Array 19 (2023) 100293

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Differential privacy in edge computing-based smart city Applications: Security issues, solutions and future directions

Aiting Yao [a], Gang Li [b], Xuejun Li [a,*], Frank Jiang [b], Jia Xu [a], Xiao Liu [b,**]

[a] School of Computer Science and Technology, Anhui University, Hefei, China
[b] School of Information Technology, Deakin University, Geelong, Australia

ARTICLE INFO

Keywords:
Differential privacy
Edge computing
Privacy-preserving
Smart city
Smart IoT systems

ABSTRACT

Fast-growing smart city applications, such as smart delivery, smart community, and smart health, are generating big data that are widely distributed on the internet. IoT (Internet of Things) systems are at the centre of smart city applications, as traditional cloud computing is insufficient for satisfying the critical requirements of smart IoT systems. Due to the nature of smart city applications, massive IoT data may contain sensitive information; hence, various privacy-preserving methods, such as anonymity, federated learning, and homomorphic encryption, have been utilised over the years. Furthermore, limited concern has been given to the resource consumption for data privacy-preserving in edge computing environments, which are resource-constrained when compared with cloud data centres. In particular, differential privacy (DP) has been an effective privacy-preserving method in the edge computing environment. However, there is no dedicated study on DP technology with a focus on smart city applications in the edge computing environment.

To fill this gap, this paper provides a comprehensive study on DP in edge computing-based smart city applications, covering various aspects, such as privacy models, research methods, mechanisms, and applications. Our study focuses on five areas of data privacy, including data transmitting privacy, data processing privacy, data model training privacy, data publishing privacy, and location privacy. In addition, we investigate many potential applications of DP in smart city application scenarios. Finally, future directions of DP in edge computing are envisaged. We hope this study can be a useful roadmap for researchers and practitioners in edge computing enable smart city applications.

1. Introduction

The term "smart city" was initially introduced in 2012 [1], with the goal of leveraging information and communication technologies to enhance the functioning of urban areas. Over the past decade, the development and implementation of smart city initiatives have proliferated, as evidenced by the rapid growth of smart home systems [2], smart healthcare systems [3,4], smart manufacturing service systems [5], intelligent transportation systems [6,7], and other similar projects. Meanwhile, advanced computing technologies are required to address issues, such as processing massive volume of data produced by the large number of IoT (Internet of Things) devices accessing the network. While cloud computing based on large-scale centralised server clusters has enabled the large-scale commercial use of the internet, enterprise IT, and smartphones, it has been shown to be insufficient for meeting the requirement of rapidly growing smart IoT systems due to issues, such as high response delay or limited bandwidth on the terminal side. In recent years, edge computing has emerged as a promising computing paradigm for smart IoT systems [8,9]. Table 1 summarizes the major factors that make edge computing an ideal computing infrastructure for smart IoT systems. Specifically, edge computing can effectively reduce the delay of the computing system and data transmission bandwidth, alleviating the pressure on centralised cloud data centres [10].

Edge computing offers the potential to offload some computation tasks of smart city applications from IoT devices to the edge of the network where abundant resources are available [11–14]. However, given the nature of smart city applications, many computation tasks may involve sensitive information and individual users' private data, making

* Corresponding author.
** Corresponding author.
E-mail addresses: yaoat@foxmail.com (A. Yao), gang.li@deakin.edu.au (G. Li), xjli@ahu.edu.cn (X. Li), frank.jiang@deakin.edu.au (F. Jiang), xujia@ahu.edu.cn (J. Xu), xiao.liu@deakin.edu.au (X. Liu).

https://doi.org/10.1016/j.array.2023.100293
Received 20 January 2023; Received in revised form 13 May 2023; Accepted 14 May 2023
Available online 31 May 2023

Table 1
Major factors of choosing edge computing for smart IoT systems.

Factors	Description
Capacity	To transmit more and more data generated by extensive connected devices to the cloud service with a centralised location, it needs super bandwidth and return capacity. However, edge computing and local data processing can reduce the amount of data to be transmitted.
Cost	There are costs associated with transmitting large amounts of data over long distances. In addition, huge amount of data generated by many devices may not be related to business, so it does not need to be transmitted to the central processing centre.
Analysis	Data are the basic asset of the digital economy. Edge computing technology having the ability to convert data into real-time (or near real-time) for analysis and operation.
Security	Many companies and users may not want sensitive data to leave the field or their servers.
Delay	Although 5G has a lower delay than 4G, it is difficult to achieve very low delay in long-distance and multi-hop networks.
Elastic	Edge computing can provide more communication paths than a centralised mode. This kind of distribution can better guarantee the flexibility of data communication.

Table 2
The categories of differential privacy study.

Categories	Contents	Years	Ref.
Network	Social Network Analysis	2021	[34]
	Cyber Physical Systems	2019	[35]
Machine Learning	Utility and Private	2020	[36]
		2014	[37]
Big Data	Cryptography	2021	[38]
	Data Release	2012	[39]
	Decision Tree	2019	[40]
	Data Aggregation	2018	[41]
	Location Pattern Mining and Health Data	2020	[42]
		2013	[43]
	Statistical Estimators	2010	[44]
		2008	[45]
	Utility and Private	2021	[38]
		2016	[46]
		2014	[47]
	Data Mining	2020	[36]
		2017	[47]
		2014	[48]
Other	Shuffle Model	2021	[49]

data privacy a crucial concern for edge computing. For example, a study analyzing friendship association accurately predicted the sexual orientation of Facebook users by examining 4080 Facebook profiles from the MIT network [15]. Additionally, cybercriminals may exploit users' private information for targeted social engineering attacks [16]. As such, a data privacy protection solution is essential for a secure smart IoT system [12,17].

In recent years, with the emergence of edge computing technology, many researchers have started to investigate the privacy protection issue for edge computing. Existing technologies on privacy protection in edge computing mainly focus on DP [18–20], homomorphic encryption [21,22], secure multiparty computing [23], and verifiability and auditability [24]. In this work, we focus on DP applications in five specific areas, including data transmission, data processing, data model training, data publishing, and location privacy, for smart city application scenarios based on edge computing. The reason for this is that the first four areas are typical data management stages, and location privacy is the most representative privacy issue in edge computing-based smart city applications.

In 2006, Dwork [25] proposed a privacy-preserving mathematical model to prevent differential attacks. The definition of DP does not depend on the attackers' background knowledge, and the technology is extensively applied in machine learning [26–28], data mining [29,30], deep learning [31–33], etc. The prominent advantages of DP are mainly the following three aspects. First, the attackers' background knowledge does not need to be considered. Second, the technology contains rigorous proof and a quantitative expression for privacy breach risk. Third, the dataset privacy is protected to a great extent by adding a few bits of noise, that is, the volume of noise added has nothing to do with the data sets' size. Currently, existing studies/reviews related to DP can be classified into four categories as summarised in Table 2. To the best of our knowledge, this paper is the first study on DP technology in smart city application scenarios based on edge computing.

In summary, this study has made the following major contributions.

- We comprehensively review the existing studies on the application of DP methods in edge computing-based smart city applications.
- We summarise and classify the security and privacy issues for edge computing-based smart city applications.
- We propose important future research directions of DP for edge computing-based smart city applications.

The rest of this paper is organised as follows. The research methodology for this paper is described in Section 2, and the main contents include the question description, data exploration, and article selection. Section 3 presents the basic concepts in this study, including the definition of edge computing, the definition of DP, and its mechanism. In Section 4, we enumerate the typical privacy challenges for edge computing-based smart city applications. Section 5 reviews the privacy-preserving technology and the advantages of DP in the environment of edge computing. In section 6 we summarise the application and implementation of DP for the smart city applications for edge computing. Section 7 envisages future directions, including anomaly detection and data attack defence. Finally, Section 8 concludes the paper.

2. Research methodology

In this section, we present the method for searching appropriate articles on DP for the smart city applications based on edge computing, including index content description, data exploration, and article selection.

2.1. Index content description

This study aims to explore the features and methods in the research articles with the main challenges associated with data privacy in the edge computing environment and smart city applications. Specifically, we focus on 5 index contents in this study, as shown in Table 3.

2.2. Data exploring and article selection

Articles pertinent to DP in the edge computing environment or smart city applications can be explored in mainstream academic databases, such as, for example, Google Scholar, the Institute of Electrical and Electronics Engineers (IEEE) Xplore, Springer Link, ACM Digital Library, etc. Fig. 1 shows the detailed distribution of the total 1397 articles found between 2014 and May 2023. As depicted in the figure, the majority of articles were published between 2019 and 2023, with 2023 having the highest proportion of published articles.

In order to identify the most relevant articles for our study, we developed a selection criteria and evaluation framework, which is depicted in Fig. 2. Firstly, we considered all articles related to smart city and edge computing published between 2019 and 2023. Secondly, we examined the articles with respect to their relevance to data privacy and security, and excluded those that were not relevant. This led to a set of 228 articles related to DP out of the initial 336 articles. Within this set, 94 articles focused on LDP, while the remaining 134 articles focused on DP. We note that while we primarily discuss the 228 articles related to DP in our analysis, we also provide a comprehensive overview by discussing some of the articles that were removed during the selection process. Fig. 3 illustrates the distribution of relevant articles on smart

Table 3
Index contents.

Index Contents	Brief Description
IC1: The typical privacy issues and challenges in smart city applications for the edge computing.	The typical privacy issues could be classified into network, data, infrastructure, and application. The index content is detailed in the description in Section 4.
IC2: The privacy-preserving techniques, and the advantage of the differential privacy method applied in the edge computing environment based smart IoT system.	Reviewed articles considered differential techniques to achieve the privacy protection mechanism. The existing privacy protection methods contain DP, secure multipart computing, private information retrieval, etc. The advantages of DP are enhancing data availability, reducing the computation complexity, etc. The index content is detailed in a description in Section 5.
IC3: The adoptions of the edge computing environment are considered in the differential privacy method research.	Some existing methods are general-purpose, but some methods are particularly suitable for specific types of applications. This relevant information is available to researchers and practitioners. The index content is detailed in a description in Section 6.
IC4: The experiment platforms/tools, datasets, and evaluated metrics are applied to implement the DP mechanism or algorithm.	The reviewed articles might apply various experimental platforms/tools to implement the DP mechanism or algorithm. Their datasets and evaluation metrics are summarised. The index content is detailed in a description in Section 6.
IC5: The future research directions of DP methods in the edge computing environment.	The directions for future research of DP methods for the smart city applications based on edge computing include data anomaly detection, data attack defence, and others. The index content is detailed in a description in Section 7.

city and edge computing, classified into five categories: thesis, conference, study/review, editorial material, and online publishing/Webpages. The highest numbers of papers were published from 2019 to 2023, followed by conference papers. Fig. 4 shows the distribution of the 228 selected articles among DP and LDP. The figure indicates that the number of articles on DP is always higher than that on LDP and that 2023 had the largest number of articles.

3. Background

In Section 3, the definition of edge computing and some related terminologies are introduced; then, we introduce the definition and mechanisms of differential privacy.

3.1. Definition of edge computing

The formal definition of edge computing was first described by the European Telecommunications Standards Institute (ETSI).

Definition 1. (Edge Computing) [50] Edge computing affords an environment of IT service and cloud computability at the mobile network edge of the radio access network (RAN) and approximate to mobile clients.

Fig. 5 shown is the full edge computing architecture which is consisted of edge devices, edge network, edge computing centre, and a core infrastructure. Edge devices, for instance, mobile phones, computers, and servers are responsible for receiving the instruction from and reporting data back to the smart gateway. The interconnection of IoT devices and accelerometers are realised by fusing several communication networks. The edge computing centre provides calculation, storage, and network forwarding resources. The core infrastructure affords computing services and management functions for mobile edge devices besides a core network. The core network mainly includes the internet, mobile core network, centralised cloud service, and data centre.

Edge computing is a new and innovative computing exemplification. Edge computing uses an open framework that provisions resources to the end devices from the network edge, which integrates network, computing, and storage and application services. Hence, edge computing can play a substantial role in smart city applications, for instance the deployment of network video cameras, the construction of intelligent sensing system platforms, data acquisition, transmission, and processing. To provide readers with a better understand the concept of edge computing, we summarise some representative terminologies [51–54] in Table 4. The related terminologies are often applied in smart city scenarios.

3.2. Differential privacy

DP is a technique that enables accurate data queries from a statistical database while minimizing the risk of identifying individual records [25]. Unlike other privacy protection methods, DP focuses on safeguarding the privacy of each individual record rather than the entire dataset [55]. Currently, DP research is primarily focused on protecting location privacy [20,56–58], data analysis [59–61], developing mechanisms [62–64], and exploring the application of federated learning [65–67]. For example, Miao et al. [56] devised a DP algorithm with the quad-tree based on Hilbert curve division to reduce the limitation of computing resources in the edge node. They built a noise query tree, and segmenting the privacy budget based on the noise tree. The retrieval

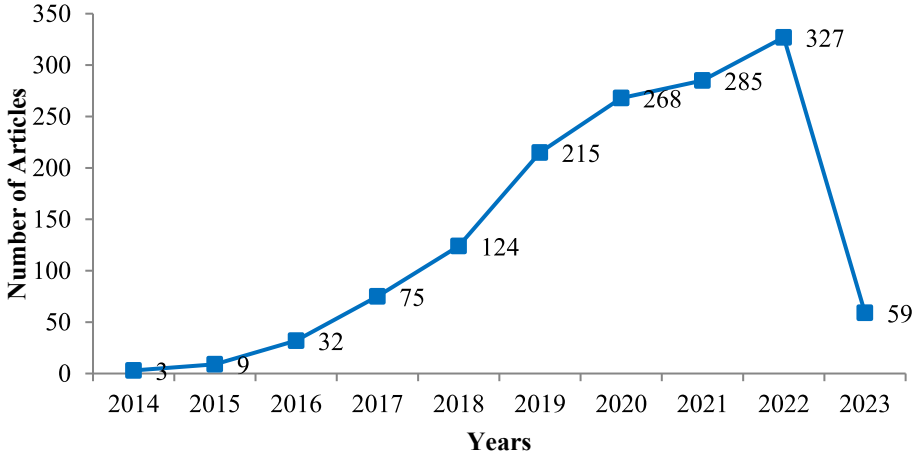

Fig. 1. Distribution of published articles with smart city and edge computing by year.

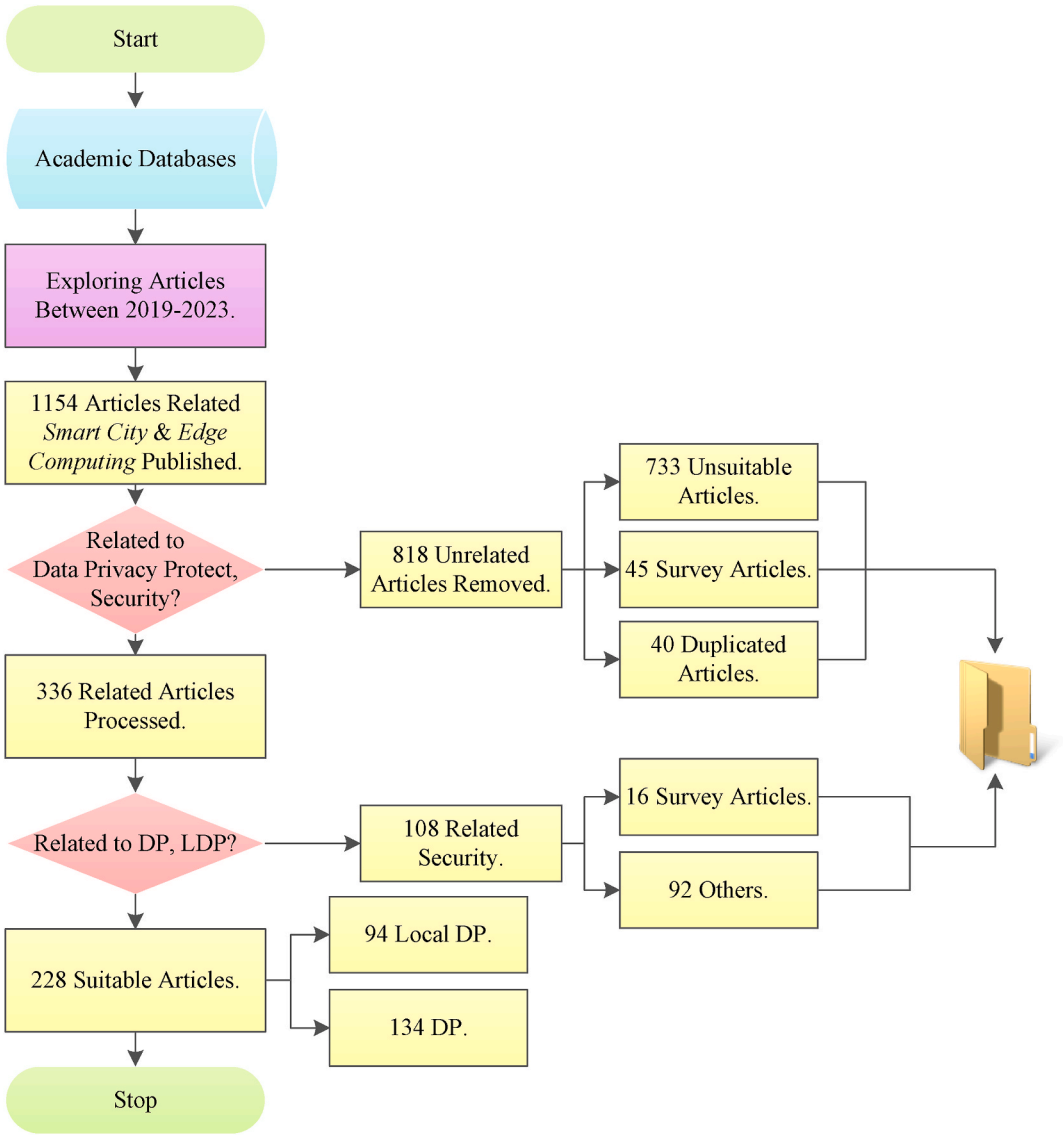

Fig. 2. Selection criteria and evaluation framework.

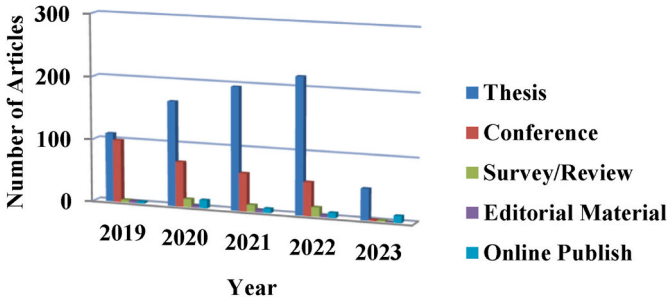

Fig. 3. The distribution diagram of article type by year.

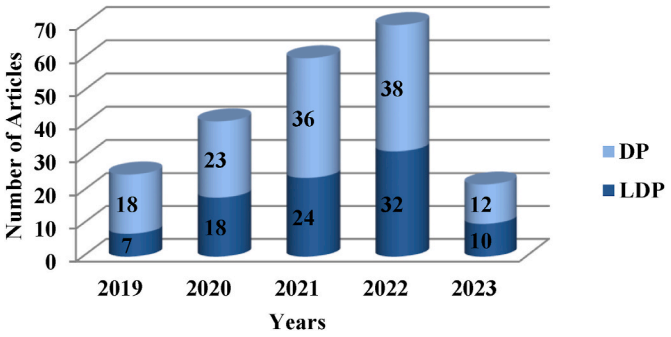

Fig. 4. The distribution diagram of articles with DP and LDP by year.

efficiency was greatly improved by the Hilbert curve. Nie et al. [59] proposed a DP tensor computing model to solve data analysis security problem in SDN-based IoT. In the tensor computing model, DP is used to protect data privacy transmitted to the cloud layer. Zhang et al. [20] designed a lightweight DP protection mechanism in order to solve the privacy leakage issue of location model training. They extended ε-DP theory to mature machine learning localisation technology, implemented privacy protection in training localisation model. In a highly untrustworthy environment Guo et al. [62] proposed a mechanism named online multi-item double auction (MIDA) to solve the method of allocating finite edge servers to IoT devices They improved the MIDA mechanism on the basis of DP to protect sensitive information from being compromised. Zhang et al. [65] put forward a federated learning framework supported by mobile edge computing that integrates the model partition technique and DP simultaneously. They reduced the

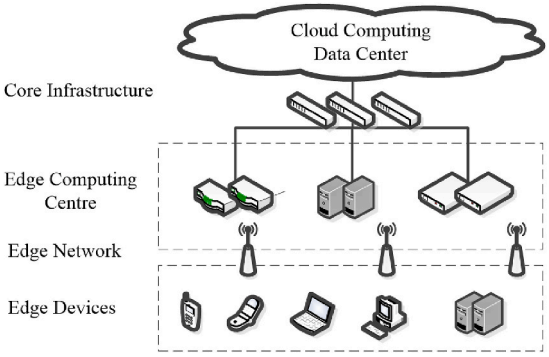

Fig. 5. The Composition of edge computing.

Table 4
The terminologies related to edge computing.

Terms	Descriptions
Application Rules and Requirements	Rules and requirements related to various edge applications, including required resources, required or useful services, maximum delay/delay, DNS (domain name system) rules, traffic rules, mobility support, etc.
Edge Applications	In the mobile edge system, edge applications have the potential to afford or consume edge services and can be instantiated on the edge host.
Edge Host	It includes an entity of edge platform and virtualisation infrastructure to afford computing resources, storage resources, and network resources for various mobile edge applications.
Edge Host Level Management	It performs the management of specific edge functions of specific edge platforms, edge hosts, and various edge applications running on them.
Edge Mainframe	It includes two level management which is edge system and edge host.
Edge Platform	It can provide and consume a variety of edge services and provide a variety of edge services for itself. It can run all kinds of edge applications on a specific edge host virtual infrastructure.
Edge Service	It can provide services according to the edge platform or application.
Edge System	The set of edge hosts and edge management entities is needed to run various kinds of edge applications in the operator network or subnet.
Edge System Level Management	A management entity with a global view of the entire edge system.
NFV (Network Function Virtualisation)	Through the abstraction of virtual hardware, the network function is decoupled from the hardware equipment needed for its operation.
UE Application	The related applications are run by the user terminal and have the ability to interact with the mobile edge system through the user application lifecycle management agent.
User Application	It is instantiated in the edge system to respond to the user's request through an application operating on the user terminal.
User Terminal Equipment	The mobile terminal equipment is used to access the basic network of mobile communication and run various applications that can transmit IP packets through the basic network of mobile communication.
Virtualised Resource	Computing resources, storage resources, and network resources are provided by the underlying virtual infrastructure for each upper mobile edge application.

heavy computational cost of deep neural network training on edge devices and provided strong privacy guarantees. In this framework, the updated edge data from a device to server are protected through the DP method perturbation.

3.2.1. DP-related definitions

In this paper, we consider $\mathscr{X}$ with size $|\mathscr{X}|$ is a limited data universe. D and $D^{'}$ named neighbouring datasets if $|D\Delta D^{'}| \leq 1$, where the two datasets are form $\mathscr{X}$ with unordered and finite. That's mean the neighbour dataset differing in one element for two datasets. In addition, a query stream is denoted as F and there is m queries ($f_i \in F, i = 1, 2, \ldots, m$), where query function f is a map D to the range $\mathscr{R}$:

$$f : D \rightarrow \mathscr{R}$$

The purpose of DP is to make attackers less certain about whether an individual is in the dataset after observing the results of a query f.

Definition 2. (Differential Privacy) [25]: For any two neighbour datasets D and $D^{'}$, a function with randomised algorithm $\mathscr{M}$ guarantees differential privacy for arbitrarily output subset $\mathscr{S} \subseteq Range(\mathscr{M})$ if $\mathscr{M}$ satisfies:

$$\Pr[\mathscr{M}(D) \in \mathscr{S}] \leq e^{\varepsilon} \times \Pr[\mathscr{M}(D^{'}) \in \mathscr{S}] + \delta$$

where the range for resultant of randomised algorithm $\mathscr{M}$ denoted $Range(\mathscr{M})$ and a privacy parameter is denoted ε. The parameter controls the privacy degree of the mechanism. If $\delta = 0$, $\mathscr{M}$ is ε-differential privacy. The schematic of DP is shown in Fig. 6. For two neighbour datasets D and $D^{'}$, the probability ratio of the output $\mathscr{S}$ by the mechanism $\mathscr{M}$ is running are all less than e^{ε}.

Definition 3. (Privacy Loss) [68]: The observed value ξ produces privacy loss is denotes as:

$$\mathscr{L}^{(\xi)}_{\mathscr{M}(x)\|\mathscr{M}(y)} = \ln\left(\frac{\Pr(\mathscr{M}(x) = \xi)}{\Pr(\mathscr{M}(y) = \xi)}\right)$$

The privacy loss reacted to the degree of protection with algorithmic $\mathscr{M}$ is, in other words, if the probability distributions of $\mathscr{M}(D)$ and $\mathscr{M}(D^{'})$ are greater, the privacy loss is larger. On the contrary, the privacy loss is smaller.

Definition 4. ((α,β)-Accuracy) [68]: A query release mechanism $\mathscr{M}$ is (α, β)-accuracy concerning queries $f \in \mathscr{\ell}$ if every database x, with probability at least $1 - \beta$, the output of the mechanism $\mathscr{M}(x)$ satisfies:

$$\max_{f \in \mathscr{\ell}} |f(x) - (x)| \leq \alpha$$

The definition expresses that the probability of malfunction is less than β and denotes $P(|f(x) - \mathscr{M}(x)| \geq \alpha) \leq \beta$, in which $|f(x) - \mathscr{M}(x)| \geq \alpha$ is "not accurate".

3.2.2. Mechanisms of differential privacy

The fundamental mechanisms include the Gaussian mechanism, Laplace mechanism, and Exponential mechanism, which are used to guarantee DP. While the Gaussian mechanism and Laplace mechanism are applied to the numerical results, and the exponential mechanism is applied to the nonnumeric results.

Definition 5. (Gaussian Mechanism) [68] A randomised algorithm $\mathscr{M}$ satisfies (ε, δ)-DP for any query function $f : D \rightarrow \mathscr{R}$, if:

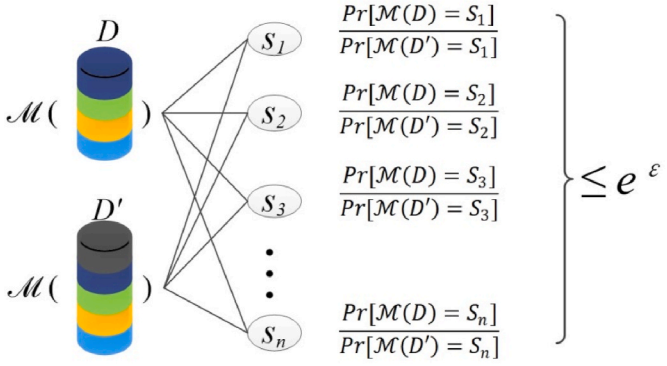

Fig. 6. Schematic of differential privacy.

$$\mathscr{M}(D) = f(D) + \mathscr{N}(0, \sigma^2)$$

where $\mathscr{N}(0, \sigma^2)$ is the Gaussian distribution, and σ ($\sigma = \frac{\|f(D)-f(D')\|_2}{\varepsilon}\sqrt{\left(2\ln\left(\frac{2}{\delta}\right)\right)}$) is the standard deviation of Gaussian distribution.

Definition 6. (Laplace Mechanism) [69] A randomised algorithm $\mathscr{M}$ satisfies ε-DP for any query function $f : D \to \mathscr{R}$, if:

$$\mathscr{M}(D) = f(D) + Lap\left(\frac{\Delta f}{\varepsilon}\right)$$

where $\Delta f = \max_{D,D'} \|f(D) - f(D')\|_1$ is the sensitivity of f andthe noise $Lap\left(\frac{\Delta f}{\varepsilon}\right)$ is defined by the Laplace distribution, which is the probability density function of a random variable $Lap(x|u, b) = \frac{1}{2b}\exp\left(-\frac{|x-u|}{b}\right)$. Here, the location parameter denoted as u, and the scale parameter denoted as b. A schematic of the Laplace distribution is shown in Fig. 7. The scale parameter b is smaller, and the curve is "thinner".

Definition 7. (Exponential Mechanism) [68]: For a query $f : D \to \mathscr{R}$ and a utility function $u : D \times \mathscr{R} \to R$, the prior distribution $\pi(y)$ of output $f(x)$, the exponential mechanism $\mathscr{M}$ is defined as follows:

$$P_{\pi,u}(y) \propto \pi(y) e^{-\beta u(x,y)}$$

DP technology involves many noise mechanisms, for instance, Gaussian, Laplace mechanism, and Exponential mechanism. In subsequent work, many mechanisms have occurred successively. The staircase mechanism and the stage Laplace mechanism have been proposed to reduce the noise power introduced by DP. These two mechanisms can reduce the influence of data query and function computation [70]. The improved matrix Gaussian mechanism (IMGM) for matrix-valued DP has been proposed to solve the influence of data correlation [71]. The DP bounds and the corresponding central limit theorem have been proposed to solve multiple queries and iterative calculations [72]. Some researchers proposed the privacy amplification approach of different scenarios to reduce the loss of the DP mechanism to performance, such as the privacy amplification based on sampling, iteration, reorganisation, etc.

4. Privacy challenges

In this section, we identify several privacy challenges in smart city applications in an edge computing environment. We begin by providing examples of privacy and security cases that have arisen in this context. We then offer a concise overview of the typical privacy challenges that smart cities face. This section focuses on addressing the following issue raised in Section 2:

IC1: Typical security issues and challenges in smart city applications for the edge computing.

To provide a comprehensive understanding of the privacy issues and challenges in smart city applications for edge computing, we have classified them into several typical categories. Fig. 8 presents the classification, and each category is described in detail in subsection 4.2.

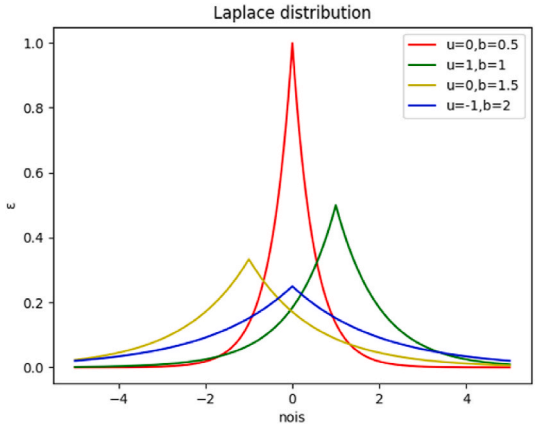

Fig. 7. Schematic of laplace distribution.

4.1. Some privacy cases in smart city applications

In this subsection, some typical privacy attacks/challenges and example cases in smart city application scenarios are listed in Table 5.

As listed in Table 6, these security and privacy incidents are closely related to people's daily lives. In other words, data privacy leaks often occur in the condition of smart cities. For example, a case of infringement of citizens' personal information was discovered in Zhenjiang, China. More than 10 provinces and cities were involved, and there were 30 suspects. The attackers used overseas chat tools and virtual currency to collect and sell more than 600 million pieces of personal information and illegally earned more than 8 million RMB [80]. Furthermore, the Elasticsearch server used by the British data analysis company, Polecat, leaked nearly 30 terabytes of data to the public network [81]. Privacy attacks are divided into two categories: black box attacks and white box attacks, which mainly contain membership inference attacks, reconstruction attacks, property inference attacks, and model extraction attacks. The main reason for this is that the server itself is not protected by any authentication or other forms of encryption. In consequence, it is of great importance to develop data privacy protection techniques for smart cities.

In the smart city application scenario, Duan et al. [82] classified data privacy issues into three categories: data privacy, information privacy, and knowledge privacy. However, the broader issue of data security has become a significant concern, encompassing data privacy, data availability, and data integrity [83]. In recent years, there has been extensive research on data privacy protection to address this concern.

Anonymity technologies such as k-anonymity, t-closeness, and l-diversity have been developed for early privacy protection. In 2002, Latanya Sweeney [84] proposed a k-anonymity model along with policies for deployment to address the issue of data owners releasing private data. However, k-anonymity has been found to be insufficient in preventing attribute leakage. To address this, Li et al. [85] proposed the t-closeness concept, which requires the distance between two distributions (the distribution for sensitive attributes in any equivalence category and the distribution of the attribute in the overall data) to not exceed a threshold t. Machanavajjhala et al. [86] introduced the l-diversity concept, which aims to publish data without revealing sensitive information. Although these technologies provide privacy to some extent, they only offer anonymity and therefore have limitations.

Encryption technology has emerged as a primary means of protecting privacy in recent years. For instance, Xia et al. [87] developed an encryption method to safeguard sensitive data and prevent unauthorized access. Liu [88] proposed methods for constructing a secure public-key encryption scheme, while Li et al. [89] presented a time-domain multi-authority outsourcing attribute encryption method to address data collection and sharing in edge computing. However, encryption technology has a significant disadvantage in that key management and distribution can be costly and complex, limiting its practicality in some contexts.

4.2. Privacy challenges

Here we analyse issues and challenges of the privacy in the edge computing environment. Privacy concerns in edge computing can be categorized into five stages of data handling in smart city applications: collection, storage, and transmission, processing, and publishing/sharing, as depicted in Fig. 8.

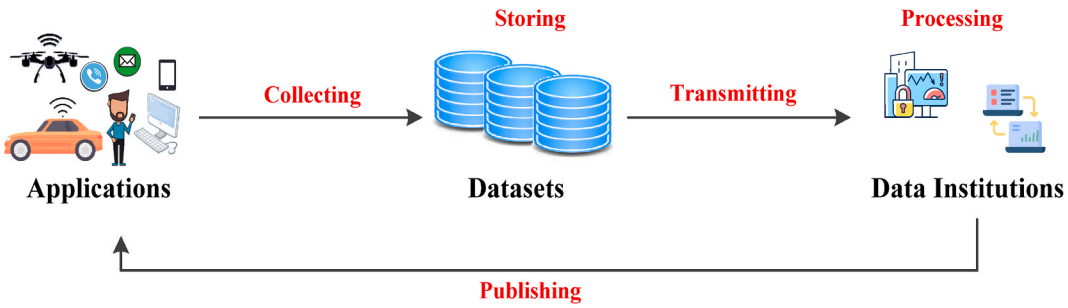

Fig. 8. The privacy challenges of edge computing.

Table 5
Typical privacy attacks and example cases.

Scenarios	Security Attacks/ Challenges	Cases	Ref.
Smart Community	Novel Forms of Cyber-attack, Denial-of-service (DoS) attack, A Man-in-the-Middle Attack, Viruses, Worms, Trojans, and Spyware, SQL Injection Attack, Social Engineering Attacks.	GPS Spoofing Attacks on Unmanned Ground Vehicles.	[73]
Smart Transportation	Privacy, Authorisation, Verification, Access Control, System Configuration, Information Storage, and Management.	Privacy of Vehicle Tracks.	[74, 75]
Smart Logistics	Leaked credentials; Hijacking attacks; Cross-cloud attacks; Service request attacks; Authentication; Man-in-the-middle attack; Response attacks; Impersonation attacks; Malicious edge device attacks; RFID tag attacks; Spoofing attacks; Blackhole attacks; Denial of service attacks; Hardware Trojan; Injection spoofing group attacks;	The recipient cannot receive the courier normally.	[76, 77]
Smart Healthcare	Information Leaks; Insertion Attacks; Minimality Attacks.	Disclosure of patient privacy information.	[78]
Smart Manufacturing Service	Communication Channels; DoS Attacks; Disclosing User Privacy.	Harm to humans and the environment, Damage to or loss of process, equipment, or any other assets.	[5]
Smart Home	Device Vulnerabilities; Trusted Codebases with Bad Security Records; Excessive Access Granted to Cloud Servers; Excessive Access Granted to Device Handlers or Smart Apps.	Determine whether there is anyone in the house based on the water or electricity situation.	[79]

As summarised in Table 6, some real security incidents are data breaches with a global impact from 2016 to 2021.

4.2.1. Data collecting

Due to the large number of perception nodes and intelligent terminal devices in the edge computing environment, these devices may leak users' personal information such as location and identity during data collection. The data collected by sensor networks is the foundation of smart systems and applications. However, these networks are generally vulnerable to attack [100]. The edge server provides access control for users, and the data collected by the end devices can be offloaded to other untrusted servers. The privacy challenges for data collecting in edge computing-based smart city application scenarios are summarised as follows.

- Unsafe communication protocol: Wireless communication technology is used to connect the edge node and mobile devices. The edge node that connects the cloud services is message middleware or network virtualisation technology. These unsafe communication protocols in edge computing environments have eavesdropped on and are easily tempered. In addition, these protocols lack encryption, authentication, and other measures.
- It is easy to initiate DDoS (distributed denial of service) for two main reasons. First, the field devices participating in the edge computing environment usually use simple processors and operating systems. Second, the computing resources and bandwidth resources of the device itself are limited, which makes complex security defence solutions impossible.
- Account hijacking: This is usually done through phishing emails, malicious pop-ups, etc. The malicious attacker performs operations, such as modifying user accounts and creating new accounts.

4.2.2. Data storing

In the edge computing environment, the influence of hardware, software, operating system of human operation, server will inevitably change or even delete users' data [101]. The attackers can obtain sensitive and private data directly from users through end devices. Therefore, we need to pay furthermore consideration to data privacy with storage in the edge computing environment. The privacy challenges for data storing are summarised as follows.

- Sensitive and private data in the edge devices. In the sensitivity data security domain, many researchers are concerned about data loss, breach, and illegal data manipulation (copying, publishing, and dissemination). In the privacy data security domain, the existing studies mainly focus on privacy data release, location privacy, and identity information.
- Data can be marred easily in the edge node. The reason for this is that the data lack effective data backup, recovery, and audit measures.
- Devices' security challenges refer to physical security issues that threaten edge computing devices, such as device damage and link failures caused by natural disasters.

4.2.3. Data transmitting

In edge computing environments, the wireless transmission of data poses a significant risk of data breaches as it is susceptible to eavesdropping and data theft by hackers. Furthermore, the distribution of devices in the network increases the potential of data breaches. As a result, ensuring the privacy of transmitted data is a major challenge in edge computing environments.

Table 6
Some real security incidents.

Time	Institutions	Incidents	Reasons	Damages
October 2016 [90]	Adult Friend Finder	Most user data (name, email, and password) are leaked.	Weak algorithm SHA-1.	412.2 million accounts
February 2018 [91]	My Fitness Pal	Users' related login information data (E-mail, IP, and login credentials) are sold on the dark web and more broadly.	Hackers accessed the network to extract data.	150 million user accounts
November 2018 [92]	Starwood (Marriott)	The users' information (name, contact, passport, travel, and other sensitive data) that was exposed for Starwood.	The access to Starwood system was gained by malicious attacker.	500 million guests.
April 2019 [93]	Facebook	The users' data (phone numbers, account names, and Facebook IDs) are posted for free.	Database leaked for free on the dark web.	533 million users
November 2019 [94]	Alibaba	Some users' data (user name, phone number) are collected by an affiliate marker.	The affiliate marker scraped customer data by crawler software that he created.	1.1 billion pieces of user data
March 2020 [95]	Sina Weibo	The users' data (name, username, gender, location, and phone numbers) are sold by an attacker.	The attacker had gathered publicly posted information by using a service.	538 million accounts
July 2020 [96]	Bitglass	Data breaches at US medical institutions.	Loss or theft of endpoint devices.	15 billion usernames and passwords
January 2021 [97]	Nitro PDF	Large-scale leakages of the user database, over 77 million pieces of data (email addresses, usernames, and passwords) were leaked.	Hackers publicly leaked user records databases for free.	77 million accounts
June 2021 [98]	LinkedIn	A dataset containing information including email addresses, phone numbers, geolocation records, genders, and social media details is posted.	A hacker used data scraping techniques by exploiting the site's and others' APIs.	700 million users
January 2022 [99]	Broward Health	The patient database of third-part medical was breach. The patients' name, birthday, home addresses, licence and so on were divulged.	Some endpoints have no Multi-Factor Authentication (MFA).	1.3 million patients
March 2022 [99]	L'Assurance Maladie	The accounts of 19 pharmacists were compromised and the data of institution were breached.	The accounts and passwords were retrieved by Hackers in a dark web forum.	510,000 people

- Data leakage: In an edge computing environment, data transmission may pass through multiple nodes, some of which may have security vulnerabilities, leading to data leakage. In addition, attackers may obtain data transmitted through methods such as network sniffing, resulting in data leakage.
- Data tampering: During data transmission, attackers may tamper with data, thereby affecting data integrity and credibility. For example, an attacker might modify the transmitted data to include malicious code or false information.
- Identity forgery: Attackers may forge identities, impersonate legitimate users or devices for data transmission, so as to obtain sensitive information or attack the system.

4.2.4. *Data processing*

In the edge computing environment, users' data become more vulnerable to security breaches and unauthorized access due to the data processing in the edge layer [102]. The privacy challenges for data processing are summarised as follows.

- Data breaches and unauthorized access are major concerns in edge computing due to the large number and wide distribution of devices. The security of data is often vulnerable to attacks and snooping, which may lead to the exposure of sensitive user information, resulting in security risks and privacy issues.
- Another challenge is the lack of control over data processing. Data is often collected, transmitted, and stored by multiple devices, making it difficult for users to understand the specific processing process of data and control it. This can lead to misprocessing or misuse of data, which can be detrimental to privacy.
- Implementing a uniform privacy policy in a distributed system is challenging. Edge computing involves data exchange and processing between multiple devices and nodes, making it difficult to implement a consistent privacy policy. This can lead to data inconsistencies and vulnerabilities, increasing the risk of data breaches and unauthorized access.

4.2.5. *Data publishing*

In the edge computing environment, the proliferation of devices and distributed computing has led to an increase in data interaction between devices. Furthermore, the development of data mining and analysis technologies has also increased the risk of privacy breaches. As a result, the issue of privacy protection in data release has become a significant challenge for edge computing [103]. The challenges to privacy in data publishing in this environment can be summarised as follows.

- Since many open devices may publish data with each other, this data may contain sensitive information about the user, such as location, health status, and so on. Therefore, data needs to be anonymized before being released to public places to protect users' privacy.

- Edge computing is usually based on wireless networks for data distribution, which is vulnerable to hacker attacks and malware infections. Therefore, security protocols and encryption technologies should be used to ensure the confidentiality and integrity of data in the release process to avoid data theft or tampering.

4.3. Summary

The use of edge computing in smart cities requires a higher level of privacy protection due to the increased interconnection of devices and data flow. As a result, there is a greater demand for data integrity, confidentiality, and availability. This distinguishes the privacy challenges faced in smart cities from those in other scenarios, which can be attributed to the following factors.

- Diversity of data sources: Smart cities have a wide variety of data sources, including sensors, surveillance cameras, mobile devices, etc., and these devices have different data collection methods and data types, so how to protect different types of data privacy need to be considered during data collection and storage.
- Security of data transmission: Data transmission in smart cities usually needs to be carried out through wireless networks, which is less secure and vulnerable to hacker attacks and eavesdropping, so a series of security measures need to be taken to protect the security of data transmission.
- Privacy protection of data processing: Data processing in smart cities usually involves a large amount of personal privacy information, such as face recognition, vehicle recognition, etc., so a series of privacy protection measures need to be taken to protect this personal privacy information.
- Transparency of data release: Data release in smart cities needs to ensure the transparency of data, that is, the source, processing process and use of data need to be open and transparent, so that the public can supervise and evaluate the legality and rationality of data use.

The distributed nature of edge computing can provide better architectural support for data privacy and security applications. However, the edge computing environment has certain requirements for privacy. Firstly, because the location of edge nodes is close to users and data are not backed up, attackers can exploit vulnerabilities in nodes to obtain advanced permissions, such as modifying or accessing users' private data. Secondly, if an attacker gains control of an edge data center, all services in the area can be compromised, and useful information can be obtained through untrusted edge servers. In addition, some edge servers are not properly managed and can be easily compromised by hackers. Unlike traditional privacy protection methods such as encryption technology, end devices in the edge computing environment may not meet system requirements, and excessive computing energy consumption is also a factor that needs to be considered. Therefore, privacy protection for both edge servers and end devices needs to be carefully considered.

The challenges of privacy protection in the edge computing environment can be summarised as follows. Firstly, the limited resources of edge servers may affect the effective implementation of some privacy protection strategies. Additionally, the computing and transmission limitations of edge servers can result in simplified services with reduced capability. Finally, there is an increasing demand for lightweight privacy protection strategies in a distributed environment, which may weaken the extent of privacy protection. Therefore, addressing these challenges is critical to ensuring the privacy and security of edge computing systems.

5. Technology research

In this section, the related research work about DP is introduced. First, we summarise and compare existing privacy-preserving methods. Afterwards, we describe the advantages of DP for edge computing in smart city application scenarios.

This index content is based on the IC2 mentioned in Section 2:

IC2: The privacy-preserving techniques and the advantage of the differential privacy method applied in the edge computing environment-based smart IoT system.

According to the study, existing privacy-preserving techniques in the edge computing environment are displayed in Table 7, and discussions are included in sub section 5.1. The advantages of DP are summarised according to the technique characteristics and the requirements of the smart city application scenarios based on edge computing in subsection 5.2.

5.1. Privacy-preserving methods

The first group of privacy protection methods include data disturbance [104], data anonymity [105], and data encryption [106,107]. Data disturbance mainly adds noise and random disturbance to the original data to distort sensitive data. However, this technology cannot guarantee that the data statistics are disturbed, as well. Therefore, the data release with DP technology [44] using data disrupting technology has also appeared one after another, but the computational cost is relatively high. Data anonymization changes or publishes the data to be used in some way to prevent key information from being identified. Data anonymization is used for the published trajectories [108] and health data [109]. Nonetheless, data anonymization is unable to handle the huge size of data processing [110]. Data encryption technology transforms a message into meaningless cipher text through an encryption key and encryption function. Then, the plaintext was restored into the cipher text by receiver through the function and key to decryption.

In recent years, numerous researchers have proposed advanced privacy protection technologies such as secure multiparty computation technology [112], private information retrieval (PIR) [113], data desensitization [114], and data cleaning technology. Secure multiparty computation is a universal cryptographic primitive that enables distributed parties to jointly compute an arbitrary functionality without revealing their private inputs and outputs [115]. Private information retrieval facilitates the retrieval of one of K pieces of information from N facsimile databases without unveiling the identity of the queried information to any single database [116]. Data desensitization reduces the sensitivity of sensitive data through substitution, distortion, and other transformations while retaining certain usability and statistical characteristics [117]. Data cleaning technology inspects and verifies the collected data, and the analyser can obtain accurate data by deleting duplicate information and correcting existing errors [118]. Table 7 presents a comparison of the existing privacy protection technologies.

5.2. The advantages of differential privacy in edge computing

Some researchers have proposed numerous effective privacy-preserving methods for edge computing environments. DP has become one of the most popular methods for privacy preservation because of its ability to define a strong privacy guarantee.

There are many methods to protect the privacy of users' data for smart devices in the smart city based on edge computing environment, such as a mobility support system (MSS) [119], a higher-level security transmission with multichannel communications [120], a dynamic customisable privacy-preserving model basis of Markov decision process [121], and a privacy-aware task offloading method (POM) [122]. However, traditional protection methods need to rely on stronger background relationships and reduce the data utility. Therefore, some privacy preservation methods with DP provide various advantages in the edge computing environment. For example, Meng et al. [19] designed a data collection method of location that imports a DP mechanism with random disturbance in the Voronoi diagram. The privacy-preserving mechanism can not only better meet users' privacy needs but also

Table 7
The comparison of different technologies for privacy-preserving.

Privacy Protection Technology	Usage Scenarios (Typical)	Protected Objects	Methods/Strategies	Privacy Requirements
Differential Privacy	Open Statistical Database	Objects in the Statistics Database	• Laplace Mechanism; • Exponential Mechanism; • Gaussian Mechanism.	• Priori-Knowledge; • Errors Exist; • Cannot be Precise.
Secure Multiparty Computing	Private Payment Channels	Private payment between merchants and customers.	• Oblivious Transfer; • Garbled Circuits.	• Confidentiality; • Correctness.
Private information retrieval	Internet Terminal.	• Documents; • Data; • Events.	• Manual Search; • CD Search; • Online Search; • Network Search.	• Confidentiality; • Correctness.
Homomorphic Encryption	• Retrieval; • Statistics; • AI Tasks.	Personal information data.	• Encrypted Neural Network; • Encrypted KNN; • Encrypted Decision Tree; • Encrypted Support Vector Machine and other Algorithms.	• Confidentiality of Data; • Availability of Data.
Format-Preserving Encryption (FPE)	• Database; • Data Masking Field.	Sensitive information in the database, network data	• Prefix; • Cycle-Walking; • Generalised-Feistel.	• Data Confidentiality; • The Format of the Data Remains Unchanged.
Data Cleaning	Media Equipment	• Disks; • Flash Memory Devices; • CDs; • DVDs.	• Covering; • Degaussing; • Encryption; • Physical Destruction.	Data confidentiality.
Data Anonymization	Medical information disclosure	Patient personal information and disease privacy.	• *K*-anonymity; • (α, k)–Anonymity [111]; • L-Diversity [86]; • T-closeness model [85].	• Unrecognisable; • Data Availability.
Data Masking	Enterprise storage, organisation, and management database.	• Personal information; • Other Sensitive Information.	• Rounding; • Quantification; • Shielding; • Truncation; • Unique Replacement; • Hashing; • Rearrangement.	• FPE Encryption Data Confidentiality; • Data Availability.

have higher data availability. Du et al. [18] designed two DP algorithms, output perturbation (OPP) and objective perturbation (OJP), to guarantee privacy protection. These algorithms can guarantee the accuracy of benchmark datasets.

The use of edge computing can provide benefits beyond data privacy protection, including energy consumption reduction, resource efficiency, and improved computing power [123]. For instance, Miao et al. [56] proposed a DP method for preserving location privacy that reduces the computational complexity and resource consumption at the edge node. Gai et al. [124] developed a privacy-preserving mechanism for IoT and blockchain in the edge computing environment to enhance trustworthiness and achieve optimal task allocation through DP. Additionally, Liu et al. [57] designed a DP framework for data release in the edge computing environment, which enhances the accuracy of queries. In summary, the main advantages of DP in edge computing are as follows.

- Meeting users' privacy needs and enhancing data availability.
- Reducing computational complexity and resource consumption.
- Enhancing the trustworthiness of the edge nodes.
- Enhancing the accuracy of the query.

6. Applications and implementation

This section focuses on the applications and implementation of DP in smart city scenarios based on edge computing [8]. We begin by introducing the detailed classification of DP research for these scenarios. Then, we explore the application of DP in various stages of data management in the edge computing environment, highlighting its benefits in protecting data privacy. Specifically, we focus on four stages, including data privacy in transmitting, processing, model training, and publishing. In addition, typical data privacy issues, such as location/trajectory privacy protection, are also discussed. Some typical applications of smart city, such as smart manufacturing service, smart logistics, and data management, are shown in Fig. 9. In this study, the complete course of data management from data collection to data publishing in smart city applications for edge computing is considered. The course of data management is shown in Fig. 9, which can be mainly divided into transmitting, processing, training, and publishing.

In recent years, with the rapid growth in the number of smart IoT devices, big data are being generated in the forms of text data (e.g., e-mails, web pages), images, audio, video, and location information (e.g., latitude, longitude, altitude) [125]. These data may contain users' private information, and hence, data privacy preservation has become a critical issue. Edge computing brings enormous benefits to analysing and mining data, perceiving location information, and localisation [126]. Nonetheless, data privacy-preserving research is important for users and managers. Chinnasamy et al. [127] reviewed the data security and privacy requirements in the edge computing environment. First, they emphasised the definition of edge computing protection and confidentiality criteria. Second, they proposed the categorisation of threats on the edge device through the definition. Then, they presented the state-of-the-art tactics used to mitigate privacy risks. At the same time, they designed the measurements for the efficiency of interventions and the related technical pattern of mitigating the attackers. Finally, they showed the development in technical approaches and research directions of potential professionals about the area of edge device privacy and security. Zhang et al. [128] proposed investigation on the privacy concerns of edge devices.

In this subsection, we introduce data privacy in data transmitting, data processing, model training, and data publishing for the smart city application scenarios based on edge computing. As summarised in Table 9, we compare the existing DP method with privacy protection for edge computing in edge computing-based smart city application scenarios, privacy mechanisms/algorithms, types of noise, types of privacy, types of data, and methods of adding noise.

This section is based on IC3 and IC4 in Section 2:

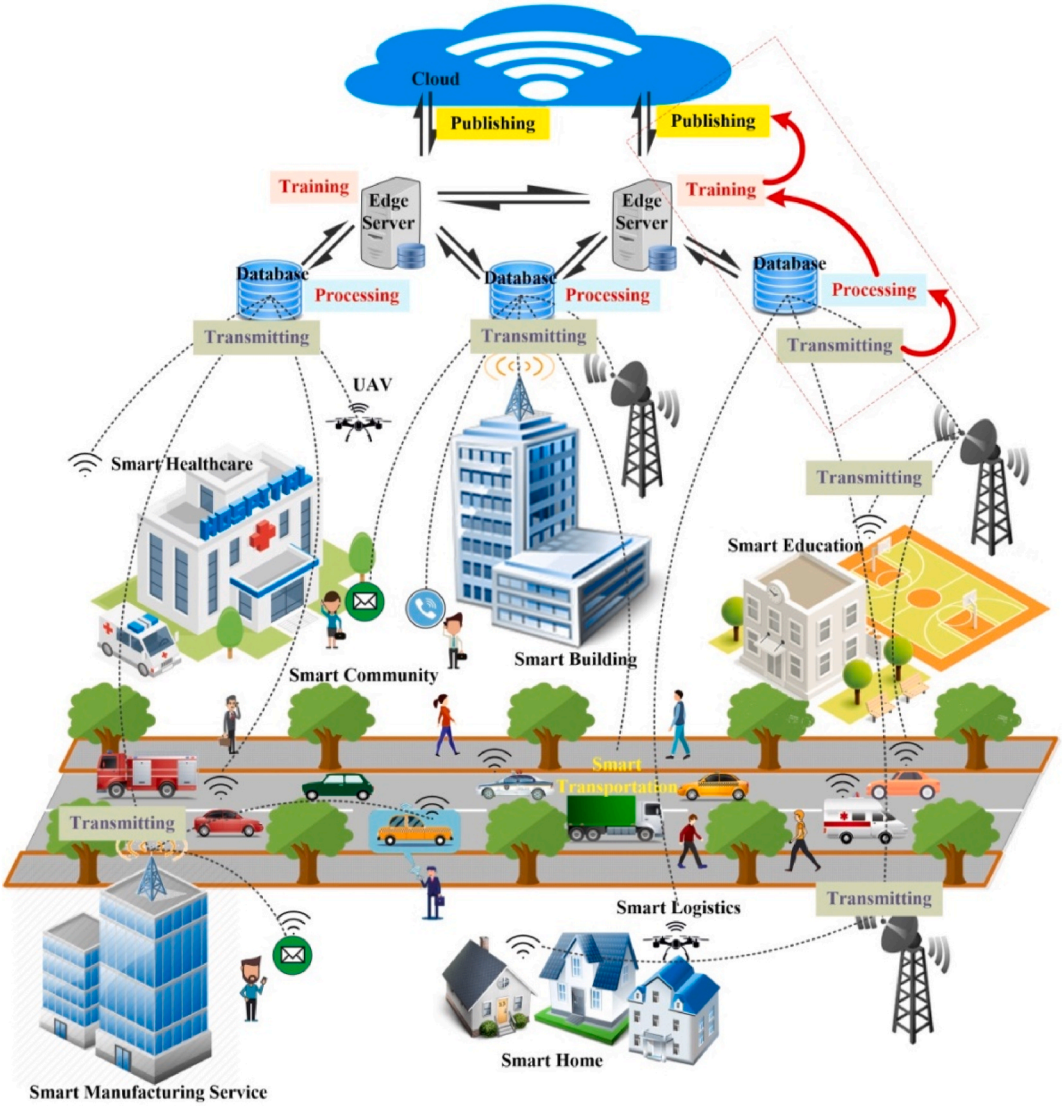

Fig. 9. Typical smart city applications and data management stages.

IC3: The applications of the edge computing environment are considered in differential privacy method research.

DP research in the edge computing environment has various applications, including protection mechanisms, algorithm design, data management, and federated learning. In this article, we focus on privacy preservation in the context of data management, which includes various stages as shown in Fig. 9.

IC4: The experimental platforms/tools, datasets, and evaluated metrics are applied to implement the DP mechanism or algorithm.

The implementation of DP techniques can be presented as generating noise data, the objective function is added noise, parameters, or gradients in model training, and the models' output is added noise. We summarise the implementation platform/tools, common datasets, and evaluation metrics of DP for some applications in Table 8. Table 9 illustrates the different characteristics of DP, and they are compared in the edge computing environment.

6.1. Data privacy in data transmitting

In the edge computing environment of smart city scenarios, the data transmitted between users and edge services or devices may expose users' privacy information [129]. Malicious attackers steal or falsify the data being transmitted, thereby mining the private information in the data. In addition, malicious attackers predict user behaviours according to the data. For example, the China Railway Service 12,306 data transmission security incident caused the leakage of 600,000 accounted and 4.1 million pieces of contact information in December 2018 [130]. Data transmission that is not properly handled can cause a great infringement on user privacy. Therefore, the privacy preservation of data transmission for cloud and edge computing has received much attention in recent years [131–133].

There are two parts to data transmission privacy preservation in the edge computing environment. For secure computing methods, Shafagh et al. [134] designed a platform named Pilatus to protect the security of data transmission. Aujla et al. [135] designed a framework named SDN-aided to provide privacy-preserving with the secure grid-based cryptograph system for edge-cloud interplay data transmitting and SDN-assisted management. Shen et al. [136] showed a scheme with bilinear mapping and homomorphic encryption to judge the positional link of multi-query keywords on the transmitted data.

On the DP method, Nie et al. [59] put forward a DP tensor computing model (DPTCM) to protect the data privacy for transmitted to the cloud. They implemented the privacy of transmitting data through the flexible computing function of edge computing. At the same time, the DP method does not generate excess overhead. Hassan et al. [35] reviewed the security of data transmission with transportation systems. They

Table 8
Platform/tools and common datasets with different articles.

Ref.	Platform/Tools	Common Datasets			Evaluation Metrics
		Name	Type	Number	
[18]	TensorFlow	MINIST	Handwritten Digital Images	6000 Samples	• Accuracy; • Data Utility; • Privacy.
		SVHN	Image Dataset	9812 Digits•	
		CIFAR-10	Colour Images	6000 Samples•	
		STL-10	Colour Images (Few Label)	6000 Samples•	
[30]	Python 3.7	Adult	Census Database	48,442 Records	• Prediction Accuracy; • Model Performance.
		General Social Study (GSS)	Personal Information Related to the Happiness	51,020 Records•	
[56]	*	Taxi Trajectories	Taxi Routes in Beijing	10,357 Routes	• Query Accuracy; • Algorithm Runtime Efficiency.
[102]	Python 3.6	Checkin from Gowalla	User Locations	6,442,890 Check-in Information	• Query Accuracy; • Algorithm Operation Efficiency.
		GPS Trajectory from Geolife	Users' Trajectory	182 Users (5 Years)	
[58]	*	Loc-Gowalla	User Locations	• 86 nodes; • 291 Edges; • 48,906 Records	• Query Accuracy; • Query Error.
		T-drive	Taxi Tracks in Beijing	10,357 Tracks	
[59]	• Python; • Cloud Server; • 64-GB RAM.	Wsdream	QoS Evaluation Result	Throughput Values of 4500 Web Services	• Complexity; • Communication Cost;
		UTS	Urban Traffic Speed	214 Anonymous Road Segments	• Accuracy.
[61]	Super Computing Center	YFCC100 M	Videos and Images	100 Million Media Objects	• Average Reward (AD); • Average Regret (AR); • Cumulative Regret (CR).
[150]	• TensorFlow; • VGG-Face Network; • MTCNN.	LFW	Labelled Face in the Wild	• 158 Identities; • 4324 Label Face Images.	• Accuracy; • Time Cost; • Memory Cost; • Energy Cost; • Network Transmission Cost.
[137]	*	Simulation Experiment.			• Average Delay Reduction; • Task Multiplier.
[142]	• Java Development Toolkit (JDK) Version 1.8; • Weka.	Adult Disease; Heart Disease;	14 Attributes 75 Attributes	48,842 Records 3030 Records	• Scalability; • Processing Time; • Accuracy.
[151]	*	SUMO	Simulation of Urban MObility	100 Trajectories	• Data Utility; • Efficiency; • Respectively.
[152]	*	SEARCH LOGS	Google Trend data and AOL search log.	32,768 Records	• Mean-squared error (MSE);
		AGE	Brazilian Census Data	100,078,675 Records	• KL Divergence (KLD);
		LOCATION	New Zealand demographic data.	7725 Blocks	
[153]	*	geom.net	Authors Collaboration Network	7343 Vertices and 11,898 Edges	• Privacy; • Utility;
		out.moreno_lesmis_lesmis	Characters Cooccurrences Network	77 Vertices and 254 Edges.	• Mean Absolute Error (MAE).
		RGD	Randomly Generated Dataset	100 Vertices and 1645 Edges	

divided DP implementation into three types in the transportation system: the networks of railway freight and vehicular, the data of automotive manufacturer. Wang et al. [137] considered the privacy of the connected vehicles, which mainly contains the process of the vehicles transmitting data to the roadside unit and base station, such as the speed and location. They proposed a system architecture named privacy-preserving vehicular edge computing (*P*P-VEC) by disturbing the context information of connected vehicles to address the privacy protect issues of the transmitting process.

In summary, the privacy preservation of data transmitted with DP for edge computing environments in smart city scenarios mainly has three characteristics.

- Low energy cost, namely, the DP mechanism, has little impact on energy consumption.
- Type diversiform, which means that the DP mechanism can address different types of data privacy.
- Low latency means that the DP mechanism requires little support of task offloading.

6.2. Data privacy in data processing

Data processing usually involves data mining and retrieval. Data mining is the process of extracting the potential information for data that is hidden and unknown but potentially useful from a large amount of data. The goal of data mining is to build a decision model that predicts future behaviour based on data from past actions.

In the smart city scenario, data mining applications have been deployed for our daily lives, such as smart transportation [138], smart healthcare [3], and smart detection systems [6]. In these scenarios, the edge devices implement the collection and pre-processing of datasets, and there is abundant personal privacy. Therefore, many researchers focus on cryptography with privacy-preserving data mining [139–141]. However, traditional cryptographic methods produce large extra

Table 9
Comparison of DP in different edge computing-based smart city application scenarios.

Classify	Application Scenario	Privacy Mechanism/Algorithm	Noise Types	Noise Distribution	Add Noise Method	Privacy Types	Ref.
Data Transmit	Large amounts of IoT devices over the network are managed by the software-defined network (SDN).	A differential private tensor computing model.	Laplace	$\tau_{G_k}^{i_j} = ([R(N-1)\Delta]/\varepsilon)\sqrt{2}\overline{e}n^{i_j}$ $\varphi_{G_j}^{i_j} \sim (0, M]$	Add noise tensors $\tau_{G_k}^{i_j}, \varphi_{G_j}^{i_j}$ to gradient $\nabla_{G_{i_{jk}}}^{i_j}$.	ε-DP	[59]
	The computing resource management and task offloading on the Internet of Vehicles (IoVs).	A system architecture named *PP*-VEC (privacy-preserving vehicular edge computing) and an algorithm named *K*-NJTA (*K*-neighbour joint optimisation of task offloading and resource allocation).	Exponential	$exp\ (-\varepsilon \times \lvert r - q(B)\rvert)$	The local differential privacy and a MWEM mechanism processes the context information for connected vehicles.	ε-DP	[137]
Data Mining	The data owners turn to collaboratively train machine-learning models in the edge computing environments.	The distributed data mining scheme with differential privacy based on tree structure.	Laplace	$Lap\left(\frac{1}{\varepsilon}\right) = \frac{\varepsilon}{2}e^{-\lvert x\rvert\varepsilon}$	Adding noise to when calculating the supports of the leaf nodes in each iteration.	ε'-DP ($\varepsilon' = \frac{P}{T}$, P is the total number of privacy budget, T is the number of iterations)	[30]
	In the public infrastructure is provisioned protection to protect an individual's information.	A fuzzy convolution neural network (FCNN) is injected noise with a Laplace mechanism.	Laplace	$Lap\left(\frac{\Delta Q}{\varepsilon}\right) = \frac{\varepsilon}{2\Delta Q}e^{-\frac{\lvert x\rvert\varepsilon}{\Delta Q}}$	$R(x) = Q(x) + Lap\left(\frac{\Delta Q}{\varepsilon}\right)$	ε-DP	[142]
Data Retrieval	The multimedia big data of the mobile social networks in the edge computing environment.	Tree-based noise aggregation algorithm	Laplace	$Lap\left(\frac{\ln n}{\varepsilon'}\right) = \frac{\varepsilon'}{2\ln n}e^{\frac{-\lvert x\rvert\varepsilon'}{\ln n}}$	$\widetilde{\mu}(n) = \sum_{(x,y)}(r_{x,y} + Lap\left(\frac{\ln n}{\varepsilon'}\right))$	ε-DP	[61]
Model Training	Training datasets in wireless big data scenario.	Output Perturbation Algorithm; Objective Perturbation Algorithm;	Laplace	$p(q) = \frac{1}{\alpha}e^{-\omega\lVert q\rVert}$	$W(D) + q$; Where $W(D)$ is objective function	ε_p-DP (p is disclosure risk)	[18]
	The training process in DNNs faces recognition models.	The client-server model-based DNNs training algorithm in a privacy-preserving manner.	Gaussian	$\mathcal{N}(0, S_f^2 \times \sigma^2)$	$d_{(x,y)}^i = c_{(x,y)}^i + \mathcal{N}(0, S_f^2 \times \sigma^2)$	(ε, δ)-DP	[150]
Data Publishing	Network data publication between two related individuals in the relevant dataset corresponding to the weighted network.	ε-correlated edge differential privacy (CEDP)	Laplace	$Lap\left(\frac{CS}{\varepsilon}\right) = \frac{\varepsilon}{2CS}e^{-\frac{\lvert x\rvert\varepsilon}{CS}}$ where CS is correlated sensitivity.	$\mathcal{M}(D_I) = f(D_I) + Lap\left(\frac{CS}{\varepsilon}\right)$	ε-CEDP(Correlated Edge DP)	[153]
	The publication data of the various sensors.	A partitioned histogram data publishing algorithm based on wavelet transform.	Laplace	$Lap\left(\frac{\Delta f}{\varepsilon}\right) = \frac{\varepsilon}{2\Delta f}e^{-\frac{\lvert x\rvert\varepsilon}{\Delta f}}$	Adding Laplace noise to the coefficient c_i with magnitude $\frac{1+log_2^k}{\varepsilon W_{Haar}(c_i)}$.	ε-DP	[152]
	The release of location data in the edge computing environment.	A differential privacy quadtree partitioning algorithm.	Laplace	$Lap\left(\frac{1}{\varepsilon_x}\right) = \frac{\varepsilon_x}{2}e^{-\lvert x\rvert\varepsilon_x}$	$x.count = C_t + Lap\left(\frac{1}{\varepsilon_x}\right)$ Where C_t is the true count value based on D.	ε_x-DP (x is the node of quadtree)	[102]
Location	The movement features of users in vehicle ad hoc network.	Trajectory partition algorithm with differential privacy and trajectory clustering algorithm.	Exponential	$\exp\left(\frac{\varepsilon}{2\Delta u}u(star_{par}, m)\right)$ where *strt_par* is a prior characteristic node.	Adding noise into the publishing trajectories set $T_1', T_2', \ldots, T_k'$.	ε-DP	[151]
	Add noise to each node of the quad-tree Laplace $\left(\frac{\varepsilon}{h}\right)$	Privacy-aware framework for mobile edge computing (MEPA).	Laplace	$Lap\left(\frac{1}{\varepsilon_x}\right) = \frac{\varepsilon_x}{2}e^{-\lvert x\rvert\varepsilon_x}$	$\widetilde{h} = h_x + Lap\left(\frac{1}{\varepsilon_x}\right)$	ε_x-DP (x is node)	[56]
	The location privacy in the edge nodes for the Internet of Things (IoT).	The geographic indistinguishable mechanism	Laplace	$Lap\left(\frac{\Delta f}{\varepsilon}\right) = \frac{\varepsilon}{2\Delta f}e^{-\frac{\lvert z\rvert\varepsilon}{\Delta f}}$ $Lap(b) = \frac{1}{2b}e^{-\frac{\lvert z\rvert}{b}}$ The noise selects different distributions when the actual positions are in different ranges.	The output of query function f is added noise. $DPQ(G, Y) = f(G, Y) + Lap(V_1, V_2, \ldots, V_r)$	εd-DP (d is the distance of two users)	[58]

(*continued on next page*)

Table 9 (*continued*)

Classify	Application Scenario	Privacy Mechanism/Algorithm	Noise Types	Noise Distribution	Add Noise Method	Privacy Types	Ref.
	The location-based service providers are distrusted in the edge computing environment.	A differential privacy quadtree partitioning algorithm.	Laplace	$Lap\left(\frac{1}{\varepsilon_x}\right) = \frac{\varepsilon_x}{2} e^{-\lvert x\rvert\varepsilon_x}$	$x.count = C_t + Lap\left(\frac{1}{\varepsilon_x}\right)$ Where C_t is the true count value based on D.	ε_x-DP (x is the node of quadtree)	[102]
Others	Add noise to each node of the quadtree Laplace $\left(\frac{\varepsilon}{h}\right)$	Privacy-aware framework for mobile edge computing (MEPA).	Laplace	$Lap\left(\frac{1}{\varepsilon_x}\right) = \frac{\varepsilon_x}{2} e^{-\lvert x\rvert\varepsilon_x}$	$\tilde{h} = h_x + Lap\left(\frac{1}{\varepsilon_x}\right)$	ε_x-DP (x is node)	[56]
	The social network attributes graphs node properties and the correlation of edge information.	The social network attributes graphs algorithm under personalised differential privacy.	Laplace	$\rho(x) = \frac{1}{2b} \exp\left(-\frac{\lvert x\rvert}{b}\right)$, $b = \frac{\Delta Q}{\varepsilon}$	$Q_{DP} = Q + \rho(x)$	ε_v-DP (v is node)	[154]
	The multimedia big data of the mobile social networks in the edge computing environment.	Tree-based differentially private and trustworthy social multimedia distributed online learning algorithm.	Exponential	$exp\left(\frac{\varepsilon^{'} B_{(h,i)}(n)}{\Delta B}\right)$ where $B_{(h,i)}(n)$ is the B-value of cluster (h,i).	Adding noise to the tree structure of edge nodes.	ε-DP	[61]
	In the public infrastructure is provisioned protection to protect an individual's information.	A fuzzy convolution neural network (FCNN) is injected noise with a Laplace mechanism	Laplace	$Lap\left(\frac{\Delta Q}{\varepsilon}\right) = \frac{\varepsilon}{2\Delta Q} e^{-\frac{\lvert x\rvert\varepsilon}{\Delta Q}}$	$R(x) = Q(x) + Lap\left(\frac{\Delta Q}{\varepsilon}\right)$	ε-DP	[142]
	The edge layer communication link on the Internet of Things (IoT).	The geographic indistinguishable mechanism	Laplace	$Lap\left(\frac{\Delta f}{\varepsilon}\right) = \frac{\varepsilon}{2\Delta f} e^{-\frac{\lvert z\rvert\varepsilon}{\Delta f}}$ $Lap(b) = \frac{1}{2b} e^{-\frac{\lvert z\rvert}{b}}$	Laplace noise is added to the output of query function f. $DPQ(G,Y) = f(G,Y) + Lap(V_1, V_2, \ldots, V_r)$	εd-DP (d is the distance of two users)	[58]

computational overhead, which further hinders the application in the edge computing environment.

In the research field of edge computing environment, some researchers have focused on DP to protect the privacy of data mining for the smart city application scenarios in recent years. To economise computing costs and increase efficiency, Sun et al. [30] considered that data mining applications complete the data mining task when the data are not shared by owners aspiration. They focused on the tree-based distributed data mining scheme with DP in the edge computing environment. The participants built a decision model with their data; then, they shared the model after being injected with noise. Sharma et al. [142] proposed DP using an algorithm named fuzzy convolution neural network (DP-FCNN) to address the problem of data accessed by an unauthorized user. The data providers were responsible for injecting noise into the datasets that the data owners uploaded. The scalability, model accuracy, and processing time of DP-FCNN are showed good efficiency.

Data retrieval extracts or queries the data in the database according to the users' needs. The results of data retrieval generate a table that can either be put back into the database or used as an object for further processing.

In the smart city scenario, because most multimedia service providers are located in remote sites, users' access delays will lead to a poor service experience when users choose a service. Edge computing technology can solve delay problems; users retrieve the required data from the nearest edge node and reduce the delay [10]. However, when retrieving the popularity of multimedia content from a massive amount of data, network administrators need a suitable retrieval system [143].

The edge nodes bypass the central trusted system and exchange data with each other, which leads to a leakage of user privacy information. Therefore, a perfect retrieval system alone is not enough, and the data privacy protection problem in the retrieval system needs to be solved. There are many privacy protection schemes, such as anonymization, DP, and desensitization. In large-scale data, DP has little influence on prediction accuracy and the algorithm does not need to be considered [144]. In the foundation of the DP, Zhou et al. [61], they proposed a multimedia content retrieval based on tree privacy-preserving and trustworthy distributed. They made personalised predictions on the edge network and deployed DP and trust mechanisms combined with edge nodes. Zhou et al. [61] found that the time of the model converges to the optimal policy when the privacy level is increased to certain degree.

In summary, privacy preservation for data processing with DP of edge computing-based smart city application scenarios mainly has three characteristics.

- Balanced computing cost and efficiency means that the DP mechanism has a good trade-off between the computation and the efficiency of data mining for the edge computing environment.
- High accuracy means that the accuracy has little impact after a certain quantity of noise is added to the data processing stage.
- Low processing time means that the DP method can disturb the original data in the data processing stage in a short time.

6.3. *Data privacy in model training*

Currently, some attackers acquire private sensitive training data to leakage privacies or attack maliciously in a series of smart city scenarios, for instance smart home, smart medical, and smart logistics,. Xu et al. [145] proposed an attack method for the training data is adjusted by conspired malicious participants in strategically. The attack made a certain dimension weight of the aggregation model is rose or fallen by the pattern. He et al. [146] designed a new attack method to compromise the inference data privacy in a collaborative deep learning system. They verified the effectiveness and generalisation by evaluating the attack method under different settings, models, and datasets. Ji et al. [147] presented a broad class of model-reuse attacks. However,

maliciously crafted models may trigger host ML systems to misbehave on targeted inputs in a highly predictable manner. Zhang et al. [148] studied generative model-inversion attacks where the privacy information of training data are inferred by the access of a model is abusing.

The model training contains some privacy data for different smart city application scenarios based on edge computing environment. Therefore, the challenges of the privacy protection field with edge computing in smart city scenarios are to insure that private information are not leaked in the model training. Some researchers focus on DP technology to solve data privacy with a training model for edge computing-based smart city application scenarios. In 2010, Michael et al. [149] proposed a method with adding Laplace noise to enhance the privacy of training data. They improved the accuracy of a general class for histogram queries while guaranteeing DP. Du et al. [18] considered the privacy of correlated datasets and designed two different algorithms, OPP (output perturbation) and OJP (objective perturbation), to protect the model training data privacy, and the algorithm satisfies DP. In order to abate the computing cost of the local device, Mao et al. [150] enabled training of deep convolutional neural networks to face recognition models with the DP mechanism. They indicated that the training accuracy is more sensitive to the noise of the deeper convolution layer.

The training data privacy-preserving for the smart city application scenarios mostly contains the model training data for edge devices and edge nodes. The noise with a Laplace or Gaussian distribution is added in the process of training to protect the training data by DP method. The privacy budget ε is confirmed by fine-tuning. Both Du et al. [18] and Mao et al. [150] believed that the model effect of privacy budget ε in a certain range is the best. The former considered the privacy budget ε in $\varepsilon \in [10^{-2}, 20]$, and the latter considered it in $\varepsilon \in [2, 5]$.

In summary, the privacy preservation of model training with DP for edge computing-based smart city application scenarios mainly has three characteristics.

- High accuracy means that the accuracy of training model can be increased by the consistent result of the DP mechanism.
- Support of different types of datasets, which means that the DP mechanism can accomplish high-quality privacy preservation in a wide variety of diverse datasets, such as correlated datasets.
- Low computing cost means that the DP mechanism requires low computing cost in the model training.

6.4. Data privacy in data publishing

Data publication for edge computing in smart city scenarios has been a hot research topic for data analysis, which has the characteristics of dynamic, multiple sources, and large amounts. However, there are many security issues with data publishing. The attackers acquire sensitivity information using some attack models, as illustrated in Table 10.

DP applies to data publishing scenarios in various fields, such as histograms and matrix mechanisms. Dwork et al. [68] designed the Laplace mechanism with a histogram, and Laplace distribution ($Lap(\frac{1}{\varepsilon})$) noise was added to the data. Su et al. [159] proposed a method named PrivPfC, which involves publishing data for classification based on DP. They use the exponential mechanism with a novel quality function. They indicate that using fewer steps to avoid spreading the privacy budget is too thin. Yan et al. [160] focused on the data published in V2G networks, and proposed DP algorithm to protect the privacy of the data release in V2G networks. They defined a variable sliding window to improve the utility of data.

In addition to these methods, Xiao et al. [161] designed the mechanism of the Haar wavelet transform, and they added noise to the Haar coefficient. Jia et al. [162] showed an approach called StructureFirst, which uses square error and an exponential mechanism to compress the original histogram. Acs et al. [163] combined the adaptive hierarchical clustering technique and the greedy bisection strategy to propose the method named *P*-HPartation. Jia et al. [162] designed a method named NoiseFirst, which obtained the optimal partitioning of data grids by adding noise and then splitting and merging. However, NoiseFirst is only applicable in the one-dimensional histogram. Therefore, Xiao et al. [164] presented a DPCube that can obtain multidimensional V-optimisation histograms by combining the KD-tree and the unit division to solve the one-dimensional histogram. These five methods have common advantages that support longer-range count queries, and the query accuracy is high. Miao et al. [56] proposed a new noise quadtree data-basedapproach to statistical data release for moving objects. The approach satisfies the DP-preserving model. Liu et al. [102] proposed a location data and distributed data release method in the edge computing environment to preserve user privacy with the release of location data. They constructed a DP complete quadtree and adjusted the quadtree by the threshold to balance the error.

Table 10
The attack models.

Attack Models	Descriptions	Ref.
Link Attacks	The attackers obtain sensitive information by analysing the relationship between an individual and attributes through links.	[155]
Similarity Attack	The attackers obtain part of individual privacy information by analysing the similarity of sensitive attributes in equivalence classes.	[156]
Skewness Attack	The attackers dope out a lot of individual privacy information by analysing the similarity distribution of sensitive attributes in equivalence classes.	[85]
Replay Attack	The attackers use the received data to cheat the system and make it pass the identity authentication.	[157]
Probabilistic Inference Attack	The attackers obtain sensitive information through the difference between before and after publishing data.	[158]
Background Knowledge Attack	The attackers infer the sensitive information of an individual for the background knowledge of the model and part of the information of an individual.	[85]
Homogeneity Attack	The attackers get the sensitive information of the individual if they know that an individual exists in an equivalence class that is the values of all attributes.	[86]

In conclusion, the privacy preservation of data release/publication with DP in edge computing environments adds noise using the Laplace and Exponential mechanism. According to the experimental results of the literature, the smaller the privacy budget ε is, the better the algorithm performance. In summary, the privacy preservation of data publishing with DP for edge computing-based smart city application scenarios mainly has three characteristics:

- Better performance, namely, the performance of publishing data can be improved by DP mechanism.
- Better data utility, namely, the DP mechanism, can improve the utility of data through data dynamics to adjust the publishing window.
- Higher accuracy, namely, the DP mechanism, can improve the accuracy of range queries over attributed histograms.

6.5. Data privacy in location/position trajectory

In smart city applications, some smart IoT devices share real-time information with edge or cloud servers to provide real-time services. For example, users' behaviour raises some security issues, and the best crucial issue is the trajectory/location leakage. The privacy of location/position trajectory belongs to the data privacy-preserving method, as well. Because some researchers often pay attention to the location privacy field, this study refers to the application of related DP methods in

the location privacy field.

Malicious attackers can infer the address, lifestyle, social relationship, etc. By obtaining location data [165]. Location-based service is one of the dominating forms of edge service, and location privacy is a concern of users [166]. Tian et al. [166] proposed an eight-category classification of location-based service for edge computing-based smart city application scenarios from three dimensions, as shown in Figs. 10 and 11.

The existing research on location/position privacy mainly divided into four types of methods which are anonymous, obfuscation, POI modelling, and DP. *K*-anonymous is often used to protect location privacy. However, in the *K*-anonymous model, the usability of data will reduce for the sparse data, and the attackers' background knowledge is explicit [58]. Then, the location's privacy is protected by obfuscation. Nevertheless, the method reduces the precision and the privacy level of location information [167]. Wang et al. [168] considered the database level and proposed a point of interest (POI) modelling method to avoid location information. However, POI modelling needs extensive pre-processing and query frequency. Therefore, the DP approach without background knowledge is a good solution to these deficiencies.

In recent years, DP has provided strong privacy guarantees in location privacy preservation for edge computing-based smart city application scenarios. Zhou et al. [151] presented a DP algorithm to protect the vehicular trajectory. They integrated an exponential mechanism and clustering algorithm. Jing et al. [58] developed the geographic indistinguishable mechanism to preserve the location privacy in the edge nodes for the Internet of Things (IoT). They regarded the edge nodes as the central server and used DP theory to realise the protection of location privacy. Miao et al. [56] presented a framework named MEPA to preserve the location privacy of edge nodes. The framework provides computing services, and an anonymous central server is acted by the edge node. Liu et al. [102] considered the data uniformity heuristic adjustment and designed an algorithm with DP quad-tree partitioning. The location privacy of the query process is protected by Hilbert curve-based the range counting query framework.

In the edge computing environment of smart city scenarios, due to the data independence and low calculation overhead of the quad-tree structure, the quad-tree method is one of the most commonly used methods. In addition, the location privacy preservation is the location of edge nodes and trajectory of vehicles in smart city scenarios based on edge computing environment.

In summary, the privacy preservation of location with DP for edge computing-based smart city application scenarios mainly has three characteristics.

- With a low loss rate, the DP method can reduce the loss rate of location information in edge computing environment. One of the methods is to use linear programming to achieve optimal location fuzzy selection to reduce data loss.
- High utility, which refers to the running utility of the algorithm and searching utility of location information.
- High accuracy, data optimisation through the compressed sensing method to improve the data accuracy.

6.6. Others

In this section, edge node privacy and edge layer privacy are discussed. There is a mass of individual and business edge nodes in the smart city scenarios based on edge computing [18]. Those nodes contain the ability to compute and process data. The edge node is a business platform built near the edge of the user's network, providing storage, computing, network, and other resources. Nevertheless, many malicious attackers cause internet attacks to threaten the security of edge nodes. For example, malicious attackers falsify edge nodes to control sensors or other services by physical attacks [169] and cooperative blackmailing attacks [170].

In addition, the edge node data contain some sensitive privacy. A variety of different big data is analysed by using machine learning approach in the edge node [171]. For instance, Vishalini Laguduva et al. [172] approached the problem of cloning physically unclonable function-based edge nodes in different settings. K.S. Mohanasathiya et al. [173] designed two encryption schemes named searchable and proxy re-encryption to provide a solution based on STFC (the Secured Two Fold Encryption Protocol in Edge Computing). They focused on data security from the double encryption method. Shiva Prasad Kasiviswanathan et al. [174] proposed several techniques for designing node differentially private algorithms. Duan et al. [82] proposed a solution based on the privacy targets of explicit and implicit divisions typed data. The method can solve the challenge of some new requests of user on the multiple sources of various integrated devices for the accumulated content or resources at the edge.

Miao et al. [56] proposed a privacy-aware framework named MEPA for edge computing, the location privacy of the edge node as an anonymous central server is protected. Du et al. [18] considered the privacy issue of any edge node dealing with the data. They strengthen the performance of privacy preservation by injecting noise to the blocked data beforehand and then computing and processing it though each edge node. Zhang et al. [175] and Zhang et al. [176] prevented the node from re-identifying the sensitive information revealed by the attackers.

In the smart city scenarios based on edge computing environment, edge layer privacy is the data privacy of edge layer servers. The data streams from the cloud and infrastructure layer are received, processed, and forwarded in the edge layer [177]. Therefore, in the operation of data processing exists some malicious attacks. For example, the data will be transmitted to malicious nodes in the edge layer [83].

Some privacy-preserving methods and algorithms have been put forward to preserve the privacy of the edge layer, for instance homomorphic encryption [178], *k*-anonymity [179], and DP [180]. Gu et al. [121] considered the data dissemination between the user and the edge

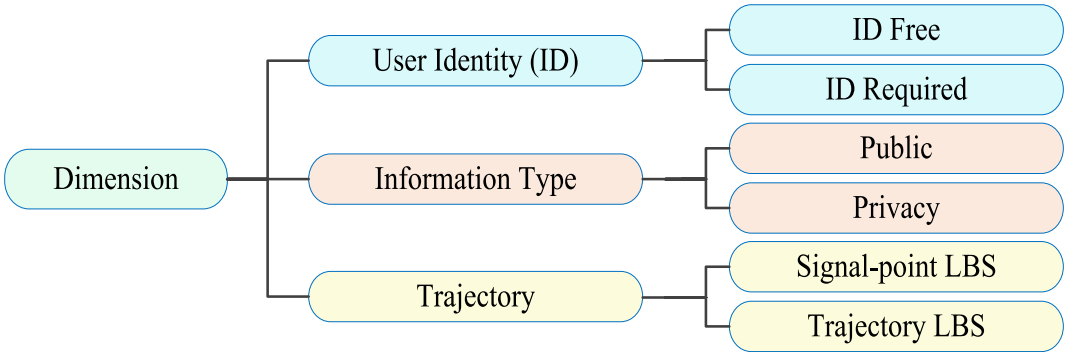

Fig. 10. The dimension classification of the LBS

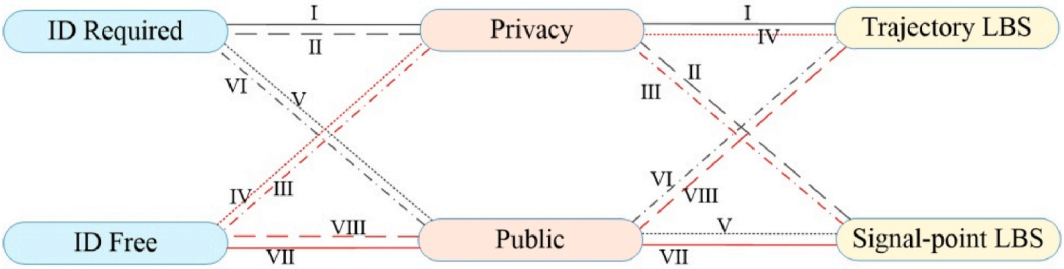

Fig. 11. The categories of the LBS

layer. They proposed the Markov decision process-based a dynamic customisable model to protect privacy. Sharma et al. [142] proposed a method that implements DP using a fuzzy convolution neural network (FCNN) with a Laplace mechanism for injecting noise to the privacy leakage at the edge layer.

In summary, for edge node privacy, most research work considered the tree structure of the node in the edge computing environment. The existing method for DP demonstrates that the smaller the privacy budget is, the higher the privacy level. However, the noise increases as the privacy budget decreases. Therefore, some researchers proposed the privacy threshold to control the mean absolute error (MAS). For edge layer privacy, first, the scalability measures the system/framework to support a building number of input data. Second, the processing time measures the individual datasets in the system/framework to evaluate the performances. Finally, the accuracy evaluates the efficiency of DP with the privacy budget value. A smaller privacy budget will normally lead to lower accuracy.

6.7. Summary

In recent years, DP-protected data privacy in the edge computing environment has made great developments. We summarised and compared the application of DP in data transmitting, data processing, model training, data publishing, location/position trajectory, and others, as shown in Table 9.

Table 9 presents a comparison of existing DP methods applied to edge computing based on application scenario, privacy mechanism/algorithm, type of noise, type of privacy, and method of adding noise. In addition, we introduce the concept of noise distribution function, which is a probability distribution function used to add noise while minimizing data distortion and protecting privacy. Differential privacy noise distribution functions, such as Laplacian and Gaussian distributions, can regulate the size and distribution of noise, achieving a balance between privacy protection and data accuracy. The choice of noise distribution function should consider the data type and application scenario.

The protection of edge computing-based smart city application scenarios with DP has been implemented by many researchers. Nevertheless, a great number of applications for the edge computing environment still need considerable attention. For example, edge intelligence is a new trend in the edge computing environment. Analogously, DP is combined with artificial intelligence algorithms or applications to ensure the privacy of the smart city. In addition, lightweight and less complex DP algorithms require fitting into the edge computing environment. In summary, DP is an effective solution for the edge computing environment, but researchers need more efforts to address all applications in smart city scenarios based on edge computing.

This paper regarding the application of DP for edge computing environments in smart city scenarios contains experiments to evaluate the performance of the mechanism or algorithm. Then, the platforms/tools, common datasets, and evaluation metrics for the articles are summarised in Table 8, where * represents undefined or unknown. The common datasets contain the name, type, and number. As illustrated in Table 8, the evaluation metrics generally contain accuracy, efficiency, utility, and privacy.

Through the summary of existing studies, we can see that DP is one of the best solutions to privacy problems in smart city applications. However, there are still many deficiencies in the environment of edge computing, which are discussed in Section 7. For example, the parameters are adjusted to improve the data utility according the datasets size. The DP algorithm cannot provide effective privacy preservation because the privacy budget cannot be adjusted. In addition, untrusted edge devices with intentionally or unintentionally leaking privacy are considered.

7. Future directions

Despite the advantages of differential privacy (DP) in edge computing, there are still several challenges that need to be addressed. One of the reasons for these challenges is the frequent operations involved in edge computing, such as data transmission, sharing, and collaborative analysis [181]. In the context of smart cities, it is particularly important to address challenges related to data anomaly detection and defense against data attacks in edge computing environments. These challenges remain an active area of research in DP for edge computing, and new solutions are needed to ensure effective data privacy and security in this context. For example, in January 2019, the server used to store data from the Oklahoma Department of Securities data without sufficient protection faced a serious data breach by the cybersecurity company, UpGuard [182]. The leaked data included up to 3 TB of data, containing millions of sensitive government documents and FBI investigation reports. In the same year, the American Banking System (ABS) was attacked by Avaddon's extortion request [183]. Although its customers were not directly attacked, the attackers may have obtained customer data and access rights through ABS. Therefore, the future research directions of data anomaly detection, data attack defence and others for implementation with DP for edge computing environments in smart city scenarios are discussed in this section.

This section is based on the IC5 mentioned in Section 2:

IC5: Future research directions of DP methods in edge computing environments.

The future research directions of differential privacy methods in the environment of edge computing can be presented as data anomaly detection, data attack defence, and others.

7.1. Anomaly detection

Anomaly detection can collect and analyse users' data for abnormal behaviours in real-time. The application applies in many scenarios, such as time-series data monitoring, fraud in the financial context, data anomaly in feature engineering, data anomaly in the ELT process, etc. Nonetheless, data anomaly detection mainly contains multi-sensor data anomaly detection and network anomaly detection in the edge computing environment.

In recent years, Fan et al. [184] considered the data that needed to be transformed before the release for privacy preservation, and they proposed the framework for anomaly detection with DP. Du et al. [185]

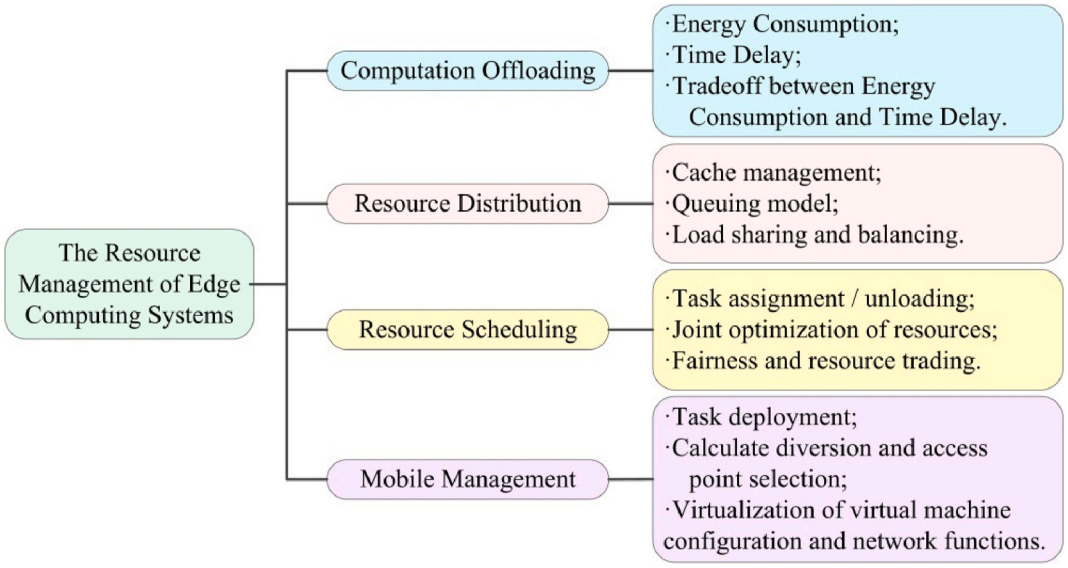

Fig. 12. The resource management of edge computing systems.

extended DP to detect poisoning samples in backdoor attacks. Randa Aliably et al. [186] proposed a privacy-preserving model that uses reconstructed data to classify user activity and detect abnormal network behaviour.

In smart city scenarios, a massive amount of data is being generated and used every day, such as trajectory data, healthcare data, business data, financial records, etc. To better develop smart cities, data anomaly detection is necessary. We believe that to solve the issue of data anomalies in the environment of edge computing, DP is a suitable solution. The reason for this is that DP contains rigorous mathematical modelling, and the method can provide a desirable level of privacy for edge computing environments in smart city scenarios.

7.2. Data attack defence

In smart city scenarios, the personal privacy information are regarded as target with an increasing number of attacks, such as data poisoning attacks [187,188]. Data poisoning attacks can manipulate the learning model to modify the training datasets.

In recent years, Ma et al. [189] considered the data poisoning attacks can be defended by DP method. They showed that private learners are resistant to that attack when the attackers are only able to poison a small number of datasets. Shadi Rahimian et al. [190] protected the machine learning model by adding noise to the gradients and as an alternative method. They evaluated the effect of two DP techniques against membership inference attacks.

In smart cities, the research direction of data attack defence has great potential and needs to be further explored. For example, the gap between the theoretical limit and the empirical performance remains an issue. We believe that the DP method can enhance the development of smart cities and can produce optimal services in smart cities.

7.3. Data storage/governance

In edge computing-based smart city application scenarios, the end device that collects data needs storage or governance.

Blockchain technology is an alternative for constructing transparent security to store or govern data. However, blockchain is a type of decentralised ledger storage system that contains tamper-proofing. It is ambiguous to develop a solution method for edge computing and blockchain for smart city applications. Gai et al. [191] proposed an implemented method for DP in blockchain systems to prevent information on blocks from data mining attacks. They introduce DP to protect the identity information of edge nodes. Sun et al. [192] designed a double disturbance localised DP algorithm to perturb the workers' location information in the transparent mechanism of blockchain. To ensure authenticity and high income in auction mechanisms, privacy and security are crucial factors to consider. Guo et al. [193] designed a differential private portfolio dual auction mechanism for edge computing platforms, where IoT devices request resource packs and edge nodes compete to provide them, maximizing revenue while ensuring privacy. Lv et al. [194] mapped the complex physical space of CPS in smart cities into virtual space and implemented differential privacy frequent subgraph-mostly regraph to secure data privacy. However, despite these efforts, DP research in edge computing still faces challenges due to frequent operations like data transmission, sharing, and task collaborative analysis, especially in scenarios where data anomaly detection and attack defense for edge computing environments are critical.

The DP mechanism protects users' privacy data by deploying disturbance mechanisms. The research direction of data storage or governance can be considered from DP and blockchain technology in edge computing-based smart city applications.

7.4. Data sharing

In smart city applications, the edge nodes need to share data effectively, although the communication between each edge node is unreliable. However, there is a series of sensitive information that leads to data owners being unwilling to share. Furthermore, users are increasingly concerned about sensitive personal information in smart city applications.

Federated learning is a collaborative method for achieving global model training without sharing any raw client data. It involves training a local machine learning model by the data owner and updating the model to an aggregator for collection and averaging. One of the advantages of federated learning is that it can handle heterogeneous local datasets that may be non-independent, identically distributed, and unbalanced among various participants [4,181,195]. Lu et al. [196] designed a data sharing scheme to preserve the devices privacy in the industrial Internet of Things. For their viewpoint, the data-sharing issue is identified as machine-learning issue by incorporating DP and federated learning. The model-sharing is parameter-sharing of the machine model instead of revealing the raw data of participant. Zhang et al. [197] proposed a

medical data privacy protection framework to protect privacy by adding DP noise into federated learning. They considered the data privacy of patients in the smart health application. Wang et al. [198] designed a control algorithm to solve the issues of low resources for the federated learning model in IoT devices.

In the smart city application scenarios, the DP mechanism can be applied in the model training stage of federated learning technology to conceal the users' private data. However, it is a key issue to improve the practicability of data models mapped from raw data in data sharing.

7.5. Resource management

In the edge computing environment, the core infrastructure provides network access and centralised cloud computing services functions for devices at the edge. However, the core infrastructures are not fully trusting in many cases, and thus, there is a high possibility of security-threatening attacks, including privacy disclosure, tampering with data, DoS (denial of service) attacks, and service manipulation. Therefore, the privacy of the core infrastructure is a research direction with DP methods for smart city application scenarios based on edge computing.

The edge data centre is one of the main components in edge computing and takes charge of virtualisation services and multiple management services. Nevertheless, edge data centres on some security issues, for instance the attacks of physical, privacy breaches, service manipulation, and data tampering. Therefore, the research of data security and privacy protection technology for the smart city application scenarios is important.

The efficient allocation of resources is a critical research direction for edge computing systems as it not only improves user experience but also optimizes bandwidth resources. However, the process of resource distribution presents a significant challenge. Resource distribution falls under the umbrella of resource management for edge computing systems, as illustrated in Fig. 12. To address this challenge, Wang et al. [137] proposed a K-neighbour joint optimisation algorithm for task offloading and resource distribution using a histogram with local differential privacy [199]. The method can protect the privacy of connected vehicle and reduce the effect of the task offloading algorithm. In future work, research direction should focus on managing privacy for resource distribution and proposing optimal privacy algorithms for the different tasks of resource management.

8. Conclusion

With the advancement of edge computing, smart city applications have become an essential part of our daily life, ranging from manual operation to automatic operation and first generation wireless cellular technology to 5th generation mobile communication technology (5G). The advantages of edge computing technology, such as the sinking of resources, ultralow latency, high bandwidth, and high real-time computing power, make it a wide framework in smart city scenarios. Nevertheless, massive data are involved in data transmission, data processing, model training, and data publishing to provide better services in smart city scenarios. However, important privacy and security issues are raising great concerns at the same time.

In this paper, we present a comprehensive summarisation and thorough comparisons of the existing DP methods for edge computing-based smart city applications. Our coverage of DP in edge computing is extensive, encompassing a wide range of application areas including data transmission, data processing, model training, data publishing, and location privacy. Compared with other data privacy protection solutions, the DP mechanism can achieve better performance in the energy cost, data types, latency, computing cost, accuracy, processing time, data utility, and model performance.

Finally, we conclude this study by identifying some future directions for the application of DP in edge computing-based smart city applications, such as data anomaly detection, data attack defence, data storage/governance, data sharing, core infrastructure, data centres, and resource distribution. We believe that this study provides a comprehensive overview for researchers and practitioners who are interested in preserving privacy for edge computing-based smart city applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Natural Science Foundation of China Project (No.61972001, No.62076002).

References

[1] Batty M, Axhausen KW, Giannotti F, Pozdnoukhov A, Bazzani A, Wachowicz M, et al. Smart cities of the future. Eur Phys J Spec Top 2012;214:481–518.
[2] Ma Q, Huang H, Zhang W, Qiu M. Design of smart home system based on collaborative edge computing and cloud computing. In: Presented at the international conference on algorithms and architectures for parallel processing; 2020.
[3] Chen M, Li W, Hao Y, Qian Y, IztokHumar. Edge cognitive computing based smart healthcare system. Future Generat Comput Syst 2018;86:403–11.
[4] Wang R, Lai J, Zhang Z, Li X, Vijayakumar P, Karuppiah M. Privacy-preserving federated learning for internet of medical things under edge computing. IEEE Journal of Biomedical and Health Informatics 2022:854–65.
[5] Qi Q, Tao F. A smart manufacturing service system based on edge computing, fog computing and cloud computing. IEEE Access 2019;7:86769–77.
[6] Wan S, Xu X, Wang T, Gu Z. An intelligent video analysis method for abnormal event detection in intelligent transportation systems. IEEE Trans Intell Transport Syst 2021;22:4487–95.
[7] Rao PM, Deebak B. Security and privacy issues in smart cities/industries: technologies, applications, and challenges. J Ambient Intell Hum Comput 2022: 1–37.
[8] Khanh QV, Nguyen V-H, Minh QN, Van AD, Anh NL, Chehri A. An efficient edge computing management mechanism for sustainable smart cities. Sustainable Computing: Informatics and Systems 2023;38:100867.
[9] Rajarajeswari S, Hema N. Edge computing in intelligent IoT. In: Convergence of deep learning and internet of Things: computing and technology. IGI Global; 2023. p. 157–81.
[10] Abbas N, Zhang Y, Taherkordi A, Skeie T. Mobile edge computing: a survey. IEEE Internet Things J 2017;5:450–65.
[11] Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: vision and challenges. IEEE Internet Things J 2016;3:637–46.
[12] Kumar A, Upadhyay A, Mishra N, Nath S, Yadav KR, Sharma G. Privacy and security concerns in edge computing-based smart cities. In: Robotics and AI for cybersecurity and critical infrastructure in smart cities. Springer; 2022. p. 89–110.
[13] Silva TPD, Batista T, Lopes F, Neto AR, Delicato FC, Pires PF, et al. Fog computing platforms for smart city applications: a survey. ACM Trans Internet Technol 2022; 22:1–32.
[14] Neves F, Souza R, Sousa J, Bonfim M, Garcia V. Data privacy in the Internet of Things based on anonymization: a review. J Comput Secur 2023:1–31.
[15] Jernigan C, Mistree BF. Gaydar: Facebook friendships expose sexual orientation. First Monday; 2009.
[16] Gupta P, Gottipati S, Jiang J, Gao D. Your love is public now: questioning the use of personal information in authentication. In: Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security; 2013. p. 49–60.
[17] Li Y, Tao X, Zhang X, Wang M, Wang S. Break the data barriers while keeping privacy: a graph differential privacy method. IEEE Internet Things J 2022: 3840–50.
[18] Du M, Wang K, Xia Z, Zhang Y. Differential privacy preserving of training model in wireless big data with edge computing. IEEE Transactions on Big Data 2020;6: 283–95.
[19] Bi M, Wang Y, Cai Z, Tong X. A privacy-preserving mechanism based on local differential privacy in edge computing. China Communications 2020;17:50–65.
[20] Zhang X, Chen Q, Peng X, Jiang X. Differential privacy-based indoor localization privacy protection in edge computing. In: IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing; 2019. p. 491–6. Internet of People and Smart City Innovation.

[21] Yan X, Wu Q, Sun Y. A homomorphic encryption and privacy protection method based on blockchain and edge computing. Wireless Commun Mobile Comput 2020:1–9. 2020.
[22] Hu F, Chen B. Channel coding scheme for relay edge computing wireless networks via homomorphic encryption and NOMA. IEEE Transactions on Cognitive Communications and Networking 2020;6:1180–92.
[23] Jolfaei A, Kant K. Data security in multiparty edge computing environments. In: Government microcircuit applications & critical technology conference; 2019. p. 17–22. United States.
[24] Wu J, Li Y, Ren F, Yang B. Robust and auditable distributed data storage with scalability in edge computing. Ad Hoc Netw 2021;117:1–14.
[25] Dwork C. Differential privacy. In: Proceedings of the 33rd international conference on automata, languages and programming - volume Part II; 2006. p. 1–12.
[26] Zhu T, Yu PS. Applying differential privacy mechanism in artificial intelligence. In: IEEE 39th international conference on distributed computing systems. ICDCS); 2019. p. 1601–9. 2019.
[27] Triastcyn A, Faltings B. Bayesian differential privacy for machine learning. ICML; 2020. p. 1–12.
[28] Zhang T, Zhu T, Xiong P, Huo H, Tari Z, Zhou W. Correlated differential privacy: feature selection in machine learning. IEEE Trans Ind Inf 2020;16:2115–24.
[29] Friedman A, Schuster A. Data mining with differential privacy. In: Presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining; 2010. New York, NY, USA.
[30] Sun X, Xu R, Wu L, Guan Z. A differentially private distributed data mining scheme with high efficiency for edge computing. J Cloud Comput 2021;10:1–12.
[31] Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I. Deep learning with differential privacy. In: Presented at the proceedings of the 2016 ACM SIGSAC conference on computer and communications security; 2016.
[32] Bu Z, Dong J, Long Q, Su W. Deep learning with Gaussian differential privacy. Harvard data science review 2020;2020:1–48.
[33] Papernot N, Thakurta A, Song S, Chien S, Erlingsson Ú. Tempered sigmoid activations for deep learning with differential privacy. In: AAAI technical track on machine learning III; 2021. p. 1–10.
[34] Jiang H, Pei J, Yu D, Yu J, Gong B, Cheng X. Applications of differential privacy in social network analysis: a survey. IEEE Trans Knowl Data Eng 2021:1–20.
[35] Hassan MU, Rehmani MH, Chen J. Differential privacy techniques for cyber physical systems: a survey. IEEE Communications Surveys & Tutorials 2020;22:746–89.
[36] Gong M, Xie Y, Pan K, Feng K, Qin AK. A survey on differentially private machine learning. IEEE Comput Intell Mag 2019;15.
[37] Ji Z, Lipton ZC, Elkan C. Differential Privacy and Machine Learning: Surv Rev 2014;7584:1–30. arXiv preprint vol. abs/1412.
[38] Huang W, Zhou S, Zhu T, Liao Y. Improving utility of differentially private mechanisms through cryptography-based technologies: a survey. 2020. p. 1–18. arXiv:2011.00976.
[39] Leoni D. Non-interactive differential privacy: a survey. Association for Computing Machinery; 2012. p. 40–52.
[40] Fletcher S, Islam MZ. Decision tree classification with differential privacy: a surveyvol. 52. Association for Computing Machinery; 2016.
[41] Shaikh A, Patil S. A survey on privacy enhanced role based data aggregation via differential privacy. In: Presented at the 2018 international conference on advances in communication and computing technology. India: ICACCT), Sangamner; 2018.
[42] Dwork C. Differential privacy : a historical survey. Springer-Verlag Berlin Heidelberg; 2008. p. 1–19. 2008.
[43] Dankar FK, Emam KE. Practicing differential privacy in health care: a review. Transactions on Data Privacy 2013;6:35–67.
[44] Dwork C, Smith A. Differential privacy for statistics: what we know and what we want to learn. Journal of Privacy and Confidentiality 2010;1:135–54.
[45] Dwork C. Differential privacy: a survey of results. In: Theory and applications of models of computation; 2008. p. 1–19.
[46] Yao X, Zhou X, Ma J. Differential privacy of big data: an overview. In: Presented at the 2016 IEEE 2nd international conference on big data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing (HPSC), and IEEE international conference on intelligent data and security. New York, NY, USA: IDS); 2016.
[47] Xiong P, Zhu T, Wang X. A survey on differential privacy and applications. Chin J Comput 2014;37:101–22.
[48] Ding L-p, Lu G-q. Survey of differential privacy in frequent pattern mining. J Commun 2014;35:200–9.
[49] Cheu A. Differential Privacy in the Shuffle Model: A Survey of Separations 2021; abs/2107.11839:1–21. ArXiv.
[50] Hu YC, Patel M, Sabella D, Sprecher N, Young V. Mobile edge computing A key technology towards 5G. ETSI white paper 2017;11:1–16.
[51] Hayes B. Cloud computing. 2008. p. 9–11.
[52] Yi S, Hao Z, Qin Z, Li Q. Fog computing: platform and applications. In: Presented at the 2015 third IEEE workshop on hot topics in web systems and technologies. Washington, DC, USA: HotWeb); 2015.
[53] Lee I, Lee K. The internet of things (IoT): applications, investments, and challenges for enterprises. Bus Horiz 2015;58:431–40.
[54] El-Sayed H, Sankar S, Prasad M, Puthal D, Gupta A, Mohanty M, et al. Edge of things: the big picture on the integration of edge, IoT and the cloud in a distributed computing environment. IEEE Access 2018;6:1706–17.
[55] Feng J, Yang LT, Ren B, Zou D, Dong M, Zhang S. Tensor recurrent neural network with differential privacy. IEEE Trans Comput 2023:1–11.
[56] Miao Q, Jing W, Song H. Differential privacy based location privacy enhancing in edge computing. Concurrency Comput Pract Ex 2019;31:1–13.
[57] Liu G, Tang Z, Wan B, Li Y, Liu Y. Differential privacy location data release based on quadtree in mobile edge computing. Transactions on Emerging Telecommunications Technologies 2020:1–17.
[58] Jing W, Miao Q, Song H, Chen X. Data loss and reconstruction of location differential privacy protection based on edge computing. IEEE Access 2019;7:75890–900.
[59] Nie X, Yang LT, Feng J, Zhang S. Differentially private tensor train decomposition in edge-cloud computing for SDN-based internet of things. IEEE Internet Things J 2020;7:5695–705.
[60] Wang T, YaxinMei, Jia W, Zheng X, Wang Guojun, MandeXi. Edge-based differential privacy computing for sensor-cloud systems. J Parallel Distr Comput 2020;136:75–85.
[61] Zhou P, Wang K, Xu J, Wu D. Differentially-private and trustworthy online social multimedia big data retrieval in edge computing. IEEE Trans Multimed 2019;21:539–54.
[62] Guo J, Wu W. Differential Privacy-Based Online Allocations towards Integrating Blockchain and Edge Computing 2021;abs/2101:1–12. 02834.
[63] Ezabadi SG, Jolfaei A, Kulik L, Kotagiri R. Differentially private streaming to untrusted edge servers in intelligent transportation system. In: Presented at the 2019 18th IEEE international conference on trust, security and privacy in computing and communications/13th IEEE international conference on big data science and engineering. Rotorua, New Zealand: TrustCom/BigDataSE); 2019.
[64] Sun Y, He Q, Qi L, Rafique W, Dou W. DPODA: differential privacy-based online double auction for pervasive edge computing resource allocation. In: Presented at the proceedings of the 2nd ACM international symposium on blockchain and secure critical infrastructure; 2020. New York, NY, USA.
[65] Zhang J, Zhao Y, Wang J, Chen B. FedMEC: improving efficiency of differentially private federated learning via mobile edge computing. In: Mobile networks and applications; 2020. p. 2421–33.
[66] Zhang J, Wang J, Zhao Y, Chen B. An efficient federated learning scheme with differential privacy in mobile edge computing. In: Machine learning and intelligent communications; 2019. p. 538–50.
[67] Nair AK, Sahoo J, Raj ED. Privacy preserving Federated Learning framework for IoMT based big data analysis using edge computing. Computer Standards & Interfaces; 2023, 103720.
[68] Dwork C, Roth A. The algorithmic foundations of differential privacy. Found Trends® Theor Comput Sci 2014;9:211–407.
[69] Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. J. Priv. Confidentiality 2006;3876:265–84.
[70] Geng Q, Kairouz P, Oh S, Viswanath P. The staircase mechanism in differential privacy. IEEE Journal of Selected Topics in Signal Processing 2015;9:1176–84.
[71] Yang J, Xiang L, Li W, Liu W, Wang X. In: Improved matrix Gaussian mechanism for differential privacy. abs/2104; 2021. p. 1–10. 14808.
[72] Lobo-Vesga E, Russo A, Gaboardi M. A programming framework for differential privacy with accuracy concentration bounds. In: Presented at the 2020 IEEE symposium on security and privacy. San Francisco, CA, USA: SP); 2020.
[73] He D, Chan S, Qiao Y, Guizani N. Imminent communication security for smart communities. IEEE Commun Mag 2018;56:99–103.
[74] Alsaffar N, Ali H, Elmedany W. Smart transportation system: a review of security and privacy issues. In: Presented at the 2018 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT). Bahrain: Sakhier; 2018.
[75] Alaba FA, Othman M, Hashem IAT, Alotaibi F. Internet of Things security: a survey. J Netw Comput Appl 2017;88:10–28.
[76] Abbas AW, Marwat SNK, Ahmed S, Hafeez A, Ullah K, Khan IU. Proposing model for security of IoT devices in smart logistics: a review. In: Presented at the 2020 3rd international conference on computing, mathematics and engineering technologies. Pakistan: iCoMET), Sukkur; 2020.
[77] Yao A, Jiang F, Li X, Dong C, Xu J, Xu Y, et al. A novel security framework for edge computing based UAV delivery system. In: Presented at the the 20th IEEE international conference on trust, security and privacy in computing and communications. TrustCom; 2021. p. 2021.
[78] Singh A, Chatterjee K. Securing smart healthcare system with edge computing. Comput Secur 2021;108:1–23.
[79] Trimananda R, Younis A, Wang B, Xu B, Demsky B, Xu G. Vigilia: securing smart home edge computing. In: Presented at the 2018 IEEE/ACM symposium on edge computing. Seattle, WA, USA: SEC); 2018.
[80] Times G. 30 suspects arrested in China for selling 600 mln pieces of personal information. 2021. Available: https://www.globaltimes.cn/page/202101/1213922.shtml?id=11.
[81] Waqas. Data analytics firm Polecat data breach – 30TB of data exposed. 2021. Available: https://www.hackread.com/polecat-data-analytics-data-breach-30tb-data-exposed/.
[82] Duan Y, Lu Z, Zhou Z, Sun X, Wu J. "Data privacy protection for edge computing of smart city in a DIKW architecture,". Eng Appl Artif Intell 2019;81:323–35.
[83] Qu Y, Yu S, Zhou W, Peng S, Wang G, Xiao K. Privacy of things: emerging challenges and opportunities in wireless internet of things. IEEE Wireless Commun 2018;25:91–7.
[84] Sweeney L, Anonymity k-. A model for protecting privacy,". Int J Uncertain Fuzziness Knowledge-Based Syst 2002;10:557–70.
[85] Li N, Li T, Venkatasubramanian S. t-Closeness: privacy beyond k-Anonymity and l-Diversity. In: 2007 IEEE 23rd international conference on data engineering. Turkey: Istanbul; 2007. presented at the.

[86] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. International Conference on Data Engineering 2006;1:1–52.
[87] Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. IEEE Trans Inf Forensics Secur 2017;11:2594–608.
[88] Liu P. Public-key encryption secure against related randomness attacks for improved end-to-end security of cloud/edge computing. IEEE Access 2020;8:16750–9.
[89] Li Y, Dong Z, Sha K, Jiang C, Wan J, Wang Y. TMO: time domain outsourcing attribute-based encryption scheme for data acquisition in edge computing. IEEE Access 2019;7:40240–57.
[90] Pagliery J. Adult FriendFinder probing claims of a second hack. Available: https://money.cnn.com/2016/11/14/technology/adult-friend-finder-hack/index.html; 2016.
[91] Armasu L. Data breach exposes 150 million MyFitnessPal accounts. 2018. Available: https://www.tomshardware.com/news/myfitnesspal-data-breach-150-million,36782.html.
[92] Fox C. Marriott hack hits 500 million Starwood guests. 2018. Available: https://www.bbc.com/news/technology-46401890.
[93] Staff AaT. Data from 533 million Facebook accounts posted. 2019. https://www.timesofisrael.com/data-from-533-million-facebook-accounts-posted-online/.
[94] Marzouk Z. June 16). Alibaba data breach exposes 1.1 billion pieces of data. Available:. 2019. https://www.itpro.co.uk/security/data-breaches/359897/alibaba-data-breach-exposes-11-billion-pieces-of-data.
[95] Cimpanu C. Hacker selling data of 538 million Weibo users. Available. 2020. https://www.zdnet.com/article/hacker-selling-data-of-538-million-weibo-users/.
[96] Lugo J. Bitglass Security Spotlight: Over 15 Billion Usernames and Passwords Are Now Available on the Dark Web. 2020. Available: https://www.bitglass.com/blog/bitglass-security-spotlight15-billion-usernames-and-passwords-available-on-dark-web.
[97] Gatlan S. Hacker leaks full database of 77 million Nitro PDF user records 2021. Available: https://www.bleepingcomputer.com/news/security/hacker-leaks-full-database-of-77-million-nitro-pdf-user-records/.
[98] Cimpanu C. Hackers leak LinkedIn 700 million data scrape. 2021. Available: https://therecord.media/hackers-leak-linkedin-700-million-data-scrape/.
[99] Kost E. 13 biggest healthcare data breaches. 2022 (Updated June 2022). Available: https://www.upguard.com/blog/biggest-data-breaches-in-healthcare.
[100] Wang T, Qiu L, Sangaiah AK, Liu A, Bhuiyan MZA, Ma Y. Edge-computing-based trustworthy data collection model in the internet of things. IEEE Internet Things J 2020;7:4218–27.
[101] Ren Y, Leng Y, Cheng Y, Wang J. Secure data storage based on blockchain and coding in edge computing. Math Biosci Eng 2019;16:1874–92.
[102] Cao K, Liu Y, Meng G, Sun Q. An overview on edge computing research. IEEE Access 2020;8:85714–28.
[103] Gheisari M, Pham Q-V, Alazab M, Zhang X, Fernández-Campusano C, Srivastava G. ECA: an edge computing architecture for privacy-preserving in IoT-based smart city. IEEE Access 2019;7:155779–86.
[104] Adam NR, Worthmann JC. Security-control methods for statistical databases: a comparative study. ACM Comput Surv 1989;21:515–56.
[105] Samarati P, Sweeney L. Generalizing data to provide anonymity when disclosing information (abstract). In: Pods '98: proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database; 1998. p. 1–15.
[106] Pinkas B. Cryptographic techniques for privacy-preserving data mining. SIGKDD Explor 2002;4:12–9.
[107] Clifton C, Kantarcioglu M, Vaidya J, Lin X, Zhu MY. Tools for privacy preserving distributed data mining. ACM SIGKDD Explorations Newsletter 2002;4:28–34.
[108] Andrienko G, Andrienko N, Giannotti F, Monreale A, Pedreschi D. Movement data anonymity through generalization. SPRINGL '09: Proceedings of the 2nd SIGSPATIAL ACM GIS 2009 International Workshop on Security and Privacy in GIS and LBS 2009;3:91–121.
[109] Fischer F. Participatory governance: from theory to practice. 2012.
[110] Al-Zobbi M, Shahrestani S, Ruan C. Implementing A framework for big data anonymity and analytics access control. Sydney, NSW, Australia: IEEE Trustcom/BigDataSE/ICESS; 2017. presented at the 2017.
[111] Wong RC-W, Li J, Fu AW-C, Wang K. (α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing. In: Kdd '06: proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; 2006. p. 754–9. Philadelphia, PA, USA.
[112] Zhao C, Zhao S, Zhao M, Chen Z, Gao C-Z, Li H, et al. Secure multi-party computation: theory, practice and applications. Inf Sci 2019;476:357–72.
[113] Kadhe S, Garcia B, Heidarzadeh A, Rouayheb SE, Sprintson A. Private information retrieval with side information. IEEE Trans Inf Theor 2020;66:2032–43.
[114] Tang Z, Wei Z, Wang C, Yang Z, Xu Y, Cui S. A data desensitization algorithm for privacy protection electric power industry. In: IOP conference series: materials science and engineering; 2020, 052059.
[115] Wang T, Ke H, Zheng X, Wang K, Sangaiah AK, Liu A. Big data cleaning based on mobile edge computing in industrial sensor-cloud. IEEE Trans Ind Inf 2020;16:1321–9.
[116] Chor B, Goldreich O, Kushilevitz E, Sudan M. Private information retrieval. In: Presented at the proceedings of IEEE 36th annual foundations of computer science; 1995. Milwaukee, WI, USA.
[117] Xiang N, Zhang X, Dou Y, Xu X, Yang K, Tan Y. High-end equipment data desensitization method based on improved Stackelberg GAN. Expert Syst Appl 2021;180:1–11.
[118] Guo J, Zong Y, Chen F, Guo C, Xie D. Data cleaning algorithm based on body area network. In: International conference on artificial intelligence and security; 2020. p. 575–85.
[119] Zhang P, Durresi M, Durresi A. Mobile privacy protection enhanced with multi-access edge computing. In: Presented at the 2018 IEEE 32nd international conference on advanced information networking and applications (AINA). Poland: Krakow; 2018.
[120] Gai K, Qiu M, Xiong Z, Liu M. Privacy-preserving multi-channel communication in Edge-of-Things. Future Generat Comput Syst 2018;85:190–200.
[121] Gu B, Gao L, Wang X, Qu Y, Jin J, Yu S. Privacy on the edge: customizable privacy-preserving context sharing in hierarchical edge computing. IEEE Transactions on Network Science and Engineering 2020;7:2298–309.
[122] Xu X, Liu X, Yin X, Wang S, Qi Q, Qi L. Privacy-aware offloading for training tasks of generative adversarial network in edge computing. Inf Sci 2020;532:1–15.
[123] Chen B, Leahy K, Jones A, Hale M. Differential privacy for symbolic systems with application to Markov chains. Automatica 2023;152:110908.
[124] Gai K, Wu Y, Zhu L, Zhang Z, Qiu M. Differential privacy-based blockchain for industrial internet-of-things. IEEE Trans Ind Inf 2020. presented at the.
[125] Iqbal R, Doctor F, More B, Mahmud S, Yousuf U. Big data analytics: computational intelligence techniques and application areas. Technol Forecast Soc Change 2020;153:1–25.
[126] Kong L, Zhang D, He Z, Xiang Q, Wan J, Tao M. Embracing big data with compressive sensing: a green approach in industrial wireless networks. IEEE Commun Mag 2016;54:53–9.
[127] Ponnusamy C, R. D., P. V., A. J. S. V., B. B.. Data security and privacy requirements in edge computing: a systemic review. In: Cases on edge computing and analytics; 2021. p. 171–87.
[128] Zhang J, Chen B, Zhao Y, Cheng X, Hu F. Data security and privacy-preserving in edge computing paradigm: survey and open issues. IEEE Access 2018;6:18209–37.
[129] Feng J, Yang LT, Zhang R. Practical privacy-preserving high-order Bi-lanczos in integrated edge-fog-cloud architecture for cyber-physical-social systems. ACM Trans Internet Technol 2019;19:1–18.
[130] TitanWolf. From the 12306 information leak, understand what hackers hit the library and drag the library to wash the library. Available:. 2018. https://titanwolf.org/Network/Articles/Article?AID=b5373fd4-6ffc-4ed1-9c93-6fe85fd6ecc4.
[131] Zhang Q, Yang LT, Chen Z, Li P. PPHOPCM: privacy-preserving high-order possibilistic c-means algorithm for big data clustering with cloud computing. IEEE Transactions on Big Data 2017:1–11.
[132] Feng J, Yang LT, Gati NJ, Xie X, Gavuna BS. Privacy-preserving computation in cyber-physical-social systems: a survey of the state-of-the-art and perspectives. Inf Sci 2020;527:341–55.
[133] Zhang Q, Yang LT, Chen Z, Li P, Deen MJ. Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning. IEEE Internet Things J 2018;5:2896–903.
[134] Shafagh H, Hithnawi A, Burkhalter L, Fischli P, Duquennoy S. Secure sharing of partially homomorphic encrypted IoT data. In: Presented at the proceedings of the 15th ACM conference on embedded network sensor systems; 2017. Delft, Netherlands.
[135] Ma Z, Liu Y, Liu X, Ma J, Li F. Privacy-preserving outsourced speech recognition for smart IoT devices. IEEE Internet Things J 2019;6:8406–20.
[136] Shen M, Ma B, Zhu L, Du X, Xu K. Secure phrase search for intelligent processing of encrypted data in cloud-based IoT. IEEE Internet Things J 2019;6:1998–2008.
[137] Wang S, Li J, Wu G, Chen H, Sun S. Joint optimization of task offloading and resource allocation based on differential privacy in vehicular edge computing. IEEE Transactions on Computational Social Systems 2021:1–11.
[138] Chen C, Liu Z, Wan S, Luan J, Pei Q. Traffic flow prediction based on deep learning in internet of vehicles. IEEE Trans Intell Transport Syst 2021;22:3776–89.
[139] Guan Z, Liu X, Wu L, JunWu, Xu R, Zhang J, et al. Cross-lingual multi-keyword rank search with semantic extension over encrypted data. Inf Sci 2020;514:523–40.
[140] Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. ACM Transactions on Intelligent Systems and Technology 2019;10:1–19.
[141] Guan Z, Lu X, Yang W, Wu L, Wang N, Zhang Z. Achieving efficient and Privacy-preserving energy trading based on blockchain and ABE in smart grid. J Parallel Distr Comput 2021;147:34–45.
[142] Sharma J, Kim D, Lee A, Seo D. On differential privacy-based framework for enhancing user data privacy in mobile edge computing environment. IEEE Access 2021;9:38107–18.
[143] Zeydan E, Ba¸stug E, Bennis M, Kader MA, Karatepe IA, Er AS, et al. Big data caching for networking: moving from cloud to edge. IEEE Commun Mag 2016;54:36–42.
[144] Klarreich E, News SS. Privacy by the numbers: a new approach to safeguarding data. Quanta Magazine 2012;10:1–15.
[145] Xu X, Wu J, Yang M, Luo T, Duan X, Li W, et al. Information leakage by model weights on federated learning. In: Proceedings of the 2020 workshop on privacy-preserving machine learning in practice; 2020. p. 31–6.
[146] He Z, Zhang T, Lee RB. Model inversion attacks against collaborative inference. In: Proceedings of the 35th annual computer security applications conference; 2019.
[147] Ji Y, Zhang X, Ji S, Luo X, Wang T. Model-reuse attacks on deep learning systems. In: Proceedings of the 2018 ACM SIGSAC conference on computer and communications security; 2018. p. 349–63.

[148] Zhang Y, Jia R, Pei H, Wang W, Li B, Song D. The secret revealer: generative model-inversion attacks against deep neural networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 253–61. 2020.
[149] Hay M, Rastogi V, Miklau G, Suciu D. Boosting the accuracy of differentially private histograms through consistency. Proceedings of the VLDB Endowment 2010;3:1021–32.
[150] Mao Y, Yi S, Li Q, Feng J, Xu F, Zhong S. Learning from differentially private neural activations with edge computing. In: Presented at the 2018 IEEE/ACM symposium on edge computing. Seattle, WA, USA: SEC); 2018.
[151] Zhou Z, Qiao Y, Zhu L, Guan J, Liu Y, Xu C. Differential privacy-guaranteed trajectory community identification over vehicle ad-hoc networks. Internet Technol. Lett. 2018;1:1–7.
[152] Qiao Y, Liu Z, Lv H, Li M, Huang Z, Li Z, et al. An effective data privacy protection algorithm based on differential privacy in edge computing. IEEE Access 2019;7:136203–13.
[153] Lu J, Cai Z, Wang X, Zhang L, Duan Z. An edge correlation based differentially private network data release method. Secur Commun Network 2017:8408253. 1-8408253:14, 2017.
[154] Yin X, Zhang S, Xu H. Node attributed query access algorithm based on improved personalized differential privacy protection in social network. Int J Wireless Inf Network 2019;26:165–73.
[155] Wang Y, Xiao S, Xiao G, Fu X, Cheng TH. Robustness of complex communication networks under link attacks. In: Icait '08: proceedings of the 2008 international conference on advanced infocomm technology; 2008. p. 1–7. Shenzhen, China.
[156] Oliver J, Forman S, Cheng C. Using randomization to attack similarity digests. In: Atis 2014: applications and techniques in information security; 2014. p. 199–210.
[157] Miao F, Pajic M, Pappas GJ. Stochastic game approach for replay attack detection. In: Presented at the 52nd IEEE conference on decision and control; 2013. Firenze, Italy.
[158] Altop B, Nergiz ME, Saygin Y. A probabilistic inference attack on suppressed social networks. In: IEEE/ACM international conference on advances in social networks analysis and mining; 2012. p. 726–7. 2012.
[159] Su D, Cao J, Li N, Lyu M. PrivPfC: differentially private data publication for classification. The VLDB Journal 2017;27:201–23.
[160] Qiu R, Liu X, Huang R, Zheng F, Liang L, Li Y. Differential privacy EV charging data release based on variable window. PeerJ Computer Science 2021;7:1–18.
[161] Xiao X, Wang G, Gehrke J. Differential privacy via wavelet transforms. IEEE Trans Knowl Data Eng 2011;23:1200–14.
[162] Xu J, Zhang Z, Xiao X, Yang Y, Yu G, Winslett M. Differentially private histogram publication. The VLDB Journal 2013;22:797–822.
[163] Acs G, Castelluccia C, Chen R. Differentially private histogram publishing through lossy compression. In: Presented at the 2012 IEEE 12th international conference on data mining; 2012. Brussels, Belgium.
[164] Xiao Y, Xiong L, Fan L, Goryczka S. DPCube: Differentially Private Histogram Release through Multidimensional Partitioning 2014;abs/1202.5358:1–14. ArXiv.
[165] Song H, Fink GA, Jeschke S. Security and privacy in cyber-physical systems : foundations, principles, and applications. 2017.
[166] Tian Z, Wang Y, Sun Y, Qiu J. Location privacy challenges in mobile edge computing: classification and exploration. IEEE Network 2020;34:52–6.
[167] Ardagna CA, Cremonini M, Damiani E, Vimercati SD Cd, Samarati P. Location privacy protection through obfuscation-based techniques. presented at the DBSec 2007: Data and Applications Security 2007;XXI.
[168] Wang Y, Tian Z, Zhang H, Su S, Shi W. A privacy preserving scheme for nearest neighbor query. Sensors 2018;18:1–14.
[169] Barak B, Halevi S. A model and architecture for pseudo-random generation with applications to/dev/random. In: CCS '05: proceedings of the 12th ACM conference on Computer and communications security; 2005. p. 203–12.
[170] Simpson SV, Nagarajan G. A fuzzy based Co-Operative Blackmailing Attack detection scheme for Edge Computing nodes in MANET-IOT environment. Future Generat Comput Syst 2021;125:544–63.
[171] Wang Y, Chen Q, Kang C, Xia Q. Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Trans Smart Grid 2016;7:2437–47.
[172] Laguduva V, Islam SA, Aakur S, Katkoori S, Karam R. Machine learning based IoT edge node security attack and countermeasures. In: Presented at the 2019 IEEE computer society annual symposium on VLSI. Miami, FL, USA: ISVLSI); 2019.
[173] Mohanasathiya KS, Prasath DrS. Security And Privacy Using Two Fold Encryption Protocol Techniques In Edge Computing 2021:2865–74.
[174] Kasiviswanathan S, Nissim K, Raskhodnikova S, Smith AD. Analyzing graphs with node differential privacy. TCC; 2013.
[175] Zhang T, Zhu Q. Dynamic differential privacy for ADMM-based distributed classification learning. IEEE Trans Inf Forensics Secur 2017;12:172–87.
[176] Zhang Z, Qin Z, Zhu L, Weng J, Ren K. Cost-friendly differential privacy for smart meters: exploiting the dual roles of the noise. IEEE Trans Smart Grid 2017;8:619–26.
[177] Wang Q, Zhang Y, Lu X, Wang Z, Qin Z, Ren K. Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy. IEEE Trans Dependable Secure Comput 2018;15:591–606.
[178] Dijk Mv, Gentry C, Halevi S, Vaikuntanathan V. Fully homomorphic encryption over the integers. In: Presented at the annual international conference on the theory and applications of cryptographic techniques; 2010.
[179] Vijayarani S, Tamilarasi A, Sampoorna M. Analysis of privacy preserving K-anonymity methods and techniques. In: International conference on communication and computational intelligence (INCOCCI); 2010. p. 540–5. 2010.
[180] Lyu L, Nandakumar K, Rubinstein B, Jin J, Bedo J, Palaniswami M. PPFA: privacy preserving fog-enabled aggregation in smart grid. IEEE Trans Ind Inf 2018;14:3733–44.
[181] Al-Huthaifi R, Li T, Huang W, Gu J, Li C. Federated learning in smart cities:: privacy and security survey. Inf Sci: Int J 2023;632:833–57.
[182] U. Team. Out of commission: how the Oklahoma department of Securities leaked millions of files. 2019. https://www.upguard.com/breaches/rsync-oklahoma-securities-commission.
[183] Lall H. American Bank Systems faces Ransomware attack 2020. Available: https://iemlabs.com/american-bank-systems-faces-ransomware-attack-2/.
[184] Fan L, Xiong L. Differentially private anomaly detection with a case study on epidemic outbreak detection. Dallas, TX, USA: IEEE 13th International Conference on Data Mining Workshops; 2013. presented at the 2013.
[185] Du M, Jia R, Song D. Robust Anomaly Detection and Backdoor Attack Detection Via Differential Privacy 2020;abs/1911:1–11. 07116.
[186] Aljably R, Tian Y, Al-Rodhaan M, Al-Dhelaan A. Anomaly detection over differential preserved privacy in online social networks. PLoS One 2019;14.
[187] Biggio B, Nelson B, Laskov P. Poisoning attacks against support vector machines. In: The 29th international coference on international conference on machine learning; 2012. p. 1467–74. Edinburgh, Scotland.
[188] Zhang X, Zhu X, Lessard L. Online data poisoning attack. the 2nd Conference on Learning for Dynamics and Control 2019;120:201–10.
[189] Ma Y, Zhu X, Hsu J. Data poisoning against differentially-private learners: attacks and defenses. In: Twenty-eighth international joint conference on artificial intelligence; 2019. p. 4732–8.
[190] Rahimian S, Orekondy T, Fritz M. Differential privacy defenses and sampling attacks for membership inference. In: Proceedings of the 14th ACM workshop on artificial intelligence and security (AISec '21). New York, NY, USA: Association for Computing Machinery; 2019. p. 193–202.
[191] Gai K, Wu Y, Zhu L, Zhang Z, Qiu M. Differential privacy-based blockchain for industrial internet-of-things. IEEE Trans Ind Inf 2020;16:4156–65.
[192] Sun Z, Wang Y, Cai Z, Liu T, Tong X, Jiang N. A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing. Int J Intell Syst 2021;36:2058–80.
[193] Guo J, Ding X, Wang T, Jia W. Combinatorial resources auction in decentralized edge-thing systems using blockchain and differential privacy. Inf Sci 2022;607:211–29.
[194] Lv Z, Chen D, Feng H, Singh AK, Wei W, Lv H. Computational intelligence in security of digital twins big graphic data in cyber-physical systems of smart cities. ACM Transactions on Management Information Systems (TMIS) 2022;13:1–17.
[195] Pandya S, Srivastava G, Jhaveri R, Babu MR, Bhattacharya S, Maddikunta PKR, et al. Federated learning for smart cities: a comprehensive survey. Sustain Energy Technol Assessments 2023;55:102987.
[196] Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Trans Ind Inf 2019;16:4177–86.
[197] Zhang H, Li G, Zhang Y, Gai K, Qiu M. Blockchain-based privacy-preserving medical data sharing scheme using federated learning. Tokyo, Japan: KSEM; 2021. p. 1–15.
[198] Wang X, Han Y, Wang C, Zhao Q, Chen X, Chen M. In-edge AI: intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Network 2019;33:156–65.
[199] Hou L, Ni W, Zhang S, Fu N, Zhang D. Wdt-SCAN: clustering decentralized social graphs with local differential privacy. Comput Secur 2023;125:103036.