How to Cite:
Seema, J., Nirmala, B. V., Malipatil, S. B., Sheetal, S. R., Shwetha, K. R., & Karuna, B. J. (2022). Encrypting security for virtualized environments in big data storage with generalized anonymization strategies. *International Journal of Health Sciences*, *6*(S9), 2645–2655. https://doi.org/10.53730/ijhs.v6nS9.13005

Encrypting security for virtualized environments in big data storage with generalized anonymization strategies

Seema. J
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: seemaj93@gmail.com

B Vijaya Nirmala
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: bvnjvit@gmail.com

Sushma B Malipatil
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: sushma.malipatil2000@gmail.com

Sheetal. S. R
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: srsheetal@gmail.com

Shwetha K R
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: kr.shwetha12@gmail.com

B. Jaya Karuna
Assistant Professor, Department of CSE, AMCEC, Bengaluru
Email: b.jayakaruna@gmail.com

***Abstract*---**Big data security has grown in importance as a result of its strong correlation to users. Because big data contains individually identifying information, it poses a considerable safety issue. To protect personal information, data confidentiality is commonly employed in non-interactive information sharing and transmission settings. This entails hiding identifiable and confidential material from the owners of acquired data.Anonymization is a computational technology that eliminates or alters personal identifying details, resulting in anonymous data which cannot be linked to every single citizen. The anonymization method enables the release of knowledge, allowing for study and assurance. Privacy protection protects sensitive data from a range of threats. Utilizing data encryption and data anonymization is

Manuscript submitted: 9 May 2022, Manuscript revised: 18 July 2022, Accepted for publication: 27 August 2022

indeed the method of making modifications to the information to be utilised or used in a method which will avoid pertinent data from being identified. Several organisations make data available anonymously to the general public. Employing data anonymization, important bits of private information are masked in order to protect private information.Anonymous data may be stored and accessed in the cloud without fear of being intercepted by others. Anonymization protects against data abuse and insider exploit hazards that lead to regulatory compliance loss. The possibility to derive personal data from findings would be limited if only anonymized data was collected and names were removed from databases. Anonymization of data improves control and reliability of outcomes. It drives digitalization by delivering secure data that can be utilised to create new value. Personal data is encrypted by converting it into a new form or code. The goal of generalisation is to remove a portion of the ids whilst maintaining a level of quality data. To safeguard huge data, many anonymization strategies have been discovered.

Keywords---Anonymization, Generalisation, Big Data, Data Encryption, Privacy Protection, Information Sharing

Introduction

Implementing cloud resources for information storage and processing functions is becoming increasingly prevalent. One of the main problems in a public cloud is data protection. After data is transferred to the cloud, the proprietor no longer has authority over it. We must rely on security precautions offered by 3rd parties to secure information.The anonymization method allows for the dissemination of data that allows for analysis and helps protect data from a number of threats. This cleans up the data. Through the use of encryption technology, it may also maintain the person's anonymity. The field of data analytics has been transformed because of big data. Data that was previously ignored a few decades ago is now regarded as a valuable resource. Almost all facets of society today employ big data significantly for information retrieval. All technological operations generate big data, which is collected and put online. This raises severe security risks. Big data confidentiality is a significant safety risk since it includes personally identifiable data.

Data anonymization is frequently used in non-interactive information dissemination and distribution situations to safeguard personal information. With respect to the proprietors of collected data, it implies concealing identity and sensitive information. A vulnerability to privacy rights exists since the most detailed version of a personal data record is shared. Private information could be properly protected even when some aggregated information is made available to data users for various kinds of analysis and extraction. The major objective of this research is to examine the limitations of extensive data privacy protection. To disclose the most usefulness, sets of data are lead anonymised till k-anonymity is broken. With aid in decision-making, these enormous amounts of information collected from different resources may be handled and evaluated.

Big data privacy is the subject of significant considerations that have necessitated the development of effective privacy protection techniques. Big data security has grown in importance as a result of its strong correlation to users. Today, a company must guarantee confidentiality when using big data analytics. Instead of concentrating on data gathering, security controls must increasingly concentrate on the applications of data. The phrase "big data analytics" refers to the practise of analysing enormous volumes of complicated information with the goal of uncovering underlying trends or locating hidden relationships. Nevertheless, there is a clear inconsistency between both the growing usage of big data and its confidentiality and safety.

The capacity to regulate access and decide whether information may be transmitted is known as confidentiality. Since the information is owned by the data user, it poses a danger to personal privacy if it exists in the public realm. Every data owner is accountable for safeguarding the confidentiality of a user's data. In addition to the information already in the digital realm, users themselves might leak information willingly or inadvertently. Anonymization can satisfy legal responsibilities while reducing security and confidentiality concerns. Anonymization isn't a defence against compromising methods. Existing anonymization approaches may reveal privileged information in databases that have been leaked. Following receiving multiple sources of data, anonymization is used. In databases, anonymization refers to hiding or removing important fields.
Data anonymization transforms plain textual information into an unreadable and irreversible format, such as defensive hashing and cryptography with a broken secret key. Information that has been transferred is at risk since cloud servers could not be totally reliable. Every owner of the information must be absolutely certain that the service is correctly preserving the information in accordance with the service level. The best way to guarantee confidentiality when utilising the cloud is to offer the system a way for the owner of the information to confirm that it remains in existence.Multiple techniques can be used by an anonymization algorithm to obtain the required level of confidentiality. While maintaining the authenticity of the data, the methods generally used in such processes are a wiser alternative. Anonymization is indeed a method that companies may employ to boost data protection within cloud platforms even while enabling analysis and usage of the data. This practise of modifying data that is to be utilised or released such that its identifying of meaningful data is prevented is known as data privacy-preserving.Every form of services required by customers is provided through cloud services. It provides diverse resources to every customer through dynamic provisioning for operations that are assured.

Consumers use one of their main storage providers the most. There are several free cloud-based storage options accessible. In contrast hand, since there are more capabilities, there are more harmful actions. Data which has been altered by malicious behavior cannot be utilized by clients. Cloud security is the use of a variety of rules, procedures, and technology that safeguard data stored in the cloud, its apps, and the infrastructure layer. A component of cloud computing called data safety, or secure storage, aims to prevent users from accessing the data and from changing someone's sensitive data.Data security is essential for both commercial and private users since third parties run the cloud services.

Regardless of whether the data holder is unaware of what occurs towards the information that's also kept and handled on a public cloud, his or her confidential communications might end up in the wrong hands. A privacy-preserving method can employ a variety of methods to achieve the necessary confidentiality level. The approaches often employed in such procedures are a smarter option while retaining the confidentiality of the information. In fact, anonymization is a technique that businesses could employ to improve data safety in the public cloud while also allowing for research and efficient utilisation of information.As a result of consumers delegating their important information to services, data protection and security systems are among the most extensive research topics in cloud technology. Current techniques which alleviate these access and security communication problems using just cryptographic methods suffer from high computational costs on the part of the data owners and the cloud provider regarding public key and administration. Since cloud services are managed by a 3rd party, it's indeed clear that information in a public cloud is not entirely safe against threats both internal and external and invasions. Because confidential data is stored in the public cloud, data protection has become a crucial factor.Cloud computing provides limitless resources for managing, storing, and analysing diverse and large amounts of data.

The demand for cloud-based services is growing quickly, and cloud services must contend with both internal and external risks to data security. Data protection has grown to be a serious problem when information is stored and transported from a server to a remote server.Cloud computing and big data may be combined at the same time. The purpose of cloud computing is to distribute virtualized resources so that resources are used as efficiently as possible to serve applications and services. Big data technologies, including analysis, processing, storing, teleconferencing, and visualisation, may all be aided by cloud technology and virtual machines. The optimum solution for supplying the crucial computing resources for big data applications while making the handling of huge databases simple is the public cloud.The most important factor for users that utilise cloud storage for both personal and professional reasons is the safety of data that is currently stored on the cloud server. To ensure security from assaults like DDoS or man-in-the-middle cyberattacks, several people integrate authenticating, intrusion, and encrypting approaches. For the security of data stored in the cloud, several academics suggested a method that employs multiple encryption techniques rather than single-level encryption. Multi-level security is challenging for an unauthorised user to exploit since he needs both encryption and decryption keys to access the information.

Literature Review

According to [1] information may be divided into three categories: private, public, and confidential. Every piece of information saved on the cloud is done so on the basis of technology. Trusting external cloud service providers with sensitive information is never easy. Indeed, the major participants in the cloud market agree that both their clients and the corporations are responsible for security. In order to prevent unauthorised users from reading information saved in the cloud, it is important from the customer's viewpoint that encryption be strong. Information security is only one of numerous problems that arise when data is

stored in the cloud. The security challenges in this research study are addressed using cryptographic methods.Security concerns are lessened in this study endeavour with the use of cryptographic methods. This suggested solution uses several encryption algorithms, such as the AES method using S-box as well as the Feistel Algorithm, to increase security, as in cloud-based architecture. To protect the massive amounts of data stored in several clouds, the architecture utilises information harvesting, pruning, sorting, encrypting, combining, decrypting, and restoration cycles. The application areas of encryption protocols are safe data transmission across networks and non-human reading storage systems in computer systems. Additionally, networking, servers, and storing at the system level are all combined in cloud technology.

[2] proposed that digital technologies, a wide range of enterprises, including healthcare, banks, e-commerce, commerce, or distribution networks, are producing enormous volumes of data. Data is gathered by both machines and humans. Instances of this include closed-circuit video streaming and website records. Social networking and cell phones create enormous amounts of data each second. With aid in decision-making, these vast amounts of data collected from many resources may be handled and evaluated. Nevertheless, analytics can violate people's privacy. Big data can aid in choice, but it can raise severe concerns about privacy. As a result, big data confidentiality has become critical. [3] introduced a paper on cloud security is the use of a variety of rules, procedures, and technology to safeguard the cloud-based data and its apps as well as the infrastructure layer. A component of cloud computing called data safety or highly secure storage aims to stop users from accessing the data and prevent the deletion or alteration of a user's private details. Most developed and irreversible encryption schemes are powerful ones. Each device is the only irreversible method among current classical or contemporary methods. This research suggests combining the three different models mentioned above in order to examine the methods for cloud storage. Because of their reduced temporal cost and spatial difficulty, the analysis recommends using RSA or OTP with some modifications.

According to [4],the personally identifiable information be kept secret in certain manner or another. In order to keep personal information hidden from prying eyes, privacy protection is essential. Anonymization techniques allow for the dissemination of data that allow for research and effectively protect data from a number of threats. This cleans up the data. Through the use of an encryption technique, it might maintain the user's anonymity. The above paper examines different anonymization methods and algorithms that are currently in use. Article introduces Datafly or Mondrian algorithms, explores their comparability, and concentrates on generalisation and suppressing strategies.

Methodology

Big Data

Big data has emerged as a current model or concept for a variety of data uses. Big data is being prioritised alongside information storage, data processing, and data gathering. Nonetheless, the growing usage of big data as a solution and data

analysis technique does not guarantee privacy and security for data. Aside from the technological advancement through the use of big data, security and confidentiality concerns must be addressed. Big data aids in the preservation of security considerations through the use of security tools such as customer service, remote monitoring, and sensitive data.Limits based on administrative and technological elements could provide safety. Privacy is characterized as an individual's right not to have their private details revealed. Procedures and guidelines can be used to ensure privacy. This paper contains thorough study material on challenges of privacy and security in large data. This allows users to carry out their personal data processing and distant computational resources allocation. Big data cloud technology is a novel method of data handling that dramatically enhances system work performance.

The extraction of large volumes of data is a characteristic of big data; this must rely on multiple processors of cloud services, decentralised databases, cloud services, and virtualized. Cloud technology provides storing, accessibility to locations, and pathways for digital assets, with the data containing the true value of those assets. Cloud technology on big data analysis and forecasting will improve decision accuracy and allow additional information to be released to reveal the hidden potential.Since the large amount of data knowledge, a few data elements have elevated ambiguity and false information elements; some relevant data in the data gathering will decrease the validity of info because of personal ability to operate variables, making the procurement of data and accurate facts challenging to correlate efficiently, and data will advise mostly in the long-term process. This data is much more distorted, which has a significant impact on the accuracy of the data management architecture.

Anonymization

Data anonymization is among the strategies that enterprises may employ to comply with stringent data confidentiality laws that demand the protection of personal information including medical files, contact details, and banking details. Although the data of IDs is cleaned, intruders can employ de-anonymization methods to reconstruct the data anonymization operation. Data de-anonymization procedures will cross-reference sources and provide personal details because data often circulates via numerous sources, a few of which are accessible to the public. Because anonymized datasets are no longer viewed as private details, they could be utilised and disseminated without requiring extra authorization under the law.The technique of maintaining secret or confidential data through removing or encrypting characteristics that connect people to the information stored is known as anonymization. Anonymization rules guarantee that a corporation recognises and upholds its responsibility to protect critical, sensitive, and private data. The possibility to derive personal data from findings would've been limited if anonymized data was gathered and names were removed from databases.

Data encryption

Encryption is accomplished by the use of techniques that change data into unreadable or cipherable codes, rendering it unusable in the modified condition. This approach is commonly used to safeguard data at rest and while in transit

when the data is not required right away. Nevertheless, because it is reversible, encryption allows you to re-identify the data whenever necessary by employing the relevant encryption keys.

Generalization

Generalization is the deliberate exclusion of certain facts in order to make them less recognisable. Data can be transformed into a number of ranges or a huge area having defined limits. This goal is to delete a few of the identities whilst keeping the data accurate. If data is purposefully made less specific by lowering data resolution, obtaining a person's data gets harder. Typically, information is adjusted by employing large ranges rather than specific data elements.

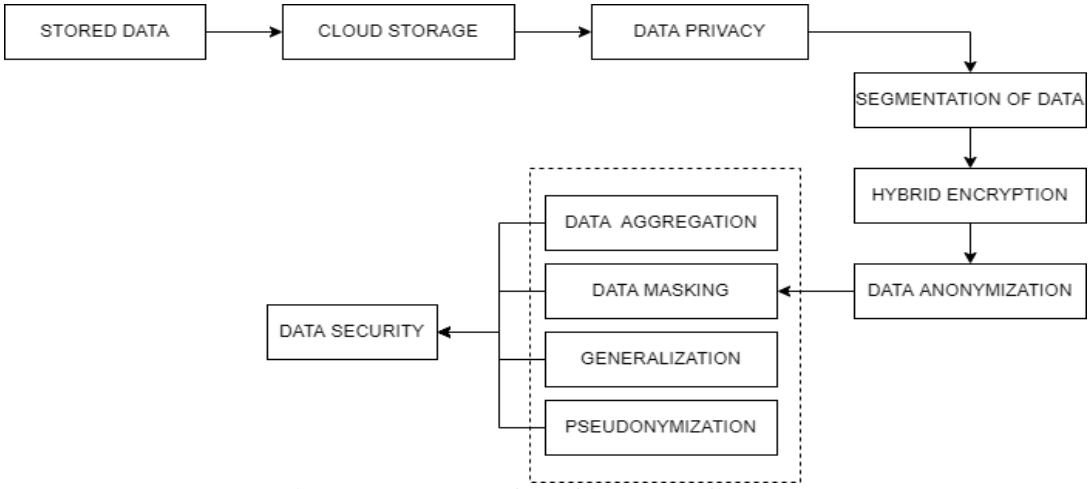

Figure 1: Proposed Architecture diagram

Data masking

Organizations can employ the masking approach to produce a mirror image of their data. This reflected data can then be handled again using encrypting and character shuffle or replacement to maintain the formatting needs of an application, like a retail invoice with the credit card details hidden.The publication of information with changed values is known as information masking. Anonymizing data is accomplished by producing a mirrored version of a database and performing altering procedures including character shuffle, encrypting, phrase or text replacement.

Construction

Data Storage

As emerging innovations are widely utilised in everyday life, such as internet services and smartphones, massive amounts of data cause the traditional data handling and storage system to breakdown. As a result, the information storage and management industries have taken a new approach to scalable data storage. This new addition must be required to address the developing needs related to

increasing constant receiving and data of all types, including pictures, textual, and audio. These shifting demands primarily involve increased data volume, increased capacity, more data sources, and app variety. When datasets and flow increase, the traditional database system encounters new issues, including becoming more effective in parallel computing, connectivity contracts, strategic planning, and high availability.The primary challenge presented by big data is the massive amount of data; a single entity cannot have strong capability with this enormous amount of data. Moreover, the volume of data is increasing significantly, necessitating the state's ability to manage both current and dynamically created data. Such characteristics encourage others to create new indexes and regular expressions in parallel computing in big data.

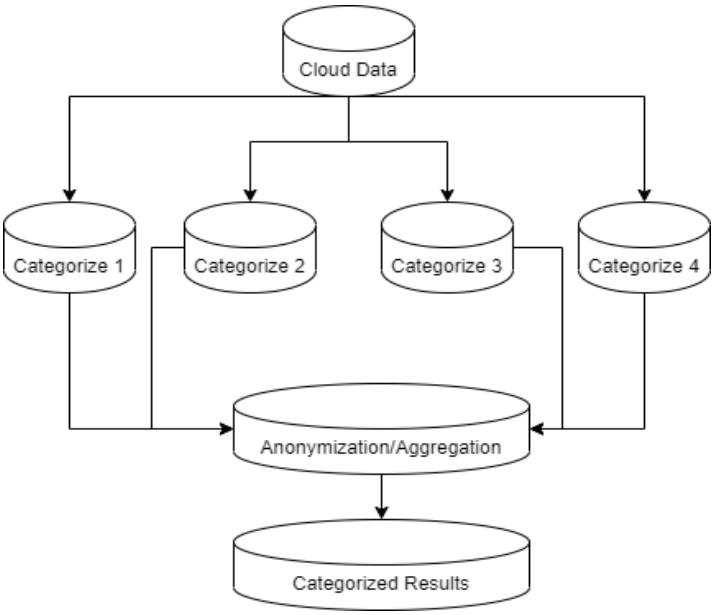

Figure 2: Categorized Features of Cloud Data

Data Privacy

Big data security entails the methods and techniques being used to protect both information and analytical operations. Its primary goal is to secure important data and prevent cyber-attacks, theft, and other destructive acts. With cloud-based businesses, big data security concerns are multifaceted. The difficult danger involves internet data breaches, malware, and DDoS assaults that can bring a system down. Such risks can also have substantial financial impact, including such loss, legal costs, and penalties or punishments levied it against organisation.Cloud computing provides all organisations with more effective communication and enhanced profit. Every transmission must be encrypted. Digital applications management, which delivers and hosts sensitive information and also analyses cloud-hosted architecture, is provided by security. Technologies do provide assistance for a variety of public cloud services. Because a breach of security could occur anywhere at the level of information processing, the potential threat and viable remedy have indeed been proposed at each stage, beginning

with information gathering and progressing through storage devices, analytical techniques, and data management.

Experimental Results

Data security refers to the process of preventing unwanted accessibility, modification, or loss of electronic content across its entire duration. Organizations could enable teams to build apps or teach individuals with real statistics while masking data.The proposed research contributes to preventing information loss in the cloud and fostering empathy between both the service provider and the data consumers. A hybrid data encryption is one that integrates multiple or even more data encryptions. This combines asymmetrical and symmetric encryption to take advantage of the advantages of both types of encryptions.

A technique of encrypting that incorporates two or even more cryptographic techniques and incorporates both symmetric and asymmetric encrypting to take advantage of the advantages of each. It conceals personally identifiable data as appropriate so that research can take place in compliant contexts.The platform has several processing components. The technical feasibility and overall throughput vary substantially amongst modules. To ensure effective collection of data, the system needs to be able to regulate the flow, add and delete nodes automatically. Furthermore, whenever the speed at which data streams flow into based on the system's overall rate at which data streams movement out of the systems, the system needs to cache content.

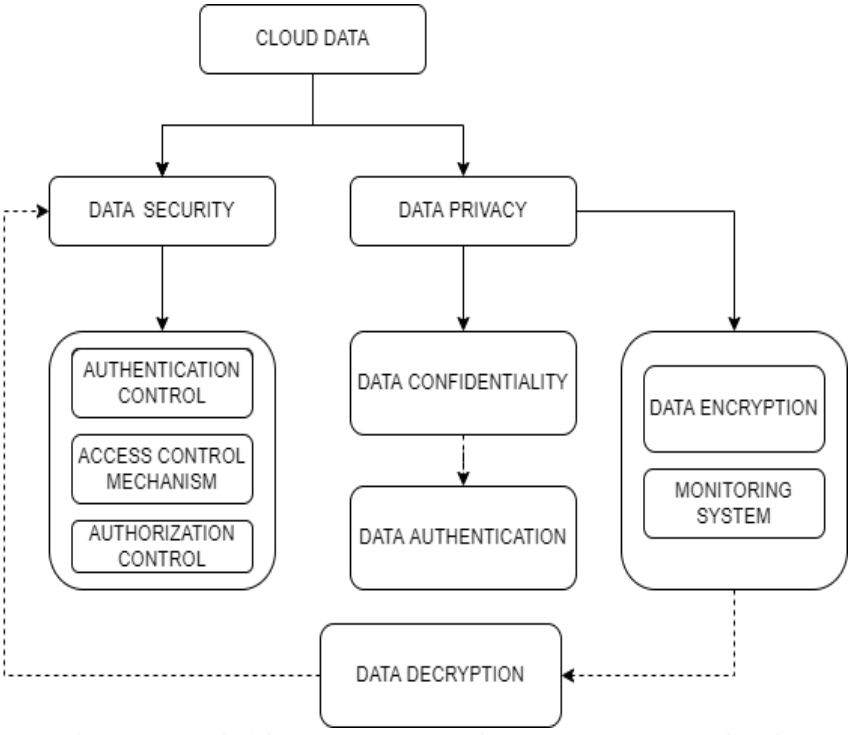

Figure 3: Hybrid Encryption Technique to secure Cloud

Table 1: Uses of Different Techniques

Approaches	Contributions	Disadvantages
Hybrid Encryption Technique	• Data transmission employing a unique shared key and symmetrical encryption results in hybrid security. • A symmetric key pair is acquired and utilised to encrypt the raw information in order to secure communication.	Slow Down in Process
Generalized Anonymization	• Removing identifying details so that the underlying data cannot be used to identify a specific individual • Data anonymization can allow for merge flow of data.	Less capacity to store anonymized data

Conclusion

Anonymization safeguards for information to obtain and insider exploitation risks which result in regulatory failure. Perhaps if anonymized information was gathered and names were deleted from systems, the ability to extract personal information on conclusions would've been reduced. Data anonymization increases result management and dependability. It promotes digitization by providing secure information that can be utilized to generate values. Sensitive data is secured by transforming it into unique conception or code. Organizations must receive consent from users before collecting sensitive data, including passwords, email addresses, and device identifiers. The capacity to extract relevant information from findings would be limited if data were gathered anonymously and names were removed from databases. Anonymization protects against data abuse and insider manipulation hazards that lead to compliance failures. The goal is to delete a few of the identities whilst keeping the data accurate. The purpose of generalization is to delete some of the ids while keeping a high degree of data integrity. Several anonymization solutions have been created to protect large amounts of data.

References

1. A. Kumar, B. G. Lee, H. Lee and A. Kumari, "Secure storage and access of data in cloud computing," 2012 International Conference on ICT Convergence (ICTC), 2012, pp. 336-339, doi: 10.1109/ICTC.2012.6386854.
   Approach in Detecting and Isolation of Malicious Nodes in MANET" Wireless Personal Communication, 119, pages21–35 (2021) Springer Jan 2021
2. Cong Wang, Qian Wang, Kui Ren and Wenjing Lou, "Ensuring data storage security in Cloud Computing," 2009 17th International Workshop on Quality of Service, 2009, pp. 1-9, doi: 10.1109/IWQoS.2009.5201385.
3. Inukollu, Venkata &Arsi, Sailaja &Ravuri, Srinivasa. (2014). Security Issues Associated with Big Data in Cloud Computing. International Journal of Network Security & Its Applications. 6. 45-56. 10.5121/ijnsa.2014.6304.
4. Jain, P., Gyanchandani, M. &Khare, N. Big data privacy: a technological perspective and review. *J Big Data* **3**, 25 (2016). https://doi.org/10.1186/s40537-016-0059-y

5. Karle, Tanashri& Vora, Deepali. (2017). PRIVACY preservation in big data using anonymization techniques. 340-343. 10.1109/ICDMAI.2017.8073538.
6. N. L. Kodumru and M. Supriya, "Secure Data Storage in Cloud Using Cryptographic Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697550.
7. Preethi, P., & Asokan, R. (2019). An attempt to design improved and fool proof safe distribution of personal healthcare records for cloud computing. Mobile Networks and Applications, 24(6), 1755-1762.
8. Preethi, P., & Asokan, R. (2021). Modelling LSUTE: PKE Schemes for Safeguarding Electronic Healthcare Records Over Cloud Communication Environment. Wireless Personal Communications, 117(4), 2695-2711.
9. Pronika and S. S. Tyagi, "Secure Data Storage in Cloud using Encryption Algorithm," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 136-141, doi: 10.1109/ICICV50876.2021.9388388.
10. R. R. Bobde, A. Khaparde and M. M. Raghuwanshi, "An approach for securing data on Cloud using data slicing and cryptography," 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 2015, pp. 1-5, doi: 10.1109/ISCO.2015.7282356.
11. R.Thiagarajan, V.BalajiVijayan, Dr.S.Arun, I.Mohan Novel Technique for Automation Billing in Smart Shopping", International Journal of Scientific & Technology Research, Vol. 9 no.4, March 2020, PP:5363-5369
12. R.Thiagarajan, Ganesan, Anbarasu, Baskar, Arthi, Rajkumar, Optimised with Secure
13. R.Thiagarajan,N.R .Rajalakshmi ,M. Baskar ,P.Jayalakshmi "A Novel Solution for EconomizingWater by a Mix of Technologies with a Low Cost Approach",International Journal of Advanced Science and Technology Vol. 29, No. 7, April 2020
14. Ram Mohan Rao, P., Murali Krishna, S. & Siva Kumar, A.P. Privacy preservation techniques in big data analytics: a survey. *J Big Data* **5**, 33 (2018). https://doi.org/10.1186/s40537-018-0141-8
15. Sedayao, Jeff. (2012). Enhancing Cloud Security Using Data Anonymization.
16. T. Devi and R. G. San, "Data security frameworks in cloud," 2014 International Conference on Science Engineering and Management Research (ICSEMR), 2014, pp. 1-6, doi: 10.1109/ICSEMR.2014.7043610.
17. Thiagarajan.R, Moorthi. M , Energy consumption and network connectivity based on Novel-LEACH-POS protocol networks,Computer Communications, Elsevier, (0140-3664), vol.149, pp. 90-98.