International Journal For Technological Research In Engineering
Volume 6, Issue 11, July-2019 *ISSN (Online): 2347 - 4718*

ROLE OF BIG DATA IN SMART GRID- A SURVEY, CHALLENGES AND SOLUTIONS

Rohit Gupta[1], Dr. KT Chaturvedi[2]
[1]PhD Scholar, [2]Associate Professor, UIT RGPV Bhopal

Abstract: This paper provides a literature review on smart grids and big data. Smart grid refers to technologies used to modernize the energy delivery of traditional power grids, using intelligent devices and big data technologies. The modernization is performed by deploying equipment such as sensors, smart meters, and communication devices, and by invoking procedures such as real- time data processing and big data analysis. A large volume of data with high velocity and diverse variety are generated in a smart grid environment. This paper presents definitions and background of smart grid and big data. Current studies and research developments of big data application in smart grids are also introduced. Additionally, big data challenges in smart grid systems such as security and data quality are discussed.
Keywords: Big data, Demand response, Electric vehicles, Quality, Renewable energy, Smart grid, Security.

I. INTRODUCTION

The traditional electric grid based on centralized generation plants and unidirectional transmission and distribution systems is transitioning to a smart grid that is decentralized and multidirectional with a high integration of information and communication technologies. The IEEE 2030 standard [1] states that the smart grid system is based on an interconnection of three systems: 1) the electric power system which emphasis the power generation, transmission, distribution and consumption of power. 2) the communication system which emphasis the communication connectivity among systems, devices and applications. 3) the information technology system which includes technologies that store, process and manage data information for decision making on the power system operation. With this development of the smart grid, large amounts of smart meters and sensors are being deployed with huge coverage. As a result, a large number of multi- sourced heterogeneous smart grid data is being produced. Enormous value can be extracted from this smart grid data that can enhance the quality of the grid, also provide better service for different types of customers. The smart grid data is large in volume, high in velocity (moving from a system that is read once every month to a system that generates readings every few minutes), and wide in variety (different types of data are generated from various resources). Interestingly, the characteristics of the nature of smart grid data can be considered as a big data challenge that requires advanced information technology systems and cyber infrastructure to handle and analyze this huge amount of data. Data mining applications have been widely used to pro- mote the reliability and automation of the electric grid, clustering [2]–[6], classification [7], [8] and prediction [8]–[10] have been main topics of research during past years. However, the traditional data management techniques and applications are not designed to handle big data. Therefore, developing frameworks that address the challenges of smart grid big data analytics are of research interest. A cloud based dynamic demand response platform for smart grid big data is presented in [11]. Also, a cloud based visual analytics framework to monitor the grid status, including micro-generators and prosumers, is presented in [12]. A recent framework that covers the life cycle of smart grid big data from generation to analytics is presented in [13]. The work in [13] introduces a framework utilizing state-of-the-art big data components to address the smart grid big data challenges. Also, various data analytical applications can be performed on top of the frame- work. However, the framework presented in [13], is not able to scale with big data applications that require low latency

II. OVERVIEW OF SMART GRID AND BIG DATA

Smart Grid

The deployment of smart meters and sensors throughout the grid results in massive amounts of data. This includes generation side data (wind farms and photovoltaic plants), consumption side data (residential homes, factories and elec-tric vehicle charging stations), prosumers data (residential photovoltaic panels and vehicle-to-grid) and, weather and natural disasters data can be included in the smart grid system. Also, images and video footage could be included to detect physical attacks (California transmission substation sniper attack [23]) or investigate power outages. The smart grid data is considered to be large in volume, high in velocity and wide in variety. The value of this smart grid big data becomes useful when integrated with multi-sourced existing smart grid data in an analytics environment, and can potentially enhance the functionality of the smart grid. Fig. (1) Shows the structure of traditional and smart grids [16]. The traditional power grid includes unidirectional transmission, meaning that power flows from power generators to consumers [17]. Smart grid systems, on the other hand include bidirectional transmission, data driven system, and renewable energy resources to offer additional utilities to customers, distributers, and providers [17]. Despite all its benefits, smart grids have difficulty in handling large

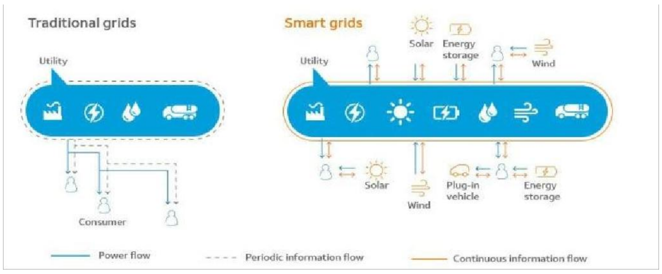

Fig. (1). Traditional grid vs. smart grid [16].

volume of data within an acceptable time limit and hardware resources [18].

Big data in Smart Grids

"Big data has high volume, high velocity, and/or high variety information assets that require new forms of processing," said Douglas Laney [19]. Smart grids require information from sources including sensors, smart meters, Phasor Measurement Units (PMUs), Geographic Information Systems (GIS), weather data, population data, internet data and energy market pricing and bidding data collected through Automated Revenue Metering systems (ARMs). Notwithstanding the extent of these informational collections, the absence of physical or worldly connection between's their components renders them past the extent of conventional examination strategies [2]. Important state data from all elements of the lattice (at all degrees of age and burden) must be spoken with insignificant inactivity to stake holding respondents that rely upon this data as working parameters [12].

Big data analytics are the key to developing modern technologies that facilitate interaction among the smart grid main components including hardware, software, network, user, server, and data [17]. Big data analytics rely on data mining and modeling algorithms that facilitate corrective, predictive, distributed and adaptive decision making techniques [18]. The diversity of information in the power grid's big data sources requires the use of batch, streaming, and interactive processing methods for optimal handling based upon the attributes of the data [17]. The big data attributes can be described by the 4V's model: volume, velocity, variety, and value [20, 21]. Big data in smart grids features similar "4V" characteristics [22, 23].

Volume

Service organizations are supplanting conventional meters with keen meters, which produce huge measure of information [24]. In a huge service organization with one million keen meters, if each 15-minute information is gathered, 35.04 billion records with volume of 2920 TBs information will be created [25]. The drastic increase in electric power systems data volume introduces several challenges which will be further discussed in section 4.

Velocity

Speed in vitality huge information setting alludes to the speed of putting away, handling and investigating the information. Dissimilar to conventional information insight gadgets, the capacity and preparing of vitality enormous information require quick and continuous ability [26]. Gushing information handling is utilized permitting social information inquiries to be consistently refreshed. High speed information is broke down as far as stream-to-connection, connection to-connection or connection to-stream inquiries [22]. Regular questioning dialects utilized incorporate Cassandra Query Language (CQL), Stream Processing Language, Spark Streaming, Storm, and Fink Framework and Apache Drill [2, 17, 22]. The outcome is the ongoing association with information enduring ostensible idleness. Specially appointed questions can be handled in PetaByte (PB) extents inside a couple of moments [2]. Along these lines, the speed of information handling can be diminished to a couple of moments enabling the vitality framework to settle on quick and brief choices, for example, deficiency location by means of PMUs and matrix self-mending reactions[18].

Variety

There are regularly three unique information types in shrewd vitality frameworks: Structured, semi-organized, and unstructured. The level of structure is characterized by the arrangement of the substance exhibited: records with qualities grouped by unmistakable classifications (for example call records from a telecom organization) are viewed as organized while graphical information getting a relationship from the plot of factors is viewed as semi-organized. A totally freestyle content passage, for example, a Twitter post or online audit is unstructured information [22]. In a savvy matrix, vitality utilization information establishes the organized information; correspondence information among clients and seller gadgets structure the semi-organized information; and vitality use email or SMS notices are instances of unstructured information[24].

Value

Worth is a consequence of the initial three V's with some calculation included. This is the reason Monica Rogati says, "More information beats sharp calculations, yet better information beats more information" [27]. Vitality enormous information has esteem once gone through calculation to help business choices or help clients [24]. For specialist co-ops, esteem renders into making focused showcasing methodologies by breaking down the client vitality utilization designs. Clients could likewise profit by vitality investment funds, straightforwardness in their vitality utilization and improved operational effectiveness [24]. Worth additionally relies upon the eye of the spectator. A matrix administrator would not think about the temperature of a solitary house or how streamlined the traffic lights are between one another. This is the reason it is so critical to incorporate Value in the depiction of what establishes enormous information.

III. RESEARCHES RELATED TO BIG DATA APPLICATIONS IN SMART GRID

Threemaincategoriesareidentifiedforsmartgridbigdataapplications: Renewable Energy(RE),Demand Response (DR), and

Electric Vehicles (EV) [2]. Renewable Energy With expanding coordination of sustainable power sources in power frameworks, information the executives of current vitality matrices turns into a mind boggling task, which ought to be tended to by enormous information examination [28, 29]. For instance, authentic climate information and GPS information can be utilized to improve gauging of sustainable power source control age, which at last upgrades the matrix vitality proficiency [30]. Information mining and handling have been utilized to concentrate highlights of time arrangement information for progressively precise determining of discontinuous sustainable assets, for example, wind and sun oriented [31 - 34].

Demand Response

Demand response refers to changes in customers' electricity consumptions in response to changes in the electricity cost and availability [37]. Flexible loads such as Heating, Ventilation and Air Conditioning (HVAC), which "need to run but their exact time of operation is not critical" and other controllable loads such as electric vehicles are the targets of demand response programs [38]. Traditional power systems do not offer real-time demand response, which degrades grid reliability and adequacy. Therefore, big data technologies are used in smart grid management to improve the electricity consumption data accessibility, which expands the demand response [39]. For example, advanced meters apply game theory and modern communication technologies enabling smart grids to provide real-time demand response capability for more efficient and reliable operation of the grid [40, 41]. A study reported that during the California electricity crisis, the price of electricity could have been halved if the demand decreased by five percent [42]. U.S. government issued Federal Energy Regulatory Commission (FERC) Order 719 to improve the electricity wholesale markets by establishing rules and regulation for demand response [43]. Additionally, the US government enacted the American Recovery and Reinvestment Act of 2009, which is a 4.5 billion U.S. dollar funding of smart grid technologies as a means to improve the U.S. electric grid systems [44].

Electric Vehicles

The International Energy Agency reports that more than 1.2 million Electric Vehicles (EVs) were operating in 2015 [45] in the world. In the US in 2015, 400,000 were operating making about 1/3 of the world's total use of EV's. EVs charge their batteries through the grids, which imposes a significant impact on electric grid systems [46 - 48]. For example, charging EVs in a populated area during the peak time may have consequences such as fuse blowouts, decreased efficiency, and transformer degradation [49 - 51]. Through its bidirectional communication technology, smart grids can address these issues by scheduling the EV charging for off-peak hours [52]. In addition, by coordinated discharging through their vehicle-to-grid (V2G) capabilities, EVs can provide several benefits such as ancillary services, mitigating uncertainties of intermittent renewable energy sources such as wind and solar, etc [53], [54 - 56]. There are several studies for coordinating the EV charging/discharging to benefit electric utilities and their customers using genetic algorithms. EV driving and charging data have been extensively analyzed by researchers to address the issues associated with high penetrations of EVs in electric grids. A team of researchers used an Estimation of Distribution Algorithms (EDAs) and population-based probabilistic search algorithms to optimally manage the enormous number of EV's charging [57]. Such algorithms require the capability to process vast and large volume of real-time data, which heavily depends on server-based processing or distributed processing networks. Another study presented a framework for EVs charging demand using big data analysis on data generated by smart meters [58]. Big data modeling for EV battery was proposed in [59] to improve estimation of driving ranges with big data cloud computing. Another study presented decision making strategies for EV charging by analyzing the predicted generation and demand through the use of queue distributions in a distributed network [60]. Table 1 offers interesting research for big data applications in smart grids.

Table 1. Big Data Applications in Smart Grids – Methods and Case Studies.

Application	Ref. #	Method(s)	Case Studies
Renewable energy	[28]	The means of communications through long distance or remote stations using energy efficient cellular communication	Off-grid or standalone base stations powered by local small-scale renewables to not require grid power for communication.
	[29]	Multiple models for current, future, and virtual energy markets used to optimize PV integration into a micro grid.	A 65 solar panel array with 15 kWH energy storage is simulated. The system operation is evaluated without any energy sales, with sales restricted to local users,
	[31]	An enhanced K-means algorithm, named Time Series Clustering (T.S.C) K-means, combined with Multilayer Perceptron Neural Networks (MLPNN) for solar	Several meteorological time-series datasets are used to assess the performance of the proposed T.S.C K-means clustering method and its comparison with other clustering techniques including K-means*, K-means++,K-means, self-organizing

	[32]	A novel time-series based K-means clustering method, named T.S.B K-means, combined with discrete Wavelet Transform (DWT), Harmonic Analysis Time Series	Wind speed, wind power, wind direction, and air temperature data from National Renewable Energy Laboratory (NREL) are used to evaluate the novel clustering and hybrid forecasting methods. A comparative analysis of the proposed
	[33]	A Transformation-based K-means algorithm, named TB K-means, combined with MLPNN for solar radiation forecasting.	Several different datasets are used to evaluate the proposed TB K- means clustering and compare it with different variants of K-means algorithm. Solar radiation time series with different characteristics are used to provide a comparative analysis between the proposed
	[34]	A novel Game Theoretic Self-organizing Map (GTSOM), combined with Neural gas (NG) and Competitive Hebbian Learning (CHL), DWT and Bayesian Neural	Historical solar radiation data are used to assess the performance of the hybrid forecasting with the proposed GTSOM and other clustering methods.
Demand response	[39], [40]	An extended framework of the Stackelberg game model for demand response optimization.	Homogeneous and heterogeneous generation supply quantities, generator profit and consumer welfare are evaluated in scenarios with few and many generation units and a large consumer population.
	[49]	Method of defining a more accurate model of electric consumption by light duty Plug-in Electric Vehicles (PEVs).	Uncontrolled home charging of EVs and uncontrolled “opportunistic” charging at public locations are simulated based on travel survey

Electric vehicle	[51]	A fuzzy expert method for online management of EVs’ charging demand.	An IEEE 38 bus distribution test feeder including charging stations at 4 nodes is simulated. .Different charging solutions/scenarios are
	[52]	A sliding horizon-based method for real-time data management and optimal coordination of EV charging with photovoltaic (PV) generation.	A 33 bus system including DG units and EV charging stations is simulated. EV charging coordination and its effect on PV power curtailment is evaluated.
	[55]	A hybrid of Auto Regressive Moving Average (ARMA), Fuzzy C-Means (FCM) clustering, Monte Carlo Simulation (MCS), and Particle Swarm Optimization (PSO) methods for optimal scheduling of EVs to increase the use of PV power for EV charging while providing economic revenues for EVs’ participation in V2G services.	A 12 MW PV system with 424 EVs is simulated. A collaborative strategy is developed between the EV aggregators and PV producers to minimize the penalty cost of PV over/under-production by charging the EVs using the PV power in excess of the scheduled output and discharging the V2G power to compensate the PV power under-production. The system performance with and without EV optimal charging/discharging are evaluated and compared.

	[56]	A hybrid of ARMA, FCM clustering, MCS, and Genetic Algorithm (GA) methods for optimal scheduling of EVs to increase the use of wind power for EV charging while providing economic revenues for EVs' participation in V2G services.	A 10 MW wind system with 484 EVs is simulated. A bilateral contract is developed between the EV aggregators and wind producers to use the extra wind power for EV charging and to discharge the V2G power during the periods of wind power deficits. The system performance with and without EV optimal charging/discharging are evaluated and compared.

IV. SMART GRID BIG DATA CHALLENGES AND PROPOSED SOLUTIONS

Three main challenges are identified for big data in smart grids: security, quality, and processing location.

Big Data Security

The use of big data technology in smart grids substantially improves the network connectivity at the price of increased security vulnerabilities [61]. In a big data context, security exposures can be divided into three main parts: privacy, integrity, and authentication.

Data Privacy

Smart meters can be a main privacy concern if their data is not securely transferred and stored [62]. Smart meters collect power consumption data of grid customers. Smart grid providers analyze such data, which provides great intuition about users' behaviors and habits, to offer intelligent customized services [63]. Several methods have been proposed to eliminate and minimize the privacy issue. These methods include, but are not limited to distributed incremental data collection method [64], and masking of consumption data embedded information [65]. Because most of the existing solutions do not consider the tradeoff between costs of lost privacy and data dissemination (utility), a new method is proposed to satisfy both privacy and utility requirements of smart metered data [66].

Data Integrity

Risk of integrity attacks is a valid concern because any violation of integrity may cause security vulnerabilities [67]. Customer and network data are usually the targets for integrity attacks, and any modification of such data interrupts the data communication exchange and reduces the entire grid functionality [2]. For example, attackers can remove the higher degree nodes and replace them with higher probability nodes in the power network, which affects the integrity of data [67]. The data integrity in smart grids and energy markets has been extensively investigated. A study presented the consequences of virtual bidding, which is a method of creating profitable integrity attacking strategies with no or minimal detection in energy markets [68]. Another investigation showed that data integrity attacks can cause unwanted energy generations and routings, which increase the grid operating costs [69]. Market revenues and their changes due to data integrity attacks are used as a measure of adversary impact of such attacks [70, 71].

Data Authentication

Users in smart grids access the communication system through authentication, a process that verifies the user credentials against the accounts credential database [2]. Authentication is used as a tool to identify valid vs non-valid identities within the majority of existing security countermeasures [72]. One critical challenge that smart grids face is message injected attacks. If such attacks are not addressed properly, they can significantly reduce the entire smart grid performance [73]. To address such challenges, a group of scientists proposed an authentication method to secure smart grid data communication exchange with the use of Merkle hash-tree techniques [73]. Another study proposed a secure message authentication mechanism by integrating Diffie-Hellman protocols and hash-based message authentication methods [74]. Such structure allows smart meters within the smart grids to complete mutual message authentication tasks with minimal signal exchange and latency [74].

Big Data Quality

Data quality refers to identifying and to removing the outliers before transferring the data to the system [75]. Energy power consumption data should have high degrees of quality to ensure correct data analysis and ultimately proper decisions. The quality issues of energy consumption data are categorized into noise data, incomplete data, and outlier data [76].

Noise Data

Generally, any data that is difficult to comprehend and/or to decode by computers is considered noise data, which degrades the data quality [76]. In a smart grid context, logical errors and inconsistent energy consumption data are considered noise [77, 78]. Logical errors are defined as the data that violates any given rules and characteristics [79]. For example, if the daily customer energy consumption data includes 25 hours, it is not logical as it exceeds the maximum 24 hours [76]. Moreover, inconsistent data occurs when data does not follow its previously agreed format [80], or it lacks sense when comparing its individual features [81, 82].

Incomplete Data

As the smart grid data complexity increases, incompleteness is occasionally observed in energy consumption data. Several methods such as delete tuple and data filing are developed to address incomplete data [82]. Delete tuple method simply removes the entire record with incomplete data. However, this method is not appropriate for cases where the data set has several incomplete observations [76]. In such cases, the incomplete data will be filled using advanced algorithms such as average value, artificial value, and regression analysis [82].

Outlier Data

In statistics, if a point of data is considerably distant from other data points, it is called outlier [83]. In energy consumption data, an outlier may be treated as noise and removed. However, they may hold valuable information and therefore, should be detected to preserve the data quality. One method of detection is data quality mining, which is to audit the data to automatically find outliers [84]. In smart grid systems, outliers should be detected, identified, and analyzed as they contain critical information such as power rationing, device failures, and suspicious indicators among others [85].

4.3. Big Data Processing Location

Processing is a key function for utilizing the algorithms required by big data. The current model for processing is that information is aggregated and sent to a data center to get processed and passed to whomever needs the resultant information. The current framework as described by H. Jiang is the three-level design with the main data processing at the center with two layers around it for aggregation and distribution [2]. There are intermediary processors called FOGs that are regional collection points that also do minimal amounts of processing before passing its collected information to the data center [87]. Edge based processing is becoming a larger part of the framework of big data. With the drop-in price to compute, researchers have started to look back when processors had limitations and are creating low power solutions that can go anywhere and still be able to process at least parts of a machine learning algorithm on small amounts of data. This helps to create the non-invasive load measuring that is only made possible with low power embedded systems [88]. Table 2 provides the literature for each category of big data challenges, their proposed solutions along with the solution's main advantage/disadvantage.

V. FUTURE OF BIG DATA IN SMART GRIDS

The future of research in big data use in smart grids is diverse. Big data offers many solutions to the bi-directional flow of information as well as processing and analyzing that information. For a smart grid, big data will be a necessity for realizing the best possible solutions for how we as a society should distribute and utilize renewables as well as how to analyze systems for abnormal conditions such as faults or power outages. The future of the smart grid will depend on building these frameworks such that they can be implemented and utilized in a meaningful way. This will include the planning to real time operation for generators and consumers for current practices to those planned for by 2050 [91].

Challenge	Ref. #	Solution	Advantage / Disadvantage
Security	[63]	A regulatory framework equivalent to Health Insurance Portability and Accountability Act (HIPPA) for smart grid privacy and consumer fraud problems	Would provide clear legislative and legal avenues should problems occur / Bureaucracy would not solve some of the problems provided
	[64]	A distributed incremental aggregation framework for smart meters to protect users' privacy by using homomorphic encryption	Unidirectional functionality not allowing for passing information back to a specific unit; Time delay of communication in possible real time operations; Does not look into malicious or fraudulent data acquisition.
	[65]	Using a battery connected between the home and the grid so that anyone looking at the power usage will see a battery charging and not the current profiles of the actual items using power	Makes power usage indistinguishable from one day to the next; Overhead of installation and usage and wear and tear costs of a battery system in a home; Difficult to hide high power usage items such as AC, washer, dryer, *etc.*
	[66]	Privacy vs utility: How to get the best of both worlds without sacrificing too much on either side.	Balanced framework / Gives up privacy information of high power item usage as well as the price of the battery
	[67]	Targeted attacks vs random attacks to smart grid: Building faster and more resilient networks to fend off attacks through the communication networks	Faster networks would entail creating a faster protocol to transfer information; Faster connections mean less encryption or protections increasing privacy and attacker problems.
	[69]	Load Redistribution (LR) attacks: Using Multi-start Benders decomposition to find the most damaging immediate attack.	Good attack prevention strategy for this specific type of attack
	[70]	Proposing strategies to detect and localize malicious attacks	Capable of detecting attacks on multiple locations / The number of locations being attacked expands computation.

Challenge	Ref. #	Solution	Advantage / Disadvantage
Quality	[75]	The data mining-based and the state estimation-based electricity consumption outlier data detection methods	Data mining algorithms are faster and better at detecting outliers than traditional methods / Does not account for missing or redundant data.
	[82]	Developing a data mining prototype system (RMINE) for fault diagnosis or system malfunction detection	Capable of obtaining the minimal diagnostic rule set to derive a logical decision in assisting maintenance engineers to diagnose faults
	[86]	Introducing a new class of attacks, called false data injection attacks against monitoring of PMUs or smart grid sensors for state estimation	N/A
Processing location	[88]	Using embedded neural networks to analyze edge-based load information.	Offers privacy concerns by identifying what is being used in a specific area.
	[89]	Creating a micro grid out of a smart home	Makes a good framework out of the smart home / Lack of intelligent connections to the grid makes it unusable.

Table 2. Big Data Challenges in Smart Grids and Proposed Solutions.

VI. CONCLUSION

This paper presents the definitions and applications of integrating big data technologies in smart grid systems based on current studies and research developments. Several research articles are reviewed to understand the current challenges and solutions of big data applications in smart grids and to identify research gaps. Thus, this survey provides new directions to further investigate such applications and challenges to propose innovative solutions for filling the identified research gaps.

REFERENCES

[1] C.W. Gellings, M. Samotyj, and B. Howe, "The future's smart delivery system [electric power supply]",*IEEE Power Energy Mag.,* vol. 2, pp. 40-48, 2004. [http://dx.doi.org/10.1109/MPAE.2004.1338121]

[2] H. Jiang, K. Wang, Y. Wang, M. GAO, and Y. Zhang , "Energy Big Data: A Survey,"", *IEEE*

Access, vol. 4, pp. 3844-3861, 2016. [http://dx.doi.org/10.1109/ACCESS.2016.2580581]

[3] J.Baek,Q.Vu,J.Liu,X.Huang,andY.Xiang,"Asecurecloudcomputingbasedframeworkforbigdatainformationmanagementofsmart grid", *IEEE Transaction on Cloud Computing,* vol. 3, no. 2, pp. 233-244, 2015. [http://dx.doi.org/10.1109/TCC.2014.2359460]

[4] B.Botte,V.Cannatelli,andS.Rogai,"TheTelegestoreprojectinENEL'smeteringsystem",*Proc.18thInternationalConferenceandExhibition on Electricity Distribution,* pp. 1-4, 2005.

[5] Modern grid benefits, Conducted by National Energy Technology Laboratory for the U.S. Department of Energy, 2007.

[6] H. Farhangi, "The path of the smart grid", *IEEE Power Energy Mag.,* vol. 8, no. 1, pp. 18-28, 2010. [http://dx.doi.org/10.1109/MPE.2009.934876]

[7] Z.Bojkovic,andB.Bakmaz,"Smartgridcommunicationsarchitecture:Asurveyandchallenges",*Proc.11thInternationalConferenceonApplied Computers and Computational Science,* pp. 83-89, 2012.

[8] J. Duff, "Smart grid challenges", *Proc. Workshop on High Performance Transaction Systems (HPTS),* 2009

[9] Z.Fan,P.Kulkarni,S.Gormus,C.Efthymiou,G.Kalogridis,andM.Sooriyabandara,"Smartgridcommunications:Overviewofresearchchallenges, solutions, and standardization activities", *IEEE Comm. Surv. and Tutor.,* vol. 15, no. 1, pp. 21-38, 2013. [http://dx.doi.org/10.1109/SURV.2011.122211.00021]

[10] Howbigdatawillmakeusmoreenergyefficient,*Available*.https://www.weforum.org/agenda/2014/05/big-data-will-make-us-energy-efficient/

[11] D.B.Rawat,andC.Bajracharya,"Cybersecurityforsmartgridsystems:Status,challengesandperspectives",*Proc.SoutheastCon2015,*pp. 1-6, 2015.

[12] J.N.Bharothu,M.Sridhar,andR.S.Rao,"AliteraturesurveyreportonSmartGridtechnologies",*Proc.2014InternationalConferenceonSmart Electric Grid (ISEG),* pp. 1-8, 2014.

[13] N.B.M.Isa,T.C.Wei,andA.H.M.Yatim,"Smartgridtechnology:Communications,powerelectronicsandcontrolsystem",*Proc.2015International Conference on Sustainable Energy Engineering and Application (ICSEEA),* pp. 10-14, 2015. [http://dx.doi.org/10.1109/ICSEEA.2015.7380737]

[14] C.Vineetha,andC.Babu,"Smartgridchallenges,issuesandsolutions",*Proc.2014InternationalConferenceonIntelligentGreenBuilding and Smart Grid (IGBSG),* pp. 1-4, 2014.

[15] R.Mahmud,R.Vallakati,A.Mukherjee,P.Ranganathan,andA.Nejadpak,"Asurveyonsmartgridmeteringinfrastructures:Threatsandsolutions", *Proc. 2015 IEEE International Conference on Electro/Information Technology (EIT),* pp. 386-391, 2015. [http://dx.doi.org/10.1109/EIT.2015.7293374]

[16] AT&TSmartGridTechnologySolutions,*Available*.https://www.business.att.com/enterprise/service/internet-of-things/smart-cities/iot-smart-grid/

[17] S.Sagiroglu,R.Terzi,Y.Canbay,andI.Colak,"Bigdataissuesinsmartgridsystem",*Proc.5thInternationalConferenceonRenewableEnergy Research and Applications,* pp. 1007-1012, 2016.

[18] M.Kezunovic,L.Xie,andS.Grijalva,Theroleofbigdatainimprovingpowersystemoperationandprotection*Proc.BulkPowerSystemDynamic and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP)*, 2013.

[19] M.Hardy,"Bigdataanalytics-whatitmeanstotheauditcommunity-part1",*Available*.www.isaca.org/chapters2/jacksonville/events/Documents/ISACA%20Big%20Data%20for%20March%202013%E2%80%93%20Part%201.pdf

[20] T.Zhu,S.Xiao,Q.Zhang,Y.Gu,P.Yi,andY.Li,"Emergenttechnologiesinbigdatasensing:Asurvey",*Int.J.Distrib.Sens.Netw.,*vol. 2015, pp. 1-13, 2015.

[21] C.K. Emani, N. Cullot, and C. Nicolle, "Understandable big data: A survey", *Comput. Sci. Rev.,* vol. 17, pp. 70-81, 2015. [http://dx.doi.org/10.1016/j.cosrev.2015.05.002]

[22] R.Gupta,H.Gupta,andM.Mohania,"Cloudcomputingandbigdataanalytics:Whatisnewfromdatabasesperspective?",*Proc.International Conference on Big Data Analytics,* pp. 42-61, 2012. [http://dx.doi.org/10.1007/978-3-642-35542-4_5]

[23] S. Madden, "From databases to big data", *IEEE Internet Comput.,* vol. 16, no. 3, pp. 4-6, 2012. [http://dx.doi.org/10.1109/MIC.2012.50]

[24] K.Zhou,andS.Yang,"Understandinghouseholdenergyconsumptionbehavior:Thecontributionofenergybigdataanalytics",*Renew. Sustain. Energy Rev.,* vol. 56, pp. 810-819, 2016. [http://dx.doi.org/10.1016/j.rser.2015.12.001]

[25] Big Data, Analytics, and Energy Consumption, *Available*. http://www.lavastorm.com/blog/post/big-data-analytics-and-energy-consumption/

[26] K.Zhou,C.Fu,andS.Yang,"Bigdatadrivensmartenergymanagement:Frombigdatatobiginsights",*Renew.Sustain.EnergyRev.,*vol. 56, pp. 215-225, 2016. [http://dx.doi.org/10.1016/j.rser.2015.11.050]

[27] Daniel Tunkelang talks about LinkedIn's data graph, *Available*.http://www.dataversity.net/daniel-tunkelang-talks-about-linkedins-data-graph/

[28] H.AlHajHassan,A.Pelov,andL.Nuaymi,"Integratingcellularnetworks,smartgrid,andrenewableenergy:Analysis,architecture,and challenges", *IEEE Access,* vol. 3, pp. 2755-2770, 2015. [http://dx.doi.org/10.1109/ACCESS.2015.2507781]

[29] W.Tushar,J.A.Zhang,C.Yuen,D.Smith,andU.Naveed,"Hassan,"ManagementofRenewableEnergyforASharedFacilityControllerin Smart Grid", *IEEE*

Access, vol. 4, pp. 4269-4281, 2016. [http://dx.doi.org/10.1109/ACCESS.2016.2592509]

[30] J.M.Andujar,andF.Segura,"Studyofarenewableenergysourcesbasedsmartgrid.Requirements,targetsandsolutions,inProc",In:*IEEE 3rdConferenceon Power Engineering andRenewable Energy,*, 2016.

[31] M.Ghofrani,M.Ghayekhloo,andR.Azimi,"Anovelsoftcomputingframeworkforsolarradiationforecasting", *Appl.SoftComput.,*vol.48, pp. 207-216, 2016. [http://dx.doi.org/10.1016/j.asoc.2016.07.022]

[32] R.Azimi,M.Ghofrani,andM.Ghayekhloo,"Ahybridwindpowerforecastingmodelbasedondataminingandwaveletsanalysis",*Energy Convers. Manage.,* vol. 127, pp. 208-225, 2016. [http://dx.doi.org/10.1016/j.enconman.2016.09.002]

[33] R.Azimi,M.Ghayekhloo,andM.Ghofrani,"Ahybridmethodbasedonanewclusteringtechniqueandmultilayerperceptronneuralnetworks for hourly solar radiation forecasting", *Energy Convers. Manage.,* vol. 118, pp. 331-344, 2016. [http://dx.doi.org/10.1016/j.enconman.2016.04.009]

[34] M.Ghayekhloo,M.Ghofrani,M.B.Menhaj,andR.Azimi,"Anovelclusteringapproachforshort-termsolarradiationforecasting",*Sol. Energy,* vol. 122, pp. 1371-1383, 2015. [http://dx.doi.org/10.1016/j.solener.2015.10.053]

[35] R.Billinton,andY.Gao,"MultistateWindEnergyConversionSystemModelsforAdequacyAssessmentofGenerating SystemsIncorporating Wind Energy", *IEEE Trans. Energ. Convers.,* vol. 23, no. 1, pp. 163-170, 2008. [http://dx.doi.org/10.1109/TEC.2006.882415]

[36] L.Kung,andH.Wang,"Arecommendersystemfortheoptimalcombinationofenergyresourceswithcost-benefitanalysis",*Proc.International Conference on Industrial Engineering and Operations Management,* pp. 1-10, 2015.

[37] M.H.Albadi,andE.F.ElSaadany,"Demandresponseinelectricitymarkets:Anoverview",*Proc.IEEEPowerEngineeringSocietyGeneral Meeting,* 2007.

[38] A.Arabali,M.Ghofrani,M.EtezadiAmoli,M.S.Fadali,andY.Baghzouz,"Genetic-algorithm-basedoptimizationapproachforenergymanagement", *IEEE Trans. Power Deliv.,* vol. 28, no. 1, pp. 162-170, 2012. [http://dx.doi.org/10.1109/TPWRD.2012.2219598]

[39] S.Maharjan,Q.Zhu,Y.Zhang,S.Gjessing,andT.Basar, "Demandresponsemanagementinthesmartgridinalargepopulationregime,'",*IEEE Trans. Smart Grid,* vol. 7, no. 1, pp. 189-199, 2016. [http://dx.doi.org/10.1109/TSG.2015.2431324]

[40] S. Maharjan, Q. Zhu, S. Gjessing, and T. Basar, "Dependable demand response management in the smart grid: A Stackelberg game approach", *IEEE Trans. Smart Grid,* vol. 4, no. 1, pp. 120-132, 2013. [http://dx.doi.org/10.1109/TSG.2012.2223766]

[41] F. Rahimi, and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm",*IEEE Trans. Smart Grid,* vol. 1, no. 1, pp.82-88, 2010. [http://dx.doi.org/10.1109/TSG.2010.2045906]

[42] S.Braithwait,andK.Eakin,"Theroleofdemandresponseinelectricpowermarketdesign",*EdisonElectricInstitute,Available:,*2002.http://www.eei.org

[43] Federal Energy Regulatory Commission, *FERC Order 719, Available:* .http://www.ferc.gov/whats-new/comm meet/2008/ 101608/E-1.pdf

[44] *AmericanRecoveryandReinvestmentActof2009,Available:,*2009.https://energy.gov/sites/prod/files/2014/12/f19/SGIG-SGDP-Highlights-October2014.pdf

[45] InternationalEnergyAgency,"GlobalEVOutlook2016:Beyondonemillionelectriccars",*Available*.https://www.iea.org/publications/freepublications/ publication/ Global_EV_Outlook_2016.pdf

[46] Z.Darabi,andM.Ferdowsi,"Aggregatedimpactofplug-inhybridelectricvehiclesonelectricitydemandprofile" ,*IEEETransactionsonSustainable Energy,* vol. 2, no. 4, pp. 501-508, 2011. [http://dx.doi.org/10.1109/TSTE.2011.2158123]

[47] L.P.Fernandez,T.G.S.Roman,R.Cossent,C.M.Domingo,andP.Frias,"AssessmentoftheImpactofplug-inelectricvehiclesondistribution networks", *IEEE Trans. Power Syst.,* vol. 26, no. 1, pp. 206-213, 2011. [http://dx.doi.org/10.1109/TPWRS.2010.2049133]

[48] Z.Jiang,L.Shalalfeh,andM.J.Beshir,"Impactofelectricvehicleinfrastructureonthecityofchatsworthdistributionsystem",*Proc.IEEEInternational Electric Vehicle Conference,* vol. vol. 1, pp. 17-19, 2014.

[49] W.Di,D.C.Aliprantis,andK.Gkritza,"ElectricEnergyandPowerConsumptionbyLight-DutyPlug-InElectricVehicles",*IEEETrans. Power Syst.,* vol. 26, no. 2, pp. 738-746, 2011. [http://dx.doi.org/10.1109/TPWRS.2010.2052375]

[50] R.Liu,L.Dow,andE.Liu,"AsurveyofPEVimpactsonelectricutilities",*Proc.IEEEPESInnovativeSmartGridTechnologies,*pp.1-8,2011.

[51] E. Akhavan-Rezai, M.F. Shaaban, E. El-Saadany, and F. Karray, "Online intelligent demand management of plug-in electric vehicles in future smart parking lots", *IEEE Syst. J.,* vol. 10, no. 2, pp. 483-494, 2016. [http://dx.doi.org/10.1109/JSYST.2014.2349357]

[52] M.F.Shaaban,A.H.Osman,andM.S.Hassan,"OptimalcoordinationforelectricvehiclesinsmartgridswithhighpenetrationofPVgeneration", *Proc. IEEE European Modelling Symposium,* 2016.

[53] S.Yoon,K.Park,andE.Hwang,"Connectedelectricvehiclesforflexiblevehicle-to-gridservices",*Proc.IEEEInternationalConferenceon Information Networking,* pp. 411-413, 2017.

[54] E.Akhavan-Rezaei,M.F.Shaaban,E.F.El-Saadany,andF.Karray,"NewEMStoIncorporateSmartParkingLotsintoDemandResponse", *IEEE Trans.*

Smart Grid.

[55] M.Ghofrani,A.Arabali,andM.Ghayekhloo,"Optimalcharging/dischargingofgridenabledelectricvehiclesfor predictabilityenhancement of PV generation", *Electr. Power Syst. Res.,* vol. 117, pp. 134-142, 2014. [http://dx.doi.org/10.1016/j.epsr.2014.08.007]

[56] M. Ghofrani, A. Arabali, M. Etezadi-Amoli, and M.S. Fadali, "Smart scheduling and cost-benefit analysis of grid-enabled electric vehicles for wind power integration", *IEEE Trans. Smart Grid,* vol. 5, no. 5, pp. 2306-2313, 2014. [http://dx.doi.org/10.1109/TSG.2014.2328976]

[57] W.Su,andM-Y.Chow, "Performance valuation of an EDA-basedlarge-scaleplug-in hybrid electricvehiclechargingalgorithm",*IEEETrans. Smart Grid,* vol. 3, no. 1, pp. 308-315, 2012. [http://dx.doi.org/10.1109/TSG.2011.2151888]

[58] X.Zhang,andS.Grijalva,"Anadvanceddatadrivenmodelforresidentialelectricvehiclechargingdemand",*Proc.IEEEPowerEnergySoc.General Meeting,* pp. 1-5, 2015.

[59] C.HLee,andC.H.Wu,"Anovelbigdatamodelingmethodforimprovingdrivingrangeestimationof EVs",*IEEEAccess,*vol.3,pp. 1980-1993, 2015. [http://dx.doi.org/10.1109/ACCESS.2015.2492923]

[60] M.F.Shaaban,M.Ismail,E.F.El-Saadany,andW.Zhuang,"RealTimePEVcharging/dischargingcoordinationinsmartdistributionsystems", *IEEE Trans. Smart Grid,* vol. 5, no. 4, pp. 1797-1807, 2014. [http://dx.doi.org/10.1109/TSG.2014.2311457]

[61] D.Dzung,M.Naedele,T.P.V.Hoff,andM.Crevatin,"Securityforindustrialcommunicationsystems",*Proc.IEEE,*vol.93,no.6,pp.1152-1177, 2005. [http://dx.doi.org/10.1109/JPROC.2005.849714]

[62] C.S.Lai,andL.L.Lai,"Applicationofbigdatainsmartgrid",*Proc.IEEEInternationalConferenceonSystems,Man,andCybernetics,*pp.665-670, 2015.

[63] P. McDaniel, and S. McLaughlin, "Security and privacy challenges in the smart grid", *IEEE Secur. Priv.,* vol. 7, no. 3, pp. 75-78, 2009.[http://dx.doi.org/10.1109/MSP.2009.76]

[64] F.Li,B.Luo,andP.Liu,"Secureinformationaggregationforsmartgridsusinghomomorphicencryption",*Proc. 1st IEEEInternationalConference Smart Grid Communication,* pp. 327-332, 2010.

[65] G.Kalogridis,C.Efthymiou,S.Z.Denic,T.A.Lewis,andR.Cepeda,"Privacyforsmartmeters:Towardsundetectableapplianceloadsignatures", *Proc. 1st IEEE International Conference Smart Grid Communication,* pp. 232-237, 2010.

[66] S.R.Rajagopalan,L.Sankar,S.Mohajer,andH.V.Poor, "Smartmeterprivacy:Autility-privacyframework",*Proc.IEEEinternationalConference Smart Grid Communication,* pp. 190-195, 2011.

[67] S. Ruj, and A.Pal, "Analyzing cascading failuresin smart gridsunder random and targetedattacks",*Proc. IEEE28th International Conference Advanced Information Networking and Applications,* pp. 226-233, 2014.

[68] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations", *IEEE Trans. Smart Grid,* vol. 2, no. 4, pp. 659-666, 2011. [http://dx.doi.org/10.1109/TSG.2011.2161892]

[69] Y.Yuan,Z.Li,andK.Ren,"Quantitativeanalysisofload redistributionattacksinpowersystems",*IEEETrans.ParallelDistrib.Syst.,*vol. 23, no. 9, pp. 1731-1738, 2012. [http://dx.doi.org/10.1109/TPDS.2012.58]

[70] O.Kosut,L.Jia,R.J.Thomas,andL.Tong,"Maliciousdataattacksonthesmartgrid",*IEEETrans.SmartGrid,*vol. 2,no.4,pp.645-658,2011. [http://dx.doi.org/10.1109/TSG.2011.2163807]

[71] L.Jia,R.Thomas,andL.Tong,"Maliciousdataattackon real-time electricity market", *Proc. IEEEInternationalConferenceonAcoustic,Speech, and Signal Processing,* pp. 5952-5955, 2011.

[72] H.Liu,H.Ning,Y.Zhang,andL.T.Yang,"Aggregated-proofsbasedprivacy-preservingauthenticationforV2Gnetworksinthesmartgrid", *IEEE Trans. Smart Grid,* vol. 3, no. 4, pp. 1722-1733, 2012. [http://dx.doi.org/10.1109/TSG.2012.2212730]

[73] H.Li,R.Lu,L.Zhou,B.Yang,andX.Shen,"Anefficient Merkle-Tree-BasedAuthenticationSchemeforSmartGrid",*IEEESyst.J.,*vol.8, no. 2, pp. 655-663, . [http://dx.doi.org/10.1109/JSYST.2013.2271537]

[74] M.M. Fouda, Z.M. Fadlullah, N. Kato, R. Lu, and X. Shen, "Towards a light-weight message authentication mechanism tailored for smart grid communications", *Proc. IEEE Conference on Computer Communications workshops,* pp. 1018-1023, 2011.

[75] B.Sasha, and D. Srivastava, "Data quality: The other face ofBig Data", *Proc. IEEE 30th International Conference on Data Engineering,* 2014

[76] W.Chen,K.Zhou,S.Yang,andC.Wu,"Dataqualityofelectricityconsumptiondatainasmartgridenvironment", *Renew.Sustain.Energy Rev.,* vol. 75, pp. 98-105, 2017. [http://dx.doi.org/10.1016/j.rser.2016.10.054]

[77] S.Bai,X.Li,Z.Qin,andL.Zhu,"TheapplicationandresearchofNoisedataacquisitionwithwirelessnetwork",*Proc.IEEEInternationalConference on Environmental Science and information,* vol. vol. 3, pp. 693-696, 2009.

[78] K.Hu,L.Li,andZ.Lu,"AcleaningmethodofnoisedatainRFIDdatastreams",*Proc.3rdInternationalConferenceonConsumerElectronics, Communications and Networks,* pp. 1-4, 2013.

[79] B.Fan,G.Zhang,andH.Li,"Multiplemodelsfusionforp

atternclassificationonnoisedata",*Proc.InternationalConferenceonSystemScience and Engineering,* pp. 64-68, 2012.

[80] E. Raham, and H.H. Do, "Data cleaning: Problems and current approaches",*Q. Bull. Comput. Soc. IEEE Tech. Comm. Data Eng.,*vol. 23, no. 4, pp. 3-13, 2000.

[81] A.Motro,P.Anokhin,andA.C.Acar,"Utility-basedresolutionofdatainconsistencies",*Proc.ACMInternationalWorkshoponInformation Quality in Information System,* vol. 76, pp. 35-43, 2004.

[82] J.R.Li,L.P.Khoo,andS.B.Tor,"RMINE:Aroughsetbaseddataminingprototypeforthereasoningofincompletedataincondition-based fault diagnosis", *J. Intell. Manuf.,* vol. 17, no. 1, pp. 163-176, 2006. [http://dx.doi.org/10.1007/s10845-005-5519-8]

[83] NationalInstituteofStandardsandTechnology,*NIST/SEMATECHe-HandbookofStatisticalMethods*.http://www.itl.nist.gov/div898/handbook/

[84] D.Luebbers,U.Grimmer,andM.Jarke,"Systematicdevelopmentofdatamining-baseddataqualitytools",*Proc.29thInternationalConference on Very Large Data bases,* vol. 29, pp. 548-559, 2003.

[85] J.Zhang,andH.Wang,"Anewpretreatmentapproachofeliminatingabnormaldataindiscretetimeseries",*Proc. IEEEInternationalGeoscience and Remote Sensing Symposium,* vol. 1, pp. 665-668, 2005.

[86] Y.Liu,P.Ning,andM.K.Reiter,"Falsedatainjectionattacksagainststateestimationinelectricpowergrids",*ACM Trans.Inf.Syst.Secur.,* vol. 14, no. 1, p. 13, 2009.

[87] F.Y.Okay,andS.Ozdemir,"Afogcomputingbasedsmartgridmodel",*Proc.2016InternationalSymposiumonNetworks,ComputersandCommunications (ISNCC),* pp. 1-6, 2016.

[88] T.Sirojan,T.Phung,andE.Ambikairajah,"Intelligentedgeanalyticsforloadidentificationinsmartmeters",*Proc.IEEEInnovativeSmart Grid Technologies - Asia (ISGT-Asia),* pp. 1-5, 2017.

[89] B.Stojkoska,andK.Trivodaliev,"Enablinginternetofthingsforsmarthomesthroughfogcomputing",*in25thTelecommunicationForum(TELFOR),* pp. 1-4, 2017.

[90] Y.Zhang,K.Liang,S.Zhang,andY.He,"ApplicationsofedgecomputinginPIoT",*Proc.IEEEConferenceonEnergyInternetandEnergy System Integration (EI2),* pp. 1-4, 2017.

[91] A.M.Annaswamy,M.Amin,A.M.Annaswamy,C.L.DeMarco,andT.Samad,IEEEVisionforSmartGridControls:2030andBeyond,Sept. 2013.

