IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Data Security and Privacy Protection in Cloud Computing: Technical Review of Emerging Trends

Km Komal[1] , Dr Puspneel Verma[2] ,Ajay Singh[3],Pawan Kumar Goel[4]

1. Research Scholar, Bhagwant Institute of Technology, Muzaffarnagar, UP, India
2. Assistant Director, Bhagwant Group of Institutions
3. H.O.D. ,Department of CSE, Bhagwant Institute of Technology, Muzaffarnagar
4. Associate Professor & Head Department of CSE, Shri ram Group of Colleges, Muzaffarnagar

Abstract: ***Recent improvements have given growth to the fame and achievement of cloud computing. However, when outsourcing the data and business presentation to a third party reasons the security and privacy issues to become a serious concern. Millions of users across the world leverage data processing and distribution benefits from cloud environment. Data security and privacy are predictable requirement of cloud environment. Massive usage and distribution of data among users unlocks door to security loopholes. This paper envisions a discussion of cloud environment, its utilities, challenges, and developing research tendencies confined to secure processing and sharing of data.***

KEYWORD: Cloud computing, Security, User-Privacy, Data Protection

I. Introduction

Cloud computing has started to emerge as a hotspot in both industry and academia; It signifies a new business model and computing paradigm, which enables on demand provisioning of computational and storage resources. Economic profits consist of the main drive for cloud computing due to the information that cloud computing offers an actual way to decrease capital expenditure (CapEx) and operational expenditure (OpEx).

To make cloud computing conceivable and existing to end-users, some services and models function behind the scenes. Fig. 1 shows two types of cloud computing models: deployment models and service models: A private cloud is normally one organization's infrastructure. The association or a service provider can manage such infrastructure to help different customer groups. A hybrid cloud is a group of private as well as public cloud computing resources. An additional model is a community cloud that shares in some organizations computing resources, and it is possible to manage either through organizational IT resources or third-party vendors [1][2][3]. The classification of cloud computing has been given in many literatures. cloud computing as: "A large-scale distributed computing model that is determined by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms, and services are distributed on demand to external customers over the Internet''. Basically two kind of cloud computing model:- Deployment models and service models.

The deployment models relate to where the cloud infrastructure is situated and managed. There are three typically used cloud deployment models: public, private, and hybrid cloud. Additional type of model is the community cloud which is commonly less used. A public cloud is a group of computing resources that third-party organizations provide. It help everything users who want a computer resource, including subscription-based hardware (OS, CPU, memory, storage) or software (application server, database) to be used.

II. CLOUD SERVICE MODELS

There are the following three kinds of cloud service models:

Infrastructure as a Service (IaaS): IaaS is also known as Infrastructure / Hardware. It is a computing infrastructure achieved over the internet. The main advantage of using IaaS is that it supports user to avoid the cost and complexity of buying and managing the physical servers. Example: - Amazon web services (AWS) , Microsoft Azure, Google Computing Engine(GCE).

Platform as a Service (PaaS): PaaS cloud computing platform is created for the programmer to develop, test run and manage the application. Example:- AWS Elastic Beanstalk, Window Azure, Google app engine.

Characteristics of PaaS: There are the following features of PaaS-

- Accessible to various user via the same development application.
- Integrates with web services and databases.
- Support many languages and framework.
- Provides ability to “Auto-scale”.

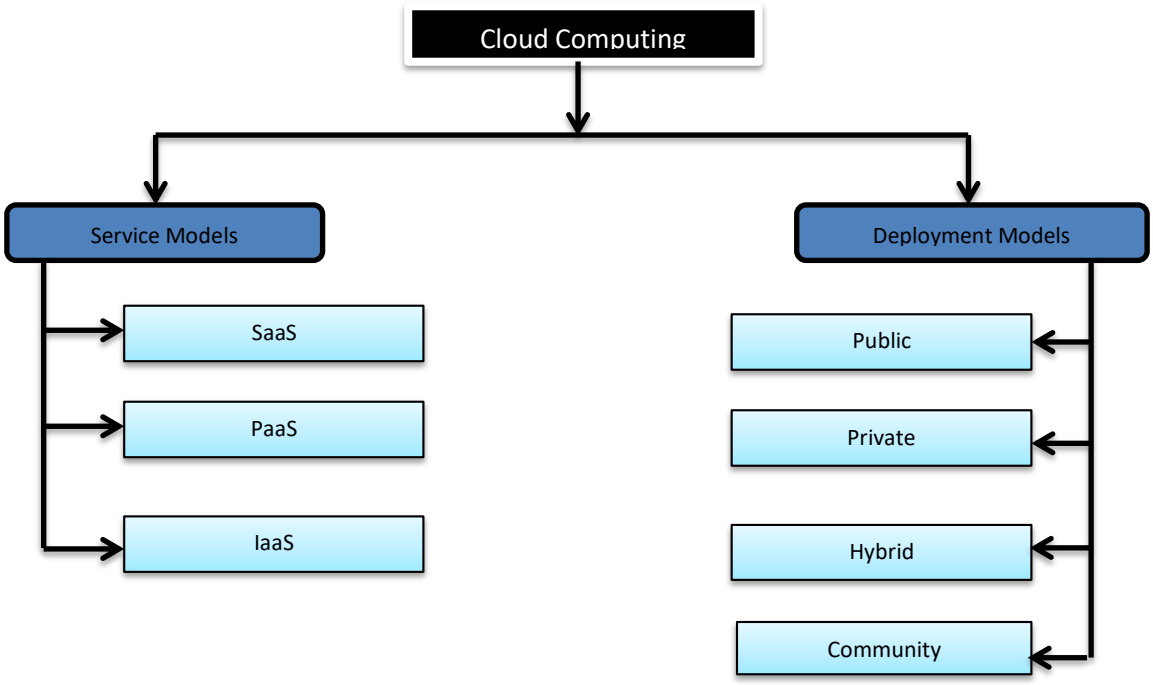

Fig-1

Software as a Service (SaaS): SaaS is also recognized as "on-demand software". It is a software in which the applications are introduced by a cloud service provider. Users can access these applications with the support of internet connection and web browser. Example:- Big Commerce, Cisco WebEx , Go To Meeting.

Characteristics of SaaS:- There are the following features of SaaS-

- Managed from a central locality
- Presented on a remote server
- Available over the internet
- The services are acquired on the pay -as- per-use basis.

III. DEPLOYMENT MODEL

Its works as simulated computing environment with a choice of deployment model depending on how much data you need to store and who has permission to the infrastructure.

- **Public Cloud:** It is available to the public. Public deployment models in the cloud are impeccable for organization with increasing and changeable demands.
- **Private Cloud:** The private cloud deployment model is the precise opposite of the public cloud deployment model. It is a one-on-one environment for a single user.
- **Community Cloud**: Community Cloud permits system and service to be available by a group of organization. it is a distributed system that is produced by integrating the services of different clouds to address the specific needs of a community, industry, or business.
- **Hybrid Cloud:** Hybrid cloud is a combination of two or more cloud architectures. A company that has critical data will desire storing on a private cloud, while less sensitive data can be stored on a public cloud. The hybrid cloud is also frequently used for "cloud bursting ".

The standard of cloud computing has allowed various users to share resources [4][5][6]. This notion has achieved broad popularity over the few years by continually increasing the number of customers and supporting substructure development [7]. For instance, Administrations like Cisco expect to have more than 24 billion gadgets connecting with the Internet by 2020. Further, Morgan Stanley has also predicted that almost 75 billion Internet devices will be in use by 2021. Cloud computing provides security to the customers, preventing their sensitive information and personal data from illegal stakeholders and several parties. Furthermore, data providers can display their outsourced data privacy at any moment. Organizations do not need to worry about data security because this facility is available as a service nowadays. Cloud computing is a rising technology that provides massive data without upfront investment to organizations with a novel business model [8][9][10] . However, most administrations and orgnization are still reluctant to explore their business across the cloud because of security. When the data is stored in the cloud server, the significances are data ownership and management separation [11]. Therefore, the Cloud service provider (CSP) can search and access data freely on the cloud server without taking the authorization of the user. Meanwhile, the cloud server might be assaulted by an aggressor to get the user's information. There are many security problems in the cloud environment, such as man-in-the-middle attacks, data leaks, etc [12][13][14] . All the above problems are enormously perils for user's information. Losses of the information of users are hampered, and there may occur data leakage problems too. If the user directly uploads data on the cloud, there may be issues like high bandwidth requirement, high latency, and a large volume of data. These are the biggest problems limiting cloud development. Therefore, it is necessary to work on it and find the right solution. Fig. 2 exhibits consumer fraud and identity theft

complaints wedged during the period ranging from 2016 to 2021. It is observed that 5.8 million identity theft and fraud complaints were received in the year 2021. Out of these, 2.4 million complaints were identity theft, 25 percent of cases also testified money loss amounting, and remaining 3.2 million were fraud complaints. The number of complaints in 2021 is increased up to 46 percent over the previous year. To address the problems mentioned above, the state-of-art encryption schemes [15][16] were used to protect the users data. These structures enable the users to encrypt their data and store it on the cloud servers. Afterward, cloud servers can carry out computations on it. But, Cloud servers are presently unable to provide advanced perspectives to users while fully maintaining their privacy.

IV. PRIVACY-PRESERVING BASED ON CRYPTOGRAPHY MECHANISM

Yuan and Yu [17] introduced a multiparty Back-Propagation Neural (BPN) network-based approach that is accurate, efficient, and secure for collaborative learning over arbitrarily divided data. To conduct operations over cipher texts, they used a doubly homomorphic encryption technique. But they focused on enhancing data processing rather than the algorithm's efficiency.

Zhang et al.[18] proposed a privacy-preserving deep computation model based on homomorphic encryption. They used the divesting of the expensive operations to improve the learning features of the cloud. The exponential process mandatory by the sigmoid function was utilized using the Taylor theorem. However, the model only contains addition and multiplication operations.

Yonetani et al.[19] proposed a privacy-preserving mechanism based on a double-permitted homomorphic encryption (DPHE) scheme, which effectually learns visual classifiers across circulated private data. This scheme provided multiparty protected scalar products while minimizing the computational cost for high-dimensional classifiers. Nevertheless, either addition or multiplication operation can only support at a time. A deep learning system based on additively homomorphic encryption was presented in [20]. The introduced system protects gradients from the curious server. Asynchronous stochastic gradient descent (ASGD) trained overall participants' joint datasets obtained the same accuracy as the connected deep learning system. However, the modified parameters are decrypted by the owner's secret key; thus, their model does not ensure parameter privacy. A basic scheme based on multi-key fully homomorphic encryption (MK-FHE) mechanism was introduced in [21]. The authors devised an advanced model for learning encrypted data in the cloud that uses the double decryption mechanism and fully homomorphic encryption (FHE) mechanism. But this scheme has a high cost in terms of computation and communication. To perform the deep neural network algorithms over encrypted data, a framework named CryptoDL was proposed by Hesamifard et al.[22] To address the existing limitations of homomorphic encryption schemes, they designed neural networking techniques. However, the proposed algorithm protects the owner's data by using keys that is not practical. A privacy-preserving outsourced classification in cloud computing (POCC) framework was presented in [23], which efficiently enable an arbitrary number of multiplication and addition operations on cipher texts. The data and query were protected by providing a proxy fully homomorphic encryption based on Gentry's scheme. Nevertheless, the cost of calculation and communication was increased in the proposed framework.

Ma et al. [24] proposed a privacy-preserving deep learning model, namely PDLM, to train the model over the data encrypted by the owners' keys. A privacy-preserving calculation toolset based on stochastic gradient descent (SGD) was utilized to accomplish the training task in a privacy-preserving way. Although the model reduced storage overhead, it has a high computation cost and lower classification accuracy.

A privacy-preserving outsourced classification scheme is presented in [25] , which delivers the classification services over encrypted data for users. They also designed two concrete secure classification

protocols for the Naive Bayes classifier and the hyperplane decision-based classification, respectively. But during the launch of a classification query, user interactions are frequently involved in this scheme.

Gao et al.[26] introduced a privacy-preserving Naive Bayes classifier scheme that prevents information leakage under the substitution-then comparison (STC) attack. A double-blinding method was adopted to protect the Naive Baye's privacy. Both the communication and processing overhead were decreased, but unable to discover the truth while continuing privacy. Phong and Phuong [27] introduced two systems, namely the Server-aided Network Topology (SNT) system and the Fully-connected Network Topology (FNT) system based on the connection with SNT and FNT server to defend the SGD privacy. The SNT and FNT systems realized an accuracy corresponding to SGD using weight parameters instead of gradient parameters. These systems are both effective and efficient in terms of computing and communication. Table 1 shows the summary of privacy-preserving of data based on the cryptography mechanism.

Literature reference	approach	Pros	Cons
Yuan et al.[17]	BGV fully and Doubly homomorphic encryption	Encrypt data efficiently secure scalar product	Low efficiency
Yonetani et al.[19]	Doubly homomorphic Encryption	Supported multi-party secure scaler product	Both addition and multiplication
Li et al.[21]	Multi- key fully homomorphic encryption	Preserve privacy of sensitive data	Low efficiency
Ma et al.[24]	Distributed two trapdoor public key cryptosystem	More efficiency	Less accuracy
Shokri and shmatikov[28]	Optimization algorithms	Protect training data	High complexity
Wang et al.[29]	Homomorphic encryption	Encrypted data by randomly splitting numerical	Either addition or multiplication
Chan et al.[30]	Homomorphic and EIGamal scheme	Partitioned data vertically	Two party
Bansal et al.[31]	Homomorphic and EIGamal scheme	Partitioned data arbitarily	Two party
Samet et al [32]	BPNN and Extreme learning machine scheme	Partitioned data horizontally ,vertically with multi parties	High communication cost
Cao et al.[33]	KNN, TF-IDE and dynamic data operation	Solve the problem of multi keyword ranked search	Does not explore checking integrity
Guo et al.[34]	Tf – IDF model	Update the outsourced data at less cost	Does not improve computational efficiency
Fu et al.[35]	Stemming algorithm accuracy improvement	Outsourced data with at less cost	Does not reflect the keyword weight
Fu et al.[36]	Keyword based search scheme	Use two cloud server for store and compute	High complexity
Huang et al.[37]	Public key authenticated encryption	Handle to resist inside keyword	Single keyword
Qi et al.[38]	Locality – Sensitive hashing	Handle the service recommendation	Low efficiency

Table -1

V. PRIVACY-PRESERVING BASED ON PERTURBATION MECHANISM

Dwork et al.[39] first introduced differential privacy and obtained complete background knowledge under the attacker's hypothesis. To prevent data privacy, the randomly generated noise is disturbed according to a specially selected distribution. Fletcher and Islam [40] introduced a differential privacy decision-making random forest algorithm to decrease the query times and sensitivity. This scheme also minimizes the amount of noise that must be appended to defend the privacy and recover data availability. However, there is no consideration for the distributed situation where multiple data owners conduct collaborative data mining. To perform privacy-preserving machine learning over cloud data from different data providers, Li et al.[41] proposed a scheme that protects the data sets of various providers and the cloud. They used the public-key encryption with a double decryption algorithm (DD-PKE) to encrypt the data sets of the different providers with different public keys and -differential privacy to add statistical noises into data to defend the privacy. Their scheme improved computational effectiveness and data analysis accuracy. But such fully homomorphic (FHE) encryption schemes are normally low efficiency. A privacy-preserving Naive Bayes learning scheme with numerous data sources is presented in[42] . The proposed scheme enabled a trainer to train a Naive Bayes classifier over the dataset provided jointly by different data owners without the help of a trusted curator. However, collaboration is permitted, or adversaries can forge and modify the data in this scheme. A distributed agent-based privacy-preserving framework, namely DADP, was proposed by Wang et al[43]. The proposed framework collects real-time spatial statistics data and publishes it with an untrusted server. To achieve global w-event -differential privacy in a distributed manner, they utilized a distributed budget allocation mechanism and an agent-based dynamic grouping mechanism. The noise is added to crowd-sourced data using the Laplace technique in DAPM. It started a batch of reliable proxies (Agents) and anonymous connection technology to safeguard users' privacy under an untrusted server. Therefore, it was regarded as a semi-centralized setting and resulted in a more complex system. An efficient privacy-preserving scheme based on machine learning was proposed by Hassan et al[44]. Authors adopted a partially homomorphic encryption technique to encrypt data, and noised is added by applying a differential privacy mechanism. It allows all parties to publicly check the ciphertext's correctness via a low-cost unidirectional proxy re-encryption (UPRE) mechanism. However, the proposed system shared fewer data. A private decision tree algorithm based on the noisy maximal vote was introduced in [45]. To strike a balance between accurate counts and noise, an effective privacy budget allocation approach was utilized. The main aim of constructing an ensemble model is to increase the accuracy and stability by using differential privacy. The proposed algorithm performs the privacy analysis on each individual tree rather than the ensemble as a whole.

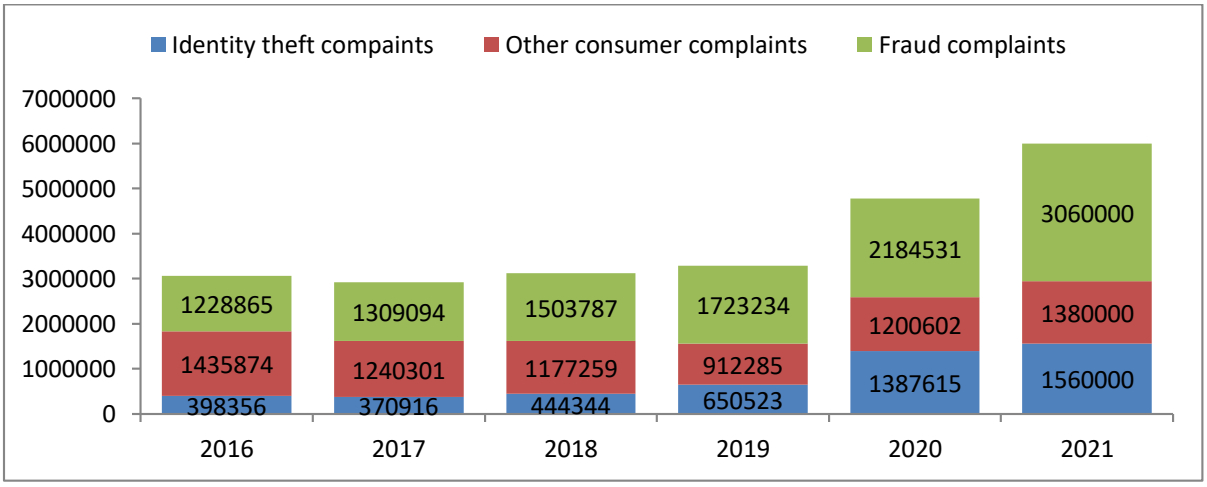

Fig-2

Gupta et al.[46] introduced a machine learning and probabilistic analysis-based model, namely MLPAM. The authors used encryption, machine learning, and probabilistic approaches to share the several participants' data and minimize the risk affiliated with the leakage for prevention and detection. However, MLPAM is not able to give security to the classifier. To preserve the privacy of the data as well as query processing, Sharma et al.[47] introduced a Differential Privacy Fuzzy Convolution Neural Network framework, namely DP-FCNN. The Laplace mechanism was used to inject the noise and encrypt the data by applying the lightweight Piccolo algorithm. The key properties were extracted using the BLAKE2s technique. But, DP-FCNN enhanced the computational overhead. Table 2 summaries the privacy-preserving data using the Perturbation Mechanism.

Literature reference	Approach	Pros	Cons
Fletchher and islam[40]	Differential privacy	Reduce query time	Less accuracy
Li et al[41]	Homomorphic encryption differential privacy	More efficiency	High computation cost
Li et al[42]	Differential privacy	Preserve privacy of data	Forge data
Wang et al[43].	Event Differential privacy	Data protection	Complex system
Hassan et al.[44]	Homomorphic encryption Differential privacy	Low cost	Limited data sharing
Liu et al.[45]	Laplace mechanism	Balance between accuracy and noise	Individual privacy
Gupta et al.[46]	Gaussian mechanism	High accuracy	No classifier protection
Sharma et al[47]	Laplace mechanism	Data protection	Computation overhead protection

Table-2

VI. RESEARCH GAPS

On the basis of the literature review, the following research gaps are identified.

1. Less security and privacy of the outsourced data.
2. The efficiency of methods to protect confidentiality must be increased.
3. Many user-based protection techniques are required.
4. Computational and communication costs during the data and information transfer must be reduced.
5. Minimization of the threats of data leakage during transmission.

VII. RESEARCH OBJECTIVES

To fill the identified study gaps, the aim described below is developed.

1. To optimize the computation and communication costs among different entities.
2. To condense data leakage by using advanced encryption techniques.
3. To solve the security problem of cloud computing.
4. To express the approach to reduce the latency while preserving data privacy during transmission.

VIII. JUSTIFICATION FOR THE OBJECTIVES

In an age of quickly rising information generation, it requires a computing infrastructure that can store and process huge amounts of information. Developing these new cloud-based methods defines present safety and privacy. Data Security & Privacy in Cloud Computing A PREPRINT demands and addresses consumer problems facing quality, effectiveness, and the best use of consumer needs. We provide complete awareness of secure cloud and efficient use of resources to authenticate our methodologies because security supports prevent data leakage and the disposal of data. With the following benefit, the advances to be carried out in research work will open an era for improved cloud facilities: -

- Additional quality of service (QoS)
- Better resource use
- Improves confidence in cryptography services
- Low the cost of computation of the procedure
- Less service level agreement (SLA) violation
- Better use of the cloud for securely sharing of data among the organizations.
- Better efficiency
- Saving the cost
- Access the file universally
- Increase the security

REFERENCES

[1] Saxena, D., Gupta, I., Kumar, J., Singh, A. K., & Wen, X. (2021). A secure and multiobjective virtual machine placement framework for cloud data center. *IEEE Systems Journal*.

[2] Saxena, D., & Singh, A. K. (2021). A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center. *Neurocomputing*, *426*, 248-264.

[3] Saxena, D., & Singh, A. K. (2021). OSC-MC: Online Secure Communication Model for Cloud Environment. *IEEE Communications Letters*.

[4] Saxena, D., & Singh, A. K. (2021). OSC-MC: Online Secure Communication Model for Cloud Environment. *IEEE Communications Letters*.

[5] Singh, A. K., & Saxena, D. (2021). A cryptography and machine learning based authentication for secure data-sharing in federated cloud services environment. *Journal of Applied Security Research*, 1-24.

[6] Kumar, J., & Singh, A. K. (2021). Performance Assessment of Time Series Forecasting Models for Cloud Datacenter Networks' Workload Prediction. *Wireless Personal Communications*, *116*(3), 1949-1969.

[7] Saxena, D., & Singh, A. K. (2021). Workload forecasting and resource management models based on machine learning for cloud computing environments. *arXiv preprint arXiv:2106.15112*.

[8] Gupta, R., Saxena, D., & Singh, A. K. (2021). Data security and privacy in cloud computing: concepts and emerging trends. *arXiv preprint arXiv:2108.09508*.

[9] Chhabra, S., & Singh, A. K. (2016, December). Dynamic data leakage detection model based approach for MapReduce computational security in cloud. In *2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS)* (pp. 13-19). IEEE.

[10] Singh, A. K., & Kumar, J. (2019). Secure and energy aware load balancing framework for cloud data centre networks. *Electronics Letters*, *55*(9), 540-541.

[11] Chauhan, A. S., Rani, D., Kumar, A., Gupta, R., & Singh, A. K. (2020, May). A Survey on Privacy-Preserving Outsourced Data on Cloud with Multiple Data Providers. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

[12] Singh, A. K., & Gupta, I. (2020). Online information leaker identification scheme for secure data sharing. *Multimedia Tools and Applications*, *79*(41), 31165-31182.

[13] Deepika, D., Malik, R., Kumar, S., Gupta, R., & Singh, A. K. (2020, May). A Review on Data Privacy using Attribute-Based Encryption. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

[14] Deepika, D., Malik, R., Kumar, S., Gupta, R., & Singh, A. K. (2020, May). A Review on Data Privacy using Attribute-Based Encryption. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

[15] Kaur, K. N., Gupta, I., & Singh, A. K. (2019). Digital image watermarking using (2, 2) visual cryptography with DWT-SVD based watermarking. In *Computational intelligence in data mining* (pp. 77-86). Springer, Singapore.

[16] Kaur, K., Gupta, I., & Singh, A. K. (2017, December). Data leakage prevention: e-mail protection via gateway. In *Journal of Physics: Conference Series* (Vol. 933, No. 1, p. 012013). IOP Publishing.

[17] Yuan, J., & Yu, S. (2013). Privacy preserving back-propagation neural network learning made practical with cloud computing. *IEEE Transactions on Parallel and Distributed Systems*, *25*(1), 212-221.

[18] Zhang, Q., Yang, L. T., & Chen, Z. (2015). Privacy preserving deep computation model on cloud for big data feature learning. *IEEE Transactions on Computers*, *65*(5), 1351-1362.

[19] Yonetani, R., Naresh Boddeti, V., Kitani, K. M., & Sato, Y. (2017). Privacy-preserving visual learning using doubly permuted homomorphic encryption. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2040-2050).

[20] Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2017). Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, *13*(5), 1333-1345.

[21] Li, P., Li, J., Huang, Z., Li, T., Gao, C. Z., Yiu, S. M., & Chen, K. (2017). Multi-key privacy-preserving deep learning in cloud computing. *Future Generation Computer Systems*, *74*, 76-85.

[22] Hesamifard, E., Takabi, H., Ghasemi, M., & Wright, R. N. (2018). Privacy-preserving Machine Learning as a Service. *Proc. Priv. Enhancing Technol.*, *2018*(3), 123-142.

[23] Li, P., Li, J., Huang, Z., Gao, C. Z., Chen, W. B., & Chen, K. (2018). Privacy-preserving outsourced classification in cloud computing. *Cluster Computing*, *21*(1), 277-286.

[24] Ma, X., Ma, J., Li, H., Jiang, Q., & Gao, S. (2018). PDLM: Privacy-preserving deep learning model on cloud with multiple keys. *IEEE Transactions on Services Computing*.

[25] Li, T., Huang, Z., Li, P., Liu, Z., & Jia, C. (2018). Outsourced privacy-preserving classification service over encrypted data. *Journal of Network and Computer Applications*, *106*, 100-110.

[26] Gao, C. Z., Cheng, Q., He, P., Susilo, W., & Li, J. (2018). Privacy-preserving Naive Bayes classifiers secure against the substitution-then-comparison attack. *Information Sciences*, *444*, 72-88.

[27] Phuong, T. T. (2019). Privacy-preserving deep learning via weight transmission. *IEEE Transactions on Information Forensics and Security*, *14*(11), 3003-3015.

[28] Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 1310-1321).

[29] Wang, Q., Hu, S., Du, M., Wang, J., & Ren, K. (2017, May). Learning privately: Privacy-preserving canonical correlation analysis for cross-media retrieval. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications* (pp. 1-9). IEEE.

[30] Chen, T., & Zhong, S. (2009). Privacy-preserving backpropagation neural network learning. *IEEE Transactions on Neural Networks*, *20*(10), 1554-1564.

[31] Bansal, A., Chen, T., & Zhong, S. (2011). Privacy preserving back-propagation neural network learning over arbitrarily partitioned data. *Neural Computing and Applications*, *20*(1), 143-150.

[32] Saeed Samet and Ali Miri. Privacy-preserving back-propagation and extreme learning machine algorithms. Data & Knowledge Engineering, 79:40–61, 2012.
[33] Bansal, A., Chen, T., & Zhong, S. (2011). Privacy preserving back-propagation neural network learning over arbitrarily partitioned data. *Neural Computing and Applications*, *20*(1), 143-150.
[34] Guo, C., Chen, X., Jie, Y., Zhangjie, F., Li, M., & Feng, B. (2017). Dynamic multi-phrase ranked search over encrypted data with symmetric searchable encryption. *IEEE Transactions on Services Computing*.
[35] Fu, Z., Wu, X., Guan, C., Sun, X., & Ren, K. (2016). Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. *IEEE Transactions on Information Forensics and Security*, *11*(12), 2706-2716.
[36] Fu, Z., Xia, L., Sun, X., Liu, A. X., & Xie, G. (2018). Semantic-aware searching over encrypted data for cloud computing. *IEEE Transactions on Information Forensics and Security*, *13*(9), 2359-2371.
[37] Qiong Huang and Hongbo Li. An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks. Information Sciences, 403:1–14, 2017.
[38] Qi, L., Zhang, X., Dou, W., & Ni, Q. (2017). A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data. *IEEE Journal on Selected Areas in Communications*, *35*(11), 2616-2624.
[39] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.
[40] Fletcher, S., & Islam, M. Z. (2017). Differentially private random decision forests using smooth sensitivity. *Expert systems with applications*, *78*, 16-31.
[41] Li, P., Li, T., Ye, H., Li, J., Chen, X., & Xiang, Y. (2018). Privacy-preserving machine learning with multiple data providers. *Future Generation Computer Systems*, *87*, 341-350.
[42] Li, T., Li, J., Liu, Z., Li, P., & Jia, C. (2018). Differentially private Naive Bayes learning over multiple data sources. *Information Sciences*, *444*, 89-104.
[43] Wang, Z., Pang, X., Chen, Y., Shao, H., Wang, Q., Wu, L., ... & Qi, H. (2018). Privacy-preserving crowd-sourced statistical data publishing with an untrusted server. *IEEE Transactions on Mobile Computing*, *18*(6), 1356-1367.
[44] Hassan, A., Hamza, R., Yan, H., & Li, P. (2019). An efficient outsourced privacy preserving machine learning scheme with public verifiability. *IEEE Access*, *7*, 146322-146330.
[45] Liu, X., Li, Q., Li, T., & Chen, D. (2018). Differentially private classification with decision tree ensemble. *Applied Soft Computing*, *62*, 807-816.
[46] Gupta, I., Gupta, R., Singh, A. K., & Buyya, R. (2020). MLPAM: A Machine Learning and Probabilistic Analysis Based Model for Preserving Security and Privacy in Cloud Environment. *IEEE Systems Journal*.
[47] Gupta, R., Saxena, D., & Singh, A. K. (2021). Data security and privacy in cloud computing: concepts and emerging trends. *arXiv preprint arXiv:2108.09508*.