A Tensor-based Big Data Management Scheme for Dimensionality Reduction Problem in Smart Grid Systems

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Engineering

in

Computer Science & Engineering

Submitted By

Devinder Kaur

(Roll No. : 801532011)

Under the Supervision of:

Dr. Neeraj Kumar

Associate Professor

CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

July 2017

Certificate

I hereby certify that the work which is being presented in the thesis entitled, *'Tensor-based Big Data Management Scheme for Dimensionality Reduction Problem in Smart Grid System'*, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Neeraj Kumar*.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Devinder Kaur)
ME-CSE
Roll no. 801532011

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Neeraj Kumar)
Associate Professor
CSED
Thapar University
Patiala

Acknowledgment

I have been waiting long for this moment to acknowledge all those who contributed in building this work. It is my pleasure to thank all of them here. First of all, I offer my sincere gratitude to my supervisor, *Dr. Neeraj Kumar* for accepting to be my supervisor. Without his help and encouraging advice, this work would have never begun. I am deeply indebted to him for providing me wonderful research atmosphere and platform to explore my research to the fullest.

I am also grateful to *Dr. Maninder Singh*, Head, CSED for providing me the opportunity to conduct my research work. I would also extend my gratitude to *Mr. Gagangeet singh Aujla* who played a significant role in clarifying my doubts related to Software Defined Networks. I would also like to thank the Director of the institute, *Prof. Prakash Gopalan* for his continuous support.

My special thanks goes to my friends for discussing thoughts and sharing all ups and downs with me during the course of this work. At the same time I would also like to thank all my colleagues for their continuous support.

Last but not the least I would like to thank my parents and family members, who made me capable of reaching this point of life and for giving me their kind support and love. I dedicate my work to them.

Devinder Kaur
ME (Computer Science and Engineering)
Roll No.:801532011

Abstract

Smart grid (SG) is an integration of traditional power grid with advanced information and communication infrastructure for bidirectional energy flow between grid and end users. A huge amount of data is being generated by various smart devices deployed in SG systems. Such a massive data generation from various smart devices in SG systems may generate issues such as-congestion, and available bandwidth on the networking infrastructure deployed between users and the grid. Hence, an efficient data transmission technique is required for providing desired QoS to the end users in this environment. Generally, the data generated by smart devices in SG has high dimensions in the form of multiple heterogeneous attributes, values of which are changed with time. The high dimensions of data may affect the performance of most of the designed solutions in this environment. Most of the existing schemes reported in the literature have complex operations for data dimensionality reduction problem which may deteriorate the performance of any implemented solution for this problem. To address these challenges, in this paper, a tensor-based big data management scheme is proposed for the problem of dimensionality reduction in big data generated from various smart devices. In the proposed scheme, firstly the Frobenius norm is applied on high-order tensors (used for data representation) to minimize the reconstruction error of the reduced tensors. Then, an empirical probability-based control algorithm is designed to estimate an optimal path to forward the reduced data using software-defined networks (SDN) for minimization of the load and effective bandwidth utilization on the network infrastructure. The proposed scheme minimizes the transmission delay occurred during the movement of the dimensionally reduced data between different nodes. The efficacy of the proposed scheme has been evaluated using extensive simulations carried out on the data traces (power consumption of appliances in different smart homes) using 'R' programming and Matlab. The results obtained depict the effectiveness of the proposed scheme with respect to the parameters such as- network delay, accuracy, and throughput.

Keywords: Big data , dimensionality reduction, smart grid, software-defined networks, tensors.

Contents

List of Figures

List of Tables

List of Abbreviations

AMI	Advanced Metering Infrastructure
API	Application Programming Interfaces
DOE	Department of Energy
DC	Data Centers
DR	Dimensionality Reduction
DRM	Demand Response Management
EPCS	Empirical Based Control Scheme
F-HOSVD	Frobenius based HOSVD
FNs	Forwarding nodes
HOSVD	High-Order Singular Value Decomposition
IDE	An integrated Development Environment
NIST	National Institute of Standards and Technology
NOS	Network Operating Systems
OFP	Open Flow Protocol
PCA	Principal Component Analysis
PMU	Phasor Measurement Units
RMT	Random Matrix theory
SCADA	supervisory control and data acquisition
SDN	Software Defined Networks
SG	Smart Grid
SVD	Singular Value Decomposition

Chapter 1

Introduction

1.1 Smart Grid (SG)

The Smart grid (SG) is an intelligent power grid which supports a bidirectional flow of energy and information between users and grid using advanced ICT based infrastructure. It envisions to optimize user's demand, energy generated, and network availability to provide reliability and efficiency using automated controls, sensors, metering devices, and distributed energy sources. It contains components such as smart meters and sensing devices connected to one another using communication infrastructure. The delivery of various services such as energy, voltage, and frequency regulations to the end users depends upon the reliable, and real-time information about the data flow between users and service providers.

According to a report given by US DOE, the energy demand increases every year by almost 4% percent. Relying upon the traditional grid systems for such extensive energy demands in the era of 21st century is quite precarious. The main purpose of traditional power grid was to light up the world when it first came into the existence in 1890s. Although, we have been relying on existing grid for so long. But with the advent of modernization and digital world, it is very difficult to cope with energy demands of end users. There are various issues with traditional grid which become features and challenges for SG and which are discussed in the next section of this chapter. SG envisions to be the modern grid of 21st century. It represents an unrivalled opportunity to move the traditional power grid towards a new era of reliability, sustainability, and efficiency. Also, it possesses the tendency to improve economic and environmental health. SG is evolving bit by bit. Once mature, it will revolutionize the lives of consumers the same way internet has done. SG consists of many components like control and data center, generation units, power lines, new technologies, smart devices, substations, etc. An overview consisting of various components of SG is given in Figure 1.1.

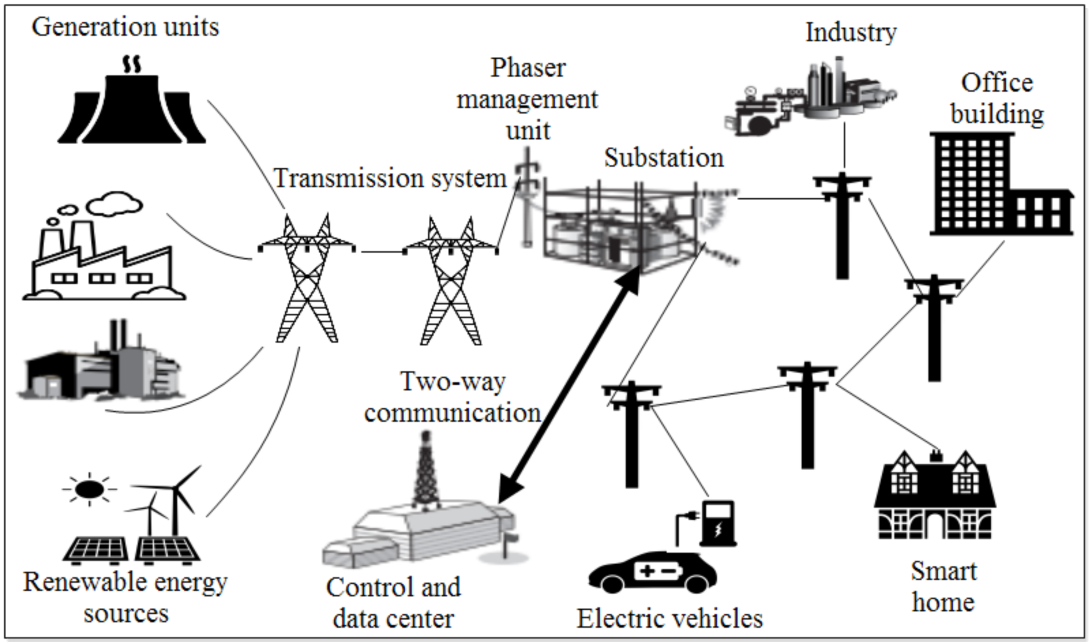


Figure 1.1: An overview of SG

1.1.1 Taxonomy of SG systems

According to the SG conceptual model given by NIST, SG is broadly divided into three systems. The systems are further categorized into various subsystems as shown in the Figure 1.2. Together, they work for various aspects of SG such as energy distribution, transmission, data generation, service providers, end users, and many other issues. The systems have been categorized as below:

1. **Smart Infrastructure System**: In SG systems, infrastructure related to the components such as energy and information handling, communication management, etc. comes under the domain of smart infrastructure system. It mainly deals with the two-way flow of energy and information between end users and grid. It plays an important role in other subsystems of SG. This system is further divided into following subsystems:

   - **Smart Energy Subsystem**: It deals with all the energy related issues such as energy generation and transmission, distribution grid. In fact, the new grid

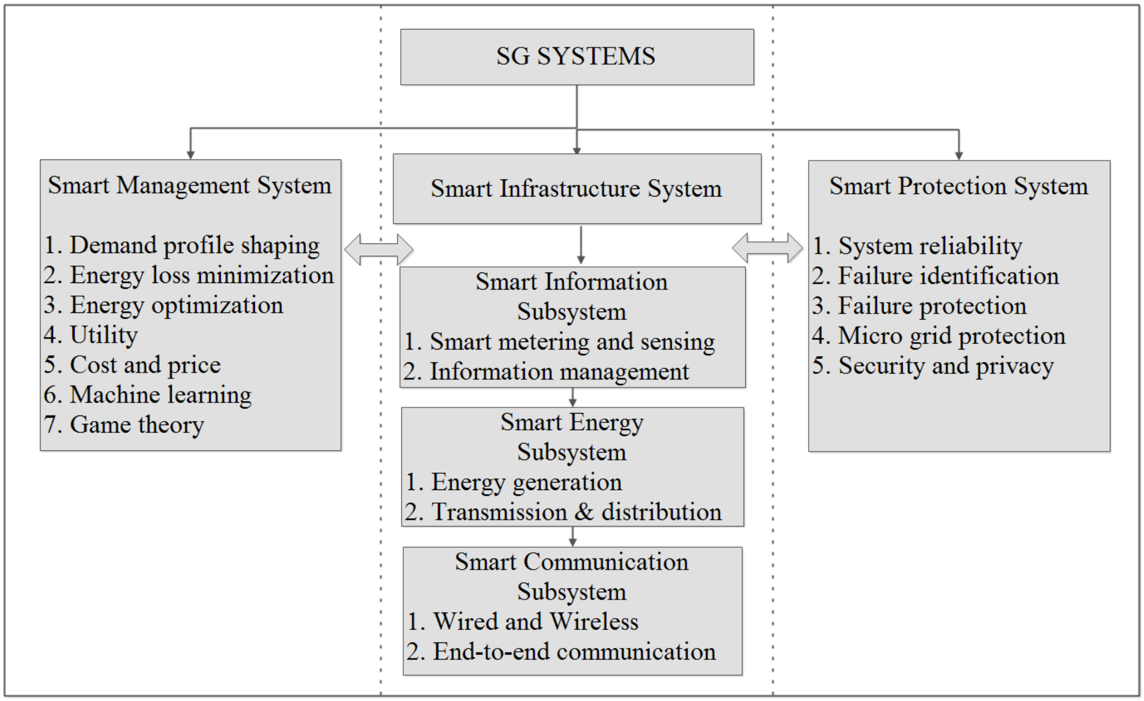


Figure 1.2: Taxonomy of SG systems and subsystems with their respective domains

paradigms such as micro grids and vehicle-to-grid (and vice versa) comes under the domain of smart energy subsystem.

- **Smart Communication Subsystem**: This subsystem as the name suggests involves type of communication that the SG involves in its systems for the flow of energy and information. Mainly the wireless, wired, and end-to-end communication types exist in SG systems for communication purposes between the end users and grid. Various communication technologies available for SG systems are named as below:

  (a) *Zigbee*

  (b) *Power line communication (plc)*

  (c) *Wireless mesh*

  (d) *Cellular network communication*

- **Smart Information Subsystem**: This subsystem mainly consists of the smart and sensing devices such as sensors, phasor measurement units (PMUs), Advanced metering infrastructure (AMIs). The metering and measurement of the generated information using above mentioned devices is usually done under the domain of smart information subsystem. Further, this captured information can be managed efficiently and can be used for data analysis and data mining purposes in the SG systems.

2. **Smart Management System**: By utilizing the benefits provided by smart infrastructure system, this system contributes towards the efficient energy management in SG systems. It involves various management tools, methods, and algorithms such as related to machine learning and energy optimization. Further, other energy related objectives such as utility cost, utility maximization, minimizing the energy loss, demand response management comes under the domain of smart management systems.

3. **Smart Protection System**: As the name suggests, this system deals with the security and privacy related aspect of SG systems. Broadly, the issues related to the grid reliability, power failures, and cyber security are the issues which are managed by the smart protection system.

1.1.2 Integrated Features and Challenges of SG Systems

There are various shortcomings of traditional power grid, that ultimately serve as the integrated features and challenges for SG systems. The detailed description of the SG features and and challenges is given as follows:

1. **Reliability**: In accordance with the recent stats given by US DOE report, there have been massive blackouts in the past years. One of the main reasons behind such outages was the poor response time of mechanical switches in the traditional grids. Today, a power outage such as a blackout or brownout put a blight effect on day to day lives of energy consumers. It may lead to a number of failures that can affect day to day applications such as banking, transactions, traffic, security, just to name a few. Today, almost anything to everything depends on electricity or some other form of power. So, reliability is one of the major concerns in existing grid systems which needs to be dealt with.
   SG makes use of modern ICT technologies and infrastructure such as AMI and PMU which work towards making the grid more reliable and efficient. The real time information generated from these infrastructure units plays a major role to deliver a reliable energy service to end users. SG systems would add resiliency to grid and make it more efficient in addressing the natural and man made calamities such as storms, earthquakes, terrorist attacks, etc. SG being bidirectional in nature, would allow for automatic rerouting when equipment fails or any disruption occurs. SG technologies would detect and isolate the outages, before they become large-scale blackouts.

2. **Flexibility and Scalability**: More flexibility in a grid system means system is able to respond to the load fluctuations in a better way. So, flexibility is one of major key features of SG systems. Using modern communication technologies such as zigbee, plc, cellular networks etc., SG systems foster the flexibility and capacity in the grid and provide advanced control and sensing services. Also, due to the presence of digitalized platform in smart transmission system it becomes more flexible to control and operate the grid operations.

3. **Sustainability**: The energy generated from the renewable energy sources such as from sun, wind, water, etc. play a vital role in energy conservation and sustainability. SG being flexible enough, renewable energy resources can be incorporated with the it and significant amount of energy can be stored and generated. eing bidirectional in nature, there are many ways to store and generate the energy at the user end as well. For example using SG techniques such as grid-to-vehicle and vehicle-to-grid. This technique involves electric vehicles which are used to generate energy at the the user end. It leads to Reducing greenhouse gas emissions.

4. **Improved Security**: Although, SG is intended to be more resilient towards natural disasters and power outages. It is self healing as well. But, security and privacy control is one of the main challenges in SG systems. SG is a complex system as it involves large number of smart devices and end users connected to each other and communicating with each other bidirectionally. The data generated from such devices can also contain sensitive information about the users which needs to be protected. So, in modern grid security and privacy is a big concern and goal as well, which SG intends to achieve.

5. **Situational Awareness**: The lack of situational awareness in traditional grid systems makes it more vulnerable, inefficient, unreliable, and least scalable. However in SG systems, to boost up the situational awareness among end users is one of the major goals. It involves understanding various grid components, problems, behaviors, and energy optimization solutions for the better management of energy and grid.

1.2 Big Data

Big data can be defined as the massive datasets which are so huge and complex in nature that it becomes very difficult to handle them using traditional database systems. While talking about big data, data with exabytes to zettabytes comes under this range. The growth of big data generation is increasing day by day due to the hype of online systems, multimedia, online gaming, and social networking sites. Today, a huge and highly heterogeneous data gets generated every second at faster speeds than ever before.

1.2.1 Characteristics of Big Data

To get a clear idea of big data and its components, it is often described using 5 Vs. Figure 1.3 shows the essential characteristics of big data. The main characteristics of big data are described in the form of various Vs which are explained in this subsection as given below:

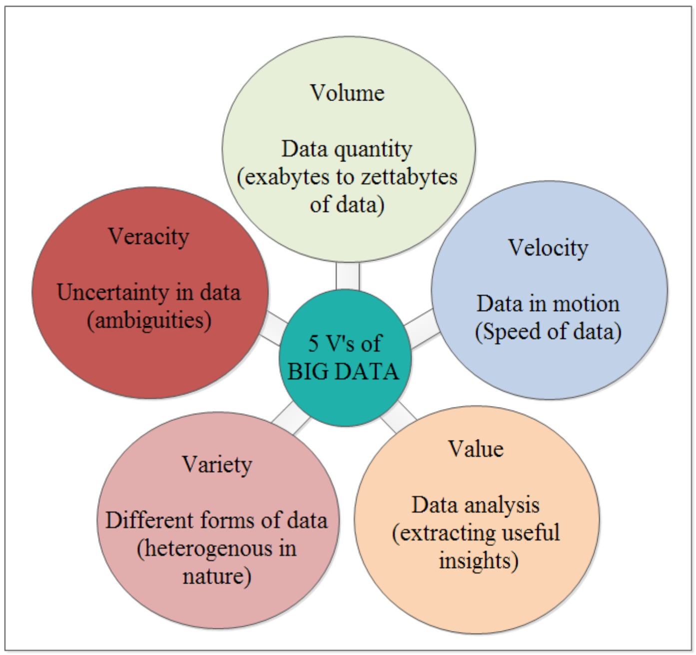


Figure 1.3: Characteristics of big data

1. **Volume**: The first and foremost essential V of big data is volume which defines the "size" of big data. It refers to the large amount of data being generated at regular events at a very fast speed. The size range of such data lies from zettabytes to yottabytes or it could be even more large. Online systems, multimedia, social networking sites like twitter and facebook generate such huge amount of data that too at a very fast speed.

2. **Velocity**: This "V" of big data refers to the speed of data with which it gets generated. Generally, the mobility of data defines its velocity. Today, a massive amount of data gets generated at real time within a fraction of seconds. Moreover, it gets analyzed

before it goes for storage. Examples for the same involve - speed of the credit card transactions, something going viral on social media, just a few to name.

3. **Veracity**: Veracity refers to the uncertainty present in data. Due to large volume and high speed of data generated, clarity or accuracy of data becomes quite challenging.

4. **Variety**: Majority of the data which gets generated is heterogeneous in nature. The category under which we can define the nature of data comes under the variety characteristics of data. Structured, unstructured, semi structured are the main types of big data.

5. **Value**: It defines the useful data insights that can be drawn from big data after applying big data techniques like data analysis. Various big data and machine learning techniques can be applied onto the data acquired to generate the value data and information, which plays role in big data applications such as decision making process.

1.3 Sources of Big Data Generation in SG Systems

Advanced metering infrastructure (AMI) and Phasor measurement units (PMU) are two main infrastructure units used for acquiring the data generated from different smart devices in SG systems and then pass the collected data to the utility which takes decisions about energy flow. AMIs are the bi-directional units which contain sensing devices, smart meters, control and monitoring systems, and data management units. On the other hand, PMUs are the energy measurement units generally used to measure energy waves and signals [2]. These devices are deployed in different smart communities in large number across the globe for effective power management generated from various distributed energy resources in SG environment. For example, approximately 50,000 smart meters are deployed by US DOE and Los Angeles department of water and power (LADWP) in the Los Angeles itself. LADWP serves approximately 4.1 million consumers which accounts to nearly 1% of total US power usage [3]. Another report by energy information administration (EIA), highlights the need of an efficient data management with higher penetration of renewable energy sources (RES) to manage demand and supply optimally [4]. For this purpose, a reliable communication infrastructure is required to manage the flow of data and information between sources of data generation and smart meters.

The basic electric grid contains various substations and a network of transmission lines. On the other hand, SG is having advanced infrastructure and information technologies equipped with the sensing and metering technologies in the from of AMIs and PMUs. It intends to manage the booming demands and needs of power in the 21st Century. It is a two way communication between the grid and its end users. End-users basically include grid and the customers on the other hand. The SG infrastructure mainly composed of the automated controls, sensors, computers, smart energy distribution and transmission systems, and communication technologies as explained in the first section of this chapter.

1.4 Challenges of Big Data in SG Systems

Handling the large amount of data generated by smart devices in different time intervals using AMIs and PMUs is one of the biggest challenges in SG systems. The enormous load of data acquired at the SG level often leads to the problems related to QoS provisioning and demand response management [2]. According to the various surveys, the data generated by smart grid qualifies for the main characteristics of big data, including volume, velocity, variety, veracity, and value. The data generated is in massive amount in the petabytes and terabytes. So it is very difficult to process this data. In this context various techniques to process SG data in literature survey has been described by various authors.

1. **Data Representation**: To represent the large amount of data generated by SG systems is one of the foremost challenge that traditional big data techniques face. PMUSs and AMIs generate big volumes of data at high speed at regular intervals. There is a need of an efficient management technique to represent this data, so that further operation on this data becomes easier.

2. **Data Storage and Management**: Now after representing the big data generated by smart devices, another challenge is to store the data and manage it in an efficient way. If managed properly, then this data can be used for various application such as data intelligence and data mining. It can be used to draw useful insights related to energy utilization. For the purpose of efficient data management, tensors can be used which are multidimensional arrays. However, for the efficient storage purposes, cloud computing is also an important paradigm.

3. **Data Processing**: After representing and storing the big data generated from SG devices, to process it with ease and efficiency is one of the main challenges offered by big data to SG systems. As of today, there are various big data techniques available to handle and process the big data over underlined network. But in the context of SG there are few. There is a need of an efficient networking paradigm, which is flexible and scalable enough to process the SG big data over the network.

1.4.1 Problem of Dimensionality Reduction in Big Data

According to the latest stats, data generation is growing at faster rates than ever before. The huge amount of big data generated at such a high speed posses multiple dimensions. Due to the heterogeneous nature of big data, problem of high dimensionality arises in data sets. The heterogeneous nature of data involves structured, unstructured, and semi structured type which comes under the variety characteristic of big data. Generally, high dimensionality comes from applications such as, multimedia, mobiles, social networking sites etc. The different variety of data such as texts, xmls, photos, videos, audios, web pages, etc. make the dimensionality of data multiple and thus highly complicated. The number of attributes in datasets has become huge along with the huge number of instances which are very difficult to handle using traditional data handling techniques. It makes the data more complex and to handle such data increases both the time and space complexity. Although various big data techniques are present to handle this problem as described in the coming chapter as well. But there is need of a more efficient and integrated technique which can deal with the problem of high dimensionality in SG systems. In general, dimensionality reduction is defined as the process of reducing the number of variables or attributes form data sets by obtaining a set of principal values. In machine learning and statistics, usually two techniques are involves during the preprocessing of data which are know as feature selection and feature extraction. Feature selection is the process of selecting a subset of relevant features or attributes form the main data sets using some specific approach depending upon the type of dataset. It is also known as attribute selection or variable selection. Basically, during the process of feature selection the redundant and null vales are removed subset of features is selected.

On the other hand, feature extraction is a technique to extract and derive the relevant features from large datasets having multiple features or attributes. Feature extraction is directly related to dimensionality reduction. Depending upon the type of data sets specific algorithms such as SVD or PCA is applied over the data and principal values are extracted. The extracted feature are in the reduced form which play a notable role in reducing the dimensionality of the data.

1.5 Applications of Big Data in SG Systems

Using the big data generated from SG devices, useful data insights and information related to grid behavior can be drawn. It contributes to various applications and processes in SG such as load balancing, DRM, decision making process for end users, etc.

1. **Demand Response Management (DRM)**: With the large number of geo-distributed smart devices deployed globally in SG, it is a challenging task to maintain a balance between demand and supply with respect to the power generation and distribution. A recent report has highlighted an annual growth of 2.5% in the demand and consumption of electricity over the last 20 years in the US only [5]. Using demand response management, the energy demand and supply can be coordinated well at real time. It may lead to reduced peak demands, which would help lowering electricity cost.

2. **Load Balancing** : SG helps to balance the load fluctuations. Load adjustment or load balancing is one of the great features that SG has to offer to the traditional power grid. For examples, using various SG technologies such as peak shaving, peak loads can be balanced. Using various other sources such as electric vehicles, and solar panels energy can be also be put back to the grid and can be used during peak hours or when required.

3. **Giving Consumers the Control**: Apart from having the advanced technologies integrated with it, SG is also about giving the consumer control of the information that he needs in order to make decisions and choices about the energy use. It is similar to the process of managing our bank transactions online and many other activities. SG would incorporate an unmatched level of participation by user end. For an instance, consumers will have to wait no longer to get to know about what number of units of electricity they used or for monthly statement to know how much power they have used. With SG, everything would become clear in the picture. There are various tools in SG such as smart meters that allow to see the electricity credentials such as quantity and related cost price. SG combines the grid with real time pricing which allows the costumers to save the power and thus money.

1.6 Software Defined Networks (SDN)

Traditional computer networks are built from the large number of devices including switches, middle boxes, and routers having various protocols implemented on them. For such networks, network administrators manually operate the network operations which leads to a very complex and time consuming process to handle. It makes the system more error prone and lesser efficient in the terms of performance. So, to cope up with the modern inter net and data processing demands which strive for faster and efficient services "programmable networks" are making their ways. Software Defined Networks (SDN) is one of such networking approaches which incorporates the programming into its control system. SDN manages and control the computer operations in a dynamic and programmatic way. SDN had disjoint data plane and control plane in a vertical manner. The hardware plane which mainly involves switches and hardware devices is disjoint from the control plane. Control plane mainly works to control the network decisions. All the logics and programs are enabled at the control plane. The forwarding nodes in the data plane simply forward the packets or traffic which is controlled centrally by control plane with the help of program logics present at the control plane. SDN breaks the vertical architecture of traditional networks into a horizontal decoupled planes connected at the center via controllers. The detailed comparison between the traditional network and SDN is given in the table(1.1) given below. In the domain of ICT, SDN has become one of the most popular issue. Figure 1.4 shows the difference between traditional networks and SDN.

1.6.1 Advantages of SDN over Traditional Networks

Table 1.1: Comparison of SDN and traditional networks (SDN) [1]

Sr. no	Feature	Traditional Networks	Software Defined Networks
1	Configuration	Static with distributed control and middle boxes.	Logical centralized network control having decoupled planes with programmatic capabilities.
2	Performance	Error prone due to manual network management.	Envisions to be a resilient, flexible, and secure network with efficient resource utilization and energy efficiency.

In SDN, network administrators can control the network operations using programming at run time via open interfaces. For the purpose of communication between network switches and network plane open flow protocol is used. Open flow has been regarded as one of the first standardized protocol to enable the SDN. The architecture of SDN varies from the traditional static network architecture in various manners. The communication between control plane and the forwarding devices of data plane takes place with the help of open flow protocol. Open flow protocol maintains a flow table which further consists of rules, actions, and statistics.

Following are the main properties of SDN architecture:

1. SDN has an automated dynamic configuration with centralized control at the control plane.

2. The data plane and control plane are decoupled form each other in a SDN architecture. Data plane consists of switching or forwarding devices. These devices are being controlled by controllers using programming logics located at the control plane.

3. The application layer of the SDN consists of various applications such as traffic engineering, security, mobility, and many more. Different plane interact with each other with some communication protocol. The standardized protocol for SDN is open flow protocol which is explained further also.

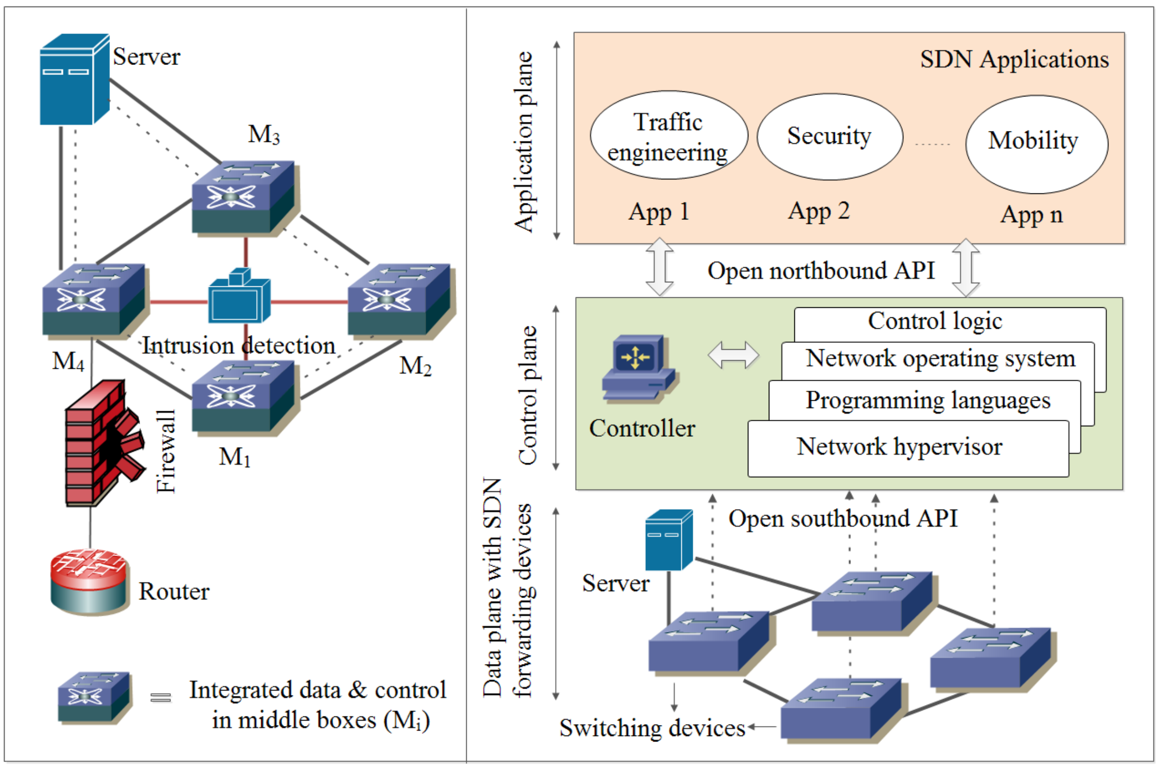


Figure 1.4: Comparison of traditional networks and SDN

As described above, handling big data today has become a daunting task for example in the case of data centers. In the context of data processing, it requires extensive parallel processing on number of servers having connected directly to each other. The network too needs to be more capable in the terms of capacity and efficiency. So, SDN has been emerging as a good candidate in revolutionizing the modern network systems and keeping the connectivity upto the mark without getting broke.

SDN and big data are interrelated to each other and attracts a lot of limelight research and industry. SDN can be used to solve the big data issues such as data storage, and data processing in data centers. On the other hand, big data too plays its role in benefiting various domains of SDN such as traffic engineering, processing of big data in data centers at cloud, scheduling of data in hadoop, and many other. Reconfiguration and programming of networks at run time is one of the main good features of SDN.

1.7 Role of SDN in SG Systems

The future grid is growing into a complex system. It involves large number of smart devices, sensors, distributed energy resources (DERs), energy transmission and distribution facilities, PMUs, AMIs, end-users. So to cope up with the communication and networking challenges using traditional networks has become quite challenging. SG needs more reliable, secure, and auto-configured networks to meet the requirements of intelligent devices and advanced infrastructure incorporated with it. In this context, programmable networks such as SDN comes into the picture. SDN would play a vital role in SG systems to increase the efficiency and quality of service as described below:

1. **Enhanced Communication and Data Exchange**: Current power systems use SCADA systems to deal with the information measurement and monitoring purposes in grid. But, due to the high demands of reliability and scalability, SCADA falls behind. However, in SG systems using PMUs and AMIs balance between supply and demand can be maintained and situational awareness can also be achieved. But, using PMUs and AMIs large amount of data gets generated at regular intervals which become every difficult to handle as described above also in the big data section. Using SDN, system can become more flexible and can offer lower latency rates with various communication modes such as multicast and unicast.

2. **Dynamic Configuration and Management**: Due to the presence of diverse number of smart devices in SG systems, dealing with multiple nodes to forward the data packets becomes quite cumbersome. However, by using forwarding rules of SDN over switching devices this process can be managed dynamically and with greater ease. Routing paths can be reconfigured and rerouted accordingly using control logics under control plane, which is a very complicated process in traditional networks due to manual management of networks.

3. **Fast Recovery for Smart Grid Communications**: An error-prone system is the ultimate vision for modern grid systems. So, error recovery and faster fault detection can be one the goals of SG. For example, using decoupled planes of SDN, routing paths can rerouted according to the shortest path available and the fault detection.

1.8 Organization of Thesis

The remaining part of this thesis is structured as follows. Chapter 2 represents the literature survey. It involves work related to existing big data techniques in SG systems, detailed survey on the role of SDN and big data in SG systems. Chapter 3 represents the problem formulation of the work. Then in chapter 4, this thesis structures the proposed scheme. Chapter 5 talks about the performance evaluation of the proposed scheme. Finally, chapter 6 concludes the work with the future scope of the work.

Chapter 2

Literature Survey

Advanced metering infrastructure (AMI) and Phasor measurement units (PMU) are the two main infrastructure units used for acquiring the data generated from different smart devices in SG systems and then pass the collected data to the utility which takes decisions about energy flow. AMIs are the bi-directional units which contain sensing devices, smart meters, control and monitoring systems, and data management units. On the other hand, PMUs are the energy measurement units generally used to measure energy waves and signals [2]. These devices are deployed in different smart communities in large number across the globe for effective power management generated from various distributed energy resources in SG environment. For example, approximately 50,000 smart meters are deployed by Department of Energy (DOE) in United States and Los Angeles department of water and power (LADWP) in the Los Angeles itself. LADWP serves approximately 4.1 million consumers which accounts to nearly 1% of total US power consumption [3]. Another report by energy information administration (EIA), highlights the need of an efficient data management with higher penetration of renewable energy sources (RES) to manage demand and supply optimally [4]. For this purpose, a reliable communication infrastructure is required to manage the flow of data and information between sources of data generation and smart meters.

The basic electric grid contains various substations, a network of transmission lines. ”Smart grid” having advanced infrastructure and information technologies equipped with the sensing technologies. It aims to automate and manage the increasing complexity and needs of electricity in the 21st Century. Its a two way communications between the grid and its end users. End-users basically include grid and the customers on the other hand. The smart grid infrastructure consists of controls, computers, automation, and various new technologies.

2.1 Challenges of Big Data in SG systems

Handling the large amount of data generated by smart devices in different time intervals using AMIs and PMUs is one of the biggest challenges in SG systems. The enormous amount of data acquired at the SG level often leads to the problems related to QoS provisioning and demand response management [2]. According to the various surveys, the data generated by smart grid qualifies for the main characteristics of big data, including volume, velocity, variety, veracity, and value. The data generated is in massive amount in the petabytes and terabytes. So it is very difficult to process this data. In this context various techniques to process SG data in literature survey has been described by various authors.

The data is generated at regular intervals depending upon the deployment of smart devices across different geographical regions. With the advent of smart homes equipped with many smart devices, the frequency of data generation has increased many folds which in turn poses many challenges. The main challenges include data representation, data storage, and data processing at SG level. As most of the smart devices generate data with high sampling rate so handling the volume and velocity of the data needs to be done in such a manner so that efficient decisions with respect to demand response can be taken on time [6]. Moreover, analyzing the SG big data may play a vital role in an intelligent power distribution such as prediction of power patterns, peak loads, demand response, fault tolerance, and RES management.

2.2 Existing Big Data Solutions in SG systems

However, the major issue with big data handling is its complexity due to the presence of multidimensional and heterogeneous attributes. Hence, the conversion of SG big data into a simplified structure is required for faster processing. In this direction, Souza *et al.* [12] presented a data compression methodology in smart distribution systems based upon the singular value decomposition (SVD) technique. The methodology presented by authors is based upon the lossy data compression method. There are basically two types of compressions to reduce the dimensionality. First one is lossy and another is lossless. In lossless compression, no amount of original data gets lost after the reduction. While, in lossy compression, after applying some DR algorithm new value of data is bound to lose its originality by a certain value. Therefore, in order to reconstruct the original data without losing valuable information, a trade-off between compression ratio and the reduction rate needs to be maintained. For this purpose, authors in [13] presented a data compression technique

Table 2.1: Existing big data techniques in SG systems

sr no.	Name	Features	Contributions
1	Random Matrix Theory (RMT) [7]	Aims to apply big data technology into smart grids, An useful framework in concern with the multivariate data.	Easy in logic and fast in speed.
2	Temporal and spatial big data computing framework [8]	Every dataset is divided into subgroups, each of which having data items shared by different tasks.	An efficient computing approach with greater convergence ratio and improved bandwidth.
3	Data compression techniques	A lossy data compression scheme in smart distribution systems using SVD.	It is Capable of reducing the data size significantly and accurately reconstructs the original data.
4	Cloud based framework [9]	A "Smart-Frame" cloud based framework for big data information management in SG systems.	Flexibility and scalability.
5	Wavelet-based data compression technique in SG environment	It compresses the size of disturbance and depresses the noise signals in data.	The scheme is implemented in SG to mitigate data congestion Without generality loss. It improves the data quality and transmission process.
6	Data decomposition techniques	Multi linear SVD and MPCA are lossy decomposition techniques which have been used to reduce the dimensionality in SG big data.	The reduction ratio is better, but has to make a trade off between reduction ration and reduction ratio.
7	Load disaggregation algorithm [10]	Decomposes smart meter power readings into a set of discrete pulses is based on low frequency sampling rates.	The proposed solution is cost efficient.
8	A Big data scale algorithm for microgrids [11]	A Big Data Scale Algorithm for of Integrated Microgrids.	Scheduling is optimal.

using SVD in smart distribution systems. Moreover, Ning *et al.* [14] proposed a data compression technique based upon the wavelet function. This technique compresses the size of noise signals, which in turn affects data transmission. However, in order to represent heterogeneous big data with reduced dimensions, tensor representation is one of the emerging techniques [15]. For example, Kuang *et al.* [16] proposed a tensor-based unified model for big data representation and size reduction. Yang *et al.* [17] introduced a similar technique named as lanczos-based high order SVD algorithm to reduce the dimensionality of unified data tensor model.

After effective storage and representation of big data, another major task is to process and transmit the reduced data over the underlying SG network efficiently with low latency. In this context, authors in [18, 19] reviewed the key aspects of smart metering process with a focus on the type of data generated and techniques required to process it. Authors highlighted that the data processing at network level is a major challenge faced by SG systems. This is due to the generation of data at regular intervals from various smart devices leading to the traffic congestion at the network infrastructure. Similarly, Plaza *et al.* [20] presented the possibility of reception and information broadcasting between smart meters and the grid, in real-time through the cellular network using AMI. Hence, after analyzing the aforementioned proposals, it is evident that efficient data flow over the existing network infrastructure is required for handling the big data generated from various smart devices in SG systems [21].

2.3 Survey on SDN Protocol and Applications

To mitigate the challenges posed by SG network, Software-defined network (SDN) has emerged as a powerful and flexible platform for efficient traffic flow for SG systems. Nunes *et al.* in [22] surveyed the history, present and future of SDN with layered taxonomy of SDN architecture. Similarly, authors in [1], [23], and [24] presented various features of SDN such as network capabilities, interfaces, and programming languages used. They highlighted the essence of SDN which lies in its decoupled planes and the programmatic behavior of SDN. The authors also surveyed various research trends in the SDN domain which have been becoming the subject of attraction both in the field of academia and research. They presented the literature with various definitions, challenges, benefits, and reference models of SDN. They highlighted the need of SDN in the modern era of ICT. ICT applications such as multimedia, big data, and cloud etc., demand more bandwidth, and dynamic management of network.

Authors in [25] discussed the key building blocks of an SDN infrastructure, southbound and northbound APIs, network virtualization layers, and NOS. Apart from this, network programming languages involved and network applications provided by SDN have also been highlighted. Authors in [26] and [27] elaborated the deployment and use of the existent communication protocol of SDN called OpenFlow. They conducted an extensive survey on the OpenFlow concept, its implementation, applications, programming languages, controller, debugging, troubleshooting, QoS, and its integration with wireless and optical networks. This survey can help both industry and R&D people to understand the latest trends of SDN/OpenFlow designs. They also brought various challenges related to privacy and security into light which mainly arise due to the process of virtualization.

Talking about the NOS and centrally located controllers, authors in [28], [29], and [30] highlighted the deployment of SDN controllers in wireless sensor networks. Kim *et al* [31] highlighted various benefits of using SDN explicitly in different environments. Due to the logic and flexibility involved, it becomes very easy to reconfigure the network changes dynamically.

In the context of SDN in SG systems as a networking paradigm, although there has not been much published work. But, authors in [32] presented a SDN based communication architecture for micro grids in order to enhance the flexibility and resilience in the system. Micro grids are basically smaller units of electricity grids which are connected centrally, but work independently.

2.4 Benefits of SDN and Big Data in SG Systems

Similarly, authors in [33] and [34] introduced the need for a scalable network in the form of SDN. The complex SG communications and computing needs require an easy to manageable and flexible network which envisions to deliver better quality of services to end users. Likewise, Cahn *et al.* [35] explored the benefits of integrating SDN with SG systems. The authors utilized SDN to design a self-managed substation network for SG systems. In an another similar work, Aujla *et al.* [36] utilized SDN for energy management of sustainable DCs. Likewise, in [37], [38], and [39] various authors have described the importance of SDN in SG systems.

Moreover, Big data and SDN [40] are interrelated to each other and utilize the properties of each other for mutual benefits. SDN benefits the big data applications by improving the performance of the network in the form of latency and throughput. In this regard, Kuang *et al.* [41] proposed a tensor-based big data approach in SDN for effective QoS provisioning.

For this purpose, a common platform in the form of SDN model is required to integrate the SG systems with SDN for big data management and processing.

Chapter 3

Problem Formulation

3.1 Research Motivation

After analyzing the aforementioned proposals, it is inferred that a huge amount of data is generated by various smart devices in the SG. However, handling of big data in an efficient manner is one of the biggest challenges in SG environment. An efficient handling and processing of big data at SG systems may lead to an effective big data analytics, which in turn can provide a better demand response management, energy consumption prediction, situational awareness, and effective communication among devices and users. Various techniques have been analyzed with respect to these issues in the existing proposals [18] - [19]. But, none of the existing proposals has focused on big data analytics in SG for an efficient QoS provisioning. Moreover, the existing proposals have not explored any unified model for data representation. In this regard, tensors have been effectively used for representation and management of big data. Tensors are basically a multidimensional arrays which can be used to represent the high dimensional data. Also, the big data represented as tensors could be reduced to a simpler form by removing the redundant and ambiguous dimensions. The dimensionality reduction leads to an efficient storage, processing and flow of big data among smart devices in SG. Hence, there is a need of an unified and intelligent SDN model using tensor representation for dimensionality reduction of big data in SG environment.

3.1.1 Motivation Examples

For better illustration of the problem formulation, let us consider an example as shown in Fig. 3.1. If a data frame of 20 Mbits is transmitted over traditional network channel, then it takes 1 second to reach destination node at a data rate of 20 Mbps (refer Fig. 3.1 (a)). However, if the data is reduced (16 Mbits) in size omitting all invaluable and unwanted

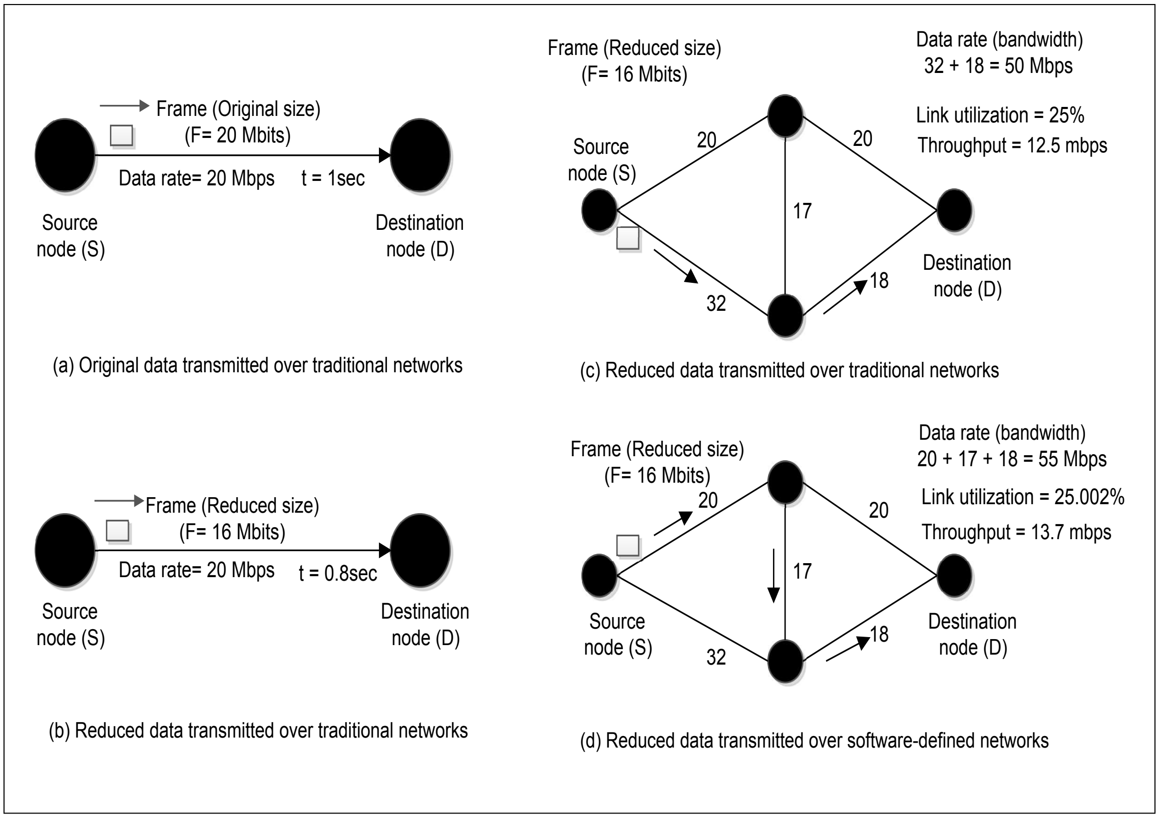


Figure 3.1: Motivation examples

values, then it takes 0.8 seconds to reach to the destination at the same data rate (refer Fig. 3.1(b)). Hence, it is quite evident, that the size of data has a strong impact on the transmission time. Since, a huge amount of data is transmitted seamlessly in SG systems so if such data is reduced then it may be beneficial for the overall performance of the network.

Now, when the underlying networks follow a dynamic network management scheme, then it can help to achieve better utilization of network resources and thereby can achieve enhanced throughput. For example, if a reduced data frame is transmitted over traditional networks, they may choose the shortest path using traditional network protocols. In such a case, a data rate of 50 Mbps is achieved along with a link utilization of 25% and throughput of 12.5 Mbps (refer Fig. 3.1(c)). However, if dynamic networks such as SDN are deployed, then a data rate of 55 Mbps is achieved along with a better link utilization of 25.2% and an enhanced throughput of 13.7 Mbps (refer Fig. 3.1(d)). Hence, it clearly shows that a better throughput, data rate, and link utilization can be achieved by deployment of SDN based network infrastructure for data management in SG systems. The better utilization of link may also help in reducing the energy consumption of network infrastructure.

3.2 Problem Statement

After analyzing the aforementioned proposals, it is inferred that a huge amount of big data is being generated by various smart devices in the SG. However, handling this big data in an efficient manner is one of the biggest challenges in SG environment. An efficient data handling and processing of big data at SG systems can lead to a better demand response management, energy consumption prediction, and effective communication among various devices. Various techniques have been analyzed with respect to these issues in the existing proposals [18] - [19]. But, none of the existing proposals have focused on big data analytics in SG systems for an efficient QoS provisioning. Moreover, the existing proposals have not explored any unified model for data representation. In this context, tensors have been effectively used for representation and management of big data using SDN [16] - [17]. Also, the big data represented by tensors could be reduced to a simpler form by removing the redundant and ambiguous dimensions. Moreover, to ease the burden of data flow on the existing network infrastructure, SDN can play an integral role in processing and forwarding the reduced data in an efficient manner over the SG network infrastructure [1]- [31], [40]- [41]. Hence, there is a need of an unified and intelligent SDN model for big data management in SG systems. Therefore, a tensor-based SDN model for efficient big data management in SG systems has been designed in the proposal.

3.3 Problem Formulation

In this paper, three different types of datasets D_ϕ, D_ψ, and D_ω are used for unstructured, semi-structured and structured data, respectively. These datasets are acquired from various smart devices incorporated in SG systems. The acquired data is represented in the tensor form. Tensors are multi-way arrays which are used to represent the data having multiple characteristics and high dimensions. A tensor T of n-order is represented as follows.

$$T \in R^{a_1 \times a_2 \times a_3 \ldots \times a_n} \tag{3.1}$$

where, a_1, a_2,.., a_n are the orders of tensor which define the dimensionality of data characteristics.

In order to represent big data as tensors, data with n number of characteristics is represented as a cross product of various characteristics having multiple dimensions. The representation of big data in tensor form is given as follows.

$$E[x_1 \otimes x_2 \otimes x_3 \otimes ..x_n] = R^{a_1 \times a_2 \times a_3 .. \times a_n} \tag{3.2}$$

Here, $x_1, x_2, ..., x_n$ represent different attributes present in big data (for example, voltage, energy consumption, meter/customer ID, and load can be described as various attributes of SG big data data generated by smart devices).

Each attribute of big data is independent of the other and can be represented as a cross product of each other. Hence, using Eq. (3.2), the acquired heterogeneous big datasets are converted into their respective tensors as given below.

$$D_\phi \rightarrow T_\phi, D_\psi \rightarrow T_\psi, D_\omega \rightarrow T_\omega \tag{3.3}$$

where, T_ϕ, T_ψ, and T_ω denotes sub-tensors for unstructured, semi-structured, and structured data, respectively.

In order to reduce the data redundancy and duplicacy, sub-tensors are converted into an unified tensor (T_υ) using a *unified data tensorization operation* as a function given below [17].

$$f\colon (D_\phi \cup D_\psi \cup D_\omega) \rightarrow T_\phi \cup T_\psi \cup T_\omega \tag{3.4}$$

$$f(x,y,z) = u \tag{3.5}$$

where x $\in T_\phi, y \in T_\psi, z \in T_\omega$, and u $\in T_{uni}$

The union operator combines the similar characteristics and reduces the redundancy from the acquired big data. However, with the presence of higher dimensionality, the complexity of big data remains high which leads to data inconsistency and data processing problems in big data analytics. To overcome such problems, the unified tensor needs to be transformed into a lower-order tensor having fewer dimensions which can be represented as the reduced tensor. The transformation of a n^{th} order tensor into n number of matrices is known as tensor unfolding or matricization [16]. For a given tensor, $T \in R^{a_1 \times a_2 \times a_3 ... \times a_n}$, the equation of unfolding n-order matrix into a mode-i matrix is given as below.

$$T \in R^{a_i \times (a_1 \times a_2 \times a_3 ... \times a_{i-1} \times a_{i+1} .... \times a_n)} \tag{3.6}$$

The number of rows and columns of each mode-i matrix are given by Eq. (3.7) and Eq. (3.8), respectively.

$$a_i, 1 \leq i \leq n \tag{3.7}$$

$$\prod_{j=1}^{n} a_j, i \neq j \tag{3.8}$$

Now, SVD is used to factorize a real or a complex matrix. An unfolded mode-i matrix (M_i) which is to be decomposed using SVD, can be represented as given below.

$$M_i = U_i S_i V_i^* \tag{3.9}$$

where, U and V are unitary matrices and are orthogonal to each other, S is a diagonal matrix, V^* is the conjugate transpose of the unitary matrix V.

The diagonal matrix S, with non-negative entries has i singular values denoted by σ_i. The singular values (σ_i) in the diagonal matrix S are as given below.

$$S = \begin{bmatrix} \sigma_1 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \sigma_i & 0 \end{bmatrix}. \tag{3.10}$$

$$S = diag(\sigma_1, \sigma_2, ..., \sigma_i, 0, ..., 0) \tag{3.11}$$

$$where;\ \sigma_1 \geq \sigma_2 \geq ...... \geq \sigma_i > 0 \tag{3.12}$$

After applying SVD on each mode-i matrix, the rank of a singular matrix is approximated to a lower rank r ($r \leq n$). The $(r-k)$ values obtained from the singular matrix are truncated and a low-rank approximation is achieved by applying the SVD incrementally. Rank is approximated based upon the threshold value of 30 percent rank reduction depending upon the target dataset. Then, the reduced tensor is obtained by projecting the orthogonal vectors obtained from the results of truncated SVD, over the initial tensor (T_{uni}). The dimensionality reduction is achieved by obtaining reduced tensor which contains the lesser dimensions, but valuable and core information as present in the initial tensor. The dimensionality of the n^{th} tensor is reduced using n-mode product operation. The n-mode product operation of a tensor (T_{uni}) by a matrix (U) is defined as follows.

$$(T_{uni} \times_n U)_{e_1 e_2 .... e_{k-1} e_k e_{k+1} .... e_n} \tag{3.13}$$

where, $e_1 e_2 .... e_{k-1} e_k e_{k+1} .... e_n$ are the dimensional attributes of n^{th} order tensor.

In order to calculate the reduced core tensor (T_{red}), n-mode product is used for a n^{th} order tensor as shown below.

$$T_{red} = T_{uni} \times_{x=1}^{n} U_n^T \tag{3.14}$$

$$T_{red} = T \times_1 U_2^T \times_2 U_3^T \ldots\ldots \times_n U_n^T \tag{3.15}$$

Moreover, from this reduced tensor, an approximated tensor ($\hat{T}_{red}$) can be reconstructed as illustrated below.

$$\hat{T}_{red} = T_{red} \times_1 U_1 \times_2 U_2 \times_3 U_3 \ldots\ldots\ldots \times_n U_n \tag{3.16}$$

The approximation can be further optimized using Frobenius-norm on the tensor values obtained after tensor product by a matrix. The Frobenius norm is one of the important matrix norms which finds the size of a multidimensional array M, by taking the square root of the sum of the squares of its elements as given below.

$$||M||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{m} (m_{ij})^2} \tag{3.17}$$

Frobenius norm on M having two dimensions m, n dimensions, is defined as below.

$$||M||_F = \sqrt{\sum_{i=1}^{min(m,n)} (\sigma_i)^2} \tag{3.18}$$

The reconstruction error defines the approximation accuracy of the reduced tensor. It occurs due to the approximation of mode-i matrices. The reconstruction error for unified tensor and approximated reduced tensor is given as below.

$$e = ||T_{uni} - \hat{T}_{red}||_F \tag{3.19}$$

With an increase in reconstruction error ratio, the accuracy of the core data or reduced tensor decreases. The reconstruction error ratio, ρ can be analyzed using Frobenius-norm of original unified tensor and final reduced tensor and is defined as below.

$$\rho = \left(\frac{||T_{uni} - \hat{T}_{red}||_F}{||T_{uni}||_F} \right) \tag{3.20}$$

Hence, the main objective function of the proposed scheme is to minimize the reconstruction error and is defined as below.

$$min\,(\rho) \tag{3.21}$$

$$s.t. \tag{3.22}$$

$$\rho \in [0,1] \tag{3.23}$$

$$T_{uni} > \hat{T}_{red} \tag{3.24}$$

$$T_{uni}, \hat{T}_{red} > 0 \tag{3.25}$$

subject to

$$T_{\upsilon} > T_{\Upsilon} \tag{3.26}$$

$$M_i > \hat{M}_i \tag{3.27}$$

$$S_{\Upsilon} < S \tag{3.28}$$

Now, the difference between T_{υ} and $\hat{T}$ is the error (e_{apx}) between the initial unified tensor and approximated tensor which needs to be reduced.

Chapter 4

Proposed Scheme

There are two main objectives or schemes involved in the proposed work of this thesis. First section of this chapter involves the first scheme which is called as the big data representation scheme. Then, in the next section SDN based data processing scheme is presented. Further, a mathematical case study is formulated to explain the working of the proposed tensor-based data representation scheme.

4.1 Big Data Representation

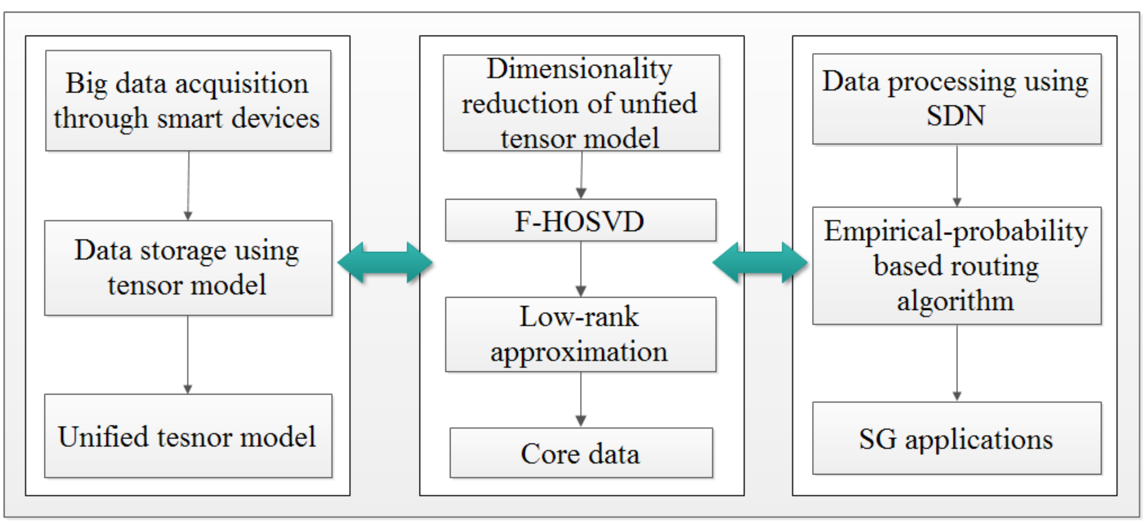


Figure 4.1: Workflow of the proposed scheme

In this section, a tensor-based data management scheme is presented to acquire raw data and reduce it to lower dimensionality thereby optimizing the reconstruction error. It involves various steps from data acquisition to the decomposition of data and finally the reduction of data. Further, reconstruction error is also minimized by keeping a trade off between reconstruction error and reduction ratio. The balance between these two ratios define the accuracy of the scheme. In this regard, algorithm 1 is designed and described as below. Figure 4.1 shows the work flow of the proposed scheme.

The acquired data (D_{ac}) is sorted in structured, semi-structured, and unstructured data

Algorithm 1 Tensor-based data management algorithm

Input: D_{ac}, Acquired raw data
Output: T_{uni}, T_{red}

1: Acquire D_{ac} from various sources
2: Sort into D_{ϕ}, D_{ψ}, and D_{ω}
3: $D_{\phi} \rightarrow T_{\phi}$, $D_{\psi} \rightarrow T_{\psi}$, $D_{\omega} \rightarrow T_{\omega}$
4: Unify T_{ϕ}, T_{ψ} and T_{ω} using Eq. (3.4)
5: Unfold T_{uni} into n matrices using Eq. (3.6)
6: **for** (i=1; i $\leq$ n; i++) **do**
7:     Apply $SVD(M_i)$
8:     Obtain U_i, S_i and V_i^*
9:     Extract singular values σ_i from the diagonal matrix S_i
10:     Calculate rank(S_i) = n
11:     **while do**$(1 < r \leq n)$
12:         Obtain ($\hat{S}$) by pruning the smallest σ_i
13:         Obtain rank($\hat{S}$)
14:         Reconstruct ($\hat{M}_i$) using Eq. (3.9).
15:         Extract new $\hat{U}_i, \hat{S}_i, and \hat{V}_i^*$
16:     **end while**
17:     Store the left truncated orthonormal vectors, $\hat{U}_i$.
18:     Perform n-mode product of $\hat{U}_i$ with initial tensor T_{uni}.
19:     Calculate T_{red} using Eq.(3.15).
20: **end for**
21: Reconstruct the approximated tensor $\hat{T}_{red}$ using Eq.(3.26).
22: Apply Frobenius-norm on $\hat{T}_{red}$ using Eq.(3.17).
23: Compute reconstruction error (e), using Eq. (3.19)
24: Obtain reconstruction error ratio (ρ), using Eq. (3.20)
25: **if** $(\rho < \rho_{th})$ **then**
26:     Send T_{red} to destination.
27: **else**
28:     Recalculate T_{red} to satisfy ρ.
29: **end if**

(line 1-2). After sorting the data, each type of data is converted into its corresponding sub-tensors (T_{ϕ}, T_{ψ}, and T_{ω}) using Eq. (3.2) (line 3). Then, all the sub-tensors are combined together to form a unified tensor (T_{uni}) using Eq. (3.4). Now, the unified tensor (T_{uni}) is unfolded into n matrices using Eq. (3.6) (line 5). Then, all the unfolded matrices are decomposed using SVD. The matrices are decomposed into a combination of unitary matrix U, conjugate transpose of unitary matrix V (V^*), and a diagonal matrix S (line 6-9). Now, lowest rank approximation is applied to keep r largest singular values and replacing other values by zero (line 10-13). Then, $\hat{M}_i$ which is used to obtain the approximated decomposed values ($U_i, \hat{S}, V_i^*$) is calculated using Eq. (3.9) (line 14-17). The n-mode product is applied to the left orthonormal column vectors with the initial tensor to obtain the reduced tensor, T_{red} using Eq. (3.15) (line 18-20). After this, $\hat{T}_{red}$ is calculated using Eq. (3.26). The Frobenius norm for minimizing the difference between original and approximated reduced tensor is applied on the reconstruction error ratio (ρ) to optimize the result. If (ρ) is less than the threshold value of error ratio (ρ_{th}), then the reduced tensor (T_{red}) is sent to the destination. Otherwise, repeat the process till it satisfies the acceptable error ratio (line 21-29).

4.2 SDN-based Control Scheme

In this section, an emerging software-centric networking paradigm called SDN is used in the proposed scheme to provide dynamic network management in SG systems. SDN is an open and programmable platform which controls the network in an intelligent and dynamic way through well-decoupled planes. It provides abstraction of underlying infrastructure from network applications, which makes it easy to manage and reconfigure according to the dynamic changes int the network configuration [35]. The growing rates of big data traffic at SG systems could be effectively handled using scalability and efficiency of SDN. Hence, the integration of big data technologies such as tensor models with SDN can led to an extensible and efficient service provisioning to the end users. In this context, a tensor-based SDN model is designed in the proposed scheme using three planes; (1) data plane, (2) control plane, and (3) application plane as shown in Figure. 4.2.

In this model, data plane mainly consists of network devices such as- switches and routers. Data is acquired from various devices such as- appliances in smart homes, and electric vehicles (EVs). This plane use open flow protocol (OFP) as a communication standard to forward the gathered data to the upper plane [35]. The acquired data is decomposed into a reduced tensor of smaller rank and size. After reduction, the core data is sent to the control plane. With the help of control algorithms, core data obtained from forwarding data plane is processed in an efficient manner. For this purpose, an empirical probability-based control scheme is designed to estimate the optimal route for transmission of reduced data over SG networks. Finally, the application plane provides various services to end users such as quality of service (QoS), mobility, security, and network virtualization.

4.2.1 Empirical Probability-based Control Scheme (EPCS)

As T_{red} needs to be transmitted over the underlined network through optimal paths. So, laying the foundation on traditional SDN routing algorithm an empirical-probability based scheme is proposed in this section to recommend routing paths with maximum likelihood for scheduling or reduced data. It would tend to maximize the channel utilization and minimize the latency. This scheme tends to shape the traffic coming from various smart devices at real time, with respect to the resource availability and given QoS constraints after the data reduction has been performed. SDN switches interact with controller which estimates the optimal routes. In order to keep the track of various network related updates like path priorities, a flow table is maintained with the help of programs or logic applied via controller at the control layer. A flow table consists of various fields such as match entry, instructions, and priority.

In this regard, a routing scheme for the SDN controller is presented in order to forward the reduced data through optimal paths with lower latencies and higher QoS. In this scheme, all the forwarding nodes (FNs) can be visualized as SDN FNs and non-SDN FNs. Data which passes through at least one SDN FN comes under controllable flow and which does not passes through any SDN FN, is considered as uncontrollable flow [42]. Now, consider a network *Z(N,L)*, having *N* number of nodes and *L* links, *c(l)* denotes the channel capacity and *f(l)* refers to the traffic flow on link *l*. Now the flow table already has old entries in it. Through our proposed logic, we tend to update the flow table to a new one (N_r) The older flow table entries have been taken as the observation set which are to be fed to the estimator at SDN controller. It further predicts or estimates the optimal path.

Empirical Probability is an estimation of the occurrence of an event, happening in an actual environment. We can estimate N_r using probabilistic approach. The empirical distribution function $\hat{\theta}$, can be given as.

$$\hat{\theta} = \frac{n_p}{o_p} \tag{4.1}$$

where, n_p represents new data packet forwarded to the controller, and o_p refers to the older observations in the flow table.

The optimality of the scheme can be checked using the *mean-square error (MSE)* of the estimator. MSE of the estimator $\hat{\theta}$, is defined as the function of the new routes to be

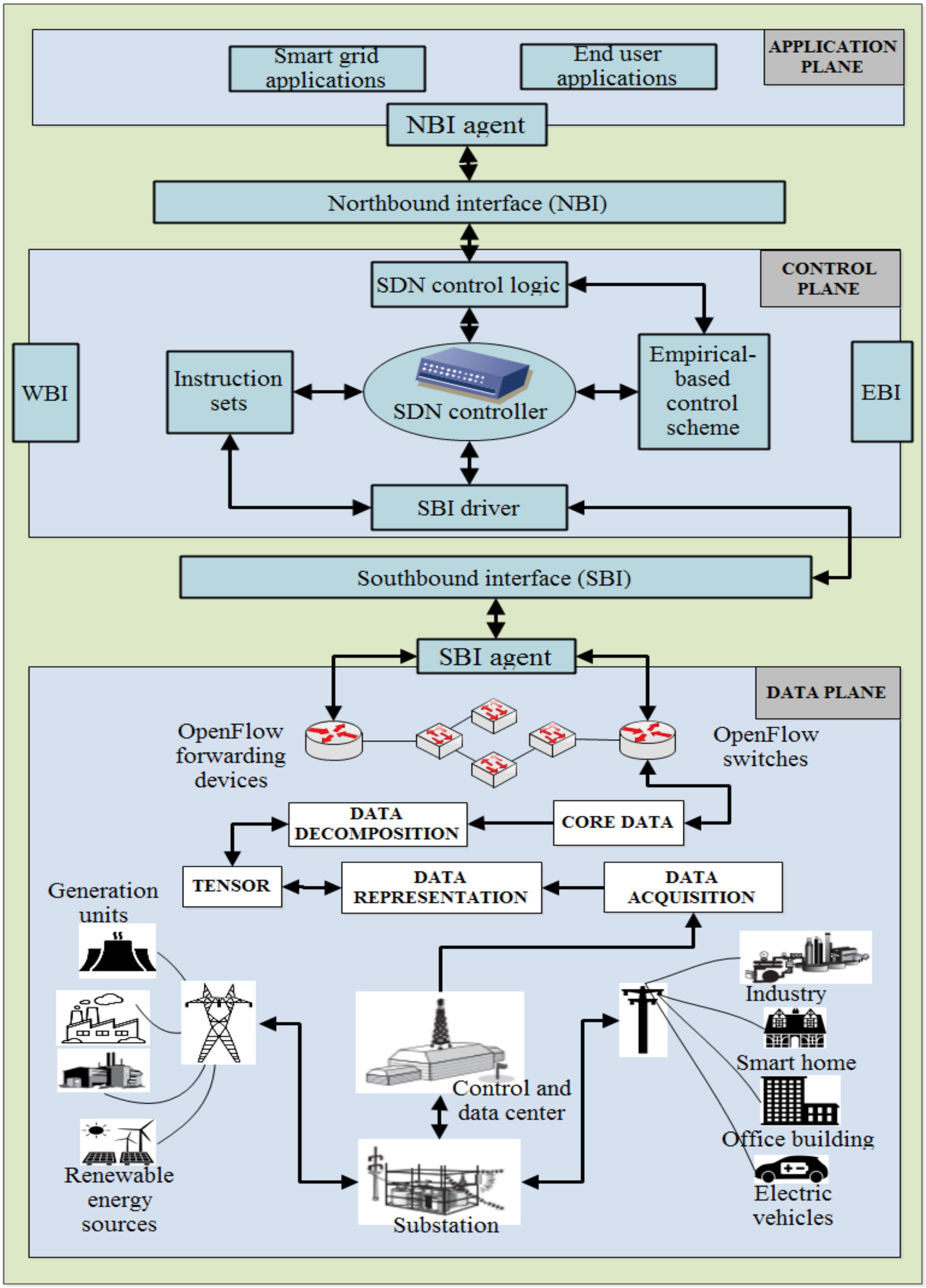


Figure 4.2: Tensor-based SDN model

predicted as shown below.

$$MSE(\hat{\theta}) = f(n_r) \tag{4.2}$$

$$E[(\theta) - (\hat{\theta})^2] = E[(\theta) - (g(n_r)^2)] \tag{4.3}$$

The algorithm for the proposed scheme is given as below.

Algorithm 2 Empirical probability-based control algorithm

Input: A topological network Z(N,L) at SDN.
Output: Maximize channel utilization, u.

1: Split the z(n,l) into SDN FNs and non-SDN FNs.
2: Forward data packets to the SDN controller.
3: For each destination d $\in$ N, apply OSPF on non-SDN FNs.
4: Obtain o_p.
5: While(SDN FNs $\subset$ N)
6:     Feed the o_p to $\hat{\theta}$.
7:     Using equation 4.1, obtain $\hat{\theta}$.
8:     Obtain N_p.
9:     Update flow table, n_p.
10: while MSE($\hat{\theta}$) $\leq threshold$
11: max u.

In this proposed algorithm, the network z(N,l) has been divided into type of forwarding nodes called SDN FNS and non-SDN FNs (line 1). Now, the data packets are forwarded to the controller (line 2). Further, using open shortest path first algorithm, entries are updated in the flow table and o_p is maintained (line3-4). After updating flow table, the empirical estimator is applied on SDN-FNs to estimate the new path for the data at the controller (line 5-8). The new flow table is then updated with new estimated values (line 9). The accuracy of the predicted routes is checked using MSE of the estimator. The MSE obtained is compared with the threshold value. If the value of MSE is less than or equal to the threshold, then the channel utilization is maximized (line 10-11).

4.3 Mathematical Case Study

The following section represents the exemplar case study for tensor-based data representation and dimensionality reduction. The high-dimensional big data is represented using tensors. Fig. 4.3, shows the visualization of a three-order tensor $R^{4\times5\times3}$ having 4, 5, and 3 instances at each order, respectively.

An n-order tensor can be unfolded into n different matrices through the process of ma-

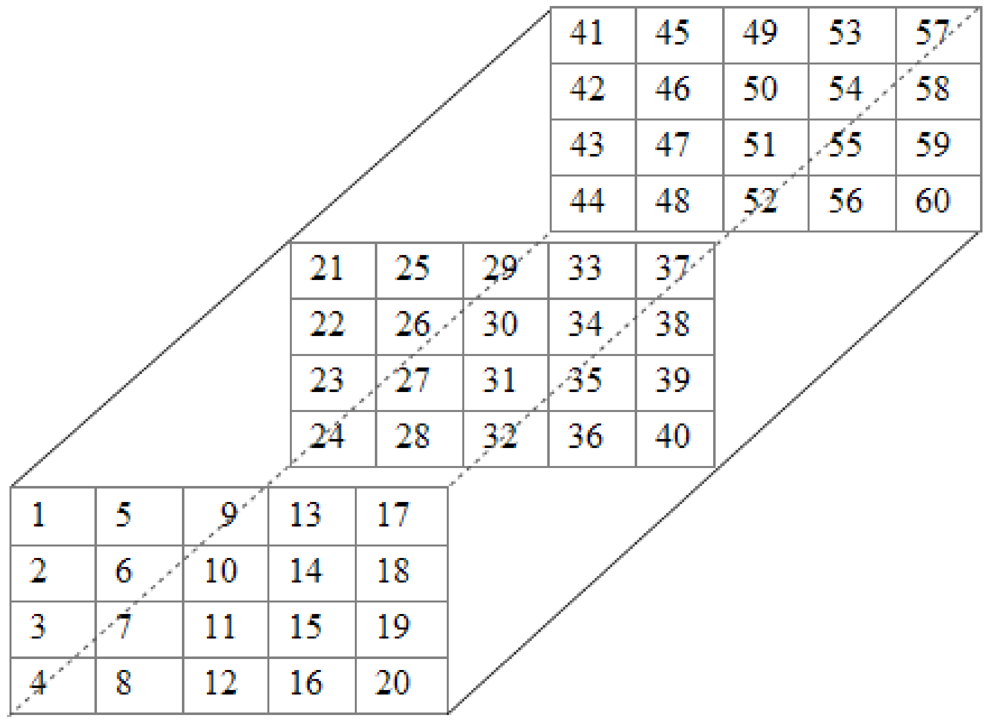


Figure 4.3: A visualization of a three-order tensor, $R^{2\times3\times4}$

tricization. The transformation of a high-order tensor into lower order matrices is known as tensor unfolding or matricization. For a given tensor $R^{4\times5\times3}$, the process of unfolding into matrices can be done using Eq. (3.6). Fig. 4.4 shows the tensor $R^{4\times5\times3}$ that has been unfolded into three different matrices named as *M1*, *M2*, and *M3*, respectively. The row and column number of each matrix is calculated using Eqs. (3.7) and (3.8), respectively. Now, M1 has been unfolded by taking first order as row number and the product of rest of the orders contribute to column number. For example, in M1, there are four number of rows and fifteen number of columns. In a similar manner, others matrices (M2 and M3) can be expanded.

After the matricization of a tensor into matrices, SVD is applied on on each matrix truncated in order to obtain singular values as shown in the Fig. 4.5. Each matrix gets decomposed further into three matrices, i.e., two orthonormal matrices ($U_i and V_i*$) and a diagonal matrix (a_i). The diagonal matrix contains singular values in descending order, i.e., $a_1 \geq a_2 \geq ....a_i \geq 0$.

After applying truncated SVD on each matrix (M_i), the null values in the diagonal

$$M1 = \begin{matrix} 1 & 5 & 9 & 13 & 17 & 21 & 25 & 29 & 33 & 37 & 41 & 45 & 49 & 53 & 57 \\ 2 & 6 & 10 & 14 & 18 & 22 & 26 & 30 & 34 & 38 & 42 & 46 & 50 & 54 & 58 \\ 3 & 7 & 11 & 15 & 19 & 23 & 27 & 31 & 35 & 39 & 43 & 47 & 51 & 55 & 59 \\ 4 & 8 & 12 & 16 & 20 & 24 & 28 & 32 & 36 & 40 & 44 & 48 & 52 & 56 & 60 \end{matrix}$$

$$M2 = \begin{matrix} 1 & 6 & 11 & 16 & 21 & 26 & 31 & 36 & 41 & 46 & 51 & 56 \\ 2 & 7 & 12 & 17 & 22 & 27 & 32 & 37 & 42 & 47 & 52 & 57 \\ 3 & 8 & 13 & 18 & 23 & 28 & 33 & 38 & 43 & 48 & 53 & 58 \\ 4 & 9 & 14 & 19 & 24 & 29 & 34 & 39 & 44 & 49 & 54 & 59 \\ 5 & 10 & 15 & 20 & 25 & 30 & 35 & 40 & 45 & 50 & 55 & 60 \end{matrix}$$

$$M3 = \begin{matrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 & 25 & 28 & 31 & 34 & 37 & 40 & 43 & 46 & 49 & 52 & 55 & 58 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 & 26 & 29 & 32 & 35 & 38 & 41 & 44 & 47 & 50 & 53 & 56 & 59 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 & 27 & 30 & 33 & 36 & 39 & 42 & 45 & 48 & 51 & 54 & 57 & 60 \end{matrix}$$

Figure 4.4: Unfolding of tensor $R^{2\times3\times4}$ into three matrices

matrix can be pruned and top largest values having rank r can be retained. Table 5.1 shows various mathematical results obtained for each matrix. Column II shows the number of singular values obtained and next three columns shows the top 3 largest singular values optimized after truncating the null values. By spanning the singular value space with orthonormal vectors of decomposed matrix, an approximate matrix with with rank (r) is obtained. After this, in order to calculate the reduced tensor T_{red}, n-mode product is used for n-order tensors as shown in the Eqs. (3.15) and (3.26). Figure 4.6 shows the orthonormal vectors decomposed after applying a specific decomposition technique onto a three-order tensor. The dimensionality of the given tensor is reduced by incrementally applying n-mode product on initial tenor with left orthonormal space. Now, the reduced tensor can be approximated by optimizing the error reduction ratio using Frobenius norm.

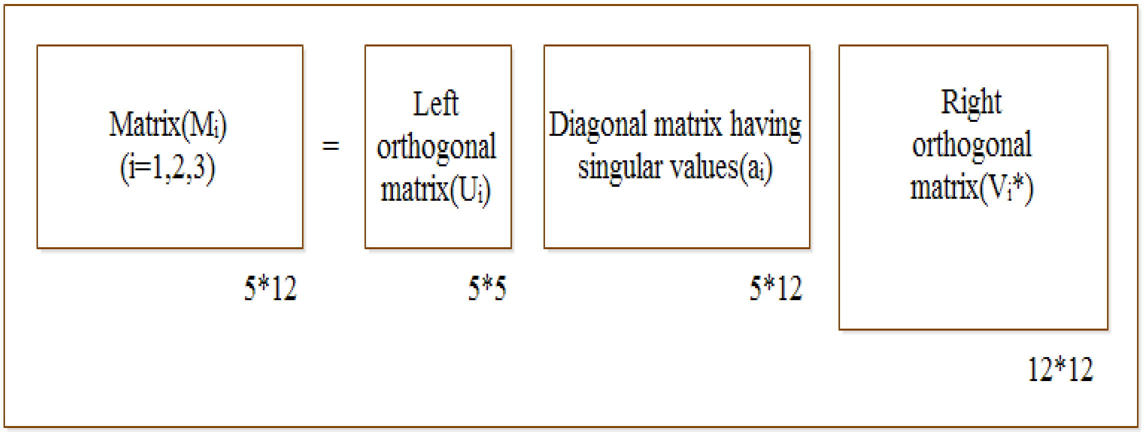


Figure 4.5: Representation of singular values decomposition

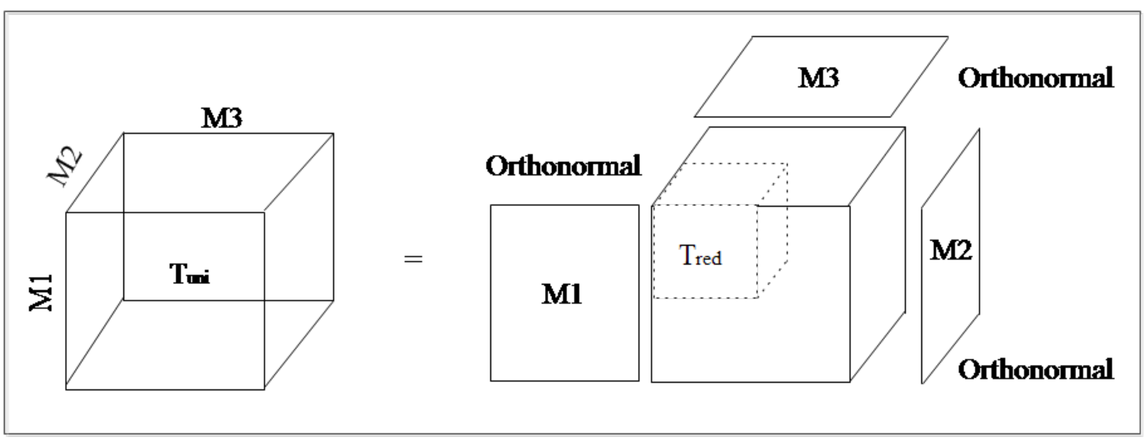


Figure 4.6: Representation of three-order tensor after decomposition

Chapter 5

Performance Evaluation

5.1 Implementation

The proposed scheme is widely implemented using matlab and R programming language. The first section of this chapter consists of the tool and techniques related to the simulation and implementation. Further, in the next section results and discussions related to the proposed work are shown involving different evaluation parameters.

5.1.1 Matlab

MATLAB with a full name of Matrix laboratory is a platform for numerical analysis that can work for multiple programming languages. It is capable of performing various functions such as data and matrix manipulations, plotting of data and functions, implementation of various algorithms using different languages, and to name a few. Matlab is a proprietary software. The main programming languages it involves are C, C#, C++, and python. Figure 5.1 shows the environment of matlab with a code snippet of matrix and tensor operations.

5.1.2 R and RStudio

R is a programming language-cum-software environment used for computational statistics and graphics. It is mainly used for statical analysis, data mining, machine learning, and graphical representation of data.

RStudio is a free and open-source IDE for R programming language. It is widely used for graphical representation and statistical computing of data sets. It was founded by JJ Allaire. RStudio is written in the C++ programming language and uses the Qt framework for its graphical user interface. RStudio is available both in open source and commercial versions. Figure 5.2 shows the IDE of RStudio along with a code snippet of R.

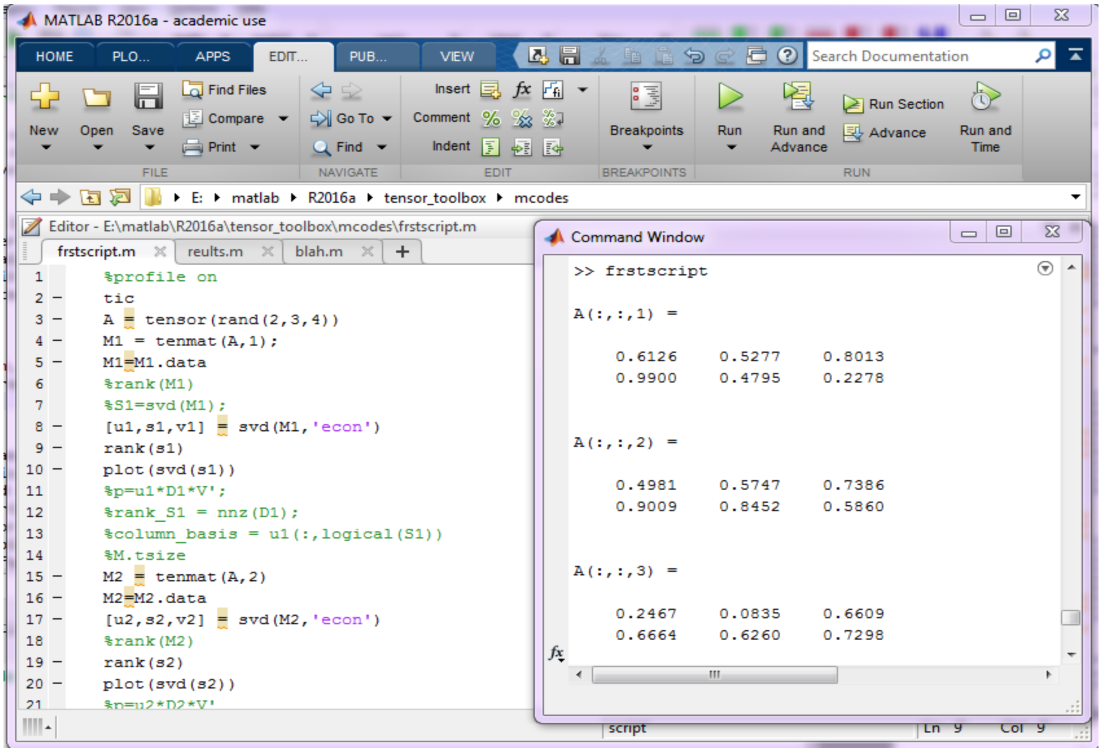


Figure 5.1: Matlab command window with an editor

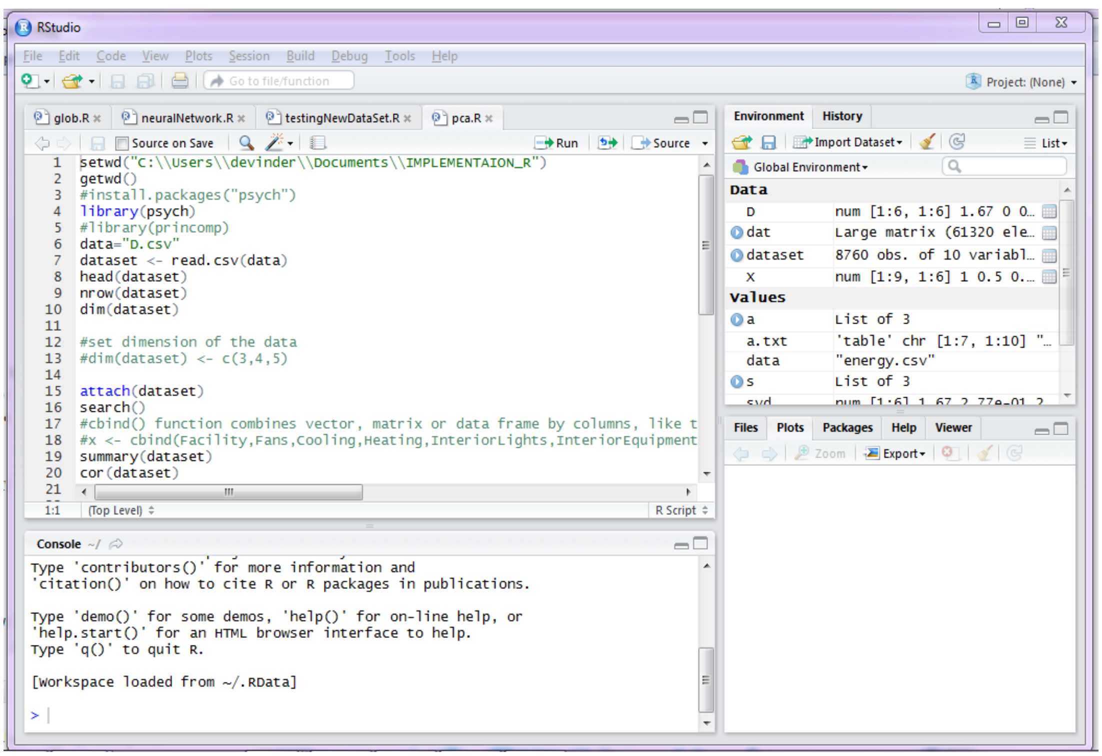


Figure 5.2: RStudio - An IDE of R programming language

5.2 Results and Discussions

In this section, the proposed tensor-based SDN model for management of big data generated by SG devices using proposed scheme is evaluated using data traces for individual household electric power consumption [43]. The dataset consist of 2075259 measurements gathered with a one-minute sampling rate for about 4 years (December 2006 to November 2010). The dataset contains some missing values along with various sub-metering and electrical quantity values [43]. The results obtained after extensive simulation are compared with HOSVD scheme using 'R' programming and Matlab. The objective of proposed scheme is to minimize the reconstruction error ratio between unified tensor and reduced tensor using Frobenius norm. To evaluate the proposed scheme, a network topology is designed in Mininet network emulator [44].

5.2.1 Evaluation Parameters

The proposed scheme has been evaluated using following parameters.

- **Dimensionality Reduction Ratio** (λ): It is the ratio of the non-zero values of the reduced tensor and orthonormal vectors to the non-zero values of the initial tensor. The λ for the initial tensor T_{uni} is obtained by using the equation as given below.

$$\lambda = \frac{nz(\hat{T}_{red}) + \sum_{i=1}^{n} nz(U_i)}{nz(T_{uni})} \tag{5.1}$$

- **Approximation Accuracy**: It is defined as the trade-off between the reconstruction error ratio (ρ) and dimensionality reduction ratio (λ). Both the ratios are inversely proportional to each other. The relation can be defined as below:

$$\rho \alpha \frac{1}{\lambda} \tag{5.2}$$

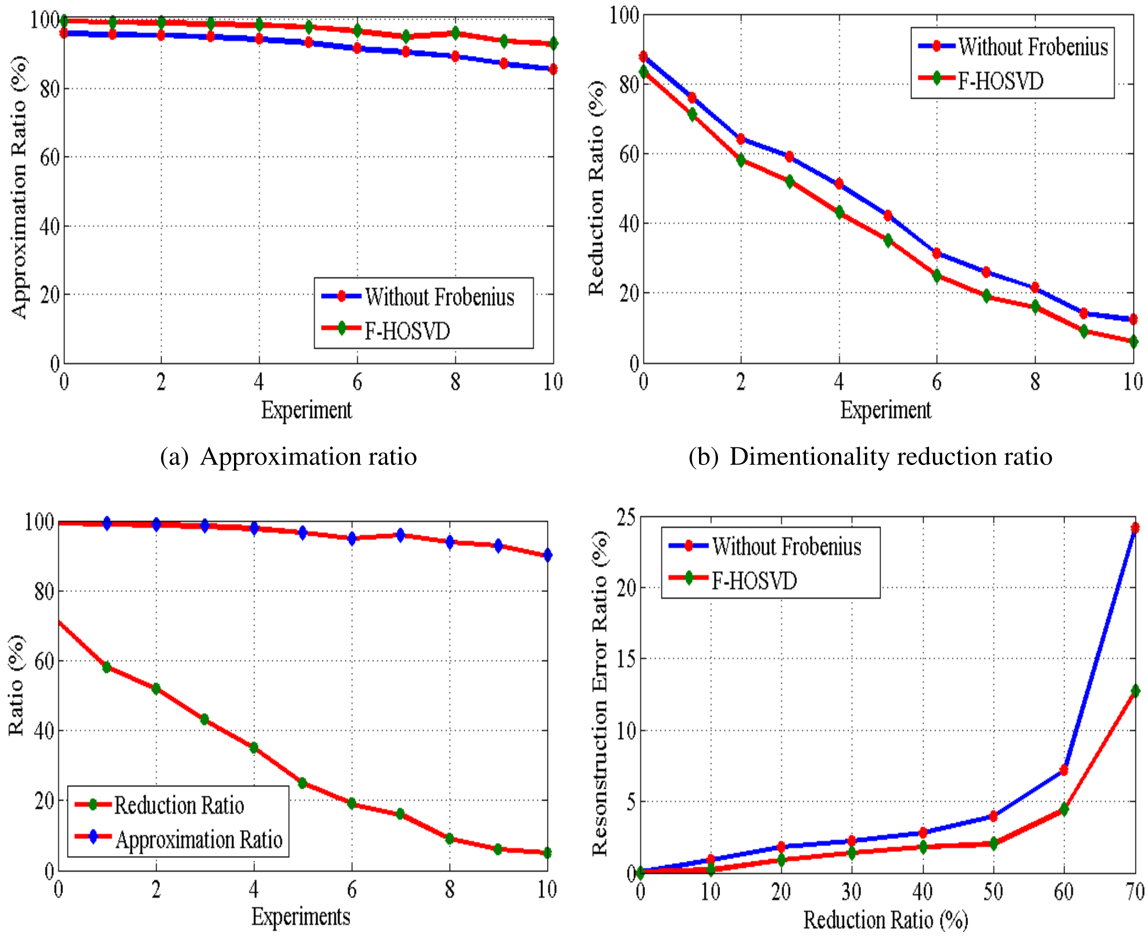
- **Delay** (d): It is defined as the latency at a specific router which comprises of processing delay (d_{pr}), queuing delay (d_q), transmission delay (d_t), and propagation delay (d_{pg}) [44].

$$d = d_{pr} + d_q + d_t + d_{pg} \tag{5.3}$$

- **Network Throughput**: It is the rate of successful delivery of message over a certain communication channel. It can also be called as the maximum rate at which data can be processed.

5.2.2 Evaluation Results

The data acquired is converted into sub-tensors (using Eq. (3.2)). Then, the sub-tensors are combined to form a unified tensor (using Eq. (3.4)). The unified tensor is obtained by applying unified data tensorization operation on the sub-tensors. The unified tensor is reduced to obtain a lower order tensor using F-HOSVD technique. The unified tensor combines the sub-tensors to remove all the ambiguities, redundancies to obtain a simplified combined tensor.



(a) Approximation ratio

(b) Dimentionality reduction ratio

(c) Reduction ratio vs Approximation ratio

(d) Recontruction error ratio

Figure 5.3: Evaluation results for the proposed scheme

The reduced tensor is an approximation of the original data which contains all the valuable information. The results obtained show that the approximation ratio obtained using the proposed scheme is more as compared to the existing technique. The approximation ratio decreases from 99.5% to 89.9% with respect to a decrease in reduction ratio from 83.4% to 5%. Hence, it is clear that the nearly 90% originality of the data is maintained even

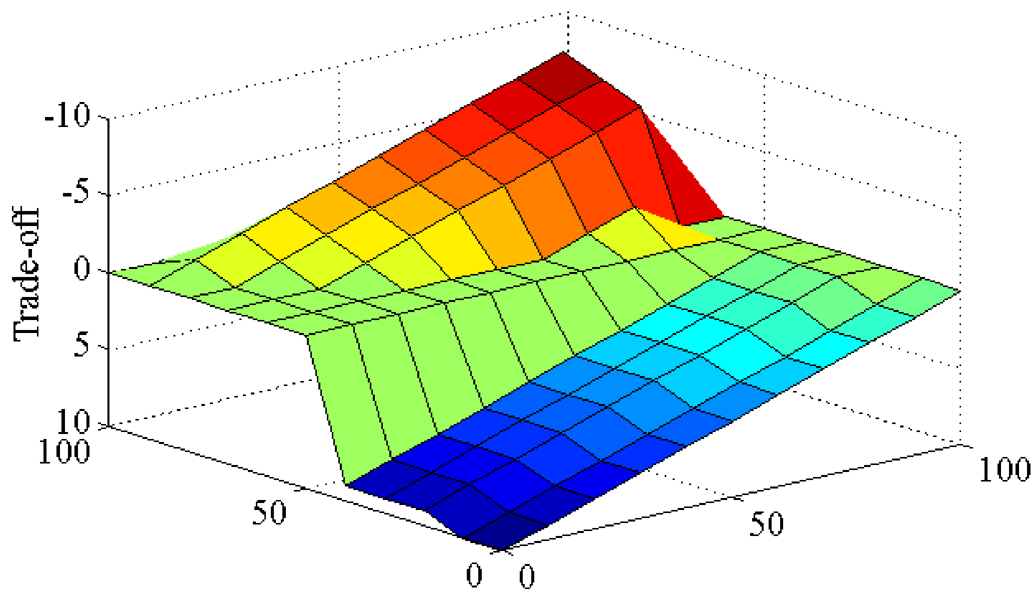


Figure 5.4: Trade-off between reduction ratio and reconstruction error

after reduction up to 5%. Data decomposition plays a significant role in the dimensionality reduction of data. In the core heart of data decomposition lies algorithms such as SVD and PCA. In the context of data decomposition, Table 5.1 shows the mathematical results for singular values carried out on samples of three samples with different modes using single value decomposition.

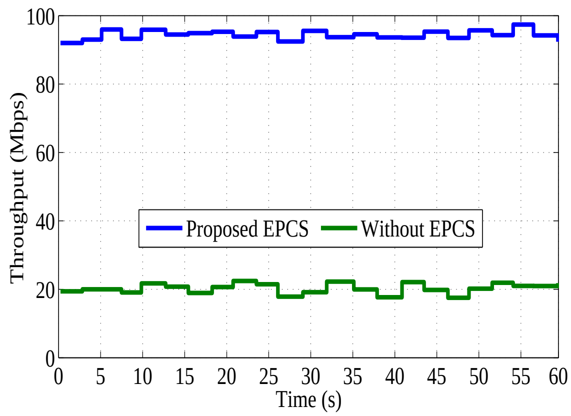
Figure 5.3(a) shows the approximation ratio obtained after performing experiments on

Table 5.1: Mathematical results

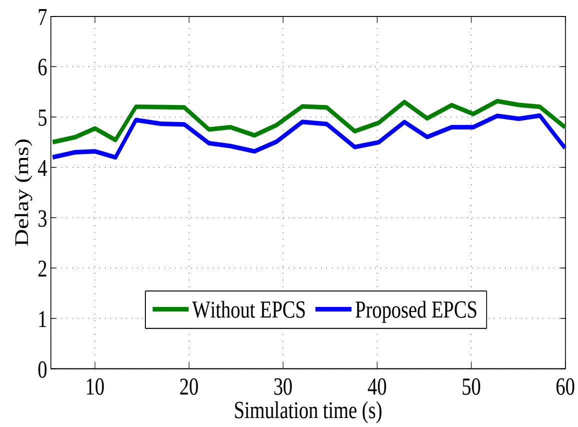
Matrix no.	i	σ_1	σ_2	σ_3
M1	5	271.626	5.391	0.301
M2	4	271.646	4.267	0.000
M3	3	271.662	3.119	0.005

the original data. Further, the reduction ratio obtained using the proposed scheme is shown in Figure 5.3(b). It shows that the original tensor is reduced to a higher extent as compared to simple HOSVD. Hence, it clearly depicts that the data is reduced to a higher ratio while maintaining originality. The comparison of reduction ratio with respect to the approximation ratio is shown in Figure 5.3(c). The above results are obtained using Frobenius norm which is applied incrementally on the singular matrix to achieve a lower rank matrix. The above conclusion is further supported by the results obtained for the reconstruction error ratio. The reconstruction error ratio obtained for the experiments using the proposed scheme is shown in Figure 5.3(d). The results depict that reconstruction error ratio is lower as compared to the existing technique. Therefore, the proposed scheme shows higher level of originality while achieving lower reconstruction error. The above results are achieved using Frobenius norm which is applied incrementally on tensors after n-mode product. Hence, the overall objective of minimizing the reconstruction error ratio is achieved using the proposed scheme. The trade-off between reduction ratio and error ratio is shown in Figure 5.4.

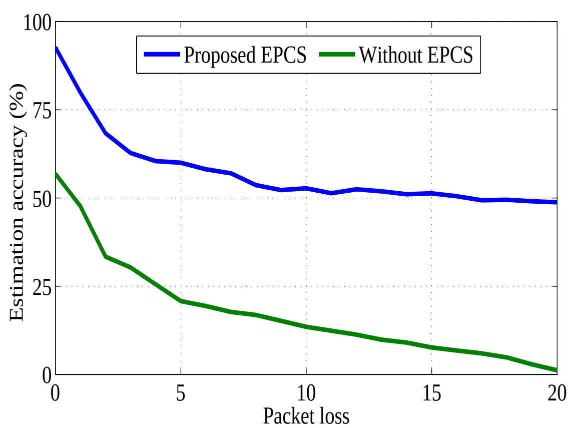
Once the data acquired from various SG devices has been reduced into core data then



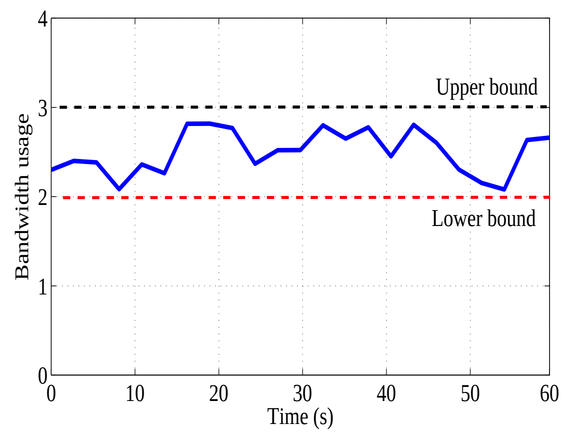
(a) Throughput



(b) Delay



(c) Estimation accuracy



(d) Bandwidth usage

Figure 5.5: Evaluation results for the proposed scheme

it has to be processed and transmitted over SG networks using SDN infrastructure. For this purpose, an empirical probability-based control scheme has been designed to estimate an optimal path for the reduced data. After evaluation of the proposed scheme, it is evident that the all the performance metrics shows a suitable growth. Figure 5.5(a) shows the throughput achieved for the proposed route estimation scheme. The results obtained shows a higher throughput is achieved by using empirical probability-based control scheme. Also, the delay incurred for transmitting the data to the destination is lower with respect to standard SDN routing scheme. Figure 5.5(b) shows the delay incurred while transmitting the reduced data over SG networks using the proposed route estimator along with SDN. The proposed scheme is evaluated for the estimation accuracy with respect to packet loss. The results obtained are evident that the optimal route estimation accuracy for the proposed scheme is better that the standard open flow scheme. It is also evident that the accuracy drop is less for the proposed scheme with respect to increase in packet loss. Figure 5.5(c) shows the achieved estimation accuracy with respect to packet loss. In this regard, Table II shows the values of RMSE obtained for estimated routes.

Finally, the proposed scheme is evaluated with respect to the bandwidth usage. The results clearly depict that the stability of bandwidth usage is maintained within the upper bound and lower bound. Hence, it shows that no congestion or bandwidth over or under utilization occurs. Figure 5.5(d) shows the bandwidth usage for the proposed scheme for the estimation of routes. This strongly shows that the available bandwidth is optimally utilized by the control scheme for transmitting reduced data over SG networks. The above discussed evaluation results depict the effectiveness and efficiency of the proposed scheme with respect to various performance metrics.

Table 5.2: RMSE values for EPCS

Samples	Data sent	RMSE	Bandwidth range
S1	8 Mbits	6.025	(10-50) Mbps
S2	16 Mbits	4.1472	(100-150) Mbps
S3	32 Mbits	18.245	(150-200) Mbps
S4	64 Mbits	13.9355	(200-250) Mbps
S5	128 Mbits	14.09	(250-300) Mbps

Chapter 6

Conclusion

6.1 Concluding Remarks

In this thesis work, a tensor-based SDN model for dimensionality reduction problem for big data acquired from various SG devices is proposed. For this purpose, a F-HOSVD algorithm is designed. The purpose of the proposed scheme is to represent the bulk data generated by SG devices in the form of tensors. The heterogeneous and highly dimensional big data is represented in the form of sub-tensors. After data representation, the sub-tensors are combined to form a unified tensor. Finally, the proposed algorithm for dimensionality reduction is applied on the unified tensor to reduce it. The proposed scheme is validated using data traces for individual household energy consumption. The results obtained show that the proposed scheme achieves higher dimensionality reduction while maintaining a high ratio of originality. The efficacy of the proposed scheme has been evaluated using extensive simulations carried out on the data traces (power consumption of appliances in different smart homes) using 'R' programming and Matlab. Also, the reconstruction error ratio of the data is minimal as compared to the existing techniques. The proposed scheme for dimensionality reduction of SG big data is also justified using a mathematical model.

The massive data generation may generate issues such as congestion and delay on the networking infrastructure deployed between end users and grid. So, an empirical probability-based control scheme for SDN is proposed to estimate path for forwarding the reduced data. The results show that the estimated path show lower latency and high throughput. Moreover, the proposed scheme maintains a high route estimation accuracy with respect to increase in packet loss. Finally, the bandwidth utilization remains stable and the proposed scheme avoid any congestion or under utilization of bandwidth. Hence, the overall results obtained for proposed schemes related to data management, dimensionality reduction, and route estimation show better performance than existing schemes.

6.2 Summary of Contributions

The major contributions of this thesis work are given as below.

1. A tensor-based data management scheme is designed for representation and dimensionality reduction of data acquired from various smart devices in SG systems. Moreover, Frobenius norm is applied to optimize the reconstruction error of the reduced tensor.

2. An empirical probability-based control algorithm is designed for estimating the optimal route to forward the reduced data over SG networks using SDN.

3. A mathematical case study is formulated to explain the working of the proposed tensor-based data representation scheme.

4. The proposed scheme is evaluated using extensive simulations on realistic traces of high dimensional SG database using various parameters.

6.3 Future Scope

In future,

1. By taking the motivation from big data handling scheme and efficient data processing, data intelligence can be pursued extensively in SG systems.

2. Various SG issues such as demand response management and load balancing can be well optimized and managed using machine learning and statistical techniques. Moreover, to enhance the Quality of Service (QoS) in SG systems using big data analytics and SDN can also become one of the major research issues in the coming days.

Bibliography

[1] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, “A survey on software-defined networking,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2015.

[2] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, “Smart grid technologies: communication technologies and standards,” *IEEE transactions on Industrial informatics*, vol. 7, no. 4, pp. 529–539, Nov 2011.

[3] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, and V. Prasanna, “Cloud-based software platform for big data analytics in smart grids,” *Computing in Science Engineering*, vol. 15, no. 4, pp. 38–47, Jul 2013.

[4] A. E. Outlook *et al.*, “Energy information administration,” *Department of Energy*, 2010.

[5] V. C. Gungor, B. Lu, and G. P. Hancke, “Opportunities and challenges of wireless sensor networks in smart grid,” *IEEE Transactions on Industrial Electronics*, vol. 57, no. 10, pp. 3557–3564, Oct 2010.

[6] Y. Demchenko, C. de Laat, and P. Membrey, “Defining architecture components of the big data ecosystem,” in *International Conference on Collaboration Technologies and Systems*, 2014, pp. 104–112.

[7] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao, and H. Liu, “A big data architecture design for smart grids based on random matrix theory,” *IEEE Transactions on Smart Grid*, 2015, doi: 10.1109/TSG.2015.2445828.

[8] W. Hou, Z. Ning, L. Guo, and X. Zhang, “Temporal, functional and spatial big data computing framework for large-scale smart grid,” *IEEE Transactions on Emerging Topics in Computing*, 2017.

[9] J. Baek, Q. H. Vu, J. K. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE transactions on cloud computing*, vol. 3, no. 2, pp. 233–244, 2015.

[10] G. C. Koutitas and L. Tassiulas, "Low cost disaggregation of smart meter sensor data," *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1665–1673, 2016.

[11] H. Nguyen, A. Khodaei, and Z. Han, "A big data scale algorithm for optimal scheduling of integrated microgrids," *IEEE Transactions on Smart Grid*, 2016.

[12] J. C. S. de Souza, T. M. L. Assis, and B. C. Pal, "Data compression in smart distribution systems via singular value decomposition," *IEEE Transactions on Smart Grid*, 2015, doi: 10.1109/TSG.2015.2456979.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[14] J. Ning, J. Wang, W. Gao, and C. Liu, "A wavelet-based data compression technique for smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 212–218, Mar 2011.

[15] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[16] L. Kuang, F. Hao, L. T. Yang, M. Lin, C. Luo, and G. Min, "A tensor-based approach for big data representation and dimensionality reduction," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 280–291, Sep 2014.

[17] L. T. Yang, L. Kuang, J. Chen, F. Hao, and C. Luo, "A holistic approach to distributed dimensionality reduction of big data," *IEEE Transactions on Cloud Computing*, 2015, doi: 10.1109/TCC.2015.2449855.

[18] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425–436, Feb 2016.

[19] D. Zhou, J. Guo, Y. Zhang, J. Chai, H. Liu, Y. Liu, C. Huang, X. Gui, and Y. Liu, "Distributed data analytics platform for wide-area synchrophasor measurement systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2397–2405, Sep 2016.

[20] F. P. Plaza and E. I. Ortega, "State of art, meter data management system using compressed sensing for AMI based on wavelet." *IEEE Latin America Transactions*, vol. 13, no. 12, pp. 3774–3780, 2015.

[21] Y. Wang, Q. Chen, C. Kang, Q. Xia, and M. Luo, "Sparse and redundant representation-based smart meter data compression and pattern extraction," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2142–2151, 2017.

[22] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[23] D. B. Rawat and S. R. Reddy, "Software defined networking architecture, security and energy efficiency: A survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 325–346, 2017.

[24] Y. Jarraya, T. Madi, and M. Debbabi, "A survey and a layered taxonomy of software-defined networking," *IEEE communications surveys & tutorials*, vol. 16, no. 4, pp. 1955–1980, 2014.

[25] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[26] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and openflow: From concept to implementation," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2181–2206, 2014.

[27] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[28] K. Jeong, J. Kim, and Y.-T. Kim, "Qos-aware network operating system for software defined networking with generalized openflows," in *IEEE Network Operations and Management Symposium*, 2012, pp. 1167–1174.

[29] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, "NOX: towards an operating system for networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 3, pp. 105–110, 2008.

[30] T. Luo, H.-P. Tan, and T. Q. Quek, "Sensor openflow: Enabling software-defined wireless sensor networks," *IEEE Communications Letters*, vol. 16, no. 11, pp. 1896–1899, 2012.

[31] H. Kim and N. Feamster, "Improving network management with software defined networking," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 114–119, 2013.

[32] L. Ren, Y. Qin, B. Wang, P. Zhang, P. B. Luh, and R. Jin, "Enabling resilient microgrid through programmable network," *IEEE Transactions on Smart Grid*, 2016, doi: 10.1109/TSG.2016.2589903.

[33] R. Narayanan, K. Xu, K.-C. Wang, and G. K. Venayagamoorthy, "An information infrastructure framework for smart grids leveraging SDN and cloud," in *Clemson University Power Systems Conference*, 2016, pp. 1–7.

[34] W. Zhong, R. Yu, S. Xie, Y. Zhang, and D. H. Tsang, "Software defined networking for flexible and green energy internet," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 68–75, 2016.

[35] A. Cahn, J. Hoyos, M. Hulse, and E. Keller, "Software-defined energy communication networks: From substation automation to future smart grids," in *IEEE International Conference on Smart Grid Communications*, 2013, pp. 558–563.

[36] G. S. Aujla, A. Jindal, N. Kumar, and M. Singh, "SDN-based data center energy management system using RES and electric vehicles," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[37] D. Jin, Z. Li, C. Hannon, C. Chen, J. Wang, M. Shahidehpour, and C. W. Lee, "Towards a cyber resilient and secure microgrid using software-defined networking," *IEEE Transactions on Smart Grid*, 2017.

[38] N. Chen, M. Wang, N. Zhang, X. S. Shen, and D. Zhao, "Sdn-based framework for the pev integrated smart grid," *IEEE Network*, vol. 31, no. 2, pp. 14–21, 2017.

[39] E. A. Leal and J. F. Botero, "Transforming communication networks in power substations through sdn," *IEEE Latin America Transactions*, vol. 14, no. 10, pp. 4409–4415, 2016.

[40] L. Cui, F. R. Yu, and Q. Yan, "When big data meets software-defined networking: SDN for big data and big data for SDN," *IEEE Network*, vol. 30, no. 1, pp. 58–65, Jan 2016.

[41] L. Kuang, L. T. Yang, X. Wang, P. Wang, and Y. Zhao, "A tensor-based big data model for QoS improvement in software defined networks," *IEEE Network*, vol. 30, no. 1, pp. 30–35, 2016.

[42] S. Agarwal, M. Kodialam, and T. Lakshman, "Traffic engineering in software defined networks," in *IEEE INFOCOM*, 2013, pp. 2211–2219.

[43] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[44] F. Ongaro, "Enhancing quality of service in software-defined networks," Ph.D. dissertation, Alma Mater Studiorum-University of Bologna, 2014.

Publications

1. D. kaur *et al.*, ”Tensor-based Big Data Management Scheme for Dimensionality Reduction Problem in Smart Grid Systems: SDN Perspective”, IEEE TKDE, under review.

Video Presentation Link

1. https://youtu.be/3MJLRSkgHQw