Optimization in Edge Computing: A Survey

Raghad Jassim Mohammed*, Shaymaa W. Al-Shammari **

* Department of Computer Engineering, College of Engineering, AL-Nahrain University, Baghdad, Iraq
Email: st.raghad.j.mohammed@ced.nahrainuniv.edu.iq
https://orcid.org/0009-0000-1873-2742

** Department of Computer Engineering, College of Engineering, AL-Nahrain University, Baghdad, Iraq
Email: shaymma.alshammari@gmail.com
https://orcid.org/0000-0003-1720-5486

Abstract

Due to advancement, there are now more smart devices connected to the internet., which causes massive data traffic in the network. Resulting in many problems such as slow response time, largely consumed energy, high load in transmission channels, and bad use of the network resources in the traditional cloud. Edge computing facilities and brings the cloud's service to the network's edge. Edge computing a distributed computing near to source of data. This technology solved and helps to solve many problems in the cloud, but also has challenges and open issues, for example, the limited lifetime of the IoT devices, limited resources, and computation offloading. The performance of edge computing is affected by offloading. So, many optimization methods are used to solve this problem and improve the performance in edge computing. This paper presents a survey of the studies related to optimizing task offloading in edge computing. The difference between this work and the previous surveys is that this survey combined offloading and optimization with the types of optimization methods and takes into consideration the three layers of edge computing architecture IoT, the Edge layer, and the cloud layer. The previous surveys did not include all types of optimization or combine the offloading with optimization. The architecture of edge computing, challenges, and open issues, optimization methods are presented.

Keywords- Optimization, Edge Computing, IoT and Task Offloading.

I. INTRODUCTION

The amount of data produced by mobile devices has expanded in recent years because of the rapid advancement of technology like the IoT and cloud computing. Massive amounts of data must be uploaded to the cloud server in the classic cloud computing approach [1]. However, issues like reaction time, network interference, and energy consumption would invariably arise during the transmission process. With its robust computational capacity, cloud computing can address issues with mobile device battery consumption and a host of other issues that cannot be addressed by mobile devices alone. Users are aiming for a more seamless experience [2]. The main problem with cloud computing is that all of the data must be sent to centralized computer clusters for computations to be done. This concerns latency on the network. Edge computing, processing, and storage resources are generally situated close to the network edge, close to the data-generating devices. Fog computing is therefore more suited for latency-aware applications like augmented reality, smart cities, and healthcare [3].

Mobile applications have significantly improved people's daily lives in recent years. Individuals use a variety of portable services offered by mobile applications to interact socially, do business, perform government affairs, and have fun [4]. Moreover, mobile devices (such as portable laptops, cell phones, and other application devices are getting more intelligent and potent, which increases the computational power available for mobile applications [5]. The size, energy, weight, and heat dissipation limitations of portable devices continue to place restrictions on the computing power of these devices [6]. Mobile applications have been progressively getting more energy consumed due to the growing need for complex functionalities of portable services. It is effective to transfer the computationally demanding functions of mobile applications from mobile devices to the cloud. The utilization of clouds is a practical solution to this issue, allowing mobile users to access high-performance applications while also significantly extending the battery life of their devices. Nevertheless, as mobile users are geographically far from the cloud, the growing mobile data traffic places significant stress on the backbone network [7]. The network's transmission delays have evolved into a new barrier as MDs and mobile applications have become more common. Several powerful mobile applications that are sensitive to response time, such as large-scale image processing, online gaming, and personal assistants, find this intolerable. Edge computing has been suggested as computing paradigm to address this issue. By placing edge servers close to or in mobile base stations, Edge computing can provide efficient quality, low-latency services, and high-bandwidth at the edge of mobile networks [8]. Mobile networs are used by mobile devices to connect to the edge servers and some of their tasks are offloading in a single hop, relieving the backbone network's load and prolonging their battery life. When compared to services offered by a cloud computing platform, MEC platform services offer an unmatched level of experience. As a result, Edge computing is getting more and more attention from both academics and businesses [9]. But also edge computing has many challenges and issues, so many optimization methods are used to solve these issues and improve the performance of edge computing. Many surveys have been shown in this field, however, these surveys missed presenting studies that introduce all network architectures and optimization methods, in addition to the main parameters that affect the offloading process.

Many surveys have been presented on the subject of edge computing optimization and computation offloading. In[10], the authors presented a survey only for six methods of optimization including, machine learning, convex optimization, Lyapunov optimization, game theory, heuristic techniques, and others. For each type, they presented the applications, evaluation methods, types of offloading methods, and objective functions. L. Liu, et al. (2021) [11] Presented a survey on computation offloading that includes different scenarios and focuses on effective variables but does not consider the optimization methods. another survey has been presented in [12] in energy efficient computation offloading, which focused on the energy consumed, however, it did not take into consideration the latency. In [13] a survey on computation offloading in the MEC system and their algorithms is introduced but did not include all metrics for computation offloading. In [14], the researchers surveyed the architecture of edge computing and explained some computation and communication models. But did not take into consideration the optimization methods. In [15], A. Shakarami, et al. (2020) discussed a survey on machine learning and computation offloading in edge computing but does not take into consideration the other computation offloading methods or optimization algorithms. In [6], a survey was presented on MEC architecture, design, and optimization methods the main objective was to determine the requirements to implement the auto- scale MEC but did not consider the metrics that have a direct effect on the performance of edge computing. In [16], a survey was presented on computation offloading techniques in MEC but does not take into consideration the optimization methods with these techniques. In [17] review the computation offloading techniques including application partitioning, task scheduling, distributed execution, and resource management but did not take into consideration the optimization methods. The researchers in [18] presented a survey on the applications of edge computing and discussed their challenges and highlighted the importance of edge computing. But they lacked considering the optimization methods and computation methods. In [19], a survey of all types of optimization methods was presented but only for artificial intelligence (AI) neglecting the other approaches of optimization.

The above aforementioned presented surveys in the literature review reveal that so many searches were presented on this topic, but not all metrics, architecture, and optimization methods were considered. This highlights the importance of presenting a work to provide the reader with complete details on this topic.

The aim of this paper is to provide the researcher in the future with a complete overview of the optimization techniques with all metrics and details. In this work, the survey merges the three layers of architecture with optimization methods and includes computation offloading of the tasks in edge computing.

The remaining parts of this survey are organized as follows. Section II presents the background and related work. Section III introduces the challenges and open issues. Section IV concludes this paper by discussing the main and possible future research directions.

II. A TAXONOMY OF OPTIMIZATION OF TASK OFFLOADING IN EDGE COMPUTING

In this section, the survey factors are presented, these factors that would affect the decision of task offloading. This includes the optimization methods that are used to make the offloading decision based on minimizing the energy consumed and elapsing time in data downloading, data uploading, and data processing in the three layers taking into consideration the three types of optimization, Machine learning, heuristic algorithms, and scheduling algorithms. The last factors included in this survey are the processing location and architecture taking into consideration the IoT layer, The edge layer, and The cloud layer. As shown in Figure 1.

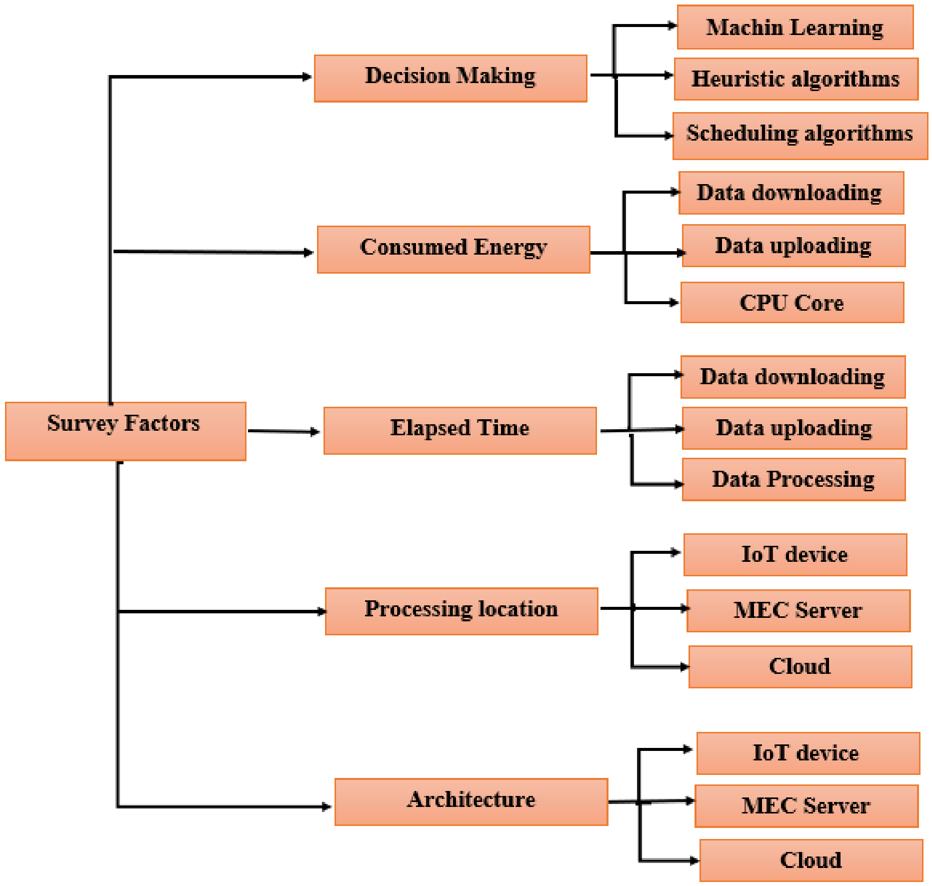

Figure 1 A taxonomy for papers related to the subject of optimization of task offloading in edge computing

- **Decision Making**

This includes classifying the decision making method into three different types which are considered in the previous literature review. Machine learning, Metaheuristic optimization, and scheduling algorithm are as follows:

  - Machine learning: is the process of iteratively optimize the accuracy and reducing the error rate of a machine learning model. Machine learning models learn to generalize, classification and prediction new real-time data based on insights from training data

  - Metaheuristic optimization: a group of stochastic algorithms that combine local search and randomization. They frequently draw inspiration from biological or natural processes. Popular algorithms include ant and bee algorithms, genetic algorithms, particle swarm optimization, and ant algorithms. Often, metaheuristic algorithms are created for global optimization.
  - scheduling algorithm
    The algorithms that presented depend on the aim of it, which are mainly used to allocate tasks to the CPU or to a certain location to improve the performance of the system.
- **Consumed Energy:** represents the consumed energy calculated for each task in terms of processing and transmission.
- **Elapsed Time:** represent the consumed energy calculated for each task in terms of processing and transmission.
- **Processing Location:** represents the location of task processing (IoT device, MEC server, or the cloud)
- **Edge Computing Architecture:**

There is no specific or standard architecture for edge computing, it is different according to the scenario. In this work, the most popular architecture (hierarchical architecture) the cloud layer, the edge layer and the IoT layer are presented and considered in this survey, as shown in Figure 2[20] . The user layer can be identified by the wireless form of communication used by wireless users devices [21]. The fundamental distinction between the edge and cloud layers is the computing power provided servers at the edge and in the cloud, respectively. Sensors, mobile phones, and automobiles are among the gadgets in the user layer[22]. Through a wireless connection, the computation-intensive operations are transmitted from these devices to the distributed edge servers for processing. [23].

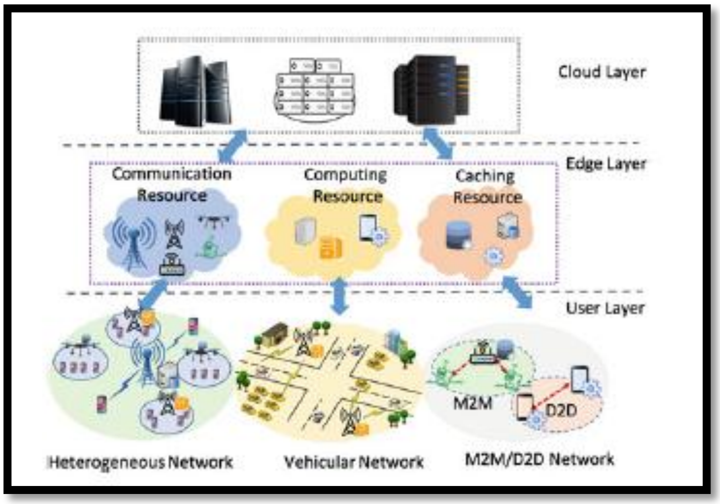

Figure 2 Edge Computing Architecture [10]

The edge tier is in the middle of the layered architecture, represented by edge servers deployed for example in base stations, access points, and mobile phones [24]. , [25].

III. MAPPING LITERATURE SURVEY ON PERFORMANCE OPTIMIZATION IN EDGE COMPUTING

Energy awareness is viewed as the need to offload tasks and allocate resources in edge computing. Several studies have presented the effect and importance of energy and latency in edge computing. Some researchers have presented studies include the IoT devices and MEC. Most of the work is moving the functionality from the cloud server to the edge server.

In [26], The online reinforcement learning problem was presented to minimize the power consumption of mobile devices, application latency was taken as a constraint, and the time life of devices was not taken into consideration. In this paper, the lifetime of devices was taken into consideration. A distributed offloading was proposed in [27] to solve the computational offloading problem in the MEC system. The system cost is reduced as result of using of this solution. The cloud layer was not taken into consideration, but in this work, the cloud layer is considered. In [28] A deep Q-network-based strategic computation offloading algorithm was proposed to solve the computation offloading problem and minimize the cost. But does not take into consideration the power consumed by channel transmission. In this work, the consumed energy of transmission depends on power consumed and time. In [29], transmit power allocation and an energy-efficient computation offloading scheme was proposed to solve the computation offloading problem and minimize the system's cost. An optimization algorithm is used to enhance the performance of the system. In [30], an efficient Joint Caching and Wireless and Backhaul Scheduling (JCWBS) algorithm is used to solve the scheduling problem and reduce the energy consumed by the cloud. However, the authors focused on the cloud layer. This work considers three layers. In [31], a series of fast numerical algorithms are used to minimize response time and consumed power, but did not consider the cloud layer. In [32], the objective was to reduce the energy and time consumption of mobile devices under deadline constraints. But the cloud layer does not consider this. In [33], a Genetic particle swarm optimization algorithm was proposed to improve data placement in the MEC system using cost effective strategy. But does not take into consideration the IoT devices layer. In this work, the IoT devices layer is considered. In [34], an online decision and computational resource management algorithm and Lyapunov optimization technique were used to reduce the energy consumed by the system and load balancing. In [35], a deep reinforcement learning-based joint optimization approach was proposed to solve the decision making problem and reduce the energy and time consumed in the MEC system. But the cloud layer was not taken into consideration. In [36], Genetic algorithm, named M-COGA was proposed to solve the task offloading selection and minimize the energy consumed by the MEC system. L. T. Hsieh et al. (2020) [37] proposed an online Double Deep Q Networks (DDQN) based learning scheme to improve the quality of service, including different parameters in dynamic networks like network delay and task arrivals. But does not take into consideration the energy consumed of the system. The authors in [38] presented an enumeration-Based Optimal Edge Server Placement Algorithm (EOESPA) and Ranking-based Near-optimal Edge Server Placement Algorithm (RNOESPA) to minimize the delay of the system.

In [39], an alternating optimization algorithm was proposed to minimize the energy and time consumed by UAVs. However, the authors did not take into consideration the cloud layer. In [40], The decomposed computation offloading and resource allocation algorithm and the random scheduling algorithm were proposed to minimize the system delay. But the cloud layer and energy consumption of the system are not considered. Z. Liao et al. (2021) [41] introduced a binary-coded genetic algorithm to get an offloading decision to minimize the delay and energy consumed in the system. But the cloud layer was not taken into consideration. In [42], three constrained multi objective evolutionary algorithms (CMOEAs) was presented to solve the problem of offloading decision and reduce the time and energy consumed in the system. Z. Kuang et al (2021) [43] introduced a joint iterative algorithm based on the Lagrangian dual decomposition to minimize the latency, and the monotonic optimization method was proposed to handle the transmission power. Ant grey wolf optimization algorithm, whale optimization algorithm, and Colony Optimization algorithm were proposed in [44] to solve the offloading problem to minimize the delay and energy of the MEC system. But the cloud layer was not taken into consideration.

Table 1 presents a summary of the presented literature review. This table compares the references taking into consideration the architecture of the networks, and the parameters taken into consideration in measuring the cost of the system. This summary helps in shedding the light on this paper’s contribution in comparison to the available studies in this field.

References	________	[26]	[27]	[28]	[29]	[30]
Decision Making	________	Joint Optimization	Distrbuted Offloading Jacobi DOJ	deep Q-network	Lagrange multiplier method,the KKT conditions	JCWBS
Enhancement	________	energy consumption, delay, system cost	System cost (energy, time)	Energy consumed, delay Cost	cost, energy, time	Energy
Consumed Energy	Data uploading	✓	✓	✓	✓	✓
	Data downloading	✓	✓	✓	✓	✓
	CPU core	✓	✓	✓	✓	✓
Elapsed Time	Data Uploading	✓	✓	✓	✓	X
	Data downloading	✓	✓	✓	✓	X
	Data processing	✓	✓	✓	✓	X
Processing	IoT devices	✓	✓	✓	✓	X
	MEC Server	✓	✓	✓	✓	✓
	Cloud	✓	X	✓	X	✓
Architecture	IoT devices	✓	✓	✓	✓	X
	MEC Servers	✓	✓	✓	✓	✓
	Cloud	✓	X	✓	X	✓

References	——	[31]	[32]	[33]	[34]	[35]	[36]
Decision Making	——	Numerical algorithms	Multi steps optimization	GAPSO	Lyapunov optimization technique	deep reinforcement learning-based joint optimization	M-COGA algorithm
Enhancement	——	Cost,time,energy	Energy, time	Cost, time	Energy consumed, delay	Energy consumed, delay Cost	Cost Energy,Time
Consumed Energy	Data uploading	✓	X	X	✓	✓	✓
	Data downloading	✓	X	X	✓	✓	✓
	CPU core	✓	✓	X	✓	✓	✓
Elapsed Time	Data Uploading	✓	X	✓	✓	✓	✓
	Data downloading	✓	X	✓	✓	✓	✓
	Data processing	✓	✓	✓	✓	✓	✓
Processing	IoT devices	✓	✓	X	✓	✓	✓
	MEC Server	✓	✓	✓	✓	✓	✓
	Cloud	No	X	✓	✓	X	✓
Architecture	IoT devices	✓	✓	X	✓	✓	✓
	MEC Servers	✓	✓	✓	✓	✓	✓
	Cloud	No	X	✓	✓	X	✓

ISSN: 2710-216

References	——	[37]	[38]	[39]	[40]	[41]
Decision Making	——	DDQN learning	EOESPA, RNOESPA	Alternative algorithm	Random scheduling algorithm	binary-coded genetic
Enhancement	——	time	System delay cost	Energy consumed , time, cost	delay ,energy, cost	Energy consumed , time, cost
Consumed Energy	Data uploading	X	X	✓	✓	✓
	Data downloading	X	X	✓	✓	✓
	CPU core	X	X	✓	✓	✓
Elapsed Time	Data Uploading	✓	✓	✓	✓	✓
	Data downloading	✓	✓	✓	✓	✓
	Data processing	✓	✓	✓	✓	✓
Processing	IoT devices	✓	✓	✓	✓	✓
	MEC Server	✓	✓	✓	✓	✓
	Cloud	✓	✓	X	X	X
Architecture	IoT devices	✓	✓	✓	✓	✓
	MEC Servers	✓	✓	✓	✓	✓
	Cloud	✓	✓	X	X	✓

ISSN: 2710-216

References	——	[42]	[43]	[44]
Decision Making	——	CMOEAs	A joint iterative algorithm	ACO, WOA, GWO
Enhancement	——	Energy and time	Energyr and latency	Energy, latency, cost
Consumed Energy	Data uploading	✓	✓	✓
	Data downloading	✓	✓	✓
	CPU core	✓	✓	✓
Elapsed Time	Data Uploading	✓	✓	✓
	Data downloading	✓	✓	✓
	Data processing	✓	✓	✓
Processing	IoT devices	✓	✓	✓
	MEC Server	✓	✓	✓
	Cloud	✓	✓	✓
Architecture	IoT devices	✓	✓	✓
	MEC Servers	✓	✓	✓
	Cloud	✓	✓	✓

IV. CHALLENGES AND OPEN ISSUES

There are many challenges and issues in Edge computing, this arises because edge computing can be present as: cloudlet, Fog computing, and Mobile Edge Computing (MEC) [45]. In this paper, the most important challenges are presented. In order to achieve a correct optimization for the task offloading in edge computing, all the factors presented in the previous section should be taken into consideration. The challenges can be summarized as follow:

- Deployment Strategy – This includes where to deploy workloads, and the location to execute the tasks, which affects the performance due to the transmission cost and the cost consumed during processing.
- Offloading Tasks

The distributed computing environment has evolved into the development of many tasks split techniques that can be performed multiple times in different geographic locations. The sample workflow has been broken down to run in different locations. Segregation of duties is usually a language or management tool. However, it uses boundary nodes to offload the computation, which isn't just a challenge for the IT department. Tasks are efficient but automated. Flexibly define computational pipelines: hierarchically sequential (first in the data center, then in the edge node, or first in the edge node, then in the data center) or possibly simultaneously on multiple edge nodes. There is an intrinsic need for distributable schedules for activities that are segmented into edge nodes [46].

- Cost of Resource allocation

Most of the research lacks considering the cost of hiring the resources from the cloud provider. The large systems required using a lot of resources provided by the cloud which required a huge cost [47].

- Latency

Latency is one of the most important factors to evaluate the performance of a system, especially in interactive applications. Providing servers with cloud computing high computing power can handle complex workloads like image processing. After calculating the time, long WAN latency can dramatically affect the behavior of real-time applications. Workload needs to be improved to reduce the latency of completing the tasks offloading on the edge of the network [48].

- Energy

Energy consumed in the edge computing system is an open issue and is affected by many factors. First, consider the performance characteristics of the workload. Data size in addition to network signal strength, and available bandwidth also affects transmission energy consumption. It is preferable to use only edge computing if the transmission cost is less than the local calculation. However, when dealing with the entire edge computing process, overall power consumption, not just endpoints is the cumulative energy cost of each layer used. Energy consumption for each layer, similar to the endpoint layer can be estimated as local computation and transmission cost. In this case, the optimal workload allocation strategy is to be changed [49].

- Architecture

Another issue is considering the three layers in estimating the optimized values of the cloud layer in addition to the edge and the user layer, which can affect the overall system cost.

- Priority of the Tasks

The priority of tasks is a critical issue to be considered in the decision making of optimization algorithms. This has to be considered as one of the parameters in calculating the cost of the system. Regarding the tasks to be offloaded, the task deadline to be included in the decision of the optimization decision is still an open issue.

- Studying the effect of using the 6G technology in the optimization process of task offloading. Most of the research so far focuses on networking and communication taking into consideration the 5G technology \, however, more efforts need to be paid to study the 6G technology.

V. CONCLUSION

Optimization algorithms and computation offloading are very important techniques to improve the performance of edge computing. In this survey, the papers about performance optimization and including the computation offloading are presented and discussed. Performance optimization can be affected by many factors. The main contribution of this work is dedicated to the optimization algorithms in task computation offloading in edge computing. The survey showed that there is still a lack of available research to cover all the factors that affect the optimization decision making in edge computing, this includes considering the three layers of edge computing and considering the cost of both energy and time of transmitting and processing the data. More work needs to be considered in this field. This can help other researchers to select suitable optimization algorithms for their scenarios and applications.

VI. REFERENCES

[1] EkramHossain, *No Fog, Edge, and Pervasive Computing in Intelligent IoT Driven Applications*, vol. 4, no. 1. 2557.

[2] Y. Chen, N. Zhang, Y. Wu, and S. Shen, *Energy Efficient Computation Offloading in Mobile Edge Computing*. Wireless Networks, 2022.

[3] Suparyanto dan Rosad (2015, *Edge Computing Fundementals,Advances and Application*, vol. 5, no. 3. 2020.

[4] Z. He and K. Li, "Server configuration optimization in mobile edge computing : A cost-performance tradeoff perspective," no. July 2020, pp. 1–28, 2021, doi: 10.1002/spe.2951.

[5] P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, and A. Mouzakitis, "A Taxonomy and Survey of Edge Cloud Computing for Intelligent Transportation Systems and Connected Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6206–6221, 2022, doi: 10.1109/TITS.2021.3084396.

[6] L. A. Haibeh, M. C. E. Yagoub, and A. Jarray, "A Survey on Mobile Edge Computing Infrastructure: Design, Resource Management, and Optimization Approaches," *IEEE Access*, vol. 10, pp. 27591–27610, 2022, doi: 10.1109/ACCESS.2022.3152787.

[7] P. Wang, K. Li, B. Xiao, and K. Li, "Multiobjective Optimization for Joint Task Offloading, Power Assignment, and Resource Allocation in Mobile Edge Computing," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 11737–11748, 2022, doi: 10.1109/JIOT.2021.3132080.

[8] Z. Wang, P. Li, S. Shen, and K. Yang, "Task offloading scheduling in mobile edge computing networks," *Procedia Comput. Sci.*, vol. 184, no. 2019, pp. 322–329, 2021, doi: 10.1016/j.procs.2021.03.041.

[9] N. Kumari, A. Yadav, and P. K. Jana, "Task offloading in fog computing: A survey of algorithms and optimization techniques," *Comput. Networks*, vol. 214, no. July, 2022, doi: 10.1016/j.comnet.2022.109137.

[10] K. Sadatdiynov, L. Cui, L. Zhang, J. Z. Huang, S. Salloum, and M. S. Mahmud, "A review of optimization methods for computation offloading in edge computing networks," *Digit. Commun. Networks*, no. April 2021, 2022, doi: 10.1016/j.dcan.2022.03.003.

[11] L. Liu, C. Chen, J. Feng, T. T. Xiao, and Q. Q. Pei, "A Survey of Computation Offloading in Vehicular Edge Computing Networks," *Tien Tzu Hsueh Pao/Acta Electron. Sin.*, vol. 49, no. 5, pp. 861–871, 2021, doi: 10.12263/DZXB.20200936.

[12] P. Cong, J. Zhou, L. Li, K. Cao, T. Wei, and K. Li, "A Survey of Hierarchical Energy Optimization for Mobile Edge Computing: A Perspective from End Devices to the Cloud," *ACM Comput. Surv.*, vol. 53, no. 2, 2020, doi: 10.1145/3378935.

[13] X. Shan, H. Zhi, P. Li, and Z. Han, "A Survey on Computation Offloading for Mobile Edge Computing Information," *Proc. - 4th IEEE Int. Conf. Big Data Secur. Cloud, BigDataSecurity 2018, 4th IEEE Int. Conf. High Perform. Smart Comput. HPSC 2018 3rd IEEE Int. Conf. Intell. Data Secur.*, pp. 248–251, 2018, doi: 10.1109/BDS/HPSC/IDS18.2018.00060.

[14] H. Lin, S. Zeadally, Z. Chen, H. Labiod, and L. Wang, "A survey on computation offloading modeling for edge computing,"

J. Netw. Comput. Appl., vol. 169, p. 102781, 2020, doi: 10.1016/j.jnca.2020.102781.

[15] A. Shakarami, M. Ghobaei-Arani, and A. Shahidinejad, "A survey on the computation offloading approaches in mobile edge computing: A machine learning-based perspective," *Comput. Networks*, vol. 182, no. August, p. 107496, 2020, doi: 10.1016/j.comnet.2020.107496.

[16] M. Maray and J. Shuja, "Computation Offloading in Mobile Cloud Computing and Mobile Edge Computing: Survey, Taxonomy, and Open Issues," *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/1121822.

[17] L. Lin, X. Liao, H. Jin, and P. Li, "Computation Offloading Toward Edge Computing," no. 1, pp. 1–15.

[18] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Futur. Gener. Comput. Syst.*, vol. 97, no. April, pp. 219–235, 2019, doi: 10.1016/j.future.2019.02.050.

[19] C. Surianarayanan, J. J. Lawrence, P. R. Chelliah, E. Prakash, and C. Hewage, "A Survey on Optimization Techniques for Edge Artificial Intelligence (AI)," *Sensors*, vol. 23, no. 3, p. 1279, 2023, doi: 10.3390/s23031279.

[20] Y. Zhang, *Edge*, 9th ed. USA: Creative Commons Attribution 4.0 International, 2022.

[21] I. Sittón-Candanedo, R. S. Alonso, J. M. Corchado, S. Rodríguez-González, and R. Casado-Vara, "A review of edge computing reference architectures and a new global edge proposal," *Futur. Gener. Comput. Syst.*, vol. 99, no. 2019, pp. 278–294, 2019, doi: 10.1016/j.future.2019.04.016.

[22] C. Martin Fernandez, M. Diaz Rodriguez, and B. Rubio Munoz, "An edge computing architecture in the internet of things," *Proc. - 2018 IEEE 21st Int. Symp. Real-Time Comput. ISORC 2018*, pp. 99–102, 2018, doi: 10.1109/ISORC.2018.00021.

[23] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges," *Futur. Gener. Comput. Syst.*, vol. 78, pp. 680–698, 2018, doi: 10.1016/j.future.2016.11.009.

[24] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge Computing in Industrial Internet of Things: Architecture, Advances and Challenges," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 4, pp. 2462–2488, 2020, doi: 10.1109/COMST.2020.3009103.

[25] P. Mach and Z. Becvar, "Mobile Edge Computing: A Survey on Architecture and Computation Offloading," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017, doi: 10.1109/COMST.2017.2682318.

[26] B. Dab, N. Aitsaadi, and R. Langar, "Q-Learning Algorithm for Joint Computation Offloading and Resource Allocation in Edge Cloud," *IEEE/IFIP Int. Symp. Integr. Netw. Manag.*, pp. 45–52, 2017.

[27] C. Wang, C. Dong, J. Qin, X. Yang, and W. Wen, "Energy-efficient Offloading Policy for Resource Allocation in Distributed Mobile Edge Computing," *2018 IEEE Symp. Comput. Commun.*, pp. 366–372, 2018.

[28] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Performance Optimization in Mobile-Edge Computing via Deep Reinforcement Learning," in *IEEE Vehicular Technology Conference*, 2018, vol. 2018-Augus, doi: 10.1109/VTCFall.2018.8690980.

[29] X. Gu, L. Jin, N. Zhao, and G. Zhang, "Energy-Efficient Computation Offloading and Transmit Power Allocation Scheme for Mobile Edge Computing," *Mob. Inf. Syst.*, vol. 2019, 2019, doi: 10.1155/2019/3613250.

[30] H. Li, S. Ci, C. Yang, X. Tan, S. Hou, and Z. Wang, "Moving to Green Edges : A Cooperative MEC Framework to Reduce Energy Demand of Clouds," *2019 IEEE Globecom Work. (GC Wkshps)*, pp. 1–6, 2019.

[31] K. Li, "Computation Offloading Strategy Optimization with Multiple Heterogeneous Servers in Mobile Edge Computing," *IEEE Trans. Sustain. Comput.*, vol. XX, pp. 1–1, 2019, doi: 10.1109/tsusc.2019.2904680.

[32] J. Fu, J. Hua, J. Wen, H. A. O. Chen, and J. Li, "Optimization of energy consumption in the MEC-assisted multi-user FD-SWIPT system," vol. 4, 2020, doi: 10.1109/ACCESS.2020.2969467.

[33] Z. Chen, J. Hu, G. Min, and X. Chen, "Effective data placement for scientific workflows in mobile edge computing using genetic particle swarm optimization," no. March 2019, pp. 1–11, 2020, doi: 10.1002/cpe.5413.

[34] L. Chen, J. Wu, and J. Zhang, "Long-term optimization for MEC-enabled HetNets with device – edge – cloud collaboration," *Comput. Commun.*, vol. 166, no. October 2020, pp. 66–80, 2021, doi: 10.1016/j.comcom.2020.11.011.

[35] P. Yan, "Optimizing Mobile Edge Computing Multi-Level Task Offloading via Deep Reinforcement Learning," in *IEEE International Conference on Communications (ICC)*, 2020, p. 7.

[36] W. Li, F. Wang, Y. Pan, L. Zhang, and J. Liu, "Computing Cost Optimization for Multi-BS in MEC by Offloading," *Springer Nat. Mob. Networks Appl.*, 2020.

[37] L. T. Hsieh, H. Liu, Y. Guo, and R. Gazda, "Quality of Service Optimization in Mobile Edge Computing Networks via Deep Reinforcement Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12384 LNCS, pp. 145–157, doi: 10.1007/978-3-030-59016-1_13.

[38] Z. Lv and L. Qiao, "Optimization of collaborative resource allocation for mobile edge computing," *Int. J. Comput. Telecommun. Ind. Int. J. Comput. Telecommun. Ind. Int. J. Comput. Telecommun. Ind. Comput. Commun. ne*, vol. 161, no. July, pp. 19–27, 2020, doi: 10.1016/j.comcom.2020.07.022.

[39] C. Zhan, H. Hu, X. Sui, Z. Liu, and D. Niyato, "Completion Time and Energy Optimization in the UAV-Enabled Mobile-Edge Computing System," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7808–7822, 2020, doi: 10.1109/JIOT.2020.2993260.

[40] S. Xiao, C. Liu, K. Li, and K. Li, "System delay optimization for Mobile Edge Computing," *Futur. Gener. Comput. Syst.*, vol. 109, pp. 17–28, 2020, doi: 10.1016/j.future.2020.03.028.

[41] Z. Liao, J. Peng, B. Xiong, and J. Huang, "Adaptive offloading in mobile-edge computing for ultra-dense cellular networks based on genetic algorithm," *J. Cloud Comput.*, vol. 10, no. 1, p. 16, 2021, doi: 10.1186/s13677-021-00232-y.

[42] G. Peng, H. Wu, H. Wu, and S. Member, "Constrained Multiobjective Optimization for IoT-Enabled Computation Offloading

in Collaborative Edge and Cloud Computing," vol. 8, no. 17, pp. 13723–13736, 2021.
[43] Z. Kuang, Z. Ma, Z. Li, and X. Deng, "Cooperative computation offloading and resource allocation for delay minimization in mobile edge computing ☆," *J. Syst. Archit.*, vol. 118, no. July 2020, p. 102167, 2021, doi: 10.1016/j.sysarc.2021.102167.
[44] A. Abbas, A. Raza, F. Aadil, and M. Maqsood, "Meta-heuristic-based offloading task optimization in mobile edge computing," *Int. J. Distrib. Sens. Networks*, vol. 17, no. 6, 2021, doi: 10.1177/15501477211023021.
[45] 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, SmartCloud 2016*, pp. 20–26, 2016, doi: 10.1109/SmartCloud.2016.18.
[46] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge Computing for Autonomous Driving: Opportunities and Challenges," *Proc. IEEE*, no. 1, pp. 1–15, 2019, doi: 10.1109/JPROC.2019.2915983.
[47] 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, doi: 10.1109/JIOT.2016.2579198.
[48] E. Ahmed and M. H. Rehmani, "Mobile Edge Computing: Opportunities, solutions, and challenges," *Futur. Gener. Comput. Syst.*, vol. 70, pp. 59–63, 2017, doi: 10.1016/j.future.2016.09.015.
[49] H. Li, G. Shou, Y. Hu, and Z. Guo, "Mobile edge computing: Progress and challenges," in *Proceedings - 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2016*, 2016, pp. 83–84, doi: 10.1109/MobileCloud.2016.16.