Privacy Preserving Approaches in Cloud Computing

Mrabet Manel
National school of Computer Sciences
University of Manouba
Tunis, Tunisia

Yosra ben Saied, Leila azzouz Saidane
National School of Computer Sciences
University of Manouba
Tunis, Tunisia

***Abstract*— Cloud computing is becoming increasingly popular and promises to revolutionize the future of IT service delivery. However, security and privacy concerns continue to hinder cloud adoption. This paper highlights major privacy issues and reviews existing privacy preserving approaches in cloud computing. Some of the challenges and open issues associated with privacy preserving concern in cloud computing are also discussed.**

***Keywords*— Cloud computing; privacy preserving; anonymity; security.**

I. INTRODUCTION

Nowadays, cloud computing is a major concern for the evolution of IT industry in favor of dematerialization of IT infrastructure and rationalization of IT investments worldwide. In fact, users benefit from different services provided by resource-intensive data centers, thus, reducing their expenses while achieving efficiency gains in every field. Most IT big companies have embraced the Cloud like Google, Microsoft, Rightscale, Rackspace, IBM, Oracle and Cisco. The various features of cloud computing which have made enterprises and other users to shift to cloud include:

- Elasticity: NIST [1] defines elasticity as the ability for customers to quickly request, receive, and later release as many resources as needed. This term is different from scalability that means the ability of the system to be enlarged to a size which was expected to accommodate a future growth [2].
- Virtualization: virtualization represents a foundational element of cloud for its various benefits such as flexibility, isolation and high resource utilization rate [3].
- Resource pooling: computing resources used to provide the cloud service are realized using a homogeneous infrastructure that is shared between all service users [1].
- Availability: availability refers to the uptime of a system, a network of systems, hardware and software that collectively provide a service during its usage [1]. It is extremely important for the cloud providers to offer environments that are high in availability anytime and anywhere.
- Pricing: completely based on usage. The users are billed based on the amount of resources they use. In [4] overviews cloud computing pricing models.

However, due to its key features, cloud computing is more prone to security threats and vulnerabilities. Following are examples of vulnerabilities [5] with root causes in one or many of these cloud's essential key features:

- Virtual machine escape: an attacker might successfully escape from a virtual environment.
- Data leakage by virtual machine replication related to the use of cloning for providing on-demand service
- Insecure or obsolete cryptography: because novel methods of breaking many cryptographic mechanisms are increasingly discovered, and cryptography is an essential factor for cloud computing broad uptake, this vulnerability must be considered as highly relevant to cloud computing.
- Data recovery vulnerability: given the resource pooling and elasticity key features, resources allocated to a one user may be reallocated to another one, at a later time. In a storage service, this last might recover data written by the previous user.
- Unauthorized access to management interface: for on-demand self-service, the probability that unauthorized user could access to the management interfaces, which enable cloud customers to monitor and manage their cloud resources and data, is extremely higher than in traditional systems.
- Vulnerabilities of shared network components such as vulnerabilities in DNS server and IP protocol vulnerabilities might enable network-based-cross-tenant attacks in cloud infrastructure environments.
- Privacy issues related to data accessed from a third party: protecting sensitive data such as patient's medical records or credit cards details from untrusted cloud providers is also a challenging issue in cloud computing
- Metering and billing evasion: relevant vulnerabilities include metering and billing the delivered service.

Most of the listed vulnerabilities are with regard to the security and privacy topics in cloud computing. Therefore, security and privacy issues should be addressed as a matter of high and urgent priority.

Commonly, the term privacy is used as synonym to anonymity, confidentiality, and security. However they are different concepts and should never be used interchangeably without regard to their respective definitions.

Security refers to measures used to protect information. It may include the concept of privacy but the two are not synonymous.

Indeed, privacy refers to a user's right to safeguard his data content from any other parties. An important aspect of privacy is the right of a user to self-control his data. It is manipulated by contrasting two guarantees: confidentiality and anonymity. Confidentiality refers to an implicit or explicit agreement that no traceable record of the user's data will be disclosed. So that, only the requestor or the user knows the response. Anonymity, on the other hand, refers to a condition in which the requestor does not know the identity of the respondent [6].

Privacy preserving has been an important concern that promotes the success of cloud computing. Many privacy preserving solutions have been recently proposed in cloud computing environments. These solutions are based on different techniques and methods such as traditional encryption, access rights and policies, access control, and data or user anonymization. Either applied to cloud consumer or cloud provider, these solutions involve different levels of services delivery in cloud computing and cover data computation, storage and communication. Despite the diversity of solutions tackling the problem of privacy preserving in cloud computing, there are no published works addressing the state-of-the-art of privacy preserving concern in cloud computing. In this sense, we propose a survey that focuses on studying privacy preserving approaches proposed in cloud computing aiming to provide a better understanding of the current research issues in this field and to give guidelines for a further enhancement of these solutions according to the cloud computing requirements.

The rest of this paper is organized as follows. In section 2 we discuss privacy issues in cloud computing environments. In section 3 we overview privacy preserving existing approaches in cloud environments. Section 4 compares the overviewed solutions based on different criteria and gives suggestions for future work. Finally, section 5 concludes the paper.

II. PRIVACY ISSUES IN CLOUD COMPUTING

Among the privacy issues of cloud [7] [8][9][10], we distinguish:

- Malicious behavior of the cloud provider: The cloud provider may be malicious or curious and may inappropriately access, use or mine the users data. Indeed, an untrusted service provider can take encrypted user's data and evaluate or analyze them repeatedly in an oblivious manner in order to gain access over user's data. Frequency analysis attack [7], repeated evaluation of user's queries and surface analysis attack can be considered as major threats that could compromise privacy of outsourced data in cloud servers. Some solutions preventing the server from learning about the user's data are presented later in this survey.
- Lack of user control: users' visibility and control over their data and applications is limited because that are stored and processed in the cloud. Indeed, the consumers no longer own the infrastructure. So, they do not manage or control the underlying cloud infrastructure, but they have limited management and monitoring capabilities offered by the service provider via management interfaces. Moreover, the control capabilities provided to the consumer are dependent of the service model.
- Malicious outsiders: Multi tenancy and resource pooling characteristics of cloud can threat privacy of cloud users. In fact, the share of virtualized and pooled resources in multi-tenant public cloud environments, free trial offers and unlimited access of network and resources at a lower price can allow vicious cloud service subscribers to target the data of legal users who share the same resources and not only violate their protections but also spread to other victims.
- Achieving regulatory compliance: difficulty to ensure regulatory compliance between different places around the world which may cause legislation problems when deciding to jump from one cloud provider to another in different location.
- Data proliferation: Majority of cloud providers ensure data duplication and backups in several datacenters in different locations [8]. In addition, data flow in the cloud and potentially across are uncontrollable by the data owners. Accordingly, it can be possible to violate data by a certain third party server.
- Information disclosure: There are many challenges related to the data encryption and data homomorphic encryption in cloud. As a matter of fact, there are many limitations of encryption solutions that exist today such as the complexity of operations, unpractical solutions, poor encryption keys, the need to perform some operations on decrypted data in cloud servers. So that, disclosure of sensitive data can cause harm to user's data privacy.
- Dynamic provision: responsible party to ensure data privacy is unclear, due to the dynamic nature of cloud. Moreover, due to the dynamic provisioning of cloud subcontractors which are involved in user's data processing, some services might have a malicious source. As a consequence, the user is no longer ensured that his data will be treated properly and he no longer completely trusts the sub providers.[10]
- Unauthorized secondary usage: There is a risk of unauthorized usage of the data stored or processed in the cloud. For example, the cloud provider can sale companies' private data to their competitors to gain income.

III. PRIVACY PRESERVING SOLUTIONS IN CLOUD COMPUTING: AN OVERVIEW

Several methods have been presented to tackle this issue of privacy preserving. It is important, that the privacy has to be preserved anytime and anywhere [9]. The following section overviews some existing solutions raised by academics researchers.

A. Privacy preserving search in data clouds

[7] proposed a technique that normalizes the histogram of the features tables. Thereafter, they encrypt the entities of the normalized tables using homomorphic encryption. This technique prevents the malicious cloud from having an

informed guess regarding the inputs and prevents from any analysis attack such as frequency analysis attack.

The suggested model hides zeros and high frequently occurred values as well as any relation between documents since it keeps the zeros and improve the retrieval efficiency. The technique was applied on three different datasets. Cao [11] and Gopal [12] techniques were discussed. Average Precision Value is used to calculate the retrieval efficiency of the compared techniques.

The simulation results showed that the scheme works better than the two discussed techniques which are summarized in table I.

TABLE I. COMPARISON BETWEEN THE DISCUSSED TECHNIQUES

Problem	**Techniques**		
	Gopal technique	***Cao technique***	***Proposed technique***
Hiding zero's	No	Yes	Yes
Relations between documents	Can be deduced	Hard to deduce	Hard to deduce
Retrieval efficiency	Higher than Cao	Lower than Gopal	Higher than both

B. *Anonymization approach for privacy preserving in cloud computing*

[13] proposed a new anonymity algorithm for the cloud computing services. This anonymity algorithm anonymizes the data before sending it to service providers. The authors claim that their method is more flexible and safe to protect users privacy than cryptographic techniques. However, their anonymization method differs from cloud service provider to another. Hence, in interconnected cloud computing environments, the clouds may collaborate to re-identify the original data easily. Furthermore, the authors do not prove or evaluate the efficiency of their method.

C. *Privacy preserving data sharing with anonymous id assignment*

[14] reviewed various algorithms based on private data sharing with anonymous IDs and proposed a new algorithm for assigning anonymous IDs.

The studied algorithms consist of the secure sum, transmitting simple data with power sums, and sharing complex data with AIDA. This last, is the best for complex data and large number of nodes and has several variants depending on the choice of the data sharing method. The expected number of rounds depends only on the selection of IDs and not on the variant chosen.

Authors present their algorithms which assign nodes iteratively and anonymously ID numbers ranging from 1 to N, based on Newton's identities and Sturm's theorem. Private communication channels are also used for collision avoidance. All the computations are performed by the nodes without requiring a trusted central authority. The advantages of the solution are the number of rounds is reduced and the anonymity of DB is not affected by inserting the records. An algorithm for the distributed solution of certain polynomials over finite field, which enhances the scalability of the algorithms, is deployed. Experimental results prove the efficiency of the algorithms.

Cloud environments are an area of interest for future extensions to AIDA. For example, a cloud consumer may seek services without revealing its identity.

D. *A privacy preserving Markov model for sequence classification*

In [15] proposed a method to train the Markov models of the first order and of order k, on sequence data distributed among parties without disclosing each party's own sequences to others. In this model, all the parties are assumed to be semi-honest which do not disrupt the protocol execution, but they would try to derive the sensitive information of others. To address this problem, they have to calculate the prior probabilities and the transition probabilities preserving privacy without the need of an authority or a server. In this research, an additive homomorphic cryptographic system is applied to perform the encryptions and the decryptions of data. The secure logsum protocol is utilized in probabilities computations. The private key is additively shared by all the parties and no party knows the complete private key. The technique was applied on two real-world datasets and simulation results show that the information loss in this algorithm is very low.

In cloud computing environments, the parties could be the cloud service providers. In fact, we assume each hospital has a datacenter which performs the calculations without revealing its sensitive data to other hospitals or mainly datacenters.

E. *Access-private outsourcing of Markov chain and random walk based data analysis applications*

Sometimes, even though data owners and data users encrypt their data and their queries respectively, a malicious service provider can take encrypted user queries, described in terms of initial state distributions and evaluates them iteratively in order to leak the input values. As a kind of this attack type, condition evaluation attack consists of derivation of the feasible region of input vectors through repeated evaluation of the conditions.

[16] focused on this attack, and developed an algorithm based on methodical addition of extra dummies states to the Markovian process in order to enlarge and unbound the feasible region for the inputs. In the enlargement, the expanded Markov model has to have the same semantics as the original one. Furthermore, the expansion process involves matrixes, which lead to computation and storage overheads. So the maximization of the feasible region should be done in minimal cost.

This research is restricted by using diagonalizable transition matrixes. However this assumption is not practically guaranteed. Therefore, it can be possible to extend this work to more general class of Markov chains with transition matrixes in Jordan normal form.

F. *Privacy-preserved access control for cloud computing*

In Miao Zhou et al. [17] presented a new scheme called" the encryption system" that achieves a flexible and fine-grained access control on the outsourced data in cloud environments. For this purpose, the scheme consisted of two phases: the base phase and the surface phase. At the first phase, the data owner performs a policy hidden attribute-set based encryption on the data files. To be more precise, he encrypts the data files and associates each data file with an access structure from which different attribute sets are generated. The attribute sets are used to generate the private keys of the privileged users which allow them to decrypt the data files after downloading them from the cloud server.

On the other hand, the second phase is performed by the cloud providers, after launching initialization by the data owner. This phase implements the Server Re-encryption Mechanism (SRM). In the SRM, the cloud server performs the re-encryption of data files when receiving request messages from the data owner, and without requiring the file content and user's information disclosure as well as the decryption keys for re-encrypted data files update. This task is transferred to the cloud server instead of the data owner due to consumers' dynamic join and leave, high availability, abundant storage capacity and computation power of cloud servers. The authors proved the security of their scheme under the standard security model. Nevertheless, the access matrix contains only zeros and ones. Hence, the cloud server may infer any information about the users and data files such as the number of users of the same class or any other kind of information. As a result, the proposed scheme is not secure against to the zero's attack and frequency attacks which can be performed by malicious or honest but curious cloud provider.

Certainly, in [16] and [17] consider both the privacy problem about the Markov model. Yet, the raised problems are different in each work. In [16], all the computations are performed by the data owners without revealing each one's private data to others and without requiring a third party server, in order to train stronger Markov models. On the other hand, in [17] the Markov model has already been learned by the data owner. The user's queries have to be tested against the model. Both of the data owner and the user encrypt and expand their information systematically before sending them to a server which is considered to be malicious.

In addition, [13], [14], and [17] are generally based on anonymization. In [13] the authors worked on data anonymity. However, in [14] and [17], users identification and users access control are anonymized.

We summarize the privacy preserving solutions discussed above in table II.

TABLE II. PRIVACY PRESERVING APPROACHES IN CLOUD COMPUTING

Approach	*Description and used techniques*
Privacy preserving search in data clouds [7].	Hiding the data ,before encrypting it using homomorphic encryption.
Anonymity-based method [13].	Anonymises the sensitive data before storing in cloud.
Privacy-preserved authentication [14]	Private data sharing with anonymous ID assignment
Privacy preserving data computation [15].	Performing multi party data computation without disclosing each party's own data. A party can be a cloud provider
Privacy Preserving Data Outsourcing [16].	Guarantees privacy by means of data enlargement to unbound feasible regions for the inputs.
Privacy-preserved Access Control [17].	Determines access rights for users and achieves access control.

IV. PRIVACY PRESERVING SOLUTIONS IN CLOUD COMPUTING: A COMPARISON AND OPEN ISSUES

Table III classifies all described privacy preserving solutions according to the platform, pros, cons, use of cryptographic techniques, scope, perspective, scalability, and open issues.

The scope characteristic defines where the privacy preserving solutions are ensured: in authentication, authorization, data outsourcing, data computation, data storage, or access control.

Privacy preserving is essential to the concept of cloud computing being actually used to ensure the privacy of cloud user or cloud provider. The perspective characteristic means that the privacy is applied to cloud provider or cloud user. It has been mostly applied to cloud users.

Most of the proposed approaches enhance the system scalability and availability because the data and the applications are outsourced to remote cloud servers, which reduce the overhead. In secure multi-party computation, the scalability is enhanced because the computation is distributed among different parties, which reduce cost and expenses.

The keys of cryptography can be used by the data owner, the data requester, or the service providers. In [7] the data owner creates his own keys and sends them to the client to decrypt the data. The data is hidden from a third party server. However, in [16] the data owner does not send his own key to the data user. He hides his data from the server and the user. Similarly, the data user hides his queries from the data owner and the server using his own keys. In [14] private secured communications channels are used instead of cryptographic keys.

Although many privacy preserving approaches have been developed in academic researches, some existing issues have not been fully addressed. We summarize in table III some of the challenges related to the use of privacy preserving approaches in cloud computing.

V. CONCLUSIONS

The objective of this paper is to present a comprehensive study about privacy preserving approaches in cloud computing environments. Firstly, we discussed privacy issues in cloud computing environments. Then, diverse related works were reviewed in order to define the state of the art of privacy preserving in clouds environments. Finally, we have discussed about the pros, the cons and many other characteristics of the existing solutions in order to solve privacy issues in cloud computing.

As we can observe in this paper, the concept of privacy in cloud computing environments is very broad and involves storage, computation, communication, access control, authorization, authentication and private information retrieval. In a perfectly private cloud computing, much of the privacy solutions, from the most basic to more elaborate ones, should be implemented together to ensure the privacy of both cloud users and cloud providers in different service models and applications. But for this to become reality, some issues must be addressed subsequently, in order to eliminate all obstacles that would prohibit the wide adoption of cloud computing.

In the near future, we will build our solution of cloud computing based on this comparative study. Then, we will focus on a large scale experimentation to assess performance of our infrastructure compared to the current solutions.

TABLE III. COMPARISON OF PRIVACY PRESERVING SOLUTIONS IN CLOUD COMPUTING

Approaches	**Characteristics**								
	Platform	***Pros***	***Cons***	***Use of cryptographic techniques***	***Scope***	***Perspective***	***Scalability and elasticity***	***Keys used by***	***Open issues***
[7]	Cloud computing	Preventing frequency attacks Improving the retrieval efficiency	The key is known by the client who can be malicious	Homomorphic encryption	Private information	Cloud user	yes	Data owner and client (the requester)	Protecting user's queries from third party server and the data owner Processing the users'queries on the data owners' data without leaking data to the users and the cloud servers
[13]	Cloud computing	The data can't be restaured by the cloud provider because he does not have the key of cryptography.	The data can be restaured by multi-cloud environments.	no	Private data storage	Cloud user	Approximately: the anonymization is very difficult in huge amounts of data and with a large number of interconnected cloud environments.	Not used	Efficiency and acuracy Proofs on effectiveness Consolidation of the method in interconnected cloud environments
[14]	distributed computing systems and cloud	Collision avoidance The number of rounds is reduced The anonymity of DB is not affected by inserting the records	Mistreatment of Sturm's theorem [18] which lead to the avoidance of the solution of a polynomial.	No	Private communications and data sharing	Cloud user/ Cloud provider	Yes	Not used	The development of a result similar to the Sturm's method over a finite field
[15]	Grid and cloud computing	Prevents from condition evaluation attacks	Increases the storage cost and thus, reduce the performance of servers	Yes	Private data outsourcing	Data owner Client(data requestor)	Yes	Data owner Data requestor	Transition matrixes in Jordan normal form
[16]	distributed computing systems and cloud	Cooperation to train stronger data models	The accuracy is approximately reduced by computations	Homomorphic encryption	Private multi party computation	Cloud provider	Yes	Parties which process the computation (interconnected cloud providers)	Accuracy
[17]	Cloud computing	Prevention from collision attack, users' information disclosure Providing flexible and fine-grained access control	Vulnerable to zero's and frequency attacks	No	Private user's access control	Cloud user	Yes	Users, data owner	Prevention from zero's and frequency attacks

REFERENCES

[1] L. Badger, R. Patt-Corner, and J. Voas, “Draft cloud computing synopsis and recommendations of the national institute of standards and technology,” Nist Special Publication, vol. 146. [Online]. Available: http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf

[2] G. Galante and L. C. E. Bona, “ A survey on cloud computing elasticity,” in Proceedings of the IEEE/ACM Fifth International Conference on Utility and Cloud Computing, UCC '12, May, 2012, Washington, DC, USA. Washington : IEEE Computer Society, 2012, pp. 263-270.

[3] N. Manohar, “A survey of virtualization techniques in Cloud Computing,” in Communication, Advanced devices, Signals and Systems and Networking: Proc. Int. Conference on VLSI, VCASAN, 2013, India, V. Chakravarthi et al. India: Springer, 2013, pp. 461-470.

[4] M. Al-Roomi et al., “Cloud computing pricing models: a survey,” IJGDC, vol. 6, no. 5, pp. 93-106, 2013.

[5] B. Grobauer, T. Walloschek and E. Stocker, “Understanding cloud computing vulnerabilities,” IEEE Security and Privacy, March/April 201, pp. 50-57.

[6] A. D. Ong, and D. J. Weiss,”The impact of anonymity on responses to sensitive questions,” J. Appl. Social Psychology ,2000, pp. 1691–1708.

[7] M. Dawoud and D. Turgay Altilar, “Privacy-preserving search in data clouds using normalized homomorphic encryption,” in L. Lopes et al. (Eds.): Euro-Par 2014 Workshops, Part II, LNCS 8806, pp. 62–72, 2014. Springer, Switzerland (2014).

[8] M. U. Shankarwar and A. V. Pawar, “Security and privacy in cloud computing: A survey,” in S.C. Satapathy et al. (eds.), Proc. of the 3rd Int. Conf. on Front. of Intell. Comput. (FICTA) 2014, 2015©Springer International Publishing Switzerland. doi: 10.1007/978-3-319-12012-6_1

[9] T. J. Neela, and N. Saravanan, “Privacy preserving approaches in cloud: A survey,” IJST, vol. 6(5), pp. 4531-4535, May 2013.

[10] S. Pearson and G. Yee, Privacy and security for cloud computing, 1st ed. London:Springer-Verlag London, 2013 .[E-book] Available: Springer, Computer communications and networks. ISBN 978-1-4471-4189-1 (eBook) ,DOI 10.1007/978-1-4471-4189-1.

[11] N. Cao et al. , “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” IEEE Transactions on Parallel and Distributed Systems vol. 25(1), pp. 222–233, 2014.

[12] G. N. Gopal and M. P. Singh, “Secure similarity based document retrieval system in cloud,” In Proc. 2012 International Conference on Data Science Engineering (ICDSE), pp. 154–159, July 2012.

[13] V. K. Saxena and S.Pushkar, “Anonymization approach for privacy preserving in cloud computing,” in Proc. International Conference on Cloud, Big Data and Trust, 2013, Nov 13-15, RGPV, pp. 179-182.

[14] L. A. Dunning and R. Kresman, “Privacy preserving data sharing with anonymous id assignment,” IEEE Trans. Inf. Forens. Security, vol. 8, no. 2, Feb. 2013.

[15] S. Guo, S. Zhong and A. Zhang, “A privacy preserving Markov model for sequence classification,” in Bioinformatics, Computational Biology and Biomedicine: Proc. the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, BCB'13, September 22 - 25, 2013, Washington, DC, USA. New York : ACM, 2013. pp. 561-568.

[16] P. Lin, and K. S. Candan, “Access-private outsourcing of Markov chain and random walk based data analysis applications,” in Proc. 22nd International Conference on Data Engineering Workshops (ICDEW’06), 2006.

[17] M. Zhou et al. , “ A privacy-preserved access control for cloud computing,” in Proc. IEEE 10th International conference on Trust, Security and Privacy in Computing and Communications: TrustCom, 16-18 Nov. 2011, Changsha: IEEE, 2011. pp. 83-90.

[18] H. Dörrie, “Sturm’s problem of the number of roots,” in 100 Great Problems of Elementary Mathematics: Their History and Solutions. New York: Dover, 1965, pp. 112–116.