iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52497

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud

Jayesh Phale[1], Asst. Prof. Soni R. Ragho[2], Rushikesh Kale[3], Priya Yelakar[4], Priyanka Sul[5]
[1, 2, 3, 4, 5]*Vidya Prasarini Sabha's College Of Engineering & Technology, Lonavala Computer Engineering*

Abstract: ***Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. First, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Second, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext.***
Keywords: Data Storage, Attribute Based Encryption (ABE), Data Deduplication.

I. INTRODUCTION

Cloud computing greatly facilitates data providers who want to outsource their data to the cloud without disclosing their sensitive data to external parties and would like users with certain credentials to be able to access the data [1], [2], [3], [4], [5]. This requires data to be stored in encrypted forms with access control policies such that no one except users with attributes (or credentials) of specific forms can decrypt the encrypted data. An encryption technique that meets this requirement is called attribute-based encryption (ABE) [6], where a user's private key is associated with an attribute set, a message is encrypted under an access policy (or access structure) over a set of attributes, and a user can decrypt a ciphertext with his/her private key if his/her set of attributes satisfies the access policy associated with this ciphertext. However, the standard ABE system fails to achieve secure deduplication [7], which is a technique to save storage space and network bandwidth by eliminating redundant copies of the encrypted data stored in the cloud. On the other hand, to the best of our knowledge, existing constructions [8], [9], [10], [11] for secure deduplication are not built on attribute-based encryption. Nevertheless, since ABE and secure deduplication have been widely applied in cloud computing, it would be desirable to design a cloud storage system possessing both properties.

We consider the following scenario in the design of an attribute-based storage system supporting secure deduplication of encrypted data in the cloud, in which the cloud will not store a file more than once even though it may receive multiple copies of the same file encrypted under • Hui Cui is with the Secure Mobile Centre, School of Information Systems, Singapore Management University. E-mail: hcui@smu.edu.sg • Robert H. Deng, Yingjiu Li and Guowei Wu are with the School of Information Systems, Singapore Management University. Manuscript received Month Day, 2016; revised Month Day, 2016. different access policies. A data provider, Bob, intends to upload a file M to the cloud, and share M with users having certain credentials. In order to do so, Bob encrypts M under an access policy A over a set of attributes, and uploads the corresponding ciphertext to the cloud, such that only users whose sets of attributes satisfying the access policy can decrypt the ciphertext. Later, another data provider, Alice, uploads a ciphertext for the same underlying file M but ascribed to a different access policy A 0 . Since the file is uploaded in an encrypted form, the cloud is not able to discern that the plaintext corresponding to Alice's ciphertext is the same as that corresponding to Bob's, and will store M twice. Obviously, such duplicated storage wastes storage space and communication bandwidth.

II. OUR CONTRIBUTIONS

In this paper, we present an attribute-based storage system which employs ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication. Our main contributions can be summarized as follows.

1) Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture [12].

©IJRASET: All Rights are Reserved

2) Secondly, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under any other access policies without revealing the underlying plaintext. This technique might be of independent interest in addition to the application in the proposed storage system.

3) Thirdly, we propose an approach based on two cryptographic primitives, including a zero-knowledge proof of knowledge [13] and a commitment scheme.

III. RELATED WORK

The template Attribute-Based Encryption. Sahai and Waters [6] introduced the notion of attribute-based encryption (ABE), and then Goyal et al. [16] formulated key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE) as two complimentary forms of ABE. The first KP-ABE construction given in [16] realized the monotonic access structures, the first KP-ABE system supporting the expression of non-monotone formulas was presented in [17] to enable more viable access policies, and the first large class KP-ABE system was International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCRACES - 2019 Conference Proceedings Volume 7, Issue 10 Special Issue - 2019 1 presented by in the standard model in [18]. Nevertheless, we believe that KP-ABE is less flexible than CP-ABE because the access policy is determined once the user's attribute private key is issued. Bethencourt, Sahai and Waters [19] proposed the first CP-ABE construction, but it is secure under the generic group model. Cheung and Newport [20] presented a CPABE scheme that is proved to be secure under the standard model, but it only supports the AND access structures. A CP-ABE system under more advanced access structures is proposed by Goyal et al. [21] based on the number theoretic assumption. In order to overcome the limitation that the size of the attribute space is polynomially bounded in the security parameter and the attributes are fixed ahead, Rouselakis andWaters [22] built a large universe CP-ABE system under the prime-order group. In this paper, the Rouselakis-Waters system is taken as the underlying scheme for the concrete construction.

IV. SYSTEM ARCHITECTURE

The architecture of our attribute-based storage system with secure deduplication is shown in Fig. 2 in which four entities are involved: data providers, attribute authority (AA), cloud and users. A data provider wants to outsource his/her data to the cloud and share it with users possessing certain credentials. The AA issues every user a decryption key associated with his/her set of attributes. The cloud consists of a public cloud which is in charge of data storage and a private cloud which performs certain computation such as tag checking. When sending a file storage request, each data provider firstly creates a tag T and a label L associated with the data, and then encrypts the data under an access structure over a set of attributes. Also, each data provider generates a proof pf on the relationship of the tag T, the label L and the encrypted message ct3 , but this proof will not be stored anywhere in the cloud and is only used during the checking phase for any newly generated storage request. After receiving a storage request, the private cloud first checks the validity of the proof pf, and then tests the equality of the new tag T with existing tags in the system. If there is no match for this new tag T, the private cloud adds the tag T and the label L to a tag-label list, and forwards the label and the encrypted data, (L, ct) to the public cloud for storage. Otherwise, let ct0 be the ciphertext whose tag matches the new tag and L 0 be the label associated with ct0 , and then the private cloud executes as follows.

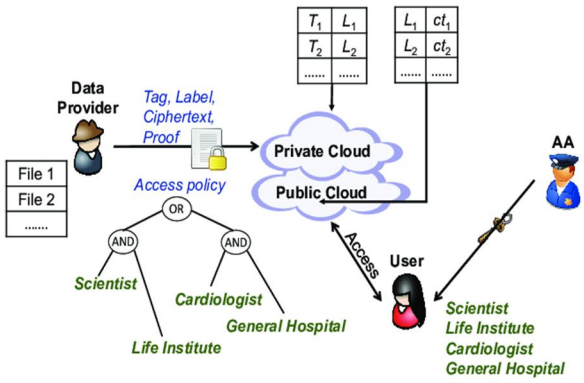


Fig. : System architecture of attribute-based storage with secure deduplication.

1) If the access policy in ct is a subset of that in ct' , the private cloud simply discards the new storage request; else, if the access policy in ct' is a subset of that in ct, the private cloud asks the public cloud to replace the stored pair (L' , ct') with the new pair (L, ct) where L = L' .
2) If the access policies in ct and ct' are not mutually contained, the private cloud runs the ciphertext regeneration algorithm to yield a new ciphertext for the same underlying plaintext file and associated with an access structure which is the union of the two access structures, and forwards the original label and the resulting ciphertext to the public cloud.

At the user side, each user can download an item, and decrypt the ciphertext with the attribute-based private key generated by the AA if this user's attribute set satisfies the access structure. Each user checks the correctness of the decrypted message using the label, and accepts the message if it is consistent with the label.

Concerning the adversarial model of our storage system, we assume that the private cloud is "curious-but-honest" such that it will attempt to obtain the encrypted messages but it will honestly follow the protocols, whereas the public cloud is distrusted such that it might tamper with the label and ciphertext pairs outsourced from the private cloud (note that such a misbehavior will be detected by either the private cloud or the user via the accompanied label). Another difference between the private cloud and the public cloud is that the former can not collude with users4 , but the latter could collude with users. This assumption is in line with the real world practice where the private cloud is trusted more than the public cloud. We assume that data users may try to access data beyond their authorized privileges. In addition to trying to obtain plaintext data from the cloud, malicious outsiders may also commit duplicate faking attacks as described before.

V. METHODOLOGY

Secure Data Deduplication: Secure Deduplication.With the goal of saving storage space for cloud storage services, Douceur et al. [23] proposed the first solution for balancing confidentiality and efficiency in performing deduplication called convergent encryption, where a message is encrypted under a message-derived key so that identical plaintexts are encrypted to the same ciphertexts. In this case, if two users upload the same file, the cloud server can discern the equal ciphertexts and store only one copy of them. Implementations and variants of convergent encryption were deployed in [24], [25], [26], [27], [28]. In order to formalize the precise security definition for convergent encryption, Bellare, Keelveedhi and Ristenpart [8] introduced a cryptographic primitive named message locked encryption, and detailed several definitions to capture various security requirements. Abadi et al. [9] then strengthened the security definition in [8] by considering the plaintext distributions depending on the public parameters of the schemes. This model was later extended by Bellare and Keelveedhi [11] by providing privacy for messages that are both correlated and dependent on the public system parameters. Since message-locked encryption cannot resist to brute-force attacks where files falling into a known set will be recovered, an architecture that provides secure deduplicated storage resisting brute-force attacks was put forward by Keelveedhi, Bellare and Ristenpart [10] and realized in a system called server-aided encryption for deduplicated storage. In this paper, a similar technique to that [9] is used to achieve secure deduplication with regard to the private cloud in the concrete construction.

VI. CONCLUSION

Attribute-based encryption (ABE) has been widely used in cloud computing where data providers outsource their encrypted data to the cloud and can share the data with users possessing specified credentials. On the other hand, deduplication is an important technique to save the storage space and network bandwidth, which eliminates duplicate copies of identical data. However, the standard ABE systems do not support secure deduplication, which makes them costly to be applied in some commercial storage services. In this paper, we presented a novel approach to realize an attribute-based storage system supporting secure deduplication. Our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage. The private cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it can transfer the ciphertext over one access policy into ciphertexts of the same plaintext under any other access policies without being aware of the underlying plaintext. After receiving a storage request, the private cloud first checks the validity of the uploaded item through the attached proof. If the proof is valid, the private cloud runs a tag matching algorithm to see whether the same data underlying the ciphertext has been stored. If so, whenever it is necessary, it regenerates the ciphertext into a ciphertext of the same plaintext over an access policy which is the union set of both access policies. The proposed storage system enjoys two major advantages. Firstly, it can be used to confidentially share data with other users by specifying an access policy rather than sharing the decryption key. Secondly, it achieves the standard notion of semantic security while existing deduplication schemes only achieve it under a weaker security notion.

REFERENCES

[1] D. Quick, B. Martini, and K. R. Choo, Cloud Storage Forensics. Syngress Publishing / Elsevier, 2014. [Online]. Available: http://www.elsevier.com/books/cloud-storageforensics/quick/978-0-12-419970-5
[2] K. R. Choo, J. Domingo-Ferrer, and L. Zhang, "Cloud cryptography: Theory, practice and future research directions," Future Generation Comp. Syst., vol. 62, pp. 51–53, 2016.
[3] K. R. Choo, M. Herman, M. Iorga, and B. Martini, "Cloud forensics: State-of-the-art and future directions," Digital Investigation, vol. 18, pp. 77–78, 2016.
[4] Y. Yang, H. Zhu, H. Lu, J. Weng, Y. Zhang, and K. R. Choo, "Cloud based data sharing with fine-grained proxy re-encryption," Pervasive and Mobile Computing, vol. 28, pp. 122–134, 2016.
[5] D. Quick and K. R. Choo, "Google drive: Forensic analysis of data remnants," J. Network and Computer Applications, vol. 40, pp. 179– 193, 2014.
[6] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, ser. Lecture Notes in Computer Science, vol. 3494. Springer, 2005, pp. 457–473.
[7] B. Zhu, K. Li, and R. H. Patterson, "Avoiding the disk bottleneck in the data domain deduplication file system," in 6th USENIX Conference on File and Storage Technologies, FAST 2008, February 26- 29, 2008, San Jose, CA, USA. USENIX, 2008, pp. 269–282.
[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings, ser. Lecture Notes in Computer Science, vol. 7881. Springer, 2013, pp. 296–312.
[9] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lock-dependent messages," in Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, ser. Lecture Notes in Computer Science, vol. 8042. Springer, 2013, pp. 374–391.
[10] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013. USENIX Association, 2013, pp. 179–194.
[11] M. Bellare and S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," in Public-Key Cryptography - PKC 2015 - 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30 - April 1, 2015, Proceedings, ser. Lecture Notes in Computer Science, vol. 9020. Springer, 2015, pp. 516–538.
[12] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin ¨ clouds: Secure cloud computing with low latency - (full version)," in Communications and Multimedia Security, 12th IFIP TC 6 / TC 11 International Conference, CMS 2011, Ghent, Belgium, October 19- 21,2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 7025. Springer, 2011, pp. 32–44.
[13] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof-systems (extended abstract)," in Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA. ACM, 1985, pp. 291– 304.
[14] M. Fischlin and R. Fischlin, "Efficient non-malleable commitment schemes," in Advances in Cryptology - CRYPTO 2000, 20th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2000, Proceedings, ser. Lecture Notes in Computer Science, vol. 1880. Springer, 2000, pp. 413–431.
[15] S. Goldwasser and S. Micali, "Probabilistic encryption," J. Comput. Syst. Sci., vol. 28, no. 2, pp. 270–299, 1984.
[16] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006, ser. Lecture Notes in Computer Science, vol. 5126. Springer, 2006, pp. 89–98.
[17] R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-based encryption with non-monotonic access structures," in Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007. ACM, 2007, pp. 195–203.
[18] A. B. Lewko and B. Waters, "Unbounded HIBE and attributebased encryption," in Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 6632. Springer, 2011, pp. 547–567.
[19] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in 2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA. IEEE Computer Society, 2007, pp. 321–334.
[20] L. Cheung and C. C. Newport, "Provably secure ciphertext policy ABE," in Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007. ACM, 2007, pp. 456–465.
[21] V. Goyal, A. Jain, O. Pandey, and A. Sahai, "Bounded ciphertext policy attribute based encryption," in Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, 2332-7790 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2017.2656120, IEEE Transactions on Big Data JOURNAL OF LATEX CLASS FILES, VOL. , NO. , MONTH 2016 13 Iceland, July 7-11, 2008, Proceedings, Part II - Track B: Logic, Semantics, and Theory of Programming & Track C: Security and Cryptography Foundations, ser. Lecture Notes in Computer Science, vol. 5126. Springer, 2008, pp. 579–591.
[22] Y. Rouselakis and B. Waters, "Practical constructions and new proof methods for large universe attribute-based encryption," in 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013. ACM, 2013, pp. 463–474.
[23] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in ICDCS, 2002, pp. 617–624.
[24] M. W. Storer, K. M. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proceedings of the 2008 ACM Workshop On Storage Security And Survivability, StorageSS 2008, Alexandria, VA, USA, October 31, 2008. ACM, 2008, pp. 1–10.
[25] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Uncovering the Secrets of System Administration: Proceedings of the 24th Large Installation System Administration Conference, LISA 2010, San Jose, CA, USA, November 7-12, 2010. USENIX Association, 2010.

[26] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, "A secure cloud backup system with assured deletion and version control," in 2011 International Conference on Parallel Processing Workshops, ICPPW 2011, Taipei, Taiwan, Sept. 13-16, 2011. IEEE Computer Society, 2011, pp. 160–167.
[27] P. Puzio, R. Molva, M. Onen, and S. Loureiro, "Cloudedup: Secure ¨ deduplication with encrypted data for cloud storage," in IEEE 5th International Conference on Cloud Computing Technology and Science, CloudCom 2013, Bristol, United Kingdom, December 2-5, 2013, Volume 1. IEEE Computer Society, 2013, pp. 363–370.
[28] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers, ser. Lecture Notes in Computer Science, vol. 8437. Springer, 2014, pp. 99–118.
[29] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in CRYPTO, ser. Lecture Notes in Computer Science, vol. 2139. Springer-Verlag, 2001, pp. 213–219.
[30] E. Fujisaki and T. Okamoto, "Secure integration of asymmetric and symmetric encryption schemes," J. Cryptology, vol. 26, no. 1, pp. 80–101, 2013.
[31] A. B. Lewko and B. Waters, "Decentralizing attribute-based encryption," in Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 6632. Springer, 2011, pp. 568–588.
[32] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," in Public Key Cryptography - PKC 2011 - 14th International Conference on Practice and Theory in Public Key Cryptography, Taormina, Italy, March 6-9, 2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 6571. Springer, 2011, pp. 53–70.
[33] A. Beimel, "Secure schemes for secret sharing and key distribution," Ph.D. dissertation, Israel Institute of Technology, Israel Institute of Technology, June 1996.
[34] J. Lai, R. H. Deng, Y. Yang, and J. Weng, "Adaptable ciphertextpolicy attribute-based encryption," in Pairing-Based Cryptography - Pairing 2013 - 6th International Conference, Beijing, China, November 22-24, 2013, Revised Selected Papers, ser. Lecture Notes in Computer Science, vol. 8365. Springer, 2013, pp. 199–214.
[35] J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin, "Charm: a framework for rapidly prototyping cryptosystems," J. Cryptographic Engineering, vol. 3, no. 2, pp. 111–128, 2013.



Assi. Prof. Soni R. Ragho
M. E. (In computer Engineering)
Email : sonirragho@gmail.com



Priya M.
Yelakar (Bachelor of Computer Engineering) in Vidya Prasarini Sabha's College of Engineering and Technology Lonavala.
Email : priyayelakar123@gmail.com



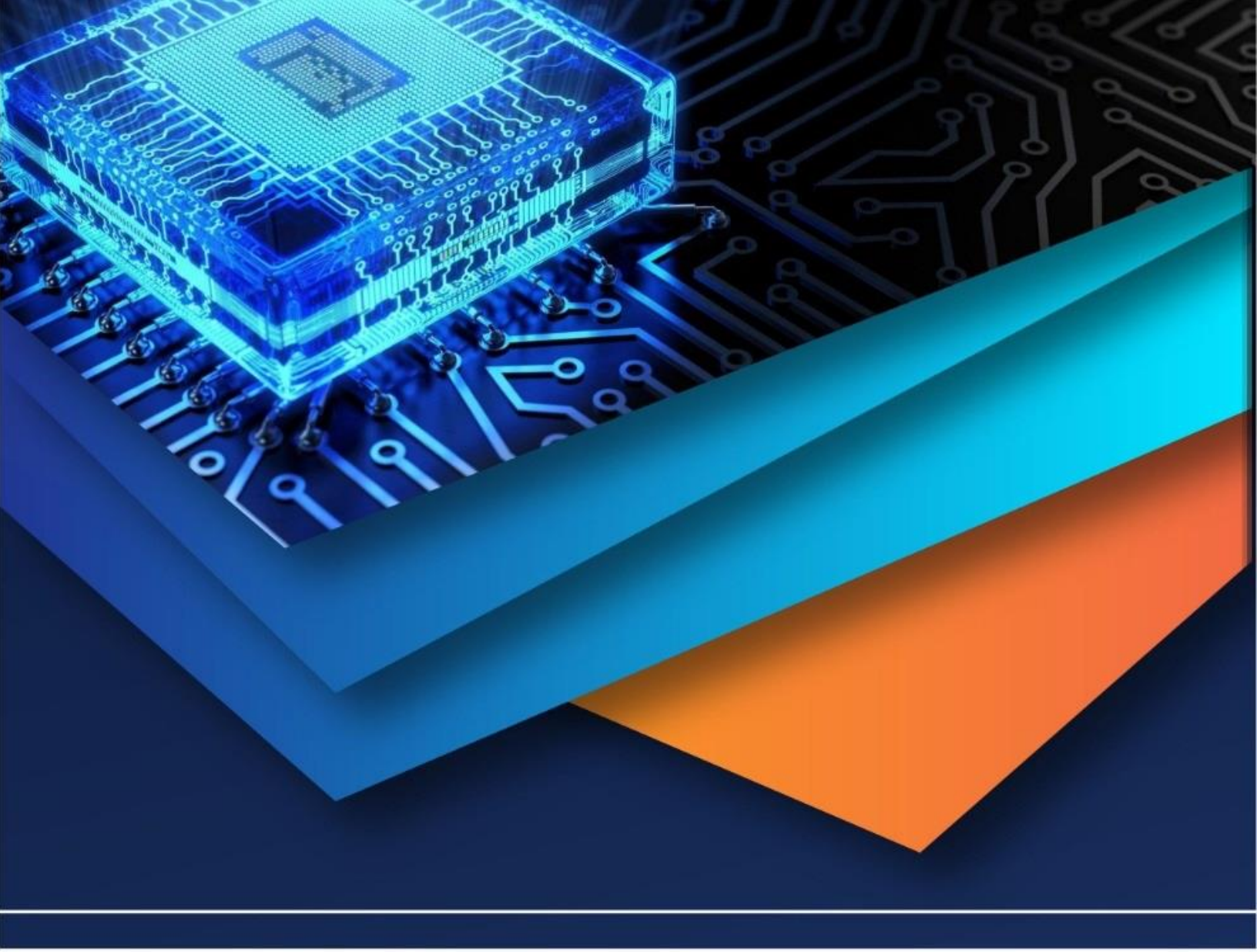
Rushikesh B. Kale (Bachelor of Computer Engineering) in Vidya Prasarini Sabha's College of Engineering and Technology Lonavala.
Email : krushikesh302@gmail.com



Jayesh B. Phale (Bachelor of Computer Engineering) in Vidya Prasarini Sabha's College of Engineering and Technology Lonavala. Email : jayeshphale4@gmail.com



Priyanka B. Sul (Bachelor of Computer Engineering) in Vidya Prasarini Sabha's College of Engineering and Technology Lonavala.
Email : priyankasul101@gmail.com



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)