IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 18, 2018, accepted January 8, 2019, date of publication January 24, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894819

Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review

EKLAS HOSSAIN[1], **(Senior Member, IEEE), IMTIAJ KHAN**[2], **FUAD UN-NOOR**[3], **SARDER SHAZALI SIKANDER**[4], **(Senior Member, IEEE), AND MD. SAMIUL HAQUE SUNNY**[3]
[1]Department of Electrical Engineering and Renewable Energy, Oregon Tech, Klamath Falls, OR 97601, USA
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[3]Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh
[4]Department of Electrical Engineering, National University of Sciences and Technology, Islamabad, Pakistan

Corresponding author: Eklas Hossain (eklas.hossain@oit.edu)

ABSTRACT This paper conducts a comprehensive study on the application of big data and machine learning in the electrical power grid introduced through the emergence of the next-generation power system—the smart grid (SG). Connectivity lies at the core of this new grid infrastructure, which is provided by the Internet of Things (IoT). This connectivity, and constant communication required in this system, also introduced a massive data volume that demands techniques far superior to conventional methods for proper analysis and decision-making. The IoT-integrated SG system can provide efficient load forecasting and data acquisition technique along with cost-effectiveness. Big data analysis and machine learning techniques are essential to reaping these benefits. In the complex connected system of SG, cyber security becomes a critical issue; IoT devices and their data turning into major targets of attacks. Such security concerns and their solutions are also included in this paper. Key information obtained through literature review is tabulated in the corresponding sections to provide a clear synopsis; and the findings of this rigorous review are listed to give a concise picture of this area of study and promising future fields of academic and industrial research, with current limitations with viable solutions along with their effectiveness.

INDEX TERMS Big data analysis, cyber security, IoT, machine learning, smart grid.

LIST OF ABBREVIATIONS

IoT	Internet of things
ML	Machine learning
SG	Smart grid
DER	Distributed energy resources
DEM	Dynamic energy management
CPL	Constant power load
MOSFET	Metal-oxide-semiconductor field-effect transistor
HDFS	Hadoop file system
LPRF	Low power radio frequency
OFDM	Orthogonal frequency-division multiplexing
HAN	Home area network
NAN	Neighbor area network
WAN	Wide area network
HG	Home gateway
ESP	Energy service provider
PDC	Phasor data concentrator
PMU	Phasor measurement unit
TCP/IP	Transmission control protocol/Internet protocol
WAMS	Wide area measurement system
UDP	User datagram protocol
NRECA	National rural electric cooperative association
NLP	Natural language processing
PCA	Principal component analysis
K-NN	k-nearest neighbors
ANN	Artificial neural network
CFD	Computational fluid dynamics
CxO	Corporate officer
BDC	Billing and debt collection
SCADA	Supervisory control and data acquisition
PLC	Programmable logic controller
EMS	Energy management system
DMS	Distribution management system

CPS	Cyber physical system
HEMS	Home energy management system
WLAN	Wireless LAN
MPO	Meter point operator
DoS	Denial of service
DDoS	Distributed denial of service
FDIA	False data injection attack
MITM	Man-in-the-middle attack
RF	Radio frequency
NERC	North American Reliability Corporation
API	Application program interface
MSE	Mean squared error
HMM	Hidden Markov model
PID	Proportional integral derivative
RMT	Random matrix theory
DNN	Deep neural network
SNN	Shallow neural network
DSHW	Double seasonal Holt-Winters
GPU	Graphics processing unit
UCSD	University of California San Diego
PV	Photovoltaic
MV/LV	Medium voltage/Low voltage
MPPT	Maximum power point tracking
AMI	Advanced metering infrastructure
ISMS	Information security management system
SoGP	Standard of good practice
IACS	Industrial automation and control systems
AES	Advanced encryption standard
TDEA	Triple-data encryption algorithm
DSS	Digital signature standard
DSA	Digital signature algorithm
RSA	Rivest, Shamir, and Adleman
ECDSA	Elliptic curve digital signature algorithm
SHA	Secure hash algorithm
CMAC	Cipher-based message authentication code
CCM	Cipher block chaining-message authentication code
GCM	Galois/counter mode
GMAC	Galois message authentication code
HMAC	Hash-based message authentication code
CFB	Cipher feedback
CBC	Cipher-block chaining
ECB	Electronic codebook
XTS	XEX-based tweaked-codebook mode with ciphertext stealing
TDES	Triple data encryption standard
TECB	TDEA electronic codebook
TCBC	TDEA cipher-block chaining
TCFB	TDEA cipher feedback
TOFB	TDEA output feedback
CTR	Counter-mode
IaaS	Infrastructure-as-a-service
SaaS	Software-as-a-service
PaaS	Platform-as-a-service
DaaS	Data-as-a-service
CVM	Core vector machine
μPMU	Micro phasor measurement unit
PPMV	Power plant model validation tool
FRAT	Flight risk assessment tool
SVM	Support vector machine
PQ	Power quality
SEMMA	Sample, Explore, Modify, Model, Assess
CRISP-DM	Cross Industry Standard Process for Data Mining
ELM	Extreme learning machine
ANFIS	Adaptive neuro-fuzzy inference system
RVM	Relevance vector machine
EMD	Empirical mode decomposition
GMR	Generalized mapping regressor
CRO	Coral reef optimization
iSSO	Improved simplified swarm optimization
HAP	Hybrid swarm technique
PSO	Particle swarm optimization
ACO	Ant colony optimization
TPSD	Three-phase signal decomposition
WRELM	Weighted regularized extreme learning machine
SSA	Seasonal separation algorithm
FEEMD	Fast ensemble empirical mode decomposition
VMD	Variation mode decomposition
PACF	Partial autocorrelation function
MLP	Multilayer perceptron
LMS	Least median square
CRO-ELM	Coral reef optimization – extreme learning machine
GPR	Gaussian process regression
LVQ	Learning vector quantization
SOM	Self-organizing map
SVR	Support vector regression

I. INTRODUCTION

The electrical power system is poised to move towards the next-generation smart grid (SG) system, and thus this topic has acclaimed significant attention in the research community [1]–[7]. SG is the integration of information and digital communication technologies with power grid systems to enable bi-directional communication and power flow that can enhance security, reliability, and efficiency of the power system [8]–[10]. Smart grid solutions aim at calculation of optimum generation-transmission-distribution pattern and storing power system data. For the growing concern about environment along with efficient generation and distribution, distributed energy resources (DER) with smart microgrid can be a potential solution [11]. It can be said that distributed smart microgrid can bring additional benefits for global power system planning [12]. In other words, SG is the integration of technologies, systems and processes to make power grid intelligent and automated [13]. Fig. 1 shows basic constructions of conventional grid and smart grid to demonstrate their dissimilarities. Unlike the unidirectional power

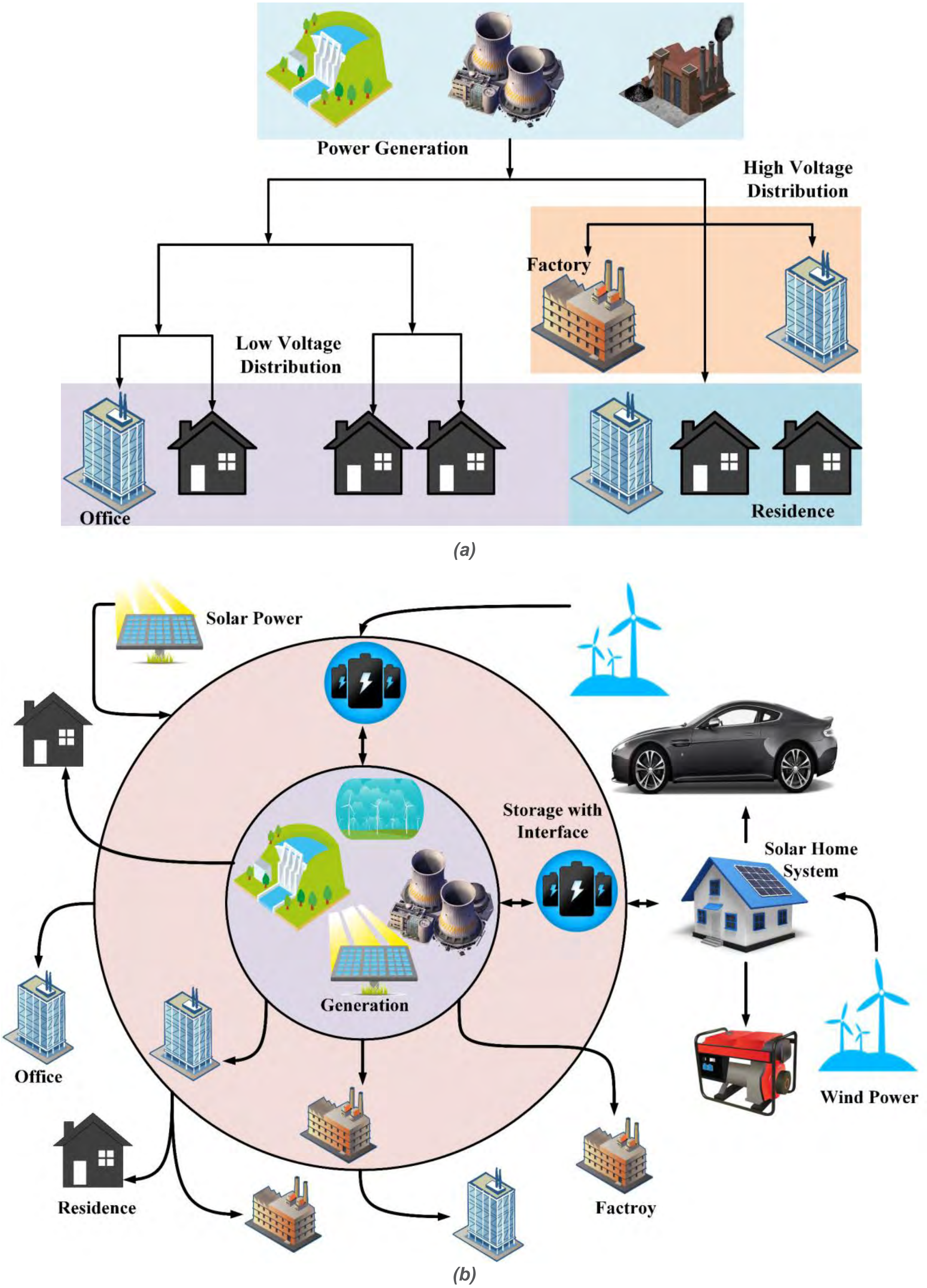


FIGURE 1. Utility grids: (a) conventional grid (b) smart grid. In the conventional system power flows from in one direction only; but for smart grid, there is no strict structure. Generation can occur at the consumer side too, such as the wind and the solar farms at the outer periphery of the topology. Power flow can also be bidirectional, demonstrated by the energy storages and the house in this illustration.

flow in the conventional system, power and information flow between the generation and distribution sides in a bidirectional manner.

Constant connectivity and communication is one of the core components of smart grid, and that requires devices equipped with such capabilities. The network created by such devices, connected to other nodes of the system through the internet, are called the ''internet of things (IoT)''. In the internet of things each object has its own identity in the digital world. Everything is connected through a complex network. IoT comprises of smart objects which possess self-awareness, interaction with the environment and data processing. Smart devices are capable to communicate with other such devices in this system [14]. Most common smart devices employed in the grid, such as the smart meter, falls into this category. These devices provide the detailed data required for accurate information and automated decision support which give the smart grid the unique capabilities it demonstrates over the legacy system. All this data need to be handled in real time, and stored to use historical data to create decisions based on certain cases. Various research works have been conducted with data obtained from intelligent devices in substations, feeders, and various databases [15]. Information sources can be market data, lighting data, power system data, geographical data, weather data etc. [16]. Optimization from generation to distribution requires reliable, accurate and efficient prediction model for electric energy consumption. For example, energy consumption data (kWh) from 100,000's of customer smart meters at 15 minutes sampling intervals shows that ensuring the quality of the collected data poses a unique challenge for prediction models and evaluation of their efficacy for SG [17]. There are several factors which require to be predicted, such as: renewable generation, power purchases from energy markets, 24 hour planning of load distribution etc. [18], [19]. These factors are the part and parcel of SG sustainability and security [20]. Predictability of electricity consumption has been studied with dynamic demand response in [21]. High volume of data from SG increases the complexity of data analysis. Dynamic energy management (DEM) is required for processing this huge amount of data for power flow optimization, system monitoring, real-time operation, and production planning [22], [23]. Data of such magnitude, which cannot be processed through traditional processes, is termed as ''big data'', and it has also become a centerpiece of current research. Researches on big data-based power generation, optimization and forecasting techniques are extended to the renewable energy based system such as wind energy system [23], [24]. A portion of the data produced in SG contains individual users' confidential information. This type of data is required to be protected under legal regulations [25], [26]. Moreover, this data contains classified and sensitive information of an organization or central grid of a country. Manipulation of such data can affect the safe operation of the grid. Therefore security and privacy is a very important issue [27]. An IoT integrated SG is a cyber-physical system [28], which makes it prone to cyberattacks. Therefore, adequate protection systems are required to ensure proper operation of the smart grid, safekeeping of data, and thwart any attack aimed at the power system. Machine learning is an attractive solution processing big data, and implementing effective security solutions.

This paper presents a concise picture of the electricity grid's transition towards the smart grid, the ensuing rise in IoT usage, and the challenges this new system brings forward. The most obvious trials are of course the handling of the huge amount of data in this connected system, their proper analysis and safety, as well as protecting this new power grid from attacks generated in both physical and cyber dimensions. This work can act as a base for future academic and industrial researchers, while pointing out the current limitations with possible solutions along with their effectiveness.

The rest of the paper is organized as follows: a short history of the power grid from its inception is presented in section II. Discussion on IoT components, its applications and issues is carried out in section III. Section IV focuses on big data, and the role of big data analysis in smart grid. Section V puts forward machine learning as a method capable of handling the big data generated in the IoT-based smart grid, and highlights its capabilities in renewable energy forecasting. Emerging security threats to the smart grid, its data, and devices are discussed in section VI, including protection and threat-detection techniques. The excerpt of this detailed study is presented in section VII, with future research directions outlined. Finally, the conclusions are drawn in section VIII.

II. CHANGES IN THE CENTURY-OLD GRID

In the early days of electric power systems, AC and DC contended to become the industry standard. The AC system prevailed and have been in use ever since. The reason of AC's dominance is its ability to use transformers for changing the voltage level, and enabling the transmission of high voltage electricity which reduces loss. The first demonstration of the AC transmission system took place in 1886, at Great Barrington, Massachusetts, USA, by William Stanley and George Westinghouse [29]. Westinghouse later formed the Westinghouse Electric Company that went on supply AC power to Buffalo, New York from the Adams Power Plant at the Niagra Falls in 1896. Thus the dominance of AC system is established, and the worldwide power grid adopted this technology as electrification expanded massively over the next century. Now, in the 21st century, technology has advanced astronomically as compared to the late 1800s; however, the grid system in the world largely resembles the century-old system that initiated the process of electrification. The advanced technologies that emerged in the power sector include power electronics, renewable energy sources, distributed generation, advanced monitoring and communication system etc. The legacy grid was not designed to accommodate these devices, and thus they create significant problems when integrated with the existing grid infrastructure. For example, power electronics based loads act differently than the generally perceived loads of resistive, capacitive, and inductive properties; electronics devices exhibit constant power load (CPL) properties that cause significant system-instabilities [30]. Distributed generation causes bi-directional power flow and thus contradicts the historically unidirectional flow of the grid. Renewable sources often generate intermittent DC power (e.g. solar energy) opposing the predominant AC systems. The renewable sources of AC (e.g. wind energy) are highly varying as well. All these which makes integrating renewable sources in the existing grid a huge challenge. However, even though these next-generation systems disrupt the grid architecture in place, adopting these technologies is the way to move forward, not the other way round. Therefore, the current time marks the transition period for the electricity grid – a metamorphosis that will supplant the archaic system with an architecture well-capable to accommodate the advanced concepts and tools. This next-generation system is called the "smart grid". The grid evolution timeline in presented in Table 1.

TABLE 1. Significant events in the evolution of electricity grid.

Year	Significant events
1800	Alessandro Volta made an electric battery to provide a steady supply
1831	Michael Faraday discovered dynamo
1878	Invention of incandescent lamp by Thomas Edison (in America) and Joseph Swan (in Britain)
1882	DC street lamps appear in New York
1886	AC system spread rapidly in the following years
1888	The induction motor got patented independently by Nikola Tesla (in USA) and Galileo Ferraris (in Italy)
1896	The Adams Power Plant supplied AC power to Buffalo, New York
1976	Power MOSFET appears as a commercial product
2017	Wind power becomes economically feasible

III. APPLICATION OF INTERNET OF THINGS (IOT) IN DISTRIBUTED POWER SYSTEM

The underpinnings those make the smart grid do so many things that the legacy grid is incapable of are a lot of connected devices, which are capable exchanging information, and receive commands to act in a certain way. This extensive communication is made possible by the internet, and all these devices are connected to their respective networks. Devices connected to the internet are currently part and parcels of the daily life, and more and more of such devices are emerging every day. An example of such devices can be smart thermostats. Such devices, which use the internet to stay connected to resources located elsewhere physically, and carry out their tasks through the resulting exchange, are termed as IoT devices. IoT stands for "internet of things", which can be defined as the interrelated system that links up such devices, and facilitate data transfer without any human intervention. According to Gubbi *et al.* [31], IoT is an interconnection of sensing and actuating devices providing the ability to share information across platforms through an unified framework, developing a common operating picture for enabling innovative applications. This is achieved by seamless ubiquitous sensing, data analytics and information representation with cloud computing as the unifying framework. Each of those objects has its own embedded computing system which enables it to be identified and to be interconnected with each other. The IoT architecture is shown in fig. 2. IoT will consist of more than 30 billion objects by 2020 [32]. The astronomic increase in number of IoT devices is visualized in fig. 3. From a mere 13 billion in 2015, their predicted population reaching 30 billion and beyond in a timespan of five years perfectly demonstrates the current trend of IoT application. These devices are able to operate with a less amount of external intervention and are capable of responding to the environment spontaneously.

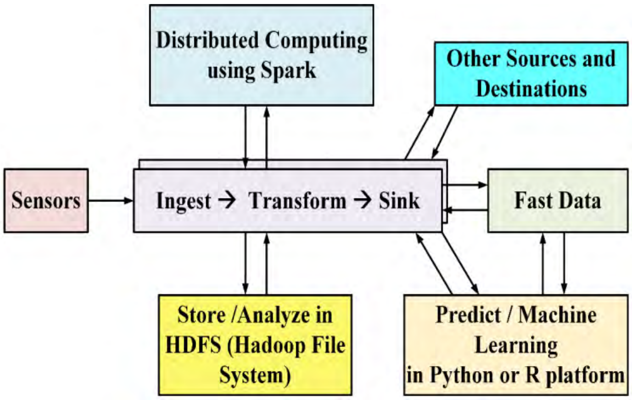

Distributed Computing using Spark
Other Sources and Destinations
Sensors
Ingest → Transform → Sink
Fast Data
Store /Analyze in HDFS (Hadoop File System)
Predict / Machine Learning in Python or R platform

FIGURE 2. IoT architecture. Data collected by sensors can be sent to different systems which use various software platforms to carry out intended tasks [38].

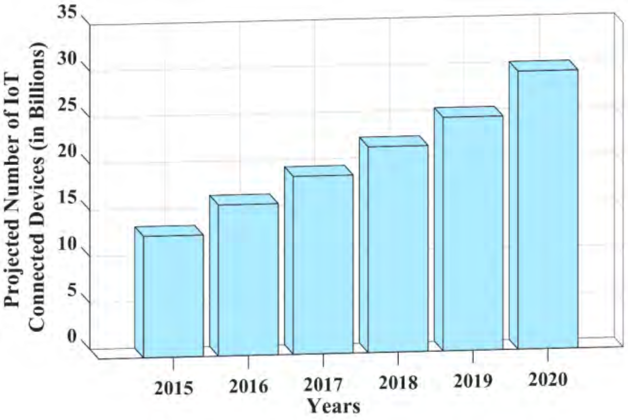


FIGURE 3. Predicted number of IoT devices over the years. The astronomic rise in this number demonstrates the recent trend of IoT application [32], [39].

IoT includes, but not limited to, technologies such as medical equipment, smart vehicles, smart grid, smart homes, and smart cities. IoT applications bring forth numerous benefits. It can reduce human intervention in the process of interconnecting devices. The most important impacts can be observed in the power sector, home appliances, and in smart cities. Smart grids which contain the attributes of IoT may be the possible solution of future global energy crisis. Efficiency at

TABLE 2. IoT Components for Monitoring Power Transmission Lines, with Corresponding Monitoring Techniques [44].

Monitoring Item	Methodology	Usage
Transmission Tower Leaning	Leaning sensor transmits the status of the transmission tower to the nearby backbone node.	Real time monitoring Early warning
Conductor Galloping	Calculation and analysis of monitoring point can determine the number of horizontal and vertical galloping conductors.	Avoiding tower collapse Possible discharge between phase conductor
Wind deviation	It can be calculated by three dimensional accelerated sensors on the conductors.	finding discharge point
Micro-meteorology	Wireless sensors can be used.	Wireless recording of temperature, humidity, wind velocity, sunshine, and rainfall
Conductor icing	It can be determined by micro-meteorology and tension sensors.	Early warning decision making Alleviating ice flashover
Wind vibration	Can be detected with acceleration sensors.	Analyzing wind parameters
Conductor Temperature	Wireless temperature sensor can be used	Real time conductor state monitoring

transmission and distribution ends can be escalated. Renewable energy sources can be more effectively utilized under IoT based networks. Currently, smart homes have monitoring systems that increases the cost effectiveness [33]. It also reduces the unwanted consumptions of energy. In a smart city, optimization of schedule for public transport can be done with IoT. However, though the general lifestyle has caught up with this technology, it is hardly present in the grid system. Incorporating these connecting devices in the grid infrastructure is a major step to advance towards smart grid – which can be evidenced by the significance put on IoT in designing microgrids [34]. Niche uses of IoT devices are also emerging with applications that are already exists, or anticipated to appear in near future. Smart homes, where household appliances can be controlled by connected intelligent devices is an example of such use. Connected vehicles, distributed energy resources (DER), green buildings are some more applications [35]–[37].

Smart energy system aims at reducing energy loss while simultaneously providing sufficient energy and services to everyone. In India more than 30 percent loss in electricity occurs during the power production process [40]. In France and Australia, 35% wastage of water occurs due to the leakage in the system; therefore, the electricity used in processing that water also goes to waste. To meet the increasing demand as well as reduce the energy wastage, a real-time tracker of supply and demand sides of distribution system needs to be developed – which IoT can provide. Centralized systems should be replaced with a distributed microgrid, which provides real-time monitoring and communication to the grid, along with remote sensing technique, two-way communication and demand response.

Useful IoT devices for using in the power sector has already been developed, smart meter being one of them. The fundamental concept of smart meters is to provide a two-way communication simultaneously while measuring power. The measurement data is transmitted to the utility suppliers through a mesh network. Low power radio frequency (LPRF) communication using a sub-1 GHz mesh network is used in USA to convey these data. Wired narrowband orthogonal frequency-division multiplexing (OFDM) powerline communication technologies are used in France and Spain. Energy information can be sent to the demand side following this same mechanism. IEEE 802.15.4 2.4 GHz ZigBee®standard is used in USA for this purpose. UK and Japan are considering sub-1 GHz RF solutions. A combined implementation with both hybrid RF and powerline communication can be a feasible way to provide necessary energy information to consumer homes. Smart meters allow better tracking of consumption and generation, and better energy management, among other things.

Similar to the smart meter, IoT can be integrated in smart grid through all of its major subsystems: generation, transmission, distribution, and utilization [41]–[43]. For example, IoT can provide monitoring services for the power transmission line, where one part of this monitoring system is deployed at the transmission line to monitor the condition and readings of the conductors; another portion of monitoring system is deployed at the transmission towers. This portion monitors the environmental conditions of the towers. A wireless communication technology is used to establish communication between the transmission line and the towers. The main monitoring components are listed in table 2 [44]. Possible integration of IoT technology in all of smart grid's subsystems are listed in table 3. The major security concerns for each of these subsystems are also mentioned in this table; such security threats are discussed in detail in section VI.

Seamless communication is a core feature of smart grid, essential for its proper functioning; and IoT integration can aid in smart grid communication too. Mainly four models are currently being used for communication technologies: device to device, device to cloud, device to gateway, and back-end data sharing pattern [44]. Three layered communication systems for IoT implemented smart grid system has also been developed, the layers being: home area network (HAN), neighbor area network (NAN), and wide area network (WAN). HAN comprises both wired and wireless technologies, e.g. wired technology is powerline communications, and wireless communications are ZigBee, Bluetooth, and WiFi. A home gateway (HG) is a key component of HAN, which collects data from home appliances. NAN requires a communication system which can cover a radius of more than thousand meters. NAN collects data from the energy meters in HAN and transmits those data to the WAN [52].

TABLE 3. IoT application in smart grid.

Application Layer	IoT application	Security Concerns
Generation	Monitoring of energy generation Controlling units, gas emission and pollution discharge [45] Prediction of power usage [46] Managing distributed power plants and microgrid [41]	Sabotage Data theft Unauthorized use of computational power
Transmission	Monitoring and controlling of transmission line and substation Protection of transmission tower [42, 43]	Sabotage
Distribution	Distribution automation [47, 48] Management and protection of equipment Fault management [49]	Sabotage
Consumption	Smart homes & appliances [50] Intelligent charging and discharging of electric vehicles [51] Power load control Multi network management	Data theft Identity theft

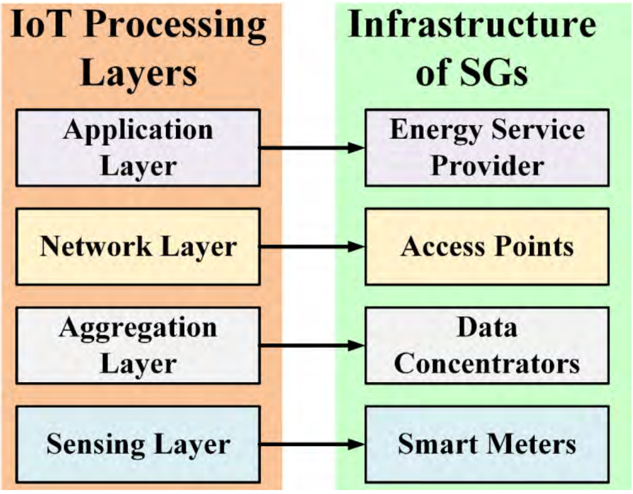


FIGURE 4. Structure for IoT implemented layers for SGs [57]. Each IoT layer corresponds to a certain layer of SG infrastructure.

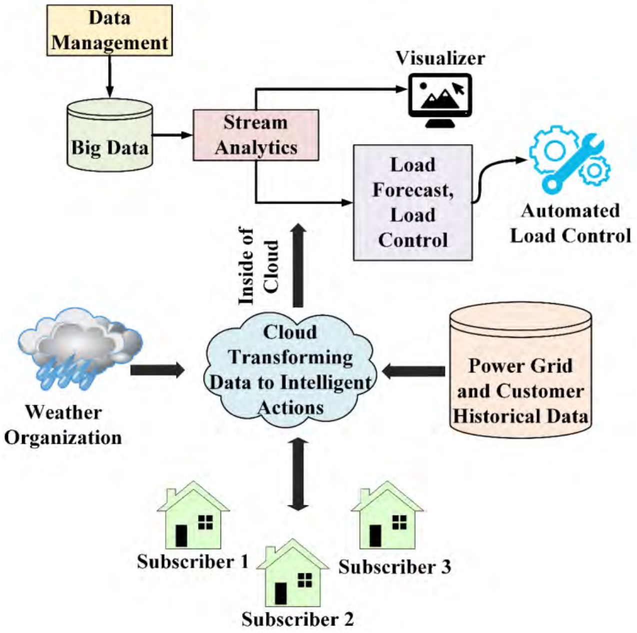


FIGURE 5. System architecture for load shedding and smart load controlling algorithm [58]. All the components of the system are connected to the cloud which makes decisions based on the system inputs, and sends out commands for execution. The directions of the arrows indicate the flow of data.

WAN serves as interconnection between every component of communication link such as network gateways, NANs, distributed grid devices, utility control centers, and substations. Core and Backhaul are two interconnected networks of WAN. Detailed discussions of HAN, NAN and WAN systems are included later in this section. Information from the physical systems in IoT integrated smart grid is fed into data concentrator [53]–[55]. From data concentrator information is met with the requirements of internet protocols for web services or cloud computing platforms. Those web services and cloud computing platforms further process the data. The energy service providers' (ESP) sites are connected with the Aggregation layer [53]–[56]. The underlying layers of IoT are depicted in fig. 4.

Efficient load management is a key benefit of employing the IoT technology. As a general practice, system disturbances that cause shortage in power generation are compensated by adjusting the amount of load from the demand side. This adjustment of load keeps the other components of grid running. Smart load control and load shedding should aim at minimizing power outage in sudden change of a load in the grid. An automated system to do such tasks with the help of IoT devices was presented in [58]. This method worked by predicting the day-ahead load, and tracking the available generation. When it found the load to be greater than the supply, it could suggest the consumers to switch off some appliances, or schedule possible loads to run at off-peak hours. Fig. 5 shows the working principle of this method. Subscriber (consumer) data, weather information, and historical data from the grid were used for the prediction in this system. All the analysis and decision-making was conducted in a cloud infrastructure; while the system components communicated through powerline communication or some wireless technology. Simulation results showed this system to be quicker in responding to emergencies, and its potential to avoid sudden power outage [58].

The integration of IoT devices bring some unique challenges with them in the smart grid scenario which are inherent to such technology, and the fact from which all that stems from is latency. Latency is defined as the difference between the time of data generation and the time when it becomes available for applications. In other words, it is the time delay for the data to become available. Latency in IoT architecture can be characterized as communication latency and phasor data concentrator (PDC) latency. Communication latency on the network is comprised of transmission delays, propagation delays, processing delays, and queuing delays. PDC latency, on the other hand, comprises of PDC device latency and PDC wait-time. Wait-time latency indicates the time each PDC has to wait for certain user-configurable time-duration so that slower PMU data can be reached and processed for time alignment operation at PDC. The maximum tolerable latency is 40 ms which includes latencies introduced by communication network, PMU, PDC, etc. This requirement must be met by any IoT architecture for the system to function deterministically [59], [60].

Communication infrastructure is a critical aspect of IoT deployment. Typical means of communication are leased telephone lines, power lines, microwave and fiber optic, among others. IoT functions require real-time processing of synchrophasor data at wide area level to aid in making informed decisions. The protection and control commands should be available for the destination in a deterministic manner. Communication channel capacity, latency and hence efficiency, are therefore important for successful implementation of IoT system. Channel capacity defines the amount of data that can be carried by a communication network. In IoT system, all PMUs send phasor data directly or indirectly to central PDC (phasor data concentrators) for concentration through time alignment and data aggregation. The aggregated data streams are used by analytic functions to identify anomalies and issue corresponding commands to rectify. With an expected 3000 PMUs transmitting 4 phasors, 6 analog quantities, and 8 digital quantities each to central PDC, all in floating-point format at 25 or 50 messages per second rate, 68 Mbps on a serial port, or 135 Mbps in TCP IP, or 122 Mbps in UDP is required. As discussed in [61], a typical serial port cannot handle the above traffic and hence an Ethernet port with TCP/IP or UDP is preferred. Essentially, the chosen communication infrastructure should enable the bandwidth requirement in a reliable way.

Since IoT data contains critical information, it is imperative that the data to be secured from all types of attacks. Attackers can modify the data to cause system instabilities or even blackout. To ensure the reliability, a two-layer communication security can be constructed: one inside substations using already deployed security measures for all data communication, and the other by secured means such as encryption for data stream outside substation. All analytical functions using IoT data assume that the incoming data is error free and continuous. But PMU measurements can become unavailable due to unexpected failure of the PMUs or PDCs or due to loss of communication links caused by congestion of communication network. This missing data will result in wrong outputs from the analytical functions. Practical counter measures for reliable and secure data transfer are: building in as much redundancy as possible in PMUs, PDCs and communication; proper PMU placement and wide area measurement system (WAMS) design; and usage of robust analytic functions [62].

IV. SMART GRID WITH BIG DATA ANALYSIS

As it is mentioned in the previous section, integrating IoT devices in every sector of the grid infrastructure is a mandatory step for moving towards smart grid. It has also been stated that the defining feature of these devices is their ability to communicate with other devices and control centers, and send useful information. Thus, an unprecedented amount of data gets generated in an interconnected network [63]–[66], posing challenges to the conventional methods of data transfer, storage, and analysis. As documented in [67], water consumption data of 61,263 houses in Surrey, Canada amounted to 5 MB, information about speeds and locations of vehicles passing through the Madrid Highway, Spain generated 450 MB of data, and monitoring a 400 square kilometer area in Cologne, Germany for a day created a dataset sized at 4.03 GB – recording information of around 700 vehicles. Monitoring of transmission line, generation unit, substation state, smart metering [68], and data acquisition from smart home - all produce a large amount of data from the smart grid, which are to be stored in a cloud-based system for proper analysis. Cloud supported IoT system has been proposed in [69] to manage all those data.

Enter big data analysis [70], which has become a buzzword in the global scientific and data analyst communities [71]–[73]. Big data refers to massive amount of data that require more advanced methods to be captured, curated, managed, and analyzed than the traditional tools and signal processing models. The amount of data that defines big data is not explicitly defined, rather it moves as the technology progresses. Generally, data demonstrating three characteristics can be labeled as big data: it has a large volume; the velocity or frequency of this data generation, storage, or transmission is high; and there is a lot of variation of data in the dataset. These features match with the data IoT devices generate, and thus the data generated in the smart grid can be considered big data. Fig. 6 shows how expanding in the aforementioned three sectors define big data. Although big data means a massive amount of data, technically it covers the predictive and behavioral analysis using those data. This huge amount of data is available at every aspect of our lives, and demands critical analysis. Scientists, businessmen, social welfare organizations, economists, and many others need to process through this large volume of information that is available online. Big data analytics is based on this massive data, and the associated

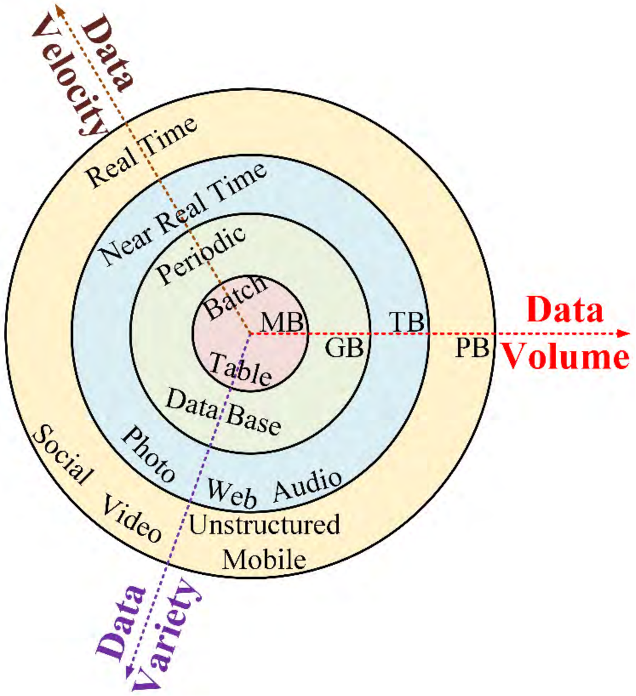


FIGURE 6. Big data characteristics: large volume of data with lots of variations which are generated, stored, or transmitted at a high at a high velocity can be labeled as big data.

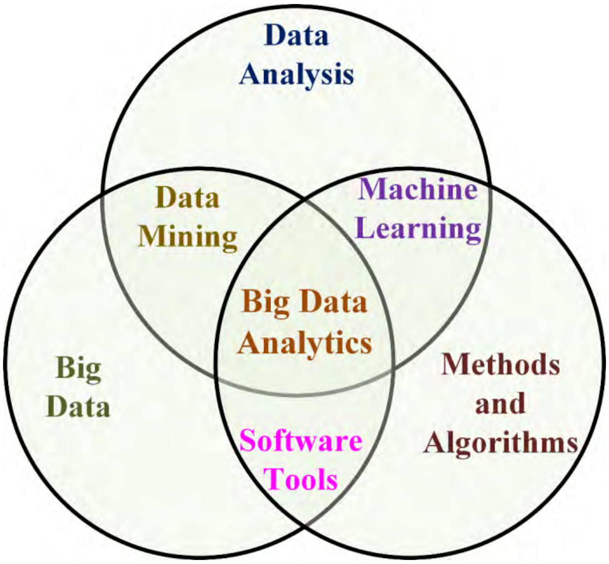


FIGURE 7. The components that create big data analytics. Big data and the techniques to analyze it has created the discipline of big data analytics [87].

analytic techniques, which is visualized in fig. 7. These techniques are based on different platforms such as Windows, Linux, Mac etc., and they require certain levels of expertise. They also have certain limitations, which hinders the rise of a single superior tool. Different tools with their platforms, required skill levels, and limitations is presented in table 4. Table 5 lists some more analysis techniques to juxtapose their

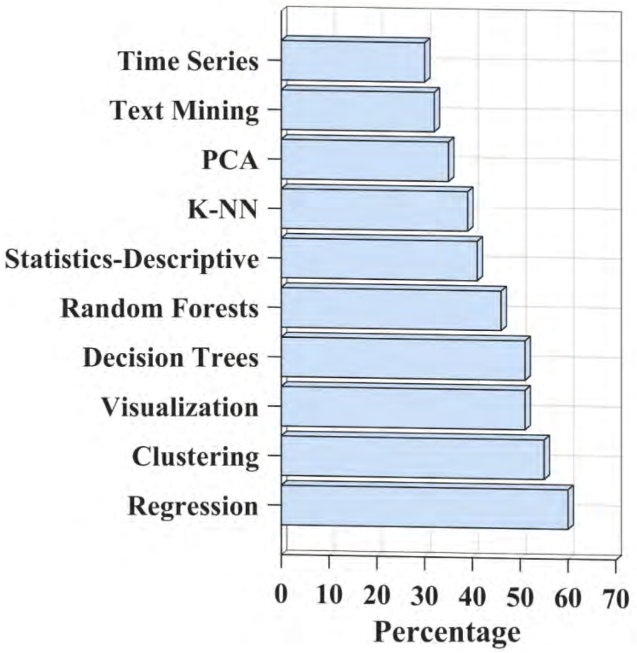


FIGURE 8. Most used data science techniques in 2017. Regression tops this list, attracting 60% of the users [94].

advantages, applications, difficulty level to master, required system to run, associated software, and financial cost. From this table, it can be seen that most of the systems has a high cost involved, and all the high-cost systems have an 'expert' level difficulty. Table 6 demonstrates some notable works on big data analysis in smart grid to point out their specific applications. Fig. 8 shows the use percentage of the most-used data science techniques in 2017, where regression appeared as the most popular one with 60% usage. Clustering was used by 55% user, while visualization and decision tree attracted 50% of all users. A 2014 report published by the National Rural Electric Cooperative Association (NRECA) of the USA enlisted big data capabilities as a crucial component of the next generation power grid, or in other words, smart grid. It states that the ever-increasing deployment of phasor measurement units (PMU), and synchrophasors at transmission, distribution, and distributed generation sectors will generate massive amounts of data - which will vary as the direction of power flow will change depending on seasonal and daily conditions [74]. Such deployments of PMU can also lead to proactive control of grids, preventing faults from taking place instead of clearing a fault after its occurrence [75]. Analyzing big data is stated as a key functionality for energy management systems (EMS) for smart grids, control algorithms, and future energy market models in [76]–[83]. Zhou and Yang [84] presented ways to determine residential energy consumption through big data analysis. Demand side management through bid data analysis has also caught much attention [85], [86]. Dynamic energy management through such data analysis is also a promising technology [23].

TABLE 4. Different analytical tools for big data, their platforms, required skill levels, and limitations (adapted from [88]).

Tool	Category	Platform	Skill Level	Limitations
Data Wrangler	Data cleaning	Browser	Advanced begineer	Sends data to an external site which compromises securtiy.
R Project	Statistical analysis	Linux, Mac, Unix, Windows	Intermediate to advanced	Takes time to adapt; interface is text-only; has limited memory.
TimeFLow	Temporal data analysis	Desktop+Java	Beginner	No option available for exporting results.
NodeXL	Network analysis	Windows	Expert	Problems with application program interface (API).
CSVKit	CSV (comma-separated values) file analysis	Linux or Mac with Python	Expert	Slow to adapt; Dependant on Python.
Tableau	Visualization app	Windows	Advanced beginner to intermediate	Sends data to public website which compromises securtiy.

TABLE 5. Advantages, application, difficulty level, system requirement, software platform, and cost associated with some big data analysis techniques (adapted from [89]).

List of Methods	Advantage	Application	Difficulty Level	System Required	Software Involved	Cost involved
Machine learning	Strategic resource usage, simplified management.	Building and training intelligent system.	Expert	Training dataset, AI algorithms.	Python, Matlab	High
Data mining	Summarizing relevant information from a large amount of data.	Data extraction, data cleaning, data labeling.	Expert	Crawler, NLP.	Python, R, Matlab	High
Genetic llgorithms	Saving time during traning, quick convergance.	Training optimization.	Expert	Genetic representation, fitness function.	Matlab, Python	High
Neural networks	Less mean squared error (MSE).	Anomaly detection, pattern recognition, prediction.	Expert	Training dataset, hidden layer, optimization.	Matlab, Python	High
Natural language processing (NLP)	Handling text data.	Analyzing and visualizing text data.	Expert	NLP tool	Python	High
A/B testing	Figuring out the best strategies.	Web analytics.	Beginner	Browers	Google analytics, Optimizely	Medium
Cluster	Gaining insights from data.	Data grouping, classification.	Intermediate	Classifers	Python, Matlab	Medium
Crowd-sourcing	Human intuition, real time analysis.	Gathering large scale data features.	Beginner	Web page	FeatureHub	Low

TABLE 6. Applications of big data analysis in smart grid.

Reference	Institute	Year	Application
Zhou et al. [84]	Hefei University of Technology, China	2016	Determining residential energy consumption
Zhou et al. [85]	Hefei University of Technology, China	2015	Demand side management
Zhou et al. [86]	Hefei University of Technology, China	2016	Demand side management
He et al. [90]	Shanghai Jiaotong University, China (with external collaboration)	2017	High-dimension smart grid modeling
Ryu et al. [91]	Sogang University, Korea (with external collaboration)	2016	Short time load side prediction
Coelho et al. [92]	State University of Rio de Janeiro, Brazil (with external collaboration)	2017	Load forecasting
Bessa et al. [93]	INESC Technology and Science, Portugal (with external collaboration)	2015	Very short-term solar power forecast

Agelidis et al. mentioned big data as one of the challenges in the information and communication technology part of smart grid [95]. Data in smart grid come from various sources. Mainly two domains provide smart grid data: generation domain and service provider domain [96]. Uncertainty in data analysis demands development of methods for long and short-term data patterns from distributed energy resources (DER). Few of the proposed models are Gaussian model [97], finite state Markov models [98], hidden Markov model (HMM) [99], [100], proportional integral derivative (PID) controller [101], and online learning techniques [102]. At the service provider domain, smart appliances provide information for energy price estimation [23], [103]. A flask framework called OASIS dashboard was developed for the visualization of real time energy data from the energy sources in Puerto Rico in [96]. Reference architecture for smart grid using big data and intelligent agent technologies was developed in [104]. In this architecture, agent-oriented programming methodologies were adopted in Hadoop platforms [104]. This system supported interoperability in smart grid systems. Application of random matrix theory (RMT) for high-dimension smart grid modeling was demonstrated in [90], which claimed to provide better accuracy, and practicality for large interconnected systems such as smart grid. Five case studies conducted in that work verified the proposed system's capabilities.

TABLE 7. Steps involved in machine learning and data mining [135].

Steps involved	Standards			
	Fayyad	Cios	SEMMA (Sample, Explore, Modify, Model, Assess)	CRISP-DM (Cross Industry Standard Process for Data Mining)
Determining objective	■	■		■
Collecting data	■	■	■	■
Cleansing data	■	■	■	■
Reducing data	■		■	■
Reformulating problem	■			
Explorating data	■		■	
Selecting tools	■		■	
Constructing model	■	■	■	■
Validating model	■	■	■	■
Interpretating result	■	■	■	■
Deployment	■	■		■

For the development of energy efficient and sustainable data processing system, a framework with robust time advanced workload and energy management was developed in [105] to integrate renewable energy sources, distributed storage unit, dynamic pricing unit etc. for green DC systems. In that work, a resource allocation problem was developed so that the net cost of the system could be minimized with spatio-temporal variations of workloads and electricity market prices. Net cost of the system comprised of network operational cost and the worst-case energy transaction cost. An optimal solution was achieved by Lagrange dual based distributed solver using strong duality of convex reformulation [105]. Large amount of data from a power system require fast and efficient computing, which has been a concern for several researchers. Task parallelism with multi-core, cluster, and grid computing can reduce the computational time in an efficient data mining algorithm [106]. A grid computing framework was developed for higher computing efficiency in [107]. In this framework, the overall architecture consisted of three layers: resource layer, grid middleware and application layer [107]. The data generated in the smart grid raises two major concerns. Firstly, the data must be processed and transferred in an efficient way within an acceptable limit of time. And secondly, security concerns are very important issues regarding IoT integrated smart grid [108]. To provide an insight on these issues, the next two sections are organized to address these concerns.

V. MACHINE LEARNING APPLICATION IN SMART GRID

The obvious question that arises from the big data generation from smart grid is efficient ways to analyze them for extracting valuable information. Without the extraction of useful information, the collected data holds little or no value. Machine learning appears as the tool required for the tall task of going through the massive amount of data generated in an IoT-based grid system. It fits in as the final piece of the smart grid system which is driven by data collection, analysis, and decision making. Machine learning techniques provide an efficient way to analyze, and then make appropriate decisions to run the grid; and thus enables the smart grid to function as it is intended to.

Machine learning (ML) is a term which refers to learning and making predictions from available data by a system. It is comprised of various algorithms which analyze the available data through a set of instructions to produce data-driven predictions and/or decisions. Machine learning undergoes the rigorous process of designing and programming explicit algorithms with expected performance. The steps and associated standards of this process are presented in table 7. Machine learning functionalities include predictions of consumption, price, power generation, future optimum schedule, fault detection, adaptive control, sizing, and detection of network intruders during a data breach [109]–[116]. Xu *et al.* [117] presented an assessment model for analyzing transient stability which employed extreme learning machine, and demonstrated impressive accuracy and computational speed when tested on New England 39 bus system. Wang *et al.* [118] pursued a similar objective with their novel core vector machine (CVM) algorithm to utilize big data generated by PMU, their system was also tested on the New England bus system. For transmission systems, machine learning can be employed to analyzed the phasor measurement unit (PMU), and micro phasor measurement unit (μPMU) data for uses such as system visualization and frequency detection. Machine learning can be used in these purposes alongside other software such as power plant model validation tool (PPMV), and free flight risk assessment tool (FRAT). Several machine learning methods are being introduced at different phases of renewable energy power system based SGs, creating a whole new prospect for research [119]–[121]. For example, the support vector machines (SVM) have been widely implemented into several problems of renewable energy power systems, which provided many optimization and prediction techniques in SG [122]–[124]. Economic optimization for smart grid prosumer node with a two-level control scheme is developed in [125]. Machine learning based fast and accurate algorithm for monitoring power quality (PQ) events in an SG has been developed recently in [126] and [127]. Li *et al.* [128] applied machine learning to analyze user predilections in a smart grid to find out usage pattern and preferences. Remani *et al.* [129] demonstrated a generalized use of reinforced learning to schedule residential load considering renewable energy sources and all possible tariff types. For distributed generation systems, islanding detection using machine learning and wavelet design was investigated in [130]. Application of particle swarm optimization (PSO) to enhance stability for unplanned islanding in microgrid is proposed in [131]. Big data analysis to monitor and detect such islanding incidents comes before this stabilization stage. A hybrid system for demand side management employing entropy based feature selection, machine learning, and soft computing was proposed by Jurado *et al.* [132]. Several algorithms such as extreme learning machine, support vector regression, improved second order, decay radial basis function neural network, and error correction to train common radial basis function networks for predicting load was investigated in [133]. Ryu *et al.* [91] proposed a deep neural network (DNN) load forecasting method for short term prediction at the load side which demonstrated as high as 29% less error compared to existing systems such as shallow neural network (SNN), and double seasonal Holt-Winters (DSHW). A graphics processing unit (GPU) based load forecasting method has been proposed in [92]. A 45 MW smart grid in University of California San Diego (UCSD) is considered in [134]. This grid supplies 54000 consumers from both renewable and non-renewable energy sources. The UCSD grid is equipped with advanced monitoring and storage techniques. In that work, big data analyses have been done leveraging large amount of data and the Hadoop system. Machine learning can also be applied for various security applications in smart grid. A concise presentation of such uses is shown in fig. 9. However, the most promising and much needed use of machine learning in the next generation energy system is the renewable energy sector. And therefore, in the following subsections, implementation of machine learning in SG with renewable energy sources is discussed.

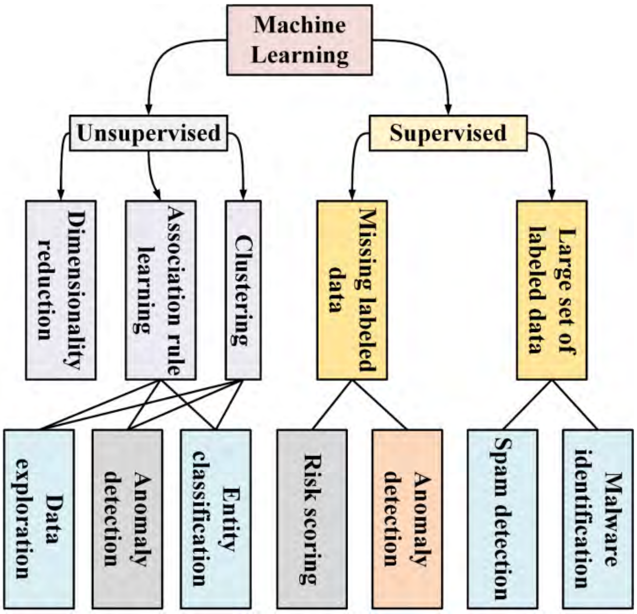


FIGURE 9. Application of machine learning in smart grid security. Unsupervised and supervised – both approaches can be used to carry out an array of tasks including threat identification and data categorization [136].

A. MACHINE LEARNING APPLICATION IN WIND ENERGY FORECASTING

Wind power is one of the fastest growing renewable energy sources in the world. About 12% of the world's electricity will be supplied by wind generation by 2020 [137]. Integration of wind power sources with the grid provides several technical, economic, and environmental benefits [138]. But due to the intermittent and stochastic nature of wind power, it provides some obstacles during power generation and distribution. Variation in wind speed causes fluctuation in the output of wind power plant, which leads to instability in the grid. Hence proper forecasting is required for wind energy based power grids, and can aid in making operational strategies [139]–[142]. This forecasting is complex, because from controlling wind turbine to integrating wind power into energy system, time duration for prediction changes from milliseconds or seconds to minutes or hours [143], [144]. Previously, several prediction models such as fuzzy modeling [145], auto regressive moving average [146], artificial neural network [147], [148], K-nearest neighbor classification [149], computational fluid dynamics (CFD) pre-calculated flow fields [150], extreme learning machine (ELM) [151]–[153], adaptive neuro-fuzzy inference system (ANFIS) [154], combination of relevance vector machine (RVM) and differential empirical mode decomposition (EMD) [155], combination of soft computing model and wavelet transformation [156], wavelet transform and SVM [157] etc. have been developed and applied. Applications of data mining for prediction of wind power was reviewed in [158]. Machine learning technique has been applied to diagnose wind turbine faults using operational data from supervisory control and data acquisition (SCADA) system of south-east Ireland [159]. Generalized mapping regressor (GMR) was employed in [160] to create steady-state model of wind farms that can help in detecting faults in case of an anomaly. Fan *et al.* [161] employed Bayesian clustering to create a dual stage hybrid forecasting model to aid in scheduling of a wind farm, and trading of wind power. This proposed system was validated by applying on a 74 MW wind farm at Oklahoma, United States. Parallel execution of Gaussian process and neural network sub models to predict wind power was presented in [162]. Short term wind power prediction using ELM and coral reefs optimization (CRO) algorithm was presented in [163] which demonstrated superior performance when applied to a wind farm in the United States. A similar objective was pursued in China through hybrid machine learning models based on variational mode decomposition and quantile regression averaging, which attained absolute error as low as 4.34% [164]. Improved simplified swarm optimization (iSSO), an improvement of simplified swarm optimization by means of bias and weight justification, showed impressive results when used to predict wind power generation at the Mai Lao Wind Farm at Taiwan [165]. Hybrid swarm technique (HAP) consisting of particle swarm optimization (PSO) and ant colony optimization (ACO) to predict wind power in short term from parameters such as ambient temperature and wind speed is presented in [166]. This system achieved a mean absolute percentage error rating of 3.5%. Wang *et al.* [167] developed a novel hybrid strategy based on a three-phase signal decomposition (TPSD) technique, feature extraction (FE) and weighted regularized extreme learning machine (WRELM) model. This model was able to do a multi-step ahead wind speed prediction. In this model, a three-phase signal decomposition framework including seasonal separation algorithm (SSA), fast ensemble empirical mode decomposition (FEEMD), and variation mode decomposition (VMD) were used to control the unstable and irregular natures of wind speed. An FE process including partial autocorrelation function (PACF) and regression analysis was used to utilize the effective and beneficial features of wind speed fluctuations. In this way, the optimal input features for a prediction model was determined. To improve the forecasting accuracy and efficiency, an improved extreme learning machine (ELM), named weighted regularized extreme learning machine (WRELM) was developed by utilizing these features. Application of reinforcement learning in energy trading in smart grids with wind energy generation was demonstrated by Xiao *et al.* [168]. Their proposed system used historical energy trading data, and energy price, to reduce power plant scheduling. The energy exchange scenarios between microgrids were also investigated using game-theory approach in [169], where it was shown that overstated trading information can result in reduced utility of smart grids.

B. MACHINE LEARNING APPLICATION IN SOLAR ENERGY FORECASTING

Solar energy is one of the most prominent renewable energy sources. Solar photovoltaic (PV) systems had 22 GW of global capacity in 2009 and almost 139 GW in 2013 [170], [171]. Similar to wind energy sources, the solar power systems too are impeded by many difficulties. Many natural and man-made impediments such as weather conditions, seasonal changes, topographic elevation, discontinuous production, and intra-hour variability have effects on solar PV system performance. As a result, solar energy information should be acquired in advance to minimize the operating costs caused by the various obstacles mentioned above. Prediction models for both meteorological forecasts and system outputs were presented in [172]–[178]. A forecasting approach aiming at very short-term solar power forecast based on the city of Évora, Portugal was proposed in [93]. This model used a vector autoregressive model fitted with recursive least squares. Data from smart meter and other smart components at medium voltage/low voltage (MV/LV) substation level were used in this model. Chaouachi *et al.* [179] proposed a neuro-fuzzy system for maximum power point tracking (MPPT) in 20 kW solar photovoltaic (PV) system. This method utilized classifier running on fuzzy logic in accordance with three artificial neural network having multiple layers. Reference [180] utilized those two methods for an intelligent energy management system that could predict PV generation 24 hours ahead. Another day-ahead forecasting method that could take weather data into consideration was presented by Yang *et al.* [181]. Their method was a hybrid one, employing three different machine learning techniques in the three stages of the prediction system, and was trained on data collected from the Taiwan Central Weather Bureau. Hossain *et al.* [182] employed machine learning techniques such as multilayer perceptron (MLP), least median square (LMS), and support vector machine (SVM) in forecasting of solar power in two-phase experiment, where the second phase concentrated on parameter optimization to find out performance enhancement margin before and after such optimization. They concluded that increased attention in parameter optimization and selection of feature subset could go a long way to increase prediction accuracy. Li *et al.* [183] proposed a solar irradiance forecasting technique employing SVM regression and hidden Markov model. A coral reefs optimization - extreme learning machine (CRO–ELM) algorithm was proposed in [184] to predict solar irradiation worldwide – which demonstrated better performance than conventional ELM and SVM. Salcedo-Sanz *et al.* [185] also employed Gaussian process regression (GPR) for such prediction. This method outperformed statistical regression algorithms in terms of robustness to prediction numbers, bias, and accuracy. A hierarchical model was proposed based on

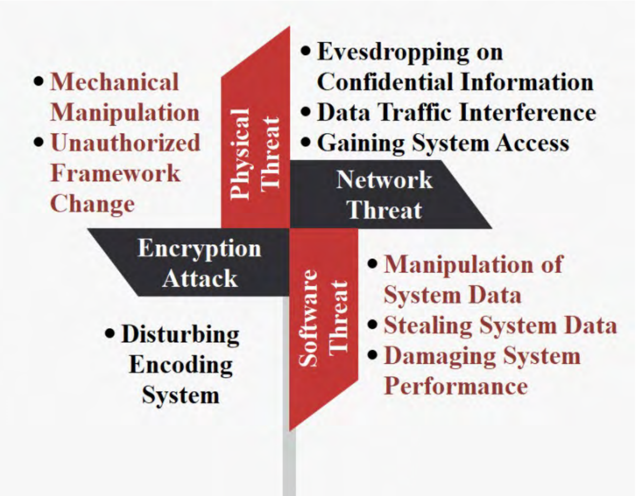

• Mechanical Manipulation
• Unauthorized Framework Change
Physical Threat
• Evesdropping on Confidential Information
• Data Traffic Interference
• Gaining System Access
Network Threat
Encryption Attack
• Disturbing Encoding System
Software Threat
• Manipulation of System Data
• Stealing System Data
• Damaging System Performance

FIGURE 10. Security concerns of IoT integrated SG system. These can be categorized into four major types: physical threat, network threat, software threat, and encryption threat. The security concerns under each type are marked in corresponding color.

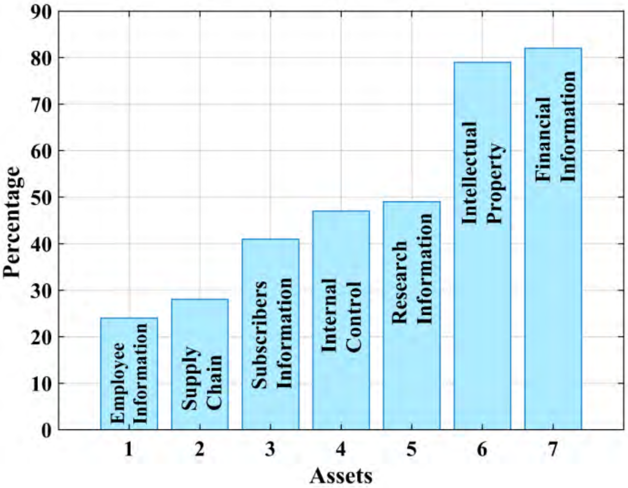


FIGURE 11. Typical targets of cyber-attacks: intellectual property and financial information are the two most sought after assets for attackers.

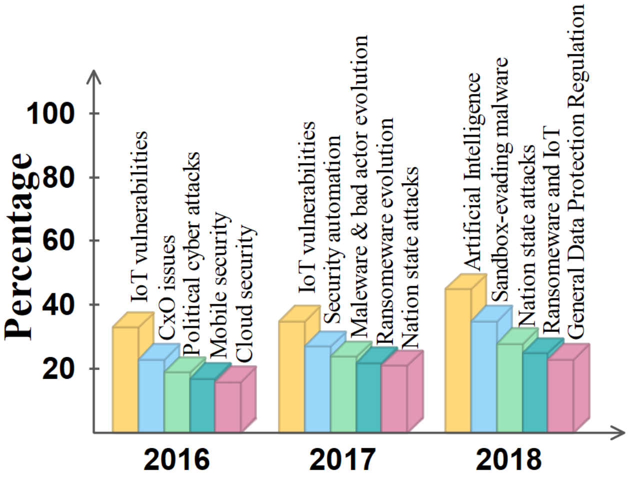


FIGURE 12. Predicted attack sources over the years. IoT always remains a major vulnerability against cyber-threats. "CxO issues" indicate data breaches at corporate officer levels.

the machine learning algorithms by Li *et al.* [186]. In this work, 15-minute averaged power measurements were collected from the year 2014. Computing error statistics models were used to test its accuracy. The hierarchical forecasting approach utilized machine learning tools at a micro level to predict each inverter performance. Then it evaluated the performances at a larger level by adding up the micro level predictions. In that way, it provided a bigger picture of the plant. This framework is visualized in fig. 36. Table 8 summarizes the applications of machine learning techniques in renewable source integrated smart grid encountered in literature.

VI. CYBER SECURITY IN SMART GRID

A. CYBER-SECURITY CHALLENGES IN SMART GRID

As IoT integrated smart grid systems create a complex interconnected web as well as a large volume of data which is often stored in cloud storages, breach of data security is a serious concern [187]. Threat to the security of this complex network is always very critical and sensitive as both the demand side and the supply side of power system are affected [188], [189]. Various types of security threats for IoT integrated SG system is depicted in fig. 10 [190]. Such cyber-attacks to smart grid systems can be carried out to cause damage to its crucial components, to gain foothold or superiority in its control system for exploitation, economic intimidation, or sabotage. The system-assets targeted in-general in cyber-attacks are depicted in fig. 11, which shows financial information and intellectual property being the most attractive targets to attackers, attacked almost 80% of the instances. Research data and other information as well as control systems are the other major targets. Information generated from the IoT devices comprise of most of the targets, and that alone can suffice to demonstrate the magnitude of the need of cyber-security for smart grid systems. Table 9 shows the information that are available in a smart grid, and can be targeted for attacks. The components vulnerable to cyber-threats in a utility's digital infrastructure are shown in table 10. The vulnerability of IoT devices to cyber-attacks rated them the most probable way of attacking both in 2016 and 2017 [191], [192]. Potential sources of attacks over the years are shown in fig. 12. Along with IoT vulnerabilities, data breaches at corporate officer levels were the second most serious threat in 2016 (shown as "CxO issues"). 2017 saw the rise of automation vulnerabilities, along with malware and ransomware attacks – quite a few of which were state-sponsored. The authors predict the use of artificial intelligence to be the most serious weapon in cyber-attacks, with advanced malwares, ransomwares, state-sponsored attacks, and weakness in data protection regulations. The most probable outcomes of cyber-attacks to smart grids can be: operational failures, synchronization loss, power supply interruption, high financial damages, social welfare damages, data theft, cascading failures, and complete blackouts [193]. Direct impacts of blackouts can be production loss

TABLE 8. Application of machine learning techniques in smart grids with renewable energy sources.

Reference	Institute	Year	Machine learning strategy	Technique	Application
Xu et al. [117]	Hong Kong Polytechnic University, Hong Kong (with external collaboration)	2011	Data categorization	Extreme learning machine	Assessment model for analyzing transient stability
Wang et al. [118]	Wuhan University, China (with external collaboration)	2016	Classification	Core vector machine (CVM)	Analyzing transient stability
Li et al. [128]	The University of Oklahoma, USA	2011	Pattern recognition	Machine learning	Finding out customer usage pattern and preferences
Alshareef et al. [130]	University of Ontario Institute of Technology, Canada	2014	Detection & Classification	Machine learning and wavelet design	Islanding detection in distributed generation systems
Jiang et al. [131]	National Renewable Energy Laboratory, USA (with external collaboration)	2017	Optimization	Particle swarm optimization	Enhancing stability for unplanned islanding in microgrid
Jurado et al. [132]	Sensing & Control Systems, Spain (with external collaboration)	2015	Data categorization & Optimization	Entropy based feature selection, machine learning, and soft computing	Demand side management
Marvuglia et al. [160]	CRP Henri Tudor/CRTE, Luxembourg (with external collaboration)	2012	Detection	Generalized mapping regressor (GMR)	Detecting faults in wind farms
Fan et al. [161]	Monash University, Australia (with external collaboration)	2009	Classification & Optimization	Bayesian clustering	Scheduling wind farm and trading of wind power
Lee et al. [162]	University of Texas at Austin, USA	2014	Prediction	Guassian process and neural network	Predicting wind power
Salcedo-Sanz et al. [163]	Universidad de Alcalá, Spain (with external collaboration)	2014	Prediction & Optimization	Extreme learning machine and coral reefs optimization (CRO) algorithm	Short term wind power prediction
Zhang et al. [164]	Wuhan University, China (with external collaboration)	2016	Data categorization & Prediction	Variational mode decomposition and quantile regression averaging	Short term wind power prediction
Yeh et al. [165]	University of Technology Sydney, Australia (with external collaboration)	2014	Prediction & Optimization	Improved simplified swarm optimization (iSSO)	Predicting wind power generation
Rahmani et al. [166]	Universiti Teknologi Malaysia, Malaysia (with external collaboration)	2013	Prediction & Optimization	Particle swarm optimization (PSO) and ant colony optimization (ACO)	Short term wind power prediction
Wang et. al. [167]	Nanjing University of Information Science and Technology, China (with external collaboration)	2018	Prediction	Hybrid strategy based on a three-phase signal decomposition (TPSD) technique, feature extraction (FE) and weighted regularized extreme learning machine (WRELM)	Wind speed prediction
Chaouachi et al. [179]	Tokyo University of Agriculture and Technology, Japan	2010	Detection	Neuro-fuzzy system	Maximum power point tracking (MPPT) in solar photovoltaic (PV) system
Chaouachi et al. [180]	Tokyo University of Agriculture and Technology, Japan (with external collaboration)	2013	Prediction	Neuro-fuzzy system	Intelligent energy management system for photovoltaic generation prediction
Yang et al. [181]	National Cheng Kung University, Taiwan (with external collaboration)	2014	Prediction	Hybrid machine learning employing learning vector quantization (LVQ), self-organizing map (SOM) network, and Support vector regression (SVR) at different stages	Forecasting solar generation considering weather data
Hossain et al. [182]	Central Queensland University, Australia	2013	Prediction	Multilayer perceptron (MLP), least median square (LMS), and support vector machine (SVM)	Forecasting solar power
Li et al. [183]	CSIRO CCI, Australia (with collaboration)	2016	Prediction	SVM regression and hidden Markov model	Forecasting solar irradiance
Salcedo-Sanz et al. [184]	Universidad de Alcalá, Spain (with external collaboration)	2014	Prediction	Coral reefs optimization – extreme learning machine (CRO–ELM) algorithm	Predicting solar irradiation worldwide
Salcedo-Sanz et al. [185]	Universidad de Alcalá, Spain (with external collaboration)	2014	Prediction	Gaussian process regression (GPR)	Predicting solar irradiance worldwide

and business shutdown, food spoilage, damage of electrical and electronic devices, data loss, inoperability of life-support systems in hospitals and elsewhere, loss of critical infrastructure such as waste-water treatment plants etc. Indirectly, blackouts may result in property loss from arson and looting – which was observed in many previous occasions, overtime payment of personnel engaged in emergency management, potential increase of insurance rates etc. [194].

In a data-based system like the smart grid, false data injection can have devastating effects, and that motivation acts

TABLE 9. Available information in a smart grid system that can be targeted in cyber-attacks [195].

Data Element	Type of Asset	Description
Name	Subscriber information	Party responsible for the account
Address	Subscriber information	Location where service is being provided
Account number	Financial information	Unique identifier for the account
Meter reading	Internal control	KWh energy consumption recorded at 15-60 minutes interval during the existing billing cycle
Current bill	Financial information	Amount due on the account
Billing history	Financial information	Past meter bills including history of late payments
Home area network	Internal control	Network information of in-home elctrical appliances and devices
Lifestyle	Subscriber information	When the home is occupied and unoccupied, when occupants are awake and asleep, usage history of different appliances
Distributed resorces	Intellectual property	The presence of on-site generation and storage devices, operational status, net supply, consumption from the grid, usage patterns
Meter IP	Internal control	The internet protocol address for the meter, if applicable

TABLE 10. Vulnerable components of digital electrical utility infrastructure that can be targeted for cyber-attacks [194].

Component	Related Assets
Billing and Debt Collection (BDC)	Customers, personal information
Supervisory Control and Data Acquisition (SCADA)	National Utility Network, Centralized system
Programmable Logic Controller (PLC)	Subsystem of SCADA, Fault Recorders
Energy Management Systems (EMS)/Distribution Management Systems (DMS)	Electric Utility Grid information
Cyber Physical System (CPS)	Overall monitoring elements, Computer based algorithms
Power Line Communication	Smart Meters, Public Cellular Network

behind false data injection attacks (FDIA). The objective of such attacks is to alter original data in an attempt to mislead the system. Load distribution attack, stealthy deception attack, covert cyber deception attack, data integrity attack, and malicious data attack – all these terms are also used to mention such attacks [196], [197]. FDIAs need to be capable of escaping bad data detection (BDD) protocols in place, and perform stealth attacks on the system state estimation mechanism [196] – which is fundamental to monitor the state of a power system [197]. Also, most of the legacy BDD systems fail to detect such attacks [197]. Along with affecting the state estimations, FDIA can disrupt electricity markets through false economic dispatch and data [196], [198], [199]. False data can occur in the cyberspace, or in the physical space to affect device operation. These can result in flooding of a communication network, corruption of data, authentication failure, replacement of data packet from communication channels connecting phasor measurement units (PMU) and control center etc. [196], [200]. FDIAs are modeled mathematically in [197] and [201]. The advanced metering infrastructure (AMI) is one of the most targeted parts of a smart grid for cyber-attacks due to its large proportions and cyber-physical properties [202]. Wei *et al.* [202] listed the attacks on AMI components such as smart meters, and communication network, FDIA, and distributed denial of service (DDoS) appeared as risks in both sectors. Thus false data injection poses a significant threat that can be carried out in stealth to misguide the state estimation process, and disrupt measurement and monitoring systems in smart grid [199]. Energy trading is also a prominent feature of smart grids, which requires the exchange of energy prices, contracts, and transactions between grid entities. Because of these, the system attracts attacks including availability attacks, integrity attacks, and confidentiality attacks. If the energy trading sector is exploited, energy, money, and data theft as well as DoS attacks are possible [203].

Among consumer-level energy appliances, home energy management system (HEMS) is a common one. In an HEMS, security and privacy of communication infrastructure is provided by the home gateway (HG) system. Network and software attacks are both capable of damaging HEMS. Smart meters which record energy data from the user-end to billing have a connection which is stable and trustable with the home gateway system. Neither of these devices can be shut down remotely. They do not have the physical access of HEMS. Smart meters use wireless LAN (WLAN) and other communication networks which should be tamper proof. Home gateway serves as a communication channel of SG. Its configuration is controlled by the suppliers. Any error in the data can be reported to the meter point operator (MPO). Network attack is the most important concern of HG. Every component of HG has cryptographic key-stores, which use different protocols for secret key generation, key exchange and management. Different protection levels are associated with each protocol. Table 11 compares some of the most common attack types to demonstrate their relative effects on smart meter systems, and associated financial impacts. It can be concluded from this comparison that availability attacks are the most severe ones, as they have the most adverse effects on the smart meter systems, while all three attack types have serious financial tolls. If compared within the availability attack categories, radio frequency jamming, and reply attacks are the most effective ones financial sabotage and smart meter communication blockade; however, denial of service (DoS) is the weapon of choice for inducing delay in the smart meter system effectively [204].

TABLE 11. Common attack types and their impacts (adapted from [204]).

Attack		Financial Impact	Delay in Smart Meter System	Duration of Smart Meter Communication Blockade
Availability attacks	Denial of service	*	**	*
	Radio frequency jamming	***	*	***
	Reply attacks	***	*	***
Integrity attacks		***	*	*
Confidentiality attacks		***	*	*

***: extreme, **: moderate, *: mild

To visualize the process of cyber-attacks in a smart grid system, a simple scenario can be considered where the state of a power system is expressed in complex magnitudes of voltages and bus angles. Taking the voltage magnitudes as V, and the angles as δ, the state vector as S can be defined as [193]:

$$S = [\delta_1\delta_2\delta_3......\delta_n V_1 V_2 V_3......V_n]^T \tag{1}$$

The state estimation can be stated as below [205]:

$$\min J(X) = \sum_{i=1}^{m} w_i(z_i - h_i(X))^2 \tag{2}$$

Here, $h(X)$ acts as the measurement function to represent the measurement of the weights: z and w, $m$is the maximum number of measurement. Without any error,

$$Z_i = h_i(X) \tag{3}$$

With error, this will be:

$$Z_i = h_i(X) + e_i \tag{4}$$

Here e indicates the measurement error. Now if a cyber-attack aimed at inserting malicious data in this power system is launched, and it succeeds in modifying the measurement data with an attack vector, α, then the control system will receive the following measurement data:

$$Z_i = h_i(X) + e_i + \alpha \tag{5}$$

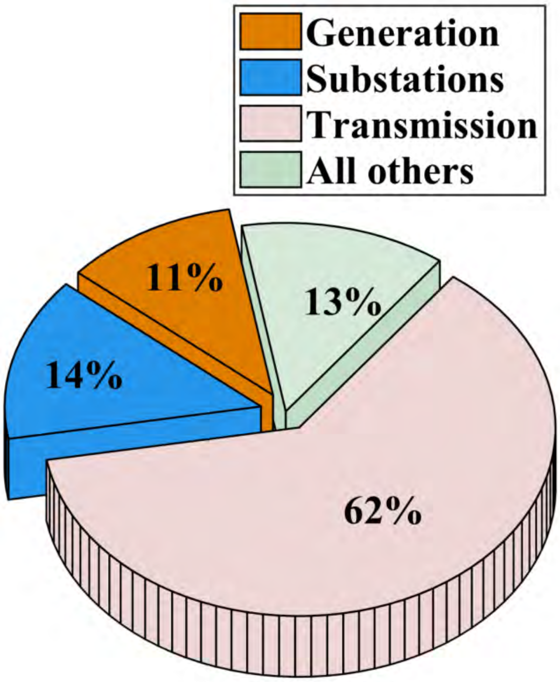


FIGURE 13. Attacks on major power grid components during 1994-2004; the transmission system faced most of the attacks, reaching 62%.

For contingency analysis [206], with W&W 6 bus system considered as the benchmark, power security was intended to be maintained for *N-1* contingencies by the North American Reliability Corporation (NERC) [207]. Even so, power systems remain exposed to damages resulting from outages in multiple branches – for example *N-k* contingencies. For *N* number of branches, total contingencies to be considered for *k* outages can be formulated as:

$$Total = (\frac{N}{k}) = \frac{N!}{k!(N-k)!} \tag{6}$$

Now if N-2 contingency is considered (taking $k = 2$), the possible combinations of simultaneous attacks will be:

$$(\frac{N}{2}) = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2} = (\frac{(N^2-N)}{2}) \tag{7}$$

Therefore, it is very much possible to cripple power systems with well-planned attacks.

Cyber-attacks on infrastructure is a very possible reality. In the fiscal year of 2014 alone, 79 such attacks on the United States energy companies have been recorded by a Department of Homeland Security division. In 2013, this number was 145. 37% of USA energy companies failed to prevent attackers in the time-period of April 2013 to 2014 [208]. During 1994-2004, the transmission system was attacked the most worldwide – a staggering 62% of all attacks in this period were aimed at this part of the power system, as presented by the Journal of Energy Security [209]. Attack percentages on all major power grid components in this time are presented in fig. 13. Some recent attacks are presented in table 12. In this

TABLE 12. Some recent cyber-attack incidents.

References	Year	Location	Attacked System	Impact of the Attack
[214]	September, 2010	Siemens systems at U.K., North America Korea, and Iran	Windows operation system	Unstable power system operation.
[215]	August, 2003	Midwest and Northwest U.S and Ontario, Canada	Software program of the cyber system	Up to 4-day-long blackouts in some parts, affecting around 50,000,000 people and 61,800 MW of electric load.
[215]	September, 2003	Italy and Switzerland	Communication system within the power grid operators	Disruption in power supply affecting a total of 56,000,000 people. 18 hours of blackout in Italy causing massive financial damage.
[215]	November, 2006	Southwest Europe	Communication system	Large blackout.
[216]	December, 2015	Ukrainian Kyivoblenergo	Computer and SCADA System	Blackout lasting 3 hours, affecting 225,000 people.
[217]	June, 2017	Ukraine	Network system	Power providers, major banks, government and airport computers taken out of service.
[213]	March, 2018	Atlanta, Georgia	Municipal system	City government's computer systems, traffic ticket system, water bill payment system, and airport WiFi were taken out by ransomware, affecting around 6,000,000 people.

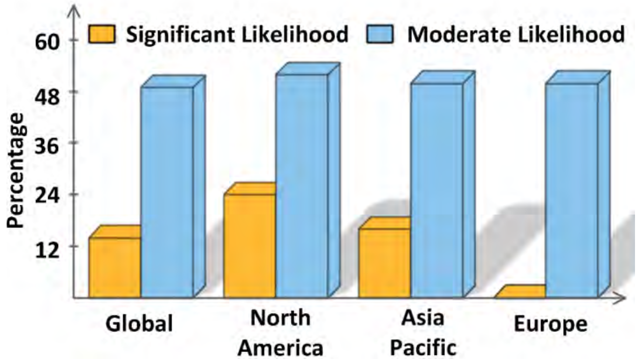


FIGURE 14. Likelihood of electric supply interruption from cyber-attacks, as predicted by utility executives. Moderate likelihood of such attacks are almost the same globally, but significant likelihood of attacks on European utilities is very low.

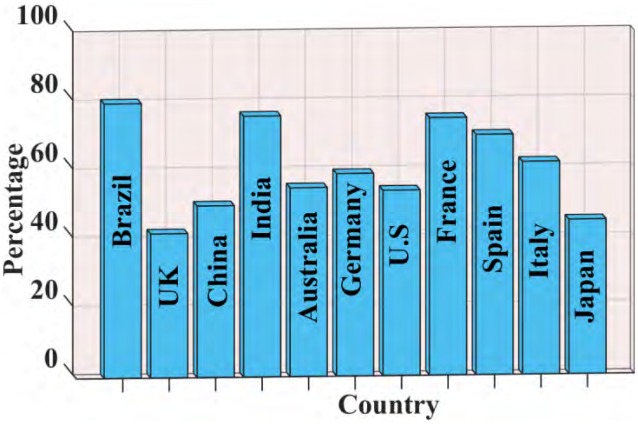


FIGURE 15. Reported large scale distributed denial of service (DDoS) attacks in different countries in 2007, Brazil was the most attacked one, while the United Kingdom and Japan stayed relatively safe.

age of connected systems, cyber-security thus appeared as a serious concern in the energy sector. 63% of utility executives believe their countries' utility grids face significant of moderate risks of being targets of cyber-attacks in the next five years, as found in a global survey conducted in October 2017. Their concerns regarding electric supply interruption from cyber-attacks are visualized in fig. 14, which shows moderate likelihood of attacks worldwide is almost 50%, with a similar scenario over North America, Asia Pacific, and Europe. However, while considering significant likelihood, Europe expects the least amount of attacks [210]. A survey conducted by McAfee in 2007 documented large-scale DDoS attacks. Frequencies of those attacks on infrastructures of different countries are shown in fig. 15. Brazil's systems appear to be the most attacked ones, hit 80% of the time, followed closely by India, France, Spain, and Italy [209]. This survey contrasts with the one presented in fig. 16, as the most three of the most hit countries (France, Spain, and Italy) are European. But these attack statistics are from 2007, and the survey visualized in fig. 15 is from 2017: which demonstrates the significant improvement in European cyber-infrastructure that almost negated significant likelihood of cyber-attacks in that region. Both of these surveys, however distant their time periods are, placed the United States as a prime target of attacks. Then it is no surprise to find this country as the one facing the most damages – $17.6 million – as documented in a 2017 report [211]. These losses faced by different countries are presented in fig. 15, where no other country faced so much penalty as the US, and Australia was the least hurt – capping the damage costs at $4.3 million. But being the target of the majority of cyber-attacks may have made the United States evolve as one the most prepared countries to face such adversaries, as shown in fig. 17 [212]. However, this 2015 preparedness index did not help the city of Atlanta in the state of Georgia, USA, when most of its municipal activities shuddered to a halt after being attacked by a ransomware on March, 2018. This situation persisted for five days, after

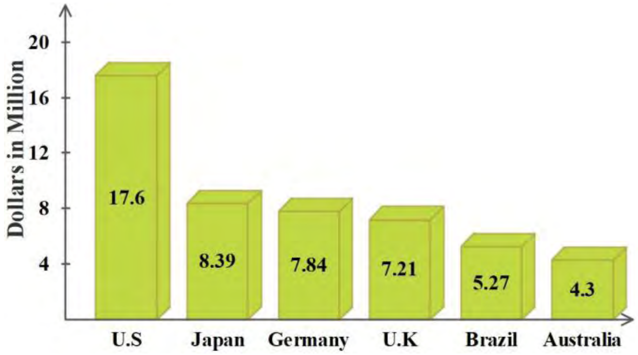


FIGURE 16. Costs of cyber-crimes in average around the world. The United States faced a huge $17.6 million cost caused by such crimes, more than double of what faced by the second most hurt country – Japan.

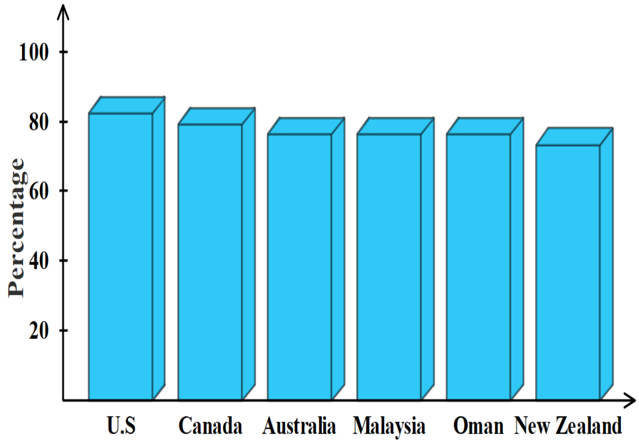


FIGURE 17. The countries most prepared to fend off cyber-attacks. All these countries have a very similar preparedness profile, though being leaded by the United States by a very small margin.

which the system recovered partially [213]. This is just a demonstration of the fast-evolving cyber-threats, and the need of better counter-measures.

B. COUNTER-MEASURES FOR CYBER-THREATS

Deploying smart grid system has far reaching impacts on an organization, and affects all the components of its technological infrastructure. Thus, security measures also need to be equally pervasive. Cyber-security strategies can be divided into two primary categories: protection and detection. Protection strategies can be hardware and administrative levels alongside the most-obvious software safeguards, while the detection can be done by applying machine learning techniques – which can predict threats as well as identify anomalies according to the features. Machine learning is applicable in most of the common tasks including classification, regression, and prediction; and thus appears as a promising solution to cyber-vulnerabilities in this age of big data and lacking cyber-defense. Security at the smart meters [218] is a good way to start on cyber protection. Also, the general approach of most organization is to enforce security at the smart meter, which is the tactical end point of their responsibility area. But such approaches fail to realize that the meter is not the only vulnerable area in the infrastructure. Thus, it is imperative for organizations to create a framework for assessing types of risks, and start this evaluation from the very top: security concerns associated with the strategy of their organization. A layered approach is needed for securing the smart grid, and the direction of strategy along with technical execution leads the way to such security layers. The driving forces and requirements of business process of any organization defines the strategic direction; while technical execution embodies data privacy, security, data integrity, network security, physical security, encryption, meter security, and associated operational procedures. In a layered security framework, the data use and security requirements are influenced by each layer according to its responsibility and accountability [219]. This layered security framework is demonstrated in fig. 18, where the security considerations of each layer are indicated.

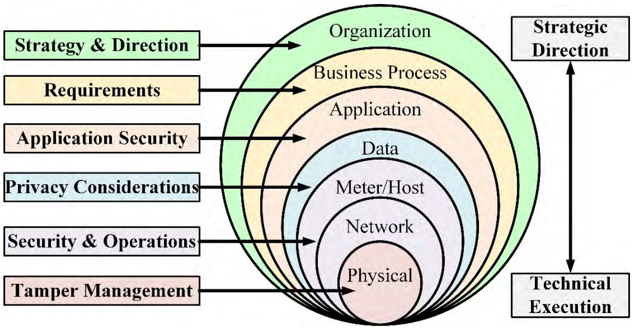


FIGURE 18. Layered security framework for smart grids. This comprehensive approach considers security at each stage of the infrastructure, rather than only smart meter placed at consumer location. Strategic direction and technical execution forms the two major contributors in defining the framework responsibilities [219].

Since an IoT-centric cyber-physical system such as smart grid provides a large volume of data, proper protection and management of this data in an SG is very critical. From the generation end to the distribution end, all kinds of data are protected with various methods. Previously, a number of work have been carried out on this purpose. Yuan *et al.* [220] developed a method for determining load distribution attack behavior. Requirements and standards of cyber security requirements have also been discussed previously in [221] and [222]. Certain standards are also enacted by various standardization organizations for cyber security which cover diverse areas such as management of information security, software size and quality, best practices, cyber security outcomes, secure integrated software and hardware testing, and industrial automation and control systems. These are presented in table 13. Data aggregation in the AMI system [223] is a target for attacks as well, and [224] presented a decentralized way of conducting that task efficiently while maintaining data privacy. Cloud computing is another important aspect of SG. This even produced the term ‘cloud grid’ in China, which integrates the nation’s power system with big data analysis facilities, IoT, information and communication technologies, and of course, cloud computing [225]. Security concerns regarding cloud based infrastructure have been addressed

TABLE 13. Current cyber-security standards and standardization organizations.

Standards	Full Form	Description
ISO/IEC 27001 [229]	ISO/IEC 27001:2013 – Information technology – Security techniques – Information security management systems – Requirements	A standard for information security management system (ISMS).
CISQ [230]	Consortium for IT Software Quality	Standards to automate measurement of structural quality of software, and their size.
ISF [231]	Information Security Forum	Issued an inclusive list of information security best-practices named 'Standard of Good Practice (SoGP)', updated every two years, except 2013-2014, the latest version came out in 2016 [232].
NERC [233]	North American Electric Reliability Corporation	Aimed at identifying the source(s) utilized
NIST CSF [234]	National Institute of Standards and Technology Cybersecurity Framework	Presents a high level taxonomy for outcomes of cybersecurity, along with a methodology for assessing and managing those outcomes.
ISO/IEC 15408 [235]	International Organization for Standardization/International Electrotechnical Commission 15408	Develops the 'Common Criteria' [236], and allows integration and secure testing of many different software and hardware products.
RFC 2196 [237]	Request for Comments	Developes security procedurs and policies for internet-connected information systems. Provides a broad and general overview of security of information which includes security policies, incident response, and network security.
ISA/IEC 62443 [238]	International Society of Automation/International Electrotechnical Commission 62443	Series of technical reports, standards, and related information to delineate procedures to implement industrial automation and control systems (IACS) that are secure electronically. Applicable to system integrators, end-users, control systems manufacturers, and security practitioners responsible for designing, manufacturing, managing, or implementing industrial automation and control systems.

TABLE 14. Encryption algorithms for cyber-security.

Category	Name	Key lengths for use between 2011-2029 (per SP 800-57 and SP 800-131)	Key lengths for use now and beyond 2030 (per SP 800-57 and SP 800-131)	Reference
Symmetric Key	Advanced Encryption Standard (AES)	AES-128, AES-192, and AES-256 with ECB, CBC, OFB, CFB-1, CFB-8, CFB-128, CTR, or XTS mode.	AES-128, AES-192, and AES-256 with ECB, CBC, OFB, CFB-1, CFB-8, CFB-128, CTR, or XTS mode.	[239]
	Triple-Data Encryption Algorithm (TDEA)	3-key TDES with TECB, TCBC, TCFB, TOFB, or CTR mode.	N/A – cannot use TDES beyond 2030	[240]
Asymetric Key	Digital Signature Standard (DSS): Digital Signature Algorithm (DSA), Rivest–Shamir–Adleman (RSA), Elliptic Curve Digital Signature Algorithm (ECDSA)	DSA with (L=2048, N=224) or (L=2048, N=256), RSA with (|n|=2048), ECDSA2 with curves P-224, K-233, or B-233	DSA with (L=3072, N=256) **, RSA with (|n|=3072) **, ECDSA2 with curves P-256, P-384, P-521, K-283, K-409, K-571, B-283, B-409, B-571	[241]
Secure Hash Standard	Secure Hash Algorithm (SHA)	SHA-224 is approved for all applications.	SHA-256, SHA-384, and SHA-512 are Approved for all applications.	[242]
Message Authentication	Cipher-based Message Authentication Code (CMAC)	CMAC with 3-key TDES	CMAC with AES-128, AES-192, or AES-256	[243]
	Cipher block chaining - message authentication code (CCM)	All algorithms/key sizes listed in the next column are approved.	CCM with AES-128, AES-192, or AES-256	[244]
	Galois/Counter Mode (GCM)/ Galois Message Authentication Code (GMAC)	All algorithms/key sizes listed in the next column are approved.	GCM with AES-128, AES-192, or AES-256	[245]
	Hash-based Message Authentication Code (HMAC)	HMAC with SHA-1, SHA-224, SHA-256, SHA-384, or SHA-512 with 112 ≤ Key Length<128 bits	HMAC with SHA-1, SHA-224, SHA-256, SHA-384, or SHA-512 with Key Length ≥ 128 bits	[246]

previously in [226] and [227]. Cryptography can provide protection of data from security breach. Cryptography is a method where data is stored in an encrypted manner. There are several algorithms in use for carrying out this task of encryption. They can be classified into various categories according to their working principles. Table 14 presents these security algorithms classified according to their techniques and security-key lengths. Use of quantum computer for encryption purpose can be very useful for data security [228]. In a quantum computer, data is stored as ''qubits'' rather than ''bits''. From the Heisenberg's uncertainty principle, the values of momentum and position of a physical system can be determined only with some characteristic uncertainty. This is the fundamental approach for quantum computer-based cryptography. External intervention or eavesdropping behavior will alter the state of the qubits. It will ultimately disrupt the integrity of the transferred information. Both private-key cryptography and public-key cryptography requires key distribution. In private-key cryptography, both ends of data transmission have a shared key to encrypt and decrypt data. However, in public-key cryptography, the key distribution is always done by a public-key sever. A method using cryptography has been proposed to approach the cyber security issue from modern quantum computing in [228].

For security measure in cloud computing based smart grid, a framework has been proposed by Baek *et al.* [247]. A flexible, scalable, and secure information management framework with cloud computing topology has been developed. The framework has three hierarchical levels: top, regional and end-user levels. A brief description of all the levels of this hierarchy is depicted in table 15. For a secure communication link between two different levels, identity of the higher level can be used for the lower level to develop an encrypted network [247]. The cloud computing centers have four major components: infrastructure-as-a-service (IaaS), software-as-a-service (SaaS), platform-as-a-service (PaaS), and data-as-a-service (DaaS). IaaS provides demand response for all applications and services in the system. SaaS provides services to the users; one such service can be optimization of power. PaaS develops tools and libraries for cloud computing applications. DaaS can be used for statistical purpose. There are four main clusters in the proposed framework: information storage, general user services, control and management services, and electricity distribution services [56].

For the detection part of cyber-security, data analysis, often employing machine learning is an obvious choice for countering cyber-attacks in the data-driven architecture of smart grid, and thus it has been heavily investigated in contemporary literature. Traditional signature based manual methods are almost useless in the current complex systems, and machine learning has non-linear analysis capabilities to detect false data injection in complex systems [200]. Efficient threat detection is particularly crucial for smart grids because of their sensitivity to delays: the system gets exposed to higher risks as the threat remains undetected for longer periods of time [197]. Buczak and Guven [248] conducted a detailed study on data mining and machine learning methods used for intrusion detection. They identified three major types of intrusion detection systems: signature-based (detects attacks from their signatures), anomaly-based (detects attacks by deviation from normal system behavior), and hybrid (combination of misused-based and anomaly-based methods). Further information on anomaly detection can be found in [249], while additional comparison of machine learning techniques for intrusion detection is available in [250]. Security threats to machine learning techniques specifically can be found in [187].

TABLE 15. Hierarchy of smart grid security framework with cloud computing topology.

Levels	Responsibility
Top	Accumulates data from regional cloud computing centers.
Regional	Manages intelligent devices.
End-users	Provides data for regional level.

For detecting false data insertion attacks (FDIA), general machine learning techniques artificial neural network (ANN) and support vector machines (SVM) were used previously, while implementation of other techniques in such detection were also conducted. Wang *et al.* [200] employed margin setting algorithm (MSA) which claimed better results than the two methods mentioned before. Other notable techniques used for this cause are Bayesian framework, particle swarm optimization (PSO), adaboost, random forests, and common path mining method [200], [251]–[253]. Ahmed *et al.* [197] proposed a machine learning approach based on Euclidean distance to detect FDIAs. They have also investigated on feature selection schemes with less complexity with improved accuracy that employed genetic algorithm for bad data detection [201]. Ozay *et al.* [254] used supervised learning to classify measurement data as secure or compromised, and thus detected FDIA. Their method was capable of identifying attacks that are unobservable, and predict attacks using observation sets. False data and stealth attack detection in wide area measurement in smart grid monitoring system was demonstrated in [199]. Xin *et al.* [255] presented a detailed study on machine learning and deep learning methods for intrusion detection, where the definitions of these areas are provided with descriptions of methods falling into each category. Wei *et al.* [202] discussed on detection of electricity theft, and provided an overview on the works done in that sector. Machine learning techniques such as principal component analysis (PCA) [256], as well as game theory approaches such as the Stackelberg game [257] can be applied in for detecting energy theft. Different types of software attacks and their counter measures [44] are depicted in table 16.

VII. OUTCOMES

The findings of this paper can be summarized as following:

- The electricity grid is currently going through the first major change from its inception almost two centuries earlier. This next-generation grid system is combining power system, information technology, communication and control systems to create a robust and adaptive infrastructure better suited to accommodate new and emerging technologies. This new grid is called the smart grid.

TABLE 16. Different software attacks and their counter measures [44], [202].

Attack Type	Description	Counter Measure
Spoofing	Unauthorized user can have access to a user's information; attacker may delete, change or control the information.	Introducing strong authentication mechanism Password encryption Using secure communication protocol
Tampering	Attackers modify user policies and device parameters; possibility of harming people physically	Strong authorization Digital signal Secure communication link Stackelberg game
Information disclosure	User privacy can be manipulated.	Strong authorization Password encryption Introducing private-enhanced protocols
Denial of service (DoS) & Distributed denial of service (DDoS)	Stopping all communications between stake holders, denial of access to EMS may be possible.	Use of home gateway to filter address with the help of firewall Using trolling technique Strong authorization Honeypot models
Elevation of privileges	The EMS includes third party plug-ins; which allow a sandboxed space in the EMS's functionality. When a malicious plug-in finds a backdoor, it could compromise EMS's assets.	Assigning minimum role for users Accurate working of entire system
False data injection attack (FDIA)	Injecting false data in the system.	Using secure encryption techniques
Unauthorized Access	Gaining access to a program, service, server, website, or other system by means of others' accounts or some other method.	Installing both spyware and virus protection programs Protect sensitive data including passwords and credit card information
Traffic Analysis	A special type of inference attack technique, monitoring communication patterns among entities of a system.	Dummy traffic approach to prevent traffic analysis attack employing Friend in the Middle (FiM)
Eavesdropping	Unauthorized interception of a private communication, for example instant message, videoconference, phone call, or fax transmission, in real-time .	Encryption
Masquerading	Attacker pretending to be an authorized user for gaining access to a system.	Enhanced key management systems
Reply attacks	Fraudulently or maliciously delaying or repeating a valid data transmission. Also known as playback attack.	Timestamping
Message modification and injection	Modifying data on a target machine or direct a message to an alternate destination by altering packet header addresses.	Use of web application firewall Regular software patches Suppressing error messages
Man-in-the-middle (MITM) attacks	Secret relaying and possible alternation of communication between two parties who believe they are directly communicating with each other by an attacker in the middle.	Secure or multipurpose internet mail extensions Authentication certificates
Flooding	A form of denial-of-service attack, executed by sending a succession of SYN (synchronize) requests to a target's system aimed at consuming enough server resources in order to make the system unresponsive to legitimate traffic.	Filtering Increasing backlog TCP half-open
Radio frequency(RF) jamming	Severe Denial-of-service attacks aimed at wireless medium. The attacker targets data packets of high importance by emitting radio frequency signals and do not follow underlying network architecture.	Anti-jamming technologies
Vulnerability attacks	Vulnerability is a weakness that allows an attacker to reduce the information assurance of a system. Vulnerabilities appear when three conditions meet: presence of system susceptibility or flaw, access of an attacker to that susceptibility, and the attacker's capability to exploit this susceptibility.	Host-based intrusion detection system Use of web proxy Use of accounts without administrative privileges

- Connectivity and exchange of information lies at the core of smart grid functionality, which made connected devices a corner-stone for this technology. These devices are called the ''internet of things (IoT)'', and enable the grid components to exchange data to maintain an up-to-date system status and receive commands to act as grid conditions change. IoT devices are increasing significantly in number each year, and are bringing unique opportunities and challenges with their wider implementation.
- IoT devices generate a huge amount of data, which cannot be handled through conventional analysis techniques. This massive data is termed as ''big data'', and it motivated the move towards new data analysis techniques. Big data generated from IoT devices are also exposed to security threats, and that have attracted a lot of attention as well.
- Machine learning is a useful way to sift through big data, and extract useful information that can extensively aid in demand and generation pattern recognition, generation

forecasting, control etc. A number of methods have already been presented in existing literature, and more novel techniques are being worked on for enhanced performance in specific use cases.

- Every sector of the smart grid – generation, transmission, and distribution – are in significant risk of cyber-attacks, and many such attacks have already been carried out. Security of data is thus a major concern in smart grid, and significant amount of work has already been conducted on detection of cyber-security threats and protection mechanisms to counter them. Many of these counter-measures have used machine learning techniques, as conventional methods are often useless in the new data-centric, non-linear system.

Based on this study, the following can be stated for future works in this field:

- The viability of the current grid infrastructure need to be validated through approaches such as mathematical modeling to find out the optimum timeframe and technological approach to move towards the smart grid architecture.
- The challenges in transitioning to the renewable energy-centric smart grid, and their feasible solutions need to be investigated. Possible business models, government initiatives, and their approach to implementation of smart grid can also be studied.
- IoT devices can be worked on to make them more compact, cheap, energy efficient, and robust. Advanced communication protocols can also be investigated to improve throughput and security. Monitoring schemes of power generation facility, pumps, and turbines can be further developed.
- Better forecasting techniques for demand and generation, especially renewable energy generation is essential for proper operation of renewable-energy based smart grids.
- Machine learning algorithms can be developed to meet power quality standards in a smart grid using the available data. Machine learning algorithms can also be applied in wind-solar hybrid system to further utilize our available resources.
- More research is required to develop viable solutions for other security concerns such as physical threats, network attacks and encryption attacks. Communication systems also need to be more efficient, with more protective measure.

VIII. CONCLUSION

The electricity grid is transitioning towards an IoT-based, connected smart grid, and with the benefits of such a system, concerns are also emerging that were unprecedented until now. The big data generated in the smart grid is requiring novel analysis techniques such as machine learning methods for proper handling and data extraction. The connected devices, and the data they generate are also bringing forth the dire necessities of proper protection, as they are being targeted to attacks of varying magnitudes which highlighted the lack of proper counter-measures in place. In an attempt to present an overall picture of these issues, this paper had presented a brief timeline of the grid's journey to the smart grid, and how internet of things (IoT) had become a part and parcel of the electricity grid. Challenges associated with IoT-generated big data, namely their analysis and protection, as well as other security concerns in the smart grid had also been discussed. The outcomes of this study had been presented finally with future research directions outlined briefly to aid researchers in this field.

REFERENCES

[1] L. M. Camarinha-Matos, "Collaborative smart grids—A survey on trends," *Renew. Sustain. Energy Rev.*, vol. 65, pp. 283–294, Nov. 2016.

[2] R. Bayindir, E. Hossain, E. Kabalci, and R. Perez, "A comprehensive study on microgrid technology," *Int. J. Renew. Energy Res.*, vol. 4, no. 4, pp. 1094–1107, 2014.

[3] E. Hossain, R. Perez, S. Padmanaban, and P. Siano, "Investigation on the development of a sliding mode controller for constant power loads in microgrids," *Energies*, vol. 10, no. 8, p. 1086, 2017.

[4] E. Hossain, E. Kabalci, R. Bayindir, and R. Perez, "Microgrid testbeds around the world: State of art," *Energy Convers. Manage.*, vol. 86, pp. 132–153, Oct. 2014.

[5] W. Y. Chiu, H. Sun, and H. V. Poor, "A multiobjective approach to multimicrogrid system design," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2263–2272, Sep. 2015.

[6] J. Rodríguez-Molina, M. Martínez-Núñez, J.-F. Martínez, and W. Pérez-Aguiar, "Business models in the smart grid: Challenges, opportunities and proposals for prosumer profitability," *Energies*, vol. 7, no. 9, pp. 6142–6171, 2014.

[7] J. Yuan, J. Shen, L. Pan, C. Zhao, and J. Kang, "Smart grids in China," *Renew. Sustain. Energy Rev.*, vol. 37, pp. 896–906, Sep. 2014.

[8] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of Internet of Things for smart home: Challenges and solutions," *J. Cleaner Prod.*, vol. 140, no. 3, pp. 1454–1464, 2017.

[9] D. S. Markovic, I. Branovic, and R. Popovic, "Smart grid and nanotechnologies: A solution for clean and sustainable energy," *Energy Emission Control Technol*, vol. 3, pp. 1–13, Jan. 2015.

[10] W. Yu, G. Wen, X. Yu, Z. Wu, and J. Lü, "Bridging the gap between complex networks and smart grids," *J. Control Decision*, vol. 1, no. 1, pp. 102–114, 2014.

[11] S. Abdollahy, A. Mammoli, F. Cheng, A. Ellis, and J. Johnson, "Distributed compensation of a large intermittent energy resource in a distribution feeder," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Feb. 2013, pp. 1–6.

[12] J. A. P. Lopes, C. L. Moreira, and A. G. Madureira, "Defining control strategies for analysing microgrids islanded operation," in *Proc. IEEE Russia Power Tech*, Jun. 2005, pp. 1–7.

[13] H. Gharavi and R. Ghafurian, "Smart grid: The electric energy system of the future [scanning the issue]," *Proc. IEEE*, vol. 99, no. 6, pp. 917–921, Jun. 2011.

[14] A. Amato, R. Aversa, B. Di Martino, and S. Venticinque, "A cyber physical system of smart micro-grids," in *Proc. 19th Int. Conf. Netw.-Based Inf. Syst. (NBiS)*, 2016, pp. 165–172.

[15] M. Kezunovic, "Data analytics: Creating information and knowledge [guest editorial]," *IEEE Power Energy Mag.*, vol. 10, no. 5, pp. 14–23, Sep. 2012.

[16] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 425–436, Feb. 2016.

[17] J. Shishido, "Smart meter data quality insights," in *Proc. ACEEE Summer Study Energy Efficiency Buildings*, 2012, pp. 277–288.

[18] A. McKane, I. Rhyne, A. Lekov, L. Thompson, and M. A. Piette, "Automated demand response: The missing link in the electricity value chain," ACEEE Summer Study Energy Efficiency Buildings, Pacific Grove, CA, USA, Tech. Rep. LBNL-2736E, 2008.

[19] X. Yu, C. Cecati, T. Dillon, and M. G. Simões, "The new frontier of smart grids," *IEEE Ind. Electron. Mag.*, vol. 5, no. 3, pp. 49–63, Sep. 2011.

[20] S. Aman, Y. Simmhan, and V. K. Prasanna, "Holistic measures for evaluating prediction models in smart grids," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 475–488, Feb. 2015.

[21] S. Aman, M. Frincu, C. Chelmis, M. Noor, Y. Simmhan, and V. K. Prasanna, "Prediction models for dynamic demand response: Requirements, challenges, and insights," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2015, pp. 338–343.

[22] M. Manfren, "Multi-commodity network flow models for dynamic energy management—Mathematical formulation," *Energy Procedia*, vol. 14, pp. 1380–1385, Mar. 2012.

[23] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, "Big data analytics for dynamic energy management in smart grids," *Big Data Res.*, vol. 2, pp. 94–101, Sep. 2015.

[24] J. Yuan, S. Sun, J. Shen, Y. Xu, and C. Zhao, "Wind power supply chain in China," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 356–369, Nov. 2014.

[25] D. A. Powner, *Electricity Grid Modernization: Progress Being Made on Cybersecurity Guidelines, But Key Challenges Remain to be Addressed*. Darby, PA, USA: DIANE Publishing, 2011.

[26] S. Simitis, "From the market to the polis: The EU directive on the protection of personal data," *Iowa L. Rev.*, vol. 80, part 3, pp. 445–469, 1994.

[27] J. Hu and A. V. Vasilakos, "Energy big data analytics and security: Challenges and opportunities," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2423–2436, Sep. 2016.

[28] P. Leitão, S. Karnouskos, L. Ribeiro, J. Lee, T. Strasser, and A. W. Colombo, "Smart agents in industrial cyber–physical systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1086–1101, May 2016.

[29] D. Apple, R. Elmoudi, I. Grinberg, S. M. Macho, and M. Safiuddin, *Foundations of Smart Grid*, 1st ed. Hampton, NH, USA: Pacific Crest, 2013, ch. 2, pp. 6–11.

[30] E. Hossain, R. Perez, A. Nasiri, and S. Padmanaban, "A comprehensive review on constant power loads compensation techniques," *IEEE Access*, vol. 6, pp. 33285–33305, 2018.

[31] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generat. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[32] A. Nordrum, "Popular Internet of Things forecast of 50 billion devices by 2020 is outdated," *IEEE Spectr.*, vol. 18, May 2018. [Online]. Available: https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated

[33] L. Liu, Y. Liu, L. Wang, A. Zomaya, and S. Hu, "Economical and balanced energy usage in the smart home infrastructure: A tutorial and new results," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 4, pp. 556–570, Dec. 2015.

[34] C. Gamarra, J. M. Guerrero, and E. Montero, "A knowledge discovery in databases approach for industrial microgrid planning," *Renew. Sustain. Energy Rev.*, vol. 60, pp. 615–630, Jul. 2016.

[35] M. Khan, B. N. Silva, and K. Han, "Internet of Things based energy aware smart home control system," *IEEE Access*, vol. 4, pp. 7556–7566, Oct. 2016.

[36] Y. Saleem, N. Crespi, M. H. Rehmani, and R. Copeland. (2017). "Internet of things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions." [Online]. Available: https://arxiv.org/abs/1704.08977

[37] J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah, and M. Sha, "An Internet of Things framework for smart energy in buildings: Designs, prototype, and experiments," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 527–537, Dec. 2015.

[38] M. Ligade. (Apr. 9, 2016). *Architecture for IOT Applications*. [Online]. Available: https://medium.com/@maheshwar.ligade/architecture-for-iot-applications-d50ece031d38

[39] J. Bauernberger. (Apr. 9, 2016). *Securing the Internet of Things*. [Online]. Available: https://blog.valbonne-consulting.com/2016/04/22/securing-the-internet-of-things/

[40] S. Thakurdesai. (Jul. 16, 2018). Smarter Metering—Now & in Future. Texas Instruments. [Online]. Available: http://electronicsmaker.com/em/admin/pdfs/free/Texas_Instruments.pdf

[41] A. Basit, G. A. S. Sidhu, A. Mahmood, and F. Gao, "Efficient and autonomous energy management techniques for the future smart homes," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 917–926, Mar. 2017.

[42] W. Shu-Wen, "Research on the key technologies of IOT applied on smart grid," in *Proc. Int. Conf. Electron., Commun. Control (ICECC)*, 2011, pp. 2809–2812.

[43] Y. F. Wang, W. M. Lin, T. Zhang, and Y. Y. Ma, "Research on application and security protection of Internet of Things in smart grid," in *Proc. IET Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, 2012, pp. 1–5.

[44] Q. Ou, Y. Zhen, X. Li, Y. Zhang, and L. Zeng, "Application of Internet of Things in smart grid power transmission," in *Proc. 3rd FTRA Int. Conf. Mobile, Ubiquitous, Intell. Comput. (MUSIC)*, 2012, pp. 96–100.

[45] M. M. Rogers *et al.*, "HERO: A smart-phone application for location based emissions estimates," *Sustain. Comput., Inf. Syst.*, vol. 8, pp. 3–7, Dec. 2015.

[46] B. Martinez, M. Montón, I. Vilajosana, and J. D. Prades, "The power of models: Modeling power consumption for IoT devices," *IEEE Sensors J.*, vol. 15, no. 10, pp. 5777–5789, Oct. 2015.

[47] J. Lin *et al.*, "Situation awareness of active distribution network: Roadmap, technologies, and bottlenecks," *CSEE J. Power Energy Syst.*, vol. 2, no. 3, pp. 35–42, 2016.

[48] V. Madani *et al.*, "Distribution automation strategies challenges and opportunities in a changing landscape," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2157–2165, Jul. 2015.

[49] S. A. A. Amin, A. Ali-Eldin, and H. A. Ali, "A context-aware dispatcher for the Internet of Things: The case of electric power distribution systems," *Comput. Elect. Eng.*, vol. 52, pp. 183–198, Mar. 2016.

[50] A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using IoT and big data analytics approach," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 426–434, Nov. 2017.

[51] F. Un-Noor, S. Padmanaban, L. Mihet-Popa, M. N. Mollah, and E. Hossain, "A comprehensive study of key electric vehicle (EV) components, technologies, challenges, impacts, and future direction of development," *Energies*, vol. 10, no. 8, p. 1217, 2017.

[52] Z. Zhu, S. Lambotharan, W. H. Chin, and Z. Fan, "Overview of demand management in smart grid and enabling wireless communication technologies," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 48–56, Jun. 2012.

[53] (Mar. 1, 2015). *RTI Presents 'How to Architect Microgrids for the Industrial Internet of Things' Complimentary Webinar*. [Online]. Available: https://www.rti.com/news/how-to-architect-microgrids-webinar

[54] M. M. Rana and L. Li, "Microgrid state estimation and control for smart grid and Internet of Things communication network," *Electron. Lett.*, vol. 51, no. 2, pp. 149–151, 2015.

[55] M. M. Rana and L. Li, "An overview of distributed microgrid state estimation and control for smart grids," *Sensors*, vol. 15, no. 2, pp. 4302–4325, 2015.

[56] S. Kinney. (Mar. 1, 2016). *The Smart Grid and Its Key Role in the Industrial IoT*. [Online]. Available: https://enterpriseiotinsights.com/20160418/channels/use-cases/smart-grids-key-role-industrial-iot

[57] R. Bikmetov, M. Y. A. Raja, and T. U. Sane, "Infrastructure and applications of Internet of Things in smart grids: A survey," in *Proc. North Amer. Power Symp. (NAPS)*, 2017, pp. 1–6.

[58] H. Mortaji, S. H. Ow, M. Moghavvemi, and H. A. F. Almurib, "Load shedding and smart-direct load control using Internet of Things in smart grid demand response management," *IEEE Trans. Ind. Appl.*, vol. 53, no. 6, pp. 5155–5163, Nov./Dec. 2017.

[59] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 269–283, Feb. 2017.

[60] A. F. Taha, J. Qi, J. Wang, and J. H. Panchal, "Risk mitigation for dynamic state estimation against cyber attacks and unknown inputs," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 886–899, Mar. 2018.

[61] Y. Seyedi and H. Karimi, "Coordinated protection and control based on synchrophasor data processing in smart distribution networks," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 634–645, Jan. 2018.

[62] Y. Zhang, Y. Xu, and Z. Y. Dong, "Robust classification model for PMU-based on-line power system DSA with missing data," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 18, pp. 4484–4491, 2017.

[63] B. K. Tannahill and M. Jamshidi, "System of systems and big data analytics—Bridging the gap," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 2–15, 2014.

[64] D. Tayal, "Disruptive forces on the electricity industry: A changing landscape for utilities," *Electr. J.*, vol. 29, no. 7, pp. 13–17, 2016.

[65] K. Wang *et al.*, "Wireless big data computing in smart grid," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 58–64, Apr. 2017.

[66] N. Zhang, Y. Yan, S. Xu, and W. Su, "A distributed data storage and processing framework for next-generation residential distribution systems," *Electr. Power Syst. Res.*, vol. 116, pp. 174–181, Nov. 2014.

[67] A. Paul, A. Ahmad, M. M. Rathore, and S. Jabbar, ‘‘SmartBuddy: Defining human behaviors using big data analytics in social Internet of Things,’’ *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 68–74, May 2016.

[68] N. Uribe-Pérez, L. Hernández, D. De la Vega, and I. Angulo, ‘‘State of the art and trends review of smart metering in electricity grids,’’ *Appl. Sci.*, vol. 6, no. 3, p. 68, 2016.

[69] Z. Guan, J. Li, L. Wu, Y. Zhang, J. Wu, and X. Du, ‘‘Achieving efficient and secure data acquisition for cloud-supported Internet of Things in smart grid,’’ *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1934–1944, Dec. 2017.

[70] B. K. Tannahill and M. Jamshidi, ‘‘Big data analytic paradigms-from PCA to deep learning,’’ in *Proc. Assoc. Adv. Artif. Intell. Symp. (AAAI)*, 2014, pp. 84–90.

[71] A. Emrouznejad, *Big Data Optimization: Recent Developments and Challenges*, vol. 18. Cham, Switzerland: Springer, 2016.

[72] W. Pedrycz and S.-M. Chen, *Data Science and Big Data: An Environment of Computational Intelligence*, vol. 24. Cham, Switzerland: Springer, 2017.

[73] L. Rodríguez-Mazahua, C.-A. Rodríguez-Enríquez, J. L. Sánchez-Cervantes, J. Cervantes, J. L. García-Alcaraz, and G. Alor-Hernández, ‘‘A general perspective of big data: Applications, tools, challenges and trends,’’ *J. Supercomput.*, vol. 72, no. 8, pp. 3073–3113, 2016.

[74] C. Miller, M. Martin, D. Pinney, and G. Walker, *Achieving a Resilient and Agile Grid*. Arlington, VA, USA: National Rural Electric Cooperative Association, 2014.

[75] J. Giri, ‘‘Proactive management of the future grid,’’ *IEEE Power Energy Technol. Syst. J.*, vol. 2, no. 2, pp. 43–52, Jun. 2015.

[76] S. Aman, Y. Simmhan, and V. K. Prasanna, ‘‘Energy management systems: State of the art and emerging trends,’’ *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 114–119, Jan. 2013.

[77] M. T. Burr *et al.*, ‘‘Emerging models for microgrid finance: Driven by the need to deliver value to end users,’’ *IEEE Electrific. Mag.*, vol. 2, no. 1, pp. 30–39, Mar. 2014.

[78] L. Collins and J. K. Ward, ‘‘Real and reactive power control of distributed PV inverters for overvoltage prevention and increased renewable generation hosting capacity,’’ *Renew. Energy*, vol. 81, pp. 464–471, Sep. 2015.

[79] J. Soares, N. Borges, M. A. F. Ghazvini, Z. Vale, and P. B. de Moura Oliveira, ‘‘Scenario generation for electric vehicles’ uncertain behavior in a smart city environment,’’ *Energy*, vol. 111, pp. 664–675, Sep. 2016.

[80] E.-K. Lee, W. Shi, R. Gadh, and W. Kim, ‘‘Design and implementation of a microgrid energy management system,’’ *Sustainability*, vol. 8, no. 11, p. 1143, 2016.

[81] N.-C. Yang and W.-C. Tseng, ‘‘Adaptive three-phase power-flow solutions for smart grids with plug-in hybrid electric vehicles,’’ *Int. J. Elect. Power Energy Syst.*, vol. 64, pp. 1166–1175, Jan. 2015.

[82] K. Zhou, S. Yang, and Z. Shao, ‘‘Energy Internet: The business perspective,’’ *Appl. Energy*, vol. 178, pp. 212–222, Sep. 2016.

[83] M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins, ‘‘A reinforcement learning approach to autonomous decision-making in smart electricity markets,’’ *Mach. Learn.*, vol. 92, no. 1, pp. 5–39, 2013.

[84] 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, Apr. 2016.

[85] K. Zhou and S. Yang, ‘‘Demand side management in China: The context of China’s power industry reform,’’ *Renew. Sustain. Energy Rev.*, vol. 47, pp. 954–965, Jul. 2015.

[86] 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, Apr. 2016.

[87] (Apr. 10, 2018). *How Can We Differentiate Big Data Analytics From Statistical Predictive Modeling Techniques?* [Online]. Available: http://www.zarantech.com/blog/can-differentiate-big-data-analytics-statistical-predictive-modelling-techniques/

[88] S. Mujawar and A. Joshi, ‘‘Data analytics types tools and their comparison,’’ in *Proc. IIJARCE*, vol. 4, 2015, pp. 488–491.

[89] P. Hassani. (Apr. 10, 2016). *An Insight Into 26 Big Data Analytic Techniques: Part 1*. [Online]. Available: https://blogs.systweak.com/2016/11/an-insight-into-26-big-data-analytic-techniques-part-1/

[90] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao, and H. Liu, ‘‘A big data architecture design for smart grids based on random matrix theory,’’ *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 674–686, Mar. 2017.

[91] S. Ryu, J. Noh, and H. Kim, ‘‘Deep neural network based demand side short term load forecasting,’’ *Energies*, vol. 10, no. 1, p. 3, 2017.

[92] I. M. Coelho, V. N. Coelho, E. J. da S. Luz, L. S. Ochi, F. G. Guimarães, and E. Rios, ‘‘A GPU deep learning metaheuristic based model for time series forecasting,’’ *Appl. Energy*, vol. 201, pp. 412–418, Sep. 2017.

[93] R. J. Bessa, A. Trindade, C. S. P. Silva, and V. Miranda, ‘‘Probabilistic solar power forecasting in smart grids using distributed information,’’ *Int. J. Electr. Power Energy Syst.*, vol. 72, pp. 16–23, Nov. 2015.

[94] (Apr. 10, 2018). *Top 10 Data Science, Machine Learning Methods Used, 2017*. [Online]. Available: https://www.kdnuggets.com/2017/12/top-data-science-machine-learning-methods.html

[95] V. G. Agelidis *et al.*, ‘‘Unlocking the smart grid: An Australian industry-university collaborative effort to address skill gaps,’’ in *Proc. IEEE PES Innov. Smart Grid Technol. Asia (ISGT)*, Nov. 2011, pp. 1–8.

[96] A. Sanchez and W. Rivera, ‘‘Big data analysis and visualization for the smart grid,’’ in *Proc. IEEE Int. Congr. Big Data (BigData Congr.)*, Jun. 2017, pp. 414–418.

[97] M. He, S. Murugesan, and J. Zhang, ‘‘Multiple timescale dispatch and scheduling for stochastic reliability in smart grids with wind generation integration,’’ in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 461–465.

[98] D. Niyato, E. Hossain, and A. Fallahi, ‘‘Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: Performance analysis and optimization,’’ *IEEE Trans. Mobile Comput.*, vol. 6, no. 2, pp. 221–236, Feb. 2007.

[99] H. Jiang, X. Dai, D. Gao, J. Zhang, Y. Zhang, and E. Muljadi, ‘‘Spatial-temporal synchrophasor data characterization and analytics in smart grid fault detection, identification, and impact causal analysis,’’ *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2525–2536, Sep. 2016.

[100] D. Li and S. K. Jayaweera, ‘‘Machine-learning aided optimal customer decisions for an interactive smart grid,’’ *IEEE Syst. J.*, vol. 9, no. 4, pp. 1529–1540, Dec. 2015.

[101] Y. Tang, J. Yang, J. Yan, and H. He, ‘‘Intelligent load frequency controller using GrADP for island smart grid with electric vehicles and renewable resources,’’ *Neurocomputing*, vol. 170, no. 1, pp. 406–416, 2015.

[102] X. Fang, D. Yang, and G. Xue, ‘‘Online strategizing distributed renewable energy resource access in islanded microgrids,’’ in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6.

[103] D. O’Neill, M. Levorato, A. Goldsmith, and U. Mitra, ‘‘Residential demand response using reinforcement learning,’’ in *Proc. 1st IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2010, pp. 409–414.

[104] H. Wang, Q. Wang, P. Liu, and L. Sun, ‘‘Big data and intelligent agent based smart grid architecture,’’ in *Proc. IEEE Int. Conf. Agents (ICA)*, Jul. 2017, pp. 106–107.

[105] T. Chen, Y. Zhang, X. Wang, and G. B. Giannakis, ‘‘Robust workload and energy management for sustainable data centers,’’ *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 651–664, Mar. 2016.

[106] R. C. Green, L. Wang, and M. Alam, ‘‘Applications and trends of high performance computing for electric power systems: Focusing on smart grid,’’ *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 922–931, Jun. 2013.

[107] M. Ali, Z. Y. Dong, X. Li, and P. Zhang, ‘‘RSA-grid: A grid computing based framework for power system reliability and security analysis,’’ in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Jun. 2006, pp. 1–7.

[108] X. Wang, Q. Liang, J. Mu, W. Wang, and B. Zhang, ‘‘Physical layer security in wireless smart grid,’’ *Secur. Commun. Netw.*, vol. 8, pp. 2431–2439, Sep. 2015.

[109] B. Dickson. (Mar. 1, 2016). *Exploiting Machine Learning in Cybersecurity*. [Online]. Available: https://techcrunch.com/2016/07/01/exploiting-machine-learning-in-cybersecurity/

[110] M. Frincu, C. Chelmis, M. U. Noor, and V. Prasanna, ‘‘Accurate and efficient selection of the best consumption prediction method in smart grids,’’ in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2014, pp. 721–729.

[111] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, ‘‘Detecting stealthy false data injection using machine learning in smart grid,’’ *IEEE Syst. J.*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.

[112] W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu, and Z. Z. He, ‘‘The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review,’’ *Renew. Sustain. Energy Rev.*, vol. 44, pp. 466–472, Apr. 2015.

[113] A. Kusiak and Z. Zhang, ‘‘Adaptive control of a wind turbine with data mining and swarm intelligence,’’ *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 28–36, Jan. 2011.

[114] A. Mellit, S. A. Kalogirou, L. Hontoria, and S. Shaari, ‘‘Artificial intelligence techniques for sizing photovoltaic systems: A review,’’ *Renew. Sustain. Energy Rev.*, vol. 13, no. 2, pp. 406–419, 2009.

[115] M. Negnevitsky, P. Mandal, and A. K. Srivastava, "Machine learning applications for load, price and wind power prediction in power systems," in *Proc. 15th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, 2009, pp. 1–6.

[116] L. Wenyi, W. Zhenfeng, H. Jiguang, and W. Guangfeng, "Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM," *Renew. Energy*, vol. 50, pp. 1–6, Feb. 2013.

[117] Y. Xu, Z. Y. Dong, K. Meng, R. Zhang, and K. P. Wong, "Real-time transient stability assessment model using extreme learning machine," *IET Gener., Transmiss. Distrib.*, vol. 5, pp. 314–322, Mar. 2011.

[118] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, "Power system transient stability assessment based on big data and the core vector machine," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016.

[119] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[120] P. Zhao, J. Xia, Y. Dai, and J. He, "Wind speed prediction using support vector regression," in *Proc. 5th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2010, pp. 882–886.

[121] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renew. Energy*, vol. 52, pp. 118–127, Apr. 2013.

[122] H. A. Kazem, J. H. Yousif, and M. T. Chaichan, "Modeling of daily solar energy system prediction using support vector machine for Oman," *Int. J. Appl. Eng. Res.*, vol. 11, no. 20, pp. 10166–10172, 2016.

[123] J.-L. Chen, H.-B. Liu, W. Wu, and D.-T. Xie, "Estimation of monthly solar radiation from measured temperatures using support vector machines—A case study," *Renew. Energy*, vol. 36, no. 1, pp. 413–420, Jan. 2011.

[124] Y. Y. Chia, L. H. Lee, N. Shafiabady, and D. Isa, "A load predictive energy management system for supercapacitor-battery hybrid energy storage system in solar application using the support vector machine," *Appl. Energy*, vol. 137, pp. 588–602, Jan. 2015.

[125] F. Liberati and A. Di Giorgio, "Economic model predictive and feedback control of a smart grid prosumer node," *Energies*, vol. 11, no. 1, p. 48, 2017.

[126] F. Ucar, O. F. Alcin, B. Dandil, and F. Ata, "Power quality event detection using a fast extreme learning machine," *Energies*, vol. 11, no. 1, p. 145, 2018.

[127] L. Morales-Velazquez, R. de Jesus Romero-Troncoso, G. Herrera-Ruiz, D. Morinigo-Sotelo, and R. A. Osornio-Rios, "Smart sensor network for power quality monitoring in electrical installations," *Measurement*, vol. 103, pp. 133–142, Jun. 2017.

[128] B. Li, S. Gangadhar, S. Cheng, and P. K. Verma, "Predicting user comfort level using machine learning for smart grid environments," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Jan. 2011, pp. 1–6.

[129] T. Remani, E. A. Jasmin, and T. P. I. Ahamed, "Residential load scheduling with renewable generation in the smart grid: A reinforcement learning approach," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2018.2855689.

[130] S. Alshareef, S. Talwar, and W. G. Morsi, "A new approach based on wavelet design and machine learning for islanding detection of distributed generation," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1575–1583, Jul. 2014.

[131] H. Jiang *et al.*, "Big data-based approach to detect, locate, and enhance the stability of an unplanned microgrid islanding," *J. Energy Eng.*, vol. 143, no. 5, p. 04017045, 2017.

[132] S. Jurado, Á. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques," *Energy*, vol. 86, pp. 276–291, Jun. 2015.

[133] C. Cecati, J. Kolbusz, P. Różycki, P. Siano, and B. M. Wilamowski, "A novel RBF training algorithm for short-term electric load forecasting and comparative studies," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6519–6529, Oct. 2015.

[134] N. Balac, T. Sipes, N. Wolter, K. Nunes, B. Sinkovits, and H. Karimabadi, "Large scale predictive analytics for real-time energy management," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2013, pp. 657–664.

[135] D. François, "Methodology and standards for data analysis with machine learning tools," in *Proc. 16th Eur. Symp. Artif. Neural Netw. (ESANN)*, Bruges, Belgium, Apr. 2008.

[136] (Apr. 10, 2018). *AI and Machine Learning in Cyber Security*. [Online]. Available: https://towardsdatascience.com/ai-and-machine-learning-in-cyber-security-d6fbee480af0

[137] *Global Wind 2006 Report*, Global Wind Energy Council, Brussels, Belgium, 2007.

[138] H. Lund and W. Kempton, "Integration of renewable energy into the transport and electricity sectors through V2G," *Energy Policy*, vol. 36, no. 9, pp. 3578–3587, 2008.

[139] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012.

[140] F. Douak, F. Melgani, and N. Benoudjit, "Kernel ridge regression with active learning for wind speed prediction," *Appl. Energy*, vol. 103, pp. 328–340, Mar. 2013.

[141] A. Clifton, L. Kilcher, J. K. Lundquist, and P. Fleming, "Using machine learning to predict wind turbine power output," *Environ. Res. Lett.*, vol. 8, no. 2, p. 024009, 2013.

[142] Z. Zhou *et al.*, "Game-theoretical energy management for energy internet with big data-based renewable power forecasting," *IEEE Access*, vol. 5, pp. 5731–5746, 2017.

[143] C. W. Potter and M. Negnevitsky, "Very short-term wind forecasting for Tasmanian power generation," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 965–972, May 2006.

[144] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 273–284, Mar. 2006.

[145] B. Zhu, M. Y. Chen, N. Wade, and L. Ran, "A prediction model for wind farm power generation based on fuzzy modeling," in *Proc. Int. Conf. Environ. Sci. Eng.*, vol. 12, Part A, 2012, pp. 122–129.

[146] J. L. Torres, A. García, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Solar Energy*, vol. 79, no. 1, pp. 65–77, 2005.

[147] M. C. Mabel and E. Fernandez, "Analysis of wind power generation and prediction using ANN: A case study," *Renew. Energy*, vol. 33, no. 5, pp. 986–992, 2008.

[148] R. E. Abdel-Aal, M. A. Elhadidy, and S. M. Shaahid, "Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks," *Renew. Energy*, vol. 34, no. 7, pp. 1686–1699, 2009.

[149] M. Yesilbudak, S. Sagiroglu, and I. Colak, "A new approach to very short term wind speed prediction using *k*-nearest neighbor classification," *Energy Convers. Manage.*, vol. 69, pp. 77–86, May 2013.

[150] L. Li, Y.-Q. Liu, Y.-P. Yang, S. Han, and Y.-M. Wang, "A physical approach of the short-term wind power prediction based on CFD precalculated flow fields," *J. Hydrodyn. B*, vol. 25, no. 1, pp. 56–61, 2013.

[151] K. Mohammadi, S. Shamshirband, P. L. Yee, D. Petković, M. Zamani, and S. Ch, "Predicting the wind power density based upon extreme learning machine," *Energy*, vol. 86, pp. 232–239, Jun. 2015.

[152] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[153] S. Wu, Y. Wang, and S. Cheng, "Extreme learning machine based wind speed estimation and sensorless control for wind turbine power generation system," *Neurocomputing*, vol. 102, pp. 163–175, Feb. 2013.

[154] M. Mohandes, S. Rehman, and S. M. Rahman, "Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS)," *Appl. Energy*, vol. 88, no. 11, pp. 4024–4032, 2011.

[155] Y. Bao, H. Wang, and B. Wang, "Short-term wind power prediction using differential EMD and relevance vector machine," *Neural Comput. Appl.*, vol. 25, no. 2, pp. 283–289, Aug. 2014.

[156] A. Ul Haque, P. Mandal, J. Meng, and M. Negnevitsky, "Wind speed forecast model for wind farm based on a hybrid machine learning algorithm," *Int. J. Sustain. Energy*, vol. 34, no. 1, pp. 38–51, 2015.

[157] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-term wind-power prediction based on wavelet transform–support vector machine and statistic-characteristics analysis," *IEEE Trans. Ind. Appl.*, vol. 48, no. 4, pp. 1136–1141, Jul./Aug. 2012.

[158] I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renew. Energy*, vol. 46, pp. 241–247, Oct. 2012.

[159] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, and A. M. Agogino, "Diagnosing wind turbine faults using machine learning techniques applied to operational data," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2016, pp. 1–8.

[160] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Appl. Energy*, vol. 98, pp. 574–583, Oct. 2012.

[161] S. Fan, J. R. Liao, R. Yokoyama, L. Chen, and W. J. Lee, "Forecasting the wind generation using a two-stage network based on meteorological information," *IEEE Trans. Energy Convers.*, vol. 24, no. 2, pp. 474–482, Jun. 2009.

[162] D. Lee and R. Baldick, "Short-term wind power ensemble prediction based on Gaussian processes and neural networks," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 501–510, Jan. 2014.

[163] S. Salcedo-Sanz, A. Pastor-Sánchez, L. Prieto, A. Blanco-Aguilera, and R. García-Herrera, "Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—Extreme learning machine approach," *Energy Convers. Manage.*, vol. 87, pp. 10–18, Nov. 2014.

[164] Y. Zhang, K. Liu, L. Qin, and X. An, "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods," *Energy Convers. Manage.*, vol. 112, pp. 208–219, Mar. 2016.

[165] W.-C. Yeh, Y.-M. Yeh, P.-C. Chang, Y.-C. Ke, and V. Chung, "Forecasting wind power in the Mai Liao Wind Farm based on the multi-layer perceptron artificial neural network model with improved simplified swarm optimization," *Int. J. Elect. Power Energy Syst.*, vol. 55, pp. 741–748, Feb. 2014.

[166] R. Rahmani, R. Yusof, M. Seyedmahmoudian, and S. Mekhilef, "Hybrid technique of ant colony and particle swarm optimization for short term wind energy forecasting," *J. Wind Eng. Ind. Aerodyn.*, vol. 123, pp. 163–170, Dec. 2013.

[167] J. Wang, Y. Wang, and Y. Li, "A novel hybrid strategy using three-phase feature extraction and a weighted regularized extreme learning machine for multi-step ahead wind speed prediction," *Energies*, vol. 11, no. 2, p. 321, 2018.

[168] L. Xiao, X. Xiao, C. Dai, M. Pengy, L. Wang, and H. V. Poor. (2018). "Reinforcement learning-based energy trading for microgrids." [Online]. Available: https://arxiv.org/abs/1801.06285

[169] L. Xiao, N. B. Mandayam, and H. V. Poor, "Prospect theoretic analysis of energy exchange among microgrids," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 63–72, Jan. 2015.

[170] "Grid integration of large-capacity renewable energy sources and use of large-capacity Electrical Energy Storage," Project Team under Market Strategy Board (MSB) IEC, International Electrotechnical Commission, Geneva, Switzerland, IEC White Paper, 2012. [Online]. Available: https://www.iec.ch/whitepaper/pdf/iecWP-gridintegrationlargecapacity-LR-en.pdf

[171] G. Masson, S. Orlandi, and M. Rekinge, "Global market outlook for photovoltaics 2014–2018," Eur. Photovolt. Ind. Assoc., Brussels, Belgium, May 2014. [Online]. Available: https://helapco.gr/wp-content/uploads/EPIA_Global_Market_Outlook_for_Photovoltaics_2014-2018_Medium_Res.pdf

[172] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," *Solar Energy*, vol. 85, pp. 746–756, May 2011.

[173] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," *Solar Energy*, vol. 84, pp. 1772–1781, Oct. 2010.

[174] D. Picault, B. Raison, S. Bacha, J. de la Casa, and J. Aguilera, "Forecasting photovoltaic array power production subject to mismatch losses," *Solar Energy*, vol. 84, no. 7, pp. 1301–1309, 2010.

[175] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881–2893, 2011.

[176] J. Lee and G.-L. Park, "Climate effect analysis on solar energy generation in Jeju City," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 19, pp. 11692–11697, 2016.

[177] Y. Gala, Á. Fernández, J. Díaz, and J. R. Dorronsoro, "Hybrid machine learning forecasting of solar radiation values," *Neurocomputing*, vol. 176, pp. 48–59, Feb. 2016.

[178] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, May 2017.

[179] A. Chaouachi, R. M. Kamel, and K. Nagasaka, "A novel multi-model neuro-fuzzy-based MPPT for three-phase grid-connected photovoltaic system," *Solar Energy*, vol. 84, no. 12, pp. 2219–2229, 2010.

[180] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka, "Multiobjective intelligent energy management for a microgrid," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1688–1699, Apr. 2013.

[181] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, Jul. 2014.

[182] M. R. Hossain, A. M. T. Oo, and A. Ali, "The combined effect of applying feature selection and parameter optimization on machine learning techniques for solar power prediction," *Amer. J. Energy Res.*, vol. 1, no. 1, pp. 7–16, 2013.

[183] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, "Machine learning for solar irradiance forecasting of photovoltaic system," *Renew. Energy*, vol. 90, pp. 542–553, May 2016.

[184] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, and M. Sánchez-Girón, "Daily global solar radiation prediction based on a hybrid coral reefs optimization—Extreme learning machine approach," *Solar Energy*, vol. 105, pp. 91–98, Jul. 2014.

[185] S. Salcedo-Sanz, C. Casanova-Mateo, J. Muñoz-Marí, and G. Camps-Valls, "Prediction of daily global solar irradiation using temporal Gaussian processes," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1936–1940, Nov. 2014.

[186] Z. Li, S. M. M. Rahman, R. Vega, and B. Dong, "A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting," *Energies*, vol. 9, no. 1, p. 55, 2016.

[187] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A survey on security threats and defensive techniques of machine learning: A data driven view," *IEEE Access*, vol. 6, pp. 12103–12117, 2018.

[188] S. Tan, D. De, W.-Z. Song, J. Yang, and S. K. Das, "Survey of security advances in smart grid: A data driven approach," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 397–422, 1st Quart., 2017.

[189] Z. Fan *et al.*, "Smart grid communications: Overview of research challenges, solutions, and standardization activities," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 21–38, 1st Quart., 2013.

[190] I. Andrea, C. Chrysostomou, and G. Hadjichristofi, "Internet of Things: Security vulnerabilities and challenges," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2015, pp. 180–187.

[191] C. McLellan. (Apr. 9, 2016). *Cybersecurity Predictions for 2016: How Are They Doing?* [Online]. Available: https://www.zdnet.com/article/ransomware-is-now-big-business-on-the-dark-web-and-malware-developers-are-cashing-in/

[192] C. McLellan. (Apr. 9, 2017). *Cybersecurity in an IoT and Mobile World: The Key Trends*. [Online]. Available: https://www.zdnet.com/article/cybersecurity-in-an-iot-and-mobile-world-the-key-trends/

[193] A. Anwar and A. N. Mahmood. (2014). "Cyber security of smart grid infrastructure." [Online]. Available: https://arxiv.org/abs/1401.3936

[194] *Smart Grid Cyber Security*, document EURELECTRIC D/2016/12.105/64, Dec. 2016.

[195] *Introduction to NISTIR 7628 Guidelines for Smart Grid Cyber Security*, NIST Special Publication 154, The Smart Grid Interoperability Panel Cyber Security Working Group, 2010.

[196] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[197] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Covert cyber assault detection in smart grid networks utilizing feature selection and Euclidean distance-based machine learning," *Appl. Sci.*, vol. 8, no. 5, p. 772, 2018.

[198] 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, Dec. 2011.

[199] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1–12, Oct. 2015.

[200] Y. Wang, M. M. Amin, J. Fu, and H. B. Moussa, "A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids," *IEEE Access*, vol. 5, pp. 26022–26033, 2017.

[201] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Feature selection–based detection of covert cyber deception assaults in smart grid communications networks using machine learning," *IEEE Access*, vol. 6, pp. 27518–27529, 2018.

[202] L. Wei, L. P. Rondon, A. Moghadasi, and A. I. Sarwat. (2018). "Review of cyber-physical attacks and counter defense mechanisms for advanced metering infrastructure in smart grid." [Online]. Available: https://arxiv.org/abs/1805.07422

[203] J. Abdella and K. Shuaib, "Peer to peer distributed energy trading in smart grids: A survey," *Energies*, vol. 11, no. 6, p. 1560, 2018.

[204] D. Tellbach and Y.-F. Li, "Cyber-attacks on smart meters in household nanogrid: Modeling, simulation and analysis," *Energies*, vol. 11, no. 2, p. 316, 2018.

[205] 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, 2011.

[206] S. Paul and Z. Ni, "Vulnerability analysis for simultaneous attack in smart grid security," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Apr. 2017, pp. 1–5.

[207] S. Poudel, Z. Ni, and W. Sun, "Electrical distance approach for searching vulnerable branches during contingencies," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3373–3382, Jul. 2018.

[208] J. Pagliery. (Apr. 8, 2014). *Hackers Attacked the U.S. Energy Grid 79 Times This Year*. [Online]. Available: http://money.cnn.com/2014/11/18/technology/security/energy-grid-hack/index.html

[209] S. M. Amin and A. M. Giacomoni, "Smart grid, safe grid," *IEEE Power Energy Mag.*, vol. 10, no. 1, pp. 33–40, Jan./Feb. 2012.

[210] J. Pyper. (Apr. 8, 2018). *DOE Forms a New Office Dedicated to 'Energy Infrastructure Security.'* [Online]. Available: https://www.greentechmedia.com/articles/read/doe-new-office-energy-infrastructure-security-cybersecurity#gs.QxgPOh0

[211] J. Desjardins. (Apr. 8, 2017). *These Are the Countries Most (and Least) Prepared for Cyber Attacks*. [Online]. Available: http://www.visualcapitalist.com/countries-least-prepared-cyber-attacks/

[212] J. Santiago. (Apr. 8, 2015). *Top Countries Best Prepared Against Cyberattacks*. [Online]. Available: https://www.weforum.org/agenda/2015/07/top-countries-best-prepared-against-cyberattacks/

[213] A. B. N. Perlroth. (Apr. 8, 2018). *A Cyberattack Hobbles Atlanta, and Security Experts Shudder*. [Online]. Available: https://www.nytimes.com/2018/03/27/us/cyberattack-atlanta-ransomware.html

[214] R. McMillan. (Apr. 10, 2010). *Siemens: Stuxnet Worm Hit Industrial Systems*. [Online]. Available: https://www.computerworld.com/article/2515570/network-security/siemens–stuxnet-worm-hit-industrial-systems.html

[215] S. Abraham *et al.*, "Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," U.S.-Canada Power System Outage Task Force, Washington, DC, USA, Tech. Rep., Apr. 2004.

[216] E-ISAC, SANS, "Analysis of the cyber attack on the ukrainian power grid," ICS Defense Use Case 5, Mar. 2016. [Online]. Available: https://ics.sans.org/media/EISAC_SANS_Ukraine_DUC_5.pdf

[217] L. Dearden. (Apr. 10, 2017). *Ukraine Cyber Attack: Chaos as National Bank, State Power Provider and Airport Hit by Hackers*. [Online]. Available: https://www.independent.co.uk/news/world/europe/ukraine-cyber-attack-hackers-national-bank-state-power-company-airport-rozenko-pavlo-cabinet-a7810471.html

[218] G. Kalogridis, M. Sooriyabandara, Z. Fan, and M. A. Mustafa, "Toward unified security and privacy protection for smart meter networks," *IEEE Syst. J.*, vol. 8, no. 2, pp. 641–654, Jun. 2014.

[219] S. L. T. Ghazi. (Apr. 9, 2011). *The Increasing Importance of Security for the Smart Grid*. [Online]. Available: https://www.elp.com/articles/powergrid_international/print/volume-16/issue-4/features/the-increasing-importance-of-security-for-the-smart-grid.html

[220] 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, Sep. 2012.

[221] E. Pallotti and F. Mangiatordi, "Smart grid cyber security requirements," in *Proc. 10th Int. Conf. Environ. Elect. Eng. (EEEIC)*, May 2011, pp. 1–4.

[222] T. Sommestad, G. N. Ericsson, and J. Nordlander, "SCADA system cyber security—A comparison of standards," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2010, pp. 1–8.

[223] Q. Chen, D. Kaleshi, Z. Fan, and S. Armour, "Impact of smart metering data aggregation on distribution system state estimation," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1426–1437, Aug. 2016.

[224] M. A. Mustafa, N. Zhang, G. Kalogridis, and Z. Fan, "DEP2SA: A decentralized efficient privacy-preserving and selective aggregation scheme in advanced metering infrastructure," *IEEE Access*, vol. 3, pp. 2828–2846, 2015.

[225] K. Zhou and S. Yang, "A framework of service-oriented operation model of China's power system," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 719–725, Oct. 2015.

[226] B. Bitzer and E. S. Gebretsadik, "Cloud computing framework for smart grid applications," in *Proc. 48th Int. Univ. Power Eng. Conf. (UPEC)*, 2013, pp. 1–5.

[227] H. Kim, Y.-J. Kim, K. Yang, and M. Thottan, "Cloud-based demand response for smart grid: Architecture and distributed algorithms," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 398–403.

[228] X. Zhang, Z. Y. Dong, Z. Wan, C. Xiao, and F. Luo, "Quantum cryptography based cyber-physical security technology for smart grids," in *Proc. 10th Int. Conf. Adv. Power Syst. Control, Oper. Manage. (APSCOM)*, 2015, pp. 1–6.

[229] *Information Technology—Security Techniques—Information Security Management Systems—Requirements*, document ISO/IEC 27001:2013, ISO/IEC JTC 1/SC 27, International Organization for Standardization, 2013.

[230] (Apr. 8, 2018). *CISQ Consortium for IT Software Quality*. [Online]. Available: http://it-cisq.org/standards/

[231] (Apr. 8, 2018). *ISF Information Security Forum*. [Online]. Available: https://www.securityforum.org/

[232] *Information Security Forum Releases the Standard of Good Practice for Information Security 2016*, Inf. Secur. Forum Ltd., London, U.K., 2016.

[233] (Apr. 8, 2018). *NERC North American Electric Reliability Corporation*. [Online]. Available: https://www.nerc.com/Pages/default.aspx

[234] (Apr. 8, 2014). *Framework for Improving Critical Infrastructure Cybersecurity (1.0 ed.)*. [Online]. Available: https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf

[235] (Apr. 8, 2018). *Standard ISO/IEC 15408, CC v3.1. Release 4*. [Online]. Available: http://commoncriteria.pl/index.php/en/common-criteria-standard/standard-documentation/standard-iso-iec-15408-cc-v3-1-release-4

[236] (Apr. 8, 2018). *Common Criteria*. [Online]. Available: https://www.commoncriteriaportal.org/

[237] B. Fraser. (2003). *RFC 2196. Site Security Handbook. 1997*. [Online]. Available: http://www.faqs.org/rfcs/rfc2196.html

[238] International Society of Automation. (Apr. 8, 2012). *Applying ISA IEC 62443 to Control Systems—ISA Automation Week 2012: Security*. [Online]. Available: https://www.isa.org/store/applying-isa-iec-62443-to-control-systems-isa-automation-week-2012-security/122496

[239] M. Dworkin, "Recommendation for block cipher modes of operation, methods and techniques," Comput. Secur. Division, Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. SP 800-38A, Dec. 2001.

[240] E. Barker, *Recommendation for Triple Data Encryption Algorithm (TDEA) Block Cipher*, document SP 800-67 Rev. 2, NIST Special Publication, 2017.

[241] E. B. Barker, "FIPS PUB 186-4-federal information processing standards publication digital signature standard (DSS)," Inf. Technol. Lab., Nat. Inst. Standards Technol., Gaithersburg, MD, USA, 2009. [Online]. Available: https://csrc.nist.gov/csrc/media/publications/fips/186/3/archive/2009-06-25/documents/fips_186-3.pdf

[242] Q. H. Dang, "Secure hash standard," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Jul. 2015.

[243] M. J. Dworkin, "Recommendation for block cipher modes of operation: The CMAC mode for authentication," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. SP 800-38B, 2005.

[244] M. J. Dworkin, "Recommendation for block cipher modes of operation: The CCM mode for authentication and confidentiality," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. SP 800-38C, 2007.

[245] M. J. Dworkin, "Recommendation for block cipher modes of operation: Galois/counter mode (GCM) and GMAC," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. SP 800-38D, 2007.

[246] J. M. Turner, "The keyed-hash message authentication code (HMAC)," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. FIPS PUB 198-1, Jul. 2008. [Online]. Available: http://csrc.nist.gov/publications/fips/fips198-1/ FIPS-198-1_final.pdf

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

[248] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.

[249] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 303–336, 1st Quart., 2014.

[250] J. Singh and M. J. Nene, "A survey on machine learning techniques for intrusion detection systems," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 11, pp. 4349–4355, 2013.

[251] 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, Dec. 2011.

[252] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.

[253] S. Pan, T. Morris, and U. Adhikari, "Developing a hybrid intrusion detection system using data mining for power systems," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3104–3113, Nov. 2015.

[254] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[255] Y. Xin *et al.*, "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.

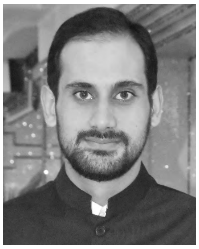
[256] V. B. Krishna, G. A. Weaver, and W. H. Sanders, "PCA-based method for detecting integrity attacks on advanced metering infrastructure," in *Proc. Int. Conf. Quant. Eval. Syst.*, 2015, pp. 70–85.

[257] S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, "Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure," *IEEE Control Syst.*, vol. 35, no. 1, pp. 66–81, Feb. 2015.

EKLAS HOSSAIN (M'09–SM'17) received the B.S. degree in electrical and electronic engineering from the Khulna University of Engineering and Technology, Bangladesh, in 2006, the M.S. degree in mechatronics and robotics engineering from the International Islamic University of Malaysia, Malaysia, in 2010, and the Ph.D. degree from the College of Engineering and Applied Science, University of Wisconsin Milwaukee, in 2016. He has been working in the area of distributed power systems and renewable energy integration for the last ten years, and he has published a number of research papers and posters in this field. He has been an Assistant Professor with the Department of Electrical Engineering and Renewable Energy, Oregon Tech, since 2015, where is currently involved in several research projects on renewable energy and grid-tied microgrid systems. His research interests include modeling, analysis, design, and control of power electronic devices; energy storage systems; renewable energy sources; the integration of distributed generation systems; microgrid and smart grid applications; and robotics and advanced control systems. He, with his dedicated research team, is looking forward to explore methods to make the electric power systems more sustainable, cost-effective, and secure through extensive research and analysis on energy storage, microgrid systems, and renewable energy sources. He currently serves as an Associate Editor of the IEEE Access.

IMTIAJ KHAN received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology, in 2017. His undergraduate thesis was on optimization and comparison of single-walled and double-walled carbon nanotube field effect transistors. His previous works were presented at the IEEE NANO 2017 and the IEEE TENCON 2017. He is currently involved in power quality improvement and noble manipulation of particles by optical force. His research interests include nanotechnology, plasmonics, photonics, and smart grids.

FUAD UN-NOOR received the B.Sc. degree in electrical and electronic engineering from the Khulna University of Engineering and Technology, in 2017. His previous works focused on electric vehicles and power systems. He is currently involved in smart grid/microgrid and energy storage systems. His research interests include electric vehicles, power and energy systems, renewable energy systems, smart grid/microgrid, and energy storage systems.



SARDER SHAZALI SIKANDER (M'12–SM'15) received the B.S. degree in electronics from International Islamic University Islamabad, in 2012, and the M.S. degree from the Department of Electrical Engineering, National University of Science and Technology Islamabad, in 2014. His research interests include robotics, power electronics converters, nonlinear control, and fuzzy logic control.

MD. SAMIUL HAQUE SUNNY received the B.Sc. degree in electrical and electronic engineering from the Khulna University of Engineering and Technology, in 2017. His undergraduate thesis was on high-performance state parameter observation with sensorless vector control of induction motor. His previous works were presented at ICEEICT 2016, ICAEE 2017, IC2IT 2017, EICT 2017, and IEEE R10-HTC. He is currently involved in data mining with AI algorithms, EEG signal for better BCI application, and structures of CNN for upgrading its performance in image recognition. His research interests include artificial intelligence, brain–computer interface, digital signal and image processing, data mining, power system stability, and robotics.