Journal of Science, Computing and Engineering Research (JSCER)
Volume-6, Issue- 3, March 2023.
DOI: *https://doi.org/10.46379/jscer.2023.060305*
ISSN: 2708-1079

Authorized Data Dedupulication Technique In Cloud Storage System

[1]S.S.Vasantha Raja, [2]R.Palson Kennedy, [3]S. Jonisha, [4]T.Bersikin Libina

[1234]Department of CSE, PERI Institute of Technology, Chennai, India.

Article Information

Received	:	08 Jan 2023
Revised	:	21 Feb 2023
Accepted	:	07Mar 2023
Published	:	17 Mar 2023

Abstract— Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this project makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered induplicate check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, the proposed work implements a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. The proposed work shows that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

Corresponding Author:
S.S.Vasantha Raja

Keywords: ***Data Deduplication, Security, Cloud Computing, Storage.***

Citation: S.S.Vasantha Raja, Dr R.Palson Kennedy, S. Jonisha, T.Bersikin Libina. "Enhancing Communication Skills of the Learners in Professional Advancement, "Journal of Science, Computing and Engineering Research, 6(3), 18-23, 2023.

I. INTRODUCTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, de duplication has been a well-known technique and has attracted more and more attention recently.

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For filelevel de duplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both inside and outsider attacks. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.

Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making deduplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible.

It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is Deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can

download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys.

Thus, convergent encryption allows the cloud to perform deduplication on the cipher texts and the proof of ownership prevents the unauthorized user to access the file. However, previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization.

Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud. For example, in a company, many different privileges will be assigned to employees.

In order to save cost and efficiently management, the data will be moved to the storage server provider (SSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control.

Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent; do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

II. RELATED WORKS

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations

In this paper, they proposed an architecture that provides secure deduplication storage resisting brute force attacks, and realize it in a system called dupLESS[1] .It enables clients encrypted data with an existing service. The encryption for deduplicated storage can achieve performance and space saving close to that of using the storage service with plaintext data.

In this System, If the duplicate of files is available on the cloud server, the application needs to map the original files to the current files that is uploaded so that the user will not get affected by the mapping but on the backend no duplicate files will be created but the files get mapped by the application, In this way the duplication files on the cloud server will be avoided thereby reducing the storage space requirement on the cloud server.

There is a mechanism to reclaim space from incidental duplication to make it available for controlled file replication. This mechanism convergent encryption, which enable duplicate files to be coalesced into the space file, even if the files are encrypted with different users keys.

It is a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud[2] However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master key.we propose several increasingly stringent levels of policy consistency constraints, and present different enforcement approaches to guarantee the trustworthiness of transactions executing on web services.

Consistency problems can arise as transactional database systems are deployed in cloud environments and use policy-based authorization systems to protect sensitive resources.

In this project , they construct a private deduplication protocol based on the standard cryptographic assumptions is then presented and analyzed. They show that the private data deduplication protocol is probably secure assuming that the underlying hash function is collision-resilient, the discrete logarithm is hard and the erasure coding algorithm can erasure up to many fractions of the bits.[3]

Deduplication method is very effective when multiple users storing the same data in outsource. This time relating security and ownership issues. The Proof-of-ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a robust way. Then many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers deduplication because of the randomization property of encryption. In this paper, they design an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. This way, data de duplication can be effective for popular data, whilst semantically secure encryption protects unpopular content.

We show that our scheme is secure under the Symmetric External Decisional Diffie-Hellman Assumption.

This system is designed to be low cost and expandable. The aim of this project is to build an android application which will store only unique contents and de-duplicated files from the server which can be accessed from anywhere in the campus.[4] Achieving Efficient Data Deduplication and Key Aggregation Encryption System in Cloud. The project implemented a scenario where the cloud service can deduplicate the uploaded data. The data users are provided with the proper data. The data was prevented from unwanted exposure and unauthorized access by placing a proper access control mechanism. Authorized data deduplication aims at data security to keep the data secured and avoid unauthorized access. Deduplication at encryption level saves a lot of memory and memory can be utilized efficiently

As a result of this the Fog must support several types of storage, from ephemeral at the lowest tier to semi-permanent at the highest tier. We also note that the higher the tier, the wider the geographical coverage, and the longer the time scale. The ultimate, global coverage is provided by the Cloud, which is used as repository for data that that has a permanence of months and years, and which is the bases for business intelligence analytics.[5] This is the typical HMI environment of reports and dashboards the display key performance indicators. Formulating a Security Layer of Cloud Data Storage Framework Based on Multi Agent System Architecture

In this paper, we investigated the problem of data security in cloud data storage, to ensure the correctness of users' data in cloud data storage; we proposed a security framework and MAS architecture to facilitate security of cloud data storage. This security framework consists of two main layers as agent layer and cloud data storage layer. The propose MAS architecture includes five types of agents: UIA, UA, DERA, DRA and DDPA.

In order to facilitate the huge amount of security, our MAS architecture offered eleven security attributes generated from four mainsecurity policies of correctness, integrity, confidentially and availability of users' data in the cloud[6]. The results show that our algorithm can upload the data from WSNs to Cloud within the limited latency and minimize the energy consumption as well. Cloud computing extends the data processing ability and storage ability of wireless sensor networks (WSNs). However, due to the weak communication ability of WSNs, how to upload the sensed data to the Cloud within the limited time becomes a bottleneck of sensor-cloud system.[7]

III. IMPLEMENTATION

A. Existing System

Data de duplication systems, the private cloud are involved as a proxy to allow data owner/users to securely perform duplicate check withdifferential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.Data de duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage.

The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, de duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy.

De duplication can take place at either the file level or the block level. For file level de duplication, it eliminates duplicate copies of the same file. De duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Identical data copies of different users will lead to different cipher texts, making deduplication impossible. The Disadvantages are traditional encryption, while providing data confidentiality is incompatible with data deduplication and Identical data copies of different users will lead to different cipher texts, making deduplication impossible.

B. Proposed System

In this proposed work, the system enhanced with security. Specifically, it present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model. Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible.

It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text.

To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. The advantages are the user is only allowed to perform the duplicate check for files marked with the corresponding privileges. We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidential.

C. Architecture of Proposed System

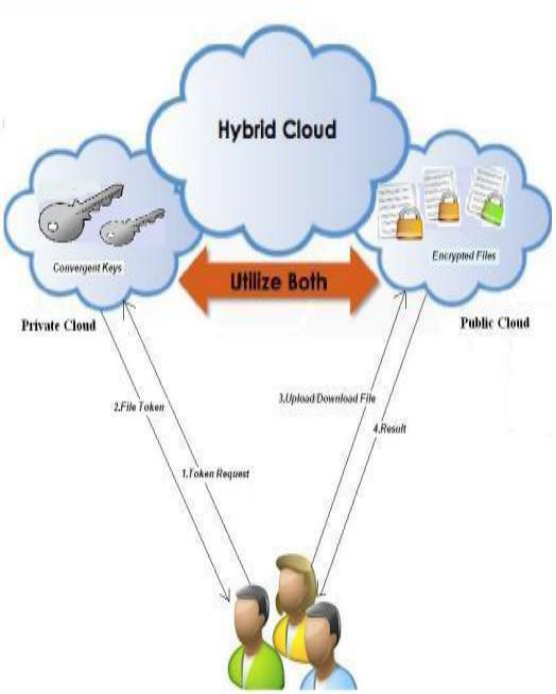

Figure 1. Architecture of Proposed System

The system architecture establishes the basic structure of the system, defining the essential core design features and elements that provide the framework for the system. The systems architecture provides the architects view of the users' vision for what the system needs to be and do, and the paths along which it must be able to evolve and strives to maintain the integrity of that vision as it evolves during detailed design and implementation.

D. Methodology

In this paper, discussed the methodology of deduplication in four different modules as follows.

1) User Module: In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first. At the very least, you need to provide an email address, username, password, display name, and whatever profile fields you have set to required. The display name is what will be used when the system needs to display the proper name of the user.

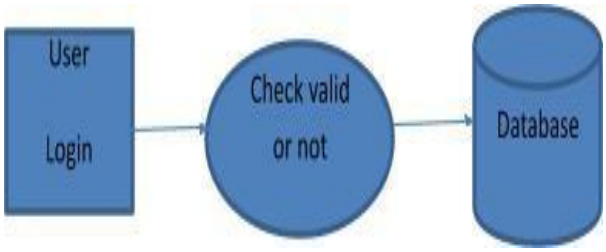

Figure 2 Use Module

2) Server start up and upload file: The user can start up the server after cloud environment is opened. Then the user can upload the file to the cloud.

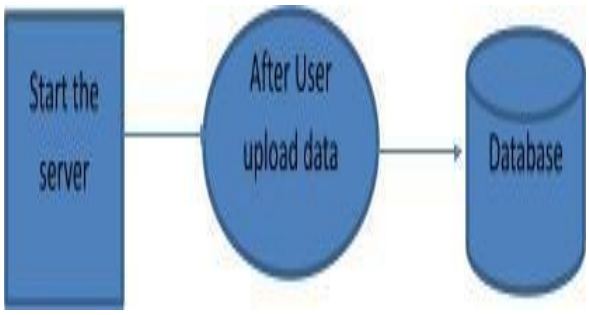

Figure 3. Server start up and upload file

3) Secure Deduplication System:To support authorized de duplication the tag of a file F will be determined by the file F and the privilege. To show the difference with traditional notation of tag, we call is file token instead. To support authorized access a secret key KP will be bounded with a privilege p to generate a file Token.

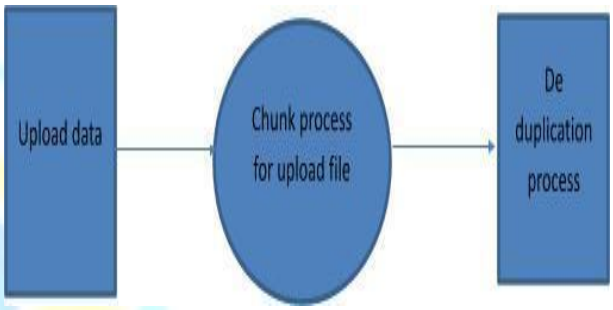

Figure 4. Secure Deduplication System

Deduplication exploits identical content, while encryption attempts to make all content appear random; the same content encrypted with two different keys results in very different cipher text. Thus, combining the space efficiency of deduplication with the secrecy aspects of encryption is problematic.

4) Download file: After the cloud storage, the user can download the file based on key or token.Once the key request was received, the sender can send the key or he can decline it. With this key and request id which was generated at the time of sending key request the receiver can decrypt the message.

Figure 5. Download File

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section discussed the experimental results how the data deduplicated from accessing the data stored in cloud in Fig 6 to 14.

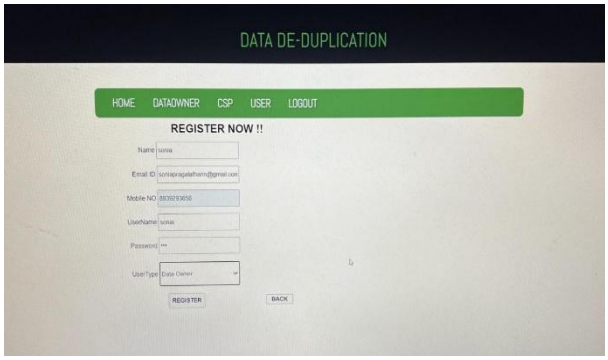

Figure 6. Registration Form

Figure 7. Login Page

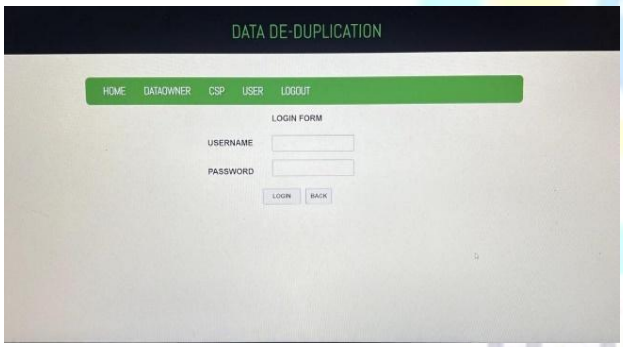

Figure 8 Data Owner Login

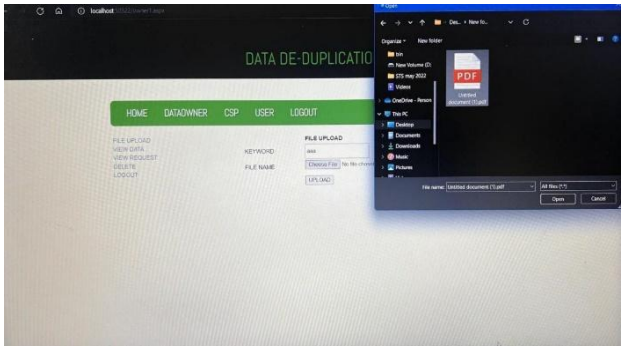

Figure 9. File Uploading

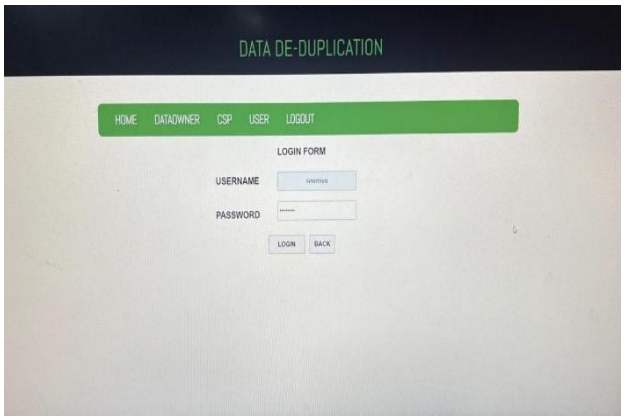

Figure 10. User Login

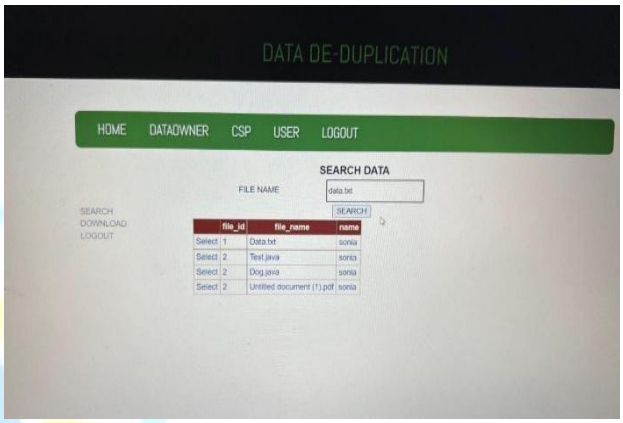

Figure 11. Search aad Results

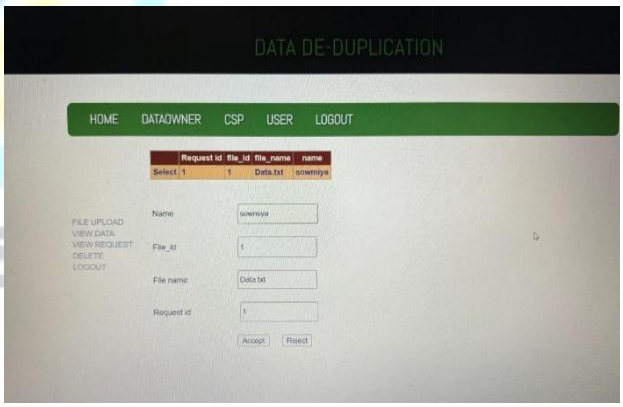

Figure 12. Accepting Request [Data Owner]

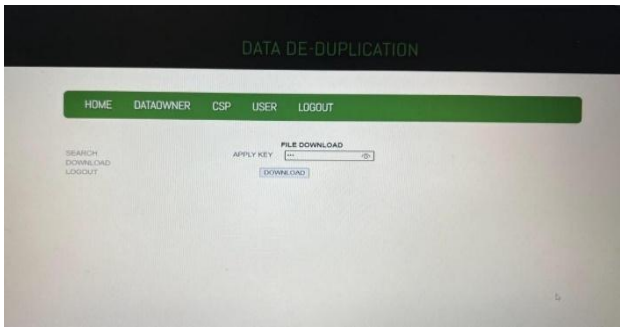

Figure 13. Download File with Key Value

Figure 14. Cloud Service Provider

V. CONCLUSIONS

In this proposal, we are providing high security to the data which is stored in cloud storage and also secures the data from attackers by using encryption technique (blow fish algorithm) by using key sharing process between user and client .De duplication of the data can be done by chunk technique, here the data is be divided into fragments, comparison of data is done with the existing data. Here who owned the data of same copy is no need to store anymore and the data can be accessed by only authorized data of data copy, by this we achieve the data confidentiality, tag consistency, data reliability.The main advantage of the proposed system was it supports both file level and block level. Here the main feature of the proposal is that data integrity, including tag consistency and also the de duplication is possible by client and also server side. The data which is deduplicated is transferred to cloud storage library. It is more efficient, also reduces the space, cost efficient. Same copy of data is tried to transfer then deduplication takes place at the source. Less electricity, Fast recoveries, reduces the overall storage cost.

In this project, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server withprivate keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

VI. FUTURE ENHANCEMENT

Finally, we believe that cloud data storage security is still full of challenges and of paramount importance, and many research problems remain to be identified. In the proposed work deduplication is done for text and image it can be further extended for audio and video.

REFERENCES

[1]. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server Aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[2]. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[3]. Prathima Chilukuri , J.R. Arun Kumar , R. Anusuya , M. Ramkumar Prabhu. “Auto Encoders and Decoders Techniques of Convolutional Neural Network Approach for Image Denoising In Deep Learning” Journal of Pharmaceutical Negative Results, 13(4), 1036–1040. https://doi.org/10.47750/pnr.2022.13.04.142 ,November 4, 2022.

[4]. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[5]. M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.

[6]. R. Anusuya, M. Ramkumar Prabhu, Ch. Prathima, J. R. Arun Kumar” Detection of TCP, UDP and ICMP DDOS attacks in SDN Using Machine Learning approach” Journal of Survey in Fisheries Sciences, Vol. 10 No. 4S (2023): Special Issue 4.

[7]. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server Aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[8]. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[9]. J.R.Arunkumar,” Chaotic African Buffalo Optimization Based Efficient Key Mechanism in Categorized Sensor Networks:, International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020.

[10]. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.