Big Data Analytics with Fog Computing in integrated Cloud Fog and IoT Architecture for Smart Devices

Sultan Ahmad[1*], Mohammad Mazhar Afzal[2] and Abdullah ALharbi[3]

[1*] Department of Computer Science, College of Computer Engineering and Sciences,
Prince Sattam Bin Abdulaziz University, Al-Kharj,11942, Saudi Arabia
[1,2]Department of Computer Science and Engineering, Glocal University, Saharanpur(U.P), India
[3]Department of Information Technology, College of Computers and Information Technology,
Taif University, Taif, 21944, Saudi Arabia
Corresponding Author: Sultan Ahmad(*s.alisher@psau.edu.sa*)

Summary
Fog Computing is also well known as Real-Time Edge Computing. It emerged from Cloud Computing and implemented closed to the end user's devices to deal with network latency. Big data services rely on the Internet of Things (IoT) systems to integrate heterogeneous devices and to deliver data to the central server for analytics. These big data in IoT is the enormous amount of data collected from IoT devices. Large number of IoT connected devices introduces new issues likes network latency, unstable bandwidth and unstable connectivity and many more issues which impact the performance and reliability of the system where big data services should be available on time. One of the possible solution which will effect the whole scenario is reasonable Fog Computing architecture, which addresses the issues of performance by shifting the computational load on to the edge devices. Finally, overall performance will be enhanced. The advancement of techniques and big data of this 5G era with low cost IoT devices implicitly introduced transition towards new paradigm of computing. The Big data collection, analysis and communication are highly impacted with inclusion of advanced IoT technologies. In this paper, authors have explored how Fog Computing enhances Big Data Analytics capabilities in Cloud – IoT architecture for smart devices of IoT environment.

Key words:
Fog Computing, Edge Computing, Big Data, IoT, Data Mining.

1. Introduction

Advances in technology are the key drivers and powerhouse of economic developments. Technological developments are currently playing a remarkable contribution in the life of modern society, and scientific knowledge has become more impactful across numerous domains of society than any time ever. Statista predicted IoT devices worldwide to be around 30 billion by 2020 with further predictions to cross 75 billion by 2025 [1]. That can be easily shown in Figure 1. Statistically, also, Big Data accretion over IoT devices and networks is apparently manifest and to clear up this trouble, various computing methods and a number of computing techniques have been known long before. There are processes like quantum computing, cloud computing, edge and fog computing. Though Quantum computing has a silvery future, it has to travel a long distance to get up and go. Developers and data scientists long before are well familiar with cloud computing and analytic method. Fogging, a brand new method and better known as Fog Computing, was first called at Cisco. Data analytic tools like Hadoop helps reduce the cost of storage which further augment the business capacity.

The current models and environments of computing have changed so dramatically following the unexpected fast improvements in technologies with big data and technologies based on cyberspace supporting as cloud computing, Fog Computing, Edge Computing and Internet-of-Things (IoT), and other large-scale computing environments [2]. IoT is being employed in various fields and is making "Everything Smart," like smart mobile, smart manufacturing, smart town, and smart transportation. Although researchers continue to tackle IoT devices issues of security and privacy, Big data generated by these growing technologies, need to be addressed more. Further, with the developing adoption of IoT devices, we will identify an advancement in the number of security and privacy problem. At present techniques related to securities are challenging with many query in these new computing environments. Thus, efforts are necessary to explore how to handle the growing big data and its effects and the safety measure and privacy problems of the above mentioned contemporary environments within the cyberspace.

Manuscript received June 6, 2020
Manuscript revised June 20, 2020

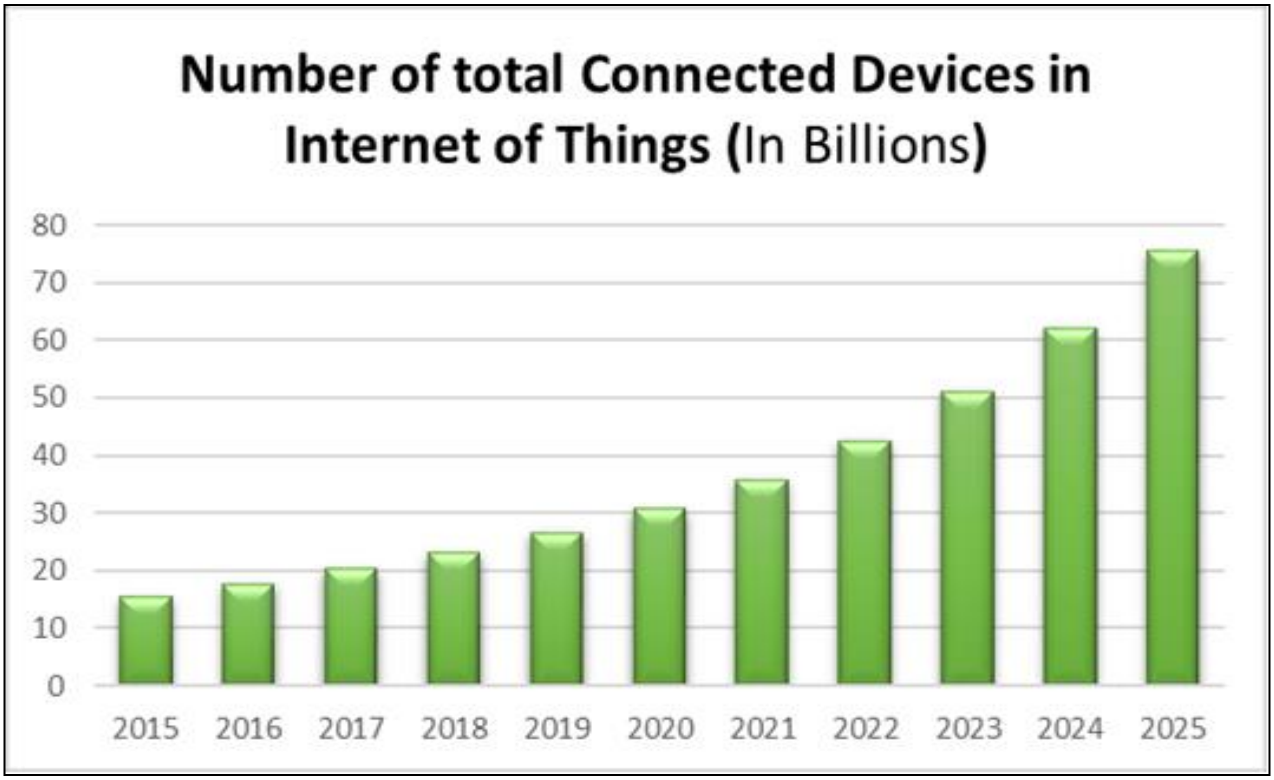

Fig. 1 No. of connected Devices worldwide in IoT (Source - Statista ,2019).

In this paper, the authors are also exploring the effects of Fog Computing and other computing technologies on the growing Big Data, Data Mining and its analytics techniques. Many aspects of data processing, storage security, and retrieval techniques are being affected and need to improve the methodologies to handle them properly in regulation to correct the effectualness and security of data storage and renascence. “Fog has some extraordinary characteristics which make it more applicable for the applications demanding low latency, mobility support, real-time interactions, online analytics, and interplay with the cloud” [3].

2. Related Works

The brand new Computing paradigms might achievable help to meet QoS requirements, application synchronization and revelation. Related problems likes storage restriction, and Management structure related issues are discussed in too many works.

Segall, R et al.(2018), has given visualization of vast data in fog computing and cloud computing in IoT environment [4]. They have done the case study for actual data from fatality analysis reporting system (FARS) handled by the National High-way Traffic Safety Administration (NHTSA) of the United States Department of Transportation (USDoT). The challenges and scope of opportunity in using Big Data along Fog Computing has been also explain in their work. In further development Segall, R et al.(2019), have explored problems along with explanation of High Performance Computing (HPC) systems, environment for storage systems for data-intensive applications, storage arrangement and storage structure for Big Data with Fog-to-cloud infrastructure [5]. Mehdipour, F et al. (2019), have explained in their book chapter about how to deploy Fog Engine for vast data analytics[6]. The FE gives as a premise data analytics along with the potentiality for Internet - of - Things (IoT) devices to interact with each other and with the cloud. They have elaborated system prototype and their evaluation with the help of two case studies. Fog Computing implementation in IoT healthcare application has been performed and evaluated by Mutlaq et al. (2018) in their work [7]. They explained latency benefit of fog computing over cloud computing with simulation that is an important aspect in real time application like healthcare. T. Kudo (2018), has proposed the distributed data model [8]. In this model of Fog Computing, it has classified three categories of data: sensor's original data, extracted data, and its analysis result data. He used distributed database concept and created this model on Fog and Cloud node. In his work out, it is being exposed that the data volume to be transferred generally to the cloud server can be decreased by using this model.

Chang et al. [9] recommended the Indie Fog framework that can be expand in many ways. It can utilize customer' network devices for giving Fog computing for IoT devices. It works for processing of data gathered from sensors and other devices. The framework Indie Fog would be expanding in static sensor devices and give the infra-structure for different services such as prompt data acquisition and processing.

Gupta H et. al. (2016), have developed a modelling and simulation/duplication tool named ' iFogSim' that can be used to do modelling and simulation/duplication of fog computing environments in order to evaluate prevising and management in edge and cloud resources in various conditions[10]. It calculates performance metrics and replicate edge devices, cloud datacenters, sensors, network links, data streams, and stream-processing applications.

In another work, Jinlai Xu et. al. (2017) have proposed a model named 'Zenith' for computing resources allocation in Edge Computing and internet of things infrastructure [11]. They evaluated their model with many experiments in order to check the effectiveness, scalability and performance efficiency. Farshad F et. al. (2018) have purposed their work and analyzed the concern and issues in Internet of Things Technologies for healthcare applications specially tele-health care application. These functions consist of wearable and body sensors, progressive pervasive healthcare systems, and the Vast Data analytics required to inform these devices.

3. Big Data and Data Mining

Big Data are in huge capacity being collected from homogeneous devices or multiple types of devices. These devices could be sensors in IoT, standard PCs and regular embedded devices. These big data include valuable information and useful knowledge. These data as in various forms and types used for further analysis and knowledge extraction. These data can be further categories as precise, unprecise and uncertain.

Data Mining tools and techniques are used in collecting and processing of Big data. Data mining is a process to discover the implicit useful data which were earlier unknown. Data mining on the Big Data, generated by Internet of Things devices helps to discover patterns for IoT applications of IoT verticals likes smart cities, smart home, smart industries and many mores analytics. It helps to find and identify different communities and groups on various social networking sites and applications.

The main famous data mining process are Frequent Pattern recognition and Supervised Learning. Pattern recognition enables us to find frequent appearance of data set and their correlation in IoT environment. Pattern Mining is help to find sets of oftentimes occurring items in IoT domains and association among them. Supervised Learning is a method to develop models from existing data for the purpose of classification and prediction. Big Data analytics revolutionized information and communication technologies. The new implementation of data analytics in organization widely enhanced and pushed the technologies advancement with time. Most of the commercial organizations focus on customer relationships and consumer applications for business [13]. As per this scenario, Big Data Analytics can be categorized in three categories as below Figure 2.

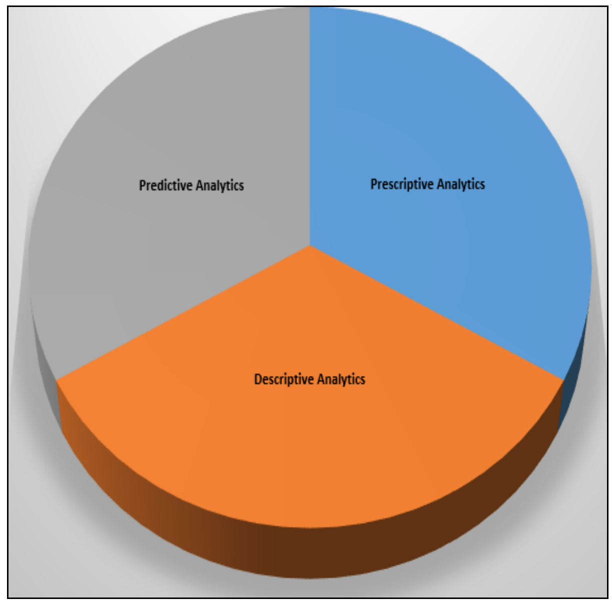

Fig. 2 Different categories of Big Data Analytics

Big Data analysis is interdisciplinary domain which is formed by integrating machine learning(ML), Data Mining(DM), NLP (natural language processing) and statistics. These data are excavated, handled and combined for better analysis for the businesses. These methods and technologies are being used nowadays in great enterprises and multinational companies in different ways for different purposes.

4. Fog Computing and its benefits towards Data Analytics

Fog Computing, implemented as a distributed system which Big Data collection, processing, and distribution as its components. In this environment edge nodes are capable of collecting, processing large quantity of data locally [14]. This makes possible to minimize the load of remote server and minimizing the bandwidth overhead of communication channel. In the middle layer of this architecture, cloudlets work as mini data center that accomplish major data processing task to support the IoT devices and applications which needs real time, low latency data processing [15]. Major data storage and processing load are shifted to Fog, which is acting as intelligent gateway to the cloud server.

Fog Computing can be performing data mining task such as pattern recognition that results in high efficiency and low latency. In this approach data is stored on IoT devices to discover regular patterns and further integrating this pattern to generate a global candidate pattern.

The main advantages of Fog Computing are as follow:

1) Optimum utilization of edge network resources instead of wide cloud network.
2) Processing and storage are made near to end users.
3) Storage and Processing are at IoT devices or middle layer between edge and central data centers.
4) It minimizes the latency problem.
5) It can efficiently work on low bandwidth.
6) Data security and reliability are improved in Fog Computing.
7) Improved users experience with low response times.
8) Power efficiency with Bluetooth, Zigbee or Z-wave.

The data processing steps and process in Cloud-Fog-IoT based infrastructure can be seen in the below given Figure 3. IoT devices which are sensor based collects the data. These data in aggregated and collaborative way, are delivered to the edge and fog server. Here data are processed locally properly. The data are being indexed and mined in order to get efficient multiple search and pattern recognition. To ensure future security data are being encrypted with secure algorithm by proxy server before sending it to the Cloud. The details subtasks in different phases can be shown in the given Figure 3.

Once the data user search the required data edge server send a query to the cloud server that search it in encrypted index data structure. Cloud server is responsible to store all the data and do the search operation for particular data, requested by data user. Finally, Cloud server send the encrypted search result to user that can decrypt the encrypted data based on their secret keys.

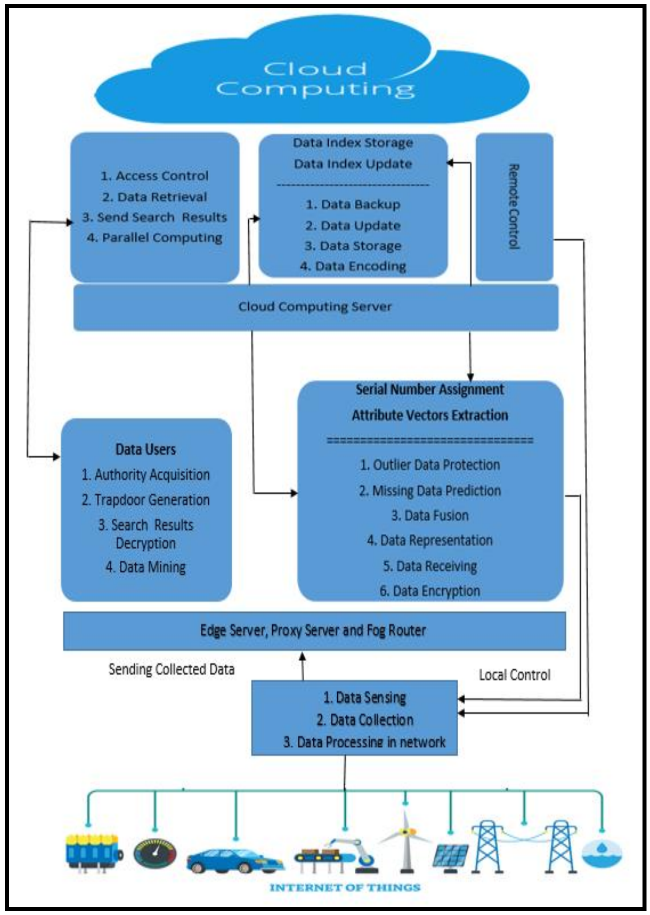

Fig. 3 Data Collection, Storage, retrieval and Mining process in Cloud-Fog-IoT Architecture

The Fog Computing is needed to help multiple conventions, for example, ZigBee, WiFi, 2G/3G/4G/5G, WiMax, 6LOWPAN and so on, whereas Cloud is for helping TCP/IP. The pushing service incorporates receiving the knowledge and converting the information. The end devices pass on information to the Fog later they produce the information [16].

Integrating cloud computing and IoT devices reduces the cost and improve the efficiency of the data analysis in IoT environment. Further, scalability of devices is also accomplished in the system. But this approach also inherits the limitation and existing security and privacy issues due to its huge distributed infrastructure [17]. That is benefits of fog computing over cloud computing paradigm.

5. Fog verses Cloud in aspect of Big Data

It is widely accepted that Fog Computing can replace cloud computing and it is merely imagination and can replace one another or take one another place. Fog computing implies an ignorant approach and is overlooked by the developers that they have a fine advantage when it is used for the local application based on IoT networks [18]. Whereas cloud computing also poses an advantage for the same but the volume of Big Data has a large potential in cloud networks and the following is well demarcated and circulated in modules/supervised in fog computing.

Table 1: Comparison of Cloud and Fog Computing for Big Data

Characteristics	Fog Computing	Cloud Computing
Hierarchy of Processing	Data are being analyzed locally	Data are being analyzed at data centers
Processing Technique	In-stream processing	Batch processing
Computing Power	GFLOPS	TFLOPS
Latency	Milliseconds	Seconds
Storage of Data	Gigabytes	Unlimited
Lifetime of Data	Hours or Days	Unlimited
Fault-Tolerance	Very High	High
Processing Resources and Granularity	Heterogeneous (e.g., CPU, FPGA, GPU) and Fine-grained	Homogeneous (Data center) and Coarse-grained
Versatility	Only exists on demand	Intangible Servers
Provisioning Process	Limited by the number of Fog-engines in the vicinity	Infinite
Nodes Mobility	Maybe Mobile (e.g. in the car)	Static
Pricing Policy	Pay once	Pay-as-you-use
Power Model	Battery-powered/Electricity	Electricity

In cloud computing, the structure of the data is placed far away from the consumer devices whereas in the fog computing the data is structured and stored in the shape of millions of tiny nodes stored adjacent to the customer devices. A comparison table is demonstrated in given below Table 1. Moreover, if one compares cloud computing and fog computing structures, cloud computing has an extensive network, large distant from end devices whereas fog computing works as a medium between data centers and hardware and it is close to end-users.

Cloud computing contact with devices directly but only without the fog layer, which sometimes consumes time. However, fog computing contacts with devices directly and it provides is real-time feedback but it works in the proximity of the devices with tiny little nodes while cloud computing works with the remote servers which is why it is time taking.

Fog computing is tiny in size as compare to cloud computing, and also fog computing has low storage capacity and power. Fog computing works on millions of tiny nodes which are linked inside and distinct as compared to cloud computing with fewer nodes. Small term data analysis can be done by the use of fog computing because it responds immediately but cloud computing is meant for long term data analysis and it provides passive responses. Because of the number of nodes in conducting data, fog computing has minor latency whereas cloud computing works on the base of few nodes and has to analyze Big Data results in incurring latency. Fog computing is feasible in a protocol like Bluetooth, Zigbee and so on, even if it faces the power cut while in contrast cloud computing cannot even imagine handling data with the power supply. Consequently, fog computing is a far safer system to store the data through the use of tiny nodes or end-devices and the use of cloud computing is not secure for data as it uses the cloud sources that need a constant power supply and is time taking.

6. Conclusion

With more IoT devices, the Big Data is becoming more intense in amount and to manage it. But, as there is distinct computing potential to every method, quite distinct from each other. Every lock has its key and similarly, every Big Data center has a distinct computing way for it. The field of Data analytics and computing has vast capabilities as there are tons of Data mined each hour through different IoT devices all over the world. There are new choices for research and analysis using data analytics. A lot of firms are spending in developing computing potentiality of their own to cope up with large amounts of Big Data and its analytics. Thus, if any institution is data-driven or it has an IoT based application, they require to have a computing solution to their data storage as data analytics can assist them to understand, customer purchase patterns, the capability pool of consumer, Market surveys, competitor strategies, and strategic marketing fronts. If their aim is for long-term analysis of data, cloud computing is the right way to go, with cloud point of supply at its core. But, if users are searching for intensify user experiences and immediate real-time responsive, computation needs, they should go for Fog Computing.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia for the assistance.

References

[1] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/(Access Available on 11 Dec,2019)

[2] Ahmad S., Afzal M.M. (2020) Deployment of Fog and Edge Computing in IoT for Cyber-Physical Infrastructures in the 5G Era. In: Karrupusamy P., Chen J., Shi Y. (eds) Sustainable Communication Networks and Application. ICSCN 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 39. Springer, Cham

[3] A. V. Dastjerdi and R. Buyya, "Fog Computing: Helping the Internet of Things Realize Its Potential," in Computer, vol. 49, no. 8, pp. 112-116, Aug. 2016. doi: 10.1109/MC.2016.245

[4] Segall, R. S., & Niu, G. (2018). Big Data and Its Visualization With Fog Computing. International Journal of Fog Computing (IJFC), 1(2), 51-82. doi:10.4018/IJFC.2018070102

[5] Segall, Richard & Cook, Jeffrey & Niu, Gao. (2019). Overview of Big Data-Intensive Storage and its Technologies for Cloud and Fog Computing. International Journal of Fog Computing. 2. 74-113. 10.4018/IJFC.2019010104.

[6] Mehdipour, F., Javadi, B., Mahanti, A., & Ramirez-Prado, G. (2019). Fog Computing Realization for Big Data Analytics. Fog and Edge Computing: Principles and Paradigms, 259-290.

[7] A.A. Mutlag, M.K.A. Ghani, N. Arunkumar, M.A. Mohamed, O. Mohd.(2018), Enabling technologies for fog computing in healthcare IoT systems, Future Generation Computer Systems (2018), https://doi.org/10.1016/j.future.2018.07.049

[8] T. Kudo, "Fog Computing with Distributed Database (2018), IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Krakow, 2018, pp. 623-630. doi: 10.1109/AINA.2018.00096

[9] C. Chang, S. Narayana Srirama and R. Buyya, Indie Fog: An Efficient FogComputing Infrastructure for the Internet of Things, in Computer, Vol. 50, no. 9, pp. 92-98, 2017

[10] Gupta H, Dastjerdi AV, Ghosh SK, Buyya R (2016) ifogsim: A toolkit for modeling and simulation of resource management techniques in internet of things, edge and fog computing environments. arXiv:abs/1606.02007

[11] Xu, J.; Palanisamy, B.; Ludwig, H.; Wang, Q. Zenith: Utility-Aware Resource Allocation for Edge Computing, In Proceedings of the IEEE International Conference on Edge Computing (EDGE), Honolulu, HI, USA, 25–30 June 2017; pp. 47–54.

[12] Farshad Firouzi, Amir M. Rahmani, K. Mankodiya, M. Badaroglu, G.V. Merrett, P. Wong, Bahar Farahani, Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics, Future Generation Computer Systems, Volume 78, Part 2, 2018, Pages 583-586, ISSN 0167-739X, https://doi.org/10.1016/j.future.2017.09.016

[13] Ajit Singh, Sultan Ahmad, Mohammad Imdadul Haque(2019), "Big Data Science and EXASOL as Big Data Analytics tool", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-9S,pp-933-937, July 2019

[14] W. Shi and S. Dustdar, "The Promise of Edge Computing," in Computer, vol. 49, no. 5, pp. 78-81, May 2016.doi: 10.1109/MC.2016.145

[15] M. Y. uddin and S. Ahmad, "A Review on Edge to Cloud: Paradigm Shift from Large Data Centers to Small Centers of Data Everywhere," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 318-322, doi: 10.1109/ICICT48043.2020.9112457.

[16] Malik, Babur & Mahmood, Faisal & Shahzad, Sohail & Arif, Muhammad & Khan, Waseem & Ilyas, Sadaf & Hassan, Muhammad. (2018). From cloud computing to fog computing in Healthcare big data. MATEC Web of Conferences. 189. 03011. 10.1051/matecconf/201818903011.

[17] S. Ahmad, Afzal M.M., Review of Assured Data Deletion Mechanism in Cloud Computing, International Journal of Engineering & Technology (UAE), Vol 7, No 4.5 (2018), Special Issue 5, 2018.

[18] Ling C, Lifang L, Xiaogang Q, Gengzhong Z (2017) Cooperation forwarding data gathering strategy of wireless sensor networks. Int J Grid Util Comput 8(1):46–52. https://doi-org.sdl.idm.oclc.org/10.1504/IJGUC.2017.10003009

Sultan Ahmad, is currently working as Lecturer in Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia. He has a unique blend of education and experience.

He has received his Master of Computer Science and Applications from the prestigious Aligarh Muslim University, India with distinction marks. He has graduated in Computer Science and Applications in 2002 from Patna University, India.

He has more than 12 years of teaching and research experience. His research and teaching interests include Cloud Computing, Big Data, Machine Learning and Internet of Things. He has presented his research papers in many national and international conferences and published research papers in many peer-reviewed reputed journals.

Mohammad Mazhar Afzal, is an Associate Professor and head of the Computer Science and Engineering Department at Glocal University, where he has been since 2015. From 2008 to 2013 he served at a Government University at KSA.

A [illegible] Masters in Computer Science in Year 1997 he served at Department of Computer Science Maulana Azad college, Aurangabad for nearly a decade. He received his PhD from Dr.Babasaheb Ambedkar Marathawada University Aurangabad(MS). His research interests span both Internet Governance and Network Security. Much of his work has been on improving the understanding, design, and performance of Security systems and various cryptographic techniques. In the networking arena, he has worked on characterizing the Internet and the World Wide Web. In addition, he is always keen on improving and imparting modern methods of teaching and imparting knowledge. He also Served various academic bodies at different capacities. Additionally, he is also working as Director (IQAC) at Glocal University.

Abdullah Alharbi received the Ph.D. degree from the University of Technology Sydney, Australia. He is currently an Assistant Professor with Information Technology department at Taif University. His research interests include human computer interaction, information systems, cybersecurity and data science.