

Smart Grid Information Processes Using IoT and Big Data with Cloud and Edge Computing

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

12-11-2021 / 18-11-2021

CITATION

Ullah, Mehar; Narayanan, Arun; Wolff, Annika; Nardelli, Pedro (2021): Smart Grid Information Processes Using IoT and Big Data with Cloud and Edge Computing. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.16995025.v1>

DOI

[10.36227/techrxiv.16995025.v1](https://doi.org/10.36227/techrxiv.16995025.v1)

Smart Grid Information Processes Using IoT and Big Data with Cloud and Edge Computing

Mehar Ullah*, Arun Narayanan*, Annika Wolff*, Pedro Nardelli*

* LUT University, Lappeenranta, Finland

mehar.ullah@lut.fi

Abstract—Smart grid applications typically use cloud computing to address the computational requirements for efficient electricity delivery. Recently, the emerging Internet of Things (IoT) has resulted in increased number of devices connected to the smart grid, including consumer gadgets, measurement equipment, and electrical and electronic devices such as smart power converters, phasor measurement units, and smart meters. These heterogeneous devices that are present in all the four stages of a smart grid—generation, transmission, distribution, and consumption—generate huge amount of structured, semi structured and unstructured data. Gathering, storing, and processing such huge data volumes using cloud computing creates problems of bandwidth, latency, disaster recovery, and cost. To overcome these problems, we present a theoretical discussion on the usage of IoT, edge computing and big data to collect the data from smart grid, process the data using edge computing and big-data analytics, and use the data for smart grid maintenance, energy information and future decisions. Further, we highlight how edge computing and big-data approaches have mitigated the above mentioned problems by shifting the control, intelligence, and trust to the edge of the network.

Keywords—*IoT, Smart grid, big data, edge computing, cloud computing*

I. INTRODUCTION

The renewable energy-based smart grid (SG) concept emerged in the early 21st century, motivated by the necessity to increase renewable energy production in the electricity grid. The transformation of the 20th century traditional electric grid to modern 21st century smart electric grids was enabled by remarkable advancements in power electronics and internet and communication technologies (ICT). The integration of the new technologies enables efficient utilization of the energy production and consumption, providing opportunities for new energy resources like wind and solar etc., allowing the exchange of generated power from different sources and also bi-directional flow of power and communication. As a result, both utility companies and customers have been installing renewable energy sources (RES), such as solar and wind energies, inside the distribution grid [1]. Further, new technological paradigms such as Internet of Things (IoT) are influencing modern SG operations by improving communication, achieving better customer relationships, and handling the huge amount of data generated from the smart devices. In this sense, IoT is increasingly being used in SG applications for data gathering, communication, and smart analytics. For example, energy-based data analytics from the user to the utility can highly improve efficiency,

reduce congestion, and improve power-supply reliability in 100% renewable-energy-based SGs in the future [1].

Recently, many industries across sectors have shifted their business operations to some form of cloud computing since cloud computing provides some benefits such as flexibility, operability, and cost savings. As a result, enormous amounts of data generated by devices are being sent to the cloud servers for processing and analysis [2]. In the industrial environment, such an increase in data from devices create many problems. First, a large amount of data is transferred to the cloud for analysis, but most of it may be irrelevant to the operations. Thus, this data transmission ends up creating high traffic to a central repository and increases costs due to extra unnecessary storage. Second, important data that need to be sent with small latency's of between seconds and millisecond can be very important for crucial industrial operations may experience costly delays. Third, sending data to cloud and retrieving that data can be very costly [3].

Edge computing is used to overcome the aforementioned problems in cloud computing [4]. The benefits of edge computing are that it moves data analysis and services away from centralized servers and a lot of data analysis is performed at the source of data collection [3]. Edge computing analyze data on the spot and filters the important data in real time; this improves the speed of data analysis and the decision-making process [5]. Edge computing is proving huge benefits to (IoT)-enabled business, but nevertheless, cloud computing remains important because having a centralized location for the data storage and analysis still has many benefits. In particular, non-time-sensitive data can be sent to the cloud, for example, for deep analysis post-hoc using machine learning (ML) methods to improve industrial operations and strategies [6]. Table I shows some differences between cloud and edge computing [6], [7].

The comprehensive sensing and processing abilities of IoT support many technologies in SG. Further, the rapid increase in IoT-enabled devices can cause explosive growth in data generation, resulting in the so-called “big data” regime, where the system generates data that is so large, fast, or complex that non-traditional methods are required for processing it. The generation of big data in SGs makes the existing data-processing capacities ineffective as edge computing does not have all the resources sufficient for the complex and intelligent big-data analytics tasks [8]. Hence, it is important to include more modern

TABLE I
DIFFERENCE BETWEEN CLOUD COMPUTING AND EDGE COMPUTING.

Point of Difference	Cloud Computing	Edge Computing
Operations	Happens on the cloud platforms such as AWS, Azure, Google	Happens on the device itself or at the gateways
Benefits	Can store massive amount of data on scalable hosting on the cloud which can be accessible anytime on the Internet	Network can be scalable independently with each new device that is added to the system, possibly working as a federation
Suitable use case	Suitable for the operations with more tolerance in terms of latency and requires high levels of computing power	This is suitable for low latency applications and that allows for distributed data storage, leading to a scalable and cost effective hosting providers

big-data analytics to improve the data-processing capacity of IoT data [9]. Big-data analytics can be defined as the process by which the variety of IoT data are analyzed to find the trends, hidden patterns, unseen correlations and new information. This huge amount of data analysis and information gathering will provide benefits to companies for current and future effective decisions and will also provide benefits to the individual users [10].

In this paper, we present a theoretical framework using a combination of IoT, edge computing, big data, and analytics for the efficient collection of data from the huge number of devices connected to SG as well as the processing, storage, and visualization of the collected data. We elaborate on some benefits of edge computing and big data, such as latency, bandwidth, disaster recovery, and price in the entire SG system, starting from data collection to data visualization. The rest of this paper is organized as follows. In Section II, we describe related studies on big-data analytics in smart grids and explain the concepts of IoT, edge computing, and big data. In Section III, we explain how these technologies—IOT, edge computing, big data, and big-data analytics—work together. Finally, Section IV concludes the paper.

II. RELATED RESEARCH

In this section, we use systematic literature review structure, similar to [11], to answer the question “what type of technologies can be used to handle the massive amount of data from SG and obtain meaningful information from it, so that it can be used for better business decisions?” The main objectives are to identify technologies that can work together to effectively extract information from big data to fulfil the requirements of SG business and operations.

A. Smart Grid

The SG technology concept has emerged to improve the flexibility and efficiency of the traditional grid and provide new opportunities for new generation methods such as wind, solar, and other RES based generation. The SG is essentially an electrical network that consists of infrastructure, software, and hardware, which enables it for two-way communication between all parts of the system and participants and efficiently generate power and enable distribution in the supply chain. As a self-sufficient distributed system that can provide energy from different sources including renewable and storage, SG also enables the suppliers and consumers access to the control and management capabilities [12].

SG allows two-way communication between the generation and consumption side with the help of devices such as smart meters, smart appliances, battery energy storage systems, power electronics converters, and other energy efficient resources. SG uses computer technologies for the improvements in automation, communications such as information exchange between consumers, transformers, and generation plants, and connectivity between many components of the power network, e.g., power gathered from different generation plants [13].

The SG works differently than the traditional grid; the network structure of the smart grid is complex having two-way communication and two-way interactions between the devices and the participant in the supply chain. The operations involves many steps from power generation to consumption, as explained below [14].

Generation: Power is generated from distributed sources that can include traditional power plants and renewable sources such as solar and wind. Electric storage can be used for generation-side management, including consumption-integrated storage such as electric vehicles.

Transmission: The generated power is transmitted using a network of transmission lines substations, and distribution systems. In the SG, transmission comprises three interactive components—smart control centers, smart power transmission networks, and smart substations. The smart transmission networks are conceptually built on the existing electric transmission infrastructure and the current advanced technologies—sensing, computing, communication and signal processing—provide services such as power utilization, power quality, network security, and reliability.

Distribution: The generated power is transmitted using a network of transmission lines that connects via substations to distribution systems that cover smaller areas and deliver power directly to the consumer. Because of the presence of both centralized and distributed power generators in modern SGs, the distribution networks have two-way electricity transmission, or, in other words, bidirectional power flow.

Consumption: The power consumption in modern SGs is often controllable and manageable by the end user using smart meters, sensors in appliances, plugs, and smart sockets. The user can control and manage their electricity consumption by using mobile phone applications or web-site applications to monitor and control the power usage.

Control and management: SGs have the capability

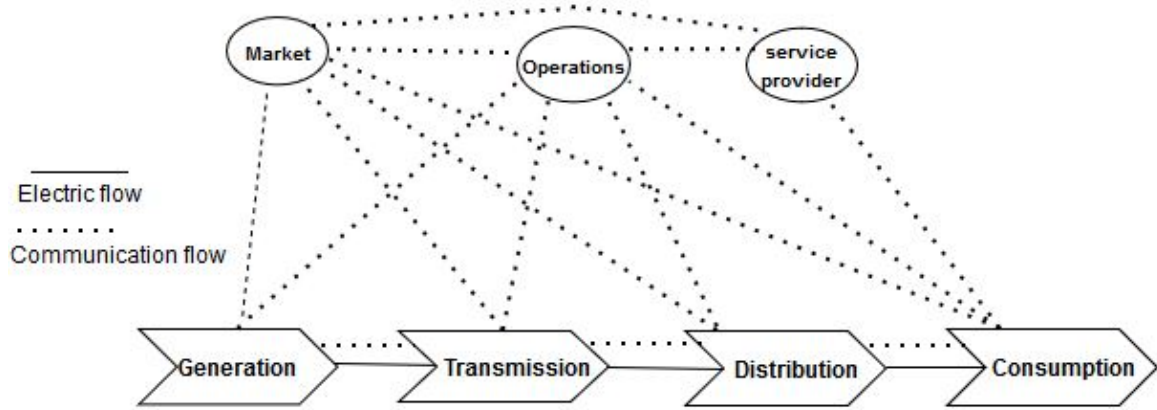


Fig. 1. Structure of Smart Grid (modified from [14]).

of control and management, and consumers, utility companies, and others in the energy industry can have a strong control on the energy usage and management. Data about the consumption and loads is generated from the connected homes, smart cities etc., and the information generated from that data is used by the companies and customers for their current and future decisions (using data analytics and visualization tools). For example, the energy companies can use the information for predictive maintenance; utility companies can use the information for demand and response programs; and residential users can use the information to reduce the energy consumption at the peak loads and reduce energy bills.

Storage: Electric storage is an important SG technology that enables generation-side management—households can store either extra produced energy or cheaper priced electricity, and later use it in the case of outage or when electricity is more expensive. Independent residential grids that are totally dependent on the renewable energy and generate a surplus of energy can store the surplus energy for future use.

B. Internet of Things

The term “Internet of Things”, or “IoT”, was introduced by the British technology entrepreneur Kevin Ashton in 1999 as the title of a presentation at Procter and Gamble. IoT and can be defined as small and complex systems that allow businesses, governments, and citizens to adopt and interconnect physical objects and virtual objects based on existing and evolving interoperable information and communication technologies [15]. IoT is a new technology paradigm that has emerged as a global network of machines and devices capable of interacting with each other and with the platform for collecting, analyzing, storing and visualizing data generated from the devices and machines using sensors, actuators, communications, and analytical tools. IoT is now playing a main role in SG by collecting data from all the main phases of SG, including generation, transmission, distribution, and consumption [16], [17].

C. Edge computing

Edge computing refers to the enabling technologies that allow computations to be performed at the edge of the network; from the cloud viewpoint, edge data is downstream data, and from the IoT services viewpoint, it is upstream data. In such a scenario, edge can be defined as any computing and networking resource that is between the data source and cloud. Smart phone is a simple example of an edge device, because it lies between the human user and the cloud. The aim of edge computing is to ensure that computations are performed at the proximity of data sources [2], [18].

The nodes at the network edge are performing many tasks such as data processing, caching, device management, and privacy protection to reduce the traffic from the devices to the cloud. In order to perform all these tasks in the network, the edge should be well designed to effectively meet the security, reliability and privacy protection requirements.

D. Big Data

In simple terms, big data can be defined as the collection of unstructured, structured, and semi-structured data generated by the social media, devices, sensors, software applications, and digital devices that are continuously generating data [19]. The data collected is so large that the normal conventional data processing software and techniques are not able to process it. Big data is characterized by the three main determinants, called as 3Vs of big data—volume, velocity, and variety. Volume is the huge amount of data generated that make the datasets too large for the normal database technology. This type of data is measured in larger units of data, such as terabytes, petabytes, and exabytes. Velocity is the speed with which the data is generated, processed, and moved around in real time. Variety is the type of data (nature of data), i.e., whether the data is structured or unstructured [20].

The main idea of IoT is to connect heterogeneous objects to the internet and collect data from these devices, analyze the collected data, and make future decisions. Recently, due to dramatic improvements in the technology and business digitization, the number of devices connected

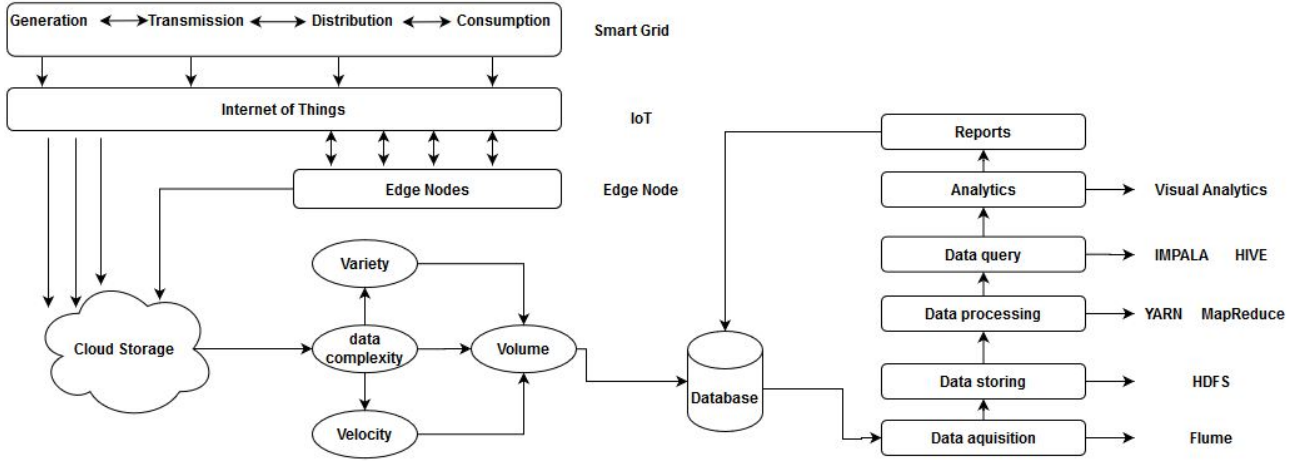


Fig. 2. IoT and Edge computing in Smart grids using Big data Analytics.

to IoT has increased tremendously; as a result, the amount of data has also increased tremendously so that there is a need to apply big data and big-data analytics to IoT. Big data and big-data analytics have high potential to extract meaningful information from the huge amount of data and improve the decision processes. The main requirements (functional and non-functional) of big data and analytics in IoT are explained below.

Connectivity: Connectivity in IoT is mostly ubiquitous with the heterogeneous objects in the network. Many objects are connected to internet via sensors in a smart environment. IoT services are mostly based on machine-to-machine (M2M) communication protocols that are required to handle a large number of streams, and it takes benefits directly from the cloud distributed storage and computing infrastructure [21]. The first and most important requirement of IoT is to provide reliable connectivity for big data and analytics. Reliable connectivity will provide big data and analytics the opportunity to efficiently combine and integrate the massive amounts of machine generated sensor data. Using the advanced wireless networks such as Wi-Fi and 4G/5G, many objects around us are able to connect to the computing and high performance infrastructure and facilitate the IoT services [22].

Storage: The amount of storage required for huge amounts of heterogeneous data in a low-cost hardware on a real-time basis has increased tremendously. The requirements of big-data storage in IoT are to handle massive amount of unstructured data and provide low latency for analytics. A challenge is that many sources of IoT data exist, for example, sensors' data, social media, etc., and they are modeled in various ways using different communication protocols and interfaces. Big-data technology provides some IoT-efficient data storage capabilities, but more robust solutions are required.

Quality of services: The ability to provide guarantee of a specific level of performance to the data flow is called Quality of service (QoS). The QoS provided by the IoT is that the IoT network should be reliable and should provide the guarantee of an efficient transfer of data from the

sources that generates the big data. The QoS in the IoT network is very important to big data and analytics [23].

Real time analytics: IoT is growing rapidly and taking key steps to improve streaming analytics and provide timely decision processes. Real-time information about the IoT-connected objects are communicated and need to be analyzed in real time. Big data uses an operational database for the streaming data, and for most of the streaming data from web-enabled objects, big-data analytics performs real-time queries to extract information quickly, make decisions, and interact with the devices and people in real time [24].

Benchmark: Due to the fast digitization of businesses, many organizations have started to shift their business online using IoT. Many organizations are now facing challenges in storing and analyzing the huge amount of data connected through the IoT devices. Finding solutions to those challenges requires some deep understanding of the problems. Benchmark plays an important role in this situation by allowing the organizations to compare the quality of the big data and analytics solutions [25].

III. THE ROLE OF IoT, EDGE COMPUTING, AND BIG DATA IN SMART GRIDS

To spur growth in businesses, effective business decisions are very important, and they are often made possible by getting information from collected data. IoT is a major source of data; by some estimates, there are currently more than ten billion devices connected to IoT networks¹, generating around trillions GB of data. These devices gather, analyze, share, and transmit data in real time. To handle such massive amounts of data, IoT needs edge computing and big data, making them the key to improve decision making [26].

In our this work, we have designed an architecture for getting the smart grid data using IoT and edge nodes as shown in Figure 2. The life-cycle of SG data starting from data generation to data analytics. The data is generated

¹<https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/>. Last access March 1, 2021.

from numerous smart meters, sensors, and digital devices with a specific time scale. The generated data may be from generation plants (wind farms, solar panels, conventional power plants, etc.), transmission and distribution networks (phasor measurement units, etc.), or customers (residential homes, electric vehicles, commercial buildings, factories, etc.). Data, such as weather, humidity, temperature, and pressure data, can also be collected from the environment. Some usable data, e.g., information about external events, can be collected from social media. Data generated from many sources increase the grid reliability. The generated data are transmitted to the IoT network using IoT devices such as sensors and actuators through the network technologies 3G/4G/5G, ZigBee, wi-fi, bluetooth, and wired communication.

The important data that needs to be processed quickly (requiring low latency) is processed by the edge nodes. The edge nodes are close to the data collection points, and therefore, require very low latency [3]. However, there are some cases in which the benefits may not be achieved, since the latency not only depends on the distance between the data collection point and edge processing server, but also on the edge server's processing power, tasks' computational complexity, and edge traffic [4]. Figure 3 shows the latency versus central processing unit (CPU) cycles that are required by a single device per bit in wireless communication by either the cloud or edge computing. To achieve the latency requirements efficiently, the edge network should be designed by keeping in mind factors such as task complexity, processing power of the servers, and the network topology used. Regarding the bandwidth, edge computing reduces the data traffic by distributing the data among different edge servers for computational workload, and thus, lower amount of data is required to be shifted to the cloud.

The cooperation between cloud and edge computing provides high bandwidth as the bits are transmitted to the cloud server when the sum of the tasks exceeds the combined computational capacity of the edge servers. Another benefit of using edge computing in SG is the reduction in failure—if there is an electricity outage problem in a particular area of the grid, the edge computing services of the other areas will operate normally, without any problem. On the other hand, if the grid relies solely on cloud computing, and there is a power supply failure due to any natural disaster in the cloud infrastructure, then the whole network will fail [27]. As shown in the Figure 4, cloud computing shows the best performance when the signal to noise ratio (SNR) is low, but the edge computing performs the best even at high-SNR regime as the number of edge servers increases, outperforming the cloud-assisted counterpart.

The data from IoT devices that is not handled by the edge nodes are directly sent to the cloud storage. The huge amount of data generated by these devices is stored in a low-cost storage at the cloud. In the second phase of data acquisition, the generated big data based on the volume, velocity, and variety is stored in a shared distributed fault

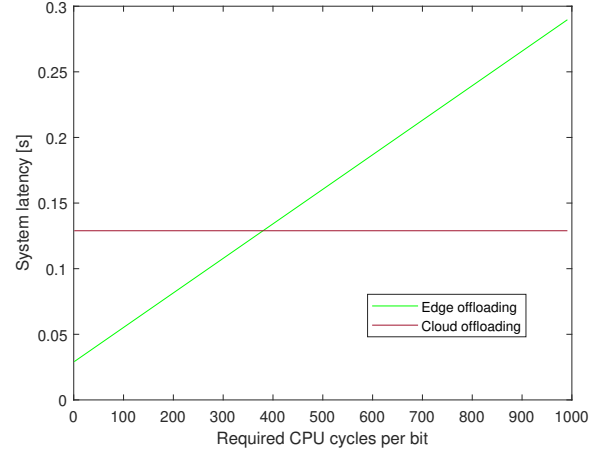


Fig. 3. System latency versus required CPU cycles per bit in wireless systems assisted by edge and cloud computing ($b = 1$ Mbit, $B^{\text{Edge}} = 10$ MHz, $B^{\text{Cloud}} = 10$ MHz, $\gamma^{\text{Edge}} = \gamma^{\text{Cloud}} = 10$ dB, $f^{\text{Edge}} = 6$ GHz, $\tau^{\text{Cloud}} = 100$ ms) adapted from [4]

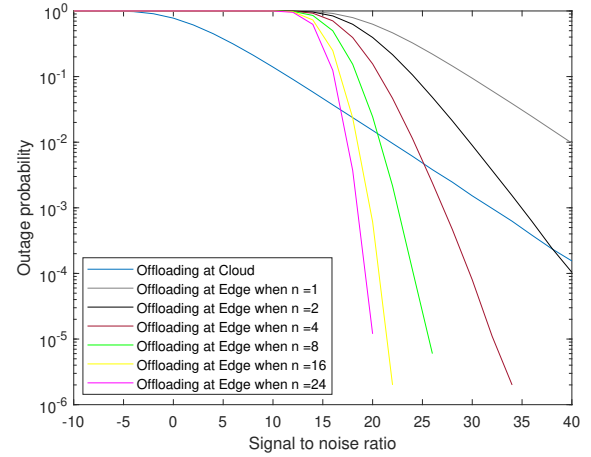


Fig. 4. Outage probability curves for wireless systems assisted by cloud and edge computing considering different numbers (n) of edge servers ($B^{\text{BH}} = 2$ MHz, $B^{\text{Edge}} = 50$ kHz, $R^{\text{target}} = 5$ Mbit/s) adapted from [4]

tolerant database. The collected data is then transferred into the master node(s) in the Hadoop cluster. As the data is collected from multiple heterogeneous devices, it may have different data formats and information, and therefore, data preprocessing will be required. In the data preprocessing, inaccurate and incomplete data are handled. Flume is used to perform the data acquisition process. The main function of Flume is to collect, aggregate, and transfer the large amount of data to Hadoop master node. The data received by Flume is stored in a single or multiple channel. The data is then sent to the external HDFS repository, where the data is written in a desired format using plug-in serializers. The serializers change and restructure the Flume data into the desired format. The data is pre-processed and a unified view of the data is achieved. The data is stored in the HDFS multiple clusters for processing. The HDFS clusters consists of DataNodes. The actual data and file system meta data are stored in those DataNodes. The data analysis is performed by the YARN on the data

stored in HDFS; these two run on the same set of nodes that allows tasks to be processed on the nodes in which the SG data is present. Hive and Impala are the tools to perform SQL queries on data residing on HDFS. HIVE is used for data querying, to select, analyze and make calculations on the data of interest. The last phase is the data analytics; the tools used in Hadoop for data analytics is Scalable Advanced Massive Online Analysis (SAMOA), a distributed streaming ML framework consisting of programming abstraction for distributed streaming algorithms for data mining and ML tasks. Data visualization (graphs, reports, etc.) is done using Tableau, a common tool for interactive data visualization and sharing of information and dashboards.

IV. CONCLUSIONS

The modern SG incorporates numerous heterogeneous devices. Due to this increase in the volume of structured, semi structured, and unstructured data, information retrieval from such a huge amount of data is a hard task. The collection, transmission, storage, processing, transformation, and analysis of large amount of data at a high rate are important for the efficient and effective function of modern SGs. The main aim of this research is to highlight the importance of IoT, edge computing, and big data for dealing with the high volume of SG data. In this paper, we have first presented the importance and requirements for big-data analytics in the SG. Subsequently, we have explained the applications of edge computing to the big data generated by the IoT devices in the SG. Edge computing is beneficial for SG in terms of latency, bandwidth, robustness to failure, and cost. In the future, we will apply big-data analytics to huge volumes of SG data and demonstrate the key requirements quantitatively.

ACKNOWLEDGEMENTS

This paper is supported by Academy of Finland via: (a) ee-IoT project n.319009, (b) FIREMAN consortium CHIST-ERA/n.326270, and (c) EnergyNet Research Fellowship n.321265/n.328869.

REFERENCES

- [1] A. Al-Ali *et al.*, "Role of internet of things in the smart grid technology," *Journal of Computer and Communications*, vol. 3, no. 05, p. 229, 2015.
- [2] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78–81, 2016.
- [3] Y. Miao, G. Wu, M. Li, A. Ghoneim, M. Al-Rakhami, and M. S. Hossain, "Intelligent task prediction and computation offloading based on mobile-edge cloud computing," *Future Generation Computer Systems*, vol. 102, pp. 925–931, 2020.
- [4] A. Narayanan, A. S. De Sena, D. Gutierrez-Rojas, D. C. Melgarejo, H. M. Hussain, M. Ullah, S. Bayhan, and P. H. Nardelli, "Key advances in pervasive edge computing for industrial internet of things in 5g and beyond," *IEEE Access*, vol. 8, pp. 206 734–206 754, 2020.
- [5] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE internet of things journal*, vol. 3, no. 5, pp. 637–646, 2016.
- [6] Stackpole Beth, "How industrial IoT is transforming the plant floor," 2016. [Online]. Available: <https://searchrpt.techtarget.com/feature/How-industrial-IoT-is-transforming-the-plant-floor>
- [7] Y. Ai, M. Peng, and K. Zhang, "Edge computing technologies for internet of things: a primer," *Digital Communications and Networks*, vol. 4, no. 2, pp. 77–86, 2018.
- [8] F. Dong, J. Yong, and X. Fei, "Emerging intelligent big data analytics for cloud and edge computing," 2020.
- [9] S. Choudhury, Q. Ye, M. Dong, and Q. Zhang, "IoT Big Data Analytics," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.
- [10] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqua, and I. Yaqoob, "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.
- [11] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," 2010.
- [12] A. A. Munshi and Y. A. Mohamed, "Cloud-based visual analytics for smart grids big data," *2016 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2016*, 2016.
- [13] S. S. Reka and T. Dragicevic, "Future effectual role of energy delivery: A comprehensive review of Internet of Things and smart grid," *Renewable and Sustainable Energy Reviews*, vol. 91, no. March, pp. 90–108, 2018. [Online]. Available: <https://doi.org/10.1016/j.rser.2018.03.089>
- [14] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid — the new and improved power grid: A survey," *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.
- [15] M. Ullah and K. Smolander, "Highlighting the Key Factors of an IoT Platform," *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 901–906, 2019.
- [16] I. Lee and K. Lee, "The internet of things (iot): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.
- [17] M. Ullah, P. H. Nardelli, A. Wolff, and K. Smolander, "Twenty-one key factors to choose an iot platform: Theoretical framework and its applications," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 111–10 119, 2020.
- [18] A. A. Munshi and Y. A. Mohamed, "Cloud-based visual analytics for smart grids big data," in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2016, pp. 1–5.
- [19] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, "Big data issues in smart grid systems," *2016 IEEE International Conference on Renewable Energy Research and Applications, ICRERA 2016*, vol. 5, pp. 1007–1012, 2017.
- [20] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, P. Zhao, M. Manic, S. Zhang, and X. Zhang, "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, vol. 2, no. 2, pp. 141–154, 2019.
- [21] G. Suciu, V. Suciu, A. Martian, R. Craciunescu, A. Vulpe, I. Marcu, S. Halunga, and O. Fratu, "Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure E-Health Applications," *Journal of Medical Systems*, vol. 39, no. 11, 2015.
- [22] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," *Journal of Internet Services and Applications*, vol. 6, no. 1, pp. 1–15, 2015.
- [23] J. Jin, J. Gubbi, T. Luo, and M. Palaniswami, "Network architecture and QoS issues in the internet of things for a smart city," *2012 International Symposium on Communications and Information Technologies, ISCIT 2012*, pp. 956–961, 2012.
- [24] Z. Ding, X. Gao, J. Xu, and H. Wu, "IOT-StatisticDB: A general statistical database cluster mechanism for big data analysis in the internet of things," *Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCom 2013*, pp. 535–543, 2013.
- [25] A. Shukla, S. Chaturvedi, and Y. Simmhan, "RIoTBench: An IoT benchmark for distributed stream processing systems," *Concurrency Computation*, vol. 29, no. 21, pp. 1–22, 2017.
- [26] A. A. Munshi and Y. A. Mohamed, "Big data framework for analytics in smart grids," *Electric Power Systems Research*, vol. 151, pp. 369–380, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.epr.2017.06.006>
- [27] J. Zhang, H.-W. Lee, and E. Modiano, "On the robustness of distributed computing networks," in *2019 15th International Conference on the Design of Reliable Communication Networks (DRCN)*. IEEE, 2019, pp. 122–129.