ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Probability Framework For Improving Efficiency And Security Of Data By Implementation Of Crowdsourcing Techniques

Kornu Ramalaxmi
M.Tech,
Department of CSE,
Sarada Institute of Science Technology & Management, Srikakulam.

Madina Jayanthi Rao
HOD,
Department of CSE,
Sarada Institute of Science Technology & Management, Srikakulam.

Abstract:

Crowdsourcing is a type of participative activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. Organizations share their data about customers for exploring potential business avenues. The sharing of data has posed several threats leading to individual identification. Owing to this, privacy preserving data publication has become an important research problem. The main goals of this problem are to preserve privacy of individuals while revealing useful information. To integrate the crowdsourcing techniques into the database engine, we must address the privacy concern, as each crowdsourcing job requires us to publish some sensitive data to the anonymous human workers.

In this paper, we study how to guarantee the data privacy in the crowdsourcing scenario. To provide data privacy we are using cryptography technique for the protection crowdsorucing database. Before provide privacy of data we are generate key for the group of members. To generate group key we are using group key management protocol. Before store and retrieve data we can encrypt and decrypt data using Block cipher Encryption Algorithm . By implementing this project we can provide more efficiency and security of data.

Keywords:

Data security, crowdsourcing, k-anonymity, 1-diversity, generalization, bucketization.

Introduction:

Today, crowdsourcing has transferred mainly to the Internet. The Internet provides a particularly good venue for crowdsourcing since individuals tend to be more open in web-based projects where they are not being physically judged or scrutinized and thus can feel more comfortable sharing. This ultimately allows for well-designed artistic projects because individuals are less conscious, or maybe even less aware, of scrutiny towards their work. In an online atmosphere, more attention can be given to the specific needs of a project, rather than spending as much time in communication with other individuals.The crowdsourced problem can be huge or very small. Some examples of successful crowdsourcing themes are problems that bug people, things that make people feel good about themselves, projects that tap into niche knowledge of proud experts, subjects that people find sympathetic or any form of injustice.

Crowdsourcing can either take an explicit or an implicit route. Explicit crowdsourcing lets users work together to evaluate, share and build different specific tasks, while implicit crowdsourcing means that users solve a problem as a side effect of something else they are doing.With explicit crowdsourcing, users can evaluate particular items like books or webpages, or share by posting products or items. Users can also build artifacts by providing information and editing other people's work. Implicit crowdsourcing can take two forms: standalone and piggyback. Standalone allows people to solve problems as a side effect of the task they are actually doing, whereas piggyback takes users' information from a third-party website to gather information.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The concept of privacy-preservation for sensitive data can require the enforcement of privacy policies or the protection against identity disclosure by satisfying some privacy requirements. In this paper, we investigate privacy-preservation from the anonymity aspect. The sensitive information, even after the removal of identifying attributes, is still susceptible to linking attacks by the authorized users. This problem has been studied extensively in the area of micro data publishing and privacy definitions, e.g., k-anonymity, l-diversity, and variance diversity. Anonymization algorithms use suppression and generalization of records to satisfy privacy requirements with minimal distortion of micro data. The anonymity techniques can be used with an access control mechanism to ensure both security and privacy of the sensitive information. The privacy is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an access control policy.

Related Work:

Based on the classic k-anonymity model, a number of privacy models have been proposed for data [2]. In this paper[3], proposing a comprehensive trajectory privacy technique and combine ambient conditions to cloak location information based on the user privacy profile to avoid a malicious LBS reconstructing a user trajectory. First an r-anonymity mechanism which preprocesses a set of similar trajectories R to blur the actual trajectory of a service user is proposed, then combine k-anonymity with s road segments to protect the user's privacy.Introducing a novel time-obfuscated technique which breaks the sequence of the query issuing time for a service user to confuse the LBS so it does not know the user trajectory, by sending a query randomly from a set of locations residing at the different trajectories in R. propose an algorithm, Seed-and-Grow in [4], to identify users from ananonymized social graph, based solely on graph structure. The algorithm first identifies a seed subgraph, either planted by an attackeror divulged by a collusion of a small group of users, and then grows the seed larger based on the attacker's existing knowledge of theusers' social relations. Presenting Anonimos in [5], a LinearProgramming-based technique for anonymization of edge weights that preserves linear properties of graphs. The paper [6] defines a proper distance metric to achieve local recording generalization with small distortion, controlling the inconsistency of attribute domains. In [7] proposing two categories of novel anonymization methodsfor sparse highdimensional data. The first category is based on approximate nearest-neighbor (NN) search in high-dimensionalspaces, which is efficiently performed through locality-sensitive hashing (LSH). In the second category, proposing two datatransformations that capture the correlation in the underlying data: 1) reduction to a band matrix and 2) Gray encoding-based sorting.A new notion of privacy called "closeness" is proposed in [8]. First present the base model t closeness, which requires that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table (i.e., the distance between the two distributions should be no more than a threshold t). Then proposing a more flexible privacy model called (n,t)-closeness that offers higher utility. In this paper [2], they proposed a new method for achieving kanonymity named K-anonymity of Classification Trees Using Suppression (kACTUS). In kACTUS, efficient multidimensional suppression is performed, i.e., values are suppressed only on certain records depending on other attribute values, without the need for manually produced domain hierarchy trees. Thus, in kACTUS, identify attributes that have less influence on the classification of the data records and suppress them if needed in order to comply withkanonymity. [9] expands the scope of perturbationbased PPDM to Multilevel Trust (MLT-PPDM). In this setting, the more trusted a data miner is, the less perturbed copy ofthe data it can access.Under this setting, a malicious data miner may have access to differently perturbed copies of the same datathrough various means, and may combine these diverse copies to jointly infer additional information about the original data that thedata owner does not intend to release. Preventing such diversity attacks is the key challenge of providing MLT-PPDM services. [10] Presents a heuristic algorithm for generating releases satisfying kw-structural diversity anonymity so that the adversary cannot utilize his knowledge to reidentify the victim and take advantages.

Existing Systems:

In the existing technique does not provide privacy of crowd sourcing database. So that the anonymous human workers are easily access the crowd sourcing database. The existing technique does not provide more for crowd sourcing database. So that so many technique are not providing security and more efficiency for crowd sourcing database. So that to overcome those problem we implement proposed system.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Disadvantages of existing system:

1. K-Anonymity affects the performance of crowd sourcing, as generalization and grouping leads to information loss. If we provide the anonymized data to the human workers, they may fail to return the correct answer. The system needs to address the tradeoff between the privacy and accuracy.

2. Because k-anonymization does not include any randomization, attackers can still make inferences about data sets that may harm individuals.

3. K-anonymization is not a good method to anonymize high-dimensional datasets.[7] For example, researchers showed that, given 4 points, the unicity of mobile phone datasets (, k-anonymity when) can be as high as 95%.

4. The process of k-anonymization reduces the effectiveness of data mining algorithms on the anonymized data and renders privacy preservation impractical.

Proposed System:

Now a day's most of the techniques are to provide privacy for crowd sourcing database. So that we must concern to provide privacy for sensitive data to the anonymous human workers. In this we are mainly discussing two concept i.e how to generate group key and cryptography technique for provide security for that data. In the generation of group key we are using group key management protocol and encryption and decryption of data we are Block cipher Encryption Algorithm. Before encrypt and decrypt the data the users will generate group key. The process generating group key is as follows.

User Authentication Schema and Key generation:

1.Each user will send request to group key manager.
2.The group key manger will send response as id of users.
3.Each user will send the nonce to group key manger. The nonce (Ni) value will be generating randomly.
4.After retrieving all nonce of users the group key manager will generate a point(Xi,Yi) for each user based on random challenge. After generating points the group key manager will send to individual users.
5. Each user will retrieve the point he/she will generate share point (Xi ,Yi ^Ni)using the point and random challenge. The generation share point each user will send to group key manager.
6.Using share point the group key manager will generate signature using hash function. Those signatures are sent individual user.
7.After retrieving the signatures each user again generate signature and compare both are equal those users verified user.
8.After completion of verification process the group key manager will secret key.
9.After generating secret key the group key manager will divide secret with n number of parts. Where any subset of parts will reconstruct the polynomial function. Before generating polynomial function the group key manager will choose random number for generation of polynomial function. The polynomial function is given below.

$$F(x)=secret+bx+ax2$$

10.After that the group key manger will send subset parts to individual user.
11.The users will retrieve those parts and again generate polynomial function and get same secret of each user.

The completion of authentication and secret key each user will store the database into database engine. Before storing data into database engine each user will encrypt the data using block cipher encryption algorithm. The procedure block cipher encryption algorithm as follows. Block cipher encryption algorithm is a 64-bit symmetric block cipher with variable length key. The algorithm operates with two parts:

i)key expansion part
ii)data encryption part.

The role of key expansion part is to converts a key of at most 448 bits into several sub key arrays totaling 4168 bytes. The data encryption occurs via a 16-round Feistel network . It is only suitable for application where the key does not change often, like communications link or an automatic file encryption. It is significantly faster than most encryption algorithms when implemented on 32-bit microprocessors with large data caches .The nature of encryption algorithms is that, once any significant amount of security analysis is done, it is very undesirable to change the algorithm for performance reasons, thereby invalidating the results of the analysis.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Thus, it is imperative to consider both security and performance together during the design phase. While it is impossible to take all future computer architectures into consideration, an understanding of general optimization guidelines, combined with exploratory software implementation on existing architectures to calibrate performance, should help achieve higher speed in future encryption algorithms.

Sub key Expansion:

Block cipher encryption algorithm uses a large number of subkeys. These keys must be pre computed before any data encryption or decryption. The P-array consists of 18 32-bit subkeys: P1, P2,..., P18. There are four 32-bit S-boxes with 256 entries each:

S1,0, S1,1,..., S1,255;
S2,0, S2,1,..,, S2,255;
S3,0, S3,1,..., S3,255;
S4,0, S4,1,..,, S4,255.

Pseudo Code of Blowfish Algorithm:

begin itemize Block cipher encryption algorithm has 16 rounds. The input is a 64-bit data element, x.
Divide x into two 32-bit halves: xL, xR.
Then, for i = 1 to 16: xL = xL XOR Pi xR = F(xL) XOR xR Swap xL and xR
After the sixteenth round, swap xL and xR again to undo the last swap. Then, xR = xR XOR P17 and xL = xL XOR P18. Finally, recombine xL and xR to get the ciphertext.

Decryption is exactly the same as encryption, except that P1, P2,..., P18 are used in the reverse order. Implementations of Blowfish that require the fastest speeds should unroll the loop and ensure that all subkeys are stored in cache.

Advantages of Proposed System:

1.To generalize the records to enforce the privacy requirement as well as maximize the utility of crowd sourcing, we show that our problem is consistent with previous K-Anonymity approaches, i.e. which target at minimizing the information loss.
2.Our scheme thus exploits the results to the synthetic samples from the crowd, during the optimization process when anonymizing the real tuples in the database.
3.Anonymization algorithms based on data mining techniques may apply different generalizations for several groups of tuples rather than the same generalization for all tuples. In this way, it may be possible to retain more useful information.
4.This model will be particularly useful when the anonymity constraints are embedded within the data mining process, so that the data mining algorithm has access to the non-anonymized data.

Conclusion:

In this paper we are proposed concepts of group key management and block cipher encryption algorithm for provide privacy of crowd sourcing database. So that using group key management protocol for generating group key of individual users. After generating group key each user will encrypt the data using block cipher encryption algorithm. After encrypting that data can be stored into data base. If any users retrieve that and decrypt that data using secret get the original data. By implementing those concepts we can provide more efficiency and more privacy of transferring data.

REFERENCES:

[1]Sai Wu, Xiaoli Wang, Shen Wang, Zhenjie Zhang and Anthony K.H. Tung, "K-Anonymity for crowdsourcing database" 2013.

[2] Slava Kisilevich, Lior Rokach, Yuval Elovici, Member, IEEE, and Bracha Shapira, "Efficient Multidimensional Suppression for KAnonymity," IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 3, March 2010.

[3] Ren-Hung Hwang, Yu-Ling Hsueh, And HaoWei Chung, "A Novel Time-Obfuscated Algorithm For Trajectory Privacy Protection", IEEE Transactions On Services Computing, Vol. 7, No. 2, April-June 2014.

[4] Wei Peng, Student Member, IEEE, Feng Li, Member, IEEE, Xukai Zou, Member, IEEE, and Jie Wu, Fellow, IEEE," A Two-Stage Deanonymization Attack against Anonymized Social Networks", IEEE Transactions On Computers, Vol. 63, No. 2, February 2014

[5] Sudipto Das, omer Egecioglu, and Amr El Abbadi, Senior Member IEEE,"Anonimos: An LP-Based Approach

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

for Anonymizing Weighted Social Network Graphs", IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 4, April 2012.

[6] Jiuyong Li, Member, IEEE, Raymond ChiWing Wong, Student Member, IEEE, Ada WaiChee Fu, Member, IEEE, and Jian Pei, Senior Member, IEEE, "Anonymization by Local Recoding in Data with Attribute Hierarchical Taxonomies," IEEE Transactions On Knowledge And Data Engineering, Vol. 20, No. 9, September 2008.

[7] Gabriel Ghinita, Member, IEEE, Panos Kalnis, and Yufei Tao," Anonymous Publication of Sensitive Transactional Data" IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 2, February 2011.

[8] Ninghui Li, Member, IEEE, Tiancheng Li, and Suresh Venkatasubramanian," Closeness: A New Privacy Measure for Data Publishing" IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No. 7, July 2010.

[9] Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang," Enabling Multilevel Trust in Privacy Preserving Data Mining" IEEE transactions on knowledge and data engineering, vol. 24, no. 9, september 2012.

[10] Chih-Hua Tai, Peng-Jui Tseng, Philip S. Yu, Fellow, IEEE, and Ming-Syan Chen, Fellow, IEEE,"Identity Protection in Sequential Releases of Dynamic Networks",IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 3, March 2014.

[2]R. J. B. Jr. and R. Agrawal, "Data privacy through optimal kanonymization", in ICDE, 2005.

[11]Sweeney.L, "Achieving k-anonymity for privacy protection using generalization and suppression," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, 2002.

[12]Machanavajjhala.A,Kifer.D,Venkitasubramaniam.M, Gehrke.J "lDiversity: Privacy Beyond k-Anonymity," ACM Transactions Knowledge Discovery Data, volume 1, issue 1, March 2007.

[13]Feng.A, Franklin.M, Kossmann.D, Kraska.T, Madden.S, Ramesh.S, Wang.A, and Xin.R, "CrowdDB: Query Processing with the VLDB Crowd," PVLDB, volume 4, issue 12, pp. 1387-1390, 2011.

[14] Li. T and Li. N, "On the Tradeoff between Privacy and Utility in Data Publishing," In KDD, pp. 517-526, 2009.

[15] K.Chandrasekhar & N.Balakrishna, Privacy Preserving Collaborative Data Publishing Using K-Anonymity, IJMETMR, http://www.ijmetmr.com/oldecember2014/KChandrasekhar-NBalakrishna-125.pdf, Volume No: 1(2014), Issue No: 12 (December) .

[16] Kiruthika.S and MohamedRaseen.M, "Suppression Slicing–using ldiversity," IJCA Proceedings on Amrita International Conference of Women in Computing, pp. 1-6, January 2013.

[17] Sorokin.A and Forsyth.D, "Utility data annotation with Amazon Mechanical Turk," in First IEEE Workshop on Internet Vision at CVPR, 2008.

[18] Marcus.A, Wu.E, Madden.S, and Miller.R.C, "Crowdsourced Databases: Query Processing with People," in CIDR, pp. 211-214, 2011.

Author's Details:

Kornu Ramalaxmi, is student in M.Tech(CSE) in Sarada Institute of Science Technology and Management, Srikakulam. He has received his B.Tech(C.S.E) from SARADA INSTITUTE OF SCIENCE TECHNOLOGY AND MANAGEMENT,Srikakulam. she is interesting areas are datamining and datawarehouse.

Madina Jayanthi Rao, working as a HOD of CSE inSarada Institute of Science, Technology and Management (SISTAM), Srikakulam, Andhra Pradesh. He is pursuing Ph.d at KRISHNA UNIVERSITY Machilipatnam in computer science. He received his M.Tech (CSE) from Aditya Institute of Technology And Management (AITAM), Tekkali. Andhra Pradesh. His interest research areas are Data mining, Image Processing, Computer Networks, Distributed Systems. He published 12 international journals and he was attended number of conferences and workshops.