A Modular approach for Big Data Management in IoT using SDN

Ruchi Bhatnagar[1]
School of Computer Science and Applications
IIMT University
Meerut, U.P., India

Prof. Dr. Deepak Sinha[2]
School of Computer Science and Applications
IIMT University
Meerut, U.P., India

Abstract— **The Collaboration of wireless sensor network, embedded systems, advanced Cloud computing and implementation of 5G technology and data analytics crafts a new shape of intelligence that predicts futuristic IoT based smart Technology. The massive integration of numerous technologies causes the considerable aspects of issues, challenges and security within the IoT infrastructure. Software Defined Networking is enabling network technology capable of supporting the dynamic nature of IoT applications while supporting the big data analysis and statistics procedures. This paper proposes an innovative scheme for secure and successful dissemination of big data between nodes and applications and for the dynamic injection of policies at real time; using modular app based edge SDN controller. The proposed algorithm deals with big data activities as per application needs; while highlighting the future research directions along with some open unresolved issues of SDN integration and deployment.**

Keywords—Internet of Things (IoT), Software Defined Networking (SDN), Edge Computing, Big Data

I. INTRODUCTION

Internet of Things; a revolution created in a technology that changes the way of devices working and transforming the life of humans and machines. As new applications are introducing day by day; it requires real-time computing power and Versioning upgrades periodically. In IoT the edge nodes are processed the data locally and then disseminate this data to central controller to reduce overhead on central controller [1]. Bandwidth, costs, speed, automation, maintenance, predictive analytics, remoteness and real time injection of periodical updates are some reasons that demands a faster, cheaper and smarter approach than the traditional one which typically goes like: gather the data, send them through networks to the cloud or other environments where they can get processed and profited [2]. Many solutions have confronted to resolve these issues one of them is proposed that to placement of SDN at edge nodes to meet the expectation of big data control and management in IoT environment.

The countless application of IoT is extending Internet connectivity among billions of devices and produce a huge amount of data, the production of such big data their storage and transformation demand intelligent network control and management solution; SDN is a paradigm considered to overcome the limitations of traditional networks; resolve big data issues and make a perfect solution for IoT scenario [3]. However, in this paper, we propose the modular app based edge SDN solution for big data management and the real time periodical injection of applications upgradation. The organization of remaining sections as follows: Section II describes problem statement; Section III overview the Big data and SDN; Section IV illustrate Proposed Method; Section V conclude the paper with future directions.

II. BACKGROUND OF THE PROBLEM

IoT is said to be a deep-sea of services that feed into big data and consists of millions of devices that collect and communicate information through sensors, but big data incorporates in much wider landscape because this data cannot be processed or even stored using traditional methods and available bandwidth thus requires new approach that makes the network more programmable and easier to configure. The real time applications of big data would not be possible without the underlying support of networking due to their extremely large volume and computing complexity [5]. Traditional networks are also not equipped and certainly weren't designed to deal with the flow of Big Data using common network portfolio. The another problem with traditional networks is that the control plane and data plane are embedded in the same device. As a result, no point in the network has a global view of complete network and this design also makes it difficult to change the network configuration. Essentially, in this configuration each device must be manually aligned to reflect a new configuration. This process becomes even more challenging for such type of network where new nodes are introducing later and becomes large networks. Implementing such new configuration can take large amounts of time and resources also [4]. IoT usually consist of hundreds of sensors generating a huge amount of data; and the acquiring and processing of such data can be very challenging. The origin of huge volumes of data are totally useless, unless they are processed to get something valuable; also it's issues vary depending on the peculiarities of incoming data, expected outcomes and more. Data comes from sensors connected to things either periodically or in streaming. The data received by streaming is essential for real-time data processing and analysing to get accurate results. In IoT scenario the sensing devices transferring the data to gateways which ensure initial data filtering pre-processing and deep analysis hence reducing the volume of data transferred to the next IoT system's blocks and also ensures real-time analytics to quickly recognize useful patterns found in a cloud. But this form of Big data analytics created the problem domain of big data processing in cloud data centres, data delivery, joint optimization, scheduling and security issues [6].

III. Overview of Big Data and Software Defined Networking

The basic concepts of big data and SDN with edge computing as well as their brief interrelationship between big data and SDN describes in this section:

1. Big Data

Big Data is an adaptable term that is universally employed for a large set of data, it contains a massive volume of heterogeneous data in the form of structured, semi-structured and unstructured data. With the capability improvement over the time, new technologies have been developed to store and process larger and larger datasets, from megabyte to gigabyte to terabyte to petabyte to Exabyte. Big data is also defined according to five salient features: volume, variety, velocity, value, and veracity. It is also called the fuel that drives Internet of Things because they both have inter-dependency relationship and hugely impact each other. The historic perspective of big data tied carefully to the capability of storing and processing data sets. In smart scenario as IoT growing, it provides rise to the demand for big data proficiencies. An increase in the amount of data every day requires more advanced and innovative storage solutions resulting in updating an organization's big data management infrastructure.

Big data is very useful in a variety of applications, such as enterprise applications, IoT-based applications, healthcare applications, and scientific research applications. Its interaction with IoT create two likely winners. The first one, the businesses organization that can profit from the information provided, and the end-user who has better information to act accordingly. Eventually, the business parties implemented IoT into their businesses products gets greater profits, more productivity, higher efficiency and reduced costs. The development of big data technologies works in favour of IoT companies, with both seeking to strategize the ways in which we see and utilize data sets. Numerous unfold research challenges remain to be addressed for big data, including analytics platforms, underlying support of networking, data processing models, data staging, distributed storage, and data privacy and security. This paper focused on un resolved issues and challenges related to the underlying support of networking for big data applications.

2. IoT Big Data Architecture

(*sponsors*).It has been clear about that the nature of big data generated by IoT devices, and their processing at IoT architectures in a large centralized fashion by cloud computers. A cloud centric structural design keeps the cloud centric activities at the centre, different applications above it, and the network of smart devices below it [12]. When IoT meets with big data its layer architecture performs following services:

i. Device Layer: This layer equipped with small, memory constraint, battery operated electronic devices with on board sensors and actuators. These devices are able to sense as well as record data.
ii. Network Layer: The services of network layer comprise transfer of device layer data with the help of network devices to the middle ware or data processing layer; that known as cloud layer.
iii. Cloud Layer: It stores, analyses, and processes huge amounts of data that comes from the network layer and it can also manage and provide a diverse set of services such as database management, cloud storage management and big data processing modules to the lower layers.
iv. Business Application Layer: The business application layer manages the whole IoT system, including applications, business and profit models and big data analytics for Business Intelligence.

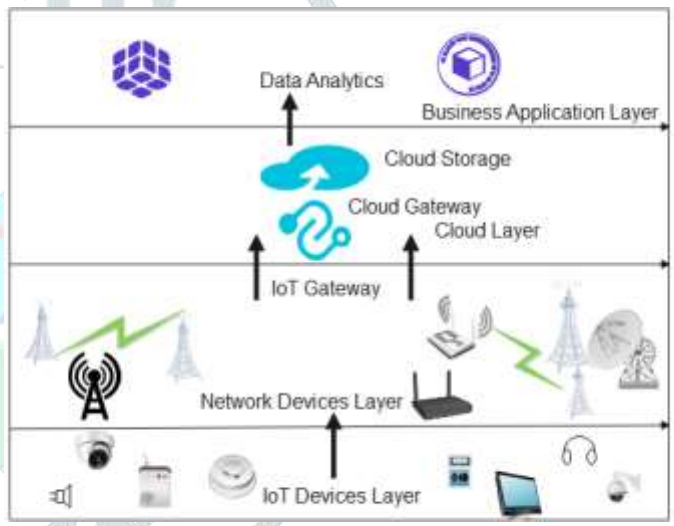

Figure1. IoT Big Data Architecture

3. Software-Defined Networking

The evolution of big data of various applications over the Internet, SDN has been chosen as one of the most exciting solutions to revolutionize the cyber world with proper network management [13]. IoT is a large scale dynamic network and needs to implement new policies or make changes across the whole network by configuring each device manually using system specific commands; In IoT network-wide updates are very complex and time-consuming, and performing the same in continually changing IoT environment makes it more difficult; Software Defined Networking (SDN) paradigm led to provide a practicable and real time solution to this problem. More and more SDN protocol standards have been also developed in real applications. These days, SDN is not partial to any one implementation, but is a general term for the platform. In the SDN architecture main emphasis on that the decoupling of control and data plane, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the application layers. This will greatly support ever changing IoT infrastructure for managing extended new devices, dealing with big data and versioning updates of real time applications.

4. SDN Architecture

This hierarchical architecture of SDN consists of three planes namely Application plane, Control plane, and Data plane. In this paradigm decoupled control plan while the communication between two planes is done through forwarding plane. The brief details for these planes are as follow.

Application or Service Plane (AP): The top most services in the SDN infrastructure; called application plane. It contains business and SDN applications to support various application domains and functionalities and helps to provide user based services. This plane communicates with the control plane using the APIs.

Control Plane (CP): The control plane contains physically separated but logically centralized controllers that run the networking operating system (NOS); maintains the over-all view of the network infrastructure, and provides hardware abstractions to business and SDN applications. The controller acts as a brain with multiple controllers in the distributed environment; and communicate with each other through interfaces, while the control plane and the data plane communicates with each other through APIs.

Data Plane (DP): The bottom up service of SDN is called the data plane contains networking devices such as routers, switches, gateways, and access points, which take actions on the incoming packets (such as forward to specific port, drop, and forward to the controller) based on the forwarding rules to which the packet matches.

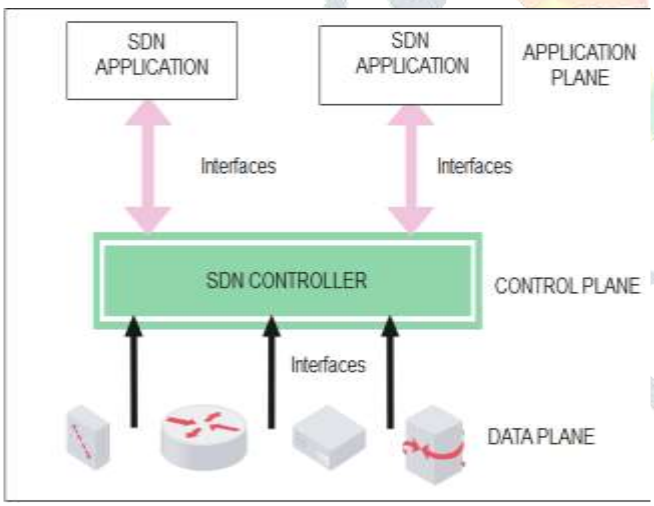

Figure 2. SDN Architecture

This network is complete programmable through software applications running on top of the NOS, which interacts with the underlying data plane devices as well as top most application plane. These good features of SDN can greatly facilitate big data acquisition, transmission, storage, and processing. Recently, SDN has been used to design data centre network architectures to provide better performance with low complexity and energy consumption [14]. Through the collection of traffic data, SDN can dynamically allocate resources to improve application performance. Moreover, there is a new trend to leverage the capabilities of SDN for application-aware networking in data centre applications and network architectures. As big data is usually processed in data centres, SDN-based data centres can benefit big data applications by dynamically. SDN can facilitate high data transmission, processing efficiency, resource allocation and network management for the IoT devices for fulfilling growing need of the customer demands.

IV. PRAPOSED METHOD

A distributed computing topology; which process information located to edge instead of centre cloud; led to the IoT data processing scenario in an innovative way. We propose the placement of modular app based SDN programmable controller on such edge nodes that manages big data generated by huge heterogeneous IoT devices; and also improves the real time injections of versioning updates for some applications. Much data gathers by different IoT devices generally consist of unstructured form and creating an issue to be processed using conventional database; the analytics on such data requires splitting of data across servers; analysis of each data block and then merging of result. Since SDN decouples the data plane from control plane; makes the network custom-designed, flexible and agile; requires for big data management. It maintains the global view of network, provisioning resources as needed to control data flow and balanced work load by changing the optimal data path according to traffic conditions on data plane.

The IoT network contains a diverse set of edge nodes such as access switches, cellular base stations, edge nodes, wireless access points and sensor devices. A controller is located at the cloud and connects to the edge nodes; these edge nodes contains numerous controllers'; and frame multiple edge SDN computing architecture where the data processing and their routing ability according to available bandwidth decided by these SDN controllers'. An only key information passed at central cloud SDN where it can meet the big data challenges and provides key solutions. These multiple edge SDN controller architecture; simple puts the part of the applications needs to be closer to the edge points; manages the data plane according to network traffic; deal with data generation; filtering and their transmission to the central cloud. It can also facilitate generation and join huge amounts of data to drive advanced analytical and artificial intelligence (AI) programs in IoT; and provide reliability and low latency.

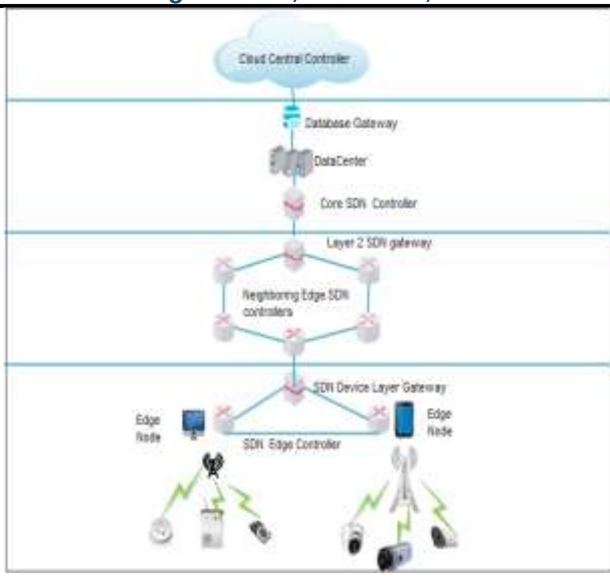

Figure 3. IoT Edge SDN Architecture.

The proposed architecture optimizes big data dealing issues and perform network management by enabling the software modules place on SDN for updating and enhancing network capabilities. The proposed architecture works as following algorithm:

ModuleApp_Algo (Edgenode, SDNcontroller, Appmodule, Datacentre)

Step1. *Collect data from sensor nodes to Edgenode at Layer1*

Step2. *Forward the collected data to SDNcontroller*

Step3. *Filter the data using Appmodule for specific applications*

if (Structured_Data)

Define format to observation and fit it into rule to get filter data

if(Semistructured_Data)

Define Tags and Marking and fit it into observation to get filter

if(Unstructured_Data)

Defined Keyword and Opinion_Tags and fit it into observation to get filter

Step4. *Store Filter data at Edge_SDN controller and forward it to Data_centers*

Step5. *Data_center forward filter data to Data_center_Gatway for Cleaning and Analysing*

Step6. *Statistical_analysed data Store to cloud for Real-time Visualization and updates.*

This modular approach of scalable designing of app data control helps to manage big data deals within the IoT infrastructure. The above ModuleApp_Algo proposes filtering, analysis and forwarding user application data as per the need of big data dealing.

V. CONCERNED ISSUES

The Edge nodes of IoT infrastructure; which provides Shifting of data processing to the edges of the network and helps growing number of IoT edge devices, improve network speeds, and enhance scalability and reliability[14-15]; while the placement of SDN controller with modular programming approach at these nodes make the network alignment as per the need and details of big data and also makes it an ideal solution for fast growing, agile IoT environment; yet faces some unresolved issues and needs future research direction to make this integration more beneficial. Some of concerned issues are:

i. Energy Constraints Nodes:

IoT is an ocean of devices where some of devices are elected as Edge nodes to carry forward the data processing capabilities while the placement of SDN controller on each of these edge nodes is not possible as they are energy constraint and becomes idle soon. Many of the researchers proposes different solutions [17,18] of their placement yet it remain the unresolved issues with many constraint and needs for optimize solutions for the selection of high energy edge nodes.

ii. Load Balancing

As current need of data centres and servers are growing in size; load balancing is a great concern in distributed paradigm [19]. The aim of load balancing is to distribute workloads to a set of reliable servers in the system. Edge nodes which are placed between traditional devices and cloud data centres' helpful for handling low latency and real-time tasks updating activities. Although; it can greatly reduce latency, the unreasonable assignment of tasks leads to the unbalanced load of each node [20].While the placement of SDN controller on these nodes used virtual switching technology monitors the data throughput of each port using variance analysis and then redirects traffic accordingly but because of the diversity and heterogeneity of edge computing nodes, their data offloading schemes of different hotspot nodes creates unresolved issues and led to failure of load balancing techniques and becomes an important research issue and that needs further exploration.

iii. Interconnecting Domain and Interoperability

IoT is not a trend of one technology, it is a pool of different domain interconnection that will shape future world but integration and interconnection of different technology is a big concerned in cloud IoT platform as device compatibility [21], their capabilities, connectivity issues, protocol standardization, security of edge nodes and servers and data intelligence are some unresolved issues taken consideration when interconnecting the different domains.

VI. CONCLUSION

In this paper, we propose the module app based algorithm for application specific data management using SDN controller placement in IoT infrastructure. This proposal yields a set of optimal controller locations and assignment of nodes to controllers with big data applications, including big data processing in cloud data centres using modular app based programming. The data filtering, cleaning, analysing and application specific data delivery to the data centres crafts a new enabling way to deal with big data management. To conclude that the joint design of big data and module app based SDN can become a promising solution for big data networking; improves the performance of big data applications and utilize big data to make SDN work better and more effectively for real time versioning updates for the applications. While some unresolved issues are their will considered in future.

REFERENCES

[1] 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.

[2] Kreutz, D., Ramos, F., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S. and Uhlig, S. 2015. Software-Defined Networking: A Comprehensive Survey. Proceedings of the IEEE. 103, 1 (2015), 14-76.

[3] H. Farhady, H. Lee, A. Nakao; Software-defined networking: a survey Computer Networks, vol. 81, Elsevier B.V. (Apr. 2015), pp. 79-95

[4] Manar Jammal, Taranpreet Singh, Abdallah Shami, Rasool Asal, Yiming Li, “Software defined networking: State of the art and research challenges”,Computer Networks, Volume 72, 2014, Pages 74-98, ISSN 1389-1286.

[5] S. Yu, X. Lin, and J. Misic, “Networking for Big Data,” IEEE Network, vol. 28, no. 4, July 2014, p. 4.

[6] Laizhong Cui, F. Richard Yu, and Qiao Yan , “When Big Data Meets Software-Defined Networking: SDN for Big Data and Big Data for SDN “, 0890-8044/16/$25.00 © 2016 IEEE

[7] D. Kreutz et al., “Software-Defined Networking: A Comprehensive Survey,” Proc. IEEE, vol. 103, Jan. 2015, pp. 14–76

[8] Ahmad I, Namal S, Ylianttila M, Gurtov A. Security in software defined networks: a survey. IEEE Commun Surv Tutor. 2015;17(4):2317-2346.

[9]] W. Hong, K. Wang, and Y.-H. Hsu, “Application-Aware Resource Allocation for SDN-Based Cloud Datacenters,” Proc. Int’l. Conf. Cloud Computing and Big Data 2013, Fuzhou, China, Dec. 2013.

[10] Y. Han et al., “Software Defined Networking-Based Traffic Engineering for Data Center Networks,” Proc. 16th Asia-Pacific Network Operations and Management Symp., Taiwan, Sept. 2014.

[11] Nehra A, Tripathi M, Gaur MS, Battula RB, Lal C. TILAK: token-based prevention plan for threats against topology discovery using random KeyâǍ˙I. Int J Commun Syst. 2018. In press.

[12] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): a vision, architectural elements, and future directions,” *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[13] D. Kreutz et al., “Software-Defined Networking: A Comprehensive Survey,” Proc. IEEE, vol. 103, Jan. 2015, pp. 14–76

[14] W. Hong, K. Wang, and Y.-H. Hsu, “Application-Aware Resource Allocation for SDN-Based Cloud Datacenters,” Proc. Int’l. Conf. Cloud Computing and Big Data 2013, Fuzhou, China, Dec. 2013.

[15] J. Li et al. (2020), A Secured Framework for SDN-Based Edge Computing in IoT-Enabled Healthcare System, IEEE Access, vol. 8, pp. 135479-135490, Available from: doi: 10.1109/ACCESS.2020.3011503.

[16] M. B. Yassein, S. Aljawarneh, M. Al-Rousan, W. Mardini and W. Al-Rashdan(2017), Combined software-defined network (SDN) and Internet of Things (IoT), International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates, pp. 1-6, Available from: doi: 10.1109/ICECTA.2017.8252003.

[17] B. Heller, R. Sherwood, N. McKeown, “The Controller Placement Problem”, in Proc. HotSDN, 2012.

[18] Hidenobu Aoki, Norihiko Shinomiya,2016 ,”Controller Placement Problem to Enhance Performance in Multi-domain SDN Networks”, IARIA, 2016. ISBN: 978-1-61208-450-3

[19] Kumar, P.; Kumar, R. Issues and challenges of load balancing techniques in cloud computing: A survey. ACM Comput. Surv. 2019, 51, 120. [CrossRef]

Li, G., Yao, Y., Wu, J. *et al.* A new load balancing strategy by task allocation in edge computing based on intermediary nodes. *J Wireless Com Network* **2020,** 3 (2020).

[20] Noura, M., Atiquzzaman, M. & Gaedke, M. Interoperability in Internet of Things: Taxonomies and Open Challenges. Mobile Netw Appl 24, 796–809 (2019https://doi.org/10.1007/s11036-018-1089-9