



User Data Protection (UDP) and Secure Content Retrieval (SCR) in OSN

¹Parveen Kumar, ²Dr. Mukesh Kumar Gupta

¹Research Scholar, Dep of CSE, Suresh Gyan Vihar University Jaipur

²Associate Dean Research, Suresh Gyan Vihar University Jaipur

ABSTRACT:

Privacy-preserving Data Mining is the main field of technology for data mining. It's a way to protect sensitive information from unintentional or unwanted disclosure. Data mining methods can help determine and forecast the most useful information. PPDM abstraction is focused on the security of private communications from access by unauthorized persons. There are a variety of ways to use PPDM to improve security and privacy. They include the Secure Data Contribution Retrieval Algorithms, Enhanced Attribute-Based encryption, Level-by-level Security Optimization and Content Visualization and Level-by-level encryption. The current issues are being addressed with SDCRA, which is a proposal SDCRA. The SDCRA algorithm develops privacy guidelines. Security is then arranged in accordance with compatibility and requirements. This algorithm is able to meet the accuracy requirements of various data sets. Online Social Networks are popular interactive media that allow users to connect, share and distribute large quantities of personal information about humans. They will also be able transmit data across OSNs by using the enhanced ABE method. Data leakage may occur during the storage phase when sensitive data is stored and collected in an unsecure location. A third advantage of the SSOV algorithm's privacy security is visualization of data and sharing of content. After analyzing the privacy compatibility of OSN, they alter their privacy settings to satisfy the user's needs on social networks. Fourthly fourth, the PPHE framework determines if tweets contain private or confidential information. It applies data suppression methods for tweets classified as private. A test analysis is conducted with the data that is gathered from the social data.

497

Keywords: OSN, Security, Data Mining and data security.

DOI Number: 10.14704/nq.2022.20.8.NQ44057

NeuroQuantology 2022;20(8):497-504

I. INTRODUCTION

Data mining is an automated or semi-automatic process that extracts non-trivial, useful, previously unknown, implicit patterns and eventually follows the knowledge or patterns of large data sets. It allows end-users to explore data from many areas, including Social networks, Bioterrorism Applications and Medical database mining. Before any data

mining technique can be applied, data summarization and exploration must be done. This allows for a better understanding of the data and helps identify the information to be extracted. There are two types of data-mining methods: Descriptive data mining, such as clustering, sequence discovery, and association rules; and Predictive data mine,



which includes Classification, Regression and Time-series analysis.

Supervised vs Unsupervised Data Mining

Techniques: Alongside the labels, which define the type of observations used in the creation of the model The observations and measurements of training data are analyzed through trained learning. The model is based on the rules for classification and mathematical equations and decision trees. The model is able to determine the kind of data that is not known. Target marketing, credit approval and detection of future events are typical supervised learning applications. You can utilize the model to create labels for the new data. The data used for training doesn't have labels for classes. This is due to unsupervised learning. It is the place where data and measures are taken to find out if there are any clusters or classes present within the data. This is an unsupervised method that group entities that have similar characteristics. It is employed to recognize patterns and document classification, image processing and analysis of spatial data.

Data mining is an essential component of KDD. Data mining is an important step in the KDD process. Data mining techniques can extract information from different sources, like databases and text files. It is an interdisciplinary procedure that is in the hands of many disciplines, such as statistical analysis, AI as well as machine-learning, Soft Computing and information retrieval. It's an iterative system which includes the following steps. Certain databases are not consistent or insufficient. Pre-processing data is one way to make unification more effective in these databases. This leads to an improved and more connected database. It is therefore essential to separate the data from the database. This involves sessions identification, data cleaning knowledge discovery, session identification, and pattern analysis. Data Cleaning In the real world, data can be messy, uncoherent and noisy. It can also be incomplete, noisy, or even unclean. Data cleaning is required to keep data consistent, eliminate noise , and ensure that data is

complete. Data cleaning is a method to erase lost values. the correct values are flattened and the absence of errors or exceptions is addressed using various methods. The data records include information on video and graphic formats. There is GIF, JPEG and CSS details. Each URL field may contain suffixes to filenames.

II. LITERATURE SURVEY

For PPDM There are a variety of variables that can be utilized or used or. They include data distribution, modification , mining algorithms rules handling, privacy preservation as well as rule-handling. The first aspect is distribution of data. This is the method by which data is distributed. One method is to use centralized data, and another are distributed databases. Data can be distributed vertically or horizontally across distributed databases. Data modification is the next step. This is the process of changing the form of data that was originally stored in an alternative format in order that sensitive information cannot be identified. The noise is then multiplied or processed to alter actual data. Data modification may involve several techniques like randomization, anonymity, swapping as well as blocking and sampling. The third aspect in which data mining takes place can be described as known as the Data-Mining algorithm. It can be utilized to guard privacy of individuals. Fourth dimension refers to protecting data from mining. Data is taken and kept hidden. The fifth dimension concerns privacy during data mining.

The objective is to share the most efficient way to protect privacy in data mining. It allows the computation of the functional statistics for all datasets without divulging the private nature of the user's data. PPDM is a technique that permits the calculation of statistical functions. The methods were created following an extensive research into data mining in the recent years. The majority of techniques employ some type of transformation to the original data to ensure that it is protected from privacy. This is essential to ensure privacy even after data is utilized for mining. The classification methods



can be described in the following manner: Srikant and Agrawal (2000) looked into the initial data allocation and created a fundamental reconstruction method to precisely estimate it. The accuracy of classifiers built from the original data and the sanitized version was evaluated to the accuracy of the classifiers by their authors. This model of predictive ability isn't able to build large quantities of user data perturbation. The authors suggested that the randomization technique be employed to create categorical data. (Verykios et. and. (2004)) suggested a hierarchical approach for classifying the processes that is used in PPDM. The author further developed the analysis and clustering methods for various PPDM algorithms. Certain data mining techniques can be used to study privacy concerns. Oliveira, Zaiane (2002a) investigated a way to conceal the most frequently used itemsets by employing different sanitization strategies by using an algorithmic framework. The design of item-restrictions can be utilized to reduce noise creation and restrict real-time dataset elimination. They also recommended that authors study the best methods for cleaning and the implications of data mining privacy within the association rule and pattern policies.

Hong et al. Hong et al. (2013) gave each of the transactions a specific weight applying the SIF-IDF method. The sanitization procedure was determined by which score is the highest and the lowest. They suggest sophisticated techniques can enhance the efficiency of the method proposed. Han as well as Ng (2007, Han and Ng) suggest the use of a secure protocol to find the most appropriate set of rules, without divulging private information. This protocol is based upon The True Positive Rate as well as Genetic

Algorithm (GA) results in relation to the false Positive Rate (FPR). This decision-making rule assessed. This author proposes that the method can be combined with optimization techniques to increase the efficiency. Lin et al. Lin et al. (2015) presented a number of GA-based techniques, such as cpGA2DT and pGA2DT. They are employed to conceal important data and differentiate the transactions of the target. In the future The victim is the collection of software that encode chromosomes and create separate gene-related transactions in chromosomes. Fitness functions are constructed with previously established weights to demonstrate the chromosome's capability. It is based on three different results. Weights that are pre-defined are still needed to ensure that algorithms are effective in identifying the most effective transactions to delete. This can affect the performance of the methods that have been developed. The authors suggested that PPDM methods be employed to increase performance.

III. PROPOSED METHODOLOGY

This section explains the procedure for implementing the proposed methodology. It is intended to help you understand the conceptualization and framework. These sections provide information about the methodology, preprocessing steps, and algorithm exploration along with logical and functional details. Level by level Security Optimization and Content Visualization Algorithm shows a diagrammatic representation to show the proposed architecture to display user profile and content in online social media networks. 1. These techniques recommend that you make a friendship or a relationship with another user, but not compromise on the privacy of your profile.



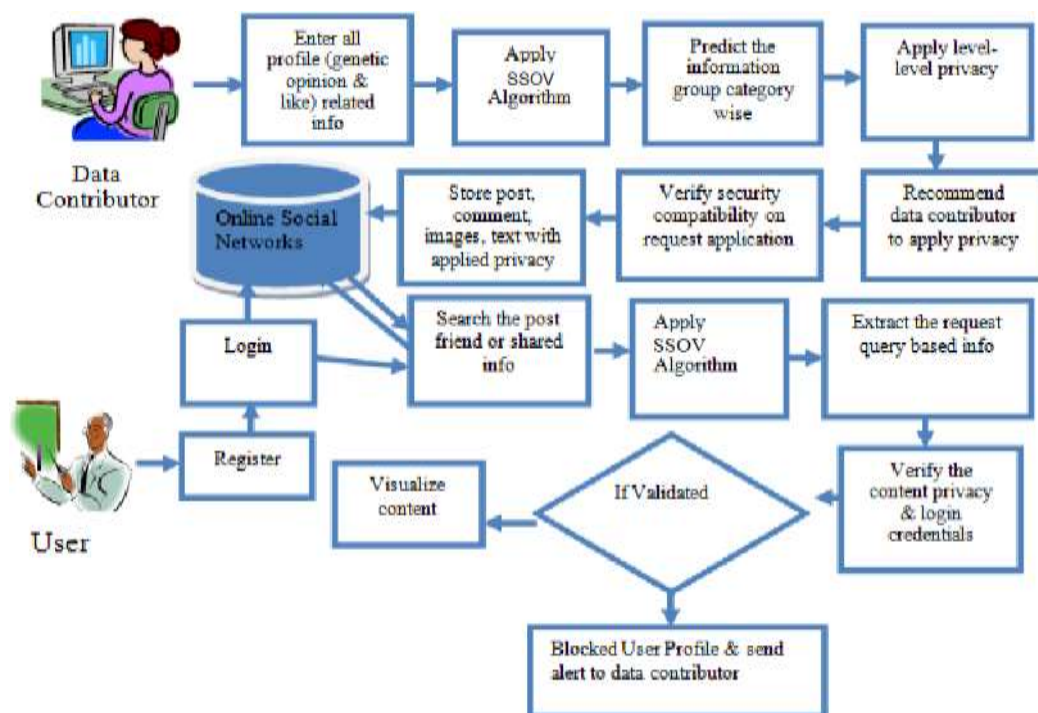


Figure 1: Step by step security, optimization and visualization (SSOV) model

System Architecture and Data Pre-processing:

This module creates a robust and efficient social network platform that uses interactive content. It allows users to connect with each other from many perspectives, such as business, academia, research, or outsourcing. This platform aims to improve the efficiency and privacy of the network. The user has complete control over the system without the need for a storage server or an online social network service provider.

Online social Network Security: Online security via Social Networks is directly related to users' concerns about potential harms. Social limitations are disrupted by prior knowledge of technologically-mediated communications. Users are, therefore "consumers" for services. They use social media to interact with their friends and family and to access information and discussion. It expands the sense of belonging and matters of the heart. These activities are made public to friends, or large numbers of people. This is a critical component of Online Social Networks. Access Control uses user modelling to create strong privacy and meet users' information management needs.

Controlling and Authentication: SSOV is used to establish the premise for potentially

adversarial entities operating or monitoring OSNs. OSNs can store a lot of user information, such as posts, photos, and interactions. An adversarial entity can acquire user details and use them in unintended ways, possibly to the disadvantages of those who are associated with the content. The step by step security between user levels is used to authenticate interaction and behavioural contents.

The Proposed SDCRA algorithm is explained as follows: The Tuples D contains the Credential Attribute A and Identifiers C. These are stored in Table T, which is used as input for the SDCRA method. Below is the pseudo-code for Secure Data Contribution Retrieval.

Algorithm 1 Proposed SSOV Algorithm
Input: User Profile (UP) and Organizational Network User Profile (ONUP)
Output: User Data Protection (UDP) and Secure Content Retrieval (SCR)

- Procedure
- Step 1: Browse the Online Social Networks (OSNs);
 - Step 2: Proceed for registration;
 - Step 3: Collect the genetic, interest, opinion, and professional information



Step 4: Store the UP or ONUP in database
Step 5: Categorize the UP information
Step 6: Apply level by level privacy based UP credential information
Step 7: Verify the UP and review the request for application in OSNs
Step 8: Authenticate the user and offer the application control and authentication;
Step 9: Contribute and share the information;
Step 10: Verify the application privacy compatibility and UP;
Step 11: If accessible
Allow user to the contribution;
Step 12: Else
Send alert to UP the verify their accessibility or login credentials;
Search the content or UP in OSNs;
Verify the accessibility of content offered data provider;
Step 13: End If
Step 14: If applicable
Permits to view the information or UP;
Step 15: Else
Block the unauthorized User profile and send alert to respective user
D0. Step 16: End If
Step 17: End Procedure

The proposed algorithm system SSOV is designed to offer End up to End security in the online social media network. It is used to determine the User profile and the Organizational network user profile. Here is a thorough explanation of the pseudocode that is used in the algorithm proposed. This proposed SSOV algorithm categorizes public information, including friends' lists comment, posts, or comments and a community-wide sharing data. In the event of data sharing to OSN, SSOV requests that UP implement privacy protections to UDP. The SSOV also offers UP restrictions for users or groups based on creditworthiness and the opinion of the content. In Secure Content Retrieval, UP can search and retrieve any kind of data. The proposed SSOV will verify the authenticity of users and requests information privacy settings for content. It will send SCR to users who have met the OSNs and credential ship of

content standards. It sends an alarm to UP if it discovers an unauthorized or insecure user. Its proposed SSOV algorithms reduces the loss of information (IL) and time for retrieving content (CRT) and enhances the security of privacy-related retrieval. The Step-by-Step security Optimization as well as Content Visualization Algorithm (SSOV) is utilized to address security concerns when sharing content and data visualization. This approach works with a framework for social networks within a web-based platform. It employs step-by-step privacy based on the user's requirements within social networks. To ensure security The proposed method evaluates privacy compatibility in this online social media. Users can only share information they intend to share with other users, but no other information is shared with the public. It makes it simpler for individuals to post and download huge amounts of data without having to worry about privacy. It lets users manage and authenticate their accounts inside the online social network. This algorithm was created to decrease the requirement for cybercrime service branches and their providers to track the activities of users. It offers a reliable security and control mechanisms that lets user's login and log out of their accounts. This technique allows users to utilize social media that is secure. This method offers secure privacy controls and authentication when data is added or deleted from the media network.

IV. RESULT ANALYSIS

Table 1 displays the PRA, CRT and IL for 500KB, 1MB and 10MB data respectively. This table can be used to assess privacy efficiency and accuracy in online social networking environments. The LSOCV proposed is calculated using the following methods: Privacy-Preserving Algorithm Clustering Algorithms (PPK-Means), (Weifeng and al. 2011, collaborative fuzzy co-cluster analytical framework (CFCAF), (Erola and al. 2011), and Cluster-based L.diversity privacy preservation [CLDPP] (Bahri, et. al. 2018, methods.



Table 1 Performance Analysis of SSOV with 500KB File

Algorithm	File size 500 KB		
	PRA (Privacy Retrieval Accuracy) (percentage)	CRT (Content Retrieval Time)(ms)	IL (Information Loss) (percentage)
CFCAF	75.56	15	18
CLDPP	81.67	9	13
PPK-MEANS	62.35	16	19
SSOV	91.57	6.9	8

Table 2 Performance Analysis of SSOV with 1MB File

Algorithm	File size 1MB		
	PRA (percentage)	CRT (ms)	IL (percentage)
CFCAF	75.23	17	19
CLDPP	83.35	15	16
PPK-MEANS	64.44	19	21
SSOV	94.12	10	10

Data mining privacy protection protects sensitive data from hackers. There are many methods to protect data, including randomization, perturbation, anonymization and randomization. Each method has its advantages and disadvantages. Standardizing methods for the PPDM can help to balance security and utility. Tables 1 and 2 show that SSOV has the highest score for each parameter across all databases.

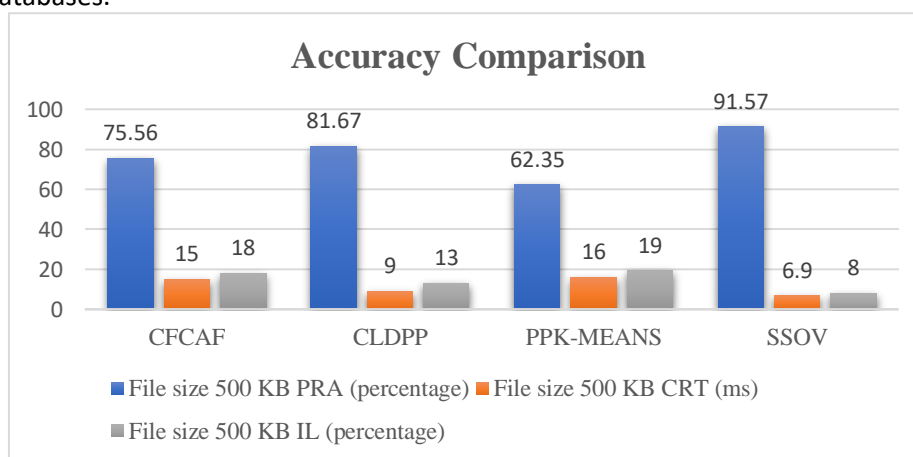


Figure 2: Accuracy comparison for all the algorithms

Figures 2 shows the graphical representations for PRA, CRT, and IL performance. Comparing the proposed SSOV method with other methods, such as CFCAF, CPK-Means, CLDPP, and CFCAF for file sizes that are different, the

results of the proposed method are shown. The accuracy, retrieval speed and information loss of the proposed SSOV method are measured.



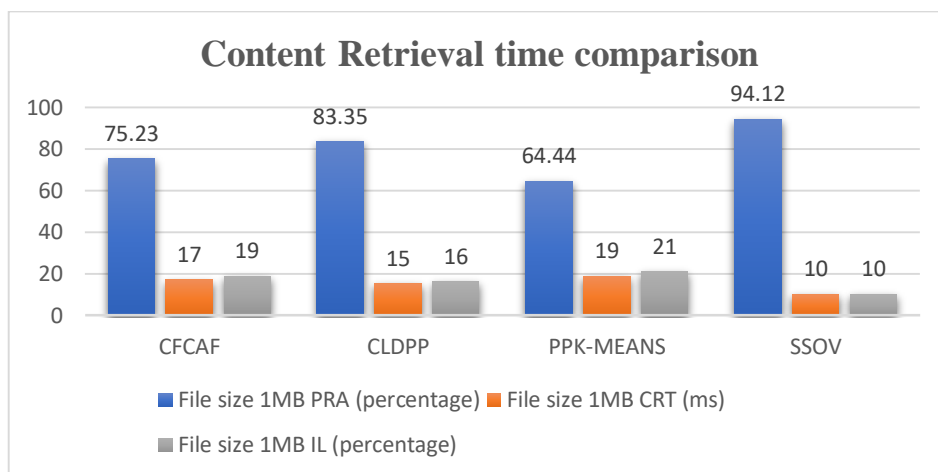


Figure 3: Comparison of content retrieval time for all the algorithms

Figure 3 shows SSOV's performance in terms of Content Recovery Time (CRT), compared to other methods with different data sizes, namely 500KB (1MB), 1MB and 10 MB. LSOCV achieved the lowest CRT, i.e. SSOV achieved the lowest CRT, taking 12 ms to process a file of 10 MB in size. Existing methods took 17-24 ms to process the same file. This clearly shows that the proposed SSOV performs better than the existing methods in all contexts.

V. CONCLUSION

This approach is designed to avoid privacy concerns when sharing content or data visualization is the subject. After assessing the security compatibility of the application, it then applies the privacy level by level to satisfy demands of users within social networks. The SSOV algorithm lets users participate in a step-by-step secure online social networking system. The public level lets users to gain access to general as well as personal information. Users also have access to private information and be protected. It gives organizations a single privacy policy and security for participants and users of data in social media environments. It helps reduce information loss (IL) as well as the time to retrieve content (CRT) and accuracy in privacy retrieval (PRA). PRA 91.97 is attained by the method proposed which is much superior to other methods which use files of 500KB. The SSOV was more efficient than other methods using 500 KB database files for CRT analysis. SSOV reduces Information Loss by up 88% when compared with conventional methods that utilize 500 KB. SSOV enhances

privacy Retrieval accuracy by 9.13 percent and reduces the Content Retrieval Time of 7 milliseconds and decreases the loss of information (IL) to 5.33 percent. The research could be further developed to include privacy-preserving strategies for big data Hadoop environments, without impacting the accuracy of data or retrieval speed.

VI. REFERENCES

- [1]. Agrawal, R. and Srikant, R. (2000), Privacy-preserving data mining, in 'ACM Sigmod Record', Vol. 29, ACM, pp. 439–450.
- [2]. Ahmed, G., Zou, J., Fareed, M. M. S. and Zeeshan, M. (2016), 'Sleep-awake energy efficient distributed clustering algorithm for wireless sensor networks', Computers and Electrical Engineering 56, 385–398.
- [3]. Andruszkiewicz, P. (2014), Hierarchical combining of classifiers in privacy preserving data mining, in 'International Conference on Hybrid Artificial Intelligence Systems', Springer, pp. 573–584.
- [4]. Attrapadung, N., Libert, B. and De Panafieu, E. (2011), Expressive key-policy attributebased encryption with constant-size ciphertexts, in 'International Workshop on Public Key Cryptography', Springer, pp. 90–108.
- [5]. Devi, S. S. and Indhumathi, R. (2019), A study on privacy-preserving approaches in online social network for data publishing, in 'Data Management, Analytics and Innovation', Springer, pp. 99–115.
- [6]. Dittrich, K. R., Hartig, M. and Pfefferle, H. (1988), Discretionary access control in "



structurally object-oriented database systems. in 'DBSec', pp. 105–121.

[7]. Fernandez, A., Garc ´ıa, S., Galar, M., Prati, R. C., Krawczyk, B. and Herrera, F. (2018), Introduction to kdd and data science, in 'Learning from Imbalanced Data Sets', Springer, pp. 1–17.

[8]. Fogues, R. L., Such, J. M., Espinosa, A. and Garcia-Fornes, A. (2014), 'Bff: A tool for eliciting tie strength and user communities in social networking services', Information Systems Frontiers 16(2), 225–237.

[9]. Giannotti, F., Lakshmanan, L. V., Monreale, A., Pedreschi, D. and Wang, H. (2012), 'Privacy-preserving mining of association rules from outsourced transaction databases', IEEE Systems Journal 7(3), 385–395.

[10]. Backstrom, L., Dwork, C. and Kleinberg, J. (2007), Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography, in 'Proceedings of the 16th international conference on World Wide Web', ACM, pp. 181–190.

[11]. Chase, M. and Chow, S. S. (2009), Improving privacy and security in multi-authority attribute-based encryption, in 'Proceedings of the 16th ACM conference on Computer and communications security', ACM, pp. 121–130.

[12]. Prasadu Peddi (2021), "Deeper Image Segmentation using Lloyd's Algorithm", ISSN: 2366-1313, Vol 5, issue 2, pp:22-34.

[13]. Huai, M., Huang, L., Yang, W., Li, L. and Qi, M. (2015), Privacy-preserving naive bayes classification, in 'International conference on knowledge science, engineering and management', Springer, pp. 627–638.

[14]. Krumholz, H. M. (2014), 'Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system', Health Affairs 33(7), 1163–1170. URL: <https://doi.org/10.1377/hlthaff.2014.0053>

[15]. Prasadu Peddi (2019), Data Pull out and facts unearthing in biological Databases, International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

[16]. Lee, S. X., Leemaqz, K. L. and McLachlan, G. J. (2019), 'Ppem: Privacy-preserving em

learning for mixture models', Concurrency and Computation: Practice and Experience p. e5208.

