# PRIVACY PRESERVING QUERY OVER ENCRYPTED GRAPH STRUCTURED DATA IN OUTSOURCED ENVIRONMENT

Nivedhitha [1],Meenaraji N[2], Roobini E[3]

[1]*Professor,Dept. of CSE, Panimalar Engineering College, Chennai , India*
[2,3] *IV Yr Dept. of CSE, Panimalar Engineering College, Chennai , India*

**Abstract**

*The expanding enthusiasm for gathering and distributing a lot of people information open for various purposes. For example, therapeutic research, showcase examination and practical measures have made significant security worries about people delicate data. To manage these worries, numerous privacy-preserving data publishing (PPDP) systems have been proposed in writing. In proposed system, the admin decides to design a smart examination system which should be conducted online. Initially, the admin allocates ten question papers for the exam. After allocation, these files should be in an encrypted form. For every file, different key will be generated and the entire folder will have single key to access .The admin shares the file to the college then they get an encrypted file as a key. This system uses a automatic technique for selecting keys randomly. That key will be merged in a QR code and if the student has registered in that college ,they can login and scan that QR code and get the key. If the entered key is correct then a request is sent to the admin. In respect to that key the admin will send the file to the student.*

**Keywords**— *PPDP, encrypted, QR code.*

## I. INTRODUCTION

Privacy preserving data mining is a method of protecting the privacy of data without sacrificing the utility of data.In this present world of internet people have become well aware that they should not share their personal data and sensitive information .This may lead to negative results of data mining.To overcome security issues PPDM is used here.There are various PPDM directions adopted to avoid disclosure of sensitive information some of them are privacy preserving data publishing(PPDP),query auditing,cryptographic methods and changing miningresults.

We first present a novel multi-variable privacy characterization and quantification model. Based on this model, we are able to analyze the prior and posterior adversarial belief about attribute values of individuals. We can also analyze the sensitivity of any identifier in privacy characterization. Then we show that privacy should not be measured based on one metric. We demonstrate how this could result in privacy misjudgment. We propose two different metrics for quantification of privacy leakage, distribution leakage and entropy leakage. Using these metrics, we analyzed some of the most well-known PPDP techniques such as k-anonymity, l-diversity and t-closeness. Based on our framework and the proposed metrics, we can determine that all the existing PPDP schemes have limitations in privacy characterization. Our proposed privacy characterization and measurement framework contributes to better understanding and evaluation of these techniques. Thus, this paper provides a foundation for design and analysis of PPDP schemes.

## II. LITERATURE SURVEY

*A. k-anonymity: A model for protecting privacy*

This paper was submitted by L. Sweeney in 2002. Consider a data holder, such as a hospital or a bank, that has a privately held collection of person-specific, field structured data. Suppose the data holder wants to share a version of the data with researchers. How can a data holder release a version of its private data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain

practically useful? The solution provided in this paper includes a formal protection model named k-anonymity and a set of accompanying policies for deployment.

*B. Robust de-anonymization of large sparse datasets*

This paper was submitted by A. Narayanan and V. Shmatikov in 2018. We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset.

*C. L-diversity: privacy beyond k-anonymity*

This paper was submitted by A.Machanavajjhala, D.Kifer, J.Gehrke and M. Venkitasubramaniam in 2007. Publishing data about individuals without revealing sensitive information about them is an important problem. In this paper we show with two simple attacks that a \kappa-anonymized dataset has some subtle, but severe privacy problems. First, we show that an attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes. Second, attackers often have background knowledge, and we show that \kappa-anonymity does not guarantee privacy against attackers using background knowledge.

*D. Privacy-preserving data publishing: A survey of recent developments*

This paper was submitted by B. C. M. Fung, K.Wang, R. Chen, and P. S. Yu in 2007. Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data privacy. Recently, PPDP has received considerable attention in research communities, and many approaches have been proposed for different data publishing scenarios. In this survey, we will systematically summarize and evaluate different approaches to PPDP, study the challenges in practical data publishing, clarify the differences and requirements that distinguish PPDP from other related problems, and propose future research directions.

## III. METHODOLOGY

In proposed system, the admin decides to design a smart examination system which should be conducted online. Initially, the admin allocates ten question papers for the exam. After allocation, these files should be in an encrypted form. For every file, different key will be generated and the entire folder will have single key to access .The admin shares the file to the college then they get an encrypted file as a key. This system uses a automatic technique for selecting keys randomly. That key will be merged in a QR code and if the student has registered in that college they can login and scan that QR code and get the key. If the entered key is correct then a request is sent to the admin. In respect to that key the admin will send the file to the student.

It uses two algorithms

1) AES Algorithm
2) Privacy Preserving Data Mining algorithm

*AES Algorithm:*

AES has been adopted by the U.S. government and is now used worldwide. It supersedes the Data Encryption Standard (DES), which was published in 1977. The algorithm described by AES is a symmetric-key algorithm, meaning the same key is used for both encrypting and decrypting the data.Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128-, 192- and 256-bits, respectively. The Advanced Encryption Standard, or AES, is a symmetric block cipher chosen by the U.S. government to protect classified information and is implemented in software and hardware throughout the world to encrypt sensitive data. The AES encryption algorithm defines a number of transformations that are to be performed on data stored in an array.

*Privacy preserving data mining algorithm:*

Privacy preserving data mining is a method of protecting the privacy of data without sacrificing the utility of data.In this present world of internet people have become well aware that they should not share their personal data and sensitive information .This may lead to negative results of data mining.To overcome security issues PPDM is used here.There are various PPDM directions adopted to avoid disclosure of sensitive information some of them are privacy preserving data publishing(PPDP),query auditing,cryptographic methods and changing mining
results. .

## IV. ARCHITECTURE

The systems architect establishes the basic structure of the system, we propose a AES Algorithm and we can put a small part of data in local machine and fog server in order to protect the privacy. Moreover, based on computational intelligence, this algorithm can compute the distribution proportion stored in cloud, fog, and local machine, respectively. Through the theoretical safety analysis and experimental evaluation, the feasibility of our scheme has been validated, which is really a powerful supplement to existing cloud storage scheme.

This system proposes the following nine modules

- User interface design
- Admin Upload the data with Secret keys.
- Admin login
- Admin Send Keys to College
- Key Verification
- Select Random Key
- College Send Bar-Code To Student
- Student Request To Admin
- Admin Response

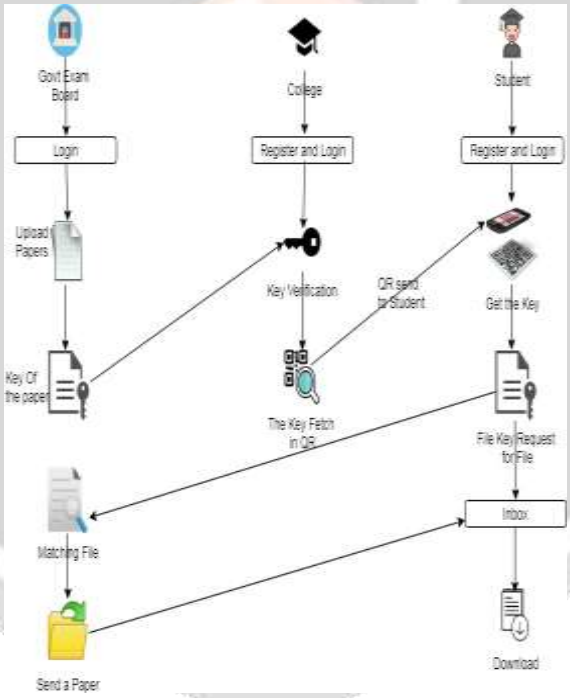


Fig 1

## V. MODULES

This system proposes the following nine modules

- User interface design
- Admin Upload the data with Secret keys.
- Admin login
- Admin Send Keys to College
- Key Verification
- Select Random Key

- College Send Bar-Code To Student
- Student Request To Admin
- Admin Response

*User registration and login details:*

This is the first module of our project. The important role for the user is to move login window to user window. This module has created for security purpose. In this login page we have to enter login user id and password. It will check username and password if it is valid or not. If we enter any invalid username or password we can't enter into login window to user window it will show an error message. So it prevents unauthorized user entering into the login window to user window. It will provide a good security . So server contains user id and password and also checks the authentication of the user.

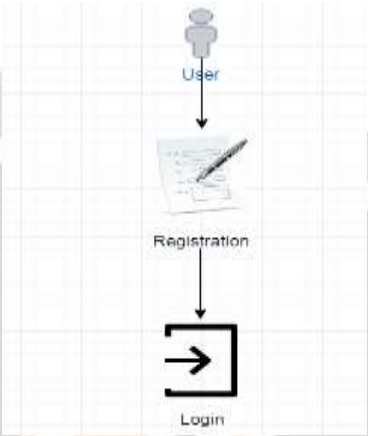


Fig 2

*Admin Upload The Data With Secret Keys:*

This is the second module of this project. Here, the Admin uploads some papers for student examination. So whenever the papers are uploaded secret keys are generated for security purpose.

*Admin Login:*

This is the third module in this project, here it symbolizes a unit of work performed within a database management system (or similar system) against a database, and treated in a coherent and reliable way independent of other transactions. Here the user will upload the file and request the admin to accept it.

*Admin Send Keys To College:*

This is the fourth module in this project. All the uploaded files will have separate keys. So the admin creates one common key for all secret keys. . The admin sends the common key to the college .

*Key Verification:*

In this module, if the correct key is entered then it has the access to open all the secret keys or else the access is denied.

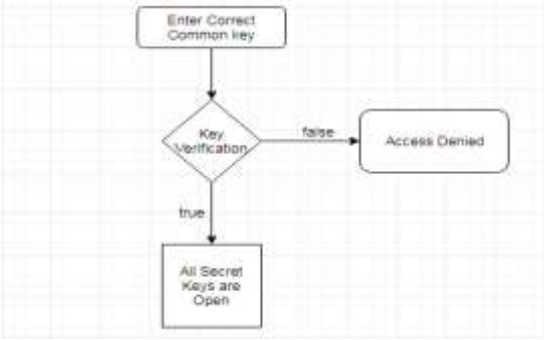


Fig 3

*Select Random Key:*

In this module, if the college enters the correct key then all secret key are appeared .This system uses an automatic technique to select keys randomly.

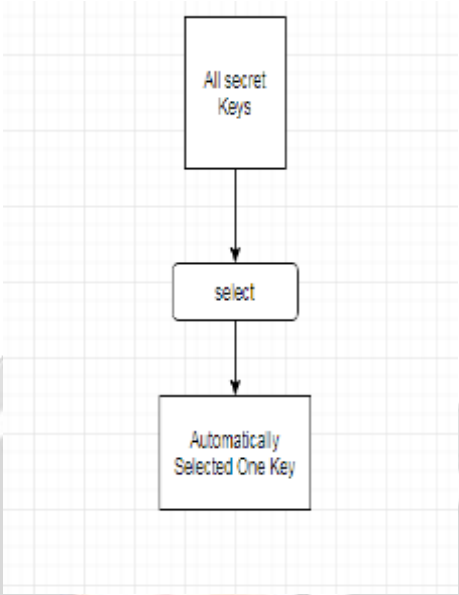


Fig 4

*College Send Bar-Code To Student:*

In this module, a key is selected automatically after that key is inserted in a QR image(BAR Code). While Scanning the QR code a key is generated . Then scanned QR code is sent to the student.

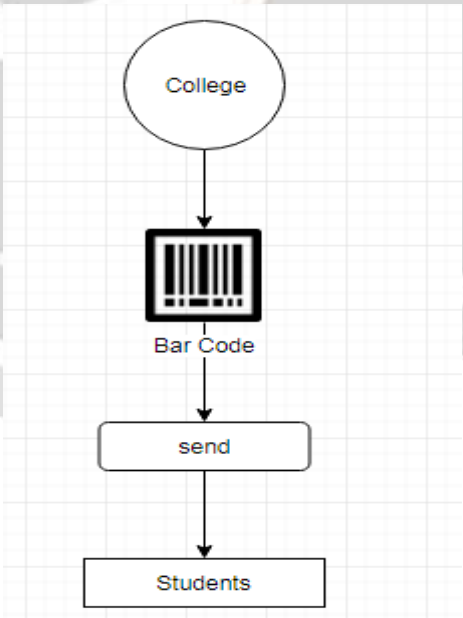


Fig 5

*Student Request To Admin:*

In this module, when the student open the inbox a QR image is appeared. Now when the students scan the QR code again a key is generated . The Students request the Admin for that key file.

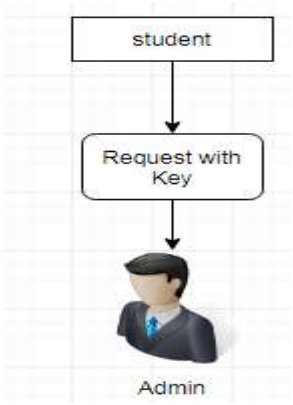


Fig 6

*Admin Response:*

In this final module, the admin match all the keys with the Student key.If any key matches with the student key, a file will be sent to the Student Inbox. If none of the key matches with the student key, he cannot get and access the paper.

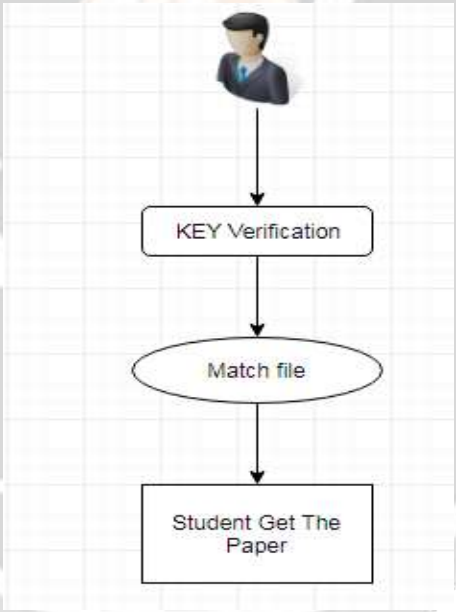


Fig 7

## VI. RESULT

Obviously, this system satisfies the main objective goal of its design that is to provide people with more security and privacy techniques related to online examination system. . In future enhancements, this system provides file and secret key Stored in different Schemas. Time Allocation for entering the key in key verification part.

## VII. CONCLUSION

In this project, we have a tendency to introduce comprehensive characterization and novel quantification ways of privacy to deal with the matter of privacy quantification in privacy preserving data commercial enterprise. so as to think about the privacy loss of combined attributes, we have a tendency to bestowed information commercial enterprise as a multi-relational model. we have a tendency to re-defined the previous and posterior beliefs of the individual. The projected model and adversarial beliefs contribute to a additional precise privacy

characterization and quantification. Supported by perceptive examples, we have a tendency to then showed that privacy couldn't be quantified based on one metric. we have a tendency to projected 2 completely different privacy outpouring metrics. supported these metrics, the privacy leakage of any given PPDP technique might be evaluated. Our experiments demonstrate however we have a tendency to might gain an improved judgment of existing techniques and facilitate analyze their effectiveness in reaching privacy.

## VIII. REFERNCES

[1] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557–570, 2002.

[2] L. Sweeney, "Uniqueness of simple demographics in the U.S. population," 2000.

[3] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in Security & Privacy, pp. 111–125, 2008.

[4] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "`-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, Mar. 2007.

[5] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in ICDE, pp. 106–115, 2007.

[6] N. Li,W. Qardaji, D. S. Purdue, Y.Wu, andW. Yang, "Membership privacy: A unifying framework for privacy definitions," in CCS, (Berlin, Germany), 2013.

[7] I. Wagner and D. Eckhoff, "Technical privacy metrics: a systematic survey," CoRR, vol. abs/1512.00327, 2015.

[8] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "Privbayes: Private data release via bayesian networks," in Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14, (New York, NY, USA), pp. 1423– 1434, ACM, 2014.

[9] M. G¨ otz, S. Nath, and J. Gehrke, "Maskit: Privately releasing user context streams for personalized mobile applications," in Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12, (New York, NY, USA), pp. 289– 300, ACM, 2012.

[10] Y. Rubner, C. Tomasi, , L. J., and Guibas, "The earth mover's distance as a metric for image retrieval," International Journal of Computer Vision, vol. 40, no. 2, pp. 99–121, 2000.