ORIGINAL PAPER

Replica selection and placement techniques on the IoT and edge computing: a deep study

Zhong-Liang Shao[1] · Cheng Huang[2] ● · Heng Li[3]

Abstract

Internet of Things (IoT) has lately been presented as a new technological transformation in which things are connected via the Internet. Several sensors and devices create data and send vital signals constantly over sophisticated networks that allow machine-to-machine interactions and monitor and manage key smart-world infrastructures. Since huge amounts of data are generated, reducing the data access costs is a critical issue. Edge computing has been developed as a novel paradigm for solving IoT demands to reduce the rise in resource congestion. One of the most significant data management challenges in the IoT is selecting suitable replication things that minimize reaction time and cost. Therefore, our goal is to examine replica selection and placement techniques in IoT and edge computing. The findings revealed that the edge computing environment might significantly enhance system performance regarding access response time, prediction accuracy, effective network, and increased data availability. Furthermore, the results illustrate that data provenance is necessary to raise the accuracy of the data by. Also, the results showed that the most important challenge in data replication and placement techniques in IoT and edge computing was the availability of data and access response time.

Keywords Replica selection · Placement techniques · Internet of Things · Edge computing · Systematic review

1 Introduction

Nowadays, competitiveness goes through several aspects: digitalization, productivity, and environmental impact. Technology is advancing fast, and helping industries obtain more detailed data about their processes and equipment. In fact, the possibility to monitor and control each part of the process is a strong base on which a more intelligent and focused control can be built. Technology advance brings innovation and the possibility to manage production in terms of the"near future" through AI prediction and decision-making support [1]. The Internet of Things (IoT), wireless systems, V2X networks, and other technologies have emerged due to advancements in the information technology domain, and they now play an important part in our everyday lives [2–4]. Although the IoT has solved many problems, some limitations will be addressed in this article. In this section, IoT and edge computing will be introduced, and the need for the role of replica selection will be considered in solving IoT problems.

1.1 What is the Internet of Things (IoT)?

The fast advancement of technology in the domains of information and communication has made the IoT unavoidable in our day-to-day activities; now, it plays an important role in every aspect of our daily routine [5, 6]. IoT is a novel term that focuses on developing computing things and devices and allows customers to use these gadgets that are linked via the Internet [7, 8]. As a contemporary concept, the IoT has been utilized to represent a future in which physical objects such as Radio-Frequency

✉ Cheng Huang
chenghuang888888@sohu.com;
wangwangho55@gmail.com

1 Department of Automation Engineering, Guangdong Polytechnic of Water Resources and Electric Engineering, Guangzhou 510635, China

2 Department of Mechanical Engineering, Guangdong Polytechnic of Water Resources and Electric Engineering, Guangzhou 510635, China

3 R & D Center, Guangdong Ruigong Intelligent Technology Co., Ltd., Guangzhou 510665, China
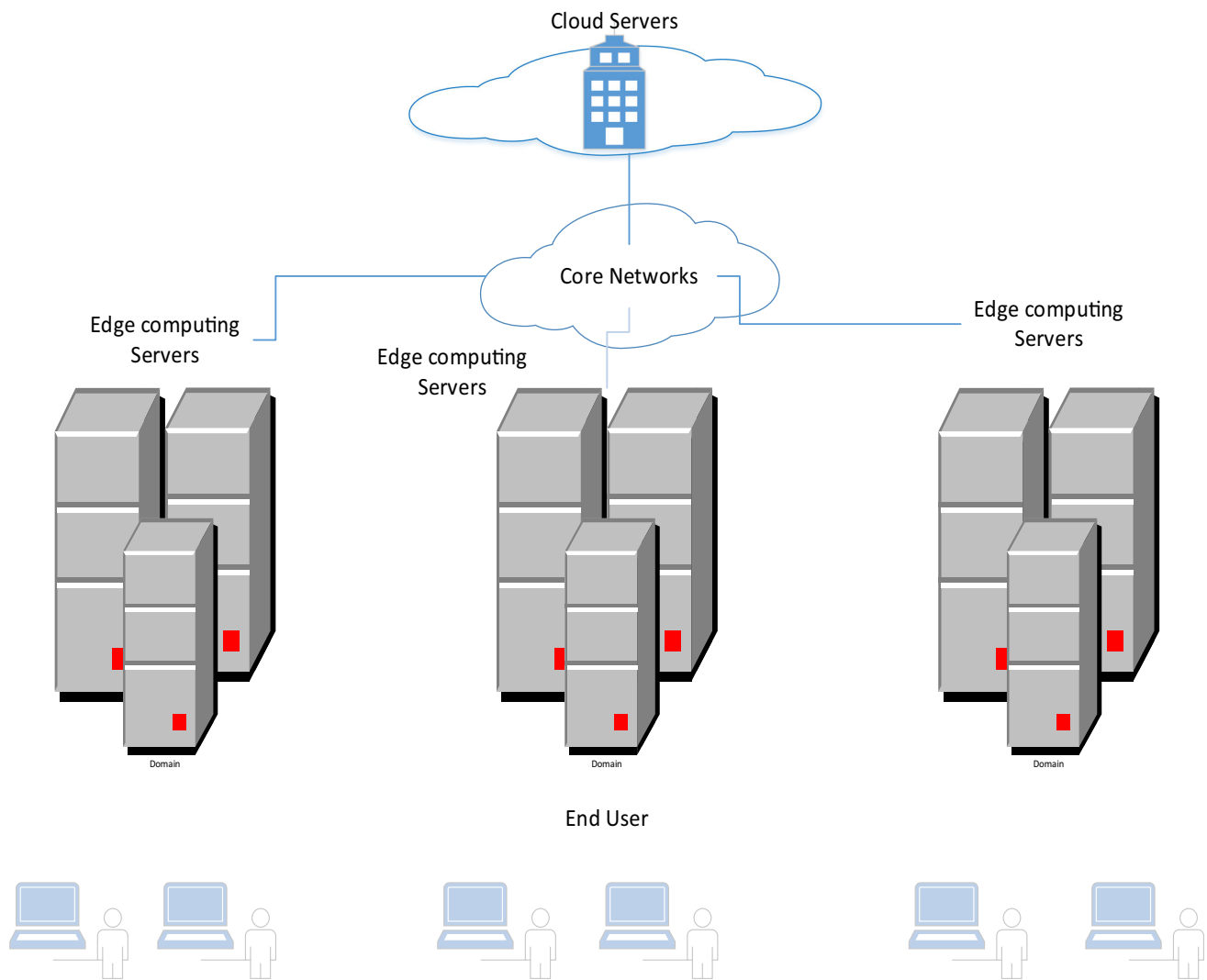
Identification (RFID) tags, sensors, actuators, and smart-phones may interact to achieve common goals [9]. By linking the tangible objects in our surroundings, the IoT allows for seamless communication, resulting in various benefits such as quick information transmission, maintenance, and monitoring [10]. The IoT's inherent benefits have made it more popular across various applications, leading to a massive data flow and management [11, 12]. There are linked devices or sensors that can gather and share data using a contemporary communication network [13, 14]. Transportation systems, environmental monitoring, home automation, social life, intelligent systems (for example, in the petroleum), energy conservation, education, entertainment, fitness, and health care (especially in the Covid-19 era) can benefit from cloud computing and IoTs [15–17]. Since IoT processes create and utilize massive quantities of data, it must be preserved and processed on edge nodes or in a distant cloud datacenter [18]. These large datasets must have been distributed once tasks were assigned to single or composite resources [19]. Because of storage and computational capacity limitations, it is difficult to store all of these massive datasets in a single edge cluster; therefore data transfer is unavoidable when IoT processes are run. As a result, under resource and delay limitations, a data manager should use an appropriate data placement approach to reduce data transfer costs among geographically scattered edge clusters [20]. Figure 1 shows the basic view of IoT and related things.

Fig. 1 The basic IoT overview and related things

1.2 What is edge computing?

Edge computing is developing as a new paradigm for addressing the needs of localized computing and the IoT. It is a technique that can minimize an escalation in resource congestion. It is a type of distributed computing in which edge servers handle the primary computation and data. Edge computing is a networking technique that allows processing on network edges, allowing cloud services to process downstream data and IoT applications to process upstream data. Hu et al. [21] defines it as a new model that deploys computing and storage resources to networks closer to mobile devices or sensors. Its architecture is shown in Fig. 2, which is divided into four-layer, core networks, end-user, edge computing servers, and cloud servers. The main feature of edge computing is to co-locate computing, network control, and storage resources at the edge computing aiming at reducing latency [22]. Edge computing, like cloud computing, will have an important impact on our civilization [23, 24]. Furthermore, providing edge servers with storage and processing characteristics can minimize edge nodes while increasing the throughput among nodes and edge nodes [25]. Edge computing reduces response time, increases application scheduling, and increases computational efficiency by monitoring network forwarding, storage, and processing [26]. In summary, micro clouds, fog computing, and mobile edge computing all fall under the umbrella of Edge computing [27]. Power, storage, and computation are brought to the network's edge to improve availability and reduce latency, eventually defeating bandwidth-hungry and delay-sensitive apps near the data source [28].

1.3 What is the necessity of data replication and placement techniques in IoT and edge computing platforms?

Replica technology, which implies storing numerous replicas of a data item on various nodes in a distributed system, is one of the most widely utilized approaches [29]. In all scattered settings, selecting the correct replica and accessing a data file is critical. It can lower the risk of an attack and increase security. The efficiency of a system can be improved with a proper replica approach [30]. Data access performance may be improved by replica [31]. Service providers and customers can optimize the use of data centers by choosing an appropriate replica [32]. It is necessary to know about the properties of a storage mechanism based on client needs. Choosing an appropriate replica increases reaction time and lowers costs [33].

Besides, in the IoT context, some large devices created a need for data storage. In sectors like traffic or smart vehicle

Fig. 2 The basic edge computing architecture

speed, personal health data, or energy consumption statistics, the accuracy of such sensitive data is critical. As a result, these distributed storage systems incorporate IoT settings to manage high demand while also ensuring data accuracy. By employing data replication, these distributed storage systems will enhance dependability and extend Mean Time to Failure (MTF) rate. Data will be split into many parts, each with a defined size determined by code blocks. Edge computing-based storage will be a distributed storage system that is well-distributed both conceptually and physically. As a result, any IoT data may be copied, and all data pieces may be saved in multiple geological regions with the use of edge computing-based storage. It can significantly reduce the risk of data loss [34, 35]. Figure 3 summarizes the role of edge computing in the IoT. As we see in Fig. 1, embedded sensor devices keep track of local environmental conditions and relay the data to an IoT gateway. The data is subsequently filtered and/or inferred

by an edge computing job at the IoT gateway. For additional analytics or control, the processed data is transmitted to other application hosts. An IoT gateway may replicate data to a backup gateway to assure data delivery by utilizing the primary backup method. If the primary IoT gateway fails, the backup edge computing activities may process the waiting data, and the embedded sensing devices can re-transmit data to the backup gateway [36].

Communication delays are a major issue for several IoT systems and can easily become a bottleneck. As a result, data replication, which brings data closer to customers, is viewed as a precise solution that may reduce network latency and bandwidth use [37]. Nevertheless, duplicated resources and objects might drive up data center costs significantly. In the IoT, replica choosing is described as a method of picking relevant things for a given service to assist both users and IoT administrators in improving data center and IoT efficiency [33].

Fig. 3 Edge computing for the IoT
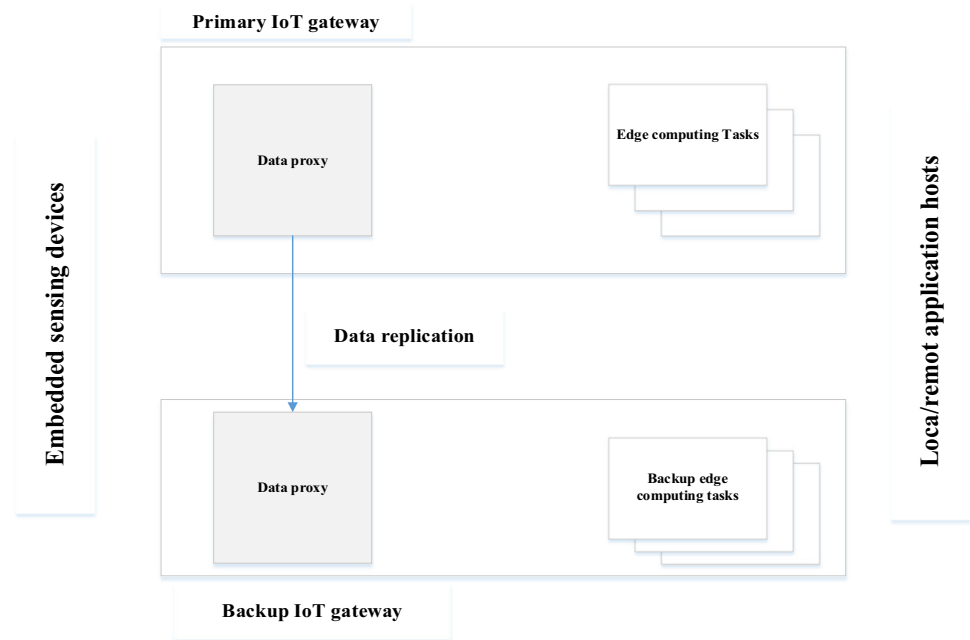
Several IoT apps rely on fast and accurate data processing from embedded sensor devices. Computing tasks are conducted on IoT gateways to obtain a quick response, and the gateways provide data replication to backup gateways for fault tolerance [36]. Edge computing and data replication set the stage for fast and dependable IoT workflow localization services. On the one side, edge computing moves computational capability to the network's edge, closer to IoT devices. Each edge cluster comprises one or more edge computing nodes linked by high-bandwidth connections [38]. As the Internet's edge devices get more capable, these devices in the edge computing setting will function even better to enhance the system's processing and storage capacities [39]. If the replica node cannot be handled efficiently in the edge cloud environment, it will generate issues like slow user access and low system fault tolerance [30, 40].

1.4 Contributions of the paper

The following sentences summarize the key contributions of this article in relation to the suggested research topics and challenges. Briefly, the goals of this paper are:

- Providing valuable information about the role of data replication and placement techniques in IoT and edge computing;
- Offering the systematic and methodical review of the data replication and placement techniques in the IoT and edge computing;

- Exploring main challenges in data replication and placement techniques and presenting the guidelines to face the current challenges;
- Outlining the key zones where the data replication and placement techniques in IoT and edge computing can be improved.

According to the goals provided, our goal is to investigate the role of data replication and placement techniques in IoT and edge computing.

1.5 Organization of the paper

In the continuation of this section, the systematic method is examined. Section 3 analyzes IoT-related articles and edge computing. Afterward, in the next section, the results are presented. The most important research challenges are discussed in the open issue section. Ultimately, the last section concludes the article. The organization of the paper is illustrated in Fig. 4.

2 Research methodology

A Systematic Literature Review (SLR) is a technique for discovering, analyzing, and constructing all related articles on a specific research topic, subject field, or critical phenomena [41, 42]. The SLR enables a precise assessment relying on a study matter utilizing a dependable, explicit, and auditable method, which comprises limited systematic faults, mitigated chance effects, increased legality and authority of the ensuing evidence, and more reliable
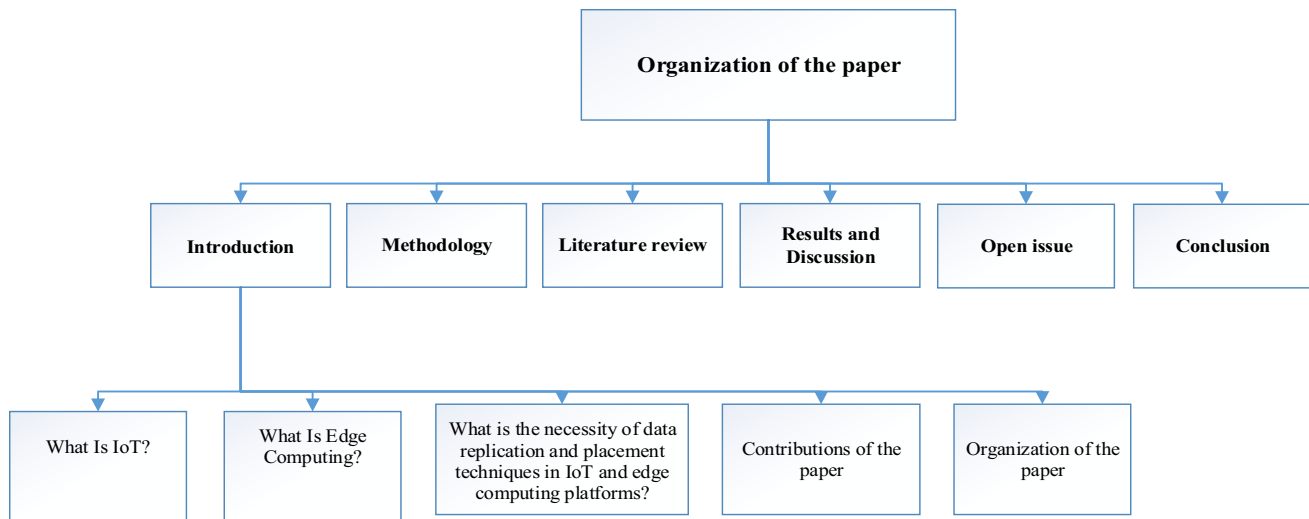
Fig. 4 The organization of the paper

outcomes [43]. The present paper aims to specify the most relevant subject and challenges in replica selection and placement techniques on IoT and edge computing. According to the study aim, the study hypothesis is designed as follows.

- IoT and edge computing are very effective in solving replica selection and placement techniques problems.

Thus, this research will try to tackle the questions below:

- What is IoT? (The response to this question can be found in the introduction section)
- What is edge computing? (The response to this question can be found in the introduction section)
- What is the necessity of data replication and placement techniques in IoT and edge computing platforms? (This question is answered in the introduction section)
- What are the challenges of data replication and placement techniques in IoT and edge computing platforms? (This question is answered in the open issue section)

In IoT and edge computing, the current research is a comprehensive review of data replication and placement strategies. The results of this study are based on articles published in prestigious international journals. Articles are selected from Google Scholar, Scopus, IEEE, Since direct, Springer, Wiley, and Sage databases. The references of identified articles were also reviewed for further study. Articles were searched using the following keywords.

1. "Internet of Thing" and "Data replication"
2. "IoT" and "Data replication"
3. "Edge computing" and "Data replication"

4. "Internet of Thing" or "IoT" and "Edge computing" and "Data replication"

First, all articles related to data replication and placement techniques in IoT and edge computing were collected. At this stage, all articles whose titles or abstracts contained the mentioned keywords were included in the initial list. Accordingly, in the initial search, 48 articles were obtained as follows.

1. "Internet of Thing" and "Data replication" = 8
2. "IoT" and "Data replication" = 14
3. "Edge computing" and "Data replication" = 26
4. "Internet of Thing" or "IoT" and "Edge computing" and "Data replication" = 0

Only the abstracts of the 8 existing articles were available; so, they have been eliminated. The following step was to eliminate 5 duplicate papers from the list of articles. Of the remaining 35 articles, 6 papers were eliminated from the list due to their Chinese and Japanese language. Finally, 29 articles reached the quality determination stage. At this stage, the text of 29 articles was totally read. The subjects of 10 articles were different, although their titles were related to our topic. Therefore, they were excluded from the study, and 19 articles were included in the final list for analysis. Figure 5 shows the process of including studies in the present investigation.

A checklist of required research info, including place of study, article title, study time, was prepared for final evaluation. Data were searched and extracted by the authors independently and evaluated based on the checklist. There were 18 cross-sectional studies and 1 descriptive study. Inclusion requirements were not limited to the time, place, and type of study but included all related articles. Table 1 shows the specifications of the selected articles.
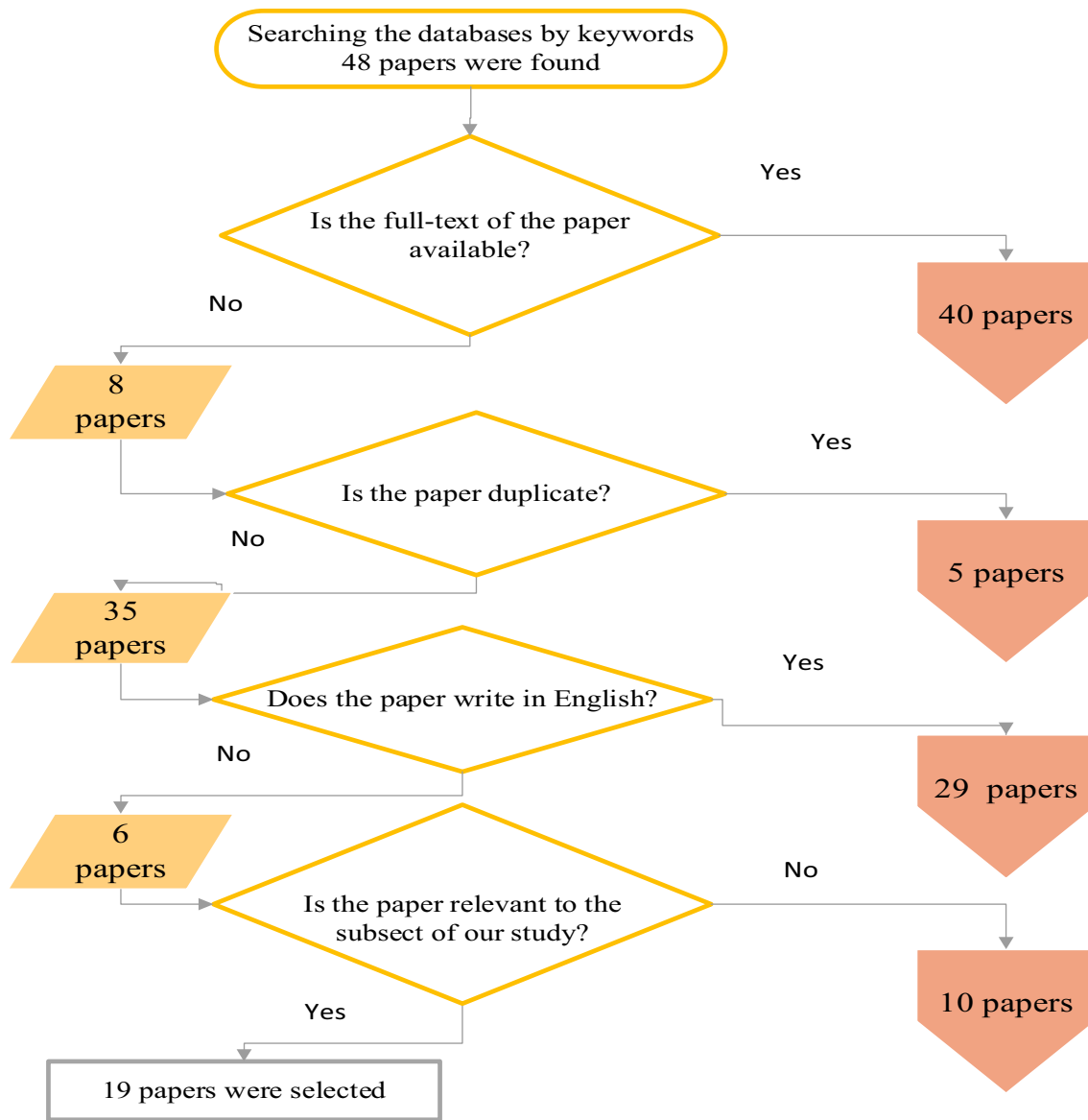
Fig. 5 The process of study inclusion into the present study

Figure 6 also shows the year-publication chart of the analyzed articles. As can be seen, most of the articles are related to the IEEE publications in 2020. Therefore, it can be concluded that this is a new and evolving subject that researchers need to study more.

Figure 6 shows the number of each article and the publications related to each one by year of the publication.

As it was observed, 19 articles were selected based on different filters in this section. Selected articles were presented in Table 1 and Fig. 6. Selected articles were analyzed in Sect. 4, and their key points were summarized in Table 2. Selected articles were analyzed in two categories. The articles in each category are shown in Fig. 7.

3 Literature review

A discussion of relevant literature is included in this section. Various recent research on replica selection in IoT and edge computing have been published. The remainder of this section examines some of the most significant replication techniques in edge computing and IoT.

3.1 Data replication and placement techniques in IoT

The idea of enabling communication between any kind of devices has led to the emergence of IoT [44]. The growing rate of data significantly outpaces the increased rate of network capacity as the Internet of Everything (IoE)

Table 1 Specifications of selected articles for analysis in this article

Source	Year	Environment	Publication	Journal	Cited
[71]	2012	IoT	CNKI	Journal of Central South University (Science and Technology)	1
[37]	2013	IoT	IET	IET communications	8
[72]	2014	IoT	IEEE	International conference on fuzzy systems and knowledge discovery (FSKD)	14
[73]	2017	Edge	IEEE	IEEE transactions on sustainable computing	20
[67]	2018	Edge	IEEE	IEEE transactions on network and service management	59
[74]	2018	IoT	IEEE	In2018 IEEE 16th Intl. Conf. on dependable, autonomic and secure computing	5
[20]	2019	IoT	Elsevier	Computer networks	33
[70]	2019	IoT	IET	IET communications	4
[30]	2019	Edge	Springer	The journal of supercomputing	5
[39]	2019	Edge	Elsevier	Computer communications	11
[75]	2019	IoT	IEEE	IEEE access	10
[9]	2020	IoT	Wiley	International journal of communication systems	1
[76]	2020	IoT	IEEE	In 2020 4th international conference on HVDC (HVDC)	–
[11]	2020	Edge	I-SMAC	Journal of ISMAC	14
[35]	2020	Edge	Springer	Wireless personal communications	6
[36]	2020	Edge	IEEE	In 2020 IEEE/ACM fifth international conference on internet-of-things design and implementation (IoTDI)	4
[77]	2020	Edge	Elsevier	Journal of parallel and distributed computing	13
[69]	2020	Edge	Elsevier	Information sciences	16
[50]	2021	Edge	Wiley	Concurrency and computation: practice and experience	1

Fig. 6 Distribution of the articles by year of publication

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
I-SMAC									1	
CNKI	1									
IET		1						1		
Wiley									1	1
Springer								1	1	
Elsevier								2	2	
IEEE			1			1	2	1	2	
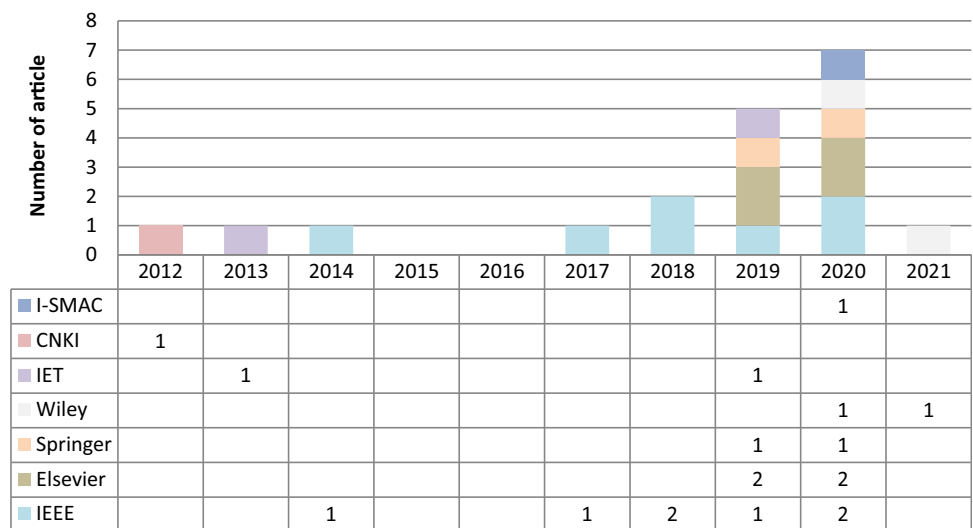
movement continues to develop. Simultaneously, the advent of a slew of new applications, like smart cities and unmanned driving, raises the stakes for delay [45]. Some interlinked devices or sensors can gather and exchange data using a contemporary communication network as an infrastructure that numerous IoT nodes have linked. Following that, there are a variety of IoT apps that may give users precise and fine-grained services. Since the number of potential applications is growing steadily, one of the common problems of these applications is handling big data. Since the number of IoT nodes can increase enormously, the volume of data generated and consumed by these nodes can grow to an unprecedented level. Storing and processing such a significant volume of data become a daunting task [46]. Cloud technology is an attractive alternative to store and process such a large volume of data. Cloud computing can assign complicated and resource-hungry tasks to capable data centers distributed worldwide

Table 2 Key points of the analyzed articles in the data replication and placement techniques in IoT

| Source	Main idea	Technique	Results
Data replication and placement techniques in IoT			
He et al. [71]	Supplying load-balancing and low-cost cloud data replica distribution strategy in IoT setting	A method for selecting low access cost storage server was presented based on access positions and frequencies	When achieving the minimum number of replicas, the replicas are adaptively distributed to the storage servers according to the data access status with low access cost
Cho et al. [37]	Investigating replica recognition in mobile applications	Efficient distributed detection methods of replicas (EDDRs)	The suggested approaches detect replicas with a high degree of accuracy and almost no detection mistakes. Moreover, the techniques' distributed and cooperative strategy decreases the amount of energy needed to identify replicas while also delivering faster replica recognition than previous systems
Zhang et al. [72]	Suggesting a data replica placement scheme for cloud storage under healthcare IoT environment	The suggested approach employs the MOX (Mosquitoes Oviposition Mating) algorithm to determine the best data replica placement answer and the SA (Simulated Annealing) technique to determine the best user access request allocation to the appropriate data replica	The suggested technique can increase system scalability and reliability while also reducing user access time and promoting load balancing
Qaim and Ozkasap [74]	Offering a totally distributed hop-by-hop data replication method for IoT-based wireless sensor systems	A data replication technique for improved data availability in IoT-based sensor systems (DRAW)	DRAW increases data availability and average replicas produced in the network by a maximum of 15% and18%, respectively, compared to a modern method. Besides, DRAW has a greater replica spread, which influences the network's data distribution quality
Shao et al. [20]	Suggesting a new data replica placement method for coordinated processing data-intensive IoT workflows in a collaborative edge and cloud computing environment	Suggesting an IoT-based method for data replica placement in collaborative edge and cloud settings	In comparison to existing algorithms, the suggested technique can not only discover a higher quality solution for data replica placement, but it also requires a smaller computational budget
Wakil et al. [70]	Decreasing the waiting time of IoT applications to access the replicated things utilizing a hybrid algorithm	A hybrid optimization algorithm (Ant Colony Optimization (ACO) and Genetic Algorithm (GA))	Compared to High-QoS (Quality of Service) First-Replication (HQFR) and the dynamic cost-aware re-replication and re-balancing technique, the hybrid method performs better
	Decreasing the replica selection cost in IoT settings utilizing a hybrid algorithm		
Wang et al. [75]	Investigating active data replica recovery for quality-assurance big data analysis in information-centric IoT	Rarity-aware data replica recovery (RADR) algorithm	The RADR technique outperforms a conventional direct data recovery approach by a substantial margin
Wang and Batiha [9]	Enhancing the load balancing among the IoT centers utilizing repellent pheromone	Simulation using the cloudsim in Improved version of ACO (IACO)	In terms of waiting time and load balancing, the suggested solution excelled the ACO, HQFR, and Response Time-oriented Replica Management (RTRM) methods
Liu et al. [76]	Suggesting edge node data replica management method for distribution IoT	The data processing architecture of distribution IoT based on edge computing	The suggested replica management technique, which combines local and cluster management, has a faster average processing time than the conventional data replica management technique and may significantly enhance IoT data processing performance with a large dispersion
		The edge computing data copy management approach is intended to achieve adequate distributed IoT data backup	
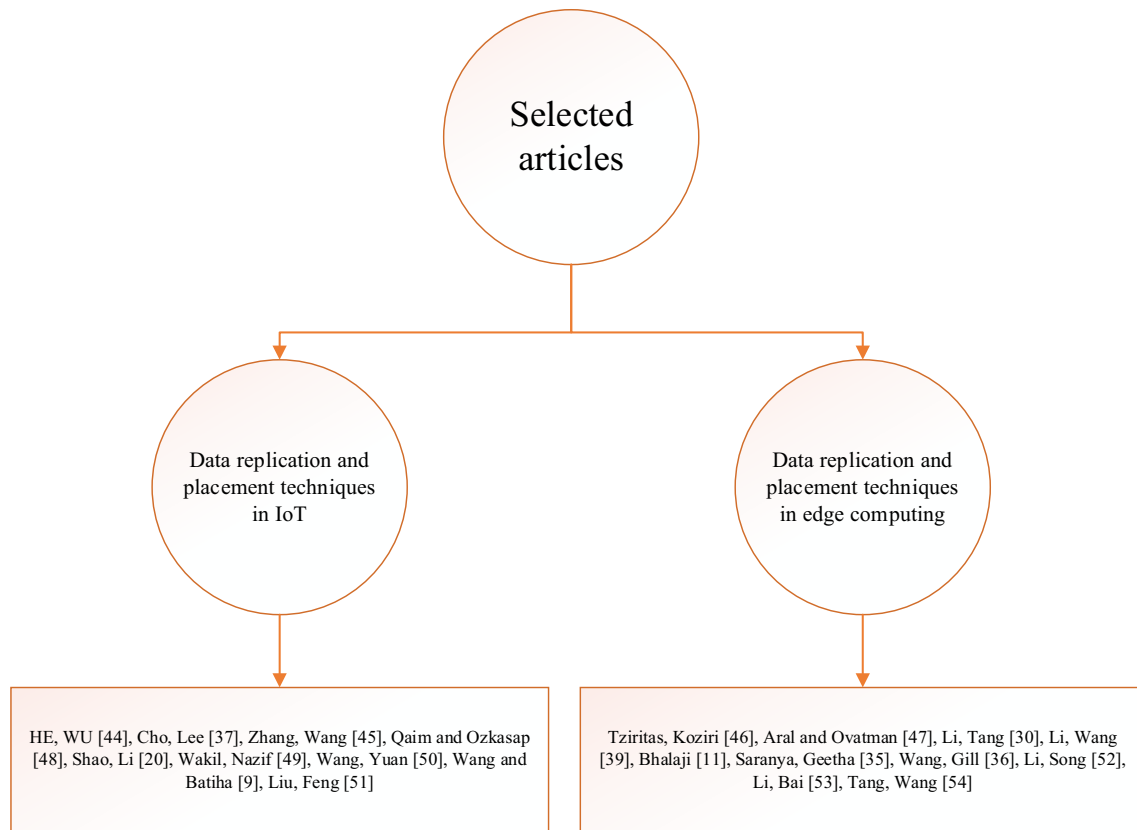
Fig. 7 The categorization of the selected article

[47]. Cloud computing systems will confront the difficulties of large equipment needs for access, inadequate bandwidth, massive data, and excessive power consumption as the IoT age unfold [48, 49]. The convergence of edge and cloud computing shares their strengths, such as unlimited shared storage and computing resources from the cloud, low-latency data preprocessing of edge computing. The collaboration of the two computing paradigms can provide a real-time and cost-effective way to deploy IoT workflows among cooperative user groups [20].

In this section, 9 articles related to data replication and placement techniques in IoT are examined. Also, Table 2 summarizes the literature on data replication and placement techniques in IoT. In addition, we provide information on their intended method, main idea, and results.

3.2 Data replication and placement techniques in edge computing

Edge computing is developing as the modern paradigm that meets the demands of localized computing and the IoT by utilizing this as a method to reduce resource congestion [35]. It is a novel technology that lets media be streamed across a network's edge [50]. There are two main advantages of edge computing: firstly, edge nodes can assist with much lower latency than the cloud because servers are physically closer to the devices, and secondly, edge nodes can shield the cloud from most requests by serving the requests locally. Edge nodes can help collect information from the devices by aggregating information from multiple devices before sending it. Edge nodes can also make information available to local devices to access this information with low latency [51]. Computing at the cloud edges is therefore critical in IoT systems for latency-sensitive apps [52]. The edge computing host may replicate data to a secondary host for fault tolerance to prevent a single point of failure and data loss. However, data replication at the pace of data arrivals is wasteful and consumes a significant amount of network bandwidth. Edge computing, which brings computing capacity to the network's edge, reduces the cloud computing center's computation and transmission load while also shortening its time for users to provide feedback. Some services, on the other hand, need access to data maintained centrally. Therefore, access latency may become a bottleneck, and the benefits of edge computing may be outweighed, particularly for data-intensive services. Furthermore, node failures are common in edge systems because of their dynamism, dispersion, and heterogeneity, leading to data inaccessibility. Data replicas are frequently utilized to ensure the

availability and reliability of data by deploying them on various nodes to minimize access failures and data loss [53]. Because replicas are stored on various nodes, other data remains available if one fails, and the service is not disrupted [39].

In this section, 10 articles related to data replication and placement techniques in edge computing are examined. Also, Table 3 summarizes the literature on data replication and placement techniques in edge computing. In addition, we provide information on their intended method, main idea, and results.

4 Results and discussion

In this section, we will review the results of reviewing articles. The contribution and innovation of the paper, open issues, and suggestions for future work will also be explained.

The IoT is a cutting-edge paradigm that aims to deliver huge apps that are already a part of everyday life. Several smart devices are deployed across complicated networks to perform vital functions like monitoring, communications, and critical infrastructure control. Nevertheless, the rapid increase of IoT devices and the resulting large data traffic produced at the network's edge imposed new demands on the modern centralized cloud computing model because of bandwidth and resource constraints. As a result, edge computing is gaining traction as a novel method for bringing data processing and storage closer to end-users, resulting in edge computing-assisted IoT [54].

Also, the results showed a typical IoT application includes several modules running together with active interdependencies, traditionally running on the cloud-hosted data centers. While all the application modules were traditionally placed on the cloud, this brought in a tremendous network cost and a high application response. With the fog/cloud computing approach, these application modules can be distributed across the fog and the cloud resources based on meeting the module requirements and network capacity constraints. Also, the logarithmic complexity of the module mapping algorithm trumps the usual solution to such problems, which tends to be NP-hard [55]. Also, performance evaluations showed that data replication IoT modules significantly outperform regular IoT modules protocols. Data replication IoT modules use less networking resources. These techniques are, therefore, very efficient with a slow network.

As we discovered, the growth of the IoT has made edge computing a developing answer to the difficulties of managing several devices and sensors and the resources required. Since the IoT is now quite prevalent in data processing at the network's edge, Edge computing has been getting a lot of traction. Several techniques, such as cloudlets, mobile edge computing, and fog computing, offer alternative cloud computing options to reduce data processing at the network edge [28]. The findings also indicated that data replication is ideal for lowering response times, worldwide traffic, and data sharing, as it can be performed even when the server is disconnected [35]. In the continuation of this section, the paradigms of edge computing are discussed.

Fog Computing decentralizes computation and storage, but its model is hierarchical, providing progressively increasing resources along a multi-tier succession of nodes located between the IoT devices and the cloud data centers.

Multi-access Edge Computing, also known as mobile edge computing, provides resources through the Radio Access Network (RAN) infrastructure. This approach is projected to gain greatly from the 5G platform, which is now in development and will make edge computing available to a wide variety of mobile devices with minimal latency [56].

Cloudlet Computing uses computers or clusters of computers, known as cloudlets, with virtualization capabilities. Tasks are taken from mobile clients and divided among the cloudlet nodes in their proximity, which in themselves might also be composed of clustered mobile devices.

Despite presenting some differences, these concepts have been used interchangeably in previous literature and demonstrate how the system models vary within the edge computing paradigm. Despite the fact that each system is, in its majority, composed of nodes that are less resourceful than the cloud data centers, there can still be found a large variance in the storage capabilities and network reach of each node, and how these communicate with one another, according to the topology and routing algorithms [51]. In the continuation of this section, we will examine the innovation of the paper.

Finally, in this section, we want to compare the main parameters that the articles reviewed in this article have studied. Hence Table 4 shows the comparison of the main parameters of the articles. Comparing the articles in Table 4 showed that the most important challenge in data replication and placement techniques in IoT and edge computing was the availability of data and access response time that most researchers have tried to increase the access response time and availability of data.

- Innovation of the paper

Based on the review studies in this paper, many articles examined replica selection on the IoT. Also, the role of replica selection in edge computing has been extensively researched. But no article systematically discusses the role of replica selection and placement techniques on the IoT

Table 3 Key points of the analyzed articles in the data replication and placement techniques in edge computing

Source	Main idea	Technique	Results
Tziritas et al. [73]	Examining data replication and virtual machine migrations to mitigate network overhead in edge computing systems	Proposing an algorithm based on hyper-graph partitioning	The algorithm yields up to 53% network overhead reduction than state-of-the-art algorithms found in the literature
Aral and Ovatman [67]	Proposing a decentralized replica placement algorithm for edge computing	A method for disseminating data depends on the dynamic generation, replacement, and removal of replicas, led by constant monitoring of data requests originating from the underlying network's edge nodes	When opposed to client-side caching, which is frequently utilized in conventional distributed systems, a decentralized replica placement technique offers considerable cost savings
Li et al. [30]	Proposing replica creation and selection strategy based on the edge cloud architecture	Suggesting the DRC-AH (Dynamic Replica Creation Algorithm based on access Heat) and replica selection algorithm -NSC (Node Service Capability) algorithms (DRS-NSC)	The suggested algorithms have substantial benefits in prediction accuracy, resource usage, and other areas, user request response time, enhancing the system's performance to a degree
Li et al. [39]	Proposing flexible replica placement for enhancing the availability in the edge computing environment	The replica placement algorithm (FNSG) and (DRC-GM)	In an edge computing context, the DRC-GM and replica placement-FNSG algorithms may significantly enhance system performance regarding prediction accuracy, effective network and storage space use, access response time, increasing data availability
Bhalaji [11]	Examining efficient and secure data utilization in mobile edge computing by data replication	Proposing method was simulated using the network simulator-2	The duplicating method improves bandwidth usage while lowering power usage and reaction time
Saranya et al. [35]	Proposing data replication in mobile edge computing systems to reduce latency in IoT	Proposing both simple and random algorithms of replication	When compared to a straightforward replication method, a random technique for replicating can reach a bandwidth that is superior in terms of savings
Wang et al. [36]	Adaptive data replication in real-time reliable edge computing for IoT	Presenting the adaptive real-time reliable edge computing architecture	The suggested data replication techniques and architecture can provide the required levels of data loss tolerance while reducing network bandwidth use and preserving latency performance
Li et al. [77]	Examining effective replica management for improving reliability and availability in an edge-cloud computing environment	Proposing a dynamic replica creation strategy based on the gray Markov chain	While guaranteeing load balancing, the suggested dynamic replica generation approach significantly decreases system response time, increases data read throughput, and enhances system storage space usage
Li et al. [69]	Proposing resource and replica management strategy for optimizing financial cost and user experience in edge cloud computing systems	Proposing a dynamic replica allocation strategy and the replica consistency preservation strategy	The suggested resource management method may lower the overall financial cost of the rented nodes and the SLA (Service Level Agreement) default rate while also improving CPU (Central Processing Unit) usage as time goes on
Tang et al. [50]	Proposing a new replica placement mechanism for mobile media streaming in edge computing	Implementing the greedy algorithm of the mechanism	The technique can minimize the customer's access time, the replica availability, and the server's load balancing

and edge computing. Therefore the novelty of our study is that this is the first paper to examine the role of replica selection simultaneously in both IoT and fog computing. So, in this study, we analyzed 19 articles (9 articles on the interface with replica selection and placement techniques on the IoT and 10 articles on replica selection and placement techniques on edge computing) using the systematic review method, and the results were being presented. Based on these studies, we identified some important challenges and provided solutions for future researchers to

Table 4 Comparison of the main parameters of the articles

Papers	Method	Decreasing network overhead	Decreasing cost	Increasing prediction accuracy	Increasing data availability	Increasing access response time	Decreasing bandwidth
Data replication and placement techniques in IoT							
He et al. [71]	IoT	No	Yes	No	No	No	No
Cho et al. [37]	IoT	No	No	Yes	No	Yes	No
Zhang et al. [72]	IoT	No	No	No	Yes	Yes	No
	Cloud computing						
Qaim and Ozkasap [74]	IoT	No	No	No	Yes	No	No
Shao et al. [20]	IoT	No	Yes	No	No	No	No
	Cloud computing						
	Edge computing						
Wakil et al. [70]	IoT	No	No	No	No	Yes	No
Wang et al. [75]	IoT	No	No	No	Yes	No	No
Wang and Batiha [9]	IoT	No	No	No	Yes	Yes	No
Liu et al. [76]	IoT	No	No	No	No	Yes	No
	Edge computing						
Data replication and placement techniques in edge computing							
Tziritas et al. [73]	Virtual machine	Yes	No	No	No	No	No
	Edge computing						
Aral and Ovatman [67]	Edge computing	No	Yes	No	No	No	No
Li et al. [30]	Cloud computing	No	No	Yes	No	No	No
	Edge computing						
Li et al. [39]	Edge computing	No	No	Yes	Yes	Yes	No
Bhalaji [11]	Mobile edge computing	No	No	No	No	Yes	Yes
Saranya et al. [35]	Mobile edge computing	No	No	No	No	Yes	Yes
Wang et al. [36]	Edge computing	No	No	No	No	Yes	Yes
	IoT						
Li et al. [77]	Cloud computing	No	No	No	No	Yes	No
	Edge computing						
Li et al. [69]	Cloud computing	No	Yes	No	No	No	No
	Edge computing						

Table 4 (continued)

Papers	Method	Decreasing network overhead	Decreasing cost	Increasing prediction accuracy	Increasing data availability	Increasing access response time	Decreasing bandwidth
Tang et al. [50]	Mobile edge computing	No	No	No	No	Yes	No

help researcher's select IoT models and edge computing. In the next section, we will examine the most important challenges in the literature and provide solutions for future research.

5 Open issue

In the following, the challenges of several studies will be explored so that future researchers can use the proposed algorithms for further research to eliminate the shortcomings in the literature.

We found that the most important challenges in data duplication are power consumption, response time, and consistency management. The battery's restriction is a key concern in data replication [57]. If the node has little power and numerous data items are accessed, it may become drained and cease functioning. The time it takes a customer to get data from a server is referred to as response time. If the server is located far away, the response time will be slower to serve customers. The replica synchronization is another difficulty in data replication. Others could develop a disability if the replica is changed often. Invalid accesses waste power and necessitate rollbacks, which may be a serious problem for nodes with limited resources [35].

Edge computing appears as a solution for supporting performance-critical IoT applications that resolves some of the challenges that conventional central data centers face. Nevertheless, resource allocation for diverse applications at a network edge of dispersed and resource-constrained nodes is difficult [58]. Although blockchain technology is offered to provide secure distributed networks, studying the possibility of combining it into the data replication process would be an exciting subject for further research [59, 60].

Mobile edge computing has emerged to manage the volume of data created, resulting in a delay in demand for IoT applications that are computationally intensive. Although mobile edge computing has progressed in terms of service latency and has been thoroughly studied, data use and security efficiency are not recognized. So, while cloud computing and other sources like edge computing are utilized to control dataflow congestion, the data security utilized is still under investigation [11].

Besides, data replication may be used for various reasons, including increased data availability and reliability. Nevertheless, handling these massive amounts of data from delay-sensitive IoT operations in collaborative edge and cloud settings remains a difficulty. One of the most important is how to put a large number of data replicas to decrease access costs while meeting a deadline. The nature-inspired algorithm can reduce data access latency, decrease bandwidth consumption, distribute storage site load, improve data access speed, and cause optimal congestion management [61, 62]. In addition, the convergence of edge and cloud computing has some advantages, including cloud's limitless shared storage and computational resources and edge's low-latency. Combining the two computing paradigms can let cooperative user groups develop IoT processes quickly and at a reasonable cost [20].

How to defend against replica attacks is one of the most difficult challenges in IoT. When a device is physically taken by an attacker or infected with malicious code, it may be reprogrammed. The secret information can be replicated into many replicas that fraudulently occupy the network and gather subscribers' private information without their knowledge [37]. As a result, selecting the appropriate replicas can improve reaction time while lowering costs [9].

Another important consideration in data replication is the accuracy of the data (also known as data consistency) in the systems. Failure of connectivity between two mobile nodes may lead to the network being partitioned into isolated sub-networks, resulting in the possibility of data inconsistency. Furthermore, several nodes may respond slowly, making it look like the network is partitioned when it is not [35].

Traditional techniques for solving NP-hard problems are inefficient, while evolutionary algorithms like ACO and GA appear extremely effective. For instance, in several current approaches for selecting a server in IoT contexts, load balancing has been deemed a key challenge. Because

the IACO has successfully resolved load balancing difficulties, it may address this issue [9]. Hence, investigating the performance of some new meta-heuristic algorithms using fuzzy logic [63, 64] and neural networks [65, 66] for solving this problem can be done in future research.

Bringing processing power to the network's edge decreases latency and allows for cloud code offloading. Some services, on the other hand, require access to data that is held centrally. As a result, data access latency can become a bottleneck, negating the advantages of edge computing, particularly for data-intensive services. The volume of data absorbed, disseminated, and processed in cloud systems is always increasing, necessitating intelligent data distribution techniques [67, 68]. Furthermore, due to its restricted capacity, the edge cloud computing system still confronts several problems in resolving the overload issue [69]. There are also several issues with file sharing across edge nodes and the files given by cloud servers.

Wakil, Nazif [70] used a hybrid optimization technique to investigate a strategy for replica selection in the IoT. The ACO is utilized to produce variety in the suggested technique, and then the GA is employed to give a comprehensive search over the search space. In terms of waiting time, the suggested technique outperformed ACO, the Response Time-based Replica Management, and the High-QoS First-Replication (HQFR) algorithms. Furthermore, according to the findings, the hybrid technique outperformed HQFR and the dynamic cost-aware re-replication and re-balancing strategy. Nevertheless, because of the suggested method's hybrid character, it takes longer to run. The upcoming study could look into the performance of some novel meta-heuristic algorithms to solve the replica selection issue in the IoT, like shark smell optimization, PSO, bee colony algorithm, and world cup optimization algorithm using a combination of fuzzy logic and neural networks. Also, Shao, Li [20] suggested a new data replica placement technique for data-intensive IoT operations in a combined edge and cloud environment. Initially, the placement of data replicas may be described as a 0–1 integer programming issue that takes into account total data dependence, data reliability, and user collaboration. The IoT method, a version of intelligent swarm optimization, is then provided as a solution to this model. The suggested approach outperforms these comparable algorithms, according to the findings of the experiments. It can discover a higher-quality data replica placement solution and requires a smaller computational budget than conventional methods. The scheduler and cluster resource management, on the other hand, are not taken into account. The suggested framework should be taken into account to integrate the scheduler and cluster resource management. Some aspects of the suggested models, such as data distribution and replication dependencies, should be addressed.

Ultimately, Wang and Batiha [9] used an IACO since the replica selection issue is NP-hard. Ants flip the fundamental logic of ACO by converting the influence of pheromone on the selected route. Because multiple IoT centers exist, the IACO has been used to choose replicated data in the IoT while load balancing across IoT centers is considered. An ant selects the optimal place for its travel in this technique, and others are not allowed to traverse the trail that the preceding ants have passed. According to the results, the method outperformed the ACO, HQFR, and RTRM methods regarding waiting time and load balancing. The suggested model, on the other hand, takes longer to run. Future challenges include linking the edges in the graph representation and considering the circumstances of IoT centers to create the graph better. Future research might improve this method by looking at additional QoS parameters. Future research can integrate it with various meta-heuristic methods to improve its performance. Ultimately, research can offer a strategy for optimally arranging replicated data to decrease overall time spent selecting services in the IoT.

6 Conclusion

On IoT and edge computing, this research offers a complete assessment of replica selection and deployment strategies. To do this, we first offer an overview of the IoT, replica selection, edge computing, and its capabilities and structure. Then we discussed articles related to replica selection on the IoT and edge computing. The results of the studies were then collected. Ultimately, open research difficulties and constraints in edge computing, IoT, replica selection, and alternative research approaches are discussed thoroughly.

The IoT follows specific requirements for inter-gadget and Internet interactions, provides real-time access to gadget data, and allows gadget management from afar. Edge computing is a promising concept to enable the IoT vision, especially for supporting time-sensitive applications. Edge or fog computing, in which edge nodes voluntarily donate computing resources, has gotten a lot of press for its ability to solve scalability issues and relieve the load on cloud computing servers in typical cloud-centric IoT architectures. The basic concept is to place smaller data centers at the backbone's edge in order to provide cloud computing resources closer to end-users. As a result, edge computing infrastructures are on the horizon, and academics and businesses alike should update their present cloud services to accommodate the unique requirements of such massively dispersed infrastructures.

The results also show that edge computing can support low-traffic, high-energy, low-bandwidth, latency-sensitive

service requests to reduce the pressure on cloud data centers. Edge computing is not a substitute for the cloud; it extends computing, communications, and storage capabilities from cloud to edge networks. Although edge computing is a viable solution for the sustainable development of the IoT market, many unresolved issues still exist.

Like any other article, this article has some limitations. One of the limitations is that non-English articles such as Chinese and Japanese were removed from the report. Therefore, future researchers can study the subject of this article in other languages, as well.

Data availability All data are reported in the paper.

References

1. Rossini, R., et al. (2020). REPLICA: A solution for next generation IoT and digital twin based fault diagnosis and predictive maintenance.
2. Aslani, R., & Rasti, M. (2020). A distributed power control algorithm for energy efficiency maximization in wireless cellular networks. *IEEE Wireless Communications Letters, 9*(11), 1975–1979.
3. Aslani, R., Saberinia, E., & Rasti, M. (2020). Resource allocation for cellular V2X networks mode-3 with underlay approach in LTE-V standard. *IEEE Transactions on Vehicular Technology, 69*(8), 8601–8612.
4. Jahantigh, M. N., et al. (2019). Integration of internet of things and cloud computing: A systematic survey. *IET Communications, 14*(2), 165–176.
5. Li, Y., & Lee, I. (2021). An 111pW Voltage Reference with a Diode-Leakage-Decoupling Replica for High-Temperature Miniature IoT Systems. In: *2021 IEEE custom integrated circuits conference (CICC)*, IEEE.
6. Sefati, S.S. & Navimipour, N.J. (2021) A QoS-aware service composition mechanism in the Internet of things using a hidden Markov model-based optimization algorithm. *IEEE Internet of Things Journal*.
7. Heidari, A., et al. (2020). Internet of things offloading: Ongoing issues, opportunities, and future challenges. *International Journal of Communication Systems, 33*(14), e4474.
8. Hajiheidari, S., et al. (2019). Intrusion detection systems in the Internet of things: A comprehensive investigation. *Computer Networks, 160*, 165–191.
9. Wang, S., & Batiha, K. (2020). A metaheuristic-based method for replica selection in the Internet of Things. *International Journal of Communication Systems, 33*(14), e4458.
10. Almusaylim, Z. A., & Zaman, N. (2019). A review on smart home present state and challenges: Linked to context-awareness internet of things (IoT). *Wireless Networks, 25*(6), 3193–3204.
11. Bhalaji, N. (2020). Efficient and secure data utilization in mobile edge computing by data replication. *Journal of ISMAC, 2*(01), 1–12.
12. Azad, P., et al. (2019). The role of structured and unstructured data managing mechanisms in the Internet of things. *Cluster Computing, 23*, 1185–1198.
13. Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics, 10*(4), 2233–2243.
14. Dehkordi, S. A., et al. (2020). A survey on data aggregation techniques in IoT sensor networks. *Wireless Networks, 26*(2), 1243–1263.
15. Sethi, P., & Sarangi, S. R. (2017). Internet of things: architectures, protocols, and applications. *Journal of Electrical and Computer Engineering, 2017*, 1–25.
16. Vahdat, S. (2020). The role of IT-based technologies on the management of human resources in the COVID-19 era. *Kybernetes*.
17. Roshani, M., et al. (2020). Proposing a gamma radiation based intelligent system for simultaneous analyzing and detecting type and amount of petroleum by-products. *Nuclear Engineering Technology, 53*, 1277–1283.
18. Heidari, A., & Navimipour, N. J. (2021). A new SLA-aware method for discovering the cloud services using an improved nature-inspired optimization algorithm. *PeerJ Computer Science, 7*, e539.
19. Hamzei, M., & Navimipour, N. J. (2018). Toward efficient service composition techniques in the internet of things. *IEEE Internet of Things Journal, 5*(5), 3774–3787.
20. Shao, Y., Li, C., & Tang, H. (2019). A data replica placement strategy for IoT workflows in collaborative edge and cloud environments. *Computer Networks, 148*, 46–59.
21. Hu, W., et al. (2016) Quantifying the impact of edge computing on mobile applications. In *Proceedings of the 7th ACM SIGOPS asia-pacific workshop on systems*.
22. Zhang, X., Cao, Z., & Dong, W. (2020). Overview of edge computing in the agricultural internet of things: Key technologies applications challenges. *IEEE Access, 8*, 141748–141761.
23. Zhu, C., et al. (2014). A novel sensory data processing framework to integrate sensor networks with mobile cloud. *IEEE Systems Journal, 10*(3), 1125–1136.
24. Guo, K., et al. (2018). Transparent learning: An incremental machine learning framework based on transparent computing. *IEEE Network, 32*(1), 146–151.
25. Zhao, Z., et al. (2018). Deploying edge computing nodes for large-scale IoT: A diversity aware approach. *IEEE Internet of Things Journal, 5*(5), 3606–3614.
26. Li, C., et al. (2019). Collaborative cache allocation and task scheduling for data-intensive applications in edge computing environment. *Future Generation Computer Systems, 95*, 249–264.
27. Haghi Kashani, M., Rahmani, A. M., & Jafari Navimipour, N. (2020). Quality of service-aware approaches in fog computing. *International Journal of Communication Systems, 33*(8), e4340.
28. Hassan, N., et al. (2018). The role of edge computing in internet of things. *IEEE Communications Magazine, 56*(11), 110–115.
29. Sun, X., et al. (2021). Data replication techniques in the Internet of Things: a systematic literature review. *Library Hi Tech*.
30. Li, C., Tang, J., & Luo, Y. (2019). Scalable replica selection based on node service capability for improving data access performance in edge computing environment. *The Journal of Supercomputing, 75*(11), 7209–7243.
31. GAO, H., et al. (2012). Replica location algorithm approach based on the small world in the Internet of Things. Journal of Sichuan University of Science and Engineering (Natural Science Edition) p. 02.
32. Ma, K., & Yang, B. (2017). Stream-based live data replication approach of in-memory cache. *Concurrency and Computation Practice and Experience, 29*(11), e4052.
33. Li, B., et al. (2012) Energy-aware replica selection for data-intensive services in cloud. In: *2012 IEEE 20th international symposium on modeling, analysis and simulation of computer and telecommunication systems*, IEEE.
34. Yu, W., et al. (2017). A survey on the edge computing for the Internet of Things. *IEEE Access, 6*, 6900–6919.

35. Saranya, N., Geetha, K., & Rajan, C. (2020). Data replication in mobile edge computing systems to reduce latency in Internet of things. *Wireless Personal Communications, 112*(4), 2643–2662.

36. Wang, C., Gill, C., & Lu, C. (2020) Adaptive Data Replication in Real-Time Reliable Edge Computing for Internet of Things. In: *2020 IEEE/ACM fifth international conference on internet-of-things design and implementation (IoTDI)*, IEEE.

37. Cho, K., et al. (2013). Energy-efficient replica detection for resource-limited mobile devices in the internet of things. *IET Communications, 7*(18), 2141–2150.

38. Taleb, T., et al. (2017). On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. *IEEE Communications Surveys and Tutorials, 19*(3), 1657–1681.

39. Li, C., et al. (2019). Flexible replica placement for enhancing the availability in edge computing environment. *Computer Communications, 146*, 1–14.

40. Wu, Z., et al. (2020). How to protect reader lending privacy under a cloud environment: A technical method. *Library Hi Tech*.

41. Heidari, A., & Navimipour, N.J. (2021) Service discovery mechanisms in cloud computing: A comprehensive and systematic literature review. *Kybernetes*.

42. Vahdat, S. & Shahidi, S. (2020). India section b: Biological sciences, D-dimer levels in chronic kidney illness: A comprehensive and systematic literature review. pp. 1–18.

43. Zou, N., Liang, S., & He, D. (2020). Issues and challenges of user and data interaction in healthcare-related IoT: A systematic review. *Library Hi Tech, 38*, 769–782.

44. Karatas, F., & Korpeoglu, I. (2019). Fog-based data distribution service (F-DAD) for Internet of Things (IoT) applications. *Future Generation Computer Systems, 93*, 156–169.

45. Qi, B., Kang, L., & Banerjee, S. (2017). A vehicle-based edge computing platform for transit and human mobility analytics. In: *Proceedings of the second ACM/IEEE symposium on edge computing*.

46. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications, 19*(2), 171–209.

47. Armbrust, M., et al. (2009). Above the clouds: A berkeley view of cloud computing, 2009, Technical Report UCB/EECS-2009-28, EECS Department, University of California.

48. Rahimi, M., et al. (2021). Toward the efficient service selection approaches in cloud computing. *Kybernetes*.

49. Yu, Z., et al. (2021). Systematic literature review on the security challenges of blockchain in IoT-based smart cities. *Kybernetes*.

50. Tang, Y., et al. (2021). A new replica placement mechanism for mobile media streaming in edge computing. *Concurrency and Computation Practice and Experience, 33*(7), 1–1.

51. Epifânio, L.M. (2020). *Replica Placement in Edge Computing*. Instituto Superior Técnico.

52. Shi, W., et al. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646.

53. Mohiuddin, M., et al. (2017). Axo: Detection and recovery for delay and crash faults in real-time control systems. *IEEE Transactions on Industrial Informatics, 14*(7), 3065–3075.

54. Alwarafy, A., et al. (2020). A survey on security and privacy issues in edge computing-assisted internet of things. *IEEE Internet of Things Journal, 2020*(8), 4004–4022.

55. Taneja, M., & Davy, A. (2017). Resource aware placement of IoT application modules in Fog-Cloud Computing Paradigm. In: *2017 IFIP/IEEE symposium on integrated network and service management (IM)*, IEEE.

56. Sadrishojaei, M., et al. (2021). A new preventive routing method based on clustering and location prediction in the mobile Internet of Things. *IEEE Internet of Things Journal, 8*, 10652–10664.

57. Mirzapour, F., et al. (2019). A new prediction model of battery and wind-solar output in hybrid power system. *Journal of Ambient Intelligence and Humanized Computing, 10*(1), 77–87.

58. Maia, A.M., et al. (2019). A multi-objective service placement and load distribution in edge computing. In: *2019 IEEE global communications conference (GLOBECOM)*, IEEE.

59. Dehghani, M., et al. (2021). Blockchain-based securing of data exchange in a power transmission system considering congestion management and social welfare. *Sustainability, 13*(1), 90.

60. Dehghani, M., et al. (2020). Blockchain-based securing of data exchange in a power transmission system considering congestion management and social welfare. *Sustainability, 13*(1), 1–1.

61. Chen, L., & Yang, P. (2008). Application research of ant algorithm to replica selection in data grid. *Computer Engineering and Design, 23*(29), 6157–6160.

62. Jalili, A., Firouz, M. H., & Ghadimi, N. (2015). Firefly algorithm based on fuzzy mechanism for optimal congestion management. *UCT Journal of Research in Science Engineering and Technology, 3*(3), 1–7.

63. Khodaei, H., et al. (2018). Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming. *Applied Thermal Engineering, 137*, 395–405.

64. Ghadimi, N. (2015). A new hybrid algorithm based on optimal fuzzy controller in multimachine power system. *Complexity, 21*(1), 78–93.

65. Roshani, M., et al. (2021). Combination of X-ray tube and GMDH neural network as a nondestructive and potential technique for measuring characteristics of gas-oil–water three phase flows. *Measurement, 168*, 108427.

66. Roshani, M., et al. (2020). Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Journal of Flow Measurement and Instrumentation, 75*, 101804.

67. Aral, A., & Ovatman, T. (2018). A decentralized replica placement algorithm for edge computing. *IEEE Transactions on Network and Service Management, 15*(2), 516–529.

68. Lu, M., et al. (2020). Smart load scheduling strategy utilising optimal charging of electric vehicles in power grids based on an optimisation algorithm. *IET Smart Grid, 3*, 914–923.

69. Li, C., et al. (2020). Resource and replica management strategy for optimizing financial cost and user experience in edge cloud computing system. *Information Sciences, 516*, 33–55.

70. Wakil, K., et al. (2019). Method for replica selection in the Internet of Things using a hybrid optimisation algorithm. *IET Communications, 13*(17), 2820–2826.

71. He, D., Wu, M., & Hu, C.-H. (2012). Load-balancing and low cost cloud data replica distribution method in Internet of Things environment [J]. *Journal of Central South University (Science and Technology), 43*(4), 1355–1361.

72. Zhang, B., Wang, X., & Huang, M. (2014). A data replica placement scheme for cloud storage under healthcare IoT environment. In: *2014 11th international conference on fuzzy systems and knowledge discovery (FSKD)*, IEEE.

73. Tziritas, N., et al. (2017). Data replication and virtual machine migrations to mitigate network overhead in edge computing systems. *IEEE Transactions on Sustainable Computing, 2*(4), 320–332.

74. Qaim, W.B. & Ozkasap, O. (2018). DRAW: Data replication for enhanced data availability in IoT-based sensor systems. In: *2018 IEEE 16th international conference on dependable, autonomic and secure computing, 16th international conference on pervasive intelligence and computing, 4th international conference on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech)*, IEEE.

75. Wang, S., et al. (2019). Active data replica recovery for quality-assurance Big Data analysis in IC-IoT. *IEEE Access, 7*, 106997–107005.

76. Liu, R., et al. (2020). Edge node data replica management method for distribution Internet of Things. In: *2020 4th international conference on HVDC (HVDC)*, IEEE.

77. Li, C., et al. (2020). Effective replica management for improving reliability and availability in edge-cloud computing environment. *Journal of Parallel and Distributed Computing, 143*, 107–128.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Heng Li is an information system project manager and senior engineer working on R & D Center in Guangdong Ruigong Intelligent Technology Co., Ltd. Her research interests include Information system development and management.

Zhong-Liang Shao is an associate professor working on Department of Automation Engineering in Guangdong Polytechnic of Water Resources and Electric EngineeringChinaHis research interests include artificial intelligence, edge computing, control theory, quantum communicationmore than 11 papers published.

Cheng Huang is an associate professor working on Department of Mechanical Engineering in Guangdong Polytechnic of Water Resources and Electric EngineeringChinaHis research interests include mechanical and electrical engineering, Internet of things technologymore than 6 papers published.