International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Novel Approach for Location Privacy Preservation of Clusters

Dr. Rashmi Amardeep [1], Nayan Prakash [2], Rishu Raj [3], Riya Gothi[4]

Sahil Singh [5]

dr.rashmi-is@dsatm.edu.in [1]
nayanprakash2001@gmail.com [2], rishurj1999@gmail.com [3],gothi.riya@gmail.com [4]
sahilsingh0903@gmail.com [5]

Faculty, Department of Information Science and Engineering, DSATM, Bangalore-88, Karnataka [1]

Student, Department of Information Science and Engineering, DSATM, Bangalore-88, Karnataka [2][3][4] [5]

Abstract - This abstract describes a novel approach to location privacy preservation algorithm that aims to protect users' sensitive location information from being exposed to unauthorized parties or malicious attacks. The proposed approach uses a combination of techniques, including spatial cloaking, k-anonymity, and differential privacy, to provide a multi-layered defense mechanism. Spatial cloaking involves grouping multiple users' location data within a particular area, while k- anonymity ensures that each location dataset is indistinguishable from at least k-1 other location datasets. Differential privacy involves adding random noise to the location data to protect the privacy of individual users. The proposed approach provides a robust defense mechanism that ensures the protection of user privacy while still allowing for the analysis of the overall location dataset. The effectiveness of the proposed approach is demonstrated through experimental evaluation, which shows that it provides a high level of privacy protection while maintaining the accuracy of location-based services.

I. INTRODUCTION

Location privacy preservation algorithms aim to protect users' sensitive location information from being exposed to unauthorized parties or malicious attacks. A novel approach to location privacy preservation algorithm involves using a combination of techniques such as spatial cloaking,k-anonymity, and differential privacy.

Spatial cloaking involves grouping multiple users' location data within a particular area to prevent an individual's precise location from being exposed. The area's size is determined by a predefined privacy threshold, which can be adjusted based on the sensitivity of the location information.

K-anonymity is another technique used to preserve location privacy by ensuring that each location dataset is indistinguishable from at least k-1 other location datasets. This means that an attacker cannot identify a particular user's location information from a group of k users.

Differential privacy involves adding random noise to the location data, making it difficult for attackers to distinguish between real and fake data. This technique ensures that the privacy of the individual user is protected while still allowing for the analysis of the overall dataset.

Overall, this novel approach to location privacy preservation algorithm provides a multi-layered defense mechanism that ensures the protection of user privacy while still allowing for the analysis of the overall location dataset.

II. IMPLEMENTATION

A. Mandarin

Mondrian is a data anonymization technique used to protect sensitive information in datasets. Similarly, Mondrian anonymization divides the dataset into a grid-like structure of rectangular partitions, each of which corresponds to a group of records that share common characteristics.

The main idea behind Mondrian is to hide sensitive information by generalizing certain attributes of the records within each partition. The attributes are generalized in a way that preserves the utility of the dataset, meaning that the anonymized data can still be used for analysis and research purposes.

Mondrian works by first identifying the sensitive attributes in the dataset that need to be protected. These can be personal identifying information (PII) such as names, addresses, and social security numbers, or other sensitive data such as medicalrecords or financial information. The anonymization process then proceeds as follows:

1. Divide the dataset into partitions: Mondrian partitions the dataset into a grid-like structure of rectangular partitions, each of which corresponds to a group of records that share common characteristics. The partitions are created based on the values of the non-sensitive attributes in the dataset.

2. Generalize the sensitive attributes: For each partition, Mondrian generalizes the sensitive attributes in a way that preserves the utility of the dataset. Generalization involves replacing specific values with more general values. For example, a specific age value of "32" could be generalized to a range of "30-40".

3. Ensure consistency: The generalization process must be consistent across all records within each partition to avoid inconsistencies in the anonymized data.

4. Evaluate the level of anonymity: The level of anonymity achieved by Mondrian depends on the size of the partitions and the degree of generalization applied to the sensitive attributes. The goal is to achieve a high degree of anonymity while preserving the utility of the data.

B. Classic mandarin

Classic Mondrian is a technique used for achieving k-anonymity in datasets.

The basic idea behind Mondrian is to partition the dataset recursively along its dimensions until k-anonymity is achieved for each resulting partition. The partitions are created by choosing a dimension with high information loss potential and a split value such that the split maximizes information gain while preserving k-anonymity.

The partitioning process can be summarized as follows:

1. Choose a dimension to partition that has high information loss potential.

2. Choose a split value that maximizes information gain while preserving k-anonymity.

3. Partition the data based on the chosen dimension and split value.

4. Repeat steps 1-3 for each resulting partition until k-anonymity is achieved for each partition or no further partitioning is possible.

For example, consider a dataset of individuals with attributes such as age, gender, income, and occupation. To achieve 3-anonymity, Mondrian partitions the dataset into subgroups such that each subgroup has at least three individuals with the same values for the chosen attributes. To do so, it might choose the "age" attribute as the dimension to partition and choose a split value such that it divides the dataset into two subgroups of similar age ranges while preserving the k-anonymity requirement.

The resulting partitions can then be used to anonymize the dataset by replacing the original attribute values with the generalizations of the attributes for each partition. For example, instead of showing the exact age of each individual, we show the age range of the partition that they belong to.

Mondrian is an effective technique for achieving k-anonymity, but it can be computationally expensive and requires careful tuning of the partitioning parameters.

C. DataFly

DataFly is an extension of the Mondrian algorithm for data anonymization, which aims to reduce the information loss caused by the generalization of sensitive attributes. It is designed to anonymize relational databases with complex structures, such as star and snowflake schemas.

The DataFly algorithm is based on the concept of a "fly," which is a subset of a database schema that consists of a fact table and all of its related dimension tables. A fly is anonymized by applying the Mondrian algorithm recursively to its fact table and dimension tables.

The DataFly algorithm works as follows:

1. Partition the database schema into non-overlapping flies.

2. Anonymize each fly independently using the Mondrian algorithm.

3. Merge the anonymized flies back into the original schema.

The key innovation of DataFly is that it maintains consistency between the different flies during the anonymization process. This is achieved by maintaining a mapping between the original and anonymized values of the dimension attributes, and propagating the changes in this mapping across the different flies.

DataFly can also handle numerical attributes by splitting them into ranges and treating each range as a separate categorical attribute. This helps to preserve the utility of the data while ensuring privacy.

One of the advantages of DataFly over other anonymization techniques is that it can handle complex relational structures with minimal information loss. However, it is computationally expensive and may not be suitable for large databases.

III. METHODOLOGY

Mondrian_ldiv is an extension of the classic Mondrian algorithm that enhances its privacy protection capabilities. It stands for "Mondrian with Local Differential Privacy".

Like the classic Mondrian algorithm, Mondrian_ldiv is a data anonymization technique that aims to protect sensitive information in a dataset by dividing it into partitions called "cells". Each cell represents a group of records that share the same values for a subset of attributes. The goal is to partition the dataset in a way that ensures the privacy of individuals by hiding sensitive information.

The Mondrian_ldiv algorithm adds an extra layer of privacy protection to the classic Mondrian algorithm by applying local differential privacy. This means that instead of simply partitioning the dataset, Mondrian_ldiv perturbs the data by adding random noise to each attribute value, in order to make it more difficult to identify individuals in the dataset.

The amount of noise added to each attribute is determined by a parameter called "privacy budget", which represents the maximum amount of privacy that can be sacrificed in order to achieve a desired level of accuracy. The higher the privacy budget, the more noise is added to the data, and the more privacy is protected. However, adding too much noise can also reduce the accuracy of the data, making it less useful for analysis.

To apply Mondrian_ldiv, the dataset is first partitioned into cells using the classic Mondrian algorithm. Then, for each cell, the algorithm adds random noise to each attribute value, such that the amount of noise is proportional to the sensitivity of the attribute (i.e., the maximum difference in the attribute values between adjacent records). The noise is drawn from a Laplace distribution, which is a probability distribution that adds noise in a waythat preserves differential privacy.

Mondrian_ldiv has been shown to be effective at protecting privacy while preserving the accuracy of the data, especially for high-dimensional datasets with many sensitive attributes. However, it can be computationally expensive, especially for large datasets and small privacy budgets.

Data Collection:
We have created a large random data set which contains the private parameters which is to be preserved.

Data preprocessing:
The sensitive data is preprocessed to remove any irrelevant or redundant information.

Encoding:
The data is encoded using deep learning methodssuch as autoencoders or variational autoencoders, which can learn the underlying distribution of the data and generate a compact representation of the data.

In the context of data analysis and machine learning, a cluster refers to a group of data points that share similar characteristics or attributes. The process of grouping these data points together is called clustering or cluster analysis.

The goal of clustering is to find structure in the data, to discover patterns or groups that are not initially apparent. By identifying clusters of similar data points, clustering can help in identifying relationships among the data and in making predictions about future data.

There are different approaches to clustering, depending on the type of data and the problem being addressed. Some of the most commonmethods include:

1. Hierarchical clustering: This method starts with each data point as its own cluster, then merges them into larger clusters based on their similarity. The result is a tree-like diagram, called a dendrogram, that shows the relationships between the clusters.

2. K-means clustering: This is a popular method for partitioning data into k clusters, where k is a pre-specified number. The algorithm works by randomly assigning each data point to a cluster, then iteratively refining the clusters until convergence.

3. Density-based clustering: This method identifies clusters as regions of high density in the data, separated by regions of lower density.

Clustering can be used in a variety of applications, such as customer segmentation, image analysis, anomaly detection, and more. However, it is important to keep in mind that clustering is an unsupervised learning technique, meaning that it does not rely on labeled data and may not always produce the most accurate or useful results.

IV. APPLICATION

The novel approach to location privacy preservation algorithm proposed in this study can be applied in various applications that require the collection and analysis of location-based data while protecting users' privacy. Here are some examples of its applications:

Transportation: The proposed approach can be applied in transportation systems, such as ride- hailing services or public transportation systems, to protect users' location information while still providing efficient transportation services.

Healthcare: Healthcare providers can use the proposed approach to protect patients' location information while still analyzing data to provide effective healthcare services.

Social networking: Social networking platforms can use the proposed approach to protect users' location information and provide a secure platform for users to interact with each other.

Environmental monitoring: The proposed approach can be applied in environmental monitoring systems to protect users' location information while collecting and analyzing data to monitor environmental conditions.

Emergency response: Emergency response systems can use the proposed approach to protect users' location information during emergency situations while still providing effective emergency response services.

V. CONCLUSION

In conclusion, the novel approach to location privacy preservation algorithm proposed in this study provides a robust defense mechanism against unauthorized access or malicious attacks on users' sensitive location information. The combination of spatial cloaking, k-anonymity, and differential privacy techniques ensures the protection of individual user privacy while allowing for the analysis of the overall location dataset. The proposed approach has been shown to provide a high level of privacy protection while maintaining the accuracy of location-based services through experimental evaluation. This approach can be applied in various location-based services, such as transportation, social networking, and healthcare, to ensure the protection of user privacy while still providing the necessary services. Further research can explore the potential of combining other privacy preservation techniques to enhance the effectiveness of the proposed approach.

VI. REFERENCES

[1] S,Dr.G.Usha."TSDLA:Algorithm for Location Privacy in Clustered LB-MCS network".Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020).

[2] M. Ghaffari, N. Ghadiri, M. H. Manshaei, and M. S. Lahijani, ‘‘P4QS: A peer-to-peer privacy preserving query service for location-based mobile applications,’’ IEEE Trans. Veh. Technol., vol. 66, no. 10, pp. 9458–9469, Oct. 2017.

[3] M. Gruteser and D. Grunwald, ‘‘Anonymous usage of location-based services through spatial and temporal cloaking,’’ in Proc. 1st Int. Conf. Mobile Syst. Appl. Services, 2003, pp. 31–42.

[4] B. Gedik and L. Liu, ‘‘Protecting location privacy with personalized k-anonymity:

Architecture and algorithms,'' IEEE Trans. Mobile Comput., vol. 7, no. 1, pp. 1–18, Jan. 2008.

[5] S. Gang, S. Liangjun, L. Dan, Y. Hongfang, and C. Victor, ''Towards privacy preservation for 'check-in' services in locationbased social networks,'' Inf. Sci., vol. 481, pp. 616–634, May 2019.

[6] J. Shao, R. Lu, and X. Lin, ''FINE: A fine- grained privacy-preserving location-based service framework for mobile devices,'' in Proc. IEEE INFOCOM, Apr./May 2014, pp. 244–252.

[7] X. Zhao, H. Gao, L. Li, H. Liu, and G. Xue, ''An efficient privacy preserving location based service system,'' in Proc. IEEE GLOBECOM, Dec. 2014, pp. 576–581.

[8] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, ''Achieving k-anonymity in privacy-aware locationbased services,'' in Proc. IEEE INFOCOM, Apr./May 2014, pp. 754–762.

[9] B. Niu, Z. Zhang, X. Li, and H. Li, ''Privacy- area aware dummy generation algorithms for locationbased services,'' in Proc. IEEE ICC, Jun. 2014, pp. 957–962.

[10] Latanya Sweeney. k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 10:557–570, October 2002.

[11] Arvind Narayanan, Narendran Thiagarajan, Michael Hamburg, Mugdha Lakhani, and Dan Boneh. Location privacy via private proximity testing. NDSS'10, 2011.

[12] Gabriel Ghinita, Keliang Zhao, Dimitris Papadias, and Panos Kalnis. A reciprocal framework for spatial k-anonymity. Inf. Syst., 35:299–314, May 2010.

[13] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for kanonymous location privacy in participatory sensing," in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 2399–2407.

[14] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k-anonymity in privacy-aware location-based services," Proceedings - IEEE INFOCOM, pp. 754–762, 2014.

[15] Y. Zhang, W. Tong, and S. Zhong, "On designing satisfaction-ratioaware truthful incentive mechanisms for k-anonymity location privacy," IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2528– 2541,2016.