A Review on Edge Computing

Rahul Patel[1] Nupur Agrawal[2] Avinash Singh Bundela[3]
[1,2,3]Assistant Professor
[1,2,3]Acropolis Institute of Technology & Research, Indore, India

Abstract— In recent years, with the proliferation of the internet of Things (IoT) and the wide penetration of wireless networks, the quantity of edge devices and the knowledge generated from the edge are growing quickly. According to International Data Corporation (IDC) prediction [1], global information can reach one hundred eighty Zetta-bytes (ZB), and 70th of the data generated by IoT will be processed on the edge of the network by 2025. IDC also forecasts that more than 150 billion devices are going to be connected worldwide by 2025. During this case, the centralized processing mode supported cloud computing isn't economical enough to handle the data generated by the edge. The centralized processing model uploads all information to the cloud data center through the network and leverages its supercomputing power to solve the computing and storage problems, that allows the cloud services to make economic advantages. This paper provides state-of-threat coverage of edge computing topics and highlights the present challenges and future opportunities in this area.

Keywords: Edge computing, Internet of Things (IoT), Amazon Web Services (AWS)

I. INTRODUCTION

Edge computing is a new paradigm within which the resources of an edge server are placed at the edge of the net, in close proximity to mobile devices, sensors, end users, and therefore the rising IoT. Terms like “cloudlets,” “micro datacenters,” and “fog” are used in the literature to seek advice from these types of small, edge-located computing hardware. all of them represent counterpoints to the theme of consolidation and massive data centers that have dominated discourse in cloud computing. Shi et al. [2] Outlined edge computing as follows. Edge computing refers to the enabling technologies allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services and upstream data on behalf of IoT services. “Edge” is defined as any computing and network resources along the trail between data sources and cloud data centers, and edge may be a continuum. The edge of the net may be a unique place. situated typically just one hop away from associated end devices, it offers ideal placement for low-latency offload infrastructure to support rising applications such as increased reality, public safety, connected and autonomous driving, Smart manufacturing, and health care. It is an optimum web site for aggregating, analyzing, and distilling bandwidth hungry sensor data from devices such as video cameras. New challenges and opportunities rise because the consolidation of cloud computing meets the dispersion of edge computing. However, within the context of IoT, traditional cloud computing has many shortcomings:

1) Latency: Novel applications within the IoT situation have high real-time requirements. Within the traditional cloud computing model, applications send data to the data center and obtain a response that will increase the system latency. as an example, high-speed autonomous driving vehicles require milliseconds of reaction time. Serious consequences can occur once the system latency exceeds expectations because of network problems.
2) Bandwidth: transmitting massive amounts of information generated by edge devices to the cloud in a real-time manner can cause great pressure on network bandwidth. as an example, Boeing 787 generates more than 5 GB/s of information, but the bandwidth between an aircraft and satellites is insufficient to support real-time transmission [10].
3) Availability: As more and more Internet services are deployed on the cloud, the availability of these services has become an integral part of everyday life. For example, smart phones users who get accustomed voice-based services, e.g., Siri, can feel frustrated if the service is unavailable for a brief amount of time. Therefore, it is an enormous challenge for cloud service providers to stay the 24 ×7 promise.
4) Energy: data centers consume a lot of energy. According to Sverdlik’s research [14], the energy consumption of all data centers in the United States will increase by4% by 2020, reaching 73 billion kilowatt-hours. With the increasing amount of computation and transmission, energy consumption will become a bottleneck restricting the development of cloud computing centers.
5) Security and Privacy: data in the interconnection of thousands of households are closely associated with users’ lives. as an example, indoor cameras sending video data from the house to the cloud will increase the chance of leaking users’ non-public data. With the enforcement of EU General Data Protection Regulation (GDPR) [15], information security and privacy issues became more important for cloud computing companies.

These challenges have pushed the horizon of edge computing, which calls for processing the data at the edge of the network. It has developed rapidly since 2014 with the potential to reduce latency and bandwidth charges, address the limitation of computing capability of the cloud data center, and increase availability as well as protect data privacy and security. Next, we describe what functions will be performed at the edges, and how edge computing fits in today’s cloud computing model by presenting a typical three-tier edge computing model. By analyzing several representative application scenarios of edge computing, in Fig. 1, we abstract a typical three-tier edge computing model: IoT, edge, and cloud. The first tiers IoT, including drones, sensors in the connected health area, devices and appliances in the smart home, and equipment in the industrial Internet. Multiple communication protocols are used to connect IoT and the second tier, edge. For example, drones can connect to a cellular tower by 4G/LTE, and sensors in the smart home can communicate with the home gateway through WiFi. Edge, including connected and autonomous vehicles, cellular tower, gateway, and edge servers, requires the huge computing and storage capabilities of the cloud to complete complex tasks. The protocols between IoT and the edge usually have the characteristics of low power consumption and short distance, while the protocols between the edge and the cloud have large throughput and high speed. The Ethernet, optical fibers, and the coming 5G are the preferred communication protocols between the edge and the cloud.

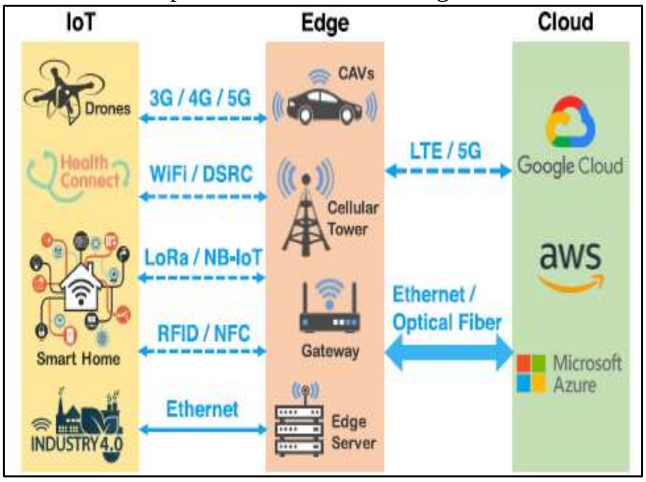

Fig. 1: Three-tier Edge Computing Model

II. LITERATURE REVIEW

The field of edge computing has developed rapidly in recent years since 2014. We categorize the development process into three stages: technology preparation period, rapid growth period, and steady development period. We use "edge computing" as the keyword to search the number of articles published per year in Google Scholar. As shown in Fig. 2, before 2015, edge computing was in the technology preparation period. Since 2015, the number of papers related to "edge computing" has grown tenfold. Edge computing has entered the rapid growth period. Note that the number of papers in 2019 is estimated based on the results of the first five months. We predict that edge computing will continue to develop rapidly until 2020. After 2020, edge computing will step into the steady development period. In this period, edge computing will realize the integration of academia and industry, bring the product into the business, and finally facilitate peoples' daily lives. Fig. 3 illustrates typical events in the development process of edge computing.

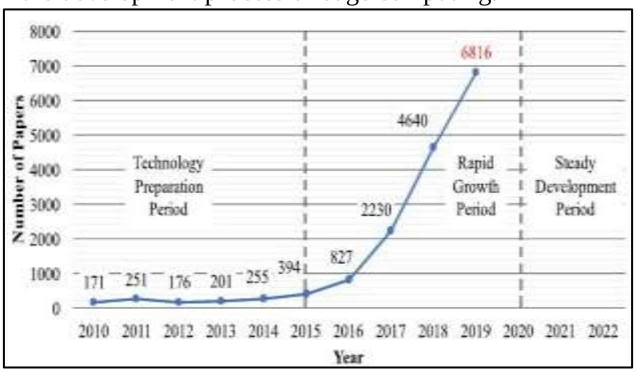

Fig. 2: Number of papers related to by "edge computing" on Google Scholar.

During the technology preparation period, edge computing went through the development process of potential, presentation, definition, and generalization. Edge computing can be traced back to the content delivery network (CDN) proposed by Akamai in 1998 [3]. CDN is an Internet-based caching network, which relies on caching servers deployed in different places and points users' access to the nearest caching server through load balancing, content distribution, scheduling, and other functional modules of the central platform. Therefore, CDN can reduce network congestion and improve user access response speed and hit rate. CDN emphasizes the backup and caching of data, while the core idea of edge computing focuses more on function caching. Ravi et al. [4] first proposed the concept of function cache at the first time and applied it to personalized mailbox management services to save latency and bandwidth. Satyanarayanan et al. [5] put forward the concept of Cloudlet, which is a trusted and resource-rich host, deployed on the edge of the network, connected to the Internet, and can be accessed by mobile devices to provide services. Cloudlet is also known as "small cloud" as it can provide services for users, similar to the cloud server. At this point, edge computing emphasized downstream, that is, it down streamed the functions from cloud servers to edge servers to reduce bandwidth and delay.

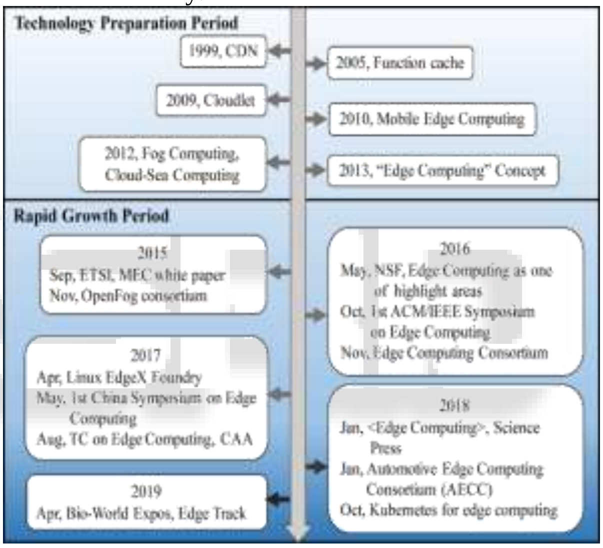

Fig. 3: Typical events in the development process of edge computing.

The rationale of edge computing is that computing should happen at the proximity of the data source with the "edge" constituting any computing and network resources along the path between data sources and the cloud [8]. In this context, sensory data are converted from raw signals to contextually relevant information in the proximity of the data source. Subsequently, edge data ushered in an explosive growth in the context of IoT. In order to address the challenges of computing offload and data transmission, researchers began to explore how to increase data processing capability near the data producer.

The representative computing models are mobile edge computing (MEC), fog computing, and cloud-sea computing. MEC [9] is a network structure that provides information services and cloud computing capabilities within the wireless access network near mobile users. Since MEC is located in a wireless access network and close to mobile users, it can achieve lower latency and higher bandwidth to improve the quality of service and user experience. MEC emphasizes the establishment of edge servers between the cloud server and edge devices to do computing, which is similar to the architecture and hierarchy of an edge-computing server, so MEC is regarded as an important part of edge computing. Cisco introduced fog computing in 2012 and defined fog computing as a highly virtualized computing platform for migrating cloud computing center tasks to network edge devices [10]. It relieves the bandwidth load and energy consumption pressure of main links by reducing the number of communications between cloud computing centers and mobile users. Fog computing and edge computing have great similarities, but fog computing focuses more on communication optimization at the infrastructure level, while edge computing pays attention to computing needs and network demand of both end devices and infrastructure, including the collaboration among end devices, edges, and clouds.

Meanwhile, in 2012, the Chinese Academy of Sciences launched a ten-year strategic priority research initiative called the Next Generation Information and Communication Technology (NICT) initiative. Its main purpose is to carry out research of the "Cloud-Sea Computing System Project" [9]. It aims to augment cloud computing by cooperation and integration of the "cloud computing" system and the "sea computing" system. "Sea" refers to an augmented client side consisting of human facing and physical world facing devices and subsystems. Cloud-sea computing focuses on the two ends "sea" and "cloud," while edge computing focuses on the data path between "sea" and "cloud." In 2013, Ryan La Mothe from the Pacific Northwest National Laboratory proposed the term "edge computing" in a two-page internal report, which is the first time modern "edge computing" [11] was formulated. At this time, the concept of edge computing includes both the downstream of cloud services and the upstream of IoT.

Since 2015, edge computing has been in a rapid growth period, attracting intensive close attention from academia and industry. At the government level, in May 2016, the National Science Foundation (NSF) of the United States listed edge computing as one of the highlighted areas in the research of computer systems. In August 2016, NSF and Intel formed a partnership in information center networks in wireless edge networks (ICN-WEN) [12]. In October 2016, the NSF held the NSF Workshop on Grand Challenges in edge computing [13]. The workshop focused on three topics: the vision of edge computing in the next five to ten years; the grand challenges to achieving the vision; and the best mechanisms for academia, industry, and the government to attack these challenges in a cooperative way. This indicates that the development of edge computing has attracted great attention at the government level. In academia, Shi et al. [2] gave a formal definition of edge computing in the paper "Edge computing: Vision and challenges." They defined edge computing as enabling technologies allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services and on upstream data on behalf of IoT services. This research pointed out the challenges of edge computing and has been cited more than 1000 times in three years. In October 2016, ACM and IEEE jointly organized the first ACM/IEEE Symposium on Edge Computing (SEC). Since then, ICDCS, INFOCOM, Middle Ware, and other important international conferences have added an edge computing track and/or workshops to their main conferences.

Although edge computing is very promising, there are still many challenges faced by the community, ranging from fundamental technologies to novel application scenarios and potential business models. To help the computing community geta better understanding of where we are and how to leverage edge computing in their own fields, we think it is important to develop special section presenting the state of the art of edge computing. We have witnessed a wide range of progress, particularly in the past five years, spanning the following topics: systems and tools, which provide the basis for the edge computing; innovative edge networks; edge computing applications in multiple domains, such as smart cities, public safety, and autonomous driving, industry IoT, and so on; and new security and privacy threats relating to edge computing, which witnesses the rapid development of relevant core technologies, including virtualization and migration, software-defined networking(SDN), computing offloading, programming models and operating systems, security, privacy, systems and tools, and fast penetration of edge computing in several applications scenarios, including video analytics for public safety, autonomous driving, deep learning, wireless communication, and edge intelligence. Although edge computing brings the computation closer to delay sensitive services, challenges that restrict the cloud model still remain as the pace of generated data continues to rise. Edge nodes can be mobile, and the rapid changes can occur anytime in dynamic networks (e.g., connectivity failure and bandwidth fluctuation); therefore, the orchestration of edge services becomes more challenging. Specifically, service discovery, resource coordination, coping with resource heterogeneity, lifecycle management, and task offloading are open research challenges. Moreover, scheduling data management and processing tasks to derive analytics insights requires "intelligent" consideration. The following six articles present state-of-the-art studies that address these challenges.

1) "A survey of virtual machine (VM) management in edge computing" provides an overview on the industrial and research projects on Management in edge computing. The paper focuses on the virtualization frameworks and virtualization techniques, server less management, and security advantages and issues that virtualization brings to edge computing.
2) "Software-defined networking (SDN) enhanced edge computing: A network-centric survey "discusses how SDN and related technologies are integrated to facilitate the management and the operations of edge servers and various IoT devices. The authors discuss the current status and present new perspectives on this topic.
3) "Dependable resource coordination on the edge at runtime "introduces a methodology and technical framework for engineering resource coordination for the edge-enabled IoT. The authors use bounded model checking as the key technique to compute coordination plans, which satisfy device, edge, and system goals.
4) "A survey on edge computing systems and tools" reviews existing systems and open-source projects for edge computing by categorizing them according to their design demands and innovations. In addition, topics that

are related to energy efficiency and deep learning optimization of edge computing systems are discussed.

5) "Ecosystem of things: Hardware, software, and architecture "surveys the state of the art in supporting smart edge computing and makes some concluding observations with respect to hardware, system software, and ecosystem architecture. Within this survey, the paper deals with the following research challenges: how much raw computing capability and energy efficiency the hardware of things provides, what abstractions the system software supports to utilize the hardware capabilities, and what ecosystem architectures are proposed to harmonize innovation and fragmentation.
6) "Computation offloading toward edge computing" surveys recent research efforts made on exploring computation offloading toward edge computing. The authors highlight the challenges of computation offloading with respect to task partitioning, allocation, and execution over the new architecture of edge computing and investigate disruptive application scenarios, such as real-time video analytics, autonomous driving, smart home, and cloud gaming. Along with the benefits that edge computing brings, there are numerous challenges we should take into account, especially with respect to data security and privacy. The limited processing power of connected devices can restrict the use of security measures making them especially vulnerable to both cyber and physical attacks, while mobility and the rapid provisioning of edge nodes require efficient mechanisms for establishing and attesting trust in the edge. The next two articles deal with security and privacy issues in edge computing:

1) "Edge computing security: State of the art and challenges "reviews the most influential and basic attacks as well as the corresponding defense mechanisms that can be practically applied to edge computing systems. Moreover, the article outlines the challenges and future directions toward securing edge computing systems.
2) "Privacy techniques for edge computing systems" discusses approaches for privacy preserving data aggregation at the edge. The article focuses on techniques by which the edge can provide services to users while assuring user privacy as well as privacy-preserving crowd-sourcing techniques .Edge computing brings together IoT, big data, and mobile computing into an integrated and ubiquitous computing platform. The capability offered to deliver on-demand computing power at the edge and the ability to process the vast amount of data coming from an abundance of devices/sensors provide a huge impetus to artificial intelligence (AI) technologies.

The following five articles focus on these topics.

1) "Deep learning with edge computing: A review" provides an overview of applications where deep learning is used at the network edge. Computer vision, natural language processing, network functions, and virtual and augmented reality are discussed as example application drivers. The authors discuss different architectures and methods to speed up deep learning inference and training deep learning models on edge devices, with an emphasis on distributed training across devices and privacy.
2) "Edge video analytics for public safety: A review" provides survey of applications, algorithms, and platforms that have been proposed to facilitate edge video analytics for public safety.
3) "Edge computing for autonomous driving: Opportunities and challenges" surveys the designs of autonomous driving edge computing systems. In addition, this article presents the security issues in autonomous driving as well as how edge computing designs can address these issues.
4) "Wireless edge computing with latency and reliability guarantees "discusses the feasibility and potential of providing edge computing services with latency and reliability guarantees. The article proceeds by presenting selected use cases that reflect the interplay between edge computing and ultra-reliable low-latency communication (URLLC).
5) "Edge intelligence: Paving the last mile of AI with edge computing" conducts comprehensive survey of the research efforts on edge intelligence. This article provides an overview of the architectures, frameworks, and emerging key technology for a deep learning model toward training and inference at the network edge.

III. Conclusion

Traditionally, the cloud is considered to be a highly centralized resource, much like an IBM mainframe. But the advent of edge computing has made the cloud truly distributed. Developers don't need to work around the limitations of latency and the performance trade-off involved in deploying applications in remote data centers. Edge computing is often used either in the context of delivering static content through CDN or moving the processing closer to IoT devices. But AWS has redefined edge computing in more than one way. It has built a continuum of compute by delivering a new form of edge computing. We hope that through this special section, we have delivered a state-of-the-art glimpse of current edge computing topics, bringing novel problems that must be investigated to the attention of the community. We also hope that it will serve as a valuable reference for researchers and practitioners working in the edge computing domain and its emerging applications. Furthermore, we envision this special section will help to establish a pathway toward a smart edge computing continuum.

References

[1] M. Zwolenski and L. Weatherill, "The digital universe: Rich data and the increasing value of the Internet of Things," Austral. J. Telecommun. Digit. Econ., vol. 2, no. 3, p. 47, 2014.
[2] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet Things J., vol. 3, no. 5, pp. 637–646, Oct. 2016.
[3] A. Vakali and G. Pallis, "Content delivery networks: Status and trends," IEEE Internet Comput., vol. 7, no. 6, pp. 68–74, Nov. 2003, doi: 10.1109/MIC.2003.1250586.

[4] J. Ravi, W. Shi, and C.-Z. Xu, "Personalized email management at network edges," IEEE Internet Comput., vol. 9, no. 2, pp. 54–60, Mar. 2005.
[5] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," IEEE Pervasive Comput., vol. 8. no. 4, pp. 14–23, Oct./Dec. 2009.
[6] M. Symeonides, D. Trihinas, Z. Georgiou, G. Pallis, and M. Dikaiakos, "Query-driven descriptive analytics for IoT and edge computing," in Proc. IEEE Int. Conf. Cloud Eng. (IC2E), Jun. 2019, pp. 1–11.
[7] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," ETSI, Sophia Antipolis, France, ETSI, White Paper 11, 2015, pp. 1–16, vol. 11, no. 11.
[8] M. Finnegan, "Boeing 787s to create half a terabyte of data per flight, says Virgin Atlantic," Computerworld UK, Mar. 6, 2013.
[9] R. LaMothe, "Edge computing," Pacific Northwest Nat. Lab., Richland, WA, USA, Tech. Rep., 2013. Accessed: Mar. 2014. [Online]. Available: https://mafiadoc.com/edge-computing-pacificnorthwest-national-laboratory_59d648481723dd08e35b7b77.html
[10] NSF/Intel Partnership on ICN in Wireless Edge Networks, Nat. Sci. Found., Alexandria, VA, USA, 2016.
[11] M. Chiang andW. Shi, "NSF workshop report on grand challenges in edge computing," Tech. Rep., Oct. 2016.[Online].Available:http://iot.eng.wayne.edu/edge/NSF%20Edge%
[12] Here's How Much Energy All US Data Centers Consume, Data Center Knowl., San Francisco, CA, USA, 2016.
[13] P. Voigt and A. von dem Bussche, "The EU general data protection regulation (GDPR)," in A Practical Guide, 1st ed. Cham, Switzerland: Springer, 2017.

