Send Orders for Reprints to reprints@benthamscience.ae

BENTHAM OPEN

CrossMark

The Open Electrical & Electronic Engineering Journal

Content list available at: www.benthamopen.com/TOEEJ/

DOI: 10.2174/1874129001812010086, 2018, *12*, 86-97

RESEARCH ARTICLE

Survey of Big Data Role in Smart Grids: Definitions, Applications, Challenges, and Solutions

Mahmoud Ghofrani[*, 1], Andrew Steeble[1], Christopher Barrett[1] and Iman Daneshnia[2]

[1]*School of Science Technology engineering and mathematics (STEM), University of Washington, Bothell, Washington, USA*
[2]*Philips, USA*

Received: June 26, 2018 Revised: September 4, 2018 Accepted: September 26, 2018

Abstract:

Objective:

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.

Methods:

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.

Conclusion:

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.

1. INTRODUCTION

Smart Grid (SG) is an important research and development direction in the energy industry. It modifies the conventional power grid by integrating advanced communication and computing methods to improve the entire system control, efficiency, reliability, and safety [1]. Smart grid carries electricity and information between suppliers and consumers, which creates a bidirectional power and information flow system [2]. Many countries have recently adopted smart grid renovation plans [3]. As an example, the ENEL Telegestore project in Italy is the first commercial project utilizing smart grid technology which brings annual savings of approximately 500 million Euros [4, 5].

Smart grids offer several benefits to electric consumers, producers, and operators. SG improves the efficiency, dependability, sustainability, and economics of electric services [6]. Despite its numerous benefits, smart grid is mainly utilized in small regions [6]. There are several roadblocks preventing smart grids from being used in larger regions such as information gathering, storing, processing, and management [7 - 9].

Smart grid requires the capability for processing large volumes of real-time data. For example, in the past, utility companies read meters monthly, but with the Advanced Meter Infrastructure (AMI), meters report data themselves

* Address correspondence to this author at the School of Science Technology engineering and mathematics (STEM), University of Washington, Bothell, Washington, USA, Tel: 4253523224; E-mail: mrani@uw.edu

every 15-30 minutes [10]. As a result, the size of electric utility systems' data reached Terabytes (TBs) [10]. Another important requirement for smart grid is real-time information processing. This is because the entire system can be interrupted by any delay [3]. Such requirements highlight the importance of applying big data, whether it is machine learning, aggregation, or analytics, into smart grids.

This survey is arranged as follows. Definition of smart grid and energy big data is presented in Section 2. Current research and studies of big data application in smart grids are reviewed in section 3. Section 4 deals with big data challenges: security, quality, and processing location. The future of research in big data applications in smart grids is in section 5. Lastly, the survey is concluded in section 6.

2. OVERVIEW OF SMART GRID AND BIG DATA

2.1. Smart Grid

Smart grid is as a complex electric grid system, which includes subsystems such as smart meters, power generations, substations, distribution, transmission, networking systems, *etc* [11]. Smart grid is a modified traditional power system with six main components: network, user, hardware, software, servers, and data [2]. Because smart grid operates and depends on two-way communication flow, reliability and security of the communication methods are critical for proper information flow and management [12].

Smart grid has several benefits such as integrated renewable energy, bidirectional power and data flow, data-driven pricing, and power consumption tracking among others [13]. Recent developments in information, communication and computation brings the smart grid vision to reality. Smart grid also has unique capabilities to perform self-coordination, self-awareness, and self-healing actions [14].

Smart grid implementation involves challenges such as outdated technology, transmission and distribution losses, power quality, renewable energy incorporation, and security vulnerabilities [14]. For example, a smart grid system must meet security requirements to prevent any vulnerabilities in its communication, control, and computation sub-systems [15].

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

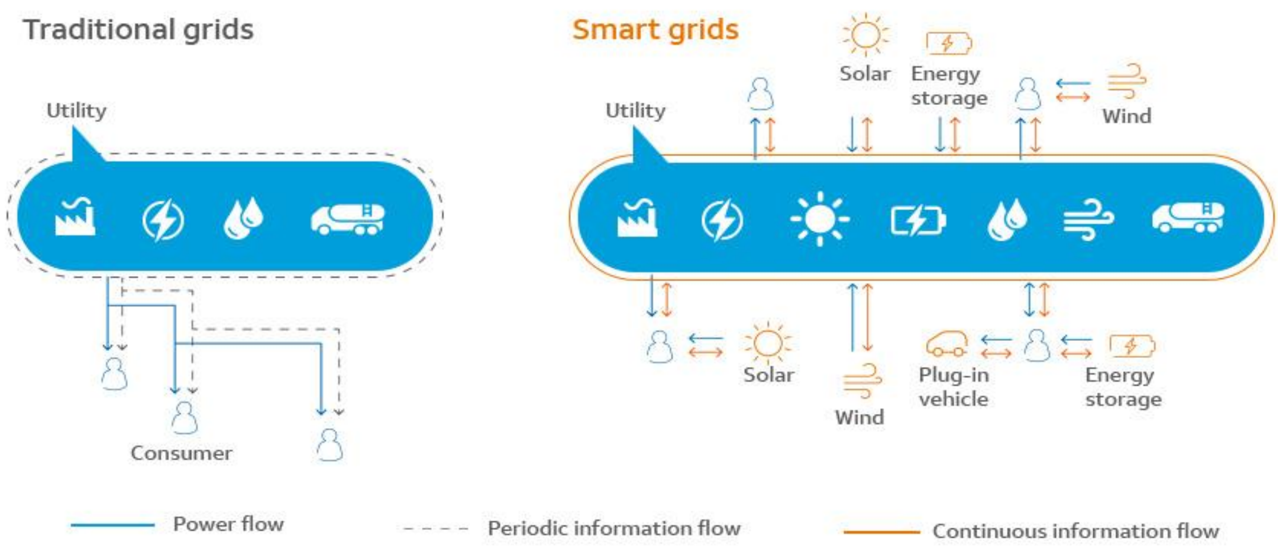

Fig. (1). Traditional grid *vs.* smart grid [16].

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

2.2. 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). In addition to the magnitude of these data sets, the lack of physical or temporal correlation between their elements renders them beyond the scope of traditional analysis methods [2]. Relevant state information from all entities of the grid (at all levels of generation and load) must be communicated with minimal latency to stakeholding respondents that depend on this information as operating 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].

2.2.1. Volume

Utility companies are replacing traditional meters with smart meters, which generate large amount of data [24]. In a large utility company with one million smart meters, if every 15-minute data is collected, 35.04 billion records with volume of 2920 TBs data will be generated [25]. The drastic increase in electric power systems data volume introduces several challenges which will be further discussed in section 4.

2.2.2. Velocity

Velocity in an energy big data context refers to the speed of storing, processing and analyzing the data. Unlike traditional data intelligence devices, the storage and processing of energy big data require fast and real-time capability [26]. Streaming data processing is employed allowing relational data queries to be continuously updated. High velocity data is analyzed in terms of stream-to-relation, relation-to-relation or relation-to-stream queries [22]. Common querying languages used include Cassandra Query Language (CQL), Stream Processing Language, Spark Streaming, Storm, and Fink Framework and Apache Drill [2, 17, 22]. The result is the real-time interaction with data suffering nominal latency. Ad-hoc queries can be processed in PetaByte (PB) magnitudes within a few seconds [2]. Thus, the speed of data processing can be reduced to a few seconds allowing the energy system to make fast and prompt decisions, such as fault detection *via* PMUs and grid self-healing responses [18].

2.2.3. Variety

There are typically three different data types in smart energy systems: Structured, semi-structured, and unstructured. The degree of structure is defined by the format of the content presented: records with values classified by distinct categories (*e.g.* call records from a telecom company) are considered to be structured while graphical data deriving a relationship from the plot of variables is considered semi-structured. A completely free-form text entry such as a Twitter post or online review is unstructured data [22]. In a smart grid, energy consumption data constitutes the structured data; communication data between customers and vendor devices form the semi-structured data; and energy usage email or SMS notifications are examples of unstructured data [24].

2.2.4. Value

Value is a result of the first three V's with some computation involved. This is why Monica Rogati says, "More data beats clever algorithms, but better data beats more data" [27]. Energy big data has value once passed through computation to support business decisions or help customers [24]. For service providers, value renders into creating competitive marketing strategies by analyzing the customer energy consumption patterns. Customers could also benefit from energy savings, transparency in their energy usage and enhanced operational efficiency [24]. Value also depends on the eye of the beholder. A grid operator would not care about the temperature of a single house or how optimized the traffic lights are between each other. This is why it is so important to include Value in the description of what constitutes big data.

3. RESEARCHES RELATED TO BIG DATA APPLICATIONS IN SMART GRID

Three main categories are identified for smart grid big data applications: Renewable Energy (RE), Demand Response (DR), and Electric Vehicles (EV) [2].

3.1. Renewable Energy

With increasing integration of renewable energy sources in power systems, data management of current energy grids becomes a complex task, which should be addressed by big data analytics [28, 29]. For example, historical weather data and GPS data can be used to improve forecasting of renewable energy power generation, which ultimately enhances the grid energy efficiency [30]. Data mining and processing have been employed to extract features of time series data for more accurate forecasting of intermittent renewable resources such as wind and solar [31 - 34].

A Danish power company improved the efficiency of their wind integration by optimizing turbine placement after analyzing the weather reports, tidal conditions, and satellite images [35]. Another study presented a system that allows for an optimum mixture of renewable energy resources while meeting the cost-benefit tradeoffs [36].

3.2. 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].

3.3. 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 networks.	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, and sales to both local users and the grid.
	[31]	An enhanced K-means algorithm, named Time Series Clustering (T.S.C) K-means, combined with Multilayer Perceptron Neural Networks (MLPNN) for solar radiation forecasting.	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 map (SOM), fuzzy C-means (FCM), and K-Medoids. Solar radiation datasets from different US states are used to evaluate the accuracy performance of the developed hybrid forecasting method and its comparison with state-of-the-art forecasting techniques.
	[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 (HANTS), and MLPNN for wind power forecasting.	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 hybrid method with other well-established forecasting models including Persistence, New Reference (NR), Adaptive Wavelet Neural Network (AWNN), and Phase Space Reconstruction (PSR) are also performed.
	[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 hybrid forecasting and benchmark forecasting models.
	[34]	A novel Game Theoretic Self-organizing Map (GTSOM), combined with Neural gas (NG) and Competitive Hebbian Learning (CHL), DWT and Bayesian Neural Network (BNN) for solar radiation forecasting.	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.
Electric vehicle	[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 data.
	[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 implemented on the test system and compared.
	[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.

4. SMART GRID BIG DATA CHALLENGES AND PROPOSED SOLUTIONS

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

4.1. 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.

4.1.1. 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].

4.1.2. 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].

4.1.3. 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].

4.2. 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].

4.2.1. 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].

4.2.2. 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].

4.2.3. 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.

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

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.

(Table 2) contd.....

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.
	[90]	Applying edge computing in the Power Internet of Things (PIOT), such as in monitoring transmission lines, managing smart homes, *etc*.	Bandwidth issues, locational solutions

5. 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].

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.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none

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, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart 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, and S. Rogai, "The Telegestore project in ENEL's metering system", *Proc. 18th International Conference and Exhibition 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, and B. Bakmaz, "Smart grid communications architecture: A survey and challenges", *Proc. 11th International Conference on Applied 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, and M. Sooriyabandara, "Smart grid communications: Overview of research challenges, 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] How big data will make us more energy efficient, *Available*. https://www.weforum.org/agenda/2014/05/ big-data-will-make-us-energy-efficient/

[11] D.B. Rawat, and C. Bajracharya, "Cyber security for smart grid systems: Status, challenges and perspectives", *Proc. SoutheastCon 2015,* pp. 1-6, 2015.

[12] J.N. Bharothu, M. Sridhar, and R.S. Rao, "A literature survey report on Smart Grid technologies", *Proc. 2014 International Conference on Smart Electric Grid (ISEG),* pp. 1-8, 2014.

[13] N.B.M. Isa, T.C. Wei, and A.H.M. Yatim, "Smart grid technology: Communications, power electronics and control system", *Proc. 2015 International Conference on Sustainable Energy Engineering and Application (ICSEEA),* pp. 10-14, 2015. [http://dx.doi.org/10.1109/ICSEEA.2015.7380737]

[14] C. Vineetha, and C. Babu, "Smart grid challenges, issues and solutions", *Proc. 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG),* pp. 1-4, 2014.

[15] R. Mahmud, R. Vallakati, A. Mukherjee, P. Ranganathan, and A. Nejadpak, "A survey on smart grid metering infrastructures: Threats and solutions", *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&T Smart Grid Technology Solutions, *Available*. https://www.business.att.com/enterprise/service/ internet-of-things/ smart-cities/iot-smart-grid/

[17] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, "Big data issues in smart grid system", *Proc. 5th International Conference on Renewable Energy Research and Applications,* pp. 1007-1012, 2016.

[18] M. Kezunovic, L. Xie, and S. Grijalva, The role of big data in improving power system operation and protection*Proc. Bulk Power System Dynamic and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP)*, 2013.

[19] M. Hardy, "Big data analytics - what it means to the audit community- part 1", *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, and Y. Li, "Emergent technologies in big data sensing: A survey", *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, and M. Mohania, "Cloud computing and big data analytics: What is new from databases perspective?", *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, and S. Yang, "Understanding household energy consumption behavior: The contribution of energy big data analytics", *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, and S. Yang, "Big data driven smart energy management: From big data to big insights", *Renew. Sustain. Energy Rev.,* 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. Al Haj Hassan, A. Pelov, and L. Nuaymi, "Integrating cellular networks, smart grid, and renewable energy: 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, and U. Naveed, "Hassan, "Management of Renewable Energy for A Shared Facility Controller in Smart Grid", *IEEE Access,* vol. 4, pp. 4269-4281, 2016. [http://dx.doi.org/10.1109/ACCESS.2016.2592509]

[30] J. M. Andujar, and F. Segura, "Study of a renewable energy sources-based smart grid. Requirements, targets and solutions, in Proc", In: *IEEE 3rd Conference on Power Engineering and Renewable Energy,*, 2016.

[31] M. Ghofrani, M. Ghayekhloo, and R. Azimi, "A novel soft computing framework for solar radiation forecasting", *Appl. Soft Comput.,* vol. 48, pp. 207-216, 2016.
[http://dx.doi.org/10.1016/j.asoc.2016.07.022]

[32] R. Azimi, M. Ghofrani, and M. Ghayekhloo, "A hybrid wind power forecasting model based on data mining and wavelets analysis", *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, and M. Ghofrani, "A hybrid method based on a new clustering technique and multilayer perceptron neural networks 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, and R. Azimi, "A novel clustering approach for short-term solar radiation forecasting", *Sol. Energy,* vol. 122, pp. 1371-1383, 2015.
[http://dx.doi.org/10.1016/j.solener.2015.10.053]

[35] R. Billinton, and Y. Gao, "Multistate Wind Energy Conversion System Models for Adequacy Assessment of Generating Systems Incorporating 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, and H. Wang, "A recommender system for the optimal combination of energy resources with cost-benefit analysis", *Proc. International Conference on Industrial Engineering and Operations Management,* pp. 1-10, 2015.

[37] M.H. Albadi, and E.F. El-Saadany, "Demand response in electricity markets: An overview", *Proc. IEEE Power Engineering Society General Meeting,* 2007.

[38] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, M.S. Fadali, and Y. Baghzouz, "Genetic-algorithm-based optimization approach for energy management", *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, and T. Basar, "Demand response management in the smart grid in a large population regime,'", *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, and K. Eakin, "The role of demand response in electric power market design", *Edison Electric Institute, 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] *American Recovery and Reinvestment Act of 2009, Available:,* 2009.https:// energy.gov/ sites/prod/ files/2014/ 12/f19/ SGIG-SGDP-Highlights-October2014.pdf

[45] International Energy Agency, "Global EV Outlook 2016: Beyond one million electric cars", *Available*.https:// www.iea.org/ publications/ freepublications/ publication/ Global_EV_Outlook_2016.pdf

[46] Z. Darabi, and M. Ferdowsi, "Aggregated impact of plug-in hybrid electric vehicles on electricity demand profile", *IEEE Transactions on Sustainable 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, and P. Frias, "Assessment of the Impact of plug-in electric vehicles on distribution 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, and M.J. Beshir, "Impact of electric vehicle infrastructure on the city of chatsworth distribution system", *Proc. IEEE International Electric Vehicle Conference,* vol. vol. 1, pp. 17-19, 2014.

[49] W. Di, D.C. Aliprantis, and K. Gkritza, "Electric Energy and Power Consumption by Light-Duty Plug-In Electric Vehicles", *IEEE Trans. Power Syst.,* vol. 26, no. 2, pp. 738-746, 2011.
[http://dx.doi.org/10.1109/TPWRS.2010.2052375]

[50] R. Liu, L. Dow, and E. Liu, "A survey of PEV impacts on electric utilities", *Proc. IEEE PES Innovative Smart Grid Technologies,* 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, and M.S. Hassan, "Optimal coordination for electric vehicles in smart grids with high penetration of PV generation", *Proc. IEEE European Modelling Symposium,* 2016.

[53] S. Yoon, K. Park, and E. Hwang, "Connected electric vehicles for flexible vehicle-to-grid services", *Proc. IEEE International Conference on Information Networking,* pp. 411-413, 2017.

[54] E. Akhavan-Rezaei, M.F. Shaaban, E.F. El-Saadany, and F. Karray, "New EMS to Incorporate Smart Parking Lots into Demand Response", *IEEE Trans. Smart Grid*.

[55] M. Ghofrani, A. Arabali, and M. Ghayekhloo, "Optimal charging/discharging of grid-enabled electric vehicles for predictability enhancement 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, and M-Y. Chow, "Performance evaluation of an EDA-based large-scale plug-in hybrid electric vehicle charging algorithm", *IEEE Trans. Smart Grid,* vol. 3, no. 1, pp. 308-315, 2012.
[http://dx.doi.org/10.1109/TSG.2011.2151888]

[58] X. Zhang, and S. Grijalva, "An advanced data driven model for residential electric vehicle charging demand", *Proc. IEEE Power Energy Soc. General Meeting,* pp. 1-5, 2015.

[59] C.H Lee, and C.H. Wu, "A novel big data modeling method for improving driving range estimation of EVs", *IEEE Access,* 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, and W. Zhuang, "Real Time PEV charging/discharging coordination in smart distribution systems", *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, and M. Crevatin, "Security for industrial communication systems", *Proc. IEEE,* vol. 93, no. 6, pp. 1152-1177, 2005.
[http://dx.doi.org/10.1109/JPROC.2005.849714]

[62] C.S. Lai, and L.L. Lai, "Application of big data in smart grid", *Proc. IEEE International Conference on Systems, Man, and Cybernetics,* 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, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption", *Proc. 1st IEEE International Conference Smart Grid Communication,* pp. 327-332, 2010.

[65] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures", *Proc. 1st IEEE International Conference Smart Grid Communication,* pp. 232-237, 2010.

[66] S.R. Rajagopalan, L. Sankar, S. Mohajer, and H.V. Poor, "Smart meter privacy: A utility-privacy framework", *Proc. IEEE international Conference Smart Grid Communication,* pp. 190-195, 2011.

[67] S. Ruj, and A. Pal, "Analyzing cascading failures in smart grids under random and targeted attacks", *Proc. IEEE 28th 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, and K. Ren, "Quantitative analysis of load redistribution attacks in power systems", *IEEE Trans. Parallel Distrib. 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, and L. Tong, "Malicious data attacks on the smart grid", *IEEE Trans. Smart Grid,* vol. 2, no. 4, pp. 645-658, 2011.
[http://dx.doi.org/10.1109/TSG.2011.2163807]

[71] L. Jia, R. Thomas, and L. Tong, "Malicious data attack on real-time electricity market", *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing,* pp. 5952-5955, 2011.

[72] H. Liu, H. Ning, Y. Zhang, and L.T. Yang, "Aggregated-proofs based privacy-preserving authentication for V2G networks in the smart grid", *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, and X. Shen, "An efficient Merkle-Tree-Based Authentication Scheme for Smart Grid", *IEEE Syst. 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 of Big Data", *Proc. IEEE 30th International Conference on Data Engineering,* 2014

[76] W. Chen, K. Zhou, S. Yang, and C. Wu, "Data quality of electricity consumption data in a smart grid environment", *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, and L. Zhu, "The application and research of Noise data acquisition with wireless network", *Proc. IEEE International Conference on Environmental Science and information,* vol. vol. 3, pp. 693-696, 2009.

[78] K. Hu, L. Li, and Z. Lu, "A cleaning method of noise data in RFID data streams", *Proc. 3rd International Conference on Consumer Electronics, Communications and Networks,* pp. 1-4, 2013.

[79] B. Fan, G. Zhang, and H. Li, "Multiple models fusion for pattern classification on noise data", *Proc. International Conference on System Science 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, and A.C. Acar, "Utility-based resolution of data inconsistencies", *Proc. ACM International Workshop on Information Quality in Information System,* vol. 76, pp. 35-43, 2004.

[82] J.R. Li, L.P. Khoo, and S.B. Tor, "RMINE: A rough set based data mining prototype for the reasoning of incomplete data in condition-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] National Institute of Standards and Technology, *NIST/SEMATECH e-Handbook of Statistical Methods*.http:// www.itl.nist.gov/ div898/ handbook/

[84] D. Luebbers, U. Grimmer, and M. Jarke, "Systematic development of data mining-based data quality tools", *Proc. 29th International Conference on Very Large Data bases,* vol. 29, pp. 548-559, 2003.

[85] J. Zhang, and H. Wang, "A new pretreatment approach of eliminating abnormal data in discrete time series", *Proc. IEEE International Geoscience and Remote Sensing Symposium,* vol. 1, pp. 665-668, 2005.

[86] Y. Liu, P. Ning, and M.K. Reiter, "False data injection attacks against state estimation in electric power grids", *ACM Trans. Inf. Syst. Secur.,* vol. 14, no. 1, p. 13, 2009.

[87] F.Y. Okay, and S. Ozdemir, "A fog computing based smart grid model", *Proc. 2016 International Symposium on Networks, Computers and Communications (ISNCC),* pp. 1-6, 2016.

[88] T. Sirojan, T. Phung, and E. Ambikairajah, "Intelligent edge analytics for load identification in smart meters", *Proc. IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia),* pp. 1-5, 2017.

[89] B. Stojkoska, and K. Trivodaliev, "Enabling internet of things for smart homes through fog computing", *in 25th Telecommunication Forum (TELFOR),* pp. 1-4, 2017.

[90] Y. Zhang, K. Liang, S. Zhang, and Y. He, "Applications of edge computing in PIoT", *Proc. IEEE Conference on Energy Internet and Energy System Integration (EI2),* pp. 1-4, 2017.

[91] A. M. Annaswamy, M. Amin, A. M. Annaswamy, C. L. DeMarco, and T. Samad, IEEE Vision for Smart Grid Controls: 2030 and Beyond, Sept. 2013.