A Review Of Fault-Tolerant Data Aggregation Schemes For Smart Grid Environment

Pooja Rani ¹ and Smruti Rekha Swain¹

¹Department of Computer Applications National Institute of Technology Kurukshetra

January 10, 2024

Abstract

The advent of smart grid in the power infrastructure of the nations has provisioned a proactive and highly efficient power management at the generation end and user end as well. The smart meters being a fundamental part of the grid system are responsible for the intermittent sharing of usage data that further assists the utility for decision making. However, the inclusion of digital technology accompanies some major challenges like data security, data loss and communication issues etc. Smart meters are lightweight devices hence susceptible to errors frequently and therefore can cause loss of data. If the unavailability of usage data is significantly high, this can be problematic for the utility in context with further power generation and distribution decisions. This article presents a comprehensive discussion of the fault-tolerant aggregation models for the smart grid environment. A thorough investigation of the most recent research on handling of unavailable data is provided. Finally, the paper identifies the major challenges and future research possibilities for an efficiently fault-tolerant aggregation in the smart grid environment.
A REVIEW OF FAULT-TOLERANT DATA AGGREGATION SCHEMES FOR SMART GRID ENVIRONMENT

A PREPRINT

Pooja Rani
Department of Computer Applications
National Institute of Technology
Kurukshetra
poojavats971993@gmail.com

Smruti Rekha Swain
Department of Computer Applications
National Institute of Technology
Kurukshetra
smruti.sai90@gmail.com

January 4, 2024

ABSTRACT

The advent of smart grid in the power infrastructure of the nations has provisioned a proactive and highly efficient power management at the generation end and user end as well. The smart meters being a fundamental part of the grid system are responsible for the intermittent sharing of usage data that further assists the utility for decision making. However, the inclusion of digital technology accompanies some major challenges like data security, data loss and communication issues etc. Smart meters are light-weight devices hence susceptible to errors frequently and therefore can cause loss of data. If the unavailability of usage data is significantly high, this can be problematic for the utility in context with further power generation and distribution decisions. This article presents a comprehensive discussion of the fault-tolerant aggregation models for the smart grid environment. A thorough investigation of the most recent research on handling of unavailable data is provided. Finally, the paper identifies the major challenges and future research possibilities for an efficiently fault-tolerant aggregation in the smart grid environment.

Keywords Data aggregation · Fault-tolerance · Data imputation · Missing Data

1 Introduction

With the incorporation of digital technology in power supply system, traditional grids are transformed into smart grids that allow two systems for bi-directional communication, electricity consumption data analysis, enhanced management of power distribution and provides self-healing capabilities to the grid system. The smart grids employ the digital technology to record the consumers’ power usage behaviours through smart sensors that help to manage power generation and supply efficiently [1–5]. A generic smart grid architecture includes a grid domain that generates, distributes and manages the electricity along with the service providers domain that include markets, operators and service providers to help consumers provide with tariff planning and other application based interfaces and the customer domain that includes house area networks with smart meters [6–9]. The customer domain communicates with other domains through different wireless technologies like wifi, zigbee and cellular GSM etc [10]. The Generation domain is the power generation end that may also include distributed energy resources along with traditional resources. Furthermore, transmission domain is responsible for carrying the long distance high voltage power and is able to produce and store the power. Distribution domain contains substations and other management entities that supply the requisite power to consumers’ outlets, where customer domain belongs to end users of electricity, including commercial, residential and industrial sub-domains. Other three layers are responsible for management of power tariffs, billing, optimization of system outcomes. These layers also include companies dealing with customers to provide them with different power usage and management plans. Fig. [1] outlines the above discussed domains of the smart grid that are defined by National Institute of Standards and Technology (NIST). Along with significant advantages of smart grid over traditional grids, they entail some vital challenges also that include users’ information privacy and security, efficiency of aggregation techniques, power generation forecasting and effective fault-tolerance mechanisms [11–13]. These all
factors are primarily based on communication system of the grid system. An efficient communication system is one of the important pillars for the appropriate balance of power generation and supply [34–54]. The major entities involved in the data communication are smart meters that record and upload the consumption data regularly to the second entity, that is aggregator node (preferably a fog node) which further transfers the aggregated consumption information to the utility. The fog node relieves the cloud from the burden of multitasking while processing enormous volumes of data, by acting as an intermediary layer between users and the utility, as depicted in Fig. 2. The real time data exchange between these entities is highly vulnerable to security and privacy threats that can potentially affect the performance of the grid system. Additionally, the efficacy of a smart grid system also depends on the integrity of the consumption data being communicated between entities [55–87]. The smart meters installed in the consumers’ outlets upload the usage information intermittently, but any fault in the devices or any network issue can significantly hinder the proper communication of the data. This will further affect the accuracy of the information being transmitted to the utility for decision making. Nonetheless, there are several fault-tolerant schemes provided by the researchers, out of which most prefer to ignore the unavailability of data and rest of them just substitute with some random values. Observing the inaccurate information being communicated to the decision makers and the cascading major issues regarding power management, there is requirement for effective and accurate fault-tolerant schemes for the smart grid environment.

![Figure 1: NIST Components of Smart Grid](image)

2 Related Work

An efficient and secure aggregation solution for multidimensional data provided by Boudia et al. [88] was based on homomorphic property of the Paillier cryptosystem. Additionally, they achieved the inter-entity authentication through BLS digital signature and ignored any unavailability of the data to realise a fault-tolerant aggregation, that further lead to incorrect information communicated to control center. Another aggregation scheme based on operations such as exclusive-OR and one-way hash functions was proposed by Gope et al. [89] that masked the consumption data for enhanced privacy. Due to usage of light-weight operations, the scheme incurs significantly low computation cost, however unavailable data is ignored that results in incorrect and partial aggregated information. An increasing range based multisubset aggregation performed by Wang et al. [90], was based on paillier cryptosystem and for enhanced privacy, a blinding factor was also used with each consumption data. However, their scheme preferred ignoring the missing data and as a result, incorrect aggregate computed on the basis of partial data and lacked effective fault-handling. Another piece of work by Mohammadali et al. [91] used Nyberg’s accumulator and additive homomorphic encryption to aggregate the data at a high computational cost. Besides that, any faults generated from unavailability of data are just handled by notifying the utility provider, that is not a sufficient solution for handling faults. Another technique for managing consumption data as groups, by Chen et al. [92] utilized paillier homomorphic encryption to accomplish scalability and multi-subset aggregation. This work supports dynamic leave and join, however, failure of a smart meter from any group will lead to elimination of that entire group from the aggregate. As a result, the information that is aggregated has a greater degree of inaccuracy. Furthermore, an aggregation technique supporting authentication through elliptic curve cryptosystem and encryption and aggregation through paillier cryptosystem was given by Li et al. [93]. Although, missing data was preferred to be ignored in order to realise fault-tolerance. Guan et al. [94] splitted the consumptions data as equisized groups along with an identical private key allocated to each correspondent smart meter from each group. Furthermore, the aggregation achieved using homomorphic encryption and fault tolerance was realised by substituting unavailable data with corresponding available data from other groups. However, aggregate transmitted to control center was still accompanied with a significant level of inaccuracy. A privacy-preserving
aggregation scheme utilizing primitive operations including pseudorandom function and XOR operation, was given by huang2021lightweight that also obfuscated the data using random blinding factors. The unavailability of data was handled by imputing randomly generated numbers in place of missing data, although that also lead to a significant error in aggregate. Zhang et al. [95] presented a secure aggregation scheme based on modified version of symmetric homomorphic encryption that masked multi-dimensional data with randomly generated secrets. Additionally, bilinear pairing based batch verification was employed in order to authenticate the received consumption data and aggregation of partially available data was performed.

Singh et al. [2] utilized two outsourced clouds for enhanced security and aggregation, that was based on paillier homomorphic encryption. Additionally, their scheme supported query processing but failure of a smart meter was not considered in the scheme and lacked fault-handling. ElGamal cryptosystem based homomorphic aggregation along with multidimensional and range-based partitioning of consumption data was proposed by Zuo et al. [96]. Furthermore, the scheme efficiently handled the internal collusion attacks through distributed decryption but any scenario with failure of smart meters was not considered in the scheme. Qian et al. [97] presented another aggregation technique where lattice-based homomorphic encryption keys $c_1, c_2, ..., c_i$, were produced through a third party and the sum of keys as $c_0 = \sum c_i$ was obtained by utility provider in order to check any discrepancy. This scheme also did not support effective fault-tolerance. Gope et al. [98] described a tree-topology based hop-by-hop aggregation method in which each succeeding smart meter collects the consumption data from its preceding smart meters and aggregates. In order to run the scheme efficiently, each smart meter is mandatorily supposed to participate and any failure in the sequence leads to breakdown of whole system, that is the biggest flaw of the scheme. The paillier cryptosystem was used by Chen et al. [99] to aggregate data from grouped smart meters, however this effort was similarly lacking in fault-tolerance. A unique key agreement protocol-based differentially private and fault-tolerant method was put out by Bao and Lu [100]. The consumption data is retrieved through a secure channel by utilizing Bonneh-Goh-Nissim cryptosystem, via a two-way transmission of data and random secret values. Any error in aggregate is limited by avoiding the unavailable data, hence somewhat fault-tolerance is achieved. Another scheme considering power consumption as an exponential distribution and performing a diverse grouping of this data on the basis of differential privacy, was presented by Shi et al. [101]. Additionally, the scheme is also able to detect any errors effectively, however not supporting the fault-handling. Furthermore, another effective aggregation scheme based on lifted elgamal encryption was given by Ni et al. [102] in which security is enhanced by utilizing laplacian noise. A limited fault-handling was provided by this scheme where any faults in smart meters were communicated to utility provider and aggregate computed over incomplete...
consumption data. Furthermore, a scheme mapping multidimensional data through Horner’s rule was proposed by Shen et al. [103]. Furthermore, a homomorphic aggregation achieved through paillier cryptosystem along with bilinear mapping for authentication, that collectively increased the computation cost. Although scheme was not considering any faults of smart meters for aggregation. Similar to this, the Yuwen, et al. [104] technique separated the meters into groups, with the same set of meters building the encryption keys. By eliminating the group containing the faulty meter from the aggregation, this plan addresses the meter failure issue. Despite that, this technique does not seem to be fault-tolerant because the inaccuracy raised from missing data is not addressed. Another fault-tolerant technique using differential privacy was introduced by Bao et al. [105]. Security and privacy of consumption data was ensured through AES symmetric cryptosystem and failure of meters and consequent missing information was managed by replacing with an auxiliary ciphertext. Although, replacement of original usage data with a random data value still leaves aggregate with a significant lack of precision. An effective summarization of above discussed works is provided in Table 1 that also entails their methods of fault-handling.

Table 1: Feature comparison of existing schemes

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Data Type</th>
<th>Architecture</th>
<th>Encryption Technique</th>
<th>FT</th>
<th>Method for FT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPI-FT [105]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>AES</td>
<td>Yes</td>
<td>imputation</td>
</tr>
<tr>
<td>DMG-MDA [104]</td>
<td>Multi-Dimensional</td>
<td>Fog Node</td>
<td>Elliptic curve</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>ESMA [88]</td>
<td>Multi-Dimensional</td>
<td>Fog-Cloud</td>
<td>Paillier</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>LPFA [89]</td>
<td>One-Dimensional</td>
<td>Fog-Cloud</td>
<td>Hash function</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>EPP-CA [103]</td>
<td>Multi-Dimensional</td>
<td>Fog Node</td>
<td>Paillier</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>FTMA [90]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>Paillier</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>NHP [91]</td>
<td>Multi-Dimensional</td>
<td>Gateway</td>
<td>Paillier</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>DiPrism [102]</td>
<td>One-Dimensional</td>
<td>Gateway</td>
<td>Differential privacy</td>
<td>Yes</td>
<td>Substitution</td>
</tr>
<tr>
<td>Chen et al. [92]</td>
<td>One-Dimensional</td>
<td>Cloud</td>
<td>Paillier</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>DG-APED [101]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>Differential privacy</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>Li et al. [93]</td>
<td>One-Dimensional</td>
<td>Fog-Cloud</td>
<td>Paillier</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>LFTA-PF [106]</td>
<td>One-Dimensional</td>
<td>Fog-Cloud</td>
<td>Hash function</td>
<td>Yes</td>
<td>Imputation</td>
</tr>
<tr>
<td>Zuo et al. [96]</td>
<td>Multi-Dimensional</td>
<td>Cloud</td>
<td>ECEG</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>PHH-DA [98]</td>
<td>One-Dimensional</td>
<td>TPA</td>
<td>Hash function</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Qian et al. [1]</td>
<td>One-Dimensional</td>
<td>Fog-Cloud</td>
<td>Lattice-based</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>APPA-FT [94]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>Paillier</td>
<td>Yes</td>
<td>Substitution</td>
</tr>
<tr>
<td>HMDA [99]</td>
<td>Multi-Dimensional</td>
<td>Fog Node</td>
<td>Paillier</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Zhang et al. [95]</td>
<td>Multi-Dimensional</td>
<td>Fog-Cloud</td>
<td>Symmetric Homomorph</td>
<td>Yes</td>
<td>Avoidance</td>
</tr>
<tr>
<td>DPAFT [100]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>Boneh-Goh-Nissim</td>
<td>Yes</td>
<td>Imputation</td>
</tr>
<tr>
<td>PP-MDA [2]</td>
<td>Multi-Dimensional</td>
<td>Cloud</td>
<td>Paillier</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>SP-DAC [97]</td>
<td>Multi-Dimensional</td>
<td>Fog-Cloud</td>
<td>Paillier</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>AFTAM [55]</td>
<td>One-Dimensional</td>
<td>Fog Node</td>
<td>ECEG</td>
<td>Yes</td>
<td>prediction</td>
</tr>
</tbody>
</table>

3 Emerging Challenges and Research Directions

In consideration of existing fault-tolerant aggregation schemes, the following challenges are identified:

1. **Lack of proper fault handling**: The fault-tolerance is concerned with dealing of unavailability of data caused by reason of faulty smart meters that can significantly affect the decision making regarding power generation and supply. More the extent of faulty devices, more will be the error rate in information being used for further power management, making it one of very important research direction. Besides that, the solutions must be efficient enough to offset the unavailable data instances and suitable for smart grid environment.

2. **Ignorance of inaccuracy in aggregation**: As the consumption information of users is communicated to utility providers in the form of aggregated information. The inaccuracy stemmed in aggregated data from any missed information in between, that is further generated from malfunctioning smart meters or any network and transmission errors, is ignored by most of the existing schemes. This can be a significant issue for utility providers, in case of households with huge power usage patterns.

3. **Misleading information towards power management**: The balance between power generation and distribution is most important aspect in a grid system that is hugely dependent on two-way usage information being shared
between users and utility, in case of smart grid. Bigger the information error, more adverse will be the impact on power management.

4. **Identification of faulty devices and missing information:** Need for light-weight solutions for detection of unavailability of data along with extent of unavailability and the underlying causes.

5. **Need to investigate machine-learning based fault-handling:** As discussed in the related work section, the Prediction based schemes can be proved to be more effective as compared to state-of-the-art works, although have not been investigated properly in context with fault-tolerance. The unavailable information can be precisely estimated using the prediction mechanism in machine learning, that provides a clear and important research direction for smart grid environment.

The subsequent research directions are formulated to deal with aforementioned challenges and to fill the research gaps:

- A practical and reliable fault handling scheme contributing to correct information communication is necessarily required, in order to prevent any mismanagement in the power generation and supply and improve the decision making of the utility.
- The unavailability of the data is pressingly required to be investigated, in accordance with the extent of unavailability and an effective scheme needs to be devised to offset the missing data with substantive accuracy. This is one of very important aspect to be considered to assist the regulatory authorities.
- The machine learning based fault-tolerant mechanisms need to be researched, in order to achieve the aim with maximum accuracy. The prediction of missing information by utilizing the previously available data at the clouds, can provide preferably more accurate estimation.
- To incorporate light-weight detection mechanisms for faulty devices along with extent of unavailability of data is essentially required.

**References**


