A STUDY OF CLOUD PROCESSING IN WIRELESS IOT NETWORKS

AVINASH KUMAR

Research Scholar, Dept. of Computer Science, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal-Indore Road, Madhya Pradesh, India,

Dr. Satendra Kurariya

Research Guide, Dept. of Computer Science, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal Indore Road, Madhya Pradesh, India.

ABSTRACT

The Internet of Things (IoT) is spreading much faster than the speed at which the supporting technology is maturing. Today, there are tens of wireless technologies competing for IoT and a myriad of IoT devices with disparate capabilities and constraints. Moreover, each of many verticals employing IoT networks dictates distinctive and differential network qualities. In this work, we present a context-aware framework that jointly optimises the connectivity and computational speed of the IoT network to deliver the qualities required by each vertical. Based on a smart port application, we identify energy efficiency, security, and response time as essential quality features and consider a wireless realisation of IoT connectivity using short range and long-range technologies. We propose a reinforcement learning technique and demonstrate significant reduction in energy consumption while meeting the quality requirements of all related applications.

KEYWORDS: Cloud Processing, Wireless, IOT Networks, IoT devices, IoT network, IoT connectivity.

INTRODUCTION

The Internet of Tings (IoT) is today's buzzword, often coupled with Big Data and Artificial Intelligence (AI). However, there is a lot of ambiguity of what is meant by that and skepticism about the actual value generated by the IoT. IoT devices have become pervasive but cover a broad range of technologies and standards. Wireless technology is key to connect these devices through gateways or aggregation points; but, similarly, a wide range of wireless protocols and standards are available and competing [1]. Once these devices are connected, they start reporting the sensed or measured data to the platform. Again, multiple choices are possible in this aspect with different strengths and weaknesses. Reporting raw data to the cloud

is very costly as every bit gets charged and may also exhaust the battery of the device; this results in massive data. On the other hand, running scripts locally in the device and reporting the resulting events to the cloud reduce the cloud service cost but limits the visibility to the actual data; this still results in big data. Moreover, local scripts result is real-time actions and do not expose the privacy of the data, whereas cloud computing incurs latency due to the transmission network and requires stringent security measures to protect the data. An environment, which is rich in IoT devices that are connected to a platform, qualifies as digitized, and often as intelligent. Analytics, which uses AI, is the added layer that transforms such an environment into a smart one. The default application of AI is to draw actionable insights from the data in order to generate value to the given vertical. In this work, we argue that IoT solutions should not be addressed through a layered perspective but, instead, a holistic optimisation approach is needed to generate the desired added value efficiently. In such a holistic approach, AI, among other machine learning tools, is employed in every stage of the solution including connectivity, storage, computing, and analytics. Since there are many usecases of the IoT paradigm [2], it should be approached from a given vertical perspective, e.g., smart health, smart cities, smart manufacturing (Industry 4.0), smart transport, etc. Each of these verticals comprises multiple IoT-based applications with various requirements. In [3], for example, signaling measurements and modelling are performed for both static and vehicular machine-to-machine (M2M) applications, asboth have different signaling overhead characteristics. As another example, remote monitoring in smart cities requires full compliance with privacy regulations, whereas security related applications rank response time highest among all key performance indicators (KPI).

EDGE COMPUTING VERSUS CLOUD COMPUTING

Existing wireless networks and Cloud/Centralized Radio Access Networks (C-RAN) under investigation are mainly designed to deliver contents without having the capability of analyzing and making use of the data-specific features in optimizing the system performance [15]. In this regard, it is important to make existing wireless networks scalable in order to handle massive data contents. Also, it is crucial to investigate suitable network architectures/mechanisms to incorporate and utilize big data awareness in wireless networks in order to enhance the system performance. As the amount of data generated by a large number of distributed sensors is highly unstructured and heterogeneous in nature, it becomes extremely complex to handle them with the conventional approaches. Mainly, how to acquire, integrate, store, process and utilize the data in highly distributed environments has been an important challenge for researchers, engineers and data scientists. To address these issues, CISCO has recently proposed the concept of fog computing which aims to support cloud computing platform in handling a part of the workload locally at the edge devices such as switches, routers, and IP-enabled video cameras instead of transmitting the whole work load to the cloud. The fog/edge layer, which may comprise of several edge servers distributed over various places such as shopping centres, parking areas and bus stations. The edge server can be regarded as a low-capacity

version of the cloud server and has communication, computing and data storage capabilities. Edge computing becomes highly advantageous for mobility support, geo-distribution, and location/context awareness. The geo-distributed nature of edge computing helps to provide rich contextual information such as event status, local network conditions, and the end-user's status. These information can be subsequently used for context-aware optimization of edge/fog applications. More specifically, edge computing can support the existing cloud computing platform in handling the following different types of applications, whose requirements cannot be met with the cloud processing.

COORDINATION MECHANISMS BETWEEN EDGE COMPUTING AND CLOUD COMPUTING

In the proposed collaborative edge-cloud architecture, communication, computing and storage functionalities need to be dynamically allocated among the edge-side units, cloud and the things (devices/sensors) in order to handle the massive IoT data in the real-time. Besides, there is a strong need to have the coordinated management of cloud computing and edge computing units in handling massive IoT connections in a reliable and secured manner [8]. In other words, effective coordination mechanisms between edge processing units and cloud processing units are crucial in order to realize this architecture effectively. In order to enable these interactions, suitable interfaces between edge processing unit and the cloud centre, among different edge processing units, and between edge processing units and IoT devices/objects need to be defined. Moreover, in the proposed coordinated platform, the cloud center is assumed to be capable of dictating the edge computing units about which parameters are important to monitor, how frequent a parameter needs to be monitored and which task should be processed to meet the real-time requirements of end-users. In this regard, various aspects such as coordination framework definition, parameters to be coordinated, control signaling design from the cloud center, and load balancing at the edge and cloud sides need to be investigated.

Some of the research challenges and future research directions under this research topic are provided below.

• **Definition of edge-cloud interaction mechanisms:** Various aspects such as what should be the suitable interfaces between cloud and edge, between cloud and things, between edge and things/devices and between different edge computing units, which computing and communication parameters will be involved in establishing relationship between edge and cloud units, and how to design common control signaling and management schemes in order to control edge units from the cloud, need to be investigated under the proposed framework.

• **Resource mapping between edge and cloud computing:** In order to enable effective collaborations between edge and cloud units, it is important to investigate suitable techniques to map edge-side computing/communication resources with the cloud-side resources, and also suitable strategies to share resources among multiple edge units for handling live data analytics.

• Load balancing among edge, cloud and things: It is crucial to investigate which tasks can be handled at the edge gateway and which tasks should be sent to the cloud, which tasks can be handled at the things/devices, in order to effectively optimize the available edge and cloud resources in the proposed platform.

• **QoS enhancement schemes:** In the considered cloud edge coordinated platform, it is important to identify the performance bottlenecks (communication bandwidth, cache size, computing power, etc.) and then to develop suitable techniques to minimize service response time and to improve reliability in the case of network/link failures.

BIG-DATA AWARE EDGE-CLOUD COLLABORATIVE PROCESSING

One promising way of dealing with the big data in wireless IoT networks is to understand the features of big data and to incorporate this awareness in order to enhance the system performance [15]. In contrast to multimedia contents, IoT data has several peculiar features such as bursty nature, small data-size and transiency, i.e., they expire in short time. Besides, in the proposed framework, only enriched data with the meaningful information can be forwarded to the cloud instead of sending all the raw data to the cloud. In this direction, it is interesting to explore model-based data processing techniques such as data compression, device cooperation, and distributed coding/encoding by extracting certain features of IoT data such as temporal and spatial correlations, and social relations. For example, the data collected from IoT sensor nodes can be compressed at the aggregator node before forwarding to the cloud, and also only certain features of the raw data can be extracted and sent to the cloud for some specific applications. One of the main tasks of IoT sensors is to sense various parameters of the environment under the practical constraints such as lowcost, low-power and weak processing capability. In the proposed architecture, it is crucial to optimize edge computing resources at the IoT gateway along with the communication resources such as transmit power, bandwidth and antennas in order to meet the desired performance metrics of the proposed system such as latency, data rate and energy efficiency. Also, in various IoT applications such as smart health and smart car, it is important to acquire and process contextual information such as location and speed to provide meaningful information. In this regard, the cloud centre can facilitate the optimization of edge-side processing under the proposed framework. For example, sampling rate to be employed at the edge-node depends on the characteristics of information (such as radio spectrum usage) to be acquired from the environment, and higher sampling rate causes more power consumption and also requires costly equipment. In this example, the cloud centre can provide guidance to the edge-nodes on the use of suitable sampling rate based on its network wide intelligence as well as history information in order to reduce the power consumption. Besides, other transmission and operating parameters at the edge-nodes such as transmit power; modulation and coding can be adapted based on the feedback provided by the cloud-centre. Another key challenge in the proposed framework is how to minimize the closed-loop (end-to-end) latency in order

to process IoT big data in the real-time. In practice, the closed loop latency of the considered system should be within the data coherence time, i.e., lifetime of the data. By employing suitable data prioritization techniques, the processing of the prioritized data can be handled at the edge computing units to provide faster response to the end-users. In addition, by employing caching techniques at the edge computing units in the proposed platform, content delivery efficiency can be maximized and front haul/backhaul bandwidth overhead can be significantly reduced as in 5G wireless networks. During the off-peak hours, popular contents can be pre-fetched to the edge-side and during the peak periods, the delivery phase has to only deal with the transmission of additional contents requested from the users. The content popularity can be predicted using the available historical big data at the cloud side, and then a suitable cache replacement policy can be employed on the basis of the predicted content popularity distribution.

Based on the above discussion, we highlight some of the research directions under this topic below.

• Energy-efficient caching strategies: Various aspects such as time-varying popularity estimation, popularity modeling/prediction based on historical data, cache placement, delivery and replacement strategies, different performance tradeoffs such as data rate versus cache size (memory), latency versus memory and energy versus delay, cooperative and coded caching can be investigated in the proposed framework.

• Edge-side data acquisition and processing techniques: At the edge-side of the proposed platform, various cloud-assisted data processing techniques suchas data filtering, compression and feature extraction techniques, cooperative sensing/monitoring/acquisition, distributed encoding/decoding, combining and decision making schemes can be investigated.

• Closed-loop latency minimization: In order to minimize closed-loop latency in the proposed framework, suitable device/node/task/data prioritization techniques can be investigated under various practical constraints such as transmit power and backhaul/front haul bandwidth based on different criteria such as residual energy, requirement, emergency level, and expected delay.

• Adaptive learning/prediction algorithms: Several issues such as how to make the best use of available prediction algorithms (Neural network, artificial intelligence, support vector machine, clustering, regression analysis, etc.) in different application scenarios and how to adapt algorithms based on time-varying situations are promising future directions.

TASK/DATA/COMPUTATION/PROGRAM OFFLOADING

It is understood that huge amount of IoT data cannot be handled at the edge-side and need to be offloaded to the cloud-side. Besides, resources at the edge side such as transmit power, bandwidth and computing power need to allocate efficiently to handle real-time applications at the edge-side. Latency tolerant and large-scale

www.jetir.org (ISSN-2349-5162)

tasks can be processed efficiently at the cloud centre while it is advantageous to process latency-sensitive tasks at the edgeside. The transfer of computational load/data/task from the edge-side to the cloud in the edge cloud framework can be enabled by employing proposed collaborative suitable task/data/computation/program offloading techniques at the IoT edge gateway/aggregator. In this regard, it is crucial to take effective decisions about which data/task to offload to cloud and which data/task to be processed at the edge-side based on the guidance from the cloud center. By using suitable data prioritization techniques, critical data (which needs real-time treatment) can be identified and processed at the edge-side to provide faster response to the end-users. The prioritized data can be later sent to the cloud for further processing and storage for future usage. In this regard, suitable data classification, data prioritization, data/task partitioning/scheduling, task/data offloading with backhaul constraints (limited bandwidth, delay, and power consumption) and delay-based node prioritization need to be investigated under the proposed framework. Besides, since there is a huge amount of cost involved in renting cloudlets from the cloud providers, it is essential to investigate the tradeoff between service execution time and the cost required to use the cloud resources.

Based on the above discussion, we provide some research topics under this theme below.

• Task scheduling techniques: In the proposed framework, suitable task scheduling techniques need to be investigated for multiple dependent and interdependent tasks under the constraints of cloud/edge computing resource cost and service/workflow execution time deadline.

• Data offloading techniques: The investigation of suitable data/task offloading techniques under backhaul and edge-cloud resource constraints and decision making strategies on which data to offload to the cloud and which to process at the edge-side are important future research directions.

• Code partitioning techniques: In order to maximize the overall computational efficiency of the proposed framework, it is crucial to investigate which parts of the code to be run locally at the edge-side, and which parts should be offloaded to the cloud and at what state of the program considering practical constraints such as battery level, delay constraint and channel state.

• **Performance tradeoffs:** Various performance tradeoffs such as offloading gain versus energy, and cost of cloud resources versus service execution time under the proposed framework are interesting aspects to be investigated in future works.

ADAPTIVE OPTIMIZATION OF COMPUTING, COMMUNICATIONS AND CACHING RESOURCES

In contrast to the traditional way of managing computing and communication resources in a separate manner, future wireless IoT networks require novel solutions towards the adaptive optimization of

www.jetir.org (ISSN-2349-5162)

computing, storage and communication resources at both the edge and cloud sides in order to deal with the massive amount of heterogeneous data. Besides, in the considered cloud-based IoT environment, the overall system dynamics vary due to different causes such as device movements, system parameter variations, and wireless channel variations, thus making the designed system unstable over the time. In this context, it is important to adapt the system model and decisions/control actions to be taken based on the global knowledge available at the cloud side and instantaneous information collected at the edge-side. Moreover, objective functions can be adapted based on the varying state of the system as well as instantaneous requirements from the end-users. Besides, in order to facilitate real-time collaboration between the cloud computing and edge computing units, it is crucial to optimize the involved backhaul/feedback links under the transmission bandwidth constraint without compromising the quality of the links. In this direction, suitable techniques for designing low-latency flexible backhaul links under the constraints of end-user QoS requirements can be developed by utilizing the concepts of software defined wireless networks. Moreover, the interactions between various layers of protocol stacks need to be considered in devising end-end system reliability. In this direction, various cross-layer techniques such as modulation and coding adaptation at the physical layer, collision free access mechanisms at the access layer and novel routing algorithms at the network layer can be investigated for the proposed architecture.

Based on the above discussion, we highlight some of the interesting research topics below.

• Characterization of system resources: In the proposed framework, it is important to identify all the involved communication, caching and computing resources, and then derive link/system capacity or any other suitable metrics by considering all these resources into account in order to provide the overall characterization of the system.

• Adaptive optimization problems: Some examples include latency minimization under the constraints of computational rate, transmit power minimization under the constraints of computational rate, preceding matrices/beam forming vector optimization under the constraints of computational rate, and the optimization of processing power under the constraints of latency, bandwidth and transmit power.

• **Performance tradeoffs:** In the considered framework, several performance tradeoffs such as caching gain and memory size/cache content size; computing and communication delays can be investigated.

• Joint optimization of system resources: Suitable joint optimization solutions can be investigated under the proposed framework in order to jointly optimize the caching (cache size), computing resources (processor speed, memory size, computational rate/power) and communication resources (bandwidth, transmit power, energy, latency etc.). Besides, suitable multi-objective optimization solutions can be developed in order to address conflicting objectives in designing a wireless IoT network.

AUTONOMOUS DEVICE COLLABORATIONS IN WIRELESS IoT NETWORKS

Device to Device (D2D) communication has got importance in various practical scenarios and plays an important role in the proposed edge-cloud coordinated framework in Fig. 3. The increasing context-aware applications require location awareness and discovery functionalities and need to communicate with other neighboring nodes. Also, D2D plays an important role in enabling the sharing of resources such as contents, applications, processing power and spectrum among spatially-closed resource-constrained nodes in order to enhance the overall system performance. In addition, D2D communication is of vital importance to form emergency communication network in disaster scenarios such as earthquake or hurricane. In such disaster scenarios, D2D communications has to take care of various aspects such as unpredictability, limited resources in disaster areas and dynamically varying environment. In wireless IoT networks involving different types of things/objects such as home appliances, body area sensors, smart-phones and environmental monitoring sensors, a series of communication is required to realize the process of creating, collecting and sharing information among nearby devices. Due to their heterogeneous capabilities, devices can help each other with the help of suitable collaborative strategies among them (for example, in a smart home, home audio, lighting bulbs, and alarming systems can collaborate to indicate the emergency level). Similar to human social network in connecting users via Internet, one device can be connected with other devices based on its social relationships with them, thus leading to the concept of social IoT. By exploring the social relations of the devices, device collaboration can be initiated without human intervention. The heterogeneity and the diversity of the connected smart devices impose challenges in handling interoperability among IoT devices. During the operation phase, there arise several challenges such as how to maintain seamless connections among the devices, how to attach a new device to the network, how to rediscover a device in the network, how to form a collaborative cluster, and which device to disconnect in case of faulty conditions and security threats. By grouping correlative devices under the same collaborative cluster, efficient management strategies can be employed at the IoT gateway in order to control them, and also to enable device collaborations. In this direction, the key issues to be considered are to understand the social relations among the devices, to classify devices based on a suitable basis, to track the location of the devices/inhabitants, and to design effective device collaborative strategies based on the inferred information with the aim of enhancing the zero configuration index of device interaction.

SOFTWARE DEFINED NETWORKING AND VIRTUALIZATION IN WIRELESS

IoT NETWORKS In the existing networks, the main problems are vendor specific interfaces and software associated with hardware, complex and expensive network operation, and the tight coupling of data and control planes. Besides, the network cannot dynamically adapt based on the network conditions. To address the aforementioned issues, the emerging concept of SDN can be employed. Besides, virtualization technology can be used to form a virtual network on the top of the existing networks, which shield the user from the underlying hardware and become adaptable to diverse technologies and protocols. However, in contrast to the traditional virtualized wired networks, radio resource abstraction and isolation in wireless

www.jetir.org (ISSN-2349-5162)

networks is challenging due to time-varying channel, broadcast nature, mobility and heterogeneous access technologies. In heterogeneous IoT environment, dynamic resource discovery and the sharing of available resources can enable the creation of effective virtual networks. However, one of the critical challenges is the efficient management of the physical resources allocated to virtual networks. How often resource discovery and allocation need to be performed is also another challenge to be addressed. Besides, mapping of the physical resources to the logical resources in order to embed a virtual network on the existing physical networks in an effective way is another key aspect to be investigated. The creation of a virtual network requires effective interactions among the involved entities at different hierarchical levels of the wireless IoT network. This requires the need of defining suitable interfaces as well as proper control signaling which can be adaptable among heterogeneous wireless access technologies. For control signalling, mechanisms such as IP-based signalling and dedicated channel assignment can be investigated considering both network overload and delay. Besides, naming the devices with logical identifiers, mapping between physical and logical addresses, slicing, device attachment and dynamic routing of the devicetraffic are other important operations to be considered. Furthermore, another key research aspect is to examine the possibility of employing various levels of slicing such as spectrum-level slicing, network-level slicing and flow-level slicing in the wireless IoT networks. In addition, investigating suitable admission control policies in order to control the admission of incoming users/devices while guaranteeing the QoS of the existing users/devices is another aspect to be considered.

CONCLUSION

Cloud computing and edge computing are considered as two emerging paradigms in handling the massive amount of distributed data generated by IoT devices. However, these paradigms have their own advantages and disadvantages. Cloud computing provides a centralized pool of storage and computing resources and has a global view of the network but it is not suitable for applications demanding low latency, real-time operation and high QoS. On the other hand, edge computing is suitable for the applications which need real-time treatment, mobility support, and location/context awareness but does not usually have sufficient computing and storage resources. Taking these aspects into consideration, this paper has proposed a novel framework of collaborative edge-cloud processing for enabling live data analytics in wireless IoT networks. The basic features, key enablers and the challenges of big data analytics in wireless IoT networks have been described and the main distinctions between cloud and edge processing have been presented. Furthermore, potential key enablers for the proposed collaborative edge-cloud computing framework have been identified and the associated key challenges have been presented in order to foster future research activities in this domain.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[2] S. K. Sharma, T. E. Bogale, S. Chatzinotas, X. Wang, and L. B. Le, "Physical layer aspects of wireless IoT," in Proc. Int. Symp. Wireless Commun. Syst. (ISWCS), Sep. 2016, pp. 304–308.

[3] P. Fan, "Coping with the big data: Convergence of communications, computing and storage," China Commun., vol. 13, no. 9, pp. 203–207, Sep. 2016.

[4] H. Liu, Z. Chen, and L. Qian, "The three primary colors of mobile systems," IEEE Commun. Mag., vol. 54, no. 9, pp. 15–21, Sep. 2016.

[5] S. Andreev et al., "Exploring synergy between communications, caching, and computing in 5G-grade deployments," IEEE Commun. Mag., vol. 54, no. 8, pp. 60–69, Aug. 2016.

[6] J. Tang and T. Q. S. Quek, "The role of cloud computing in contentcentric mobile networking," IEEE Commun. Mag., vol. 54, no. 8, pp. 52–59, Aug. 2016.

[7] P. Corcoran and S. K. Datta, "Mobile-edge computing and the Internet of Things for consumers: Extending cloud computing and services to the edge of the network," IEEE Consum. Electron. Mag., vol. 5, no. 4, pp. 73–74, Oct. 2016.

[8] X. Masip-Bruin, E. Marn-Tordera, G. Tashakor, A. Jukan, and G. J. Ren, "Foggy clouds and cloudy fogs: A real need for coordinated management of fog-to-cloud computing systems," IEEE Wireless Commun., vol. 23, no. 5, pp. 120–128, Oct. 2016.

[9] C. Vallati, A. Virdis, E. Mingozzi, and G. Stea, "Mobile-edge computing come home connecting things in future smart homes using LTE deviceto-device communications," IEEE Consum. Electron. Mag., vol. 5, no. 4, pp. 77–83, Oct. 2016.

[10] M. Satyanarayanan, "The emergence of edge computing," Computer, vol. 50, no. 1, pp. 30–39, Jan. 2017.

[11] S. H. Park, O. Simeone, and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks," IEEE Trans. Wireless Commun., vol. 15, no. 11, pp. 7621–7632, Nov. 2016.

[12] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," IEEE Internet Things J., vol. 3, no. 6, pp. 854–864, Dec. 2016. [13] S. Yin and O. Kaynak, "Big data for modern industry: Challenges and trends [point of view]," Proc. IEEE, vol. 103, no. 2, pp. 143–146, Feb. 2015.

[14] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," IEEE Access, vol. 2, pp. 652–687, Jul. 2014.

[15] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," IEEE Commun. Mag., vol. 53, no. 10, pp. 190–199, Oct. 2015.

[16] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, and R. C. Qiu, "Big data analytics in mobile cellular networks," IEEE Access, vol. 4, pp. 1985–1996, 2016.

[17] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-based big data storage systems in cloud computing: Perspectives and challenges," IEEE Internet Things J., vol. 4, no. 1, pp. 75–87, Feb. 2017.

[18] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "Threats to networking cloud and edge datacenters in the Internet of Things," IEEE Cloud Comput., vol. 3, no. 3, pp. 64–71, May 2016.

[19] J. A. Stankovic, "Research directions for the Internet of Things," IEEE Internet Things J., vol. 1, no. 1, pp. 3–9, Feb. 2014.

[20] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," IEEE Trans. Signal Process., vol. 31, no. 5, pp. 32–43, Sep. 2014.