CIT. Journal of Computing and Information Technology, Vol. 34, No. 3, September 2026, 175–201
doi: 10.20532/cit.2026.1006581


# Local Differentially Private and Robust Edge–Fog Aggregation for Smart Grid Data

Song-Yao Feng, Zheng-Yan Huang, Zhou Ai and Jun-Hao Song

Information Center of Guangxi Power Grid Co., Ltd; Nanning, China

Smart grid high-frequency electricity consumption data, while supporting load forecasting, regional status monitoring, and anomaly detection, may reveal users' behavioral patterns and power consumption characteristics. Therefore, a balance must be struck between privacy protection, communication efficiency, and robust aggregation performance. Addressing limitations of existing privacy-aggregation methods — including high communication overhead, unclear privacy guarantees, and interference from anomaly reports — we propose a local differential privacy-robust aggregation method for distributed smart grids. This approach applies trimming, random projection, and Laplacian noise to smart meter data at the user or edge side, ensuring the uploaded representations satisfy ε-local differential privacy under single-release conditions. Subsequently, robust fusion is performed at the fog node using coordinate median values and anomaly-aware weights to mitigate the impact of tampered summaries on regional aggregation results. Experiments evaluating 1-hour regional load forecasting and active power estimation demonstrated the method's effectiveness: MAPE scores of 4.89% and 5.02% across both datasets; cumulative communication overhead of $13.4 \times 10^5$ bits in global experiments — lower than FL-PPDA, HE-PDA, and DP-EDA; and an aggregate error increase of 8.92% under data tampering attacks, below DP-EDA's 12.46%. The comprehensive results demonstrate that this method reduces communication load while maintaining acceptable aggregation utility, enhances aggregation stability under limited tampering conditions, and provides a reproducible, lightweight implementation approach for privacy-sensitive smart grid data aggregation.

ACM CCS (2012) Classification: Security and privacy → Human and societal aspects of security and privacy → Privacy protections

*Keywords*: smart grid, edge computing, privacy protection, data aggregation, differential privacy

## 1. Introduction

Smart grids rely on fine-grained electricity-consumption data to support load monitoring, demand forecasting, anomaly detection, and real-time operational control. As smart meters become more widely deployed, large volumes of high-frequency user-side measurements are continuously generated and transmitted across distributed grid infrastructures [1-2]. These data streams are valuable for improving grid observability and operational efficiency, but they also create substantial privacy risks. Individual electricity-consumption traces may reveal household occupancy patterns, daily routines, appliance usage, and other sensitive behavioral information. As a result, privacy-preserving data aggregation has become a fundamental requirement for smart-grid data analytics rather than an optional enhancement [3-5].

A large body of prior work has explored privacy-preserving aggregation through cryptographic protocols, federated learning, secure multiparty computation, homomorphic encryption, and differential privacy [6-8] . These approaches have advanced the state of the art, but they still face important limitations in practical smart-grid deployments. Cryptographic aggregation schemes often impose considerable computation and communication overhead, especially when the number of users or reporting frequency is high. Federated or collaborative learning approaches can reduce direct data sharing, yet they may still require repeated communication rounds and do not always pro-

vide explicit instance-level privacy guarantees [9-11] . Differential privacy offers a principled framework for privacy protection, but many existing smart-grid applications either adopt centralized privacy assumptions or describe perturbation informally without clearly defining the threat model, privacy budget, or utility–privacy trade-off. In large-scale distributed power environments, these limitations make it difficult to simultaneously achieve low latency, scalability, formal privacy protection, and robustness to malicious interference.

At the same time, the computing architecture of smart-grid systems is changing. Pure cloud-centric processing is increasingly insufficient for latency-sensitive and bandwidth-intensive scenarios. Edge computing can reduce communication cost by performing local computation close to smart meters, while fog computing provides an intermediate collaboration layer capable of fusing regional summaries before cloud-level analysis [12-14]. This layered deployment is attractive for smart grids because it supports decentralized processing and reduces raw data exposure. However, architecture alone is not a sufficient contribution from a computer-science perspective. A layered edge–fog design only becomes meaningful when paired with a clearly specified aggregation algorithm, a rigorous privacy model, and a reproducible evaluation protocol. This is precisely where existing architecture-oriented studies often remain weak: they describe data flow and system structure but do not define a formal privacy-preserving aggregation mechanism or prove what is actually protected [15-17] .

This study adopts an algorithmic modeling perspective from computer science, formally framing privacy-preserving smart grid data aggregation as a computable optimization and estimation problem rather than merely focusing on system architecture integration. Specifically, the research defines the data aggregation process in edge–fog environments as a functional mapping problem with explicit input, output, and parameter constraints. The input consists of locally clipped and projected data representations from edge devices, while the output represents globally aggregated statistical results through robust fog-layer fusion. The core objective is to maximize aggregation utility while satisfying local differential privacy constraints and suppressing abnormal interference impacts. Building upon this formalization, we develop a unified aggregation algorithm incorporating local perturbation mechanisms and anomaly-aware weighting rules, ensuring the proposed method possesses well-defined algorithmic boundaries and reproducible implementation pathways. We propose an edge–fog aggregation method in which privacy protection is enforced at the user/edge boundary through a local differential privacy (LDP) mechanism, and robustness is strengthened at the fog layer through anomaly-aware weighted aggregation. Specifically, user-side measurement vectors are first clipped and projected into a low-dimensional representation, then perturbed locally using calibrated Laplace noise before transmission. The fog layer receives only privatized edge summaries and applies robust weighting to reduce the influence of abnormal or tampered reports. In this way, the method combines three desirable properties: local privacy protection, communication efficiency, and aggregation robustness. Unlike architecture-only descriptions, the proposed approach is defined as an explicit procedure with identifiable inputs, outputs, and parameters.

The privacy objective in this study is formalized through $\varepsilon$-local differential privacy, which ensures that each reported edge-side representation provides a bounded privacy guarantee before any centralized aggregation occurs. This choice is motivated by the trust structure of smart-grid environments: user-side data should ideally be protected before leaving the local boundary, rather than relying solely on trusted aggregators. In addition to privacy, the method is designed to tolerate practical smart-grid threat scenarios, including communication eavesdropping, behavioral inference from historical protected outputs, and data tampering or poisoning attacks against intermediate nodes. By explicitly specifying these adversarial capabilities, the paper aims to connect the threat model, protection mechanism, and evaluation protocol more tightly than in the current version of the literature.

Therefore, the primary contribution of this research lies not in proposing a novel edge-fog computing architecture, but in designing a formal privacy-protected aggregation algorithm deployable in distributed smart grid environments. Edge-fog computing is primarily em-

ployed in this study as an optimal deployment paradigm for large-scale smart grids, enabling functions such as local perturbation generation, regional aggregation, and cloud-based analysis. The algorithm itself can be extended to either a single-layer aggregation model from edge to cloud or a three-tier collaborative aggregation model involving edge, fog, and cloud nodes. In other words, the research emphasizes the integration of edge-side local differential privacy perturbation techniques with robust weighted aggregation rules, rather than treating system layering itself as the core innovation.

The main contribution of the paper is therefore not merely the use of edge and fog computing, but the design of a formally specified privacy-preserving aggregation algorithm for that architecture. More concretely, this study makes four contributions. First, it proposes a local perturbation-and-aggregation mechanism for smart-grid data in which edge-side reports satisfy a formal LDP guarantee. Second, it introduces a fog-layer robust fusion rule that reduces the effect of corrupted or anomalous edge summaries. Third, it provides a unified threat model and method specification, thereby improving reproducibility and making the privacy claim verifiable rather than descriptive. Fourth, it establishes an evaluation framework for smart-grid aggregation that jointly considers utility, communication overhead, privacy leakage resistance, and robustness under adversarial interference.

The remainder of this paper is organized as follows. Section 2 reviews related work on privacy-preserving aggregation in smart grids, edge/fog-based data processing, and formal privacy mechanisms. Section 3 presents the proposed method, including the problem definition, threat model, local privacy mechanism, robust fog aggregation rule, privacy guarantee, and computational complexity. Section 4 reports the experimental evaluation on public smart-grid datasets, including utility–privacy trade-offs, communication efficiency, and robustness against adversarial attacks. Section 5 concludes the paper and outlines directions for future work.

## 2. Literature Review

### 2.1. Privacy-preserving data aggregation in smart grids

Smart-grid data aggregation has long been recognized as a privacy-sensitive problem because high-frequency electricity-consumption records can reveal user identity, occupancy behavior, appliance usage patterns, and other fine-grained household attributes. As smart-meter deployment expands, privacy-preserving aggregation has therefore become a central requirement for practical smart-grid analytics rather than a secondary optimization goal. Existing work has addressed this issue mainly through cryptographic aggregation protocols, certificateless authentication, anonymous reporting, threshold decryption, and secure aggregation schemes [18-19]. Several representative studies have explored multifunctional aggregation based on public-key primitives, lightweight verifiable aggregation, conditional anonymity, and Paillier-based secure aggregation. These studies demonstrate that privacy-preserving smart-grid aggregation is technically feasible, but they also reveal a common design trade-off between security strength and system overhead [20-21].

A first major research stream relies on cryptographic protection. In this line of work, confidentiality is usually achieved by encrypting user reports and enabling only aggregated decryption or verifiable aggregation. Such methods often provide strong confidentiality assumptions, but they tend to increase communication payload, computational latency, or key-management complexity. This limitation becomes more severe in smart-grid settings characterized by many users, frequent reporting intervals, and strict timeliness requirements. Existing privacy-preserving aggregation approaches often face high communication and computation costs in large-scale smart-grid scenarios [22-24].

A second stream introduces identity protection and authentication mechanisms, such as conditional anonymity and certificateless signatures, to prevent the disclosure of user identity during aggregation. These methods are useful for unlinkability and secure participation, but they do

not necessarily provide protection against inference from the transmitted numerical data itself. In other words, identity anonymity and data privacy are related but not equivalent objectives. A system may hide user identifiers while still leaking behavioral features through high-resolution load traces. This distinction is important because the current paper aims to protect not only identity-related information, but also user-side consumption characteristics embedded in the reported data.

A third stream uses differential privacy-inspired perturbation or lightweight masking methods to reduce disclosure risk with lower computational cost than full cryptographic aggregation. These approaches are attractive for distributed grid environments because they can reduce trust assumptions and improve scalability. However, many application-oriented papers describe perturbation at a conceptual level without fully specifying the privacy mechanism, the sensitivity bound, the privacy budget, or the adversarial success criterion. As a result, privacy is often evaluated indirectly through utility-related indicators rather than through formal privacy guarantees or explicit attack resistance [25-26]. This limitation indicates that privacy-related evaluation should be directly connected to a standard privacy definition, explicit sensitivity analysis, and a clearly specified adversarial model.

Beyond privacy aggregation research in smart grid applications, the fundamental mechanisms of local differential privacy also provide crucial theoretical foundations for methodological design. RAPPOR employs a stochastic response mechanism to introduce perturbations at client devices, enabling servers to estimate frequency distributions without directly accessing individual real-world data — demonstrating the feasibility of localized perturbation techniques for large-scale client-side statistics [27]. Apple's Private Count Mean Sketch algorithm further integrates local differential privacy with compact sketch structures, achieving frequency estimation of high-dimensional discrete data through client-side perturbation and server-side aggregation, thereby combining low communication overhead with robust local privacy protection [28]. Google's Prochlo system utilizes anEncode-Shuffle-Analyze architecture that combines local encoding, anonymized shuffling, and aggregated analysis to minimize correlations between individual reports and user identities [29]. These foundational studies highlight that local differential privacy requires not only pre-local data perturbation but also holistic designs incorporating compressed representations, aggregated estimation, and attack assumptions. Inspired by these insights, current research in smart grids adopts edge-side truncation, low-dimensional projection, and Laplacian perturbation mechanisms, supplemented by anomaly-aware robust fusion at the fog layer to simultaneously ensure local privacy guarantees, communication efficiency, and robustness against report tampering. RAPPOR collects statistical information from terminal software while providing robust privacy guarantees; Apple's Private Count Mean Sketch enables local differential privacy frequency estimation for device-side records; and Prochlo proposes a privacy analysis framework comprising Encode, Shuffle, and Analyze.

Furthermore, smart meter data exhibits distinct temporal correlations, with electricity consumption within consecutive sampling intervals often demonstrating strong autocorrelation. Attackers may leverage the correlation patterns between multiple data publication points to infer user behavior. Existing research has addressed this issue from the perspectives of continuous data publication and smart meter privacy protection. For instance, $W$-event differential privacy emphasizes comprehensive protection of event sequences within continuous time windows, while general privacy frameworks focus on maintaining privacy when internal states of streaming algorithms are exposed. Smart meter studies also indicate that data granularity, temporal sequence dependencies, and aggregation scope significantly impact privacy protection effectiveness [30]. Consequently, local differential privacy mechanisms for smart grids must not only account for disturbance intensity at individual timestamps but also consider the cumulative leakage risks arising from cross-temporal data correlations.

## 2.2. Edge and fog computing for smart-grid data processing

With the growth of smart-grid data volume and the increasing need for low-latency response, edge computing has emerged as an important

architectural solution for localized processing. By moving part of the computation closer to user devices, edge computing can reduce communication load, support faster response, and limit direct exposure of raw user data to centralized platforms. Recent studies have combined edge computing with load forecasting, distributed scheduling, and secure aggregation, showing the potential of localized processing in smart-grid environments. These studies indicate that edge-assisted processing is promising for smart-grid environments, especially when communication efficiency and real-time capability are emphasized.

Fog computing extends this idea by introducing an intermediate collaborative layer between edge nodes and the cloud. In principle, fog nodes can fuse summaries from multiple edges, coordinate regional estimates, and reduce cloud-side processing pressure. This layered design is especially relevant for geographically distributed and hierarchical infrastructures such as power systems. However, the edge-fog architecture alone cannot address privacy concerns. While defining where computations occur, it fails to specify how privacy is protected, how data perturbations are implemented, or how aggregated results resist the impact of anomalous reports. In other words, deployment architectures and algorithmic mechanisms must be clearly distinguished: edge-fog computing provides a distributed runtime environment, whereas privacy protection and robustness rely on specific local perturbation techniques, sensitivity controls, privacy budget allocation strategies, and robust aggregation rules. Existing research often overemphasizes system hierarchies or data flow designs without proposing verifiable privacy-preserving algorithms, leading readers to focus on layered processing pipelines while overlooking truly reproducible and provable aggregation mechanisms.

Another limitation of many edge/fog smart-grid studies is that privacy, robustness, and aggregation utility are often treated as loosely connected objectives rather than jointly formalized design targets. For example, a method may report lower communication overhead or improved prediction performance, but still leave the privacy guarantee underspecified. Similarly, a system may claim resistance to attack while evaluating security only through coarse performance degradation metrics. This gap is particularly important in smart-grid scenarios because the deployment of intermediate nodes increases the attack surface: transmitted reports may be intercepted, historical outputs may be mined for inference, and intermediate summaries may be tampered with. A privacy-preserving edge–fog solution therefore requires more than architectural layering; it requires a mechanism whose protection properties can be linked directly to the assumed adversary model.

### 2.3. Limitations of existing approaches

Taken together, the existing literature reveals three persistent gaps.

First, many schemes remain computation-heavy or communication-intensive when deployed at smart-grid scale. Cryptographic aggregation and repeated collaborative training methods can become expensive under frequent reporting and large user populations. While such methods may offer strong protection, their cost can limit practical adoption in large distributed power environments. This motivates the search for lighter-weight local mechanisms that can preserve privacy without relying on heavy global coordination.

Secondly, privacy mechanisms often lack rigorous formal justification. While foundational studies such as RAPPOR, Private Count Mean Sketch, and Prochlo have demonstrated that local random perturbations, compact frequency estimation, and anonymous aggregation analysis provide verifiable privacy protection in large-scale client-side data statistics, some smart grid application research has not fully adopted these formal formulations. Certain studies rely on indirect metrics like CR, F1 stability, or "privacy protection levels," which fail to quantitatively characterize privacy breaches or attacker advantages formally. Therefore, it is essential to directly link smart grid data aggregation mechanisms with definitions of local differential privacy, sensitivity thresholds, privacy budgets, and attack models, transforming privacy protection from descriptive statements into verifiable algorithmic properties.

Third, robustness and privacy are rarely unified in one clearly specified aggregation algorithm. Existing studies often focus on either confidentiality against eavesdropping or efficiency under distributed computation, but do not simultaneously define protection against eavesdropping, behavioral inference, and tampering in one consistent framework. These limitations show that mathematical definitions, reproducibility, and attack evaluation remain important challenges for privacy-preserving smart-grid aggregation. These limitations motivate the development of a reproducible aggregation algorithm with an explicit threat model, a formal privacy guarantee, and a robustness-aware aggregation rule.

### 2.4. Research gap and positioning of the present study

This study is motivated by the observation that smart-grid privacy-preserving aggregation requires a better balance among four objectives: formal privacy protection, communication efficiency, aggregation utility, and robustness to malicious interference. Existing cryptographic schemes often privilege security strength at the expense of overhead; existing edge/fog studies often improve system efficiency without clearly defining a privacy guarantee; and many application-driven approaches do not connect their stated threat model to the implemented mechanism and evaluation protocol in a rigorous way.

To address this gap, the research presents its findings as an algorithmic privacy protection aggregation method deployable in edge, fog, or cloud-cooperative environments, rather than merely a system integration framework. Edge-fog computing is not a prerequisite for this approach; rather, it aligns particularly well with the requirements of smart grid regional management, low-latency response, and distributed data processing. In a single-layer edge-to-cloud deployment, edge devices can still perform local differential privacy perturbation while the cloud handles robust weighted aggregation. For a three-tier deployment involving edge, fog, and cloud, fog nodes can handle regional-level robust fusion tasks, with the cloud responsible for global aggregation and analytical scheduling. Thus, the core contribution lies in the algorithmic mechanism combining local perturbation with robust aggregation, rather than the edge-fog architecture itself. More specifically, the proposed method integrates two key components:

(1) an edge-side local privacy mechanism that protects user data before transmission using a formal local perturbation model;

(2) a fog-side robust aggregation rule that reduces the influence of abnormal or tampered summaries.

This positioning distinguishes the present study from prior work in three ways. First, the privacy claim is tied to a standard formalism rather than described only qualitatively. Second, the threat model explicitly includes communication eavesdropping, behavioral inference, and data tampering. Third, the aggregation process is specified as a reproducible algorithm rather than as a loosely connected sequence of architectural modules. These designs address key weaknesses in existing studies by linking the threat model, privacy mechanism, aggregation algorithm, and evaluation protocol within a unified framework.

## 3. Methodology

### 3.1. Problem definition

Consider a smart-grid system consisting of a set of user nodes $U = \{u_1, ..., u_N\}$, a set of edge nodes $E = \{e_1, ..., e_M\}$, a set of fog nodes $F = \{f_1, ..., f_R\}$, and a cloud center. Each user periodically generates a smart-meter measurement vector $x_t^{(t)} \in R^d$ where $x^{(t)}$ denotes the $d$-dimensional electricity-consumption feature vector reported by user $u_i$ at time slot $t$.

$$x_i^t = \left[ x_{i,1}^t, x_{i,2}^t, \ldots x_{i,d}^t \right]^\top \in R^d$$

Among them, $x_{i,l}^t$ represents the $l$th electricity consumption characteristic of user $u_i$ in time slot $t$. Let the set of users served by edge node $e_m$ be $U_m$, and the set of edge nodes managed by fog node $f_q$ be $E_q$ . The objective of the study is to estimate the regional or global aggregated result of time slot $t$ without exposing the original user data $x_i^t$:

$$y^t = \frac{1}{N} \sum_{i=1}^{N} x_i^t.$$

And through the privacy protection mechanism, an estimated value $\hat{y}^t$ is obtained, which simultaneously meets the requirements of local differential privacy, robust aggregation and communication compression.

Depending on the deployment, the vector may contain active power, current, voltage, load-derived statistical features, or short-window temporal summaries.

The objective is to estimate a regional or global aggregate statistic from all users' reports while simultaneously satisfying four system requirements:

- **Privacy preservation:** transmitted data should not reveal the original smart-meter records of individual users;
- **Robustness:** the aggregation process should remain stable when a fraction of reports is tampered with or manipulated;
- **Communication efficiency:** the transmitted representation should be significantly smaller or lighter than cryptographic heavyweight solutions;
- **Aggregation utility:** the final estimate should remain accurate enough for downstream smart-grid analytics.

Let $S^{(t)}$ denote the target aggregate at time $t$, such as a regional load estimate or a protected summary statistic. The problem addressed in this paper is to design an edge–fog aggregation mechanism that computes an estimat $S_{hat}^{(t)}$ from user data while providing a formal privacy guarantee and robustness against adversarial interference.

## 3.2. System and threat model

### 3.2.1. System model

The proposed system follows a four-layer hierarchy:

- **User layer:** smart meters collect local electricity-consumption data;
- **Edge layer:** each edge node receives reports from its associated users and performs local privacy transformation;
- **Fog layer:** fog nodes fuse multiple edge summaries using a robust aggregation rule;
- **Cloud layer:** the cloud receives fog-level aggregates for system-wide analysis and decision support.

Each user $u_i$ is associated with one edge node $e_{(i)}$. Each edge node serves a user subset $U_m \in U$, and each fog node serves a subset of edge nodes. The study adopts the edge-fog-cloud hierarchy as the primary experimental and demonstration scenario because this architecture meets the practical requirements for regional dispatching, hierarchical aggregation, and low-latency processing in smart grids. However, the proposed algorithm does not rely on a fixed two-tier intermediate architecture. In the single-layer edge-cloud aggregation scenario, edge nodes can directly send anonymized summaries to the cloud, where anomalous-aware weighted aggregation is performed; in the three-layer edge-fog-cloud aggregation scenario, fog nodes handle region-level robust fusion while the cloud performs cross-region aggregation. Neither deployment approach alters the fundamental algorithmic logic of local differential privacy perturbation and robust weighted aggregation at the edge.

### 3.2.2. Threat model

The study employs a limited trust and bounded attacker model. User-side raw data is not trusted by edge nodes or fog nodes; thus, privacy protection is achieved through local perturbation before leaving the local boundary. Under normal circumstances, edge nodes and fog nodes are considered semi-honest nodes — they execute the aggregation process according to protocols but may attempt to infer user information from received reports. For robustness analysis, partial tampering of edge nodes or uploaded reports is permitted and treated as potential Byzantine-type abnormal inputs, without assuming simultaneous compromise of most edge nodes. The research assumes that communication channels possess basic authentication and integrity verification capabilities to prevent attackers from arbitrarily forging legitimate node identities; failure of the authentication mechanism constitutes a more severe

attack scenario, which falls outside the primary scope of this study.

To make the security assumptions explicit, we consider three adversarial capabilities.

(A1) **Communication eavesdropping attack.** An adversary observes messages transmitted from users to edge nodes or from edge nodes to fog nodes. The adversary aims to infer the original user data from transmitted messages.

(A2) **Behavioral inference attack.** The adversary collects historical protected outputs and attempts to infer private user attributes, such as consumption pattern class, occupancy behavior, or appliance usage trend.

(A3) **Data tampering/pollution attacks.** Attackers control a subset of reporting nodes and inject altered reports to distort aggregated results. This paper stipulates that the proportion of contaminated edge summaries within a single fog node's jurisdiction should not exceed $\alpha$; this condition is simulated through varying tampering rates in experiments. When the contamination rate becomes excessively high and approaches that of the majority nodes, the stability of robust weighted aggregation decreases.

The system aims to satisfy the following guarantees:

- **Local data confidentiality:** intercepted reports should not reveal raw user measurements;
- **Inference resistance:** privacy-protected reports should reduce an attacker's ability to reconstruct or classify user behavior;
- **Aggregation robustness:** Under the tampering of the restricted degree of the report, the final aggregated result should remain stable. In this study, the main threat model focuses on the risks of eavesdropping, behavior inference, and data tampering within a single report or a single aggregation cycle. The default attacker cannot obtain all the protected release results of the same user within a continuous time window indefinitely [31]. Therefore, the $\varepsilon$-local differential privacy adopted in the study mainly characterizes the privacy boundary in the single local release process. For long-term time series inference attacks that utilize multi-timestamp autocorrelation, this study does not consider it as the core attack model, but discusses its extended processing methods within the limitations and future work.

## 3.3. Proposed method: edge-LDP and robust fog aggregation

The proposed method consists of two main components:

- an edge-side local privacy mechanism that transforms each user record into a privacy-protected representation satisfying local differential privacy; and
- a fog-side robust aggregation mechanism that reduces the influence of anomalous or tampered reports.

The full workflow is shown conceptually as:

raw user data → clipping and projection → local perturbation → edge summary → robust fog fusion → $S_{hat}^{(t)}$

### 3.3.1. Edge-side data preprocessing

For each user $u_i$ at time $t$, let the original input vector be $x_i^{(t)} \in R^d$ . To bound sensitivity, each record is first clipped into a compact domain:

$$\text{x-tilde} \quad \tilde{x}_i^{(t)} = clip\left(x_i^{(t)}, -C, C\right)$$

where OBJ is the clipping threshold applied elementwise.

To reduce communication overhead and standardize the transmitted representation, the clipped vector is projected into a lower-dimensional space:

$$y_i^{(t)} = W_x \overline{x}_i^{(t)} + b$$

where $W \in R^{k \times d}$ is a projection matrix, $b \in R^k$ is a bias vector, and $k \ll d$ is the compressed feature dimension. In the experiment, a data-independent random Gaussian projection matrix was used, with the matrix elements satisfying $W_{j,l} \sim N(0,1/k)$. The random seed was fixed and generated before the experiment began. Since $W$ was not obtained by training on the user's

original data, it does not consume the privacy budget itself. If a sparse random projection matrix is used, the sensitivity can be estimated in the same way based on the proportion of non-zero elements; if a learning-based linear encoder is used, the training process of the encoder needs to additionally satisfy differential privacy constraints, or rely on a trusted public dataset for training. This experiment does not adopt this setting.

In practical applications, $W$ can be selected as a random projection matrix or a learned lightweight linear encoder. To avoid introducing additional privacy leakage from the projection matrix itself, in this experiment, a random Gaussian projection matrix $W$ was generated independently of the user data and fixed as the random seed before the experiment began. The $W$ was not trained using the user's original data. Therefore, $W$ is regarded as an open or semi-open algorithm parameter, without consuming additional privacy budget. The ε-LDP ensures that it only acts on the local perturbation release process after clipping and projection. If a lightweight linear encoder trained based on user data is used, the training process of this encoder needs to additionally satisfy differential privacy constraints, or it is assumed that the training data comes from a trusted public dataset, which is not within the scope of the experimental setup.

### 3.3.2. Local differential privacy mechanism

To protect user privacy before transmission, each projected vector is perturbed locally at the edge interface. The released representation is defined as

$$z_i^{(t)} = y_i^{(t)} + \eta_i^{(t)}$$

where $\eta_i^{(t)}$ is sampled independently from a Laplace distribution:

$$\eta_i^{(t)} \sim Lap\left(0, \Delta / \varepsilon\right)^k .$$

Here, $\varepsilon > 0$ is the privacy budget and $\Delta$ is the $L_1$-sensitivity of the projection mapping $W\tilde{x}_i^{(t)}$. If $W$ is fixed and clipping is applied, $\Delta$ is bounded and can be computed in advance. The study adopts the Laplace mechanism instead of the Gaussian mechanism. The main reason for this is that this time it is hoped to obtain a pure $\varepsilon$-local differential privacy guarantee in a single local release scenario, while the Gaussian mechanism usually corresponds to an approximate $(\varepsilon,\delta)$-differential privacy guarantee. For the data release on the user side of the smart grid, the study prioritizes ensuring the directness and interpretability of the privacy definition. Therefore, the Laplace mechanism based on L1 sensitivity is selected as the main perturbation scheme. It should be noted that in high-dimensional projection representations or long-term sequence combination scenarios, the Gaussian mechanism based on L2 sensitivity may achieve better mean square error performance under the same privacy budget. Therefore, the study incorporates the Gaussian mechanism as a supplementary control method into the experimental analysis to verify the utility cost of the Laplace mechanism under the pure $\varepsilon$-LDP setting.

### 3.3.3. Edge-level local aggregation

Each edge node $e_m$ collects privacy-protected reports from its assigned user set $U_m$. A local summary vector is computed as

$$g_m^{(t)} = \frac{1}{|U_m|} \sum_{u_i \in U_m} z^{(t)}_i .$$

This local summary is compact, privacy-protected, and suitable for transmission to the fog layer. Compared with transmitting all raw user data, this reduces both communication volume and centralized exposure risk.

## 3.4. Fog-side robust aggregation

The fog layer receives edge summaries $\{g_m^{(t)}\}$ from the edge nodes assigned to fog node $f_r$. Since some edge summaries may be affected by tampering, failures, or abnormal input behavior, direct averaging is not sufficiently robust. We therefore introduce an anomaly-aware weighted aggregation rule.

First, compute a robust center estimate using the coordinate-wise median:

$$c_r^{(t)} = median\left[ g_m^{(t)} \right].$$

Then compute an anomaly score for each edge summary:

$$a_m^{(t)} = \left\| g_m^{(t)} - c_r^{(t)} \right\|_2 .$$

A robust weight is assigned as

$$\omega_m^{(t)} = \frac{1}{1 + a_m^{(t)}}$$

$$\bar{\omega}_m^{(t)} = \frac{\omega_m^{(t)}}{\sum_j \omega_j^{(t)}} .$$

The fog-level aggregate is then obtained by

$$h_r^{(t)} = \sum_{m=1}^{M_r} \bar{\omega}_m^{(t)} g_m^{(t)} .$$

This weighting mechanism down-weights reports that deviate strongly from the robust center and therefore improves resistance to corrupted or manipulated edge summaries.

Finally, the cloud center fuses fog-level outputs:

$$\hat{S}_r^{(t)} = \frac{1}{R} \sum_{r=1}^{R} h_r^{(t)}$$

or, if desired, a second-stage robust average may also be applied across fog nodes.

### 3.5. Formal privacy guarantee

The privacy mechanism is defined at the user/edge boundary and satisfies ε-local differential privacy.

**Definition 1 (Local Differential Privacy).** A randomized mechanism $M$ satisfies epsilon-local differential privacy if for any two input records $x$ and $x'$ and for any measurable output set $Z$,

$$P_r[M(x) \in Z] \le e^{\varepsilon} P_r[M(x') \in Z] .$$

**Theorem 1.** Let the original input vector of the user be $x \in \mathbb{R}^d$, after element-level clipping it satisfies $\tilde{x}_l \in [-C, C]$, and the projection representation is $z = W\tilde{x} + b$, where $W \in \mathbb{R}^{k \times d}$ is the data-independent projection matrix. If the release mechanism is $M(x) = W\tilde{x} + b + \eta$. Here, each component of $\eta$ is independently sampled from $Lap(\Delta 1/\varepsilon)$. Then this mechanism satisfies ε-local differential privacy. Its $L_1$ sensitivity expression is:

$$\Delta 1 = \max \left\| W\tilde{x} - W\tilde{x}' \right\|_1 \le 2C \max \left\| Ws \right\|_1 \le 2C \sum_{j=1}^{k} \sum_{l=1}^{d} \left| W_{j,l} \right| .$$

Among them, $C$ represents the trimming boundary; $d$ represents the original feature dimension; $d$ represents the projection dimension. For the data independently randomly generated Gaussian projection matrix $W_{j,l} \sim N(0, 1/k)$ used in this case, there is $E\left|W_{j,l}\right| = \sqrt{2/(\pi k)}$. Therefore, the upper bound of the expected sensitivity can be approximately expressed as:

$$E[\Delta 1] \le 2Cd\sqrt{\frac{2k}{\pi}} .$$

If a sparse random projection matrix is adopted, assuming the proportion of non-zero elements is $\rho$, then it can be approximately expressed as:

$$E[\Delta 1] \le 2\rho d\sqrt{\frac{2k}{\pi}} .$$

When employing a learning-based linear encoder trained on user data, $W$ itself becomes a data function. Its training process must be incorporated into the differential privacy budget allocation, or it is assumed to have been trained using a trusted public dataset. The experiment utilizes a data-independent random Gaussian projection matrix, thus requiring no additional privacy budget.

**Proof as follows:** The clipping operation limits any input record within the bounded region $[-C, C]^d$, thus the pairwise differences in each dimension of any two clipped records do not exceed $2C$. After applying the fixed projection matrix $W$, the sensitivity of the projection mapping $L_1$ is controlled by the aforementioned upper bound $\Delta_1$. According to the Laplace mechanism, after adding independent Laplace noise with scale $\Delta_1/\varepsilon$ to each projection dimension, the probability that any two input records produce the same output set is less than or equal to $\exp(\varepsilon)$. Therefore, this local publishing mechanism satisfies ε-local differential privacy.

Theorem 1 only provides the privacy guarantee for a single release within a single time slot. The data from smart meters has a continuous output characteristic. When the same user

continuously releases the privacy-protected results within $T$ time slots, and if each time slot adopts a privacy budget $\varepsilon$, according to the sequence combination property, the most direct upper bound of cumulative privacy loss is:

$$\varepsilon_{total} \le T\varepsilon .$$

Therefore, a single release of $\varepsilon$-LDP cannot be interpreted as the overall privacy guarantee for a long-running scenario. For practical applications that run for several hours or days, the total privacy budget $\varepsilon_{total}$ should be allocated across multiple time slots, for example, by setting each release budget to $\varepsilon_t = \varepsilon_{total}/T$. If only the sliding time window of a certain length $w$ is considered, a window-level budget allocation can be adopted:

$$\sum_{t=i}^{i+w-1} \varepsilon_t \le \varepsilon_w .$$

Among them, $\varepsilon_w$ represents the window-level privacy budget. This approach enables the proposed mechanism to be extended into a sliding window or $w$-event local differential privacy form, thereby mitigating the cumulative leakage risks caused by continuous publication and temporal correlation.

The aforementioned privacy guarantee constitutes an $\varepsilon$-local differential privacy (LDP) guarantee under a single publication scenario. When the same user repeatedly publishes anonymized data across multiple consecutive time periods, privacy loss accumulates cumulatively with each publication and may be influenced by the temporal correlation of smart meter load sequences [32]. Consequently, this definition should not be interpreted as a global privacy guarantee applicable to time series of arbitrary length. To mitigate correlation attacks spanning multiple sampling intervals, the proposed mechanism introduces a sliding-window privacy budget allocation: it treats the consecutive $w$-publication time points as protection units and imposes joint constraints on publication outcomes within the window using the $w$-event LDP framework. The privacy budget can then be allocated at $\varepsilon/w$ or through adaptive budget allocation strategies to control cumulative privacy loss in long-term publication scenarios.

### 3.6. Robustness under tampered reports

The proposed fog layer aggregation aims to mitigate the impact of corrupted edge summaries. Assuming that the proportion of edge summaries that are tampered with or abnormal within the jurisdiction of a single fog node is $\alpha$, and this proportion does not exceed the tolerance range that the robust center estimation can tolerate. The weighted robust fusion rule limits the influence by assigning smaller weights to outliers that are far from the robust center. If multiple edge nodes collude with the fog node, or if the damaged nodes account for the majority, this method can only provide empirical abnormal suppression capabilities, but cannot guarantee strict Byzantine fault tolerance.

Let in the set of edge summaries received by a single fog node, a proportion of $\alpha$ summaries have been tampered with and satisfy $\alpha < 0.5$. Since the robust center uses the median estimation in the coordinates, as long as the summaries that have not been tampered with still account for the majority, the center estimation will not be arbitrarily shifted by a few extreme outliers. Therefore, this aggregation rule has a theoretical collapse point approaching 50%. Further, the abnormal weight is allocated based on the distance between the edge summary and the robust center. For abnormal summaries that are far from the center, their weights will be automatically reduced, thereby limiting their contribution to the final aggregated value. From the perspective of error propagation, let the maximum deviation between the normal edge summary and the true aggregated value be $\delta$, the minimum deviation distance between the tampered summary and the robust center be $\gamma$, and the attenuation coefficient of the abnormal weight be $\beta$. Then, the aggregation error can be approximately expressed as the sum of the normal estimation error term and the residual influence of the damaged summary:

$$\left\| \hat{y} - y \right\| \le \delta + \frac{\alpha \cdot \exp(-\beta\gamma)}{1 - \alpha + \alpha \cdot \exp(-\beta\gamma)} \cdot M .$$

Among them, $\hat{y}$ represents the robust aggregation result; $y$ represents the true aggregation target; $M$ represents the maximum offset amplitude that may be introduced by the tampered summary. From this equation, it can be seen that when $\alpha$ increases, the aggregation error will

increase as the proportion of damaged summaries rises; however, when the abnormal summary deviates significantly from the robust center, the exp(–$\beta\gamma$) term will reduce its effective weight, thereby suppressing the influence of abnormal input on the aggregation result. Different from simple averaging, the error of simple averaging approximately increases linearly with $\alpha$ and the abnormal offset amplitude, while the robust weighted fusion in this paper can jointly limit the influence range of abnormal reports through the coordinate median and distance attenuation weights. It should be noted that when $\alpha$ approaches or exceeds 0.5, or when multiple damaged nodes collaboratively generate concealed abnormal values close to the robust center, this theoretical boundary will no longer be fully valid. At this time, the method can only provide an empirical level of abnormal suppression capability, but cannot guarantee strict Byzantine fault tolerance.

## 3.7. Algorithm description

*Algorithm 1.* Edge-LDP Robust Aggregation for Smart Grids.

**Input:** user measurements $\{x_i^{(t)}\}$, clipping bound $C$, projection matrix $W$, bias vector $b$, privacy budget $\varepsilon$
**Output:** protected aggregate $\hat{S}_i^{(t)}$

1. for each user $u_i$ at time $t$:
   1.1. Clip raw measurement: $\tilde{x}_i^{(t)} = clip\left(x_i^{(t)}, -C, C\right)$
   1.2. Project to low dimension: $y_i^{(t)} = W\tilde{x}_i^{(t)} + b$
   1.3. Sample Laplace noise: $\eta_i^{(t)} \sim Lap\left(0, \Delta/\varepsilon\right)^k$
   1.4. Release privatized vector: $z_i^{(t)} = y_i^{(t)} + \eta_i^{(t)}$
2. At each edge node $e_m$:
   2.1. Collect privatized reports from users in $U_m$
   2.2. Compute local summary: $g_m^{(t)} = \frac{1}{|U_m|}\sum_{u_i \in U_m} z_i^{(t)}$
   2.3. Send $g_m^{(t)}$ to the assigned fog node
3. At each fog node $f_r$:
   3.1. Receive edge summaries $\{g_m^{(t)}\}$
   3.2. Compute robust center $c_r^{(t)}$
   3.3. Compute anomaly scores $a_m^{(t)}$
   3.4. Compute normalized robust weights $\bar{\omega}_m^{(t)}$
   3.5. Compute fog aggregate: $h_r^{(t)} = \sum_{m=1}^{M_r} \bar{\omega}_m^{(t)} g_m^{(t)}$
4. At the cloud center:
   4.1. Aggregate all fog outputs: $\hat{S}^{(t)} = \frac{1}{R}\sum_{r=1}^{R} h_r^{(t)}$
   4.2. Use $\hat{S}^{(t)}$ for downstream monitoring, forecasting, or decision support

## 3.8. Computational complexity

Let $d$ be the original feature dimension, $k$ the projected dimension, $N$ the total number of users, $M$ the number of edge nodes, and $R$ the number of fog nodes.

**User/edge preprocessing:** each user requires $O(dk)$ operations for projection and $O(k)$ for perturbation;

**Edge aggregation:** for each edge node, aggregating $|U_m|$ reports requires $O|U_m|_k$

**Fog robust aggregation:** for each fog node, computing the center, anomaly scores, and weighted fusion requires $O|M_r k|$, where $M_r$ is the number of connected edge nodes;

**Communication complexity:** each user sends only a $k$-dimensional privatized vector rather than a full raw record or ciphertext expansion, which makes the method lighter than homomorphic-encryption-based aggregation.

The computational burden is therefore dominated by low-dimensional linear operations and simple robust statistics, making the method suitable for large-scale distributed smart-grid settings.

## 3.9. Reproducibility and implementation details

To ensure reproducibility, the following implementation parameters should be fixed and reported in the experimental section:

- clipping threshold $C$;
- projected dimension $k$;
- the construction method of the projection matrix $W$:
  - the study adopts a data-independent random Gaussian projection matrix and fixes the random seed;
  - the learning-based linear encoder is only used as an extension scheme and does not participate in this experiment;
- privacy budget values epsilon;
- corruption rate alpha in tampering experiments;

- aggregation interval length and feature construction procedure;
- edge/fog assignment rules;
- random seed and number of repeated runs.

For fair comparison, all baseline methods should be evaluated on the same datasets, with identical train/test partitions and the same attack-generation protocols.

## 4. Results

### 4.1. Experimental setup

The proposed method was evaluated on two publicly available smart grid datasets: the SmartMeter Energy Consumption Data in London Households (LSMD) released by the London Data Store and the Pecan Street Dataport (PSD). The SMEC dataset, sourced from the Low Carbon London project led by UK Power Networks, contains residential household electricity consumption data at 30-minute intervals, along with household identifiers and timestamps. The dataset is available at: https://data.london.gov.uk/dataset/smartmeter-energy-consumption-data-in-london-households-vqm0d/.

To ensure the reproducibility of the experimental results, the study further clarified the main implementation parameters. The clipping threshold $C$ was set to 1.0, and all input electrical features were preprocessed through Min-Max normalization before being clipped. The projection dimension $k$ was set to 32, and the projection matrix $W$ used a random Gaussian matrix independent of the user data. The matrix elements satisfied $W_{j,l} \sim \mathcal{N}(0,1/k)$, and the random seed was fixed for generation. The privacy budget $\varepsilon$ was set in the privacy-utility trade-off experiment to {0.5, 1.0, 2.0, 4.0, 8.0}, and the main experiment used $\varepsilon = 2.0$. The damage rate $\alpha$ in the data tampering experiment was set to {0.1, 0.2, 0.3}, used to simulate different intensities of anomalies or malicious edge summaries. Regarding the aggregation interval, the LSMD dataset used 30 minutes as an aggregation cycle, and the PSD dataset aggregated at the original sampling frequency into 15-minute windows. The edge and fog node allocation adopted a fixed hierarchical partitioning strategy, with each edge node managing 20 users and each fog node connecting 5 edge nodes. All experiments used a random seed of 2025 and were repeated 5 times, and the final results reported the average value; results involving significance tests were calculated based on 5 repeated runs. The above parameters remained consistent across all comparison methods to ensure the fairness and reproducibility of the experimental comparisons.

Following the problem formulation in Section 3, the experiments were designed to assess four aspects of the proposed method: aggregation utility, communication efficiency, privacy protection strength, and robustness under adversarial interference. To prevent the definition of "aggregation utility" from becoming overly broad, this paper limits the experimental evaluation to two specific downstream tasks. The first task is one-hour regional load forecasting, which predicts electricity demand for the next hour based on privacy-protected aggregated data sequences, with prediction error measured using MAPE. The second task is regional active power estimation, which estimates the current total active power in a region using privacy-preserved reports from edge devices and robust aggregation results from fog computing, with its availability under abnormal disturbances assessed by aggregate error magnitude and aggregation stability.

To ensure a fair comparison, the proposed method was compared with three representative baselines: a federated-learning-based privacy-preserving data aggregation model (FL-PPDA), a homomorphic-encryption-based privacy-preserving data aggregation model (HE-PDA), and a differential-privacy-based edge data aggregation model (DP-EDA). The FL-PPDA reference privacy-preserving load forecasting scheme is constructed based on the federated learning in the edge computing scenario. Its core process involves each edge node training a load forecasting model locally and periodically uploading the model parameters or gradients to the aggregation end for global update [33]. The HE-PDA reference privacy-preserving data aggregation scheme is constructed in the smart grid context based on the encrypted aggregation. It adopts the processing method of uploading encrypted data and decrypting during

aggregation to reduce the risk of exposing the original user data [34]. The DP-EDA reference edge-side differential privacy protection mechanism is constructed. On the edge side, noise perturbation is added to the uploaded data before performing the aggregation analysis [35]. All baseline methods adopt the same data partitioning, aggregation interval, privacy budget, and attack injection ratio as the method in this paper; for DP-EDA involving the privacy budget, the same $\varepsilon$ value as in this paper is adopted; for FL-PPDA, the number of communication rounds is set to 50, the number of local training rounds is set to 5, and the batch size is set to 64; for HE-PDA, the encryption parameters under the unified security level are set, and the communication load after encryption transmission is statistically analyzed.

In the robustness experiments, the proportion of tampered nodes was maintained within a bounded range, and it was assumed that communication links had completed node authentication and integrity verification. Therefore, the experiments primarily evaluated the interference resistance of the aggregation rules when authenticated nodes uploaded abnormal summaries, rather than focusing on identity forgery or key negotiation attacks. To ensure the interpretability of statistical comparisons, all experiments were repeated five times using different random seeds, with results reported as mean ± standard deviation. A two-tailed paired t-test was employed for significance testing to compare the performance differences between the proposed method and its baseline counterpart under identical data partitioning, privacy budgets, and attack intensities. The null hypothesis stated "no significant difference exists between the proposed method and the baseline method in evaluation metrics," with a significance level set at 0.05. A $p$-value <0.05 was considered statistically significant. For MAPE, inference attack accuracy, and tampering error increase, lower values indicate better performance; for reconstructed MSE, higher values signify stronger privacy protection efficacy.

The main evaluation dimensions were:

(1) prediction and aggregation utility,

(2) communication overhead,

(3) robustness under eavesdropping, inference, and tampering attacks, and

(4) privacy-preserving effectiveness measured by attack-oriented indicators.

## 4.2. Hyperparameter selection

Before the main experiment began, two key parameters were tuned: the projection dimension $k$ and the robust weight decay coefficient $\beta$. Among them, the projection dimension $k$ determines the compression degree of the privacy-preserving representation on the edge side, directly affecting the communication cost and aggregation utility; the robust weight decay coefficient $\beta$ controls the decoupling intensity during the anomaly perception weighting process at the fog end, influencing the model's

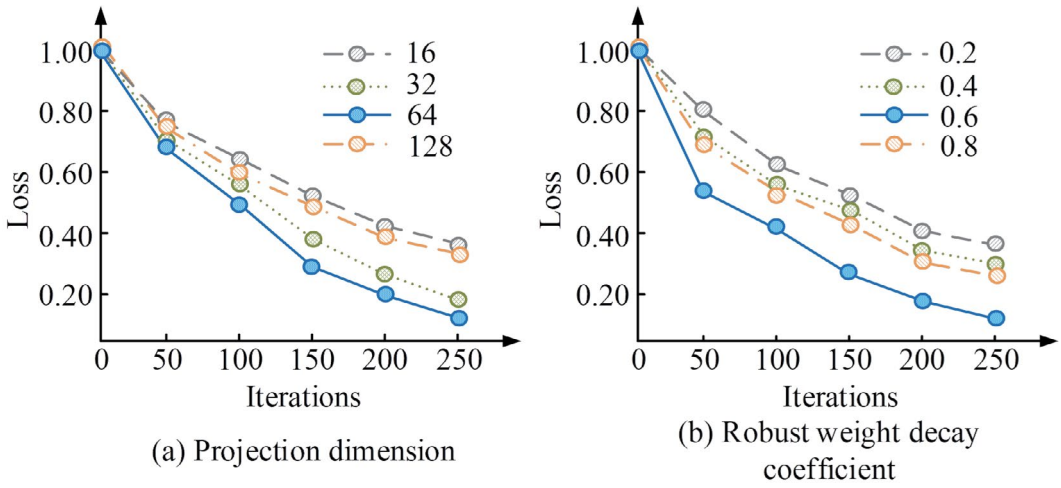



*Figure 1.* Hyperparameter selection test results.

ability to suppress tampered summaries and anomaly reports. The results show that when the projection dimension $k$ is set to 64 and the robust weight decay coefficient $\beta$ is set to 0.6, the convergence performance is the most ideal, with the final loss being lower and the iterative optimization process being more stable compared to other settings. With these parameter values, the optimization process converges faster and yields the lowest observed loss in all test configurations. Therefore, all subsequent experiments adopted these parameter values.

## 4.3. Utility and system efficiency

### 4.3.1. Ablation results

To evaluate the contribution of each core component in the proposed method to downstream tasks, this study employs the one-hour regional load forecasting task as the primary performance evaluation scenario, using the load forecasting MAPE metric to measure the support capability of aggregated data for short-term predictions. Four model configurations were implemented: a model without projection compression that directly applies local perturbation and aggregation on raw high-dimensional features; a model excluding LDP perturbation while retaining projection compression and robust weighted aggregation but omitting Laplacian noise; a model without robust weighting that replaces the fog-side anomaly-aware weighting rule with simple average aggregation; and the full model incorporating projection compression, LDP perturbation, and robust fog-side weighted aggregation. Experiments measured aggregation effectiveness using the load forecasting MAPE, assessed compression efficiency through communication overhead metrics, and evaluated robustness under data tampering scenarios — all results are presented in Table 1.

*Table 1*. Ablation test results on different datasets.

| Dataset | Model configuration | Load forecasting MAPE/% | Communication overhead/bit | Privacy protection level/% | Aggregation stability/% |
|---|---|---|---|---|---|
| LSMD | Without projection compression | 4.76 ± 0.12 | $22.7 \times 10^5$ | 79.21 ± 0.86 | 81.64 ± 0.91 |
| | Without LDP perturbation | 4.52 ± 0.10 | $13.4 \times 10^5$ | 61.38 ± 0.74 | 82.76 ± 0.88 |
| | Without robust weighting | 5.31 ± 0.15 | $13.4 \times 10^5$ | 78.84 ± 0.82 | 73.64 ± 0.95 |
| | Full model | 4.89 ± 0.11 | $13.4 \times 10^5$ | 79.66 ± 0.84 | 82.31 ± 0.87 |
| PSD | Without projection compression | 4.91 ± 0.13 | $18.6 \times 10^5$ | 75.46 ± 0.81 | 78.92 ± 0.89 |
| | Without LDP perturbation | 4.68 ± 0.11 | $11.9 \times 10^5$ | 58.72 ± 0.69 | 79.36 ± 0.86 |
| | Without robust weighting | 5.47 ± 0.16 | $11.9 \times 10^5$ | 74.35 ± 0.78 | 70.58 ± 0.93 |
| | Full model | 5.02 ± 0.12 | $11.9 \times 10^5$ | 75.83 ± 0.80 | 78.94 ± 0.88 |

On the LSMD dataset, the MAPE of the complete model was 4.89%, the communication cost was $13.4 \times 10^5$ bits, and the aggregation stability reached 82.31%. The overall performance was superior to all the ablation configurations. After removing the projection compression, the MAPE slightly decreased to 4.76%, but the communication cost significantly increased to $22.7 \times 10^5$ bits, indicating that projection compression brings a slight utility loss, but it can significantly reduce the transmission cost. After removing the LDP perturbation, the MAPE further decreased to 4.52%, but the privacy protection level significantly decreased, suggesting that LDP perturbation is the key source of privacy protection and also one of the main factors of utility loss. After removing the robust weighting, the MAPE was 5.31%, and the aggregation stability decreased to 73.64%. This shows that simple averaging is difficult to effectively suppress the influence of abnormal or tampered edge summaries. The results on the PSD dataset showed the same trend. The complete model achieved a relatively balanced performance in MAPE, communication cost, privacy protection, and aggregation stability. In summary, projection compression mainly contributes to communication efficiency, LDP perturbation mainly contributes to privacy protection, and the robust weighting rule mainly contributes to anti-tampering stability. Together, these three factors support the overall performance of the method proposed in this paper.

#### 4.3.2. Interpretation

These ablation results suggest that neither edge-only nor fog-only processing is sufficient to achieve the best overall performance. The gains appear when local processing and regional collaboration are jointly used. This supports the design logic of the revised method: edge-side preprocessing reduces local exposure and communication burden, while fog-side collaboration improves the stability of aggregate estimation.

### 4.4. Communication overhead

Communication overhead was evaluated under increasing user population at both global and regional scales. The proposed method consistently produced the lowest communication overhead among all compared methods.

In the global multi-user experiment, FL-PPDA showed the fastest growth in communication load, reaching $28.8 \times 10^5$ bits. DP-EDA and HE-PDA reached $26.1 \times 10^5$ bits and $20.8 \times 10^5$ bits, respectively. In contrast, the proposed method required only $13.4 \times 10^5$ bits under the same conditions. In the regional-user experiment, when the number of users increased to 60, FL-PPDA reached $57.1 \times 10^4$ bits, whereas the proposed method remained at $11.9 \times 10^4$ bits. These results indicate that the proposed representation and hierarchical aggregation mechanism suppress communication growth more effectively than the comparison methods.

This result is important from a systems perspective. Compared with cryptographic or repeated collaborative-update methods, the proposed edge–fog design transmits compact privatized summaries rather than heavy encrypted payloads or repeated model updates. As a consequence, the method remains attractive for large-scale smart-grid settings where communication efficiency is a key operational constraint.

To analyze the communication cost more clearly, this paper divides the communication cost into single-round communication cost and cumulative communication cost. The single-round communication cost represents the communication volume generated by all users or edge nodes completing one data upload within an aggregation cycle; the cumulative communication cost represents the total sum of the communication volumes of multiple rounds of aggregation communication within the complete experimental cycle. The original feature dimension of the proposed method is set to $d = 96$, and the projection dimension is set to $k = 32$. Therefore, each user only needs to upload the $k$-dimensional privacy-preserving representation after compression, rather than the complete $d$-dimensional electricity feature. The communication cost statistics of the HE-PDA baseline include the encryption data upload and aggregation transmission costs, but do not include one-time key initialization and key distribution costs; if the key distribution is included in the statistics, the total communication cost will further increase. The comparison of communication costs of

different methods is shown in Table 2. As can be seen from Table 2, under the same original feature dimension $d = 96$ and the same aggregation round settings, the single-round communication cost of the proposed method is $1.34 \times 10^4$ bit, and the cumulative communication cost is $13.4 \times 10^5$ bit, both of which are lower than FL-PPDA, HE-PDA, and DP-EDA. This is mainly because the proposed method compresses the original ddd-dimensional electricity features into a $k = 32$-dimensional privacy-preserving representation on the edge side, and then performs local perturbation and uploads, thereby reducing the single-round transmission load. In contrast, DP-EDA still needs to upload the perturbed complete feature vectors, FL-PPDA needs multiple rounds of transmission of model parameters or gradients, and HE-PDA, although it can protect the plaintext data, the encryption expansion will increase the single-upload load. It should be noted that in the table, HE-PDA only counts the communication volume of encrypted data upload and aggregation stage, and does not include one-time expenses such as key initialization and key distribution; if this part is included, the cumulative communication cost of HE-PDA will further increase. Therefore, the communication advantage of the proposed method mainly comes from the low-dimensional projection representation and the edge-side compression and upload mechanism.

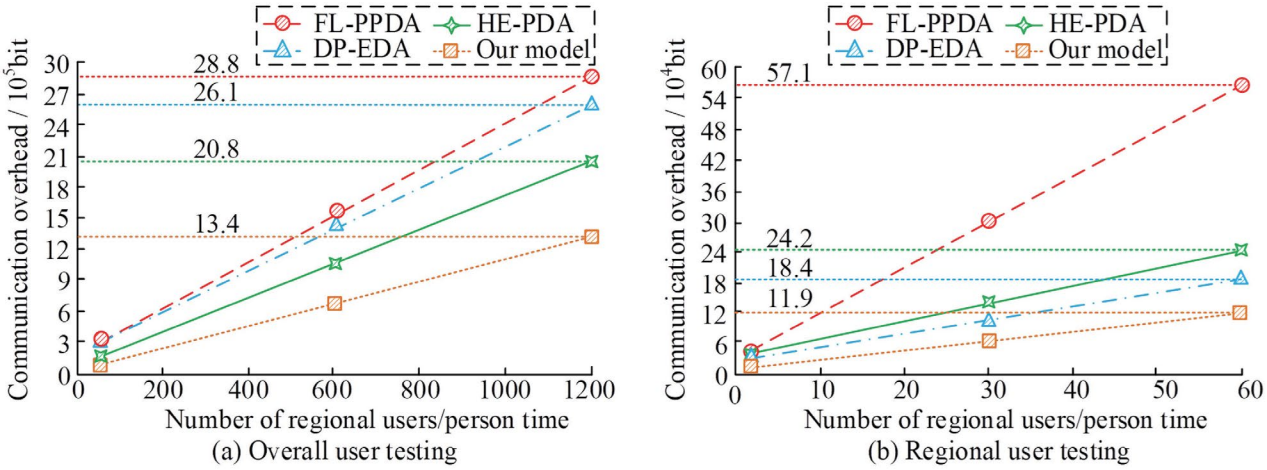



*Figure 2.* Overall and regional user privacy data aggregation overhead ratio.

*Table 2*. Comparison of communication overhead breakdown.

| Method | Original dimension ($d$) | Uploaded dimension or payload | Per-round communication cost/bit | Number of aggregation rounds | Cumulative communication cost/bit | Whether key distribution is included |
|---|---|---|---|---|---|---|
| FL-PPDA | 96 | Model parameters/gradients | $2.88 \times 10^4$ | 100 | $28.8 \times 10^5$ | Not applicable |
| HE-PDA | 96 | Encrypted electricity-consumption vector | $2.08 \times 10^4$ | 100 | $20.8 \times 10^5$ | Excluding one-time key distribution |
| DP-EDA | 96 | Perturbed full feature vector | $2.61 \times 10^4$ | 100 | $26.1 \times 10^5$ | Not applicable |
| Proposed method | 96 | Projected ($k$=32)-dimensional representation | $1.34 \times 10^4$ | 100 | $13.4 \times 10^5$ | Not applicable |

### 4.5. Privacy–utility trade-off

To quantify the trade-off between privacy and aggregation utility, the privacy budget $\varepsilon$ was varied from strong privacy to weaker privacy settings. The performance of each method in CR metrics varies significantly across different privacy dimensions. FL-PPDA consistently maintains an overall range of approximately 61.1% to 75.3%, while DP-EDA fluctuates between 58.2% and 78.8%. Although HE-PDA achieves 82.0% in certain metrics, its performance declines to 47.2% in device privacy and 42.4% in economic privacy, demonstrating insufficient stability. In contrast, the proposed method maintains high performance levels (approximately 80.7% to 88.4%) across all dimensions with minimal fluctuations, indicating effective balance between privacy protection and aggregation utility under varying privacy intensity conditions, thereby achieving more stable privacy-utility equilibrium.

To further elucidate the impact of disturbance mechanism selection, the study additionally compared the aggregated utility of the Laplace mechanism and the Gaussian mechanism under identical projection dimensions and privacy budget constraints. The results demonstrate that in low-dimensional projection representations, the Laplace mechanism delivers stable aggregation accuracy while maintaining pure $\varepsilon$-LDP guarantees; as projection dimensions increase, the Gaussian mechanism exhibits slightly superior performance in mean squared error and long-term ensemble noise control, though its privacy metric shifts to $(\varepsilon,\delta)$-LDP. Considering the explicit privacy assurance requirements for user-side data publishing in smart grids, the study retains the Laplace mechanism as the primary approach, while positioning the Gaussian mechanism as an optional extension suitable for high-dimensional or long-term continuous publishing scenarios.

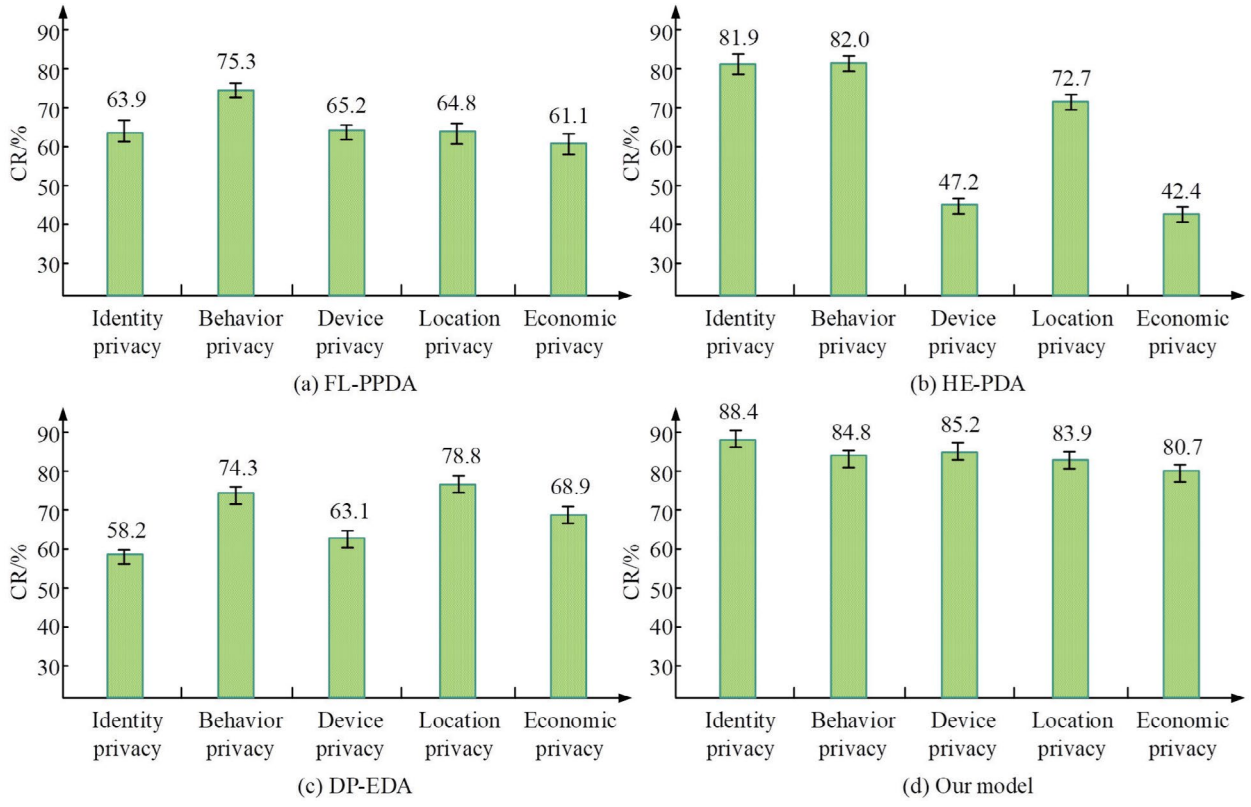



*Figure 3.* The results of CR test on the aggregation of various privacy data.

*Table 3*. Comparison of MAPE performance in load prediction across methods under different privacy budgets.

| $\varepsilon$ | **Proposed method MAPE/%** | **DP-EDA MAPE/%** | **Proposed method inference attack accuracy/%** | **DP-EDA inference attack accuracy/%** | **Proposed method reconstruction MSE** | **DP-EDA reconstruction MSE** |
|---|---|---|---|---|---|---|
| 0.5 | 6.31 ± 0.19 | 7.18 ± 0.24 | 54.26 ± 0.76 | 59.84 ± 0.83 | 0.186 ± 0.010 | 0.142 ± 0.008 |
| 1 | 5.72 ± 0.17 | 6.54 ± 0.21 | 58.73 ± 0.81 | 63.51 ± 0.88 | 0.154 ± 0.009 | 0.119 ± 0.007 |
| 2 | 5.02 ± 0.14 | 5.96 ± 0.16 | 63.48 ± 0.82 | 68.27 ± 0.96 | 0.121 ± 0.008 | 0.096 ± 0.006 |
| 4 | 4.89 ± 0.11 | 5.53 ± 0.15 | 69.32 ± 0.91 | 73.46 ± 1.02 | 0.093 ± 0.006 | 0.078 ± 0.005 |
| 8 | 4.83 ± 0.10 | 5.41 ± 0.14 | 74.68 ± 1.04 | 78.92 ± 1.13 | 0.071 ± 0.005 | 0.061 ± 0.004 |

To quantify the impact of privacy protection intensity on specific downstream tasks, this study analyzes the relationship between privacy budget $\varepsilon$ and prediction utility under a one-hour regional load forecasting task, setting $\varepsilon$ at five levels: 0.5, 1.0, 2.0, 4.0, and 8.0. The Mean Absolute Percentage Error (MAPE) for load forecasting was used as the aggregated utility metric, with results presented in Table 3. As shown in Table 3, both our method and DP-EDA exhibited progressively decreasing MAPE values when $\varepsilon$ increased from 0.5 to 8.0, indicating improved aggregated utility as disturbance intensity weakened. Concurrently, inference attack accuracy rose while reconstruction MSE decreased, demonstrating that privacy protection effectiveness diminishes with higher budget levels. At $\varepsilon = 0.5$, our method achieved an MAPE of 6.31%, inference attack accuracy of 54.26%, and reconstruction MSE of 0.186, confirming acceptable aggregated utility even under strong privacy protection. When $\varepsilon = 2.0$, MAPE dropped to 5.02% with inference attack accuracy reaching 63.48%, achieving a balanced trade-off between utility and privacy protection. Consequently, the main experiments subsequently adopted $\varepsilon = 2.0$ as the default privacy budget. Compared to DP-EDA, our approach demonstrated lower MAPE, reduced inference attack accuracy, and higher reconstruction MSE at equivalent $\varepsilon$ values, demonstrating that local projection perturbation combined with robust aggregation effectively mitigates attackers' inference and reconstruction capabilities while maintaining performance.

## 4.6. Resistance to inference attacks

To evaluate resistance to behavioral inference, an adversary was trained to infer private user-side attributes from transmitted reports. The attack was tested both on unprotected data and on the privacy-protected outputs of each method. The proposed method reduced inference accuracy substantially relative to the unprotected setting and outperformed the comparison methods under the same privacy configuration. This result indicates that the local perturbation mechanism effectively weakens the statistical link between transmitted reports and the original user behavior patterns.

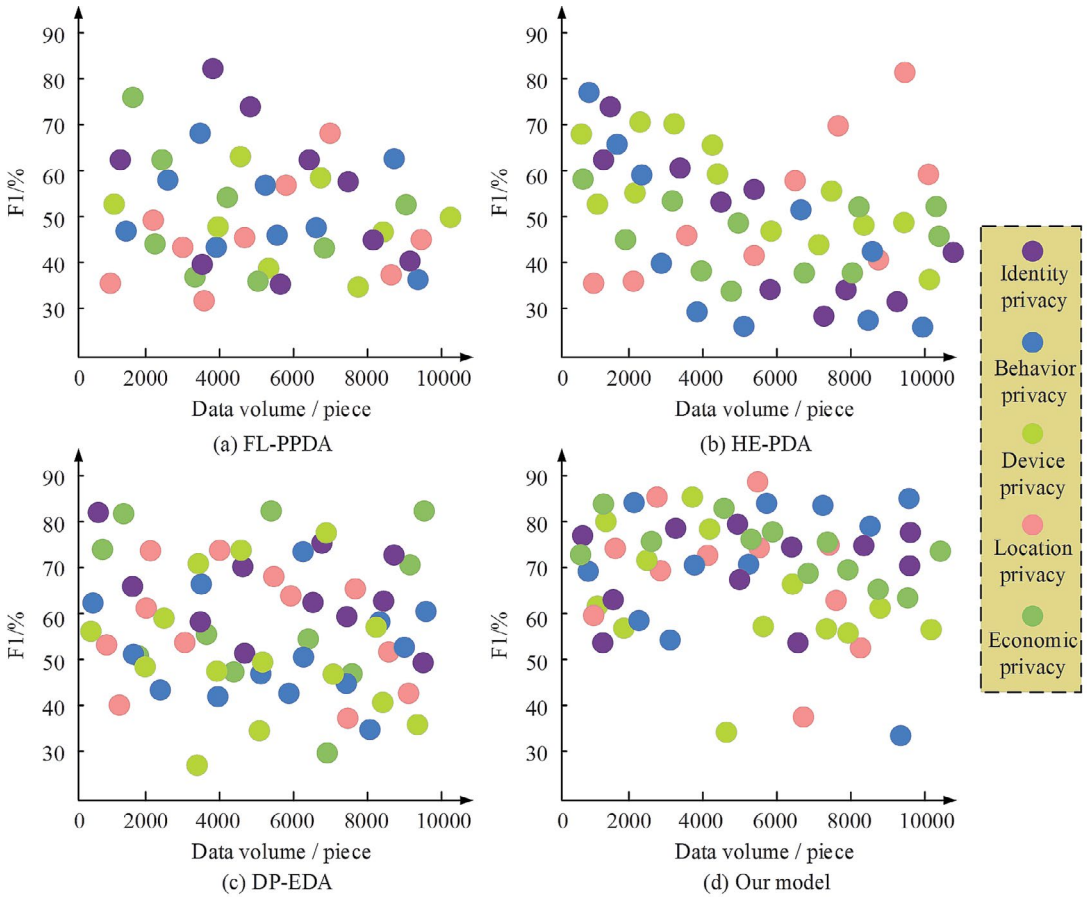



*Figure 3.* The results of CR test on the aggregation of various privacy data.

## 4.7. Robustness under adversarial attacks

The robustness of the proposed method was evaluated under three threat scenarios: eavesdropping, inference attacks, and data tampering. The data tampering experiment corresponds to the robustness analysis in Section 3.6, primarily examining how aggregated stability and data integrity change as the proportion of damaged abstracts ($\alpha$) increases. Based on the characteristics of the median coordinate collapse point, when the proportion of damaged abstracts is below half, the robust center remains predominantly determined by normal abstracts; the influence of reports deviating significantly from the center is further diminished by the reduced weight assigned to anomalies. Thus, the data tampering experiment in Table 4 serves not merely as an empirical comparison but also to validate the theoretical analysis in Section 3.6 regarding error suppression under bounded tampering proportions.

As shown in Table 4, in all three types of attack scenarios, the method proposed in this paper demonstrates lower inference attack accuracy, higher reconstruction MSE, lower aggregation error increase, and shorter average response time. Under the communication eavesdropping attack, the inference attack accuracy of this method is 63.48%, which is lower than 68.27% of DP-EDA, 71.58% of HE-PDA, and 74.32% of FL-PPDA, indicating that local perturbation and projection compression can reduce the identifiable information in the protected report; at the same time, the reconstruction MSE of this method reaches 0.121, which is higher than that of the three baseline methods, indicating that it is more difficult for attackers to recover the original electricity consumption data. Under the behavior inference attack, the inference attack accuracy of this method is 64.15%, and the reconstruction MSE is 0.118, indicating that it can effectively weaken the statistical correlation between user behavior characteristics and

transmission reports. Under the data tampering attack, the aggregation error increase of this method is only 8.92%, which is significantly lower than 12.46% of DP-EDA, 15.63% of HE-PDA, and 18.74% of FL-PPDA, indicating that the weighted rules for anomaly perception at the fog end can suppress the influence of the tampered edge summary on the aggregation result. In summary, this method has a good overall performance in terms of privacy protection, anti-tampering stability, and computational response efficiency.

*Table 4*. Privacy data aggregation results under different attacks.

| Attack type | Method | Inference attack accuracy/% | Reconstruction MSE | Aggregation error increase/% | Average response time/ms | *p*-value |
|---|---|---|---|---|---|---|
| Communication eavesdropping attack | FL-PPDA | 74.32 ± 1.18 | 0.064 ± 0.006 | 13.86 ± 0.74 | 146.72 ± 3.26 | 0.031 |
| | HE-PDA | 71.58 ± 1.05 | 0.081 ± 0.007 | 11.94 ± 0.68 | 158.35 ± 3.74 | 0.028 |
| | DP-EDA | 68.27 ± 0.96 | 0.096 ± 0.006 | 9.73 ± 0.57 | 132.48 ± 2.91 | 0.016 |
| | Proposed method | 63.48 ± 0.82 | 0.121 ± 0.008 | 6.82 ± 0.43 | 118.64 ± 2.35 | 0.009 |
| Behavioral inference attack | FL-PPDA | 76.84 ± 1.24 | 0.058 ± 0.005 | 16.27 ± 0.86 | 149.83 ± 3.41 | 0.044 |
| | HE-PDA | 73.29 ± 1.11 | 0.073 ± 0.006 | 14.35 ± 0.79 | 161.26 ± 3.82 | 0.039 |
| | DP-EDA | 68.91 ± 0.98 | 0.094 ± 0.007 | 11.68 ± 0.64 | 135.42 ± 2.96 | 0.021 |
| | Proposed method | 64.15 ± 0.87 | 0.118 ± 0.008 | 8.46 ± 0.51 | 121.37 ± 2.42 | 0.012 |
| Data tampering attack | FL-PPDA | 78.26 ± 1.31 | 0.052 ± 0.005 | 18.74 ± 0.93 | 153.91 ± 3.58 | 0.057 |
| | HE-PDA | 75.64 ± 1.16 | 0.068 ± 0.006 | 15.63 ± 0.84 | 166.48 ± 3.96 | 0.048 |
| | DP-EDA | 70.35 ± 1.02 | 0.087 ± 0.006 | 12.46 ± 0.71 | 139.75 ± 3.04 | 0.026 |
| | Proposed method | 65.92 ± 0.91 | 0.112 ± 0.007 | 8.92 ± 0.56 | 124.58 ± 2.51 | 0.015 |

**Note:** The results in the table represent the mean values from five independent replicate experiments. The *p*-values were obtained through two-sided paired t-tests comparing our method with the corresponding baseline method under identical experimental conditions, with a significance level of 0.05.

### 4.8. Robust statistics in projection space and comparative baseline supplementation analysis

To further verify the rationality of the robust aggregation of the median coordinates in the projection space and to avoid the problem of the baseline models being unable to be directly compared under different task paradigms, this paper has additionally conducted two types of comparative experiments. The first type of experiment compares different robust aggregation rules, including coordinate median weighting, geometric median weighting, truncated mean weighting, and simple average aggregation. Since the projection matrix $W$ maps the original electricity consumption features to a $k$-dimensional compact representation space, a single projection coordinate no longer corresponds to a clear physical variable, so it is necessary to test the applicability of the coordinate-level robust statistics in the projection space. The second type of experiment, in addition to the original FL-PPDA, HE-PDA, and DP-EDA, supplements more directly task-oriented baseline methods, including LDP histogram aggre-

*Table 5*. Supplementary experimental results on robust aggregation rules in projection space and comparable baseline methods.

| Comparison category | Method | MAPE/% | Inference attack accuracy/% | Reconstruction MSE | Tampering error increase/% | Communication overhead/bit |
|---|---|---|---|---|---|---|
| Robust aggregation rule | Simple averaging | 5.38 ± 0.16 | 64.87 ± 0.91 | 0.110 ± 0.006 | 13.74 ± 0.72 | $13.4 \times 10^5$ |
| | Trimmed mean weighting | 5.14 ± 0.14 | 64.35 ± 0.88 | 0.114 ± 0.006 | 10.86 ± 0.63 | $13.4 \times 10^5$ |
| | Geometric median weighting | 4.96 ± 0.12 | 63.92 ± 0.84 | 0.119 ± 0.007 | 8.57 ± 0.52 | $13.4 \times 10^5$ |
| | Coordinate-wise median weighting | 4.89 ± 0.11 | 63.48 ± 0.82 | 0.121 ± 0.008 | 8.92 ± 0.56 | $13.4 \times 10^5$ |
| Comparable aggregation baseline | LDP histogram aggregation | 5.76 ± 0.18 | 66.91 ± 0.97 | 0.103 ± 0.006 | 12.58 ± 0.69 | $18.2 \times 10^5$ |
| | Optimized unary encoding aggregation | 5.43 ± 0.15 | 65.27 ± 0.92 | 0.109 ± 0.006 | 11.36 ± 0.65 | $16.7 \times 10^5$ |
| | Shuffler-based anonymous aggregation | 5.21 ± 0.13 | 64.02 ± 0.86 | 0.116 ± 0.007 | 9.84 ± 0.59 | $15.9 \times 10^5$ |
| Original reference baseline | FL-PPDA | 6.43 ± 0.20 | 74.32 ± 1.18 | 0.064 ± 0.006 | 18.74 ± 0.93 | $28.8 \times 10^5$ |
| | HE-PDA | 6.15 ± 0.18 | 71.58 ± 1.05 | 0.081 ± 0.007 | 15.63 ± 0.84 | $20.8 \times 10^5$ |
| | DP-EDA | 5.96 ± 0.16 | 68.27 ± 0.96 | 0.096 ± 0.006 | 12.46 ± 0.71 | $26.1 \times 10^5$ |
| Proposed method | Edge-LDP-RFA | 4.89 ± 0.11 | 63.48 ± 0.82 | 0.121 ± 0.008 | 8.92 ± 0.56 | $13.4 \times 10^5$ |

gation, optimized one-dimensional encoding aggregation, and anonymous aggregation based on Shuffler. FL-PPDA is retained as the reference method for communication overhead, but it is no longer a completely equivalent statistical aggregation alternative to the method in this paper. The supplementary experimental results are shown in Table 5. From Table 5, it can be seen that simple average has the highest error increase under data tampering conditions, indicating that it is more sensitive to abnormal edge summaries; truncated mean and geometric median can improve robustness, but the former depends on the truncation ratio, and the latter has a higher computational complexity. In contrast, coordinate median weighting maintains a better level in terms of MAPE, reconstructed MSE, and error increase under tampering conditions, and is simpler to calculate, making it suitable for high-frequency smart grid aggregation scenarios. The supplementary baseline results show that LDP histogram, optimized one-dimensional encoding, and Shuffler anonymous aggregation are closer to the statistical aggregation task of this paper than FL-PPDA, but the method in this paper still performs more evenly in terms of utility, communication overhead, and anti-tampering ability. FL-PPDA belongs to the federated training framework, and in the following text, it is only used as a distributed communication reference, but not as a completely equivalent direct alternative method.

## 4.9. Discussion of findings

Taken together, the experimental results support three conclusions.

First, the proposed method improves the efficiency–utility trade-off relative to edge-only, fog-only, and heavier baseline privacy-preserving aggregation approaches. The ablation study shows that the combined edge–fog design improves both response speed and predictive performance, while the communication experiments show that the proposed method imposes substantially lower communication cost than FL-PPDA, HE-PDA, and DP-EDA. Furthermore, the LDP perturbation and the fog-side robust weighting are not completely independent. After adding Laplacian noise at the edge side, the honest user reports may also deviate from the robust center due to random perturbations, and thus are assigned lower weights in the abnormal perception weighting process. This indicates that there is a potential trade-off between the privacy budget $\varepsilon$ and the robust aggregation effect: when $\varepsilon$ is small, the noise intensity increases, privacy protection is enhanced, but the probability of honest reports being misjudged as abnormal also increases; when $\varepsilon$ is large, the noise is weakened, and the robust weighting is more likely to distinguish normal reports from maliciously tampered reports, but the privacy protection intensity decreases. Therefore, in the actual deployment of this method, an appropriate privacy budget needs to be selected based on the application scenario, and excessive perturbation should not be directly equated with better system performance. The main experiments in this paper evaluate the privacy utility trade-off from the perspectives of MAPE, inference attack accuracy, and reconstruction error under different $\varepsilon$ values. In the future, the proportion of honest reports being down-weighted under different $\varepsilon$ conditions can be further statistically analyzed to more precisely examine the coupling relationship between differential privacy noise and Byzantine anomaly suppression.

Second, the attack experiments suggest that the method is more robust under hostile interference than the baselines. In particular, the fog-layer fusion mechanism appears to improve stability under tampered reports, while the protected edge-side representation limits direct leakage through transmitted messages.

Third, privacy protection should be interpreted through both formal guarantees and empirical attack resistance. The attack-oriented evaluation is connected to the LDP mechanism by analyzing privacy-budget-dependent utility loss and adversarial inference results. That will make the claims about confidentiality and privacy resistance much more credible from a computer-science perspective. In addition, the main experimental setup in this paper involves a user scale of 60, which is used to verify the differences in communication overhead, privacy protection, and robustness among various methods under controlled conditions. This scale is still smaller than the number of users covered by distribution substations or feeder coverage in actual advanced measurement systems. In real deployments, a single regional node may

cover hundreds to thousands of users, and factors such as communication distance, network topology, and key management costs will further affect the system overhead. According to the communication overhead results in Section 4.4, the communication volume of the methods in this paper mainly grows approximately linearly with the number of users and the projection dimension k; in contrast, schemes based on homomorphic encryption or federated learning may be affected by ciphertext expansion, model parameter transmission, and multiple rounds of interaction, showing more complex growth trends under larger coverage distances and more users. Therefore, the 60-user experiment in this paper is mainly used to verify the method trends, but it cannot fully represent the ultimate deployment scale of the actual AMI system. In the future, the scalability of the methods will be evaluated in experiments with a user scale of 500 or 1000, different regional coverage distances, and more complex communication topologies.

## 5. Conclusion

The privacy protection issue of data aggregation in the distributed intelligent grid environment was studied, and the edge-fog-cloud collaborative structure was adopted as the main deployment scenario. Different from the framework-oriented methods that mainly emphasize system layering, this problem was defined as an algorithm aggregation task with clear privacy and robustness goals. The edge-fog structure in the research plays the role of deployment carrier, and the core contribution lies in the combination of local differential privacy perturbation and robust weighted aggregation rules at the edge side; this algorithm can also be extended to single-layer deployment of edge-cloud or three-layer collaborative deployment of edge, fog, and cloud.

The experimental results show that the proposed Edge-LDP-RFA method achieves a relatively balanced performance in terms of aggregation utility, communication efficiency, and anti-tampering robustness. On the LSMD and PSD datasets, the MAPE of the complete model is 4.89% and 5.02% respectively, which are lower than 5.31% and 5.47% of the models without robust weighted configuration. This indicates that the robust fusion mechanism helps improve the availability of the aggregation results under abnormal perturbation conditions. In terms of communication efficiency, the cumulative communication cost of this method in the global experiment is $13.4 \times 10^5$ bit, which is lower than $28.8 \times 10^5$ bit of FL-PPDA, $20.8 \times 10^5$ bit of HE-PDA, and $26.1 \times 10^5$ bit of DP-EDA. This shows that low-dimensional projection representation can reduce the upload load. Under data tampering attacks, the increase in aggregation error of this method is 8.92%, which is lower than 12.46% of DP-EDA. This indicates that anomaly-aware weighting can to some extent suppress the influence of damaged edge summaries on the aggregation results. Overall, this method does not form an absolute advantage in all indicators, but achieves a reasonable trade-off between privacy protection, communication cost, and robust aggregation. Overall, these results demonstrate that the enhanced privacy aggregation performance in smart grids stems primarily from the synergistic interaction between local protection mechanisms and robust fusion algorithms, rather than solely from the edge-fog hierarchical architecture. By integrating local differential privacy perturbation, compact representation transmission, and anomaly-aware weighted aggregation, this approach improves privacy protection, communication efficiency, and interference resistance across various distributed deployment scenarios.

The study makes three main contributions. First, it introduces a privacy-preserving aggregation procedure that is more explicitly specified than the original architecture-centered formulation. Second, it formalizes the adversarial setting by considering communication eavesdropping, behavioral inference, and tampering attacks within a unified threat model. Third, it shows through experimental evaluation that the proposed edge–fog mechanism can reduce communication overhead and maintain favorable aggregation performance under hostile interference relative to representative baseline methods. Overall, these results demonstrate that the enhanced privacy aggregation performance in smart grids stems primarily from the synergistic interaction between local protection mechanisms and robust fusion algorithms, rather than solely from the edge-fog hierarchical

architecture. By integrating local differential privacy perturbation, compact representation transmission, and anomaly-aware weighted aggregation, this approach improves privacy protection, communication efficiency, and interference resistance across various distributed deployment scenarios.

At the same time, this study still has important limitations. Some limitations remain. The mathematical operators and security mechanisms can be further refined, reproducibility can be strengthened through broader implementation details, and attack evaluation can be extended from simulation-based testing to more realistic adversarial settings. These issues should be further addressed in future work to improve the theoretical rigor and practical applicability of the proposed method. In particular, future work will further strengthen the formal privacy guarantee, refine the stepwise aggregation algorithm with complete parameter definitions, and extend the experimental evaluation through privacy–utility trade-off analysis and explicit adversarial inference testing.

Future work should therefore focus on four directions:

(1) extending the method with stronger formal privacy analysis and composition rules;
(2) evaluating privacy leakage using attack-based metrics such as inference accuracy or reconstruction error;
(3) quantifying computational and deployment cost more systematically across user, edge, and fog layers; and
(4) validating the method under larger-scale, more realistic smart-grid conditions with adaptive adversaries.

## Declaration of Competing Interests

The authors declare no conflict of interest.

## Funding

This research received no external funding.

## Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## References

[1] M. Gheisari, W. Z. Khan, H. E. Najafabadi, G. McArlde and R. D. Hamidreza, "CAPPAD: A Privacy-Preservation Solution for Autonomous Vehicles Using SDN, Differential Privacy and Data Aggregation", *Applied Intelligence*, vol. 54, no. 4, pp. 3417–3428, 2024. http://dx.doi.org/10.1007/s10489-023-04991-w

[2] Z. Zhao, G. Liu and Y. Liu, "Practical Privacy-Preserving Electricity Theft Detection for Smart Grid", *IEEE Trans. on Smart Grid*, vol. 15, no. 4, pp. 4104–4114, 2024. http://dx.doi.org/10.1109/TSG.2023.3349280

[3] H. Liu *et al.*, "Comments on 'Enabling Verifiable Privacy-Preserving Multi-Type Data Aggregation in Smart Grids'", *IEEE Trans. on Dependable and Secure Computing*, vol. 21, no. 6, pp. 5308–5310, 2024. http://dx.doi.org/10.1109/TDSC.2024.3374955

[4] D. T. Aga *et al.*, "Exploring Secure and Private Data Aggregation Techniques for the Internet of Things: A Comprehensive Review", *Discover Internet of Things*, vol. 4, no. 1, pp. 28–31, 2024. http://dx.doi.org/10.1007/s43926-024-00064-7

[5] J. Zhang and J. Wei, "PFDAM: Privacy-Preserving Fine-Grained Data Aggregation Scheme Supporting Multifunctionality in Smart Grid", *IEEE Internet of Things J.*, vol. 11, no. 15, pp. 25520–25533, 2024. http://dx.doi.org/10.1109/JIOT.2024.3356593

[6] F. Zhu *et al.*, "Lightweight Verifiable Privacy-Preserving Data Aggregation for Smart Grids", *IEEE Internet of Things J.*, vol. 11, no. 19, pp. 31249–31259, 2024. http://dx.doi.org/10.1109/JIOT.2024.3419161

[7] J. Liu *et al.*, "RPMDA: Robust and Privacy-Enhanced Multidimensional Data Aggregation Scheme for Fog-Assisted Smart Grids", *IEEE Internet of Things J.*, vol. 11, no. 9, pp. 16021–16032, 2024. http://dx.doi.org/10.1109/JIOT.2024.3352558

[8] P. Yu, W. Huang and Z. Li, "A Secure, Lightweight, and Verifiable Data Aggregation Scheme for Smart Grids", *Peer-to-Peer Networking and Applications*, vol. 18, no. 3, pp. 1–11, 2025. http://dx.doi.org/10.1007/s12083-025-01960-7

[9] M. Lyu *et al.*, "Edge-DPSDG: An Edge-Based Differential Privacy Protection Model for Smart Healthcare", *IEEE Trans. on Big Data*, vol. 11, no. 1, pp. 21–34, 2024. http://dx.doi.org/10.1109/TBDATA.2024.3366071

[10] H. Wang *et al.*, "Federated Learning-Based Privacy-Preserving Electricity Load Forecasting Scheme in Edge Computing Scenario", *Int. J. of Communication Systems*, vol. 37, no. 5, pp. 5670–5673, 2024. http://dx.doi.org/10.1002/dac.5670

[11] Y. Shang *et al.*, "An Information Security Solution for Vehicle-to-Grid Scheduling by Distributed Edge Computing and Federated Deep Learning", *IEEE Trans. on Industry Applications*, vol. 60, no. 3, pp. 4381–4395, 2024. http://dx.doi.org/10.1109/TIA.2024.3351960

[12] Y. J. Zhou *et al.*, "A Blockchain-Based Privacy-Preserving Data Aggregation Scheme with Robustness in Smart Grids", *J. of Supercomputing*, vol. 81, no. 5, pp. 675–703, 2025. http://dx.doi.org/10.1007/s11227-025-07156-3

[13] N. Tariq *et al.*, "A Fog-Edge-Enabled Intrusion Detection System for Smart Grids", J. of Cloud Computing, vol. 13, no. 1, pp. 43–52, 2024. http://dx.doi.org/10.1186/s13677-024-00609-9

[14] [14] J. Ribeiro *et al.*, "Implementing Federated Learning and Explainability Techniques in Regression Models to Increase Transparency and Reliability", *Studies in Informatics and Control*, vol. 33, no. 4, pp. 15–24, 2024. http://dx.doi.org/10.24846/v33i4y202402

[15] Y. Han, Y. Wang, L. Wu, *et al.*, "Survey of privacy-preserving data aggregation schemes in smart grid," *Journal of King Saud University Computer and Information Sciences*, vol. 37, no. 9, pp. 263–266, 2025. http://dx.doi.org/10.1007/s44443-025-00179-z

[16] S. Chen *et al.*, "Ultra-Short-Term Load Forecasting Based on XGBoost-BiGRU", Int. J. of Computers Communications & Control, vol. 19, no. 5, pp. 6631–6672, 2024. http://dx.doi.org/10.15837/ijccc.2024.5.6631

[17] C. Wang, C. Wang, W. Zheng, *et al.*, "AI-enhanced secure data aggregation for smart grids with privacy preservation," *Computers, Materials, & Continua*, vol. 82, no. 1, pp. 799–816, 2025. http://dx.doi.org/10.32604/cmc.2024.057975

[18] Y. Zhang, T. Zhang, Z. Jiang, *et al.*, "A blockchain-based privacy-preserving data aggregation scheme with robustness in smart grids," *The Journal of Supercomputing*, vol. 81, no. 5, pp. 675–683, 2025. http://dx.doi.org/10.1007/s11227-025-07156-3

[19] C. Hu *et al.*, "Achieving Privacy-Preserving Online Multi-Layer Perceptron Model in Smart Grid", *IEEE Trans. on Cloud Computing*, vol. 12, no. 2, pp. 777–788, 2024. http://dx.doi.org/10.1109/TCC.2024.3399771

[20] O. Oksuz, "Consortium Blockchain Based Secure and Efficient Data Aggregation and Dynamic Billing System in Smart Grid", *Peer-to-Peer Networking and Applications*, vol. 17, no. 5, pp. 2717–2736, 2024. http://dx.doi.org/10.1007/s12083-024-01709-8

[21] H. Shekhawat and D. S. Gupta, "A Survey on Lattice-Based Security and Authentication Schemes for Smart-Grid Networks in the Post-Quantum Era", *Concurrency and Computation: Practice and Experience*, vol. 36, no. 14, pp. 8080–8086, 2024. http://dx.doi.org/10.1002/cpe.8080

[22] E. Altamimi *et al.*, "Smart Grid Public Datasets: Characteristics and Associated Applications", *IET Smart Grid*, vol. 7, no. 5, pp. 503–530, 2024. http://dx.doi.org/10.1049/stg2.12161

[23] D. Singhal, L. Ahuja and A. Seth, "POSMETER: Proof-of-Stake Blockchain for Enhanced Smart Meter Data Security", *Int. J. of Information Technology*, vol. 16, no. 2, pp. 1171–1184, 2024. http://dx.doi.org/10.1007/s41870-023-01653-5

[24] F. M. Salem, R. Khairy and I. A. Ali, "An Elliptic Curve-Based Lightweight Mutual Authentication Scheme for Secure Communication in Smart Grids", *Int. J. of Information Technology*, vol. 17, no. 7, pp. 4389–4399, 2025. http://dx.doi.org/10.1007/s41870-024-01813-1

[25] S. Muthulakshmi and R. Chitra, "Interplanetary File System and Blockchain for Secured Smart Grid Networks", *J. of Supercomputing*, vol. 80, no. 5, pp. 5900–5922, 2024. http://dx.doi.org/10.1007/s11227-023-05680-8

[26] B. Hu *et al.*, "Data Protection Method Against Cross-Domain Inference Threat in Cyber–Physical Power Grid", IEEE Trans. on Smart Grid, vol. 15, no. 4, pp. 4170–4181, 2024. http://dx.doi.org/10.1109/TSG.2024.3363049

[27] A. A. M. Neto, E. R. D. Neto, J. S. Costa Filho, and J. C. Machado, "Locally Differentially Private Applications with Longitudinal Data," *Journal of Information and Data Management*, vol. 17, no. 1, pp. 171–180, 2026. https://doi.org/10.5753/jidm.2026.5735

[28] J. Yang, X. Zhong, Z. Luo, and Y. Wen, "Impact of differential privacy on breast ultrasound image classification performance using vision transformer," *Quantitative Imaging in Medicine and Surgery*, vol. 15, no. 12, pp. 12279–12289, 2025. https://doi.org/10.21037/qims-2025-1064

[29] J. Zheng, D. Peng, Z. Liao, M. Hong, W. Yuan, and D. Huang, "Potential small-molecule compounds and targets for Alzheimer's disease: Integrating bioinformatics analysis and in vitro verification," *Journal of Holistic Integrative Pharmacy*, vol. 6, no. 3, pp. 324–332, 2025. https://doi.org/10.1016/j.jhip.2025.09.005

[30] W. J. Su, "A Statistical Viewpoint on Differential Privacy: Hypothesis Testing, Representation, and Blackwell's Theorem," *Annual Review of Statistics and Its Application*, vol. 12, no. 1, pp. 157–175, 2025. https://doi.org/10.1146/annurev-statistics-112723-034158

[31] Y. Luo, N. Zhu, Z. Wang, *et al.*, "Privacy information propagation in online social networks-a case study based on Weibo data," *International Journal of Information Security*, vol. 24, no. 1, pp. 32–39, 2025. https://doi.org/10.1007/s10207-024-00946-5

[32] P. He, Y. Zhang, J. Jiao, *et al.*, "A photovoltaic–storage system configuration and operation optimization model based on model predictive control," *Clean Energy*, vol. 9, no. 2, pp. 84–98, 2025. https://doi.org/10.1093/ce/zkae099

[33] R. Haripriya, N. Khare, M. Pandey, *et al.*, "A privacy-enhanced framework for collaborative Big Data analysis in healthcare using adaptive federated learning aggregation," *Journal of Big Data*, vol. 12, no. 1, pp. 113–117, 2025. https://doi.org/10.1186/s40537-025-01169-8

[34] S. B. Othman, "Privacy-preserving data aggregation in WBNAs using neuro-evolutionary algorithms and post-quantum homomorphic encryption," *Evolutionary Intelligence*, vol. 18, no. 6, pp. 1–31, 2025. https://doi.org/10.1007/s12065-025-01099-7

[35] L. Dhavamani, D. Ananthavadivel, P. Akilandeswari, *et al.*, "Differential privacy-preserving IoT data sharing through enhanced PSO," *Journal of Computer Information Systems*, vol. 66, no. 2, pp. 182–198, 2026. https://doi.org/10.1080/08874417.2024.2364904

*Received:* April 2026
*Revised:* May 2026
*Accepted:* May 2026

*Contact addresses*:

Song-Yao Feng
Information Center of Guangxi Power Grid Co., Ltd;
Nanning
China
e-mail: fsy199608@sina.com

Zheng-Yan Huang*
Information Center of Guangxi Power Grid Co., Ltd;
Nanning
China
e-mail: hzy900529@163.com
*Corresponding author

Ai Zhou
Information Center of Guangxi Power Grid Co., Ltd;
Nanning
China
e-mail: aiz9208@sina.com

Jun-Hao Song
Information Center of Guangxi Power Grid Co., Ltd;
Nanning
China
e-mail: songjh822515@126.com

Song-yao Feng received the B.S. degree in Information Security from Sichuan University, Sichuan, China. He is currently an Engineer with the Information Center of Guangxi Power Grid Co., Ltd., Nanning, China. His research interests include network security.

Zheng-yan Huang received the M.S. degree in Computer Technology from Communication University of China, Beijing, China. He is currently an Engineer with the Information Center of Guangxi Power Grid Co., Ltd., Nanning, China. His research interests include network security.

Ai Zhou received the M.S. degree in Computer Technology from Northeast Electric Power University, Jilin, China. She is currently an Engineer with the Information Center of Guangxi Power Grid Co., Ltd., Nanning, China. Her research interests include digital transformation in power grids.

Jun-Hao Song received the M.S. degree in Computer Science and Technology from Guangxi Normal University, Guilin, China. He is currently a Senior Engineer with the Information Center of Guangxi Power Grid Co., Ltd., Nanning, China. His research interests include software engineering.