A Comprehensive Review on Security and Privacy Preservation in Cloud Environment

Rajesh Bingu, S. Jothilakshmi, and N. Srinivasu

Abstract With the widespread growth of the Internet of Things (IoT) devices and emergent data generated at the edge computing network, the conventional cloud computing (CC) model encounters various bottleneck issues like resource constraints and bandwidth constraints. Thus, edge computing devices help in processing and storing data at the edge level and intended as a promising solution for the past few years. Moreover, the unique nature of computing devices like parallel processing, content perception, and real-time computing have introduced various challenges in privacy preservation and data security. These two factors are considered as the key factors in this review. In this survey, an empirical analysis of the privacy and security threats, protection methodologies, and related countermeasures are inherited in cloud computing. Specifically, an overview of basic CC definitions, applications, and architectural models are discussed. Subsequently, an extensive analysis is performed specifically with data security, privacy, challenges, requirements, and mechanisms in CC. Various cryptographic-based methods based on privacy preservation and security are discussed in this review. The existing issues related to security and privacy in CC are reviewed and finally, various open research directions related to privacy and security in the CC field are also discussed.

Keywords Cloud Computing · Resource constraints · Bandwidth constraints · Privacy · Security · Edge computing · Threats · Cryptography methods

1 Introduction

In the fast-growing technological advancements, the Cloud Computing concepts are raised from the distributive software architectural model [1]. Cloud Computing

R. Bingu (✉) · S. Jothilakshmi
Department of Information Technology, Annamalai University, Chidambaram, Tamilnadu, India

N. Srinivasu
Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, India
e-mail: srinivasu28@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_54

(CC) technology attempts to offer host services on the Internet. Recently, CC is considered as the technological advancements with the rise of various industrial markets and user communities. CC services are offered from small/huge data centres provided in various regions of the world. For instance, Google applications and Microsoft are some common examples that offer CC services where security plays a substantial role during the establishment of CC services [2]. Various prevailing literature concentrates on diverse security solutions that include the execution of security policies and technological advancements. Some researchers concentrate on various security issues due to attacks over the CC environment from the attacker's perspective [3]. The solutions anticipated by the researchers to handle these attacks rely on the security theories to protect the computing environment. Aluvalu et al. [4] describe various security issues that influence the CC attributes. Also, the author intends to provide the solution to the problems identified and directly related to cloud security. Some security-based guidelines are provided by the author and facilitate the computing organizations needs to be aware of vulnerabilities and various approaches are provided to handle these vulnerability in an efficient manner [5].

Various threats and related challenges are grown from the wider use of CC services. At present, CC models are the preliminary source of these vulnerabilities and challenges [6]. The attackers exploit the feebleness of the computing models while the service providers initiate to access the users' private data and influence the computer systems' processing power. Recently, cloud-based intrusion detection and prevention (CIDS) mechanisms are anticipated to handle the problems discussed above. Before the design of the cloud intrusion detection model, the 'Network Intrusion detection and prevention (NIDS)' mechanism is used and it handle the risks over the CC virtual networks [7]. When compared to the NIDS, CIDS is finest during the mitigation of risks and security challenges towards the network model.

Recently, there is a rising growth of CC techniques in Information Technology. Moreover, various services are still unwilling to completely adapt towards the CC as appropriate security technologies which are not matured yet [8]. Therefore, various literature provides the preliminary requirements to invest the amount for the CC-based device security. Few studies concentrate on offering a solution to the evaluation of the CC-based security model. One research model initiates a novel 'attack tree map' to examine the security threats and vulnerabilities. Xue et al. [9] discuss various CC facets merged with the trusted computing environment to offer security services like integrity, authentication, and confidentiality.

Owing to the distinctive characteristics and benefits of the CC paradigm like mobility support requirements, location-awareness, parallel computation, heterogeneity-based distributive architecture, and huge data processing shows some conventional privacy preservation and data security mechanisms in CC [10]. It is no longer appropriate for protecting massive data in the CC paradigm. Specifically, secure data computation, secure data storage, privacy protection, and authentication access control factors are specifically pre-dominant [11]. For instance, CC is a distributed computing system with huge trust domains where multiple functional entities, the authentication mechanism not only needs the identity validation for all entities in the trusted domain; however, it needs every cloud entity need to authenticate

mutually among various trusted domains. However, for certain resource-constraint devices, it is not possible to preserve the huge amount of data or execute the complex security algorithm [12]. To be specific, the privacy preservation and data security in CC specifically concentrates on the following confronts [13]:

(1) **Fine-grained and Lightweight model**: Some novel requirements for fine-grained data sharing systems and lightweight data encryption approaches are based on multiple authorized parties in CC.
(2) **Distributed access control**: The heterogeneous secure data management and multi-source data dissemination issues in the distributed environment.
(3) **Resource constraints**: Security confronts resource-constrained devices and large-scale edge services.
(4) **Efficient privacy-preserving**: Some novel requirements of effectual privacy preservation methods for diverse cloud computing models and edge services faced in the CC environment.

The above-mentioned privacy preservation and security challenge of the CC paradigm motivates us to offer comprehensive research analysis. The significant contributions of this review are summarized as follows.

An extensive analysis is performed for constructing a factor in CC that is discussed holistically. A comprehensive analysis of CC definition and architectural model is provided. Some promising CC applications are also discussed [14]. The privacy preservation and data security requirements are provided based on the critical metrics include privacy requirements, access control, authentication, integrity, confidentiality, and availability. Some comprehensive analysis of the potential privacy and security challenges in CC is pinpointed [15]. Specifically, the prevailing privacy preservation and data security methods are discussed in detail. The architectural level towards the security model is also anticipated.

A comprehensive analysis of the cryptographic-based model for resolving privacy and data security issues is discussed. It includes proxy re-encryption, identity-based encryption, searchable encryption, attribute-based encryption, and homomorphic encryption. Moreover, a comprehensive analysis and comparison of various privacy and data security solutions are provided and the solutions related to these issues are discussed.

Some open issues are also discussed with the future research challenges like the modelling of fine-grained privacy preservation model, dynamic data processing, cross-domain authentication, multi-authority access control mechanism, and distributive data encryption.

The rest of the work is organized as: Sect. 2 discusses the CC-based privacy preservation framework. Section 3 elaborates the security threat mechanism with secure outsourcing computation and computational requirements. Section 4 explains the privacy threat mechanism encountered in the cloud that includes intrusion, public disclosure, appropriation, and false light. Section 5 discusses the regulations of privacy preservation which are followed by the privacy factors established in Sect. 6. Here, factors like security, data access and utilization, data location, and data status are discussed. Sections 7 and 8 discusses privacy-enhancing technologies and data

encryption technologies. Section 9 discusses the research gaps with the research questions that need to be resolved in future. Section 10 provides the summary of the comprehensive survey.

2 CC-Based Privacy Preservation Framework

In CC, the user's privacy preservation data needs to be more confidential. Some data outsourcing mechanisms like storage, multi-tenant and virtualization, big data storage, and other technologies are considered as the major factors of risk privacy disclosure [16]. Privacy preservation mechanisms are considered to eliminate sensitive information and personal information exposure in the cloud environment, i.e., computing, data sharing, delete, integrity verification, and other functionalities. Various solutions are considered to protect personal privacy and user's information like trusted technology, encryption, access control, or various other combinations of these methods [17]. Storage services in the CC platform show substantial variations in the IT field and give a tremendous impact on the privacy and security factors [18]:

(1) User data is completely centralized and security methods need to be fulfilled during the process of CC requirements.
(2) The cloud users' do not have any trustfulness in the CC platform, therefore, the end-users do not believe that the data is stored in a secured manner.
(3) Generally, the user data is highly confidential and centralized, therefore, security must be fulfilled during the data processing in the cloud environment.

Security and privacy are the major factors in cloud where the efficient access control and secure data resources are the key factors to be achieved in an efficient manner [19]. After predicting the legitimate user identity, the access control mechanism denies the request generated from the data access and generally, it is used for preserving the data resources and avoids the illegal functionalities of the intruders. There are diverse access control mechanisms like Mandatory Access Control (MAC), Attribute-based Access Control (ABAC), Discretionary Access Control (DAC), Task-based Access Control (TBAC), Usage Control (UCON), and Role-based Access Control (RBAC). The performances of these mechanisms are evaluated. Some set of procedures and rules can facilitate the authentication of legitimate users to access data by fulfilling the data privacy, data confidentiality, and integrity in the security model [20]. When compared to the conventional network environment, the access control mechanism is extremely essential in the CC environment. While using the computing and storage services, the users need to provide authentication to the Cloud Service Providers (CSP) and consider the essential policies to access services and data. Access control and mutual authentication among the CSP are essential to fulfil the CC security [21]. The user not only requires a side-channel attack; however, it requires appropriate methods to fulfil privacy and security.

At present, CSP uses various access control methods to offer security protection and privacy. But, various issues need to be resolved [22]:

(1) In cloud security, both objects and subjects are not distinguished clearly and the access control method needs to be expanded from the user's authority to protect the data, and access virtual resources in the cloud environment. Additionally, there is some differentiation among the conventional centralized resource management system and distributed access mechanism in the CC and certain access policies are generated as a test for security management.
(2) In the security mechanism, there are diverse applications that come under various security management issues and the user needs to access the certification services and resources in the domain boundaries to deal with the public policies. When the tenants encounter a side-channel attack, it leads to privacy leakage.
(3) In the access control mechanism, the idea behind the object and the subjects in the cloud are extremely influenced by the conventional access control mechanism and it is unable to fulfil the cloud requirements. In the CC environment, the relationship among the various mechanisms is complex and the user scalability frequently varies, the number of administrators is huge and complex, and the authoritative assessment of the cloud authority varies from the conventional models.

In the conventional model, the user needs to deal with the data in a trusted and reliable manner; however, it is complex in the CC environment. Both the server and the users are not credible completely and the user needs to fulfil the data security and avoids CSP malicious activities of the private data. It is essential to merge the diverse technological concepts to build the privacy preservation concept.

3 Security Threat Mechanism

Generally, outsourcing computation is not completely satisfying; there is some preliminary cause for handling the privacy and security factors during the data analysis like user privacy and revealing the data content. It should predict and handle security threats. Thus, it reduces the risk of avoiding privacy and data security [23]. Two diverse threats are considered in the CC environment. First are the external attackers that include threats of hardware attacks and remote software in CC. It uses various kinds of approaches like malware attacks and network overlapping to evaluate unauthorized data and server intrusion. Secondly, security threat is identified from the internal participants. When the user submits the computational tasks, the computational process and data are generally controlled. Therefore, cloud server mortality plays a substantial role in data privacy and security [24]. Moreover, the server does not as a trustful environment. It shows that the server can handle the computational tasks inappropriately and pretends to learn the realistic features. Based on the server's nature, some users need to categorize the adversarial models into two diverse levels: semi-honest and honest level [25]. In the former model, the server carries out the functionality with the essential computational process, however, crucial regarding the

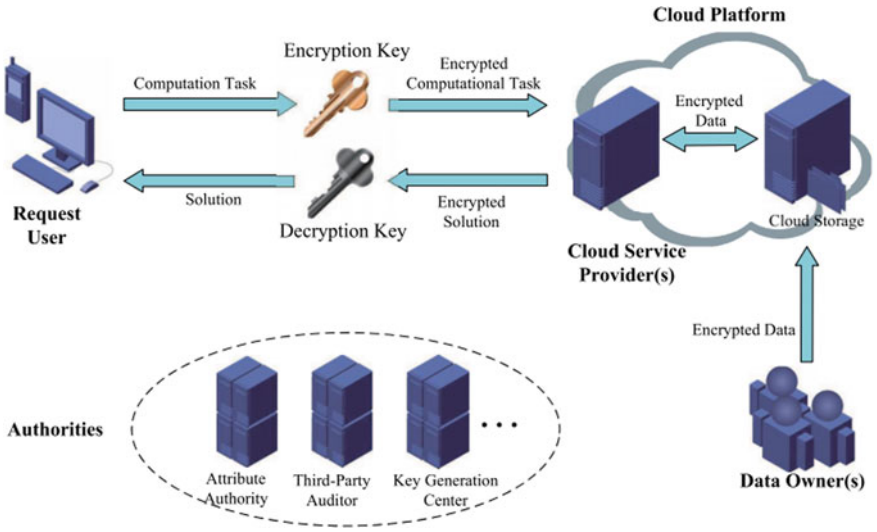

Fig. 1 Security-based cloud system model

sensitive user's information. The malicious servers move against the computational power and attain certain benefits [26]. There are two kinds of threats like internal and external attacks that destroy the user's integrity and confidentiality. Figure 1 depicts the security-based cloud system model.

(a) **Need for secure outsourcing computation**: The security information is also substantially essential in various computational schemes and the efficiency specifies the communication and computational cost for performing the task outsourcing. The evaluation process includes three diverse factors like access controllability, data integrity, and data confidentiality [27]. Here, confidentiality specifies the competency to preserve the user's data to be exposed in the CC environment or to unauthorized parties. Data integrity specifies the fulfilment of data correctness and completeness. It is also termed as the data check ability and verification. Finally, access controllability is a property that executes the permission process towards the valid users and restricts the set of users to attain the computational outcomes and predict the data sources from the data owners. It is a fine-grained process that offers feasible access control mechanisms and thus shows higher significance towards reality and theory.

(b) **Essential requirements**: Another pre-dominant requirement for computational outsourcing is the system performance. It is due to the higher computational cost and it selects the outsourcing of various complex tasks over the cloud environment [28]. The significant factors to evaluate the system efficiency are computational overhead before and after the outsourcing process, i.e., verification, processing, encryption and decryption, and communication cost during the transmission process.

Moreover, outsourcing needs to attain the requirements essentially and shows huge effect over the CC environment which is more impractical. Generally, security needs added operational support and diminish efficiency [29]. However, an excess process with high efficiency generally causes insufficient security factors. Therefore, superior and feasible performance is generally considered to offer a superior trade-off between efficiency and security. Also, it is a huge inspiration to model the design algorithm efficiently. It provides an extensive scope on the certain computational process with the application-specific and fundamental process as the outsourcing process shows limited functionalities.

4 Privacy Threat Mechanisms

The need for privacy arises when the user's data is accessed without any knowledge or consent of the data owners. It happens during the attack, data breaching, eavesdropping, and other forms. It is classified in various forms:

(a) **Public disclosure**: Realising the private or unknown information to the external members is more determined as a public disclosure [30]. The information is offensive while exposed to unauthorized users. Thus, if the data does not offer any public concern, the individual who is accountable for releasing the data is more liable for privacy invasion. Some typical forms of examples are publicly disclosing the private information by the officers, politicians, and other celebrities.

(b) **Intrusion**: The intrusion towards the privacy data includes the process of direct/indirect access towards the institutional or individual private data, i.e., telephonic conversations that are recorded devoid of the knowledge of the owners and trespassing the private data sources [31]. The injections of malicious activities are also identified in the CC environment.

(c) **Appropriation**: It specifies the appropriation of organizational or individual identity. Usually, it happens with the use of the individual name or personal features devoid of knowledge and authorization. It is generally noted in the reference, media cases, marketing, and stories. It is most probable to occur and the issue is more recurrent and happens in the online profiles.

(d) **False light**: It is like public disclosure. It is a form of malicious statements and public disclosure of false statements. Usually, it is performed with trust distortion and fictional factors.

There exist some added privacy threats, for instance, fine-grained taxonomy like information processing, information collection, invasion, and information dissemination [32]. An attack is a process of attempt, alters, expose, disable, destroy, gain unauthorized access, steal, or unauthorized access over a specific environment as stated by ISO. In the privacy preservation process, various forms of consensus specify the form of attack. The attackers evaluate the kind of information that is available and reflects the other form of information methods.

- **Marketer—**The attacker is not interested in re-identifying the individuals.
- **Journalist—**The attacker does not have any prior knowledge of the attack that needs to be held.
- **Prosecutor—**The attacker is aware of the form of data available in the target region and sometimes data is available in the form of a dataset.

Also, it is necessary to ensure certain attack models that operates in the specific environment. There are some forms of attack identification process [33]. They are:

- **Probabilistic attack—**It relies on the uninformative principle that concentrates on actual records. It is known that access to data does not change significantly.
- **Table linkage—**It happens when the attack derives the presence or absence of the data successfully in the targeted record of the table owner.
- **Record linkage—**It happens when the attacker matches the record owner successfully towards the sensitive attributes from the published dataset.
- **Attribute linkage—**It happens when there is no specific record identification process; the attacker infers some sensitive information with a group of data ownership.

Table 1 depicts some sample instances of the privacy invasions with the exposed private data and compromised organizational privacy. The credit card credentials, political interest, and some personal details are disclosed publicly. In certain cases, like Uber and Yahoo, data breaching is identified due to some security reasons. Moreover, in the case of AOL and Netflix, incorrect utilization of anonymization is considered [34]. Similarly, linkage attacks and cross-references are identified with a high risk for data anonymization. These security risks have to be reduced and avoided by considering the attack model.

5 Regulations for Privacy Preservation

Various countries adopt certain laws and regulations for privacy preservation, data handling and sharing, and data access [35]. In European countries, certain essential directives are enforced and they should be adopted by the European Union. Data Protection Regulation needs to be enforced for some businesses or services that handle data from the citizens and enforced to execute it. In the USA, an act known as the Gramm-Leach-Bliley Act is adopted and the Personal Information Protection and Electronic document act are adopted by Canada. In the Russian Federation, the Personal data protection act is used, while in China, the Personal Information Protection act and Computer processed personal data are some examples of regulations [36]. It may vary from one country to the other with some common objective for provisioning legal regulation and protection to the user's private and personal data.

Table 1 Samples of exposed data

Organization	Description	Threat invasion model	Year	Exposed data
Microsoft	Roughly 250 million organizational data are exposed	Security problems	2020	Appropriation and intrusion
Facebook	Roughly 540 million organizational data are exposed	Third-party security problems	2019	Appropriation and intrusion
Cambridge Analytica	Unauthorized way of profiling	Scrapped personal information from the personal user account	2018	Only intrusion
Equifax	Exposed some sensitive and confidential data of around 140 million users	Happens due to hacking	2017	Appropriation and intrusion
Uber	Driver details and 57 million uber customer details are exposed	Happens due to hacking	2016	Appropriation and intrusion
JPMorgan	Driver details and 87 million customer details are exposed	Happens due to hacking	2014	Appropriation and intrusion
Yahoo	Roughly 500 million user account details are exposed	Happens due to hacking	2014	Appropriation and intrusion
Netflix	The exposed dataset helps to identify the user's information	Cross-referring	2006	Disclosed publicly
AOL	The exposed dataset helps to identify the user's information	Cross-referring	2006	Disclosed publicly

6 Establishing Privacy Over the Cloud Environment

The provisioning of cloud services varies based on conventional Internet services. The data processing over the servers or the server's location is concerned with the privacy policies. The preliminary requirements of privacy are discussed below [37]:

(a) **Security—**For establishing security, some essential features like infrastructure, communication, and security features play a substantial role in securing the data. Some general aspects like anti-virus, strong password, and certain regular software updates can improve the security in both the service providers and users side.

(b) **Data status—**The aspect related to CSP needs to consider the method for disclosure of some specific method utilized for preserving data. The data

status during the handling and processing stages needs to be specified, i.e., pseudonymized, anonymized, plain text, and encrypted form.

(c) **Data access and utilization—**It is completely essential to fulfil the appropriate handling and accessing of data. The system service might not be compromised even in case of the least security policies and measures. Fulfilling appropriate cloud utilization policies like logical and physical access are not provided due to diligence. Some specific form of data processing is required. It is necessary to disclose the usage policies in two diverse directions: service provider to user and user to service provider. It is completely crucial for efficiently defining the access rules. Some sets of questions need to be resolved while considering data access and usage like how, who can access the data, where, why, and time.

(d) **Data location—**The privacy regulations and laws differ from one country to other. Thus, compliance in various locations is more challenging. Some organizations that process data from International users face some constraints, i.e., the servers with computing power and databases are distributed in various countries. Some added aspects need to be handled in an efficient manner: customer data management and local laws; and laws based on the origin of the country. Failure leads to some losses in the organization.

As depicted in Fig. 2, there are some requirements like availability, portability, and audibility that needed to be considered. However, there is still some vulnerability in association with CSP to fulfil the higher privacy standards. Generally, data owners with these services have no physical access over the system mode. Thus, indeed of complete trust establishment, there is a semi-trusted relationship [38]. Generally, CSP is used for some common outsourcing data purposes. Users need to consider some proactive approaches like data anonymization. Similarly, various privacy-based approaches and tools are restricted with the exposure of sensitive information.

7 Privacy-Enhancing Technologies

The privacy-enhancing technologies carry out data transformation and operations that help to deal with a higher level of data privacy, i.e., encryption or data anonymization. Various technologies are accessible in handling privacy preservation. Some of the evaluation indicators are provided based on the quality of the technology [39]. They are discussed below:

- **Readiness evaluation**: There are six diverse stages of the readiness evaluation process. They are product, idea, outdated, proof-of-concept, pilot, and research.
- **Quality evaluation**: It is based on nine diverse indicators with varying weights. They are: scope, protection, transferability, trust assumptions, maintainability, side effects, performance efficiency, reliability, and operability

It is chosen based on the score offered by the indicators. The scores are essential for providing a systematic model, i.e., selection of appropriate combinations of the

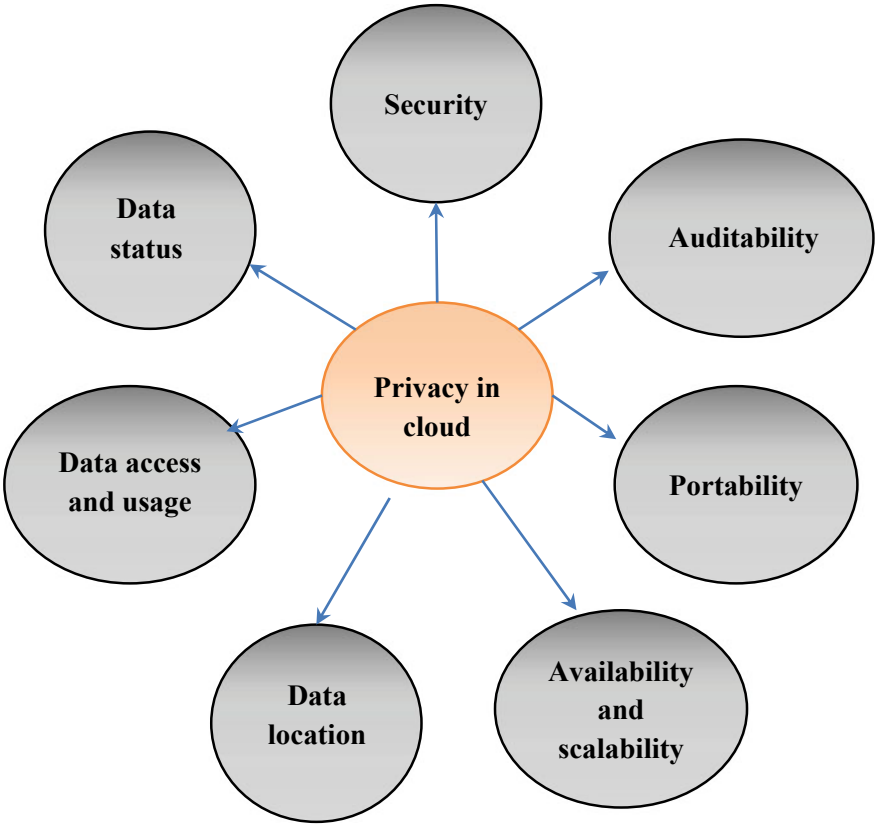

Fig. 2 Privacy preservation consideration in CC

enhancing technologies, for instance, data protection, authorities, online users and researchers, and software developers [40]. Some common types of privacy provisioning technologies are provided in-between the services and the users. Onion routers and the corresponding network provide increased online privacy using web traffic anonymization. It is improved and provided as a modern version of Firefox browser with improved usability. However, it is extremely possible to predict the client data by performing browser attacks [41]. In the communication and network domain, Domain Name Server and encryption models are utilized to improve privacy. However, for the improved privacy preservation process, the technology is complementing one another. Data publishing provides challenges to researchers and authorities owing to the trade-off among the disclosure risk and data utility and inherits the re-identification process by linkage or cross-reference attacks. Attacks rely on the published data or matches with the data anonymization. Chen et al. [42] perform hybridization of Internet Protocol (IP) and timestamp using some distinctive methods and the outcomes provide protection against the re-identification process. In the

timestamp process, multiplicative noise and enumeration are used for preserving the structural flow. Some preservation techniques are suited for applications like unstructured and structured data, real-time, offline application, and reversible data. Based on these applications, data owners need to characterize suppressiveness to operate the unstructured or structured data.

8 Data Encryption Technologies

The security is vulnerable when the provided data is outsourced to the cloud environment. Thus, encryption is considered an efficient technique to preserve data. The significance of the encryption process is to convert the original form of plain text to a string format with unreadable code. It is known as ciphertext [43]. The garbled code attains the original content which significantly preserves the data confidentiality and avoids the data being tampered with. The authorized users can access that encrypted data with the appropriate private key and later update or modify the content. The encryption process is further divided into asymmetric and symmetric encryption processes. In symmetric encryption, the process makes use of the secret key for encrypting and decrypting the data. Moreover, before symmetric encryption, users have to determine the consensus key. However, it is extremely inefficient for multi-user sharing files. Subsequently, asymmetric encryption is also known as the public key encryption process and it is extremely convenient. It has a pair of keys and it is disclosed with the encrypted files. However, the private key is utilized for ciphertext decryption. There are diverse encryption technologies that are extensively adopted in cloud storage systems.

(a) **Identity-based encryption (IBE)**: In the conventional Public key Infrastructure, the identity information is more constant with the public key and used for the encryption process. The sender has to authenticate the identity information of the receiver via a trusted third-party certificate [44]. It may lead to the increase of the sender's workload when the sender needs to share confidential data with multiple receivers. The idea behind the identity-based cryptographic model is anticipated to associate the identity information with the available public key. Therefore, there is no need to validate the receiver's certification before performing the encryption process. The identity-based encryption (IBE) model is designed by Franklin et al. to establish a security model using bilinear mapping to design a secure scheme. For example, 'A' is a sender who needs to transfer an encrypted message to the receiver 'B'. For this purpose, a trusted third party and Private Key Generator (PKG) is needed for generating the corresponding private key and public key. Here, 'A' uses unique identity information for generating the public key and sends the encrypted message to 'B'. Then, 'B' needs to communicate with the key generator for establishing authentication and to attain related private keys (See Fig. 3).

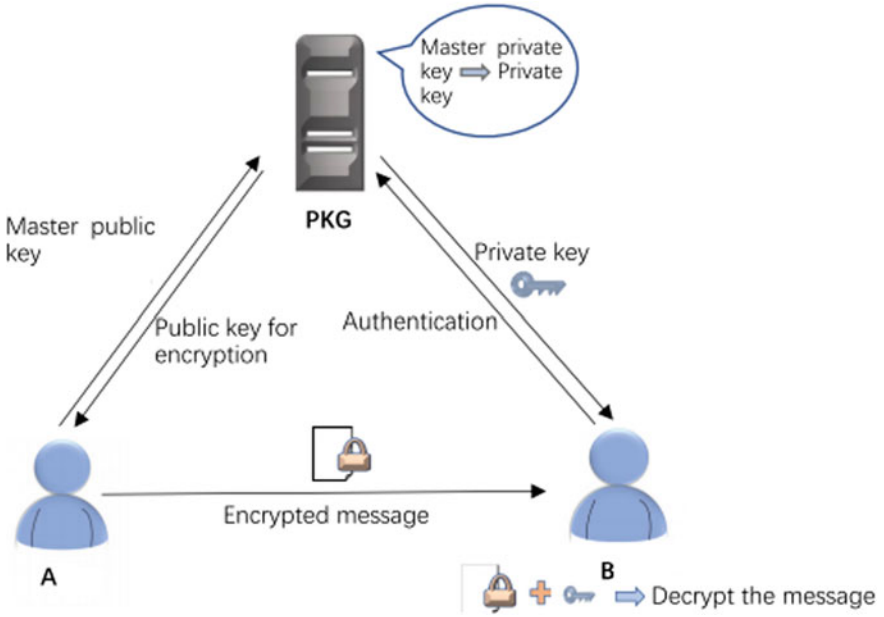

Fig. 3 Identity-based encryption process

(b) **Attribute-based encryption (ABE)**: In the existing IBE scheme, identity is measured as the string which is divergent from one another. Moreover, the feasibility of the model shows some bottlenecks when the ciphertext intends to be legally accessed by the users [45]. The identity of the Integrated Development Environment (IDE) model is substituted by the set of attributes in the ABE model. Here, the user's attribute access policy has the competency to access the encrypted data. It is composed of four diverse parts:

(1) **Setup phase**: It is also known as the system initialization phase with certain security parameters. They are input data, public parameters and the master key.
(2) **KeyGen phase**: Here, the data owner needs to submit their attributes for attaining private key related to the attributes.
(3) **Encryption phase**: The data owners encrypt the data using the public key and attain ciphertext and transfer it to the receiver or the public cloud environment.
(4) **Decryption phase**: here, the users get ciphertext and decrypt with the private key.

The attribute-based encryption model is considered to be a promising solution to offer fine-grained access control towards the encryption files in data sharing applications. The data owner needs to specify who can access these encrypted data. It is subdivided into two phases: Ciphertext-policy attribute-based encryption model

(CP-ABE) and Key-policy attribute-based encryption model (KP-ABE). In the KP-ABE, the ciphertext is related to the attribute set while the private key is associated with the access policy of the attributes. While in CP-ABE, the policy is merged with the ciphertext and the data owner needs to determine the access policy. Here, the private key is associated with the corresponding attributes. The user's attribute can be changed based on various reasons. The malicious nature of some authorized users is disclosed to spoil the privacy and confidentiality of the data [46]. Thus, secure and protective revocation is highly essential in the ABE model. Some prevailing revocation models shows indirect revocation and some may have direct revocation. The trusted party interacts with the non-revoked users periodically and updates the decryption key. Here, the decryption key is composed of two diverse parts. They are the update key and long-term secret key where the update key needs to be regularly updated. The difference among the attributes are partitioned into two diverse disjoint sets and merged with the master key for generating the secret key. These secret keys are diverse and show the re-randomization property. Therefore, resistivity is achieved with the decryption key. Later, a tree-based model is designed for diminishing the computational complexity and key generation phase.

In the direct revocation process, the trusted authority needs to generate the revocation list that includes the revoked users with the public key. The data owner represents the revoked users and cannot decrypt the ciphertext even in the case of matched attributes. Here, the authority revokes the user by updating the revocation list and interacts with the non-revoked users. After receiving the list, the third party needs to update the ciphertext with some public information and ensure the ciphertext is not decrypted by the revoked users. At last, the authorized user has the pleasure to validate the update of the third party appropriately. This model does not forbid the revoked users to decrypt the ciphertext; however, it provides verifiable functions to fulfil ciphertext under the revocation list.

(c) **Homomorphic encryption model**: The introduction of attribute-based encryption (ABE) and Identity-based encryption (IBE) is provided for fulfilling the data confidentiality; there exist certain drawbacks in those models. When the user intends to update the encryption files over the cloud storage, there are two options. One is the modification of ciphertext in the cloud. After this decryption of the modified ciphertext, it is generally considered as the meaningless garbled code that causes data damage. Next is to update the decrypted file and transfer it to the new cloud location. However, these two modifications are complex. Also, the data transmission from the cloud storage system to the cloud storage leads to data leakage risk [46]. The homomorphic encryption model is used for handling these issues with higher significance (See Fig. 4). It is a kind of public key encryption process that facilitates the users to carry out certain algebraic functions and acquires encrypted text and results in the consistent form of decrypted data. The data owner performs file encryption using this homomorphic model and transfers it to the cloud server. The authorized user needs to decrypt the ciphertext with the appropriate private keys. This homomorphic encryption model is more efficient to protect outsourced data. Table 2 depicts

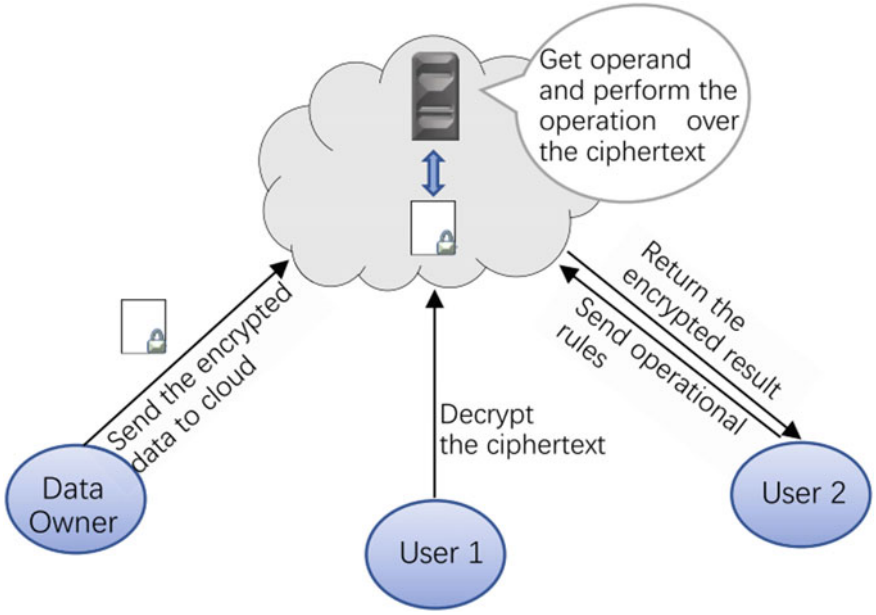

Fig. 4 Homomorphic encryption process

the comparison of the ABE model.

(d) **Searchable encryption:** It is a symmetric key encryption approach that outsources the confidential data from one person to another using searching process. It adopts proxy re-encryption that shares data from the cloud with E2E encryption and confines the authentic recipients. It serves as a privacy and security factor with various cryptographic approaches like key search model and time-dependent proxy re-encryption. It helps to search the unique keyword preserved over the server and uses non-iterative and iterative schemes. These schemes assist in index generation, searching and updating while non-iterative scheme helps in hash chain. It gives superior security model over the huge database and reduces the storage overhead.

(e) **Proxy re-encryption:** It is a cryptographic model provides semi-trusted server (proxy) to re-encrypt the provided ciphertext (See Fig. 5). It is encrypted by one public key to another ciphertext. The proposed proxy re-encryption model facilitates decryption of user's data over the cloud server and credentials devoid of secret key disclosure. A time-based proxy re-encryption is proposed that facilitates users to access the user's record in a pre-defined time interval. The target is to achieve access control, time-based revocation, efficiency, and user revocation.

Table 2 ABE model comparison

Category	Sub-category	Model	Techniques	Benefits
Revocation	Direct	The key-policy attributes are revocable directly with the ciphertext delegation verification process	(1) Linear secret key sharing (2) KP-ABE	(1) Appropriate security (2) Universal construction (3) Update verification (4) Direct revocation
		ABE-based verifiable key concept and direct revocation	(1) Linear secret key sharing (2) KP-ABE (3) Sub-set difference method	(1) Outsource decryption process verification (2) Partially hidden policy (3) Direct revocation
		Partial secure delegation and policy-hidden attributes	(1) Linear secret key sharing (2) Broadcast encryption (3) Outsourcing model (4) CP-ABE	(1) Decryption key resistive exposure (2) Ciphertext delegation (3) Randomized key generation (4) Indirect revocation
	Indirect	Ciphertext delegation and revocable ABE with exposed decryption key	(1) CP-ABE (2) ABE (3) Tree-based revocation	(1) Complete security (2) Ciphertext delegation (3) Storage revocation
Revocable storage system		Ciphertext delegation and dynamic credentials for ABE	(1) CP-ABE and KP-ABE (2) Linear secret sharing model	(1) Least computational complexity (2) Effectual keyword searching process (3) Secured CPA
Computational overhead reduction	Outsource computation	Outsourced ABE with key search function	(1) KP-ABE (2) secret key sharing with interpolation	(1) Reduced redundancy (ciphertext) (2) Lower computational overhead and storage
	Compact policy	Compact ABE	(1) CP-ABE (2) greedy compact model (3) Policy compact	(1) Reduced computation overhead (2) Reduced ciphertext redundancy

(continued)

Table 2 (continued)

Category	Sub-category	Model	Techniques	Benefits
	Enhanced policy management	Scalable ciphertext policy-based ABE	(1) CP-ABE (2) Linear secret key blocking	(1) Collision resistivity (2) Reduced computational overhead (3) Reduced storage cost

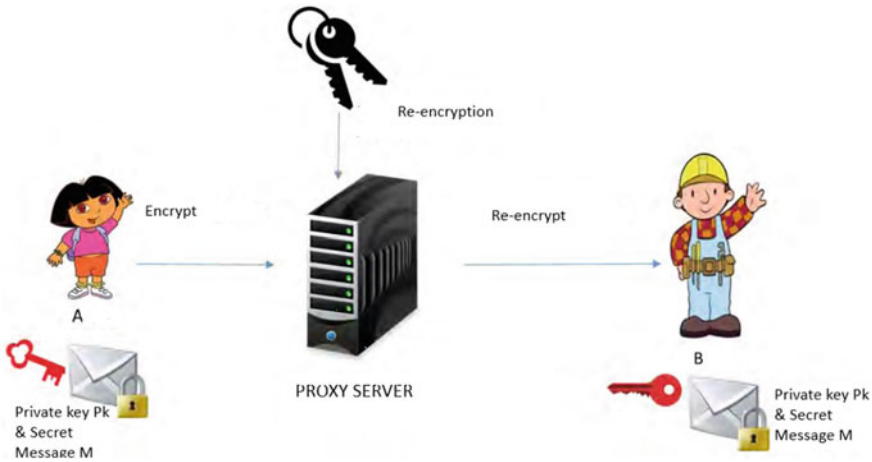

Fig. 5 Proxy re-encryption scheme

9 Research Gaps

Based on the extensive research analysis, it is observed that some research constraints need to be resolved in future. The trust evaluation criteria are insufficient and lack standardization. The access control methods are extremely complex with the selective constraint towards flexibility and scalability. Generally, the access control policies are formulated by the end-users. Thus, it increases the risk of service access and strictly restrictive to third-party access [47, 48]. There is a diverse issue that needs to be resolved efficiently and it is related to policy and key management, unauthorized access, attribute management, and so on. When dealing with encryption-based challenges, ABE is effective but has various constraints. How efficiently the ciphertext can be searched and control the privacy leakage is extremely complex. The construction of traceable, revocable, and feasible attribute encryption is challenging.

(a) **Research questions**

There is a need for a trustful, secure, and reliable service to eliminate illegal access to cloud systems. The researchers need to concentrate on the below-given research

questions to improve the standard of privacy preservation and data security over the cloud [49].

(1) How to compute and store the privacy preservation data efficiently in the cloud environment?
(2) Which standard access control method is efficient for secure data transmission in the cloud?
(3) How efficiently the data is shared with multiple Cloud Service Providers?
(4) Which encryption scheme is used for protecting the data in a secured and private manner?
(5) How to handle the computational complexity efficiently?
(6) How to integrate the encryption scheme with parallel cloud architecture in case of emergency conditions?
(7) How to construct an efficient keyword search mechanism in storage services?
(8) How to offer identity and location-based privacy for the CC environment?

10 Conclusion

This survey provides a comprehensive analysis of privacy preservation factors and security establishment over the cloud storage system. Here, a baseline analysis of the cloud performance at the organizational level, digital economy, and transformation are discussed with the privacy preservation framework. An extensive analysis of the data security elements in cloud storage like confidentiality, integrity, fine-grained access, availability, access control, data sharing, leakage resistivity, delegation, and privacy protection methods are performed. Some encryption concepts like ABE, IBE, and homomorphic encryption models are evaluated by analysing the significance of the modelling and validating whether the security is achieved notably. However, some research constraints need to be resolved which is a hot research topic in the field of security and privacy preservation in the cloud. In future, this research is extended to provide a solution to ensure the security and privacy of the data hierarchically while data sharing occurs. Thus, a constructive hierarchical data sharing (CHDS) method is proposed with the adoption of symmetric encryption over the rooted hierarchical graph structure. The hierarchical graph model deals with the features of incoming data to establish the privacy and authenticity of the model.

References

1. Y. Zhang, Research on the security mechanism of cloud computing service model. Autom. Control Comput. Sci. **50**(2), 98–106 (2016)
2. P.G. Shynu, K.J. Singh, A comprehensive survey and analysis on access control schemes in the cloud environment. Inf. Technol. **16**(1), 19–38 (2016)

3. D. Stevenson, J. Pasek, Privacy concern, trust, and desire for content personalization, in *Proceedings of Research Conference on Communication, Information and Internet Policy* (2015), pp. 1–30
4. R.K. Aluvalu, L. Muddana, A survey on access control models in cloud computing, in *Proceedings of 49th Annual Convention of Computer Society of India (CSI)*, vol. 1 (2015), pp. 653–664
5. J. Li, Y. Zhang, X. Chen, Y. Xiang, Secure attribute-based data sharing for resource-limited users in cloud computing. Comput. Sec. **72**, 1–2 (2018)
6. R. Zhang, H. Ma, Y. Lu, Fine-grained access control system based on fully outsourced attribute-based encryption. J. Syst. Softw. **125**, 344–353 (2017)
7. A. Beimel, A. Ben-Efraim, Multi-linear secret-sharing schemes, in *Proceedings of Theory of Cryptography Conference*. Lecture Notes Comput. Sci. **8349**, 394–418 (2014)
8. M. Bellare, D. Hofheinz, E. Kiltz, Subtleties in the definition of IND-CCA: when and how should challenge decryption be disallowed? J. Cryptol. **28**(1), 29–48 (2015)
9. L. Xue, Y. Yu, Y. Li, M.H. Au, X. Du, B. Yang, Efficient attribute-based encryption with attribute revocation for assured data deletion. Inf. Sci. **479**, 640–650 (2019)
10. A. Balu, K. Kuppusamy, Ciphertext-policy attribute-based encryption with user revocation support, in *Proceedings of International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness* (2013), pp. 696–705
11. J. Wei, W. Liu, X. Hu, Secure and efficient attribute-based access control for multiauthority cloud storage. IEEE Syst. J. **12**(2), 1731–1742 (2018)
12. Q. Li, J. Ma, R. Li, X. Liu, J. Xiong, D. Chen, Secure, efficient and revocable multi-authority access control system in cloud storage. Comput. Sec. **59**, 45–59 (2016)
13. K. Yang, X. Jia, Expressive, efficient, and revocable data access control for multi-authority cloud storage. IEEE Trans. Parallel Distrib. Syst. **25**(7), 1735–1744 (2014)
14. J. Hao, C. Huang, J. Ni, H. Rong, M. Xian, X.S. Shen, Fine-grained data access control with attribute-hiding policy for cloud-based IoT. Comput. Netw. **153**, 1–10 (2019)
15. J. Li, W. Yao, Y. Zhang, H. Qian, J. Han, Flexible and fine-grained attribute-based data storage in cloud computing. IEEE Trans. Serv. Comput. **10**(5), 785–796 (2017)
16. N. D. Hua, M.J. Feng, Enhanced cloud storage access control scheme based on an attribute. J. Commun. **34**(Z1) (2013)
17. Z. Wang, D. Huang, Y. Zhu, B. Li, C.-J. Chung, Efficient attribute-based comparable data access control. IEEE Trans. Comput. **64**(12), 3430–3443 (2015)
18. J. Li, X. Huang, J. Li, X. Chen, Y. Xiang, Securely outsourcing attribute-based encryption with checkability. IEEE Trans. Parallel Distrib. Syst. **25**(8), 2201–2210 (2014)
19. Z. Liu, Z. Cao, D.S. Wong, White-box traceable ciphertext-policy attribute-based encryption supporting any monotone access structures. IEEE Trans. Inf. Forensics Sec. **8**(1), 76–88 (2013)
20. H.-J. Seo, H.-W. Kim, Attribute-based proxy re-encryption with a constant number of pairing operations. Int. J. Inf. Commun. Eng. **10**(1), 53–60 (2012)
21. H. Li, L. Pang, Efficient and adaptively secure attribute-based proxy re-encryption scheme. Int. J. Distrib. Sens Netw. **12**(5) 2016. Article No. 5235714
22. H. Wang, L. Wu, Unbounded anonymous hierarchical identity-based encryption in the standard model. J. Netw. **9**(7), 1846–1853 (2014)
23. G. Wang, Q. Liu, J. Wu, M. Guo, Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers. Comput. Sec. **30**(5), 320–331 (2011)
24. Z. Wan, J. Liu, R.H. Deng, HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing. IEEE Trans. Inf. Forensics Sec. **7**(2), 743–754 (2012)
25. S. Chenthara, K. Ahmed, H. Wang, F. Whittaker, Security and privacy-preserving challenges of e-Health solutions in cloud computing. IEEE Access **7**, 74361–74382 (2019)
26. J. Zhang, B. Chen, Y. Zhao, X. Cheng, F. Hu, Data security and privacy-preserving in edge computing paradigm: survey and open issues. IEEE Access **6**, 18209–18237 (2018)
27. M. Adjedj, J. Bringer, H. Chabanne, B. Kindarji, Biometric identification over encrypted data made feasible, in *Proceedings of 5th International Conference on Information Systems Security* (2009), pp. 86–100

28. D. Cash, J. Jaeger, S. Jarecki, C.S. Jutla, H. Krawczyk, M.-C. Rosu, M. Steiner, Dynamic searchable encryption in very-large databases: data structures and implementation, in *Proceedings if NDSS Symposium* (2014), pp. 23–26
29. S. Kamara, C. Papamanthou, T. Roeder, Dynamic searchable symmetric encryption, in *Proceedings of ACM Conference on Computer and Communications Security* (2012), pp. 965–976
30. B. Zhu, B. Zhu, K. Ren, PEKsrand: providing predicate privacy in public-key encryption with keyword search, in *Proceedings of IEEE International Conference on Communications*, June 2011, pp. 1–6
31. W. Sun, S. Yu, W. Lou, Y.T. Hou, H. Li, Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. IEEE Trans. Parallel Distrib. Syst. **27**(4), 1187–1198 (2016)
32. M.R. Clark, K. Stewart, K. Stewart, Dynamic, privacy-preserving decentralized reputation systems. IEEE Trans. Mob. Comput. **16**(9), 2506–2517 (2017)
33. N. Busoma, R. Petrlic, F. Sebé, C. Sorge, M. Valls, A privacy-preserving reputation system with user rewards. J. Netw. Comput. Appl. **80**, 58–66 (2017)
34. Y. Lai, Z. Liu, Q. Pan, J. Liu, Study on cloud security based on trust spanning tree protocol. Int. J. Theor. Phys. **54**, 3311–3330 (2015)
35. Q.I. Xia, E.B. Sifah, K.O. Asamoah, J. Gao, X. Du, M. Guizani, MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access **5**, 14757–14767 (2017)
36. V.V. Rajendran, S. Swamynathan, Hybrid model for dynamic evaluation of trust in cloud services. Wirel. Netw. **22**(6), 1807–1818 (2016)
37. C. Xu, J. Wang, L. Zhu, C. Zhang, K. Sharif, PPMR: a privacy-preserving online medical service recommendation scheme in eHealthcare system. IEEE Internet Things J. **6**(3), 5665–5673 (2019)
38. Y. Dou, H.C.B. Chan, M.H. Au, A distributed trust evaluation protocol with privacy protection for intercloud. IEEE Trans. Parallel Distrib. Syst. **30**(6), 1208–1221 (2019)
39. F.A.M. Ibrahim, E.E. Hemayed, Trusted cloud computing architectures for infrastructure as a service: survey and systematic literature review. Comput. Sec. **8**(2), 196–226 (2019)
40. L. Chen, R. Urian, DAA-A: direct anonymous attestation with attributes, in *Proceedings of 8th International Conference on Trust and Trustworthy Computing (TRUST)*, Heraklion, Greece, August 2015, pp. 228–245
41. I. Khalil, A. Khreishah, M. Azeem, Consolidated identity management system for secure mobile cloud computing. Comput. Netw. **65**(2), 99–110 (2014)
42. M. Chen, W. Li, Z. Li, S. Lu, D. Chen, Preserving location privacy based on distributed cache pushing, in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Turkey, April 2014, pp. 3456–3461
43. Z. Brakerski, V. Vaikuntanathan, Efficient fully homomorphic encryption from (standard) LWE. SIAM J. Comput. **43**(2), 831–871 (2014)
44. M. Sookhak, F.R. Yu, M.K. Khan, Y. Xiang, R. Buyya, Attribute-based data access control in mobile cloud computing: taxonomy and open issues. Fut. Gener. Comput. Syst. **72**, 273–287 (2017)
45. G. Ateniese, K. Fu, M. Green, S. Hohenberger, Improved proxy re-encryption schemes with applications to secure distributed storage. ACM Trans. Inf. Syst. Sec. **9**(1), 1–30 (2006)
46. J. Ni, X. Lin, X.S. Shen, Toward edge-assisted internet of things: from security and efficiency perspectives. IEEE Netw. **33**(2), 50–57 (2019)
47. S. Bragadeesh, U. Arumugam, A conceptual framework for security and privacy in edge computing, in *Edge Computing*. (Springer, 2019), pp. 173–186
48. W.Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, A. Ahmed, Edge computing: a survey. Futur. Gener. Comput. Syst. **97**, 219–235 (2019)
49. D. Liu, Z. Yan, W. Ding, M. Atiquzzaman, A survey on secure data analytics in edge computing. IEEE Internet Things J. **6**(3), 4946–4967 (2019)