JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3057-3066
https://publishoa.com
ISSN: 1309-3452

Survey on Various Cloud Security Approaches

K.V.Daya Sagar[1], PSG Aruna Sri[2], Chinta Venkata Murali Krishna[3], Dr. BALA BRAHMMESWARA[4],[5]Sridevi Sakhamuri

[1]Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation

[2]Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur,india

[3]Associate professor, Department of Computer Science and Engineering,

NRI Institute of Technology, Agiripalli, Vijayawada, Andhra Pradesh 521212

[4]Assistant Professor,Department of CSE, Sesadri Rao Gudlavalleru Engineering College, Gudlavalleru.

5Assistant Professor,Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur,india

Corresponding author: sagar.tadepalli@gmail.com

Received 2022 April 02**; Revised** 2022 May 20**; Accepted** 2022 June 18**.**

Abstract

Cloud computing plays a significant role in effective data handling based on the increase in data usage in various real-time applications. Data auditing is performed on certain files and the authenticator with deduplication. It addresses the problem of key management to deduce the file content based on the malicious activities performed on the cloud. So, based on effective auditing of the integrity of the data and authenticator, the data is checked properly and minimizes the overhead of cloud storage overhead. In this work, the cloud audit and authenticator approach is proposed based on a certain file system, which makes the malicious user get authenticate the data auditing verification as the existing algorithm has low security based on entropy. They propose a data auditing approach that integrates with file management and the authenticator of data deduplication. The proposed approach performs the authenticator process and new form of file tag, which helps guarantee effective security based on the random generation of message key. The proposed approach achieves minimum computational overhead based on the authenticator and data block generation in the performance analysis. Then the security verification is performed on various attacks, such as brute force attack, a man-in-the-middle attack, etc., to check whether the approach is safe or unsafe against the attacks.

Keywords: Deduplication, Authenticator, File Auditing, Entropy, Data Block Generation.

1. Introduction

Cloud computing plays a significant role, as a large amount of data is needed to provide several services while using the applications of real-time one. Therefore, data deduplication has been highlighted by cloud services based on specific providers to minimize the cost of services [1]. A vast survey has been conducted related to securing data deduplication considering the several attacks. Then the models related to data deduplication are increasing drastically. It may lead to several other security and privacy problems when a large amount of data is delivered through the cloud [2].

The wireless network is growing rapidly due to the increase in smartphone usage and becoming the daily element of human life. As humans are physically and mentally dependent on mobile phones, various applications are needed to process different activities such as bank transactions, chat interaction, social networking, etc. To utilize the above mobile applications, there is a need for a vast internet and huge data storage with mobile phones where the cloud plays a significant role. As there is a purpose of installing various applications, malware detection has a major role, and it extracts useful information. It happens day to day [3]. As the statistical survey has predicted, several thousand new threats are based on mobile applications. Therefore, the mobile user faces the significant challenge that new intruder threats are a major problem.

In the solution concern, the data deduplication is secured by applying various cryptographic techniques and security protocols and helps secure the data exchanged between various users. Although several cryptographic techniques are applied, they may help secure the data transferred in the cloud. Thus, it must be more effective in managing the data in various variants such as small data, larger data, and personal data to increase data reliability, scalability of resources, accessing the network, security control, and cost.

Before securing the data, there is an effective process of classifying the duplication of data. It should be done in the stored environment. It should be more effective in processing data deduplication to reduce costs. The data deduplication will consider all types of storage, face several challenges, and provide vast solutions through effective data deduplication. There are certain factors such as granularity, indexing, timing, locality, scope, and techniques to classify the data through data deduplication. It helps to decide whether the data deduplication model is more effective in data classification or not [4]. Therefore, an effective technique related to data deduplication is required to reduce data cost as a large amount of data is prevalent in the process of services related to real-time applications. Several existing data deduplication models are available to classify the data based on the taxonomy of data storage to make effective decisions as a large amount of data are needed to provide services with high effectiveness in the cloud to reduce the service and data cost and able to improve the performance of the model related to cloud computing taxonomy [5]. As the cloud has a flexible infrastructure, which is simplified to a network-centric approach and able to access the data easily, it needs the cloud services, and computing those services is more essential.

Cloud-based data services address the issue of security, and may lead to effective security of data that are transferred and exchanged among users for certain applications. It should also increase the quality of Service (QoS) and provide personal and confidential data with better privacy [6]. The cloud provides several services to their customers based on the effective utilization of data specific to resource sharing. As there must be concern about the effectiveness of the cloud services and cloud users, security issues play a significant role, mainly focusing on data security and providing services in the cloud platform.

The effective intrusion detection and prevention system analyses the effectiveness of data related to data scalability and reliability of data [7]. In securing the data, the authorization process plays an important role in securing the data based on encryption and decryption. It will perform the creation of data files, data storage, and processing of those data as it increases daily. To process those data in the cloud, certain factors must be considered, such as data space, data processing, power, data bandwidth, and cost. In the current scenario, a large amount of data is prevailing. It needs to be processed because the generated data are duplicated in a larger amount.

In this case, the data deduplication technique is needed to compress the data by removing or eliminating the duplicated of multiple instances of data and maintaining the unique content of data. It should help maintain effective bandwidth, resource utilization, and improved data storage. The data privacy factor must be considered in preserving the larger amount of data while performing the data deduplication.

2. Related work

To secure the data through the process of data deduplication, there should be the consideration of certain issues related to security. This will organize the data storage, data transfer, and data backup as there is a vast increase in data usage and it provides low-cost data with effective usage of accessing the resource in the cloud platform. To provide effective usage of data storage, data cloud storage will make use of cloud services effectively based on the data duplication with better instance of data and removing duplicated data. It will manage the elimination of data storage overhead and uploading bandwidth saving. Generally, the client makes their data in cloud to be secure, data integrity, data privacy, and data confidentiality factors. These factors make the data in the cloud more secure and provide the services in an effective manner [8]. To ensure effective data security, standard encryption is applied to perform data encryption when outsourcing the data. In the case of protecting data confidentiality and integrity, standard encryption and proof of ownership are needed [9].

[10] proposed a new deduplication model to mitigate the key server in the distributed environment. [11] proposed a modified DupLESS system using the bow fish optimization algorithm. The proposed system used a key server, which is shared between the group of clients and helps to process the deduplication based on message generated key and is also protected against the external attacks.

[12] proposed the Proof of Storage with Deduplication (POSD), which combines the various functionality Proof of Data Possession (POD), Proof of Retrievability (POR) and Poof of Work (PoW) to secure cloud storage. The data is outsourced into the cloud server has a lot of duplicated data, which needs data deduplication to remove the duplicated files and takes only the unique data content. During the process of data deduplication, the limitation of malicious attacks happened and must be mitigated against the intruders.

[13] consider the concept of cloud storage, which plays a significant role in making users to outsource their data and share confidential data to their legitimate users. Based on the cloud environment, the process of data deduplication is needed to mitigate the data redundancy through the process of encryption and highlights the parameters of minimizing the space storage and communication overhead. As the existing approaches focus on the confidential data, consistency, access control policies against brute-force attack.

Here, the author proposed the effective secure data deduplication with an access control mechanism. On behalf of the owner of the data, the cloud will provide the access on the confidential information, and it will mitigate the duplication of data with enhanced security and data privacy on the cloud. In the performance analysis, the proposed algorithm will perform well related to communication overhead, communication cost, effective deduplication, and overhead storage.

[14] discuss cyber physical systems to rely on mobile users to exchange the data with the concept of cloud. With the concept of cloud, data deduplication techniques are used to store the data storage and bandwidth related to real-time applications and services. Encrypting the key is not possible on the data deduplication as it suffers from the lack of security and high performance and data applicability. Here author has proposed message lock encryption with neVer-decrypt homomorphic encRyption (LEVER) protocol to remodify the encryption and perform data deduplication. The author has analysed the proposed protocol to be more effective in terms of data redundancy and breaching the user privacy. In the future they plan to consider the enhanced LEVER protocol to consider side-channel attacks by creating the relation while the message gets exchanged.

[15] introduced the concept of cloud integrated with big data related to some real-time applications. In recent days, there are several applications will provide surplus number of files and their shared ownership participants. In this regard, data deduplication is a significant process to improve the storage process and cost. The author has highlighted the issue of securing the ownership and shared data against some of the attacks using effective data deduplication. Here the author has proposed the novel Proof of Shared Ownership (PoSW), which enables one to deploy the construction of multi-tier based PoSW to provide enhanced security on the ownership and data deduplication scheme. As the proposed scheme uses convergent encryption, secret sharing, and bloom filter, which deploy the key sharing and provide secure interaction on the shared owner and ownership. The performance analysis is made on the proposed system to minimize the computational cost and evaluate the security performance.

[16] perform data deduplication in the cloud to remove the redundant data on some of the data blocks makes copy as one copy on the cloud. For the process of deduplication, where the encryption process is performed on the data files. Here the author has proposed the data deduplication scheme based on data certificate on proxy re-encryption. The data certificate contains proof of ownership-based signature, which uses cryptography operations and make the key Generation to decrypt the cipher text. In the performance analysis, to verify the proposed approach is against dictionary attack and will increase the security.

[17] propose the secure data deduplication-based authorization scheme used blockchain to ensure data confidentiality on the personal information's and maintain security among the users while the data are stored on the cloud. Based on the blockchain concept, data are protected from unnecessary modification based on smart contract, user data integrity of the user are verified, and a hierarchical hash tree is deployed as a key to upload the data by the user and allow to access the data by legitimate users [18]. Here the author has made the security analysis on the proposed protocol against brute force

attack and collision attack. Then the analysis is done on the proposed protocol to check its effectiveness related to computational overhead. In the future, blockchain technology is used to ensure data reliability and secure key management on the cloud systems.

[19] proposed the scheme, which is user-friendly and no need to interact with the third party in the phase of auditing. In securing the proposed system, they use a cloud-based file system and an authenticator to verify the reliability of the data [20]. It helps to minimize the overhead on cloud storage. The proposed system uses low entropy security with malicious data. To determine the efficiency, storage performance, and verification on the malicious attacks.

S.No.	Title of the paper	Techniques used	Merits	Gap Identified
1.	Lightweight mobile phone app certification [21]	An in-depth security study of the Android set of rules was performed to match the malware characteristics.	To protect against viruses using surety rules. Users will feel more at ease installing software because viruses will be less targeted.	Continue with the security requirements engineering approach to find more malware-defending rules.
2.	Machine learning classifiers for mobile malware detection are being evaluated. [22]	Anomaly-based malware detection using machine learning classifiers allows for secure data-sensitive malware detection.	Effective selection of appropriate network features for malware detection inspection. Using true-positive rate (TPR) numbers, choose the best classifier.	Machine learning classifiers on the cloud are being used to provide real-time mobile virus detection.
3.	Malware detection on Android using latent network behaviour analysis [23]	Extracted network spatial properties of Android apps and independent component analysis: an automatic malware detection technique (ICA)	Polymorphism tolerance Determine the behaviour of domain name resolution based on spatial information. Malicious Android apps are automatically detected.	Malware app dataset for Android. From the Android Market, we have gathered some of the most popular and benign apps. Detection and effectiveness of Android malware
4.	A Gaussian Mixture Model for Dynamic Abnormal Behavior Detection in Smartphone Apps [24]	The Gaussian mixture model is a combination of probabilistic models for the dynamic detection of anomalous behaviour in smartphone applications.	The effectiveness in detecting anomalous application behaviour. To estimate the models of the behavior's application, a Gaussian mixture model was used.	System for decentralised data management
5.	Malicious code detection approaches for cellphones based on power [25]	Effectiveness - two smartphone-specific strategies for detecting malicious code behaviors based on time and location-based power consumption profiles	Malware and rootkits are detected and removed. Battery consumption is significantly reduced. Enhance security by significantly speeding up the scanning process.	The accuracy of power consumption-based detecting approaches In terms of power signature.
6.	An Examination of on-platform versus	In terms of power consumption, machine learning-based	Trade-offs when executing anomaly detection	Allow for the distribution of computational

	externalized operation in power-aware anomaly detection in smartphones [26]	detection methods outperform on-platform solutions.	components in terms of power consumption. Scenarios in which one device interacts with the cloud	workloads in order to extend the total battery life of all devices.

3. Methods

The cloud audit and authenticator approach is proposed based on certain file system, which makes the malicious user to get authenticate the data audit verification as the existing algorithm has low security based on entropy. The proposed data auditing approach integrates with file management and authenticator of data deduplication. The proposed approach performs the authenticator process and new form of file tag, which helps guarantee effective security based on the random generation of message key

Initially the encryption process is performed on a file, which gets identified based on the cloud.

Algorithm 1: Encryption Algorithm

```
Input: Data ‘d’ & File ‘$F_d$’
Output: File Tag and Authenticator
The encryption process gets initiated on the file.
Identify the File ‘$F_d$’ on cloud.
If (File = exists in database)
{
    File Found in the database
Else
    File not found in the database
            Initiate key generation process
            To generate the key tag and authenticator.
            Move the encrypted file into the cloud.
            Move the file tag and the authenticator into the cloud.
}
```

Algorithm 2: User Process in cloud

```
Initially, the file identity is sent to the cloud.
To check the malicious behaviour of the user to access the cloud.
First, the users → the ownership of the file access).
Proof work is checked and sent to the user.
```

```
If (Proof of Work = Valid)
{
    User can execute the file uploading operation
            File gets tag in the cloud
Otherwise
    Users' operation to access the file gets refused
}
```

User choose signing key to get verified and here secure digital signature algorithm is applied. The user is randomly selected based on the secret key and computes the public key. The user makes secret and publish the system information.

```
Algorithm 3: Generation of key and Authenticator
Input: User, Third Party Authenticator
Output: Update the key Generation
Initially,
The user chooses and computes the secret tag key
To compute the authenticator by the user
User chooses and computes the public tag key
To compute the rekey auditing by the user
User to compute the re-key process auditing → Third Party Authenticator
User choose and fix the file tag → Public Information
Third Party Authenticator → File Tag (Proof Verification Algorithm)
File Tag → Public Tag key
Subsequent user → File Tag Recompilation
Compute the authenticator to check the data integrity in cloud
User to upload the data to the cloud
The user processing Algorithm is applied by the user to compute the requested upload and it is sent to the cloud.
If (File = Cloud)
{
   User = User Processing Algorithm
}
```

When the user does not need upload the duplicate file as the cloud checks for that particular file to access. The cloud selects randomly by sending the PoW to cloud.

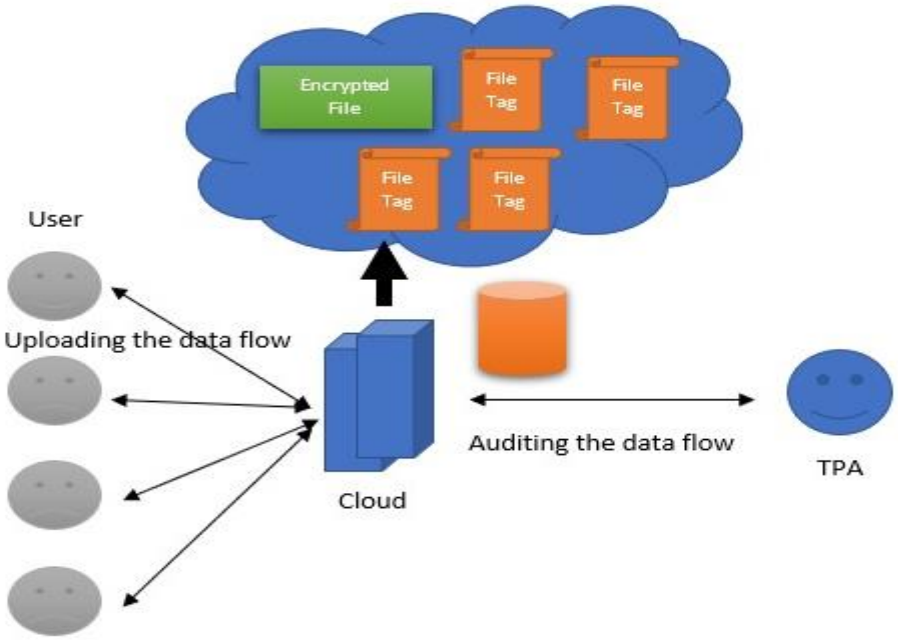

Fig. 1. Proposed Approach

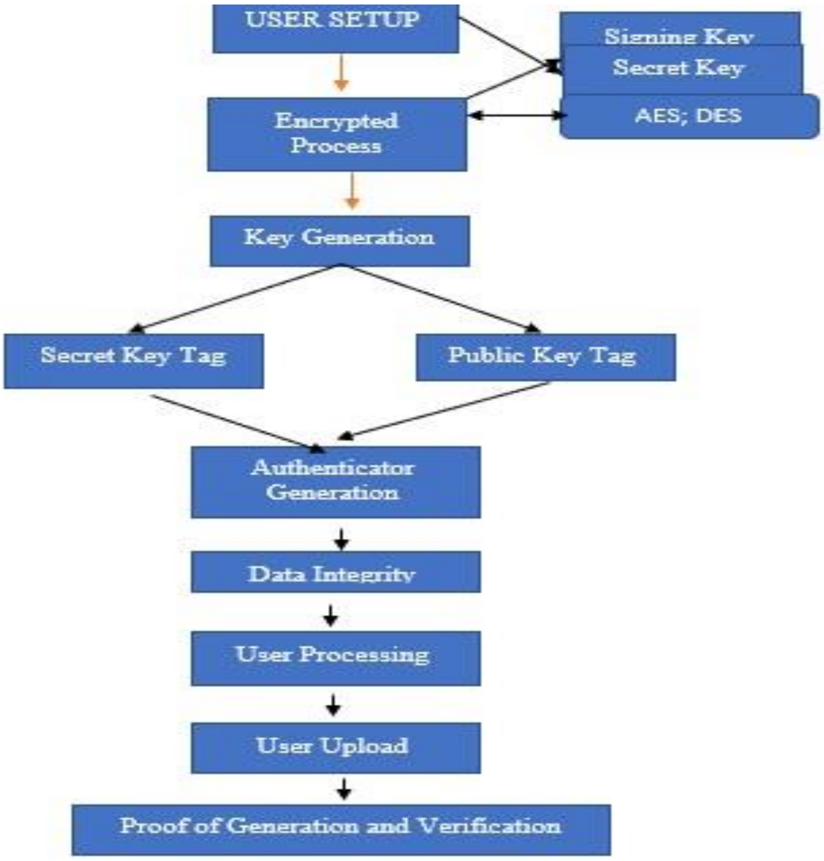

Fig. 2. Flow Chart for the Proposed Work

Table 1. The comparison of various existing techniques'

Security approach	**Suggested approach**	**Strengths**	**Limitations**
Data Storage and security [5]	It employs a homomorphic token with distributed verification of erasure-coded data to guarantee data storage security and to locate the server being attacked..	Data block operations such as update, remove, and add are supported without data loss or corruption with this software. Able to withstand data tampering and server collusion attacks, as well as the torturous failures that go along with them.	In terms of dynamic data storage, security has been thoroughly examined. The fine-grained data error location has not yet been dealt with.
In cloud computing, the user's identity is protected. [9]	The active bundles approach compares predicates across encrypted data and multiparty computing.	When there is no need for a trusted third party (TTP) to verify or approve a user's identification, there is no need for a trusted third party. This also protects the user's identity because his or her identity remains hidden. Even if it is now used for other purposes, such as decryption, the TTP is still free and accessible to the public.	The requested service's host may refuse to execute the active bundle. It could jeopardize the system. Because his requests are not granted authorization, the user's name is hidden.
Interoperability and security trust paradigm for cross-cloud [16]	For each domain, there is a provider and a user, and both have a different trust agent. The two groups, service providers and customers, use different tactics to establish trust. In addition, time and transaction factors are considered in the trust assignment process.	Virtualization is widely used to secure clouds.	It is still in the development stages and will require further testing to ascertain its effectiveness.
Virtualization that is safe [23]	A specialised security system to protect cloud-based virtual machines and distributed computing middleware is available as an option. By logging and reviewing system executable files on a regular basis, you can track cloud component behaviour.	Virtualized networks are at risk of a number of security threats, both to their infrastructure and to their virtual machines. An ACPS system keeps an eye on the guest VM, alerting the system's security system whenever there is questionable activity.	A small performance penalty is accrued. ACPS systems cannot obtain approval due to this.

In a cloud context, a secure virtual network is essential. [21]	When cloud providers focus on limiting the danger of information leaking, they should focus on internal structures, placement policies, and side-channel vulnerabilities.	Helps us in locating a distant location for an attacking party from its target, which reduces the threat level of other VMs.	The attacker will try to learn where the other virtual machines are located in order to launch an attack on them. This could create issues for the VMs that are in the middle.

4. Conclusion

In this research work, the Secure Data Auditing approach is proposed along with file and authenticator deduplication. Here we have applied novel way of generating the authenticator and file tags and avoid duplications through the deduplication process. In this process, malicious users are not allowed to pass the authenticator process to get verified the data integrity. The performance is analysed based on computational overhead with respect to authenticator, proof verification and Generation.

References

[1]. Ravneet Kaur ; Inderveer Chana; Jhilik Bhattacharya; ‘Data deduplication techniques for efficient cloud storage management: a systematic review’ The Journal of Supercomputing, 74, 2035-2085, 2018.

[2]. Youngjoo Shin; Dongyoung Koo; Junbeom Hur; “A Survey of Secure Data Deduplication Schemes for Cloud Storage Systems,” ACM Computing Surveys, 49 (4), 1-38, 2017.

[3]. Jidong Xiao; Zhang Xu; Hai Huang; Haining Wang; “Security implications of memory deduplication in a virtualized environment,” in Proc. Of 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2013.

[4]. Dirk Meister and André Brinkmann, “Multi-level comparison of data deduplication in a backup scenario,” In Proc. Of SYSTOR '09: Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference, 8, 1-12, 2009.

[5]. Junbeom Hur; Dongyoung Koo; Youngjoo Shin; Kyungtae Kang; “Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage,” IEEE Transactions on Knowledge and Data Engineering, 28 (11), 3113-3125, 2016.

[6]. Waraporn Leesakul; Paul Townend; Jie Xu; “Dynamic Data Deduplication in Cloud Storage,” IEEE 8th International Symposium on Service Oriented System Engineering, 2014.

[7]. Qing Liu; Yinjin Fu; Guiqiang Ni; Rui Hou; “Hadoop Based Scalable Cluster Deduplication for Big Data,” IEEE 36th International Conference on Distributed Computing Systems Workshops (ICDCSW), 2016.

[8]. Silambarasan Elkana Ebinazer, Nickolas Savarimuthu & Mary Saira Bhanu S; “An efficient secure data deduplication method using radix trie with bloom filter (SDD-RT-BF) in cloud environment,” Peer-to-Peer Networking and Applications, 14, 2443-2451, 2021.

[9]. Xiaoyu Zheng; Yuyang Zhou; Yalan Ye and Fagen Li; “A cloud data deduplication scheme based on certificateless proxy re-encryption,” Journal of Systems Architecture, 102, 101666, 2020.

[10]. Haonan Su; Dong Zheng; Yinghui Zhang; “An Efficient and Secure Deduplication Scheme Based on Rabin Fingerprinting in Cloud Storage,” IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017.

[11]. Meixia Miao; Jianfeng Wang; Hui Li; Xiaofeng Chen; “Secure multi-server-aided data deduplication in cloud computing,” Pervasive and Mobile Computing, 24, 129-137, 2015.

[12]. Qingji Zheng. Shouhuai Xu, “Secure and efficient proof of storage with deduplication,” in Proc. Of second ACM conference on Data and Application Security and Privacy, 1-12, 2012.

[13]. Zheng Yan; Wenxiu Ding; Xixun Yu; Haiqi Zhu; Robert H. Deng; “Deduplication on Encrypted Big Data in Cloud,” IEEE Transactions on Big Data, 2(2), 138-150, 2016.

[14]. Zahra Pooranian; Mohammad Shojafar; Sahil Garg; Rahim Taheri; Rahim Tafazolli; "LEVER: Secure Deduplicated Cloud Storage With Encrypted Two-Party Interactions in Cyber--Physical Systems," IEEE Transactions on Industrial Informatics, 17(8), 5759-5768, 2020.

[15]. Ismail Hababeh; Ammar Gharaibeh; Samer Nofal; Issa Khalil; "An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility," IEEE Access, 7, 9153-9163, 2018.

[16]. B. Rasina Begum &P. Chitra, "SEEDDUP: A Three-Tier SEcurE Data DedUPlication Architecture-Based Storage and Retrieval for Cross-Domains Over Cloud," IETE Journal of Research, 2021.

[17]. Ruba S, A.M. Kalpana, "An Improved Blockchain-Based Secure Data Deduplication using Attribute-Based Role Key Generation with Ecient Cryptographic Methods," Research Square, 2021.

[18]. Praveen Kumar Premkamal, Syam Kumar Pasupuleti, Abhishek Kumar Singh & P. J. A. Alphonse, "Enhanced attribute-based access control with secure deduplication for big data storage in cloud," Peer-to-Peer Networking and Applications, 14, 102-120, 2021.

[19]. Xiang Gao; Jia Yu; Wen-Ting Shen; Yan Chang; Shi-Bin Zhang; Ming Yang; Bin Wu; "Achieving low-entropy secure cloud data auditing with file and authenticator deduplication," Information Sciences, 546, 177-191, 2021.

[20]. Wenting Shen; Ye Su; Rong Hao; "Lightweight Cloud Storage Auditing With Deduplication Supporting Strong Privacy Protection," IEEE Access, 8, 44359 – 44372, 2020.

[21]. Enck W, Ongtang M, McDaniel P (2009) On lightweight mobile phone application certification. In: Proceedings of the 16th ACM conference on Computer and communications security. ACM, pp 235–245

[22]. Narudin FA, Ali F, Nor BA, Abdullah G (2016) Evaluation of machine learning classifiers for mobile malware detection. Soft Comput 20(1):343–357

[23]. Wei T-E, Mao C-H, Jeng AB, Lee H-M, Wang H-T, Wu D-J (2012) Android malware detection via a latent network behavior analysis. In: 2012 IEEE 11th international conference on trust, security and privacy in computing and communications. IEEE, pp 1251–1258

[24]. El Attar A, Khatoun R, Lemercier M (2014) A Gaussian mixture model for dynamic detection of abnormal behavior in smartphone applications. In: 2014 global information infrastructure and networking symposium (GIIS). IEEE, pp 1–6

[25]. Dixon B, Mishra S (2013) Power based malicious code detection techniques for smartphones. In: 2013 12th IEEE international conference on trust, security and privacy in computing and communications. IEEE, pp 142–149

[26]. Suarez-Tangil G, Tapiador JE, Peris-Lopez P, Pastrana S (2015) Power-aware anomaly detection in smartphones: an analysis of on-platform versus externalized operation. Pervasive Mob Comput 18:137–151.