

Privacy Preserving User Data and Load Management in a Smart Grid



MCS

by

Yasir Mehmood


A thesis submitted to the faculty of Information Security Department, Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfilment of the requirements for the degree of MS in Information Security


Aug 2023

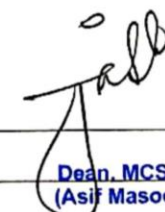
THESIS ACCEPTANCE CERTIFICATE

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by Mr. Yasir Mehmood, Registration No. 00000318101, of Military College of Signals has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/MS Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members and local evaluators of the scholar have also been incorporated in the said thesis.

Signature: 
Name of Supervisor Dr. Fawad Khan
Date: 13/9/23

Signature (HOD): 
Date: 13/9/23
HoD
Information Security
Military College of Sigs

Signature (Dean/Principal) 
Date: 13/9/23
Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Declaration

I hereby declare that no portion of work presented in this thesis has been submitted in support of another award or qualification either at this institution or elsewhere

Dedication

“In the name of Allah, the most Beneficent, the most Merciful”

I dedicate this thesis to my mother, sister, and teachers who supported me each step of the way.

Abstract

The conventional Power grid also known as the traditional power grid is used for the distribution of electricity in a country. The traditional power grid is the interconnection of various electrical equipment like conventional meters, wires, conventional transformers, and other load distributor equipment. In conventional power grid uses a one-way electricity flow from the power generation station to the consumers. Conventional grid systems exhibit a dual drawback. Initially, they compromise user privacy, thereby placing users at risk due to inadequate data confidentiality.

To reduce the above-mentioned problems Smart grid is developed from a traditional grid, but it is an intelligent grid that monitors all the activities in a real-time. Smart grid is more efficient in reliability, efficiency, better demand management and real time monitoring. In order to fulfill the various electrical needs of end users, a smart grid is an electricity network that employs digital and other cutting-edge technology to monitor and regulate the transmission of electricity from all generation sources. A smart metering system is supported by three main components: a smart meter (SM), Aggregator Node (AN), and Smart Grid (SG).

A smart meter is an electronic device that is used to monitor consumers' usage detail voltage level, consumption, usage of electricity, etc in a much more efficient way. Smart meters help to increase efficiency and submit every report to next central device after every 15, 30 or 60 minutes. For privacy preserving, we will use the Pailliar Homomorphic Encryption and forecasting demand of electricity we use LSTM.

Acknowledgments

All praises to Allah for the strengths and His blessing in completing this thesis.

First and foremost I praise and acknowledge **ALLAH**, the most beneficent and the most merciful. Secondly, my humblest gratitude to the Holy Prophet Muhammad (SAW) whose way of life has been a continuous guidance for me.

I would like to convey my gratitude to my supervisor, Dr. Fawad Khan, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and thesis works are major contributions to the success of this research. Also, I would thank my committee members; Dr. Shahzaib Tahir, and Dr. Imran Makhdoom for their support and knowledge regarding this topic.

Last, but not the least, I am highly thankful to my parents, brothers and sisters. They have always stood by my dreams and aspirations and have been a great source of inspiration for me. I would like to thank them for all their care, love and support through my times of stress and excitement.

Table of Contents

THESIS ACCEPTANCE CERTIFICATE.....	i
Declaration.....	ii
Dedication	iii
Acknowledgments	iv
Chapter 1	1
Introduction.....	1
1.1 Background	1
1.2 Motivation / Justification for the Selection of the Topic	4
1.3 Problem Statement	4
1.4 Objectives.....	5
1.5 Thesis Contribution.....	5
1.6 Thesis Organization.....	5
Chapter 2	7
Preliminary Background and Related Work	7
2.1 Introduction	7
2.2 Smart Meters	7
2.3 Data Encryption.....	8
2.4 Aggregator Node	9
2.5 Machine Learning	9
2.7.1 Auto-Regressive (AR)	10
2.7.2 Moving Average (MA).....	11
2.7.3 Autoregressive Moving Average (ARMA)	11
2.7.4 Autoregressive Integrated Moving Average (ARIMA).....	12

2.6	Related Work.....	13
2.7	Homomorphic Encryption.....	17
2.7.1	Partially Homomorphic Encryption:	18
2.7.2	Somewhat Homomorphic Encryption:	18
2.7.3	Fully Homomorphic Encryption:.....	18
2.8	Paillier Homomorphic Cryptosystem.....	19
2.9	ECDSA (Elliptic Curve Digital Signature Algorithm):	20
2.10	ElGamal Encryption-Based Privacy:	22
2.11	Privacy and Security.....	24
Chapter 3		25
Methodology		25
3.1	Introduction	25
3.2	Data Sets.....	27
3.3	Description about Data Set.....	27
3.4	System Model.....	29
3.5	Collection of data (using smart meter).....	30
3.6	Paillier homomorphic encryption.....	31
3.7	Aggregator Node	31
3.8	Encrypted Data.....	32
3.9	Machine Learning Algorithm.....	32
3.10	Advantages of LSTM over ARMA and ARIMA:.....	34
Chapter 4		35
Security and Performance Analysis		35
4.1	Data aggregation in secure form	35
4.2	Grid Station	36

4.3	Time to Encrypt Data	36
4.4	Size of Encrypted and Non-Encrypted Data	37
4.5	Graphs and Results	38
4.6	Mean Absolute Percentile Error (MAPE)	40
4.7	User Data Privacy.....	42
Chapter 5	43
Conclusion and Future Work	43
5.1	Future Work	43
References	44

List of Figure:

Figure 1: Conventional Power Grid.....	2
Figure 2: Smart Grid.....	3
Figure 3: Central Device in Smart Grid (Aggregator Node)	9
Figure 4: Privacy Preserving User Data and Load Management in a Smart Grid.....	26
Figure 5: Detailed model Flow Chart	30
Figure 6: LSTM Model.....	33
Figure 7: Encrypted Data received by Aggregator node.....	35
Figure 8: Encrypted sum of data received Grid Station.....	36
Figure 9: Time taken to Encrypt data.	37
Figure 10: Size of Encrypted and Non-Encrypted Data	37
Figure 11: Energy Consumption According to year	38
Figure 12: Energy Consumption According to year	39
Figure 13: Energy Distribution	39
Figure 14: Machine Learning the Pattern Predicting Future Value	40
Figure 15: MAPE Graph.....	41

List of Table:

Table 1: Related Work 15

Table 2: Data Set useful information. 28

List of Equations:

Equation 1: Auto Regressive (AR)	10
Equation 2: Moving Average (MA).....	11
Equation 3: Autoregressive Moving Average (ARMA).....	12
Equation 4: Autoregressive Integrated Moving Average (ARIMA)	13
Equation 5: MAPE Equation	41

Introduction

1.1 Background

Providing electricity for homes, businesses, and industries: The primary use of the conventional power grid is to provide electricity to homes, businesses, and industries. The conventional Power grid also known as the traditional power grid is used for the distribution of electricity in a country. The traditional power grid is the interconnection of various electrical equipment like conventional meters, wires, conventional transformers, and other load distributor equipment. In conventional power grid uses a one-way electricity flow from the power generation station to the consumers. Conventional grid systems exhibit a dual drawback. Initially, they compromise user privacy, thereby placing users at risk due to inadequate data confidentiality. The traditional electrical grid is an electricity distribution network that connects distributed electric energy customers to a few central generators. It employs a demand-driven strategy centered on forecasting consumption and reacting to any residual gaps between forecasted and actual consumption [23]. An opportunity to narrow the supply-demand gap in electricity distribution is evident through improved utilization of electricity, as highlighted in [31]. The conventional grid does not perform effective and smart distribution as there is lack of new technologies and traditional power grid also fails in real-time load monitoring and management of automated systems for billing and electricity distribution. According to the United States' National Energy Technology Laboratory, power outages are causing annual losses amounting to approximately \$100 billion [32].

Although this model has been effective for the past century or more, there is a growing need to transform the electric power industry in order to fulfill the expectations of the digital age society as well as to address issues with aging infrastructure and new societal and environmental problems.

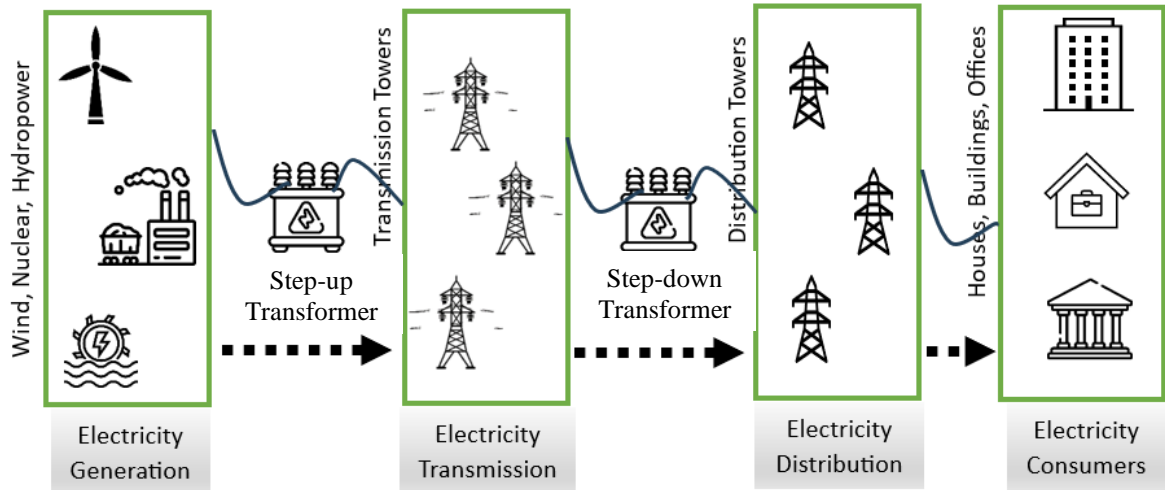


Figure 1: Conventional Power Grid

To reduce the above-mentioned problems Smart grid is developed from a traditional grid, but it is an intelligent grid that monitors all the activities in a real-time. Smart grid is more efficient in reliability, efficiency, better demand management and real time monitoring. In order to fulfill the various electrical needs of end users, a smart grid is an electricity network that employs digital and other cutting-edge technology to monitor and regulate the transmission of electricity from all generation sources. Smart grids involve two-way communication from the power generation station to the consumers and vice versa. The main components used in smart grids are intelligent applications, smart meters, smart transformers, smart distribution centers, and different types of IOT sensors.

The deployment of a heterogeneous infrastructure, including metering devices, data collection, processing, and communication networks, as well as the installation and administrative tasks associated with it, are required for a smart metering system. A smart metering system is supported by three main components: a smart meter (SM), Aggregator Node (AN), and Smart Grid (SG).

A smart meter is an electronic device that is used to monitor consumers' usage detail voltage level, consumption, usage of electricity, etc in a much more efficient way. Smart meters help to increase efficiency and submit every report to next central device after every 15, 30 or 60 minutes. The smart meter delivers a wealth of data on electricity usage to lower bills, improve customer service, and increase efficiency.

Smart meters frequently monitor energy usage in close to real-time and report it at regular, brief intervals throughout the day [1]. These smart meters are used to send the information to the central device. The Aggregator node is responsible for the collection of data from different smart meters and based on these data some action is taken like bill generation and help in effective load distribution etc. Despite smart metering networks' control and administration capabilities, the collected usage data is exploited by various automated and intelligent systems including Distributed Generation and Distributed Storage, Billing, Load Monitoring and Control System [8].

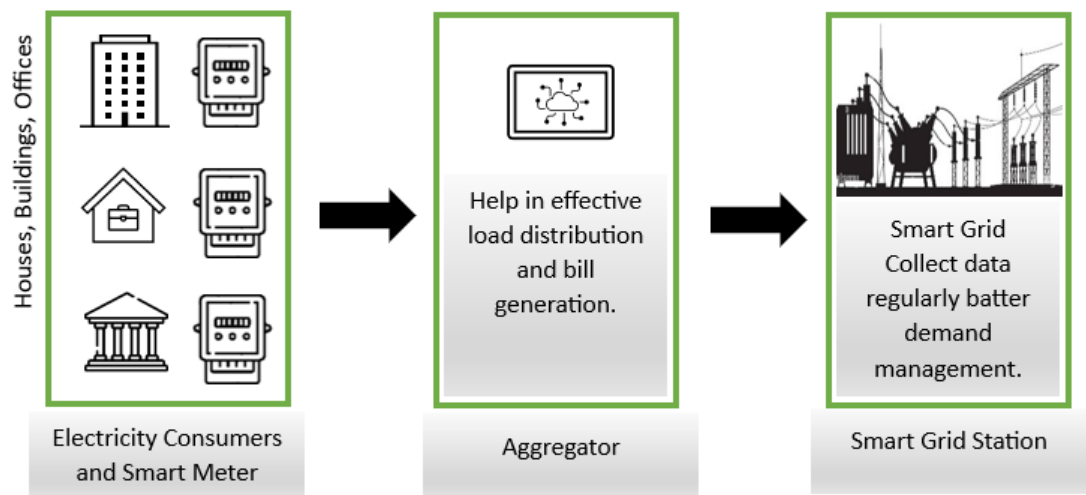


Figure 2: Smart Grid

Smart meters gather specific data about a household's energy usage, which might give rise to concerns about security and privacy issues. People can feel awkward disclosing this information to their utility company because it can reveal insights into the routines and conduct of the household members. From energy usage patterns anyone can easily judge the customer's daily routine, usage of appliances, and occupancy. There are many difficulties to collect data from different smart meters because they may contain some sensitive data. Indeed, it's widely recognized that data transmitted over a network in plaintext form is inherently insecure [2].

1.2 Motivation / Justification for the Selection of the Topic

Smart grids are emerging technology used now a days and show improvement day by day. Smart grids greatly improve reliability by managing and routing power. This allows the grid to reduce the blackouts by reducing maximum power demands. Electrical Grid with Automation is known as Smart Grid, the automated smart grid uses different equipment that keep track of the electricity's journey from production to consumption and control the power flow on real time or near real-time.

To ensure the privacy of a consumer we have to reduce the traces of the smart meter by securing the communication between the smart meter and smart grid. To overcome all these problem machine learning using federated learning is used where model is trained without sharing their personal data.

1.3 Problem Statement

The conventional grid does not perform effective and smart distribution as there is lack of new technologies and traditional power grid also fails in real-time load monitoring and management of automated systems for billing and electricity distribution. To overcome the problem of cascading failures and power losses smart grid is developed from a traditional grid, but it is an intelligent grid that monitors all the activities in a real-time.

The smart grid working is efficiently as compared to conventional system, but the disadvantage is it reveals consumer privacy in terms of consumption of electricity. Sensitive information on electricity trends, and personal information of customers would be kept and delivered in an unencrypted way. As a result, there is a higher chance that third parties will gain access to, intercept, or steal customer data. The privacy of consumers may be compromised if such security breaches expose personal data and energy use trends.

From the standpoint of consumers, ensuring privacy is of utmost importance, as the data transmitted by smart meters could potentially divulge intricate real-time consumption patterns, thereby possibly unveiling personal details of the owner. For instance, discerning low or high-power usage might suggest whether residents are away or present at home.

1.4 Objectives

The main objectives of this thesis are:

- Ensuring the confidentiality of data communication between the smart meter and smart grid.
- Privacy of consumers about the usage of electricity.
- Improve the load distribution of the smart grid.

1.5 Thesis Contribution

It's the 21st century and it is the era of machine learning; machine learning is applied in almost all fields of life in the world. Smart meters often record energy in close to real-time, and report periodically, at regular intervals throughout the day, as was previously mentioned [1]. As there is much shortage of electricity in our country so the smart grid will distribute electricity more efficiently according to their need and reduce the blackouts in a different area. The hospitals and the industrial area need more power and continuous power which can be done by real-time monitoring.

In order to predict electricity demand effectively, machine learning algorithms can examine past consumption trends, weather information, and other related elements. Power grid operators can optimize electricity generation and distribution by projecting demand, providing an adequate supply of electricity for hospitals and industrial sectors during peak hours.

Smart meter data collected in real-time can potentially be analyzed by machine learning algorithms to determine how various areas and industries use power. By moving power resources from low-demand locations to high-demand ones, this information can assist in balancing the load on the grid and guarantee a constant and dependable supply of electricity where it is most required.

1.6 Thesis Organization

The arrangement of the thesis is organized in the following manner:

- Chapter 1 contains an introduction, a problem statement is highlighted, followed by the motivation behind the research, and research objectives are enumerated. Furthermore, the contributions made through this research are highlighted.
- Chapter 2 contains an overview of existing aggregation schemes, followed up by pros and cons of each technique.
- Chapter 3 contains an overview of the preliminary cryptographic primitives employed to design the proposed scheme provided in this chapter, also system model, the assumed adversarial model under which the scheme will function is discussed.
- Chapter 4 covers results of scheme and graph and working and communication is discussed. Mean Absolute Percentile Error (MAPE) is also discussed.
- Chapter 5 covers recommendation, conclusion and future work areas are revealed in this chapter.

Preliminary Background and Related Work

2.1 Introduction

This chapter is related to literature review. In this we explained the preliminary background and related work. In this chapter we explained important topics in detail. We explained Smart Meters, Aggregator, Grid Station, Machine Learning and its classification algorithms, machine learning and its algorithms, and secure encryption methodologies.

2.2 Smart Meters

Smart meters are innovative digital devices that take the place of conventional utility meters in a smart grid to measure and record electricity use in real-time. They are an essential part of the energy distribution system's upgrading and digital transformation. Smart meters allow utility companies and customers to communicate with each other in two directions, which has several advantages for both parties and the whole power system.

Smart meters are installed in word in past decade, the number of meters installed in US is 2.9 million while in China reached to 96 million.[16] Smart meters continually track and record energy use in order to periodically gather data. The utility company's preferences, the kind of smart meter technology being utilized, and the needs of the smart grid installation can all affect how frequently data is gathered. Here is how smart meters normally periodically gather data:

- **Regular Intervals:** Smart meters are set to gather data on usage at regular intervals, such as every 15 or 30 minutes, every hour, or every day. Depending on the utility's requirements and the amount of granularity needed for invoicing and grid management, the interval can be modified.

- **Interval Data Recording:** Throughout the day, the smart meter records the energy use at scheduled intervals. For instance, it may record the amount of energy used every 15 minutes, yielding 96 data points each day.
- **Data Storage:** The smart meter's memory or an integrated storage system maintains the gathered data. The time and date of each consumption reading may be included in this data, which may also be timestamped.
- **Communication with Utility Company:** To transfer the recorded consumption data, the smart meter communicates with the utility company regularly or at the conclusion of each collecting interval. The communication can take place via a variety of techniques, including cellular networks, radio frequencies (RF), powerline communication, and other means of communication.
- **Statistics Processing:** Several smart meters provide the electricity provider with usage statistics. The information is prepared, compiled, and examined to determine each consumer's energy use, produce bills, and assess the functioning of the grid as a whole.
- **Consumer Access:** A few smart utility billing systems additionally give customers access to their usage information via web portals or mobile applications. This gives users the ability to keep an eye on their consumption habits and decide how best to conserve energy.

Smart meters allow utility providers to obtain real-time insights into energy consumption trends and system performance by regularly collecting data. The use of demand response systems, improved grid management, and effective load balancing are all encouraged by the data in order to optimize energy distribution and consumption throughout the smart grid.

2.3 Data Encryption

Ensuring the privacy of consumers about the usage of electricity is the big hurdle in way of smart grid [17]. The user's data must be secure from malicious attack [18]. Recent research has suggested privacy-preserving approaches to prevent the leakage of users' private information. The true identity of users being kept secret is one direct way to ensure privacy protection.

Li et al. [18] proposed an effective privacy-preserving demand response (EPPDR) system that combines efficient response with privacy-preserving power demand aggregation.

2.4 Aggregator Node

The Aggregator Node plays a crucial role in a smart grid, ensuring reliable and efficient power distribution by coordinating and controlling its diverse elements. Referred to as the central controller or central control system, it serves as the intelligent core of the smart grid, facilitating seamless communication, billing and load monitoring among the grid's various components.

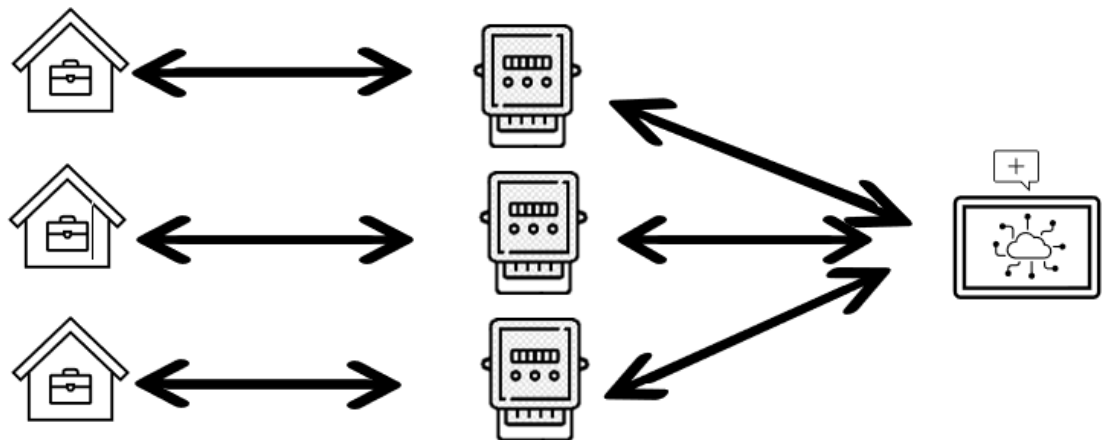


Figure 3: Central Device in Smart Grid (Aggregator Node)

2.5 Machine Learning

Machine learning represents a segment within the realm of Artificial Intelligence (AI). In machine learning, a computer is assigned a task to complete, and machine learning code learns from its experiences and tries to complete the task. Machine learning refers to learning on its own without writing lengthy code to complete a task. Machine learning emphasizes on code that gets large datasets and trains itself on them using different machine learning algorithms. Machine learning learns itself from its experiences. More experience will give us more accurate results. After training an algorithm, the same algorithm is used for making decisions, predictions, or forecasting based on data. There are different examples in our real life for which machine learning is used, like to predict cancer disease from different medical

reports. Machine learning is used in wide variety of fields like robotic, business, computer games, google map, healthcare, online fraud detection, pattern recognition etc. [13]

2.7.1 Auto-Regressive (AR)

The Auto-Regressive (AR) model is a sort of time series model used in machine learning that forecasts future values based on its own historical values. The AR model makes the assumption that, up to a specific lag order, a time series' present value is linearly reliant on its prior values.

An auto-regressive model of order "p" is represented mathematically as follows:

Equation 1: Auto Regressive (AR)

$$X_t = C + \sum_{i=1}^p \phi_i \cdot X_{t-i} + \varepsilon_t$$

Where:

X_t is the time series' value at time "t."

C constant term or intercept is called c.

ϕ_i indicates the estimated coefficients or weights that were applied to the historical values during model training.

$X_{t-i} + \varepsilon_t$ is a random error term, reflecting the noise or unexplained variance in the time series, where "i" runs from 1 to "p." represents the previous values of the time series.

The lagged values of the time series themselves serve as the predictors in the AR model, which is a linear regression model. The AR model's order "p" tells us how many previous time steps were considered while predicting the currently occurring result.

When employing the AR model in machine learning, historical time series data is utilized to train the model, and a variety of methods, including the method of least squares, are employed to estimate the coefficients. After the model has been trained, it may be used to forecast future time steps based on historical data from the time series.

2.7.2 Moving Average (MA)

In statistics and econometrics, the Moving Average (MA) model is a time series model that is used to evaluate and forecast data that changes over time. The MA model models the value of a time series based on its previous forecast errors or residuals, as opposed to the Autoregressive (AR) model, which models the value of a time series based on its prior values.

The MA model is denoted by "MA(q)", where "q" is the model order, or the quantity of lag prediction errors taken into account. An MA(q) model is represented mathematically as follows:

Equation 2: Moving Average (MA)

$$X_t = C + \varepsilon_t + \sum_{i=1}^q \phi_i \cdot \varepsilon_{t-i}$$

X_t reflects the time series value in this equation at time "t."

C An intercept or constant term is c .

ε_t signifies the stochastic noise or unpredictable variation at time "t."

The coefficients or weights, denoted as ϕ_i , are associated with previous forecast errors, encompassing the range of "i" from 1 to "q."

ε_{t-i} symbolize the forecast errors at time "t-i."

In the MA model, the white noise errors from the preceding "q" time periods are combined linearly to represent the current value of the time series. The MA model is very effective for identifying transient patterns or oscillations in time series data.

2.7.3 Autoregressive Moving Average (ARMA)

In the realm of statistics and econometrics, the Autoregressive Moving Average (ARMA) model is employed to evaluate and forecast time-varying data. The Moving Average (MA) and Autoregressive (AR) models' traits are combined in the ARMA model.

The following is a mathematical representation of the order (p, q) ARMA model:

Equation 3: Autoregressive Moving Average (ARMA)

$$X_t = C + \sum_{i=1}^p \phi_i \cdot X_{t-i} + \sum_{i=1}^q \phi_i \cdot \varepsilon_{t-i} + \varepsilon_t$$

X_t is the time series' value at time "t."

C A constant term or intercept is called c .

ϕ_i indicates the previous values of the time series, where "i" runs from 1 to "p."

X_{t-i} represents the coefficients or weights applied to the time series' past values.

ϕ_i denotes the coefficients or weights applied to the time series' previous forecast mistakes, where "i" spans from 1 to "q."

ε_t represents the white noise or random error at time "t."

ε_{t-i} represents the historical forecast errors at time "t-i," where "i" might be anything between 1 and "q."

The ARMA model incorporates both the moving average links between previous forecast errors and the autoregressive correlations between past values of the time series. It can manage time series data that have temporal dependencies and brief fluctuations.

In several disciplines, including finance, economics, engineering, and environmental sciences, the ARMA model is often used for time series analysis, forecasting, and deciphering underlying patterns. The Autoregressive Integrated Moving Average (ARIMA) model, which incorporates differencing to handle non-stationary time series, is one of the most complex models built on top of it.

2.7.4 Autoregressive Integrated Moving Average (ARIMA)

Within the domains of statistics and econometrics, the Auto-Regressive Integrated Moving Average (ARIMA) model holds a prominent position as a favored method for assessing, predicting, and enhancing understanding of temporally changing data. ARIMA amalgamates the attributes of three constituent components: Auto-Regressive (AR), Integrated (I), and Moving Average (MA).

The following is a mathematical representation of the ARIMA model of order (p, d, q):

Equation 4: Autoregressive Integrated Moving Average (ARIMA)

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - 1)^d X_t = C + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

X_t is what the time series is worth at time "t."

C An intercept or constant term is c.

L is difference between two successive time periods is represented by the lag operator.

$\phi_1, \phi_2, \dots, \phi_p$ corresponding to the autoregressive terms' coefficients or weights.

$\theta_1, \theta_2, \dots, \theta_q$ are connected to the moving average terms via coefficients or weights.

d stands for the degree of differencing, or the number of times the time series must be differenced before reaching stationarity.

The ARIMA model is very helpful for addressing non-stationary time series data, when the mean and variance of the series are statistical constants. By eliminating trends and seasonality, the "Integrated" component (differencing) aids in transforming the data into a stable form.

Numerous fields, including finance, economics, environmental sciences, and business analytics, use ARIMA models extensively. When the data shows temporal relationships and fluctuations, they are useful tools for time series analysis, trend forecasting, and decision-making.

2.6 Related Work

The vulnerability of data during its transmission from smart meters to grid stations is a notable concern. Recent studies have demonstrated that data leaks occur through local model parameters, and attackers might take advantage of the vulnerability to collect data about the participant. [5]. Aono et al. utilized a Homomorphic encryption (HE) scheme to protect local gradients trained with local data in [5]. A lightweight framework (federated

multi-task learning) to robustness against a poisoning attack that reduces learning accuracy [6]. A Privacy-Preserving Multi-subset Aggregation (PPMA) scheme in smart grid is used in [12] according to their electricity usage during each period, the users in a residential area are divided into multiple subsets, and the control center can calculate the total electricity consumption, and the number of users for each subset. To lessen the loss of energy and learning efficiency caused by frequent up-linking and far-off central servers, a federated learning architecture leveraging device-to-device communication was developed. [7].

Yining et al. [26] present a privacy-preserving aggregation method that does not require the use of a Trusted third-party. To safeguard a single user's data, the technique encrypts consumption data using the EC-EIGamal cryptosystem and creates a virtual aggregation area instead of a physical one. In the work [27] by Zhitao Guan et al., EFFECT is proposed to achieve both the source authentication and aggregation using the Paillier cryptographic scheme and Secret Sharing Scheme. While ensuring individual privacy, the scheme also guarantees fault-tolerance.

In [28], researchers introduced a P2DA approach employing Boneh Goh-Nissim encryption to safeguard against internal threats. However, while this method offers enhanced security potential, achieving substantial security necessitates a large 'n,' resulting in elevated communication costs. Conversely, Elliptic-Curve Cryptography (ECC) achieves comparable security levels with a smaller key size. An alternative approach [29] presents a lightweight framework for aggregating electricity consumption utilizing lightweight lattice-based homomorphic cryptosystems. Notably, this method involves Smart appliances, rather than smart meters, performing the aggregation of readings. Moreover, study [30] outlines a technique for geographically aggregating load monitoring data through simple cryptographic primitives like XOR operations and one-way hash functions.

The algorithm proposed in [5], All participants utilise the same private key for HE, hence key management is critically necessary and trustworthy channels must be developed to transfer ciphertexts. Several open-source FL systems, e.g., TensorflowFL, and Pysyft are now intensively used by both research communities.

In recent years much research work used deep neural networks and Long Short-term memory (LSTM) to handle short-term forecasting. Authors proposed a solution using a variant of LSTM which shows significant improvement for one-minute resolution but not in one hour compared to the LSTM standard [4]. Some authors complimentary add the weather conditions, where weather conditions is playing an important role in the individual load and short-term load forecasting [10]. Distinct methodologies are employed by various authors to achieve proficient load distribution. These approaches encompass K-nearest neighbor (KNN), neural networks (NNs), decision tree classifier (DTC), logistic regression (LR), and support vector machines (SVM) [11].

Table 1: Related Work

Authors	Methods	Limitation	Load Monitoring	Privacy issues
A. Ahmad, N. Javaid [13]	Enhanced differential evolution (EDE)	Only suitable for the bulk power generation	✓	
Yining et al. [26]	encrypts consumption data using the EC-EIGamal cryptosystem	creates a virtual aggregation area instead of a physical one		✓
Zhitao Guan et al [27]	EFFECT, an efficient flexible privacy-preserving aggregation scheme	Increased Complexity		✓

<p>Peng Kou, Feng Gao [14]</p>	<p>heteroscedastic Gaussian</p>	<p>The model relies heavily on external variables, which may be difficult to obtain or may not be available in real-time</p>	<p>✓</p>	
<p>D. L. Marino, K. Amarasinghe [4]</p>	<p>LSTM-based Sequence to Sequence and Standard LSTM architecture</p>	<p>The conventional LSTM architecture struggled to provide accurate load predictions at a one-minute resolution.</p>	<p>✓</p>	
<p>Aono, Y.; Hayashi, T [5]</p>	<p>All gradients are encrypted and stored on the cloud server.</p>	<p>Cost in computation and communication</p>		<p>✓</p>
<p>Zhang, C.; Li, S.; [8]</p>	<p>BatchCrypt utilizes batch encryption to substantially lower encryption overhead and the aggregate amount of ciphertext.</p>	<p>Increased Complexity</p>		<p>✓</p>

Mothukuri, Viraaji; Parizi, [9]	FL was proposed to extend machine learning benefits to domains with sensitive data.	Increased Complexity		✓
M. Akgün, E.U. Soykan [15]	Data is encrypted in the household smart appliances; Encrypted database's keys are protected by Trusted Execution Environment	Do not discuss how data is encrypted. Paper discussed the security of user data.		

Based on the historical records of earlier time spots recorded for the same observation, the AR, MA, ARMA, and ARIMA models are used to anticipate the observation at $(t+1)$. However, it is crucial to confirm that the time series remains stationary during the course of the historical observational data. Applying the differencing factor to the records would determine if the time series' graph represents a stationary overtime period if it were not stationary.

2.7 Homomorphic Encryption

Homomorphic encryption is a cryptographic technique that enables computations to be performed on encrypted data without the need to decrypt it first. This advanced encryption approach is particularly valuable for maintaining data privacy and security while allowing

computations to be carried out on sensitive information. Three primary categories of homomorphic encryption exist: partially homomorphic encryption, somewhat homomorphic encryption, and fully homomorphic encryption.

2.7.1 Partially Homomorphic Encryption:

Partially homomorphic encryption allows for computations of a specific type on encrypted data, but not arbitrary operations. There are two primary types of partially homomorphic encryption:

a. Additive Homomorphism: This type of encryption enables encrypted numbers to be added or subtracted, producing a new encrypted result. The operation is:

$$\text{Encrypted}(a) + \text{Encrypted}(b) = \text{Encrypted}(a + b)$$

b. Multiplicative Homomorphism: This type permits encrypted values to be multiplied by a plaintext constant, yielding an encrypted product:

$$\text{Encrypted}(a) \times b = \text{Encrypted}(a \times b)$$

2.7.2 Somewhat Homomorphic Encryption:

Somewhat homomorphic encryption builds upon partially homomorphic encryption by allowing both addition and multiplication operations on encrypted data. However, the number of operations that can be chained is limited, and complex computations may not be feasible due to accumulation of noise during the operations. The most well-known example of somewhat homomorphic encryption is the RSA cryptosystem.

2.7.3 Fully Homomorphic Encryption:

Fully homomorphic encryption (FHE) is the most powerful form of homomorphic encryption. It allows arbitrary computations to be performed on encrypted data, including addition, multiplication, and more complex operations. FHE enables iterative computations without requiring decryption at any point, maintaining data privacy throughout the process.

However, FHE is computationally intensive and can be significantly slower compared to other encryption methods.

The significance of homomorphic encryption lies in its applications in secure data processing, such as secure cloud computing, privacy-preserving data analysis, and confidential computation outsourcing. It provides a strong solution for scenarios where data privacy is of utmost importance and computations need to be performed on encrypted data without revealing the underlying information. While FHE is a cutting-edge advancement, research and development are ongoing to improve its efficiency and practical usability.

2.8 Paillier Homomorphic Cryptosystem

Homomorphic encryption empowers users to manipulate encrypted data directly, eliminating the need for prior decryption. The Paillier Cryptosystem constitutes a variant of partial Homomorphic Encryption (HE) [24], facilitating the execution of additive operations on data that has been encrypted using homomorphic techniques.

The Paillier cryptosystem, a probabilistic cryptographic construction introduced by Paillier in 1999, is rooted in the composite residuosity problem. Widely employed in privacy-preserving applications, including those cited in [25], the Paillier scheme is especially adept at achieving additive homomorphic encryption. Employing an asymmetric encryption approach, this system enhances homomorphic traits with greater efficiency compared to existing algorithms. Its security against chosen plaintext attacks is well-established, with its correctness and security demonstrated in [25].

Below is a sequential elucidation of key generation, encryption, and decryption processes within the Paillier Homomorphic Cryptosystem:

Key Generation:

Choose two large prime numbers, p and q , randomly and independently.

Calculate the modulus $n=p \times q$.

Compute Carmichael's lambda function: $\lambda(n)=\text{lcm}(p-1,q-1)$.

Choose an integer g such that g is in the multiplicative group modulo n^2 and has order $n \times \text{lcm}(p, q)$.

Compute the modular inverse μ of $L(g^{\lambda(n)} \bmod n^2)$, where $L(x) = (x - 1)/n$, using the provided public key (n, g) and private key $(\lambda(n), \mu)$.

Encryption:

1. Select a plaintext message m satisfying $0 \leq m < n$.
2. Generate a random integer r such that $0 \leq r < n$ and $\text{gcd}(r, n) = 1$.
3. Calculate the ciphertext $c = g^m \times r^n \bmod n^2$.

Decryption:

Compute $L(c^{\lambda(n)} \bmod n^2)$.

Decrypt the ciphertext to obtain the plaintext $m = L \times \mu \bmod n$.

The Paillier Homomorphic Cryptosystem provides additive homomorphism, allowing operations on encrypted data without decryption. Specifically, you can perform homomorphic addition and scalar multiplication on encrypted values.

2.9 ECDSA (Elliptic Curve Digital Signature Algorithm):

ECDSA is a widely used digital signature algorithm based on elliptic curve cryptography. It offers a method to generate digital signatures for messages, which can be authenticated by others through the signer's public key.

ECDSA is commonly used for authentication, data integrity, and non-repudiation in various applications, including secure communication, blockchain technology, and digital certificates. It does not provide homomorphic encryption capabilities like the Paillier Cryptosystem.

This encompasses the Elliptic Curve Digital Signature Algorithm (ECDSA), covering its procedures for key generation, signature generation, and verification using both public and private keys:

- **Key Generation:**

Select an Elliptic Curve: Choose an elliptic curve E defined over a finite field, often represented as F_p , where p is a prime number, and select a base point G on the curve. The curve's parameters are publicly known.

Private Key Generation: Choose a private key d randomly from a secure random number generator such that $1 \leq d < \text{order of } G$. The order of G (n) is the number of distinct points on the curve.

Public Key Derivation: Compute the corresponding public key $Q = d \times G$, where d is the private key and Q is a point on the elliptic curve. The public key Q is used for verification and is made public.

- **Signature Generation:**

Hash the Message: Compute a cryptographic hash (e.g., SHA-256) of the message m to be signed: $H(m)$.

Generate a Random Number: Choose a random number k such that $1 \leq k < n$, where n is the order of the base point G .

Compute k-times Point: Compute the point $k \times G$ on the elliptic curve.

Calculation of r Component: Find the r component by obtaining the x -coordinate of $k \times G$ modulo n , symbolized as r : $r = (k \times G)_x \bmod n$.

Computation of s Component: Determine the s component using the formula $s = k^{-1} \times (H(m) + d \times r) \bmod n$, with d representing the private key.

Generation of Signature: Form the signature for the message m as the pair (r, s) .

- **Signature Verification:**

Message Hashing: Commence signature verification by computing the hash of the received message, denoted as $H(m)$.

Calculate Inverse of s: Compute the inverse of s , denoted as s^{-1} , modulo n .

Calculate u_1 and u_2 : Calculate $u_1 = H(m) \times s^{-1} \bmod n$ and $u_2 = r \times s^{-1} \bmod n$.

Compute $u_1 \times G + u_2 \times Q$: Compute the point $u_1 \times G + u_2 \times Q$ on the elliptic curve.

Verify r : If the x-coordinate of $u_1 \times G + u_2 \times Q$ modulo n matches r , the signature is valid. Otherwise, it is invalid.

2.10 ElGamal Encryption-Based Privacy:

In smart grid systems, where sensitive data must be safely delivered and stored, privacy protection is a crucial problem. To deal with these privacy issues, ElGamal encryption has emerged as a potential cryptographic method. This note presents a summary of ElGamal encryption-based privacy preservation methods for the smart grid, covering such schemes' guiding concepts, processes, and essential elements.

Working Principle:

For safe transmission of data, the public-key cryptosystem ElGamal makes use of discrete logarithms' mathematical features. ElGamal encryption is employed in the context of smart grids to safeguard sensitive data, such as energy consumption statistics, user behavior, and grid operation details.

Key Elements:

Key Generation:

The process of creating a key pair, which consists of a public key (PK) and a private key (SK), is known as key generation.

While the private key is kept confidential and is used for decryption, the public key is used for encryption.

Encryption:

Using the recipient's public key (PK), the sender (for example, a smart meter) encrypts the plaintext data.

Choosing a random number (k), calculating the components of the ciphertext, and sending the ciphertext are all steps in the encryption process.

Decryption:

To retrieve the original plaintext and decrypt the ciphertext, the recipient (for instance, a utility company) utilizes their private key (SK).

Procedure:**Generating a Key:**

A huge prime integer (p) and a simple root (g) modulo p are generated by the user.

The user chooses a private key (SK) at random.

The public key (PK) is calculated by the user as $g^{\text{SK}} \bmod p$.

Encryption:

Gets the recipient's public key (PK) from the sender.

The sender chooses a number at random (k).

The sender generates the ciphertext's various parts:

- Text cipher $C1 = g^k \bmod p$
- Ciphertext $C2 = \text{Plaintext} \bmod p * (\text{PK}^k)$

Decryption:

The recipient calculates the shared secret using their private key (SK) as follows:

$\text{shared_secret} = (C1^{\text{SK}}) \bmod p$.

The recipient calculates the shared secret's inverse ($\text{shared_secret}^{-1}$) modulo p).

Receiving party obtains plaintext using the formula $\text{plaintext} = (C2 * \text{shared_secret}^{-1}) \bmod p$.

Challenges:

To maintain secure communication, efficient key management and dissemination are essential.

Complex mathematical calculations used in ElGamal encryption might result in processing overhead.

ElGamal encryption is mathematically proven to be secure when subjected to chosen plaintext attacks (CPA), but it is vulnerable under chosen ciphertext attacks (CCA)[33].

2.11 Privacy and Security

In the previous related work, there were shortcomings in adequately addressing privacy concerns and load management issues. The research or projects might have overlooked or inadequately considered the impact of these important aspects. Sensitive information on electricity trends, and personal information of customers would be kept and delivered in an unencrypted way. As a result, there is a higher chance that third parties will gain access to, intercept, or steal customer data. The privacy of consumers may be compromised if such security breaches expose personal data and energy use trends.

In our thesis we will address privacy and load management at the same time. All the data of the user will be encrypted so no third party misuses the user's data. We will also perform different operations using python for the purpose of load management without losing user's privacy.

Methodology

3.1 Introduction

This chapter present the description of research process. It provide detail information about the conduct of research, including the justification for choosing that method. This chapter include various stages of thesis, which include participants, dataset, machine leaning algorithms, models and data analysis. This study have the following research questions:

- How data is collected from different clients?
- How we can protect client data?
- How clients can participate in training process?
- How all the data of different clients are used for prediction?
- Will the prediction accuracy up to the mark of machine learning?

This research is based on prediction of electricity usage or load management and user privacy. In the modern era data privacy in the most serious concern for users. The record of user is very sensitive and cannot be shared with some, as other entity can get much more information from the user data which is very threaten to user. Two primary categories employed in load forecasting are mathematical models and artificial intelligence-based computational models[34].

Here we are using 3 basic components for smart grid, for the collection of consumption of units we are using smart meters (SM), the SM sends data to next device called aggregator node (AN) the data is receive by the AN is in encrypted form so that no one intercept users important data to make sure the privacy of the user. The Aggregator Node (AN) aggregates the data from all meters and shares it with Grid Station. The 3rd and last step is grid station

(GS), it receives the summed encrypted data and only GS will be only part of the smart grid who can decrypt data without knowing about individual consumption.

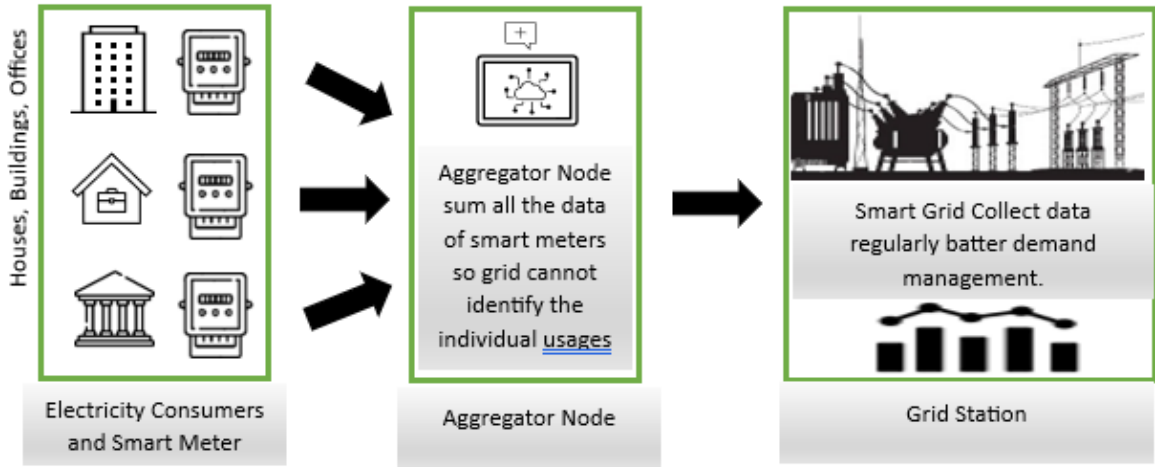


Figure 4: Privacy Preserving User Data and Load Management in a Smart Grid

Machine learning is a technique that used different algorithm for predicting or forecasting based on given data. Traditional machine learning used the centralized training approach, where all data should reside on a centralized location or server in plain text. The data is available in plaintext form which is big security concern for user or clients. Without using machine learning algorithms or sharing data, smart grid cannot achieve the high accuracy prediction of usage of electricity.

In this chapter we use the Homomorphic Encryption and Long Short-Term Memory Recurrent neural network (RNN) architecture. When long-term dependencies is a concern, LSTM, a modified recurrent neural network (RNN) with a chain-like topology, is appropriate for making predictions [19]. Homomorphic Encryption is a cryptographic technique that allow to perform computation on an encrypted data. In other word the data is encrypted while doing computation. With traditional encryption techniques, the data must first be decrypted before calculations can be performed on it. This might provide privacy and security problems, especially when working with sensitive data. Homomorphic encryption overcomes this limitation by allowing computations to be performed on encrypted data directly, thereby maintaining privacy and confidentiality throughout the process. Recurrent neural network

(RNN) architecture that uses Long Short-Term Memory (LSTM) was created with the goal of overcoming the shortcomings of conventional RNNs in capturing long-term dependencies in sequential input. LSTMs find frequent utilization across various sequential data domains, encompassing applications such as time series analysis, speech recognition, and natural language processing.

3.2 Data Sets

A dataset is an organized group of data that depicts a specific collection of things, happenings, or entities. A dataset is an organized collection of data samples or observations that are used in data analysis and machine learning to train and test algorithms, create statistical models, conduct analysis, and come to data-driven recommendations.

This research is related to prediction of electricity usage and user data security. So we use data set of a city having different clients like offices, houses, hospital. The data usage of single client is kept private in whole process so now one can misuse the user's data.

The dataset available at the provided Kaggle link offers valuable insights into hourly energy consumption trends. Analyzing this dataset could contribute to research focused on understanding and optimizing energy usage patterns.

The following is the link of the data set used in reserch

https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption?select=AEP_hourly.csv

3.3 Description about Data Set

This dataset contains 29 columns having different clients, clients include House hold data set, offices and Charging points. These all data in kept private in whole process, but the data in a data set is not in encrypted form. The electricity of the clients are calculated in **MegaWatt (MW)**. The power unit watt (W) is used to quantify the rate of production or

consumption of electrical energy. There is 121273 entries of different of years (2004-2018). The missing values are removed before submitting the data to the model to prevent any computation errors. Some of basic information of table is given below

Table 2: Data Set useful information.

Feature	Value
Total Entries	121273
Mean	15499.513
Std	2581.3959
Min	9579
Max	25696
No of unique years	15

Data set contains 29 columns having different clients, The basic table is shown

	Shopping Mall	Hospital	Fire Station	Sum
Datetime				
12/31/2004 1:00	437	378	38	
12/31/2004 2:00	417	361	37	
12/31/2004 3:00	408	353	36	
12/31/2004 4:00	406	351	36	
12/31/2004 5:00	411	355	36	
12/31/2004 6:00	423	366	37	
12/31/2004 7:00	444	384	39	
12/31/2004 8:00	463	401	41	
12/31/2004 9:00	477	413	42	
12/31/2004 10:00	484	419	42	

House 7	House 8	House 9	House 10	House 11	House 12	H
808	317	105	7	938	431	
771	303	101	6	895	412	
754	296	98	6	875	403	
750	295	98	6	871	401	
759	298	99	6	881	406	
781	307	102	7	907	417	
820	322	107	7	952	438	
857	336	112	7	995	458	
882	346	115	7	1024	471	
895	352	117	7	1039	478	

	House 13	House 14	House 15	sum	Office Building	Office Building	C
MW							
1	168	904	236		934	815	
2	160	863	225		892	778	
3	157	844	220		872	760	
1	156	840	219		868	757	
5	158	850	222		878	766	
7	163	875	228		904	788	
3	171	919	240		949	828	
3	178	959	250		991	864	
1	184	988	258		1020	890	
3	186	1003	261		1036	903	

sum	Slow	Slow	Fast	Fast	sum
	787	232	275	422	
	751	221	263	402	
	734	216	257	393	
	731	215	256	392	
	740	218	259	396	
	761	224	266	408	
	799	235	280	428	
	835	246	292	447	
	859	253	301	460	
	872	257	305	467	

3.4 System Model

The LSTM-based demand forecast model utilized in this study has its parameters adjusted for greater accuracy. The model needs a large quantity of data on the smart grid's power use in order to make precise forecasts. The training data set and the validation dataset are then separated from this data. The training dataset is utilized by the model. Following that, the validation dataset was used to test each model to make sure it was functioning as planned. Following the validation phase, both machine learning and statistical models were used to make predictions.

Data security is ensured through encryption at various stages. Smart Meters (SM) send encrypted data to the Aggregator Node (AN) to protect user privacy from interception. The data is encrypted with public key of Grid Station. The AN aggregate encrypted data from all meters before transmitting it to the Grid Station (GS). Only the Grid Station possesses the

decryption capability using the private key of Grid Station, allowing it to receive and decrypt the summed data while maintaining individual consumption privacy within the smart grid framework.

At last to make sure the valid comparisons Mean Absolute Percentage Error (MAPE) is also calculated. Mean Absolute Percentage Error (MAPE) is statistic used to assess the accuracy of a forecasting or prediction model. The detailed model flow chart is illustrating every step and required method for the complete paper.

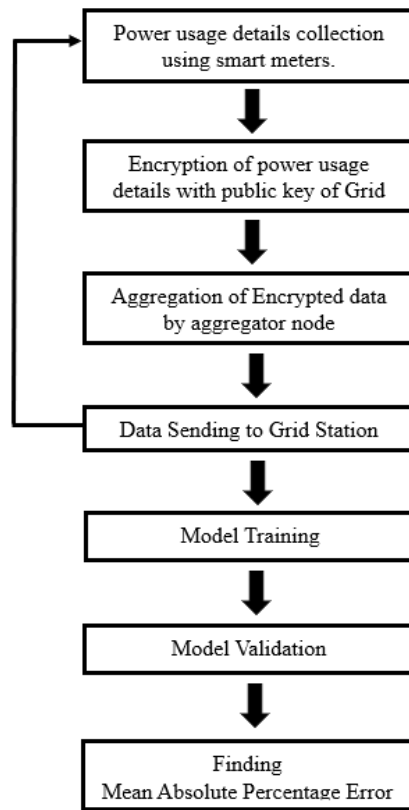


Figure 5: Detailed model Flow Chart

3.5 Collection of data (using smart meter)

Smart meters use a variety of communication methods to gather data from users in a smart grid. These communication channels enable the smart meters to report regular or actual time usage statistics to the utility provider.

Smart meters are responsible for the collection of details of usage of electricity. Smart meter collects the usage details in megawatt at every 60 minutes. Smart meters collect the data and sent to aggregator for further processing. The smart grid shares the data in encrypted form so that no one can access the data or get any valuable information that leads to destruction to the clients.

3.6 Paillier homomorphic encryption

Paillier homomorphic encryption is a specific type of homomorphic encryption scheme known as an "additive homomorphic encryption" scheme. Within an additive homomorphic encryption scheme, operations like addition and subtraction can be directly executed on encrypted data, yielding decrypted outcomes identical to the results of those operations on the original plaintext data.

Additive Homomorphism: Paillier encryption supports addition of encrypted values without the need for decryption. When two ciphertexts are added together, their corresponding plaintexts are added modulo the encryption modulus, preserving the homomorphic property.

3.7 Aggregator Node

A different method of protecting privacy is data aggregation, which involves combining power consumption statistics for an area without disclosing individual customer use information. The majority of the current data aggregation solutions combine the electrical readings while ensuring user privacy through different encryption methods.

In the context of a smart grid, a "aggregator" is often a piece of software or hardware that collects and aggregates data from various smart meters. Multiple smart meters from different parts of the grid are gathered by aggregators. They compile this data into a centralized repository to produce a thorough and current picture of the energy usage trends throughout the whole locations. Rapid consolidation of data and resources plays a pivotal role within smart grid systems [22].

3.8 Encrypted Data

The data that is sent by smart meter to aggregator node is in encrypted form, to protect the privacy and security of consumers and ensure the integrity of the grid system. The user data is homomorphically encrypted. A smart grid uses homomorphic encryption to handle the difficulty of safely processing encrypted data without the need for decryption. It enables calculations to be made directly on encrypted data, protecting the secrecy and privacy of sensitive data.

1. Homomorphic encryption ensures data privacy by enabling computations on encrypted data, keeping sensitive information confidential.
2. Encrypted data can be securely shared and processed without revealing the original content, promoting secure data sharing.
3. Data integrity is maintained through homomorphic encryption, as any unauthorized alterations to encrypted data are detectable during decryption.
4. Computation outsourcing is facilitated, allowing third-party service providers to process data without accessing the raw information.
5. Homomorphic encryption aids in complying with data privacy regulations and mitigates legal concerns for organizations handling sensitive data.

3.9 Machine Learning Algorithm

As our research is based on predicting the future based on previous data, so we are using Time Series Analysis LSTM (Long Short-Term Memory) Recurrent Neural Network (RNN).

A recurrent neural network (RNN) called LSTM (Long Short-Term Memory) is made to handle the vanishing gradient problem and identify long-distance relationships in sequential input. Sepp Hochreiter and Jürgen Schmidhuber first launched it in 1997.

Important characteristics of LSTM neural networks:

Memory Cells:

LSTMs feature memory cells that can continuously store and update data. The memory cells are equipped with three gates: the input gate, the forget gate, and the output gate.

The **input gate** chooses how much of the incoming data should be stored in each memory cell. It chooses which brand-new data to add to the cell state.

The **forget gate** selects which data to remove from memory cells, enabling the model to ignore irrelevant or out-of-date data.

The **output gate** controls how much of the state of the memory cell should be revealed to the following layer or utilized to make predictions.

Handling Long-Term Dependencies: LSTMs excel at capturing extensive temporal relationships within sequential data, rendering them valuable for tasks such as natural language processing, speech recognition, and time series analysis.

Vanishing Gradient Issue: Traditional RNNs frequently experience the vanishing gradient issue, which makes it difficult to train deep networks over lengthy periods. LSTMs address this issue.

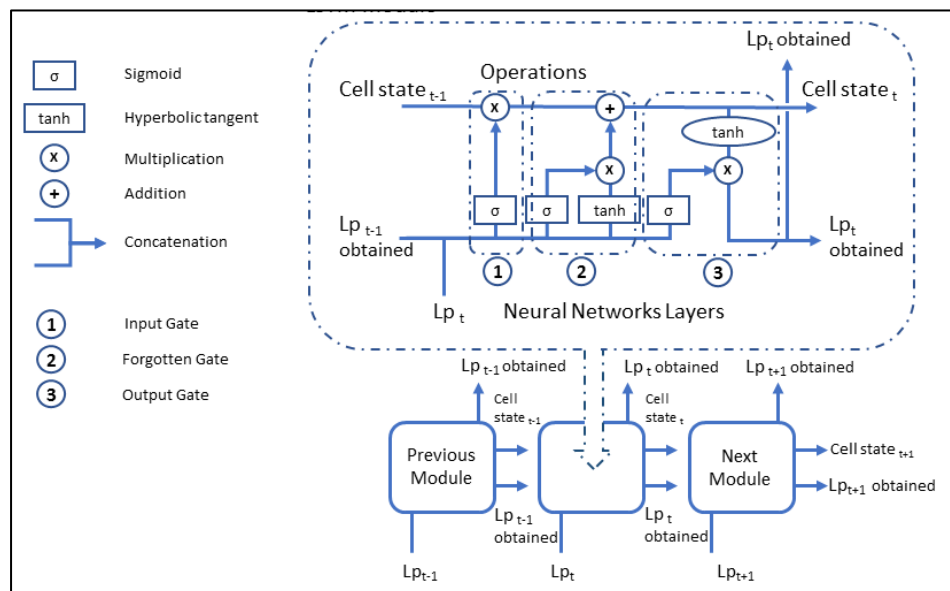


Figure 6: LSTM Model

3.10 Advantages of LSTM over ARMA and ARIMA:

- **Linearity:**

LSTM can handle non-linear relationships more successfully than ARMA and ARIMA, which are linear models and may have trouble capturing complicated non-linear patterns in time series data.

- **Memory:**

LSTM can capture long-term dependencies by exploiting its recurrent nature and memory cells in contrast to ARMA and ARIMA, which have limited memory and can only take into account a set number of prior observations.

- **Data preprocessing:**

While LSTM can handle non-stationary data directly, ARMA and ARIMA require data stationarity and differencing, which may result in information loss.

- **Handling Irregular Sampling:**

While LSTM can accept erratically sampled data without further preprocessing, ARMA and ARIMA assume normal time intervals between observations, making them less suited for such data.

- **Training Difficulty:**

LSTM training can be computationally costly and necessitates more data for good training, whilst ARMA and ARIMA may require additional effort in parameter estimation.

Security and Performance Analysis

4.1 Data aggregation in secure form

Here in first step the data is firstly encrypted homomorphically to securely send data from smart meters to aggregator node. Aggregator can be software or hardware to collect all the data (readings) from different smart meters and then combine them.

All the data received by the aggregator is in encrypted form to achieve data privacy, data integrity, prevent from unauthorized access and to achieve the consumer trust. Homomorphic encryption enables computation on encrypted material without the need to first decode it. As a result, data may be securely aggregated without the exact data values being made public.

Datetime	Shopping MW	Hospital	Fire Static	Sum	House 1	House 2	House 3	House 4	House 5	House 6	House 7	House 8	House 9
2004-12-31 1:00:00	6.26E+37	6.59E+36	1.17E+37	6.92E+37	1.43E+36	7.41E+37	7.13E+37	2.05E+37	7.94E+37	1.6E+37	2.55E+37	3.1E+37	3.85E+35
2004-12-31 2:00:00	6.74E+37	3.85E+37	9.47E+36	1.06E+38	3.74E+37	9.3E+37	1.56E+37	4.37E+36	7.98E+36	8.97E+37	8.83E+37	4.62E+37	4.03E+37
2004-12-31 3:00:00	3.47E+37	1.06E+38	1.57E+37	1.41E+38	9.18E+37	1.17E+38	4.3E+36	6.37E+37	6.38E+36	9.64E+37	1.05E+38	6.06E+37	1.06E+38
2004-12-31 4:00:00	7.33E+37	2.9E+36	3.69E+37	7.62E+37	4.28E+37	5.89E+37	6.82E+37	6.57E+37	3.92E+37	4.51E+37	2.44E+37	8.43E+37	1.82E+37
2004-12-31 5:00:00	3.32E+36	4.02E+37	1.46E+37	4.35E+37	1.08E+38	4.71E+37	3.49E+37	4.89E+36	3.74E+37	4.47E+37	9.97E+37	4.92E+36	2.32E+37
2004-12-31 6:00:00	4.69E+37	5.45E+37	4.44E+37	1.01E+38	2.89E+37	1.15E+38	6.35E+37	7.29E+37	7.45E+37	4.14E+37	3.84E+37	1.34E+37	3.7E+37
2004-12-31 7:00:00	2.13E+36	7.6E+37	1.16E+38	7.82E+37	6.19E+37	1.09E+38	7.41E+36	2.97E+37	7.02E+36	2.97E+37	8.76E+37	2.32E+37	1.59E+37
2004-12-31 8:00:00	1.35E+37	1.04E+38	1.07E+38	1.18E+38	8.42E+37	6.1E+37	1.03E+38	1.1E+38	1.06E+38	3.4E+37	4.75E+37	8.58E+37	8.07E+37
2004-12-31 9:00:00	3.23E+37	1E+38	1.03E+38	1.32E+38	4.79E+36	4.85E+37	1.14E+38	6.27E+37	1.82E+37	7.19E+37	1.38E+37	8.72E+37	7.42E+37
2004-12-31 10:00:00	1.09E+38	6.3E+37	1.09E+38	1.72E+38	1.91E+37	5.11E+37	7.28E+37	2.25E+37	1.12E+38	5.47E+37	7.46E+37	8.35E+37	9.54E+37
2004-12-31 11:00:00	1.38E+37	1.89E+37	9.86E+37	3.27E+37	4.83E+37	6.17E+36	7.95E+36	1.14E+38	5.45E+37	8.41E+37	1.89E+36	1.89E+36	6.87E+37
2004-12-31 12:00:00	4.3E+36	5.35E+36	9.65E+37	9.65E+36	8.37E+37	8.74E+37	6.38E+37	8.55E+37	1.11E+38	5.23E+37	6.57E+37	1.09E+38	6.54E+37
2004-12-31 13:00:00	9.37E+37	1.38E+37	4.23E+37	1.07E+38	5.9E+37	1.07E+38	6.55E+37	2.79E+37	1.12E+38	7.93E+37	7.11E+37	2.64E+37	7.7E+37
2004-12-31 14:00:00	3.93E+37	9.63E+37	9.04E+37	1.36E+38	1.27E+37	8.87E+37	4.05E+37	1.13E+38	4.85E+37	7.38E+37	8.53E+37	7.17E+37	4.35E+37
2004-12-31 15:00:00	9.73E+36	6.55E+37	7.57E+37	7.53E+37	4.63E+37	9.87E+37	9.85E+37	2.44E+37	9.79E+37	1.11E+38	7.31E+37	1.06E+38	9.19E+37
2004-12-31 16:00:00	3.1E+37	6.8E+37	7.11E+37	9.9E+37	9.55E+37	8.31E+37	5.38E+35	1.73E+36	2.72E+36	1.08E+38	2.46E+37	1.16E+38	1.11E+38

Figure 7: Encrypted Data recived by Aggrigator node

4.2 Grid Station

Grid stations are vital parts of a smart grid because they act as hubs for the transformation of power, distribution management, and integration of cutting-edge technology. Grid stations support a more trustworthy, efficient, and sustainable power supply for customers in a smart grid environment by regulating energy distribution and ensuring system stability.

All the data received from aggregator is in encrypted form by homomorphic encryption using public key of grid station. Firstly, data is decrypted by the grid using grid's public key. When the data is in plain text the grid is trained using the Long short-term memory (LSTM).

Datetime							WATT
2004-12-31 1:00:00	6.92E+37	6.71E+38	4.15E+38	1.01E+38		7.01E+37	13479
2004-12-31 2:00:00	1.06E+38	5.72E+38	3.83E+38	1.82E+38		9.15E+37	12867
2004-12-31 3:00:00	1.41E+38	9.48E+38	3.2E+38	1.22E+38		7.61E+37	12577
2004-12-31 4:00:00	7.62E+37	7.86E+38	2.58E+38	1.98E+38		2.54E+37	12519
2004-12-31 5:00:00	4.35E+37	6.33E+38	4.23E+38	2.22E+38		6.2E+37	12671
2004-12-31 6:00:00	1.01E+38	7.23E+38	2.85E+38	1.63E+38		4.51E+37	13040
2004-12-31 7:00:00	7.82E+37	6.14E+38	3.05E+38	1.83E+38		4.61E+37	13692
2004-12-31 8:00:00	1.18E+38	8.01E+38	2.77E+38	1.44E+38		2.01E+37	14298
2004-12-31 9:00:00	1.32E+38	7.51E+38	2.37E+38	1.47E+38		4.75E+37	14720
2004-12-31 10:00:00	1.72E+38	9.17E+38	2.43E+38	1.53E+38		7.59E+37	14940
2004-12-31 11:00:00	3.27E+37	6.31E+38	4.11E+38	2.09E+38		5.61E+37	15184
2004-12-31 12:00:00	9.65E+36	9.56E+38	3.84E+38	2.01E+38		4.42E+36	15007
2004-12-31 13:00:00	1.07E+38	1.05E+39	2.32E+38	1.42E+38		8.68E+37	14806
2004-12-31 14:00:00	1.36E+38	8.69E+38	3.71E+38	2.32E+38		2.16E+37	14522
2004-12-31 15:00:00	7.53E+37	1.01E+39	3.4E+38	2.81E+38		7.48E+36	14348
2004-12-31 16:00:00	9.9E+37	7.65E+38	4.08E+38	2.49E+38		3.3E+37	14108
2004-12-31 17:00:00	1.04E+38	8.32E+38	2.55E+38	1.54E+38		1.74E+36	14409
2004-12-31 18:00:00	1.35E+38	1.02E+39	2.27E+38	1.64E+38		8.05E+37	15176
2004-12-31 19:00:00	6.48E+37	7.65E+38	3.62E+38	2.61E+38		1.34E+37	15260

Figure 8: Encrypted sum of data received Grid Station.

4.3 Time to Encrypt Data

The phrase "Time to Encrypt Data" describes how long it takes to use cryptographic techniques to transform plaintext data into an encrypted, safe version. It includes the

processing time needed for encryption operations and might change depending on the encryption technique, the amount of data, and the technology available.

```
D:\pythonProject3\research\Scripts\python.exe "D:\pythonProject3\check 5.py"  
Time taken for encryption: 141.221133 seconds
```

Figure 9: Time taken to Encrypt data.

Total time taken to encrypt 121273 x 29 values is 141.221133seconds.

Time taken to encrypt a single value is:

$$\begin{aligned} \text{Encryption time for single value} &= \text{total time} / \text{values} \\ &= 141.221133 / (121273 * 29) \\ &= 141.221133 / 3516917 \end{aligned}$$

Encryption time for single value = 0.00004015481 second

4.4 Size of Encrypted and Non-Encrypted Data

The phrase "Size of Encrypted and Non-Encrypted Data" refers to a comparison between the amount of storage space needed to store data in its unencrypted, original form and the amount of space needed when the data is encrypted using cryptographic methods. The storage capacity, communication bandwidth, and overall system performance may all be impacted by this size disparity.

```
Run  
D:\pythonProject3\research\Scripts\python.exe "D:\pythonProject3\ol  
Non-encrypted value: 437  
Size of non-encrypted value: 28 bytes  
Size of encrypted value: 44 bytes
```

Figure 10: Size of Encrypted and Non-Encrypted Data

From the above fig the size of encrypted data is more then the non-encrypted data.

- Size of Encrypted data: 28 bytes

- Size of Non- Encrypted data: 44 bytes

Encrypted data typically occupies a larger storage footprint compared to its plaintext counterpart due to the inclusion of all necessary decryption information, excluding the encryption key itself. In many cases, encrypted data necessitates approximately 33 percent more storage capacity than unencrypted data.

4.5 Graphs and Results

The result of a thesis embodies the culmination of rigorous research, offering a profound contribution to knowledge within a specific academic domain. It showcases the author's analytical prowess and the transformative impact of their insights on the subject.

Energy consumption is show of different years from 2004 to 2018, x-axis represents the Dates in year while y-axis represent the Energy in MegaWatt.

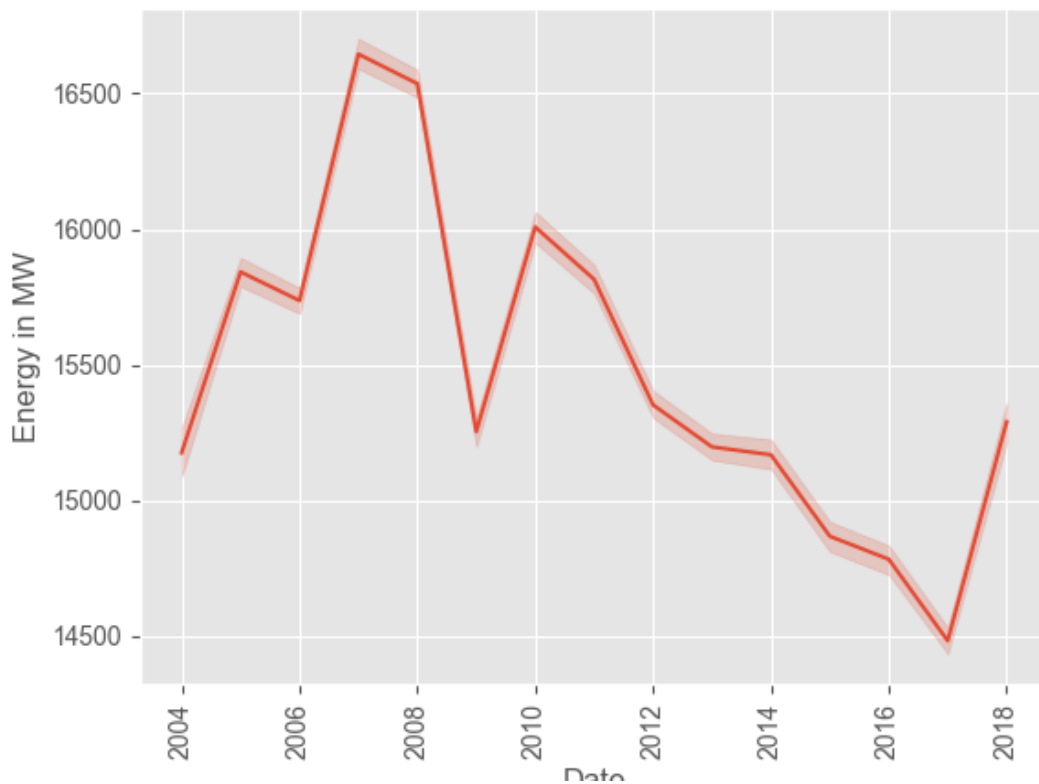


Figure 11: Energy Consumption According to year

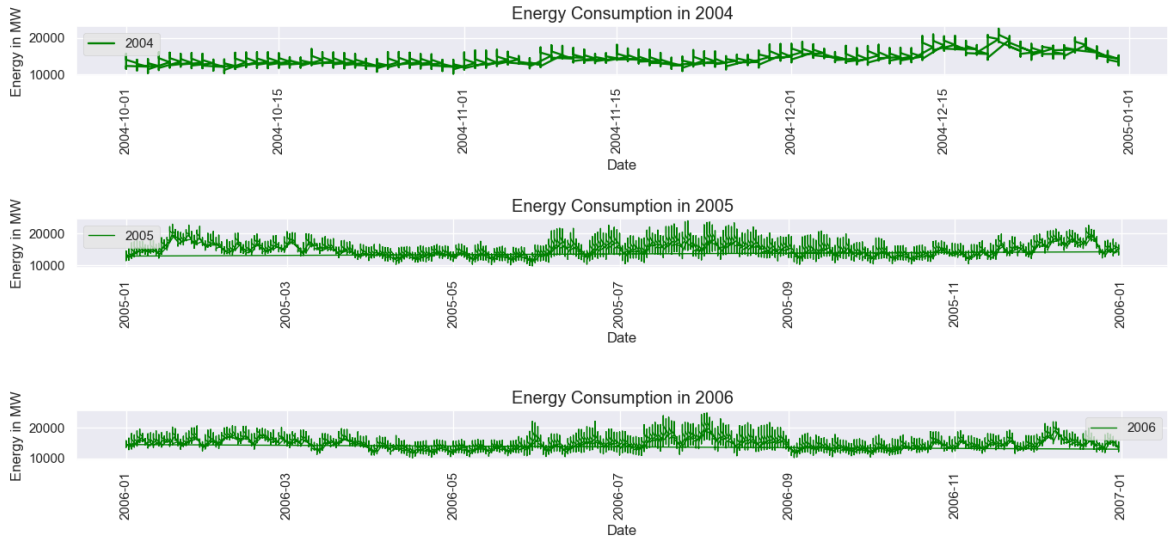


Figure 12: Energy Consumption According to year

In Fig 17 the graph show the magnitude of MegaWatt. X-axis represents the Dates in year while Y-axis represent the Energy in MegaWatt.

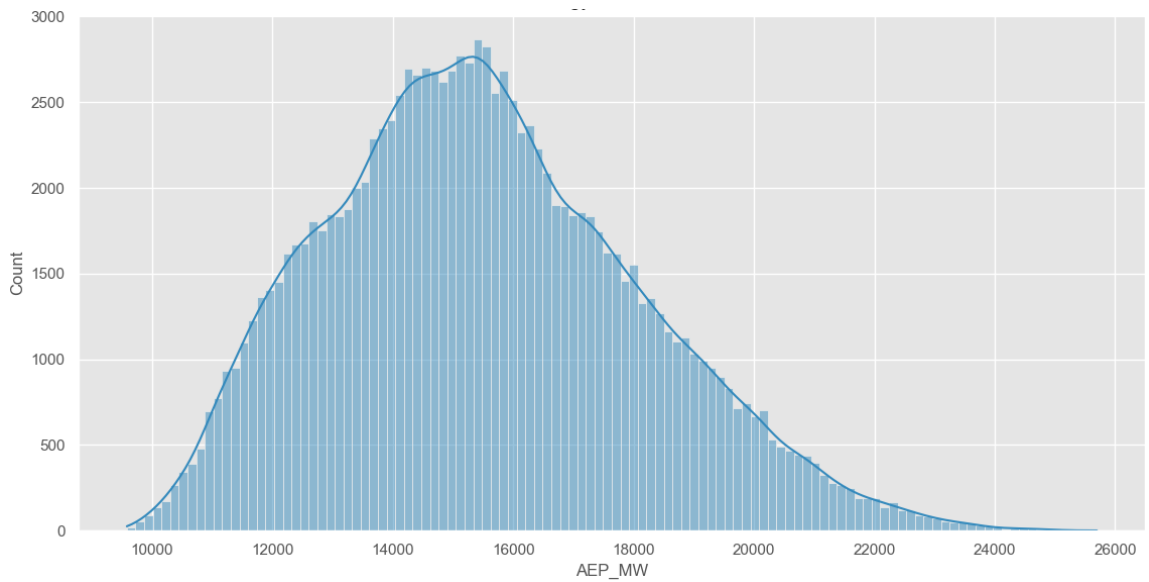


Figure 13: Energy Distribution

The "Machine Learning the Pattern Predicting Future Value" graph visually represents the application of machine learning techniques to predict future values based on historical data patterns. It typically consists of two main components:

Historical Data Points:

This part of the graph displays the historical data points or observations that have been collected over time. These data points represent the past values of a particular variable or phenomenon that we want to predict in the future.

Predicted Future Values:

The graph also includes a line or curve that represents the predictions made by the machine learning model. This line is generated by training the model on the historical data, learning the underlying patterns, and then extrapolating those patterns into the future.

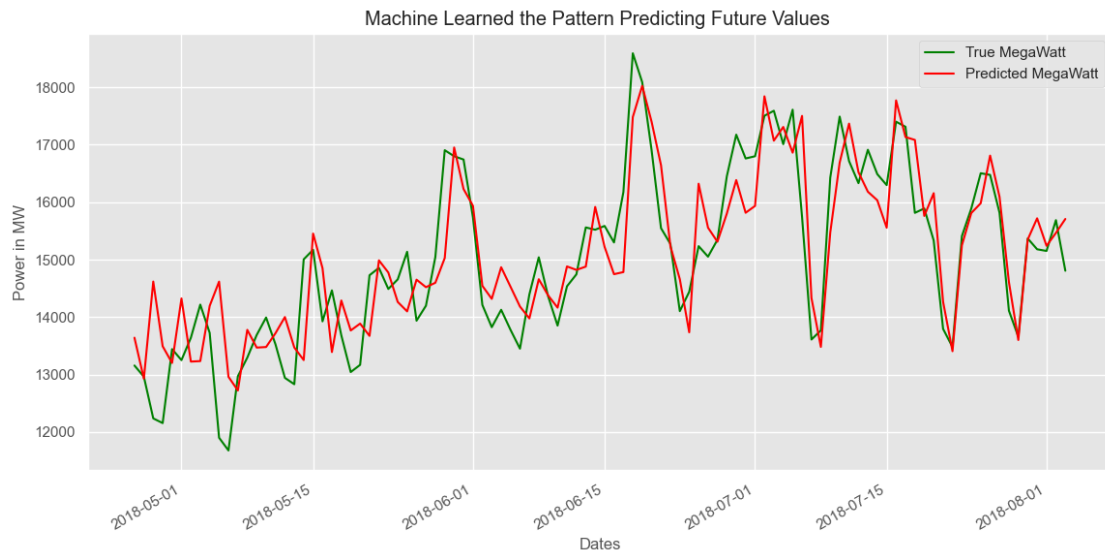


Figure 14: Machine Learning the Pattern Predicting Future Value

4.6 Mean Absolute Percentile Error (MAPE)

A statistic called Mean Absolute Percentage Error (MAPE) is used to assess how accurate a forecast or prediction model is. It calculates the percentage that separates a dataset's actual values from its anticipated values. When assessing a model's performance in scenarios where the data's size fluctuates much, MAPE is very helpful.

The following formula may be used to determine MAPE:

Equation 5: MAPE Equation

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|$$

- n is the total number of data points in the dataset.
- A_i is the actual value of the i th data point.
- F_i is the predicted or forecasted value of the i th data point.

As a percentage, the MAPE value is the average percentage difference between the actual and anticipated values.

Since the predicted values are closer to the actual values, a model with a lower MAPE value is considered more accurate.

The range of MAPE is 0% to infinity. Greater MAPE values indicate bigger prediction errors, whereas a value of 0% indicates that the model's predictions exactly match the actual data.

In the context of time series analysis and regression issues, the Mean Absolute Percentage Error (MAPE) is a statistic used to assess the accuracy of a forecasting or prediction model.

Mean Absolute Percentile Error (MAPE): 4.60%

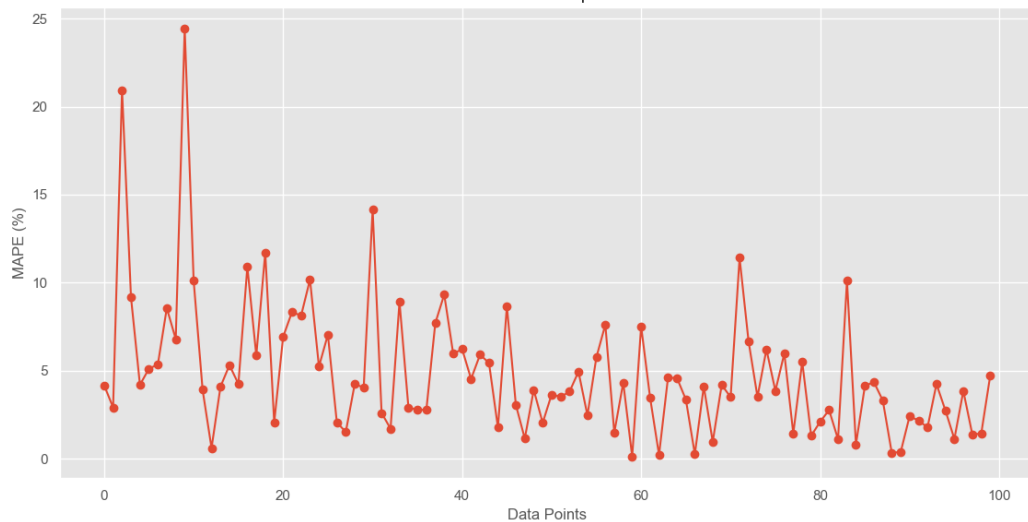


Figure 15: MAPE Graph

4.7 User Data Privacy

By leveraging the mathematical properties of Paillier Homomorphic Encryption, the preservation of user data privacy within the smart grid becomes achievable. The utilization of Paillier encryption involves the following strategies to ensure privacy:

Secure Data Transmission:

Prior to transmission, user-specific data, such as energy consumption patterns, undergoes Paillier encryption using the Grid Station's Public Key (PK). This encrypted data is then forwarded to the Aggregator Node, guaranteeing the confidentiality of consumption details during communication.

Aggregated Data Handling:

The Aggregator Node performs data aggregation, consolidating encrypted energy consumption information from various users without decryption. By capitalizing on Paillier's homomorphic property, computations like data averaging or summation can be directly executed on the encrypted data, effectively safeguarding the individual consumption patterns' privacy.

Preservation of Demand Response Privacy:

The implementation of demand response initiatives is achievable without compromising individual user behaviors. Encrypted data empowers the smart grid to optimize energy usage while upholding user privacy.

Through these Paillier Homomorphic Encryption mechanisms, sensitive user data privacy within the smart grid is upheld, allowing for efficient grid management, secure communications, and data analysis while maintaining the utmost privacy standards.

Conclusion and Future Work

According to this research, power demand forecasting using LSTM demonstrates superior accuracy and lower error rates compared to conventional statistics-based forecasting models like AR, ARMA, and ARIMA. Therefore, using LSTM to estimate power consumption with further hyper tuned parameters and optimizations can be successful and produce superior outcomes. The accessibility of the long-term historical data will have a significant impact on the model's effectiveness.

5.1 Future Work

The future advancements of smart meter technology are poised to significantly enhance load profiling and forecasting capabilities. Concurrently, research efforts will be directed towards the creation of intuitive user interfaces and visualization tools designed to offer customers lucid insights into their energy consumption patterns and associated costs. Moreover, paramount attention will be given to addressing privacy apprehensions stemming from the usage of smart meter data. In this context, the development of robust and secure billing mechanisms, which uphold the sanctity of consumer data while ensuring confidentiality, will be of paramount importance. Through these multifaceted initiatives, the smart meter ecosystem is poised to achieve a harmonious convergence of technological innovation, customer engagement, and data security.

References

- [1] "Federal Energy Regulatory Commission Assessment of Demand Response & Advanced Metering" (PDF). FERC.gov. Retrieved 16 January 2018.
- [2] Zhang, Tuo, Lei Gao, Chaoyang He, Mi Zhang, Bhaskar Krishnamachari, and A. Salman Avestimehr. "Federated Learning for the Internet of Things: Applications, Challenges, and Opportunities." *IEEE Internet of Things Magazine* 5, no. 1 (2022): 24-29.
- [3] Asad, Muhammad, Ahmed Moustafa, and Takayuki Ito. "Federated Learning Versus Classical Machine Learning: A Convergence Comparison." arXiv preprint arXiv:2107.10976 (2021).
- [4] D. L. Marino, K. Amarasinghe, and M. Manic. Building energy load forecasting using Deep Neural Networks. In IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, pages 7046– 7051, October 2016.
- [5] Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans. Inf. Forensics Secur.* 2017, 13, 1333–1345.
- [6] Li, T.; Hu, S.; Beirami, A.; Smith, V. Ditto: Fair and robust federated learning through personalization. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 6357–6368.
- [7] F. P. -C. Lin, S. Hosseinalipour, S. S. Azam, C. G. Brinton and N. Michelusi, "Semi-Decentralized Federated Learning With Cooperative D2D Local Model Aggregations," in *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3851-3869, Dec. 2021, doi: 10.1109/JSAC.2021.3118344.
- [8] Asmaa Abdallah and Xuemin Shen, "Lightweight Security and Privacy Preserving Scheme for Smart Grid Customer-Side Network", *IEEE TRANSACTIONS ON SMART GRID*, VOL. 8, NO. 3, MAY 2017.
- [9] Mothukuri, Virraji; Parizi, Reza M.; Pouriye, Seyedamin; Huang, Yan; Dehghantaha, Ali; and Srivastava, Gautam, "A survey on security and privacy of federated learning" (2021). *Faculty Publications*. 5650.

- [10] Dhaou Said et al. Advanced scheduling protocol for electric vehicle home charging with time-of-use pricing. pages 6272–6276, June 2013. ISSN: 1938-1883
- [11] T. Alquthami, M. Zulfiqar, M. Kamran, A. H. Milyani and M. B. Rasheed, "A Performance Comparison of Machine Learning Algorithms for Load Forecasting in Smart Grid," in *IEEE Access*, vol. 10, pp. 48419-48433, 2022, doi: 10.1109/ACCESS.2022.3171270.
- [12] S. Li, K. Xue, Q. Yang and P. Hong, "PPMA: Privacy-Preserving Multisubset Data Aggregation in Smart Grid," in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 462-471, Feb. 2018, doi: 10.1109/TII.2017.2721542.
- [13] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh and Z. A. Khan, "An Accurate and Fast Converging Short-Term Load Forecasting Model for Industrial Applications in a Smart Grid," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2587-2596, Oct. 2017, doi: 10.1109/TII.2016.2638322.
- [14] Peng Kou, Feng Gao, A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises, *International Journal of Electrical Power & Energy Systems*, Volume 55, 2014, Pages 144-154, ISSN 0142-0615
- [15] M. Akgün, E. U. Soykan and G. Soykan, "A Privacy-Preserving Scheme for Smart Grid Using Trusted Execution Environment," in *IEEE Access*, vol. 11, pp. 9182-9196, 2023, doi: 10.1109/ACCESS.2023.3237643.
- [16] Department for Business Energy and Industrial Strategy, "Smart Meters, Quarterly Report to end December 2016, Great Britain," Tech. Rep., 2017.
- [17] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security and Privacy*, vol. 8, no. 1, pp. 81–85, 2010.
- [18] H. Li, X. Lin, H. Yang, X. Liang, R. Lu and X. Shen, "EPPDR: An Efficient Privacy-Preserving Demand Response Scheme with Adaptive Key Evolution in Smart Grid," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2053-2064, Aug. 2014, doi: 10.1109/TPDS.2013.124.
- [19] S. Yan, "Understanding lstm and its diagrams," *MLReview.com*, 2016.
- [20] S. Ray, "A Quick Review of Machine Learning Algorithms," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

- [21] Ibrahim, Ibrahim, and Adnan Abdulazeez. "The role of machine learning algorithms for diagnosing diseases." *Journal of Applied Science and Technology Trends* 2, no. 01 (2021): 10-19.
- [22] W. H. Sanders, "Progress towards a resilient power grid infrastructure," in Proceedings of the IEEE Power & Energy Society General Meeting, July 2010.
- [23] <https://www.sageautomation.com/blog/traditional-grids-vs-smart-grids-why-were-making-the-shift>
- [24] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in International conference on the theory and applications of cryptographic techniques, Berlin, Heidelberg, Springer, 1999, pp. 223-238.
- [25] Paillier, Pascal. "Public-key cryptosystems based on composite degree residuosity classes." International conference on the theory and applications of cryptographic techniques. Springer, Berlin, Heidelberg, 1999.
- [26] Yining Liu, Wei Guo, Chun-I Fan, Liang Chang, and Chi Cheng, "A Practical Privacy-Preserving Data Aggregation (3PDA) Scheme for Smart Grid" IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 15, MARCH 2019.
- [27] Zhitao GUANI, Yue ZHANG¹, Liehuang ZHU, Longfei WU & Shui YU, "EFFECT: an efficient flexible privacy-preserving data aggregation scheme with authentication in smart grid" SCIENCE CHINA Information Sciences, Vol. 62, March 2019.
- [28] Debiao He, Neeraj Kumar, Sherali Zeadally, Alexey Vinel, and Laurence T. Yang, "Efficient and Privacy-Preserving Data Aggregation Scheme for Smart Grid Against Internal Adversaries" IEEE TRANSACTIONS ON SMART GRID, VOL. 8, NO. 5, SEPTEMBER 2017.
- [29] Asmaa Abdallah and Xuemin (Sherman) Shen, "A Lightweight Lattice-Based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart Grid" IEEE TRANSACTIONS ON SMART GRID, VOL. 9, NO. 1, JANUARY 2018.
- [30] P. Gope and B. Sikdar, "Lightweight and Privacy-Friendly Spatial Data Aggregation for Secure Power Supply and Demand Management in Smart Grids," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 6, pp. 1554-1566, June 2019, doi: 10.1109/TIFS.2018.2881730.
- [31] Echelon Corporation, Echelon System Software, 2013.

- [32] P. Yeung, and M. Jung, Improving Electric Reliability with Smart Meters, White Paper, Silver Spring Networks, 2013.
- [33] Y. Tsiounis and M. Yung. On the security of ElGamal based encryption. In H. Imai and Y. Zheng, editors, Public Key Cryptography, volume 1431 of Lecture Notes in Computer Science, pages 117–134. Springer, 1998.
- [34] Khan AR, Mahmood A, Safdar A, Khan ZA, Khan NA. Load forecasting, dynamic pricing and DSM in smart grid: A review. Renew Sustain Energy Rev. 2016;54:1311-1322.1430270684635 Nasir_786