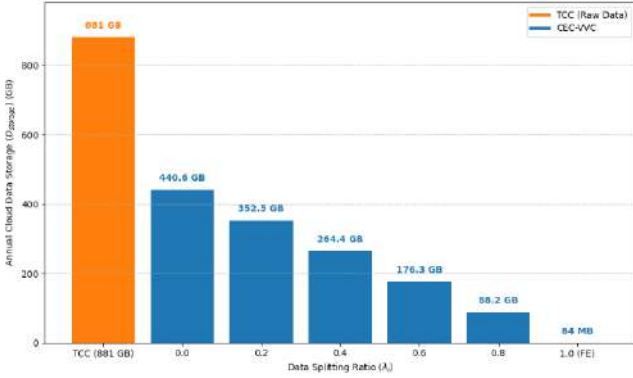


Figure 5. Data storage capacity at the Cloud Server between the CEC and TCC methods

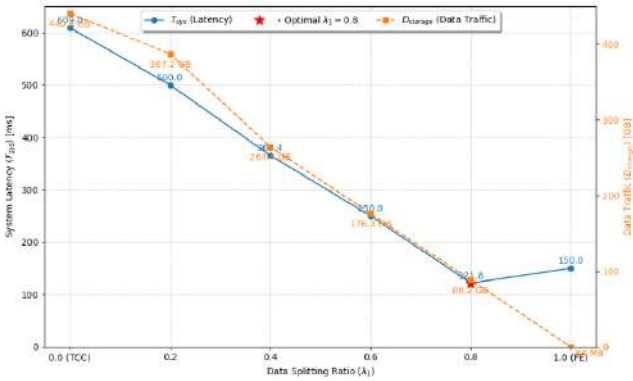


Figure 6. Trade-off between system latency and data storage capacity (CEC-VVC)

Figure 6 illustrates the trade-off relationship between the T_{sys} and D_{storage} in the CEC-VVC architecture when the data splitting ratio λ_i is varied. The analysis results show that the CEC-VVC architecture provides an optimal balance between speed and resource efficiency. The optimal operating point is $\lambda_i = 0.8$. At this point, the T_{sys} reaches its lowest level of 121.8 ms (an 80% reduction compared to TCC). Simultaneously, the D_{storage} at the Cloud is also significantly reduced to 88.2 GB (a 10-fold reduction compared to TCC's 881 GB), accounting for only 20% of the event traffic. Although increasing λ_i to 1 maximizes the reduction of D_{storage} , it increases T_{sys} to 150 ms and results in the loss of

the Cloud's global optimization capability. This demonstrates that $\lambda_i = 0.8$ is the ideal balanced solution to ensure immediate response while still optimizing storage costs.

C. Improving Response Time based on Data Transmission Time

To evaluate the robustness and adaptability of the CEC-VVC architecture to changes in the communication infrastructure, we conducted a sensitivity analysis of the minimum latency $T_{\text{sys,min}}$ based on the optimal data splitting ratio (λ_{opt}) according to variations in data transmission speed using different technologies, such as 4G/satellite networks with higher latency or fiber optic/5G networks with lower latency. Where the initial assumed values are:

$$T_i^{\text{comp,ES}} = \lambda_i \cdot 150 \text{ ms}$$

$$T_i^{\text{tran,ES}} = (1 - \lambda_i) \cdot T_{\text{tran,max}}$$

$$T_i^{\text{comp,CS}} = 569 \text{ ms}$$

Based on equation (16), the λ_{opt} occurs when the processing time at the Edge equals the total transmission and processing time at the Cloud:

$$T_{\text{sys,min}} = \{\lambda_{\text{opt}} \cdot T_i^{\text{comp,ES}}, (1 - \lambda_{\text{opt}}) \cdot T_{\text{tran,max}} + (1 - \lambda_{\text{opt}}) \cdot 569 \text{ ms}\} \quad (32)$$

Balance Equation:

$$\lambda_{\text{opt}} \cdot T_{\text{Max}}^{\text{comp,ES}} = (1 - \lambda_{\text{opt}}) \cdot T_{\text{tran,max}} + (1 - \lambda_{\text{opt}}) \cdot T_{\text{Max}}^{\text{comp,CS}} \quad (33)$$

The final equation to determine the theoretical λ_{opt} is:

$$\lambda_{\text{opt}} = \frac{T_{\text{tran,max}} + T_{\text{Max}}^{\text{comp,CS}}}{T_{\text{tran,max}} + T_{\text{Max}}^{\text{comp,ES}} + T_{\text{Max}}^{\text{comp,CS}}} = \frac{T_{\text{tran,max}} + 569}{T_{\text{tran,max}} + 150} \quad (34)$$

Solving equation (34) to find λ_{opt} based on $T_{\text{tran,max}}$, The calculated results are presented in Table 4.

Table 4. Calculation results based on the comparison of different wireless data transmission technologies [27, 28].

Technology	4G Low Speed	4G High Speed	5G	6G (Projected)
Maximum Data Latency (ms)	50	25	10	1
Maximum Transmission Capacity	20 Mbps	100 Mbps	1000 Mbps	10 Gbps

$T_{tran,Max}$ (ms)	50	40	10	1
λ_{opt}	3.095	3.205	3.619	3.775
$\lambda_{sys,min}$ (ms)	150	150	150	150

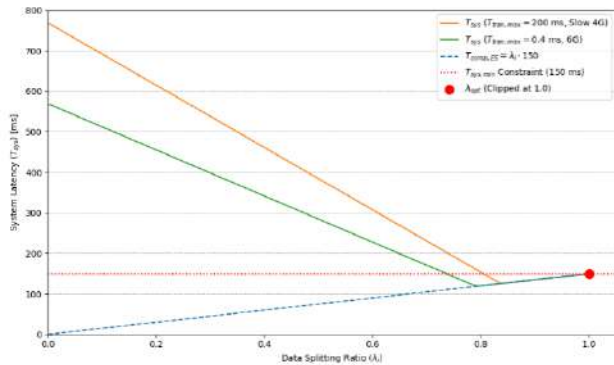


Figure 7. Sensitivity analysis of T_{sys} on $T_{tran,max}$

Based on the calculation results in Table 4 and Figure 7, the robustness and stability of CEC are assessed as follows:

- **Edge Constraint:** In every case, the calculated λ_{opt} value is greater than 1. This calculated value is only theoretical. In practice, λ_{opt} is constrained by $\lambda_{opt} \in [0,1]$. Therefore, if the result is greater than 1, it will be set to 1. This means that, even when the transmission latency $T_{tran,max}$ is very large (50 ms) or very small (1 ms), the system always attempts to shift nearly 100% of the workload to the Edge Server to achieve maximum speed.
- **Robustness:** Since λ_{opt} is limited to 1.0, the actual $T_{sys,min}$ is always fixed at the physical limit of the ES, which is 150 ms (when $\lambda_{opt} = 1.0 \rightarrow T_{sys,min} = 1 \cdot 150$ ms).
- **Stability:** This result proves that large changes in communication technology (from high-latency 4G to 6G) do not change the optimal operating point (λ_{opt} still maintains a narrow range near 1.0). This ensures that CEC-VVC provides a stable and reliable solution for future smart grids, regardless of advances in network infrastructure.

IV. DISCUSSION

This study demonstrates the superior efficiency of the CEC-VVC architecture through an analysis of the trade-off relationship between system latency and data storage capacity based on the data splitting ratio. Experimental results and the mathematical model have indicated that the optimal operating point is at $\lambda_i = 0.8$. At this point, T_{sys} reaches a minimum of 121.8 ms, establishing a significant improvement level of 80% compared to the TCC mode (609 ms). Concurrently, $\lambda_i = 0.8$, the Cloud storage capacity is reduced to only 88.2 GB (a 10-fold reduction compared to TCC's 881 GB), demonstrating high resource efficiency. Furthermore, the sensitivity analysis

confirms the strong robustness of CEC-VVC. T_{sys} remains fixed at 150 ms regardless of large changes in the transmission latency $T_{tran,max}$, proving that this solution is stable and not overly reliant on the advancement of future network technology, starting from 4G mobile networks. Therefore, $\lambda_i = 0.8$ represents the ideal balanced solution, maximizing the immediate response speed while maintaining efficient storage costs and system scalability.

V. CONCLUSION

This study proposed and analyzed the performance of the CEC-VVC architecture aimed at optimizing the control process of smart grids in a distributed computing environment. By developing a mathematical model and conducting sensitivity analysis, we demonstrated the dual benefits and superior robustness of CEC-VVC.

Latency Performance: We identified the optimal balance point between latency components at the data splitting ratio $\lambda_i = 0.8$. At this operating point, T_{sys} reached a minimum of 121.8 ms, establishing an 80% improvement over the TCC mode (609 ms). This strictly meets the requirement for ultra-low latency in real-time control applications.

Robustness: Sensitivity analysis based on $T_{tran,max}$ confirmed the strong stability of the architecture. The maximum transmission latency varied significantly (from 200 ms to 0.4 ms), yet the actual T_{sys} remained constant due to the physical limit of the ES at 150 ms. This affirms that CEC-VVC is a robust solution, unaffected by the advancements or limitations of communication network technology. In summary, CEC-VVC provides a flexible and efficient framework, with $\lambda_i = 0.8$ being the ideal point to maximize instantaneous response speed while still achieving superior resource cost efficiency, thereby reinforcing the feasibility and importance of Edge Computing within the context of next-generation smart grids.

REFERENCES

1. S. Chen and G. Heilscher, "Integration of distributed PV into smart grids: A comprehensive analysis for Germany," *Energy Strategy Reviews*, vol. 55, p. 101525, 2024/09/01/ 2024.
2. N. B. G. Brinkel et al., "Impact of rapid PV fluctuations on power quality in the low-voltage grid and mitigation strategies using electric vehicles," *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105741, 2020/06/01/ 2020.
3. D. Jiandong, L. Hao, L. Yang, and S. Di, "Over-voltage regulation capability analysis of rooftop photovoltaic grid-connected inverters," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 2017, pp. 1-6.
4. Y. Varetsky, M. Fedorcak-Cisak, and B. Kushka, "Studying Voltage Fluctuations in the MV Distribution Grid with a Renewable Energy

- Source," *Energies*, vol. 18, no. 16, p. 4217. doi: 10.3390/en18164217
5. D. Hua, F. Peng, S. Liu, Q. Lin, J. Fan, and Q. Li, "Coordinated Volt/VAR Control in Distribution Networks Considering Demand Response via Safe Deep Reinforcement Learning," *Energies*, vol. 18, no. 2, p. 333. doi: 10.3390/en18020333
6. M. S. Hasan, P. Paudyal, H. V. Padullaparti, M. Mills, and M. Obi, "Centralized Volt-VAR-Watt Optimization for Real Distribution Grids," in *2024 IEEE Power & Energy Society General Meeting (PESGM), 2024*, pp. 1-5.
7. R. Direya, V. Ollikainen, F. Ponci, and A. Monti, "Remote Monitoring and Control of Distributed Energy Resources via 5G," in *2025 7th Global Power, Energy and Communication Conference (GPECOM), 2025*, pp. 941-946.
8. K. E. Ojo, A. K. Saha, and V. M. Srivastava, "Microgrids' Control Strategies and Real-Time Monitoring Systems: A Comprehensive Review," *Energies*, vol. 18, no. 13, p. 3576. doi: 10.3390/en18133576
9. D. Peharda, I. Ivanković, and N. Jaman, "Using Data from SCADA for Centralized Transformer Monitoring Applications," *Procedia Engineering*, vol. 202, pp. 65-75, 2017/01/01/ 2017.
10. H. Faraji, B. Vahidi, A. Khorsandi, and S. Hossein Hosseini, "Multiple control strategies for smart photovoltaic inverter under network voltage fluctuations and islanded operation," *International Journal of Electrical Power & Energy Systems*, vol. 156, p. 109723, 2024/02/01/ 2024.
11. G. Li et al., "A cloud-edge collaborative optimization control strategy for voltage in distribution networks with PV stations," *International Journal of Electrical Power & Energy Systems*, vol. 167, p. 110632, 2025/06/01/ 2025.
12. R. Shan, Z. Ma, and H. Lu, "A voltage control method for distribution networks based on TCN and MPGA under cloud edge collaborative architecture," *Measurement: Sensors*, vol. 31, p. 100969, 2024/02/01/ 2024.
13. H. Hua, Y. Li, T. Wang, N. Dong, W. Li, and J. Cao, "Edge Computing with Artificial Intelligence: A Machine Learning Perspective," vol. 55, no. 9 %J *ACM Comput. Surv.*, p. Article 184, 2023.
14. B. Hu and H. Gharavi, "Artificial Intelligence-Assisted Edge Computing for Wide Area Monitoring," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1659-1667, 2023.
15. M. Chen, Y. Miao, H. Gharavi, L. Hu, and I. Humar, "Intelligent Traffic Adaptive Resource Allocation for Edge Computing-Based 5G Networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 499-508, 2020.
16. H. Hua, Y. Li, T. Wang, N. Dong, W. Li, and J. J. A. C. S. Cao, "Edge Computing with Artificial Intelligence: A Machine Learning Perspective," vol. 55, pp. 1 - 35, 2022.
17. F. Wang, M. Zhang, X. Wang, X. Ma, and J. Liu, "Deep Learning for Edge Computing Applications: A State-of-the-Art Survey," *IEEE Access*, vol. 8, pp. 58322-58336, 2020.
18. Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-Efficient Edge AI: Algorithms and Systems," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2167-2191, 2020.
19. X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of Edge Computing and Deep Learning: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869-904, 2020.
20. W. Gil-González, O. D. Montoya, L. F. Grisales-Noreña, C. L. Trujillo, and D. A. Giral-Ramírez, "A mixed-integer second-order cone model for optimal siting and sizing of dynamic reactive power compensators in distribution grids," *Results in Engineering*, vol. 15, p. 100475, 2022/09/01/ 2022.
21. H.-J. Lee, K.-H. Yoon, J.-W. Shin, J.-C. Kim, and S.-M. Cho, "Optimal Parameters of Volt-Var Function in Smart Inverters for Improving System Performance," *Energies*, vol. 13, no. 9, p. 2294. doi: 10.3390/en13092294
22. R. Wagle, P. Sharma, C. Sharma, M. Amin, and F. Gonzalez-Longatt, "Real-time Volt-Var control of grid forming converters in DER-enriched distribution network," (in English), *Original Research* vol. Volume 10 - 2022, 2023-January-06 2023.
23. D. Hu, Z. Li, Z. Ye, Y. Peng, W. Xi, and T. Cai, "Multi-agent graph reinforcement learning for decentralized Volt-VAR control in power distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109531, 2024/01/01/ 2024.
24. J. Ren, G. Yu, Y. He, and G. Y. Li, "Collaborative Cloud and Edge Computing for Latency Minimization," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 5031-5044, 2019.
25. "IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces," *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pp. 1-138, 2018.
26. B. Zhang and C. He, "Decentralized Voltage and Var Control of Active Distribution Network Based on Parameter-Sharing Deep Reinforcement

- Learning," *IEEE Access*, vol. 13, pp. 115090-115101, 2025.
27. M. Ergen et al., "Edge computing in future wireless networks: A comprehensive evaluation and vision for 6G and beyond," *ICT Express*, vol. 10, no. 5, pp. 1151-1173, 2024/10/01/ 2024.
28. P. Song, J. Lee, A. M. Abdelmoniem, and L. Mukhanov, "Unlocking the power of 4G/5G mobile networks: An empirical dive into quality and energy efficiency in YouTube Edge services," *Computer Networks*, vol. 267, p. 111344, 2025/07/01/ 2025.