Edge Computing: Classification, Applications, and Challenges

Gagandeep Kaur
School of Computer Science & Engineering
Lovely Professional University
Phagwara, India
gagandeep.23625@lpu.co.in

Ranbir Singh Batth
School of Computer Science & Engineering
Lovely Professional University
Phagwara, India
ranbir.21123@lpu.co.in

2021 2nd International Conference on Intelligent Engineering and Management (ICIEM) | 978-1-6654-1450-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICIEM51511.2021.9445331

***Abstract*— Edge computing is a relatively recent phenomenon in the computing world, which takes cloud computing services closer to the end user and is distinguished by fast processing and application response time, which leads to many advantages such as faster and more efficient data processing; Safety; reduced leakage on existing networks. Moreover, delay-sensitive applications may benefit from the Edge computing paradigm's low latency, agility, and location awareness. Significant research has been conducted in the field of Edge computing, which is reviewed in terms of recent technologies such as Mobile Edge Computing, Cloudlet, and Fog computing, allowing researchers to gain a better understanding of current and potential solutions. This article summarizes the edge computing paradigm classification, applications, and various challenges in detail.**

Keywords—Edge computing; Cloud computing, Mobile Edge Computing, Fog Computing, Mobility, Security.

I. INTRODUCTION

With the escalation of the Internet of Things (IoT), Cloud computing, which is a collection of networks, has tremendously changed the living and working style. Traditional cloud computing as centralized is accosting severe major challenges, such as high latency in real-time applications, low spectral efficiency (SE), and a non-adaptive machine type of communication [1]. Propelled to explain all these difficulties, new technology is driving a pattern that moves the capacity of incorporated distributed computing to the edge devices of the organization or network. Edge computing is also recognized as edge processing. It is a type of network communication technique to distribute the load on the system by placing a huge number of servers near users and devices, In simple words "putting servers at edge network closer to the device". In other words, an alternative to storing and processing all the data on the cloud, a fragment of it is handled by the administering platform near the terminal. This makes it possible to extremely digest traffic on the Internet and eliminates communication delays. There are various types of network techniques, and in addition to edge computing, there are "cloud computing" and "fog computing "and all of these computing techniques processes the data differently, so let us see how it differs from edge computing [1]. Cloud computing is named a "cloud" because it remotely processes a computer in a location away from the user via the Internet. As shown in fig. 1, In cloud computing, services such as servers, storage, databases, and applications are all located in the cloud, so you can access them on any device if you have the Internet [2]. Besides, there is no need to prepare a server or use a data center, which can significantly reduce costs. It is possible to store a huge amount of information in the cloud, but on the other hand, there is a concern that data processing will take time. Cloudlet addresses the issue of end-to-end responsiveness between mobile devices and cloud which was a challenge in the cloud. Fog computing "fog" means "fog" and was named because it is "closer to the device than the cloud". In edge computing, processing power is at the edge of the network, cloud computing has processing power on the Internet, while fog computing has processing power in the LAN. Therefore, scale of data handled is larger than that of edge computing, which is good at personal data processing, and it can perform functions like mini-cloud computing that can optimize the amount of data. Mobile Edge computing placed the edge nodes on the network and enhanced to include the fixed network as shown in fig 2.

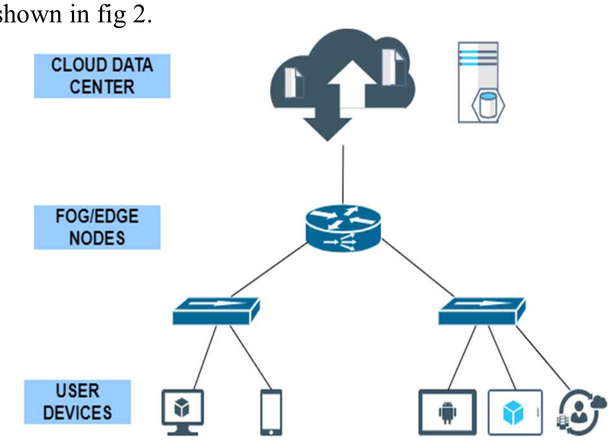

Fig. 1: Edge Computing Paradigm

The IoT, defined as a distributed network of heterogeneous devices (sensors, actuators, and processors) that exchange messages between them without human intervention, has attracted particular interest from researchers. With the increase in the number of services, other needs also arise. Some of them, and perhaps the most critical in the architecture intended for IoT, are the latency of data and efficiency in how resources are used, and data is manipulated.

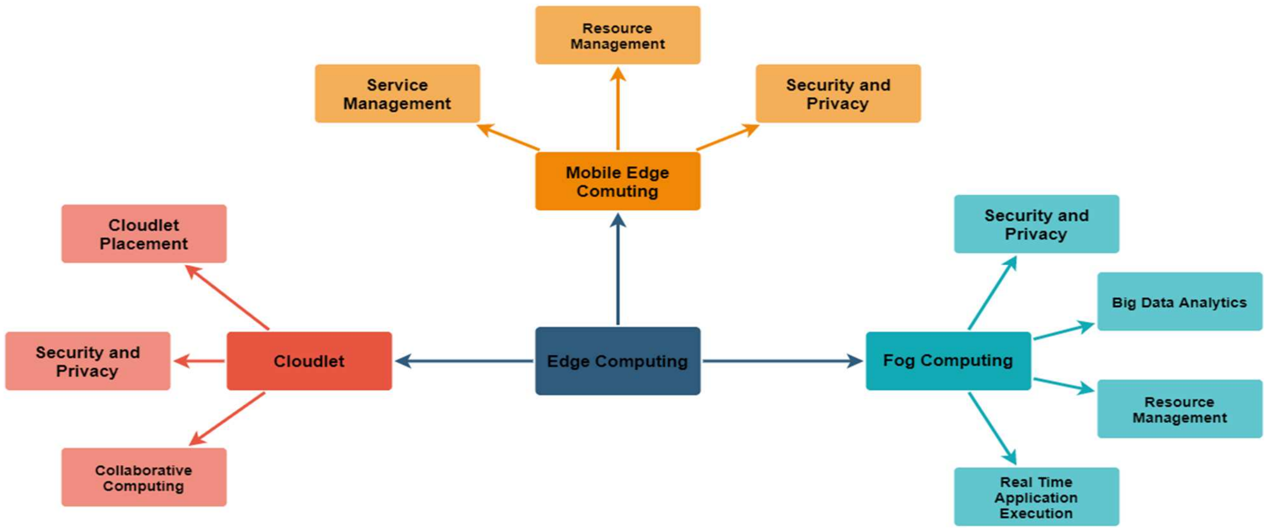

Fig.2: Classification of Edge Computing

Tab.1: Three Edge Computing Technologies

	Cloudlet Computing	Mobile Edge Computing	Fog Computing
Architecture	Single layer	Single layer	Single or multilayer
Latency	Medium	Medium	Low
Mobility	Medium	Medium	High
Context Awareness	Low	High	Medium
Management	Centralized	Hierarchical	Hierarchical
Usage	Resource Intensive applications	5G requirements in telecommunication network	IoT

II. Related Work

In this research article a review has been provided on the overview of edge computing based on various issues and challenges of edge computing for diversified applications. This main work aims to summarize edge computing and to make aware of its challenges with modern technologies. **Weisong Shi et al. (2016)** the authors explained the definition of edge computing which addresses the concern of response time where concerns are latency, resource like battery-life constraint, bandwidth cost-saving, as well as data safety and privacy [2]. **Fang Liu et al. (2019)** summarized the existing edge computing systems and related tools. The authors divided the paper into two parts: System View and Application View. In system View, Open Source Edge computing projects and edge computing systems & tools are discussed wherein application view, deep learning optimization at the edge are discussed. [3]. **Yuan Ai et al. (2017)** presented the major three edge computing technologies: mobile edge computing, cloudlets, and fog computing. The authors explained application areas, architectures, standardization efforts for mobile edge computing, cloudlets, and fog computing. [4]. **S. Bhattacharyya et al. (2017)** the authors described the Edge computing that processes the gathered data from end devices at the edge of the network. By covering a large range of technologies, edge computing addresses the various concern as battery life constraint, bandwidth usage, latency, data security, and data privacy. The need for edge computing (Push from Cloud Services and Pull from the Internet of Things) is discussed [6]. **Salman Taherizadeh et al. (2017)** the authors explained the auto-scaling applications in edge computing which maintained the online services at a decentralized location. They broadly explained two aspects of this paper. In the first section, they major focused on the different types of edge computing applications (IoT Applications, Micro-service applications, Time-critical applications). For these applications, auto-scaling challenges when the workload dynamically changes are discussed. In the second section of this paper, container-based visualization auto-scaling technologies are discussed. Self-adaptive application at runtime enhances the performance. In the classification of auto-scaling applications in edge computing, the authors explained the cloud framework, Virtualization Technology, Monitoring approach, Operational behavior, Adjustment ability, Architectural support, Image Delivery, and scalability techniques [7]. **JunjieXie et al. (2020)** addressed the issue of low latency requirements in mobile edge computing. The author proposed a fast data-sharing framework HDS (Hybrid data sharing) to meet the requirements of low latency by dividing the gathered data location service into two regions: Intra-region and inter-region. With the Hybrid information sharing system which comprises 100 areas, the creator accomplished low latency, low usage overhead, and 50.21% more limited query ways, and 92.75% fewer false positives. In this whole network, the total edge server used was 1000 to 10000 [14]. **Batyr Charyyev et al. (2020)** the authors performed the cloud edge latency comparison. They performed an extensive measurement to assess the latency characteristics of end-users to the edge servers and cloud data centers. It estimated latency from 8,456 end-clients to 6,341 Akamai edge workers and 69 cloud areas. At last, the paper's outcome is that while 58% of end-clients can arrive at a close by edge worker in under 10 ms, just 29% of end-clients get a comparable dormancy from a close by cloud area. [18]. **N.Saranya et al. (2020)** the authors reviewed the various data latency techniques in Mobile Edge Computing. As in the centralized cloud, one cannot achieve low latency. Mobile Edge Computing makes efficient use of the resources and decreased the movement of large data generated by the edge devices. Edge computing is important for solving fatal situations such as Conflicts in Autonomous vehicles, Fire, Environmental Hazards. [20].

Tab.2: Comparison of Techniques on various Parameters

Ref.	**Scalability**	**Latency**	**Resource Constraint**	**Security and Privacy**
[8]	✓	×	×	✓
[13]	✓	✓	×	×
[14]	×	✓	×	×
[15]	✓	✓	×	×
[17]	×	×	✓	×

(✓:Observed, ×: Not Observed)

Tab.3: Edge computing system and tools based on their Properties

Edge Computing System	**End Device**	**Functionality**	**Architecture**
Cloudlet	Mobile	VM Migration	Hybrid 3-tier
CloudPath	Mobile	Path Computing	Hybrid Multi-tier
AirBox	IoT	Security	Hybrid 3-tier
Pcloud	Mobile	Dynamic Allocation	Hybrid 3-tier

III. EDGE MULTI-LAYER ARCHITECTURE

The general architecture of edge computing is having terminal, edge, and cloud layers which is 3-tier architecture. In this, the terminal layer deals with the end devices and nodes which include both tasks as data consumer and data provider. In the terminal layer, all the devices like smartphones, sensors, wearable devices gathered all the data and provide it to the upper layer for computations. The second layer which is the edge layer plays an important role as processing unit of system. This layer contains routers, switches, gateway, access points which store and compute all the data provided by the end devices. High-performance servers are placed in the third layer, the cloud layer which is widely used in global applicable situations and organizations that require large-scale centralized processing. Edge computing industry alliance (ECC) released framework 3.0. Multi-layer functionality as reference architecture is shown in fig 3.

IV. APPLICATION AREAS OF EDGE COMPUTING

It is proposed to bring some of the cloud services closer to the level at which the data are found as well as using multiple levels, according to the needs of the different applications. Due to multiple benefits of edge computing like speed, security, scalability, versatility, and reliability edge computing is used in many real time applications. Basically, edge computing is used to push applications and data to logical extremes by centralized hubs in a network. We use edge computing in real time applications because better responsiveness and robustness make this system more effective. Edge computing is distributed kind of system where large data volume is processed and compiled. Distribution properties gives features like supply chain tracking, point of sale system and distributed artificial intelligence processing.

A. Edge computers and smart home devices

Increasingly common appliances, such as thermostats and smart speakers, are using edge computing technology to respond quickly to changing conditions and controls. Smart city and personal houses, and augmented reality can also benefit from fast and localized data processing. The same principle works for any activated device, with the processing center either directly on the device or in a near location. Hypothetically, a system can mutually use the cloud, for less time-sensitive processing and the edge for more urgent requirements. Edge computing also needs dedicated hardware for edge nodes and these edge nodes ultimately provides services.

B. Edge computing in business

An example of a business is a retail store with a collection of activated security cameras. Utilizing the cloud-based technology, the camera sends data to a server, which is thousands of KMs away that identifies the movement and saves those clips to the store data. By utilizing edge computing, the processing unit itself is a part of the camera, categorizing the clips that should be saved and which should not, and transferring only the relevant clip to the server. This procedure reduces the load on the store's computing power, allowing for more efficient operation, and protecting other company data from possible security attacks. Edge computing can be used in the variability of industries beyond retail, including manufacturing, energy, utilities, transportation, defense, healthcare, and the media [3].

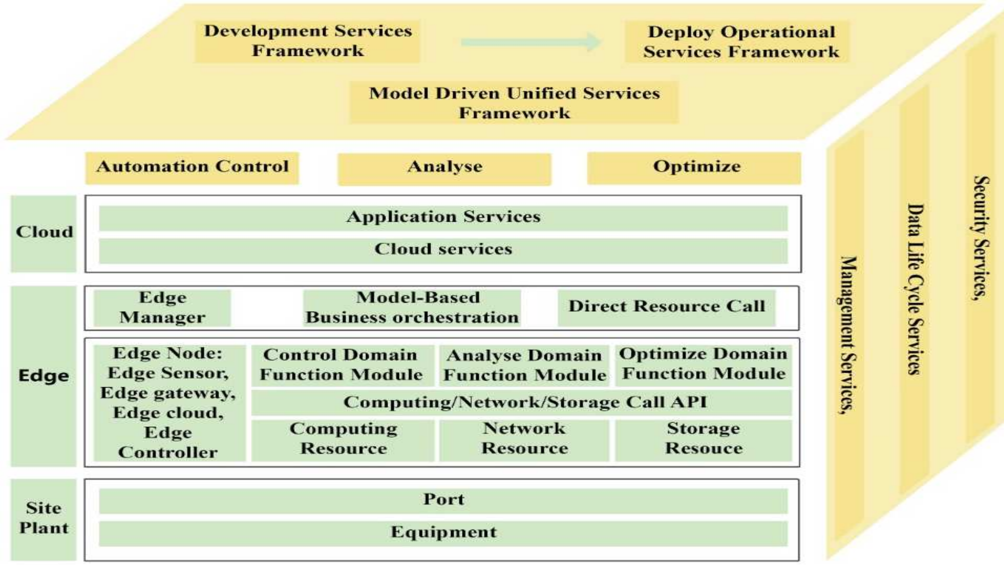

Fig. 3: Edge computing reference architecture 3.0

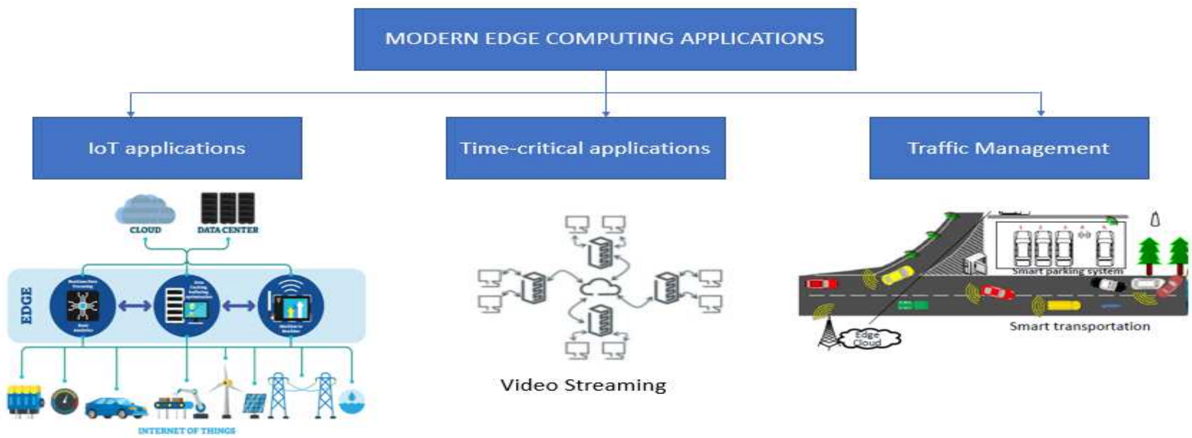

Fig. 4: Application Areas of Edge Computing

C. Edge computing for 5G technology

Edge computing takes a big role to play in 5G wireless networks**,** which will be able to host several times the number of connected devices that use them at a period.5G network can be able to transmit a large amount of data, hundreds of times faster than 4G, will incorporate the number of cell stations several times, and reduce latency. Telecommunications providers using 5G will use local data centers on or near 5G towers, giving the network an extra boost.

D. Edge computing for smart technology

Experts consider edge computing to be the next great thing in technology, which will transform our living way. Several types of equipment fall under the edge computing umbrella, including 5G cell stations, data center infrastructure, and applications to analyze data faster. The number of sensors associated with cities, homes, and smart vehicles are expected to increase exponentially in the coming years. 5G wireless networks, supported by peak computing, will be able to manage all data generated by these devices in ways that 4G cannot.

V. EDGE COMPUTING IN REAL-LIFE

With the progress of IoT year by year, the merits of edge computing are attracting a great deal of attention and are already being used in many scenes.

A. Face Recognition System

The face recognition system is a system that automatically recognizes a person from the digital image of the camera and has been introduced as a security enhancement such as immigration control, company entry/exit, and PC logon. It may also be used to enhance corporate services. Face recognition systems require high real-time performance, but the use of edge computing has made it possible to perform high-speed authentication.

B. Wearable Device

A wearable device can be worn on the part of the body, such as the head, wrists, or ankles. Since it can be worn like glasses or a wristwatch, it is used in fields such as health management, medical fields, and industry. Besides, smart watches can be used in conjunction with smartphones act as wearable devices and are becoming popular among consumers as edge computing that utilizes mobile devices.

C. Human Behaviour Analysis Service

As the face recognition system, the wearable device is a mechanism that performs image processing of face recognition at the edge part and analyses it in cooperation with the data on the cloud. Wearable devices can predict and track the movement of a person extracted from camera images; the analysis data of the tracking is processed in the temporary memory of the computer. The created coordinate data is aggregated on the cloud for each camera, and besides this becomes a flow line of people, and it is used for the analysis of purchasing behavior at retail stores.

D. Edge Computing as the savior of the IoT Era

In modern times, vast amounts of data are stored in cloud and processed on the cloud, such as the spread of IoT, digitization, diversification of data, and quality improvement. However, since the amount of data is increasing year by year, the processing on the cloud cannot be completed in time, and scalability problems (delay in data processing) will occur. Meanwhile, edge computing, which distributes communication traffic, has come to be used, and needs are increasing significantly in scenes where real-time data processing is required. Therefore, edge computers can be said to be the savior of the IoT era and will increase the value of big data [5].

VI. CHALLENGES IN EDGE COMPUTING

Edge computing is still in its infancy. The major challenges in Edge Computing have been classified as:

A. Bandwidth

The more data at the network bandwidth-edge, the more the bandwidth will change. It should balance the greater bandwidth across the network. From the latency's point of view, the high bandwidth can reduce the network transmission time, especially for the huge amount of data. In cloud higher bandwidth provided at data centers and lower at edge but this will be challenge at edge network to provide higher bandwidth to end devices. If the workload can be handled at the edge, the delay can be improved, and bandwidth will also be saved [6].

B. Multi-domain Management and Cooperation

Multiple types of resources are involved in Edge computing, where all resources belong to various owners. How to access all the required resources for application and allocate them according to the need of the user's application, services. In an emergency, a problem is that the mobile edge computing system needs to consider. In the initial phase, systems considered the functionalities rather than implementation issues like the price model. With the upgrade of edge processing, the genuine formation concern ought to be illuminated before we appreciate all the advantages.

C. Security and Privacy

Security plays important role in every field as every user is concerned about privacy for their confidential information. While working in a network, various attacks can be performed on the edge network. The following is different types of security threats:

- *Phishing* - An attempt to extract sensitive or confidential information from the user of cloud-connected IoT equipment. Cybercriminals use this method to distribute electronic content to a wide range of victims, prompting them to take specific actions, such as clicking a link or sending a reply to an email. If they do so, they will provide personal information such as username, password, bank or financial details, or credit card details without their knowledge**.** Phishing distribution methods include email, online advertising, and SMS. A phishing type is *spear phishing* - an attack aimed at specific people or groups of people, e.g., corporate management boards, celebrities, or high-ranking government officials.
- *Malware* - Malware that is specially designed to harm an edge device, system, or data. It includes several types of tools, including adware, ransom ware, Trojans, viruses, and worms.
- *Rootkit* - Malware that enables secretly accessing parts of an edge device, software, or system. It often modifies the hardware operating system in such a way that it remains hidden from the user.
- *Trojan* - A form of non-replicating malware that contains hidden functionality. The Trojan usually does not try to propagate or inject itself into other files.
- *Backdoor* - Malware targeting remote entry into a system, device, edge device, or software, or bypassing traditional security measures installed in them.
- *Spyware* - Malware that spies on an edge device user, intercept keystrokes, emails, documents, and even turns on a video camera without their knowledge.

D. Uncompromising Quality of Service (Latency)

Locating the compute at the edge is one of the parameters that provide QoS. In edge computing, the focus is performing the computing anywhere close to the data collected by the computer. To check the network performance, usually Latency is vital QoS metrics which is not only determined by computation time. The computation of the workload should be performed at closer layer. Also, this layer must have enough computation capability to do the calculations at the edge network.

E. Partition and Offloading tasks (resource allocation)

Workload allocation is not an easy task. Distributed computing should consider location as an additional aspect of computation. Resource mobility defines how to dynamically discover the resources in the network and manage all the required available resources, including long-term resources and short-term resources. Also, in some situations, few edge devices or end devices are damaged, then at that situation challenge is how the system can be resumed as soon as possible with the best replacements [6]. At the edge, the resources are resource-constrained as having smaller processors and a limited power budget, also resource heterogeneous where the processors are with different architectures (Data Flow, Control and Tenancy) and workload need to change dynamically [29].

VII. CONCLUSION

In this paper, we came up with the introduction of edge computing, Applications, Architecture, and benefits of edge computing over centralized computing. Most of the applications require prompt response, so edge computing process the data at the edge of the network that eliminates the roundtrip travel time provides a real-time response and native authority. At last, we put forward the edge computing challenges. In the future, with the persistent advancement of the Internet and human culture, edge processing will assume a more significant job in various applications like smart homes, smart widgets, intelligent transport, quick audio, and video streaming.

REFERENCES

[1] P. Singh, A. Kaur, G. S. Aujla, R. S. Batth and S. Kanhere, 2020 "DaaS: Dew Computing as a Service for Intelligent Intrusion Detection in Edge-of-Things Ecosystem," in *IEEE Internet of Things Journal*, https://doi.org/10.1109/JIOT.2020.3029248

[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, 2016.

[3] F. Liu, G. Tang, Y. Li, Z. Cai, X. Zhang, and T. Zhou, “A Survey on Edge Computing Systems and Tools,” *Proc. IEEE*, pp. 1–24, 2019.

[4] Y. Ai, M. Peng, and K. Zhang, “Edge computing technologies for Internet of Things: a primer,” *Digit. Commun.Networks*, vol. 4, no. 2, pp. 77–86, 2018.

[5] B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, and D. S. Nikolopoulos, “Challenges and Opportunities in Edge Computing,” *Proc. - 2016 IEEE Int. Conf. Smart Cloud, Smart Cloud 2016*, pp. 20–26, 2016.

[6] S. Bhattacharyya, “Research on Edge Computing : A Detailed Study,” *Academia.Edu*, vol. 2, no. 6, pp. 9–13, 2016.

[7] S. Taherizadeh and V. Stankovski, “Auto-scaling applications in edge computing: Taxonomy and challenges,” *ACM Int. Conf. Proceeding Ser.*, pp. 158–163, 2017.

[8] C. Feature and E. Deep, “Private and Scalable Personal Data Analytics Cloud Deep Learning,” 2018.

[9] T. G. Rodrigues, K. Suto, H. Nishiyama, N. Kato, and K. Temma, “Cloudlets Activation Scheme for Scalable Mobile Edge Computing with Transmission Power Control and Virtual Machine Migration,” *IEEE Trans. Comput.*, vol. 67, no. 9, pp. 1287–1300, 2018.

[10] R. Bruschi, F. Davoli, P. Lago, and J. F. Pajo, “A Multi-Clustering Approach to Scale Distributed Tenant Networks for Mobile Edge Computing,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 499–514, 2019.

[11] S. Maheshwari, D. Raychaudhuri, I. Seskar, and F. Bronzino, “Scalability and performance evaluation of edge cloud systems for latency constrained applications,” *Proc. - 2018 3rd ACM/IEEE Symp.Edge Comput.SEC 2018*, pp. 286–299, 2018.

[12] S. H. Mortazavi, M. Salehe, C. S. Gomes, C. Phillips, and E. De Lara, “CloudPath: A multi-Tier cloud computing framework,” *2017 2nd ACM/IEEE Symp. Edge Comput.SEC 2017*, 2017.

[13] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. P. C. Rodrigues, and M. Guizani, “Edge Computing in the Industrial Internet of Things Environment: Software-Defined-Networks-Based Edge-Cloud Interplay,” *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 44–51, 2018.

[14] J. Xie, D. Guo, X. Shi, H. Cai, C. Qian, and H. Chen, “A Fast Hybrid Data Sharing Framework for Hierarchical Mobile Edge Computing,” *Proc. - IEEE INFOCOM*, vol. 2020-July, pp. 2609–2618, 2020.

[15] W. Xia, J. Zhang, T. Q. S. Quek, S. Jin, and H. Zhu, “Mobile Edge Cloud-Based Industrial Internet of Things: Improving Edge Intelligence with Hierarchical SDN Controllers,” *IEEE Veh. Technol. Mag.*, vol. 15, no. 1, pp. 36–45, 2020.

[16] M. Uddin, S. Mukherjee, H. Chang, and T. V. Lakshman, “SDN-Based multi-protocol edge switching for IoT service automation,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 12, pp. 2775–2786, 2018.

[17] T. P. Raptis, A. Passarella, and M. Conti, “Industrial Edge Networks,” vol. 38, no. 5, pp. 915–927, 2020.

[18] B. Charyyev and M. Gunes, “Latency Characteristics of Edge and Cloud,” no. June 2020.

[19] A. Singh, G. S. Aujla, and R. S. Bali, “Intent-Based Network for Data Dissemination in Software-Defined Vehicular Edge Computing,” *IEEE Trans. Intell. Transp. Syst.*, pp. 1–9, 2020.

[20] Z. Ning *et al.*, “Intelligent Edge Computing in Internet of Vehicles: A Joint Computation Offloading and Caching Solution,” *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2020.

[21] X. Ren, G. S. Aujla, A. Jindal, R. S. Batth and P. Zhang, "Adaptive Recovery Mechanism for SDN Controllers in Edge-Cloud supported FinTech Applications," in *IEEE Internet of Things Journal (2021).* https://doi.org/10.1109/JIOT.2021.3064468

[22] T. G. Rodrigues, K. Suto, H. Nishiyama, N. Kato, and K. Temma, “Cloudlets Activation Scheme for Scalable Mobile Edge Computing with Transmission Power Control and Virtual Machine Migration,” *IEEE Trans. Comput.*, vol. 67, no. 9, pp. 1287–1300, 2018.

[23] S. Kumar and R. H. Goudar, “Cloud Computing – Research Issues, Challenges, Architecture, Platforms, and Applications: A Survey,” *Int. J. Futur. Comput. Commun.* vol. 1, no. 4, pp. 356–360, 2012.

[24] Kumar VDA, Kumar A, Batth RS, “Efficient data transfer in edge envisioned environment using artificial intelligence based edge node algorithm”, *Transactions on Emerging Telecommunications and Technologies.,2020;e4110*. https://doi.org/10.1002/ett.4110

[25] Q. Zheng and Z. Ping, “Simulation study on latency-aware network in edge computing,” *2019 6th Int. Conf. Control. Decis. Inf. Technol. CoDIT 2019*, pp. 150–155, 2019.

[26] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, “Mobile Edge Computing A key technology towards 5G,” no. 11, 2015.

[27] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. De Foy, and Y. Zhang, “Mobile edge cloud system: Architectures, challenges, and approaches,” *IEEE Syst. J.*, vol. 12, no. 3, pp. 2495–2508, 2018.

[28] Wang, C., Batth, R.S., Zhang, P., Aujla, G.S., “VNE solution for network differentiated QoS and security requirements: from the perspective of deep reinforcement learning”. *Computing (2021).* https://doi.org/10.1007/s00607-020-00883-w

[29] M. C. Computing, X. Chen, L. Jiao, and W. Li, “Efficient Multi-User Computation Offloading for,” *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, 2016.

[30] C.-H. Hong and B. Varghese, “Resource Management in Fog/Edge Computing,” *ACM Comput. Surv.*, vol. 52, no. 5, pp. 1–37, 2019