IoT Big Data Analytics with Fog Computing for Household Energy Management in Smart Grids

Shailendra Singh[1] and Abdulsalam Yassine[2(✉)]

[1] Department of Electrical and Computer Engineering, Lakehead University, 955 Oliver Road, Thunder Bay, ON P7B 5E1, Canada
ssingh59@lakeheadu.ca

[2] Department of Software Engineering, Lakehead University, 955 Oliver Road, Thunder Bay, ON P7B 5E1, Canada
ayassine@lakeheadu.ca

Abstract. Smart homes generate a vast amount of data measurements from smart meters and devices. These data have all the velocity and veracity characteristics to be called as Big Data. Meter data analytics holds tremendous potential for utilities to understand customers' energy consumption patterns, and allows them to manage, plan, and optimize the operation of the power grid efficiently. In this paper, we propose a unified architecture that enables innovative operations for near real-time processing of large fine-grained energy consumption data. Specifically, we propose an Internet of Things (IoT) big data analytics system that makes use of fog computing to address the challenges of complexities and resource demands for near real-time data processing, storage, and classification analysis. The design architecture and requirements of the proposed framework are illustrated in this paper while the analytics components are validated using datasets acquired from real homes.

Keywords: Internet of Things · Cloud computing · Fog computing
Big data analytics · Energy management · Smart grids

1 Introduction

The combination of IoT and big data analytics technology is expected to shape the decision-making processes in various industries [1]. As IoT systems expand to smart city applications that demand instantaneous actions, processing enormous data in near real-time to satisfy the stringent requirements of smart city functions becomes a challenging prospect. One solution is to use cloud-based systems since there is an abundance of computing and storage resources for various computationally intensive applications that need processing of high volume of data on the fly. However, attaining real-time responses from a cloud system is practically difficult due to the inherited latency of the underlying transport

A.-S. K. Pathan et al. (Eds.): SGIoT 2018, LNICST 256, pp. 13–22, 2019.
https://doi.org/10.1007/978-3-030-05928-6_2

communication network which has significant impact on time-sensitive applications [2–5]. Fog computing fundamentally resolves latency issues by processing and storing data at the edge of the cloud system [3,6]. Furthermore, fog computing nodes are resource-efficient because they are equipped with virtual machine technologies capable of continuously processing fresh IoT streams of data and transfer the processed data to the cloud for further processing. These nodes play a key role in the IoT ecosystem to support the processing of big data for near real-time responses. As a result, IoT big data analytics begin to leverage fog computing infrastructure to handle the data on the fly, with low latency.

In this paper, we propose a unified architecture that enables innovative operations for near real-time processing of enormous fine-grained data. As a compelling application of such architecture, we focus on developing a scalable IoT big data analytics with fog computing for processing and analyzing household energy consumption data for smart grids. Through smart meters and sensor devices, households generate continuous streams of massive amount of data in short time intervals. A large part of these data is attributed to home appliances and plug-in electric vehicles. Processing and analyzing these data is vital for smart grid energy management applications that aim at reducing cost and greenhouse gas emissions [7,8]. However, the implementation of home data analytics can be quite costly for a large number of consumers. The total computational effort required to perform data analytics for each consumer over time at fine-grained intervals is enormous. This is extremely challenging for utilities trying to adopt analytics to find the right consumers for an energy management program, let alone the heterogeneity of consumers' energy consumption behaviors [12,13]. To cope with such analytics complexities, several research studies, such as those in [10,11,15,18] and [19], have proposed IoT platforms with dedicated resources from fog and edge computing nodes to perform the analytical computations. The main idea is to be as close as possible to the source where data is generated. While such approaches are genuine, they tackle the latency issue only, but not necessarily applicable for handling a large volume of incoming data that requires orchestration of various application requirements. For example, residential Automatic Demand Response (ADR) applications require energy consumption data about appliances in residential homes to be analyzed in near real-time to engage them in demand response signals effectively [9]. Other smart grid applications that require predictive analytics need access to historical data which must be stored in a large database that only can be provided by a cloud system [14].

We present an Internet of Things (IoT) big data analytics platform with fog computing that is capable of managing, analyzing and transforming household energy consumption data into actionable insights. The proposed system, which acts as a hub for metered consumption and event data originating from household energy systems, is well suited to support huge data and computationally-intensive, always-on applications. It addresses the challenges of complexities and resource demands for near real-time data processing as well as the requirements of scalability with the growing volume of data and the temporal granularity of decision making. The advantages of such a platform lie in the ability of serving multiple households within a neighborhood at the same time which means that we can process multiple home appliances in parallel. Thus allowing us to analyze data faster and engage home appliances in smart grid applications (e.g. ADR) in a timely manner.

The rest of the paper is organized as follows: In Sect. 2, we present the components of the proposed platform followed by a study case in Sect. 3. Finally, in Sect. 4 we conclude the paper and provide direction for future work.

2 Platform Overview

2.1 Requirements and Functionalities

As mentioned earlier, utility companies are facing simultaneous streams of data from multiple household devices and metering systems. The design of an innovative platform that is suited to support a huge amount of energy for smart grid energy management applications posses peculiar requirements, functionalities, and design structures.

- IoT energy metering streams should be handled in a parallel manner to boost the performance of data analytics and to optimize the smart grid dynamic energy management operations. Depending on the analytics activity, the specific requirements include elastic resource acquisition, efficient data network management, data reliability, and functional data abstractions.
- Data processing should make full use of all computational resources to address performance challenges of near real-time computation algorithms such as finding hidden patterns and produce new, faster and richer knowledge.
- Home appliance data changes over time due to the changing consumption behavior of consumers. As a result, an automated data ingestion pipeline must support dynamic data acquisition at variable rates and volumes and be adaptive to current data sources and operational needs. This high-speed pipeline should process all incoming time-series data, applies simple data transformations, and outputs the processed data.

The main challenges for satisfying the above requirements especially for near real-time energy management applications are in the development of a platform capable of processing and analyzing large volume of energy consumption data streaming from various sources. Next section describes the proposed platform.

2.2 Platform Components

In Fig. 1, we present an IoT big data analytics with fog computing platform that supports complex operation of continuous integration, processing and analytics of multiple household energy consumption data.

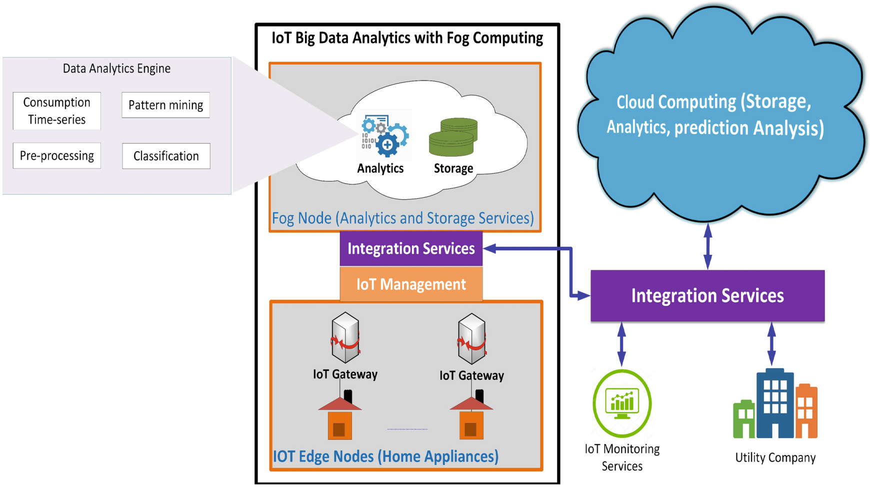

Fig. 1. IoT big data analytics with fog computing

The architecture is mainly composed of four sub-systems that enable the management of household energy consumption data in near real-time. These modules are as follows:

- **IoT data acquisition**: In the proposed model, data is acquired from household appliances through data acquisition modules or the smart meter disaggregated data services. In either case, IoT protocols such as machine-to-machine (M2M)/Internet of Things (MQTT) are responsible for transporting the data between home appliances and the IoT gateway. An IoT gateway acts as a broker that connects IoT home devices to the IoT management entity that bridges the communication to the fog nodes.
- **IoT Management and Integration Services**: The IoT management services play a mediating role for transferring data from IoT devices to Fog computing nodes and then to the cloud system. They are protocol independent and are mainly responsible for maintaining continuity and flexibility for the whole IoT ecosystem. The frequency of data transfer is generally application specific. For example, smart meter data are collected in intervals of 15 min with various resolutions. The integration services provide application programming interfaces (APIs) with external systems. There are many benefits for decoupling the analytical components of the fog nodes from the external systems. Such decoupling assures security since users would not be able to have any direct access to the analytical engine. Also, it adds abstracting data and interoperability by enabling the use of data for various user-specific applications including mobile and desktop applications.
- **Fog Computing Nodes**: The fog node is a resource-efficient computational entity that supports rapid analytics of energy consumption data for near real-time smart grid applications. Among the main function of the fog node

is pre-processing collected IoT data and sending the aggregated results to the cloud or directly to the serviced applications. By doing so, the fog nodes increase the ability of the platform to manage an integrated array of analytics for smart grid applications in highly automated ways which result in significant savings for the grid operator. Also, utilities can design and develop their applications using fog nodes that offer abundance elasticity to enhance performance, redundancy and storage devices that make the scaling problem of energy consumption analytics much easier to handle. We should note that the method of allocating fog nodes to households is beyond the scope of this paper. However, optimization mechanisms such as those in [16,17] may be employed to determine the optimal distribution and configuration of fog nodes while taking into consideration the computational resources and capability of processing the required data from multiple homes. In the case of configuring more than one household to a single fog node, privacy and security of information is often considered an issue which can be tackled by mechanisms such as those in [20,21,23,24] and [25] should be considered.
- **Cloud System**: Household energy consumption management and data analytics is a complex operation that requires continuous integration of multiple sources to a common processing system with easy access to data. In the proposed platform, energy consumption from many fog nodes is aggregated at the cloud system which provides additional computations for large data processing.

2.3 Data Analytics Engine

The data analytics engine in our platform performs all the short-term analytics at the edge of the cloud system. Energy consumption time-series data acquired from IoT streams are processed as they arrive at the analytical engine. The processing of this data can be divided into three main stages: pre-processing, pattern mining, and classification.

In pre-processing stage all IoT streams are filtered, parsed and translated into a unified data structure for further analysis. At this stage, raw data which contains millions of high time-resolution energy records are transformed into a pre-defined resolution for each appliance, while recording usage duration, average load, and energy consumption. The decision for determining the resolution (5 min, 15 min, 30 min, etc.) is provided and configured by the user. In the second stage, frequent pattern mining techniques are conducted on the data to discover the occurrence of appliance correlation in a dataset. The main idea here is to uncover appliance relationships that affect energy consumption behavior. Frequent pattern mining searches for these recurring patterns in a given dataset to determine associations and correlations among patterns of interest [22]. In our platform, the data analytics engine uses both the FP-growth mechanism and the Apriori algorithm to discover appliance associations in the form of frequent patterns and association rules, respectively.

In clustering stage, we employ an unsupervised form of classification which is capable of distinguishing classes of appliances which are learned from the data [22]. There are various clustering approaches such as hierarchical clustering, centroid-based or partitional distribution clustering, distribution-based clustering, and density-based clustering. In our model, we extend the k-means, which is a partitional distribution clustering algorithm, to discover appliance-time associations. Our goal is to provide a critical analysis of consumer energy consumption behavior concerning preferences on time of energy usage. Appliance-time associations can be defined with respect to hour of day, time of day, weekday, week, month and/or season. Determining the appliance-time associations, for an appliance, can be considered as a grouping of sufficiently close appliance usage time-stamps, when that appliance has been recorded as active or operational, to form classes or clusters. The clusters or classes constructed will describe appliance-time associations while respective size of clusters, defined as the count of members in the cluster, will establish the relative strength for the clusters. The strength or size of the cluster will indicate how frequently and when a given appliance has been used by consumer, which indicate personal preferences. Therefore, discovery of appliance-time associations can be translated into clustering of appliances' operating time-stamps into brackets of time-spans, where each cluster belongs to an appliance with respective time-stamps (data points) as members of the cluster. Finally, the results of the above mentioned stages are send to the cloud system which is freed for computationally intensive tasks, especially where the analysis of historical data and large datasets is concerned.

3 Case Study and Analysis

We conducted extensive experiments using household energy consumption data from smart meters using the real dataset [26]. We present sample results indicative of our observations. Figures 2 and 3 show the energy consumption patterns for Toaster and Home Theater over hour-of-day, time-of-day, weekday, month and season. The outcome of the frequent pattern mining is the association among appliances that are the result of the simultaneous use of the appliance by occupants. Figure 4 show the result of clustering operation that is part of possible behavioral predictive analytics, which can be conducted at fog nodes. The figure exhibit, appliance to hour-of-day associations for Toaster that were determined by clustering of the time-stamps where the toaster was registered operating. Further, when we compare Figs. 2 and 4, we notice that the outcome of clustering analysis successfully captures the appliance usage patterns that are direct reflection of occupants energy consumption behavior. This is in addition to the appliance-to-appliance associations learned through frequent pattern mining that comprehensively represent energy usage behavioral traits of occupants. Similar results were obtained for appliance to time-of-day, weekday, month and season associations. Further, it can be done while taking into account the local energy generation through renewal sources at individual house or a neighborhood level. By facilitating distributed near real-time smart meter big data analytics at

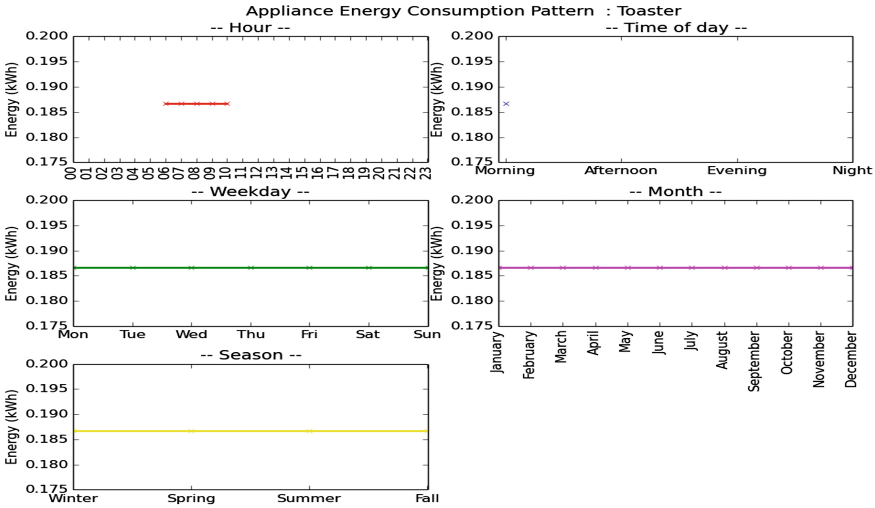

Fig. 2. Energy consumption pattern - Toaster

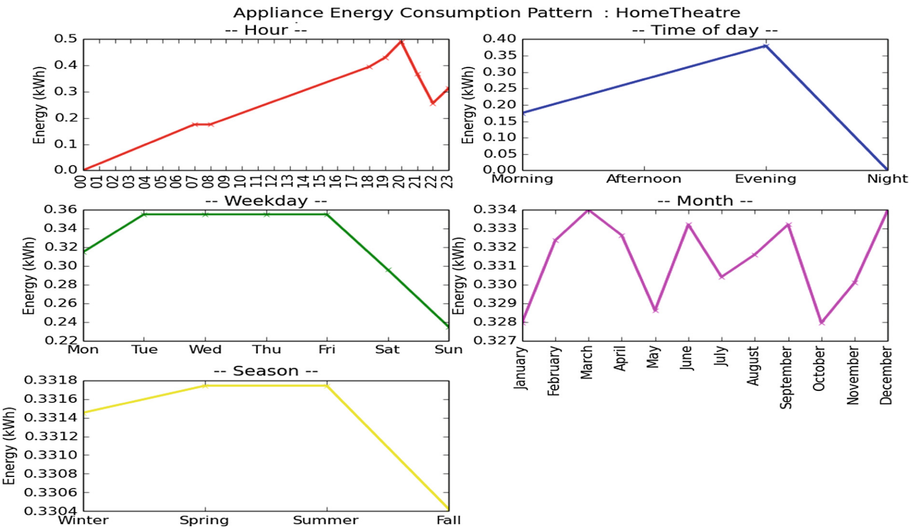

Fig. 3. Energy consumption pattern - Home theater

the edge of the cloud using fog computing the proposed platform can aid effective and in-time decision making for individual house owners, distribution whereas large-scale data analytics in the cloud can facilitate various energy management programs at producers level.

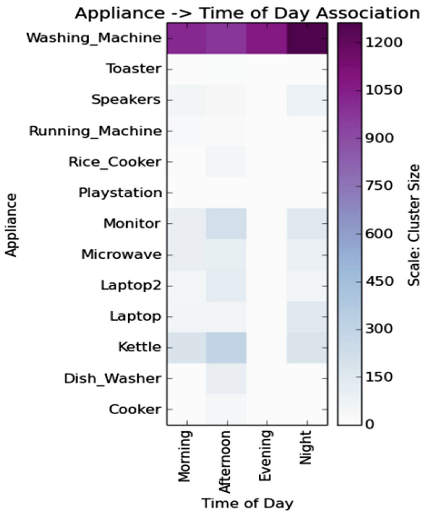

Fig. 4. Appliance time of the day associations. It shows the co-utilization of appliances during the day

4 Conclusion and Future Work

In this paper, we have discussed a platform to effectively exploit IoT and Big Data Analytics towards developing energy management strategies for efficient and effective household energy management. We have described the requirements and components of this platform and discussed a use case scenario of real-life data from actual homes. Further, the proposed platform can be tailored by domain experts for the evaluation of various smart grid applications of interest. Our plan for future work is to refine the platform component and test with different datasets from various homes. This is crucial to validate the applicability of the platform and its robustness in dealing with all kind of energy consumption measurements. We also plan to conduct predictive analytics for energy consumption based on real-time streaming of IoT data from smart homes. Thus, enabling the platform to perform short and long term forecasting for smart grid applications.

References

1. Chhabra, A.S., Choudhury, T., Srivastava, A.V., Aggarwal, A.: Prediction for big data and IoT in 2017. In: International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dubai, pp. 181-187 (2017)
2. Ge, Y., Liang, X., Zhou, Y.C., Pan, Z., Zhao, G.T., Zheng, Y.L.: Adaptive analytic service for real-time internet of things applications. In: IEEE International Conference on Web Services (ICWS), San Francisco, CA, pp. 484–491 (2016)

3. El-Sayed, H., et al.: Edge of Things: The big picture on the integration of edge, IoT and the cloud in a distributed computing environment. IEEE Access **6**, 1706–1717 (2018)
4. Pouladzadeh, P., Kuhad, P., Peddi, S.V.B., Yassine, A., Shirmohammadi, S.: Mobile cloud based food calorie measurement. In: IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Chengdu, pp. 1–6 (2014)
5. Peddi, S.V.B., Kuhad, P., Yassine, A., Pouladzadeh, P., Shirmohammadi, S., Shirehjini, A.A.N.: An intelligent cloud-based data processing broker for mobile e-health multimedia applications. Future Generat. Comput. Syst. J. **66**, 71–86 (2017)
6. Mebrek, A., Merghem-Boulahia, L., Esseghir, M.: Efficient green solution for a balanced energy consumption and delay in the IoT-Fog-Cloud computing. In: IEEE 16th International Symposium on Network Computing and Applications (NCA), Cambridge, pp. 1-4 (2017)
7. Al-Ali, A.R., Zualkernan, I.A., Rashid, M., Gupta, R., Alikarar, M.: A smart home energy management system using IoT and big data analytics approach. IEEE Trans. Consum. Electron. **63**(4), 426–434 (2017)
8. Berouine, A., Lachhab, F., Malek, Y.N., Bakhouya, M., Ouladsine, R.: A smart metering platform using big data and IoT technologies. In: 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), Rabat, pp. 1-6 (2017)
9. Yassine, A.: Implementation challenges of automatic demand response for households in smart grids. In: 3rd International Conference on Renewable Energies for Developing Countries (REDEC), Zouk Mosbeh, pp. 1–6 (2016)
10. Sultan, M., Ahmed, K.N.: SLASH: self-learning and adaptive smart home framework by integrating IoT with big data analytics. In: Computing Conference, London, pp. 530–538 (2017)
11. Yang, S.: IoT stream processing and analytics in the fog. IEEE Commun. Mag. **55**(8), 21–27 (2017)
12. Singh, S., Yassine, A.: Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies **11**, 452 (2018)
13. Singh, S., Yassine, A.: Mining energy consumption behavior patterns for households in smart grid. IEEE Transactions on Emerging Topics in Computing (2017). https://doi.org/10.1109/TETC.2017.2692098
14. Cai, H., Xu, B., Jiang, L., Vasilakos, A.V.: IoT-based big data storage systems in cloud computing: perspectives and challenges. IEEE Internet Things J. **4**(1), 75–87 (2017)
15. He, J., Wei, J., Chen, K., Tang, Z., Zhou, Y., Zhang, Y.: Multi-tier fog computing with large-scale IoT data analytics for smart cities. IEEE Internet Things J. **5**(2), 677–686 (2018). https://doi.org/10.1109/JIOT.2017.2724845
16. Taneja, M., Davy, A.: Resource aware placement of IoT application modules in Fog-Cloud Computing Paradigm. In: IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, pp. 1222-1228 (2017)
17. Minh, Q.T., Nguyen, D.T., Van Le, A., Nguyen, H.D., Truong, A.: Toward service placement on Fog computing landscape. In: 4th NAFOSTED Conference on Information and Computer Science, Hanoi, pp. 291–296 (2017)
18. Gonzalez, N.M., et al.: Fog computing: data analytics and cloud distributed processing on the network edges. In: 35th International Conference of the Chilean Computer Science Society (SCCC), Valparaiíso, pp. 1-9 (2016)

19. Cao, H., Wachowicz, M., Cha, S.: Developing an edge computing platform for real-time descriptive analytics. In: IEEE International Conference on Big Data (Big Data), Boston, MA, pp. 4546–4554 (2017)
20. Yassine, A., Nazari Shirehjini, A.A., Shirmohammadi, S.: Smart meters big data: game theoretic model for fair data sharing in deregulated smart grids. IEEE Access **vol. 3, no**, 2743–2754 (2015)
21. Yassine, A., Shirmohammadi, S.: Measuring user's privacy payoff using intelligent agents. In: IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA 2009, pp 169–174 (2009)
22. Han, J., Pei, J., kamber, M.: Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods. Data mining: Concepts and techniques, Chap. 6, 3rd edn, pp. 243–278. Morgan Kaufmann, Waltham (2011). http://www.sciencedirect.com/science/book/9780123814791. ISBN: 9780123814791
23. Paverd, A., Martin, A., Brown, I.: Security and privacy in smart grid demand response systems. In: Cuellar, J. (ed.) SmartGridSec 2014. LNCS, vol. 8448, pp. 1–15. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10329-7_1
24. Yassine, A., Shirmohammadi, S.: Privacy and the market for private data: a negotiation model to capitalize on private data. In: IEEE/ACS International Conference on Computer Systems and Applications, Doha, pp. 669-678 (2008)
25. Yassine, A., Shirehjini, A.A., Shirmohammadi, S., Tran, T.: Knowledge-empowered agent information system for privacy payoff in ecommerce. Knowl. Inf. Syst. **32**(2), 445–473 (2012)
26. Makonin, S., Ellert, B., Bajic, I.V., Popowich, F.: AMPds2 - Almanac of minutely power dataset : electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. Sci. Data **3**, 1–12 (2015). https://doi.org/10.1038/sdata.2016.37