IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

Privacy Preserving Data Publishing Using Cryptographic Techniques

Shridhar V. Karad, Dr. Manikrao L. Dhore

Department of Computer Engineering Vishwakarma Institute of Technology, University of Pune, Maharashtra, India

ABSTRACT: In this paper privacy preserving data publishing is the important in day to day life. Data publishing has generated much concern on individual privacy. Recent work has focused on different background knowledge and their various threats to the privacy of published data.

Privacy is an important issue when one wants to make use of data that involves individuals' sensitive information. Research on protecting the privacy of individuals and the confidentiality of data has received contributions from many fields, including computer science, statistics, economics, and social science. Driven by mutual benefits, or by regulations that require certain data to be published, there is a demand for the exchange and publication of data among various parties. Data in its original form, however, typically contains sensitive information about individuals, and publishing such data will violate individual privacy. The current practice in data publishing relies mainly on policies and guidelines as to what types of data can be published, and agreements on the use of published data. This approach alone may lead to excessive data distortion or insufficient protection. 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.

KEYWORDS: Steganography techniques, AES, DES, Encryption, Privacy, Security.

I. INTRODUCTION

In recent time globally networked society places great demand on the collection and sharing of person-specific data for many new uses. This happens at a time when more and more historically public information is also electronically available. When these data are linked together, they provide an electronic image of a person that is as identifying and personal as a fingerprint even when the information contains no explicit identifiers, such as name and phone number. Other distinctive data, such as birth date and postal code, often combine uniquely and can be linked to publicly available information to re-identify individuals.

Data Encryption is the process of converting the plaintext into Encoded form (non-readable) and only authorized person/parties can access it. Data security is an essential part of an Individual/organization; it can be achieved by the using various methods. The encrypted data is safe for some time but never think it is permanently safe. After the time goes on there is chance of hacking the data by the hacker. Fake files are transmitted in the same manner as one can sends the encrypted data. Network security prevents the data in a network from unauthorized access. It involves the authorization of access to information throughout a network and it is measured by network administrator. The need for security is to protect the information as well as provide authentication and access control for resources, guarantee availability of resources.

Data mining research deals with the extraction of potentially useful information from large collections of data with a variety of application areas such as customer relationship management, market basket analysis. Privacy preserving has originated as an important concern with reference to the success of the data mining. Privacy preserving data mining (PPDM) deals with protecting the privacy of individual data or sensitive knowledge without sacrificing the utility of the data. People have become well aware of the privacy intrusions on their personal data and are very unwilling to share their sensitive information.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

II. LITERATURESURVEY

Ding qi Yang et. al [1] Personalized recommendation is crucial to help users find pertinent information. It often relies on a large collection of user data, in particular users' online activity (e.g., tagging/rating/checking-in) on social media, to mine user preference. However, releasing such user activity data makes users vulnerable to inference attacks, as private data (e.g., gender) can often be inferred from the users' activity data. In this paper, we proposed PrivRank, a customizable and continuous privacy-preserving social media data publishing framework protecting users against inference attacks while enabling personalized ranking-based recommendations. Its key idea is to continuously obfuscate user activity data such that the privacy leakage of user-specified private data is minimized under a given data distortion budget, which bounds the ranking loss incurred from the data obfuscation process in order to preserve the utility of the data for enabling recommendations.

JIANZHE ZHAO et.al [2] Large-scale spatiotemporal data mining has created valuable insights into managing key areas of society and the economy. It has encouraged data owners to release/publish trajectory datasets. However, the ill-informed publication of such valuable datasets may lead to serious privacy implications for individuals. Moreover, as a major goal of data protection, balancing privacy and utility remains a challenging problem due to the diversity of spatiotemporal data. However, the user dimension was not considered for traditional frameworks, which limits the application at the global level as opposed to the user level. Many researchers overcome this issue by assuming that a user in the dataset generates only one trajectory. Actually, a user always generates multiple and repetitive trajectories during observation. Only considering one trajectory for one user may cause insufficient privacy protection at the trajectory level alone, as a user's privacy can be manifested in many trajectories collectively. In addition, it demonstrates strong user correlation when using multiple and repetitive trajectories. If not considered, additional information will be lost, and the utility will be decreased. In article, author proposes a novel privacy-preserved trajectory data publishing method, i.e., IDF-OPT, which can reduce global least-information loss and guarantee strong individual privacy. Comprehensive experiments based on an actual trajectory publishing benchmark demonstrate that the proposed method maintains high practicability in trajectory data mining.

Jingcheng Song et.al [3] The development of the Internet of Things (IoT) and 5th generation wireless network (5G) is set to push the smart agriculture to the next level since the massive and real-time data can be collected to monitor the status of crops and livestock, logistics management, and other important information. Recently, COVID- 19 has attracted more human attention to food safety, which also has a positive impact on smart agriculture market share. However, the security and privacy concern for smart agriculture has become more prominent. Since smart agriculture implies working with large sets of data, which usually sensitive, some are even confidential, and once leakage it can expose user privacy. Meanwhile, considering the data publishing of smart agriculture helps the public or investors to real-timely anticipate risks and benefits, these data are also a public resource. To balance the data publishing and data privacy, in authors paper, a privacy-preserving data aggregation scheme with a flexibility property uses ElGamal Cryptosystem is proposed. It is proved to be secure, private, and flexible with the analysis and performance simulation.

Akash Siddhpura & Prof. Daxa V. Vekariya [4] Due to the wide deployment of information technology, privacy concern has been major issue in data mining. So for that new path is identified which is known as Privacy Preserving Data Mining (PDDM). Available PDDM techniques are Perturbation, Generalization, Anonymization, Randomization and Cryptography. All of them have some advantages as well as disadvantages also. If apply only cryptography PDDM using symmetric key encryption algorithm, then there will chances of losing data, because if anyone knows the key then data is available to anyone. If author apply perturbation PDDM only then it will not give you accurate result. So if author will use cryptography and perturbation then it will achieve security as well as very less chances of losing data after applying the privacy preserving.

R. Ramya Devi1 & V. Vijaya Chamundeeswari1 [5] Triple DES offers a fairly simple technique of increasing the key size of DES shield against such attacks, devoid of necessitates to design an entirely new block cipher algorithm. Data anonymization work as an information sanitizer whose target is to defend the data privacy. It encrypts or takes away the personally recognizable data as of the data sets in order that the persons about whom the data designate remain anonymous. In authors work, a combination of anonymization and Triple DES are utilized that are shortly called as the A3DES algorithm. Experimental outcome reveals that the approach performed well when contrasted with all other related approaches.

R. Srinivas et.al [6] The important aspect in research is to publish data by conserving one's privacy. Enormous techniques have been proposed to tackle this issue. The main concept is to concentrate on how cogently one protects

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

individual privacy in publishing data without the exact attribute missing in the research. As per our knowledge is concerned, there raised a great scope for research on vertically partitioned distributed databases.

S. Sharma and A. S. Rajawat [7] nowadays, data mining techniques are massively used via organizations intended for converting huge amount of data into information. Due to the advancement in database technology data are present at distributed sites consequently for carrying data mining analyzing in cost effective way we need to integrate these distributed data at one site. The predicament of anxiety here is that privacy of individual is at risk and author find out sensitive information upon integration so a secure data model is needed to accomplish this task.

V. S. Susan and T. Christopher [8] Privacy preservation data is an emerging field of research in the data security. Numerous anonymization approaches have been proposed for privacy preservation such as generalization and bucketization. Recent works show that both the techniques are not suitable for high-dimensional data publishing. Several other challenges for data publishing are speed and computational complexity. Slicing is a novel anonymization technique which partitions the data both horizontally and vertically and solves the problem of high-dimensional complexity. To overcome the other challenges of speed and computational complexity, a new advanced clustering algorithm is used with slicing.

S. A. Onashoga et.al [9] Privacy preservation methods for anonymizing multiple sensitive attributes (MSA) data in the field of privacy-preserving data publishing (PPDP) mostly seek enforcement of the i-diversity privacy model on MSA coupled with quasi-identifier (QID) generalization and tuple suppression, resulting in high data degradation of the published releases. Most existing work produces static releases that are not dynamic and web-based. In this article, we propose KC-Slice, which is amodified LKC-privacy model and slicing technique, for anonymizing MSA data dynamically, to produce releases that preserve the dataset content from most attack models and reduce data degradation, through cell suppression and QID random permutation. Experimental results and evaluation using data metrics and information entropy show remarkable reduction in data degradation and suppression ratio.

S. Goryczka et,al [10] The collaborative data publishing problem for anonymizing horizontally partitioned data at multiple data providers. Author consider a new type of “insider attack” by colluding data providers who may use their own data records (a subset of the overall data) in addition to the external background knowledge to infer the data records contributed by other data providers.

III. PROBLEMSTATEMENT

The present work focuses on cryptography to secure the data while transmitting in the network. Firstly the data which is to be transmitted from sender to receiver in the network must be encrypted using the encryption algorithm in cryptography, user selects via which technique he wants to encrypt/decrypt data. Secondly, by using the decryption technique (as per the user’s choice), the receiver can view the original data. Key is sent via Email using Steganography.
Organizations that are collecting and maintaining data are being challenged to protect the data while using and sharing over the internet because data become an important source of profits, but also the threat to individual.
Many organizations can sale users’ data for financial gain or to do user behavior analytics so this will leads to a user’s privacy breach.

IV. PROPOSED SYSTEM

In this system privacy preserving data publishing cryptography by using AES and DES algorithms in cryptography. In these algorithms keys are private in the steganography technique. These means the data is private and safe in this system by using steganography technique.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

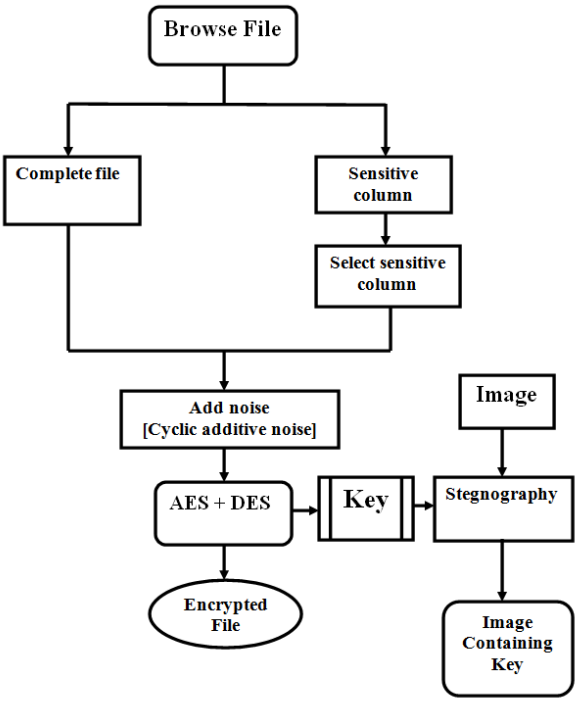

Fig 1: Architecture of proposed system.

In above Data Flow Diagram, the input of this DFD is browse file that means XL sheet or file where public data is collected on that XL file. Input gives to that system then one is the complete file otherwise sensitive column is in that system. If this file is in sensitive column then it can select the sensitive column. Then adding noise [Cyclic additive noise] in the system. Because of this adding noise there not chance for any information loss. So the data remain private in the system. AES and DES are the algorithms where by using the key we can save our private data. In the AES and DES algorithm steganography technique will be used for the data security. By using Steganography technique we get image containing key.

Data Publishing Formats: Data is usually published in table format with some specific attributes. In data privacy, attributes are categorized into:

1) Direct Identifiers (DI): as the name suggests, individual patient can be directly identified without applying any reverse engineering methods. The example of direct identifiers can be name, email-id, address etc.

2) Quasi Identifiers (QI): certain attributes in dataset can be combined with other attribute/s which can uniquely identify an individual in the data publishing table. For example, individual gender as an attribute is very difficult to identify but combination of gender, sex, zip code, date of birth can be identified uniquely.

3) Sensitive attributes (SA): the attributes which supposed to be not published with an individual. The example includes genomic information, special diseases etc.

Steganography is formed from the two Greek words 'Steganos' meaning 'Covered' and 'Graphene' meaning 'writing', which refers to 'Covered Writing'. It is the science of hiding the existence of information into information. The information is embedded into a cover or carrier object so that no one can understand the presence of information. A key is used for embedding procedure without which the adversary cannot be able to detect the embedded message. The altered new object is called stego object. Image, audio, video etc. can be the cover objects. Fig shows the general concept of steganography.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

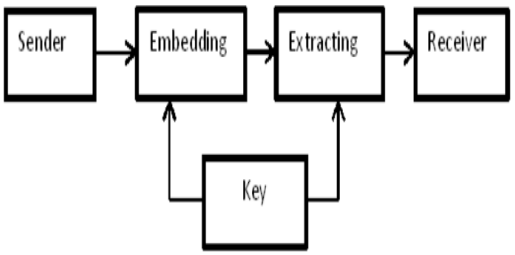

Fig 2:General Concept of Steganography

ALGORITHM

AES Algorithm

AES (acronym of Advanced Encryption Standard) is a Symmetric encrypt algorithm. AES bits for encrypt/decrypt the data and fortifies lengths are 128,192 and 256 bits.

Byte Supersede (Sub Bytes) The 16bit of info information is fine-turned layouts and result in network shapes lines and section.

Circular byte Shift rows Every four lines of matrix network are moved to left positions for each round other.

Mix Columns The yield of another framework is store of 16 nascent bytes and in last round this progression in not rehashed.

Add round key the 16bytes of input matrix and round key and output will stored in cipher text 128 bits and 16 bytes homogenous round of interpreted the data

Decryption the tasks of decode of an AES cipher text activity in the inconsistency request. All round comprises of the four stage directed in the logical inconsistency request.

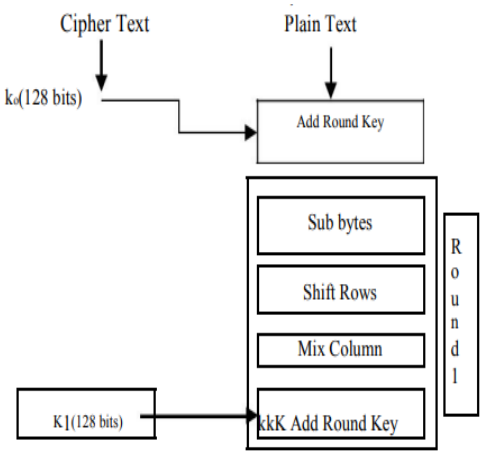

Fig 3: AES Encrypt

DES Algorithm

Information Encryption Standard (DES) may be a symmetric-key. DES is a utilization of a Fiestel Idea having a cleared out half that's a culminate reflect image to right half of the correct half. It utilizes 16 circular Fiestel structure. The 64 bit information and key length is 56 bit for scramble information 8 and 64 bit are not utilized. DES is anticipate a head on the Fiestel Cipher, code all that must outline DES is

Round function

Key schedule

Any ads citations processing − Inceptive and eventual organization combination of ordering

Introductory and Last Stage: In to begin with and objective p-boxes are inverses orchestrate of each Circular function. The heart of this cipher is the DES work, f. The DES work petition a 48- bit key to the farthest right 32 bits to cause a 32-bit yield. In cryptography strongly produced the sequenced of cipher. While Fiestel multilevel round divide into cipher for Fiestel network.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

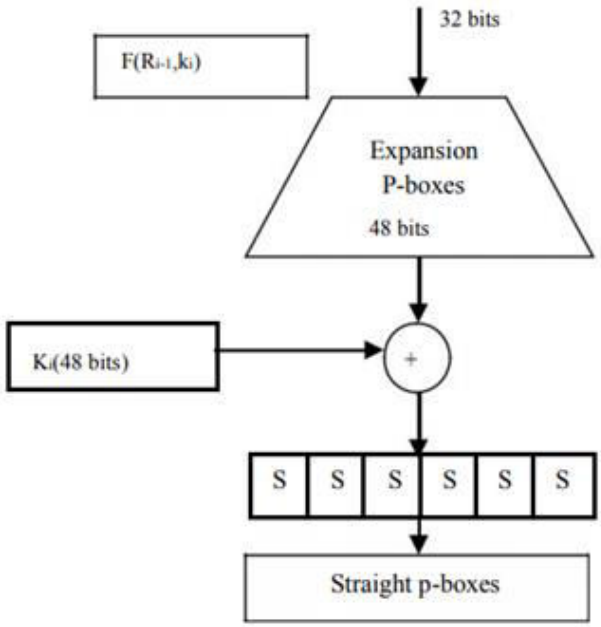

Fig. 4: DES Encrypt

V. CONCLUSION

Privacy preserving can be achievedby using two techniques, by adding the noise and using the cryptography we can protect the data. Here Data loss will be 0%.But it will some time while performing steganography. There are no chances of data loss. While if we apply only any other technique then there will be chances of data loss. If we apply only AES and DES technique then quality of data will not that much good, we had improved quality of the data here also .

VI. FUTURE WORK

Our future work will focus on SLSB which replace LSB technique (Steganography technique).

REFERENCES

[1] Dingqi Yang, Bingqing Qu, and Philippe Cudr´e-Mauroux "Privacy-Preserving Social Media Data Publishing for Personalized Ranking-Based Recommendation "JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

[2] JIANZHE ZHAO 1,4, JIE MEI2, STAN MATWIN3,5, YUKAI SU1, AND YUANCHENG YANG1 "Risk-Aware Individual Trajectory Data Publishing With Differential Privacy "Received December 24, 2020, accepted December 28, 2020, date of publication December 31, 2020, date of current version January 13, 2021.

[3] Jingcheng Song, Qi Zhong, Weizheng Wang, Chunhua Su, Zhiyuan Tan, and Yining Liu "FPDP:Flexible Privacy-preserving Data Publishing Scheme for Smart Agriculture "IEEE SENSORS JOURNAL, VOL. XX, NO. XX, 0601 2020

[4] Akash Siddhpura & Prof. Daxa V. Vekariya " An approach of Privacy Preserving Data mining using Perturbation & Cryptography Technique "International Journal on Future Revolution in Computer Science & Communication Engineering IJFRCSCE |APRIL 2018

[5] R. Ramya Devi1 & V. Vijaya Chamundeeswari1 "Triple DES: Privacy Preserving in Big Data Healthcare "Received: 20 May 2018 / Accepted: 30 July 2018 © Springer Science+Business Media, LLC, part of Springer Nature 2018

[6] R. Srinivas, K. A. Sireesha, and S. Vahida, "Preserving Privacy in Vertically Partitioned Distributed Data Using Hierarchical and Ring Models," Advances in Intelligent Systems and Computing Artificial Intelligence and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001022 |

Evolutionary Computations in Engineering Systems, pp. 585–596, 2017.
[7] S. Sharma and A. S. Rajawat , "A secure privacy preservation model for vertically partitioned dist ributed data," 2016 International Conference on ICT in Business Industry & Government (ICTBIG), 2016.
[8] V. S. Susan and T. Christopher, "Advanced Cluster-Based At tribute Slicing: A New Approach for Privacy Preservation," Proceedings of the International Conference on Soft Computing Systems Advances in Intelligent Systems and Computing, pp. 205–213, 2016.
[9] S. A. Onashoga, B. A. Bamiro, A. T . Akinwale, and J. A. Oguntuase, "KC-Slice: A dynamic privacy-preserving data publishing technique for multisensitive attributes," Information Security Journal: A Global Perspective, vol. 26, no. 3, pp. 121–135, Apr. 2017.
[10] S. Goryczka, L. Xiong, and B. Fung, "m-Privacy for Collaborative Data Publishing," Proceedings of the 7th International Conference on Collaborative Computing: Networking, Applications and Work sharing, 2011.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details