International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 1 Issue 5 ǁ Sep. 2013 ǁ PP.51-60

A Study On Performance Analysis Of Privacy Preservation Data Mining Techniques For Versatile Digital Data Publishing

Kesavaraj. G[1], Dr. Sukumaran. S[2]
[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelvelli,Tamil Nadu,INDIA
[2]Associate professor, Department of Computer Science, Erode Arts and Science College, Erode,Tamil Nadu,INDIA

Abstract: *Many organizations need publishing micro data for personal information sharing and assisting the research work in favor of helping the public interests. Despite the contrary, privacy is concerned for the individual vendor as publications of tuples. Privacy Preserving Data Publishing (PPDP) is addressed to produce the published tuple table. Privacy tuple table enables examining tasks such as collective query answering, data mining over the published data, and finally protecting the privacy of distinct data vendor. The three kinds of attributes namely Personal Identifiable Attributes (PIA), Quasi Identifier (QI) and Sensitive Attributes (SA) are available in micro data table.*

An Algorithm-Safe Publishing (ASP) model is designed with PPDP algorithm in the form of disclosure to remove threats. The eliminate algorithm in the form of disclosure remove threats and it is used in data publishing algorithms. Two generic tools worst-case eligibility test and stratified pick-up is used for correcting the design. The effectiveness of ASP tools is to convert the two accepted ℓ-diversity algorithms namely Mondrian 1 and Hilb. The Mondrian 1 and Hilb are converted to SP-Mondrian and SPHilb safer algorithm. SP-Mondrian and SP-Hilb efficiency is measured in terms of information utility and efficiency.

Privacy requirements are particularized as a subjective set of privacy rules over characteristics in the micro data table by designing an adaptable publishing. The adaptable publishing is enabled with the Guardian Normal Form (GNF) method for declaring various sub-tables. Each sub-table is destroyed by QI and SA publishing algorithm. The aggregation of all published tables promises all isolation policy. Guardian Decomposition (GD) method and Utility Aware Decomposition (UAD) decompose a micro data table into GNF designed model.

Group based anonymization is the generally considered approach for privacy preserving data publishing. Privacy models using group based anonymization includes k-anonymity, l-diversity, and t-closeness. The group based anonymization methods in the form of bucket which fundamentally hide every distinct record after a group to preserve data privacy. If not correctly anonymized, patterns are derived from the published data and used by the opponent to break entity privacy.

Data publishing has generated much concern on entity privacy. The privacy threat from the Full Functional Dependency (FFD) is analyzed for adversary knowledge. The cross-attribute correlation by FFDs brings potential vulnerability. The privacy model, (d, ℓ) inference is formalized to combat the FFD based privacy attack. Robust algorithms efficiently anonymize the micro data with low information loss when the unsafe FFD present.

Keywords : *Privacy Preserving Data Publishing, Full Functional Dependency (FFD), Guardian Normal Form (GNF)*

I. INTRODUCTION

Recent years observed huge quantity of data to be collected on a large scale. Induced by common benefits and regulation requires certain data to be published. There is a demand for publishing collected data to the public. Even though it offers convincing merits for ad-hoc analysis in a collection of area such as public population studies and health, it directly violates individual privacy. Violation of individual data takes place particularly when the data is of microdata format. Micro data contain elaborated self specific data in its genuine structure.

In recent years, publishing microdata with privacy preserving gained attention. Quietly detaching unambiguous identifiers (IDs) e.g., name and SSN, from the unconfined data is lacking to keep privacy. The presence of a position of non-ID attributes, called quasi-identifiers, differently recognizes individuals (e.g., the sequence of zipcode, date of birth and gender). Quasi-identifier acquires information from varied extrinsic sources to re-identify the individuals in the unconfined data. The initial principle, k-anonymity needs every identical record from at least k−1 additional records with recognition to their quasi-identifiers. An enhanced

principle named ℓ-diversity further desires each group of identical tuples should contain at least l separate sensitive principles.

Published table, an outcome of a data publishing algorithm concludes whether an isolation model is accurately satisfied or not. The privacy disclosure designs a data publishing algorithm which is opposed by Algorithm in the form of disclosure. Reject the algorithm in the form disclosure to method and useful safe algorithm for privacy-preserving data publishing. Worst-case eligibility test and stratified pick-up are the two methods accomplished to eliminate algorithm in the form of disclosure the design of data publishing algorithms. Elimination of algorithm in the form of disclosure leads a high level of utility saved for the published table. Algorithm-Safe data Publishing (ASP) is introduced to formalizes algorithm in the form of disclosure elimination. Bucketized data are published using versatile publishing scheme.

PPDP enforces a single privacy rule QI -> SA. An opponent with information of QI misunderstand the SA of a tuple if ℓ-diversity is pre-defined where QI and SA are two displace sets of attributes in the micro data table. Versatile publishing is designed to accurately model an authentic world isolation requirement. Versatile distributing is a structure which particularizes the privacy necessity of distributing a micro data table as a subjective set of isolation rules. Versatile distributing is the absolute utilization of the particular tables multiple SA distributing method. Bucketized dataset are published using versatile publishing scheme.

A primary method handled in privacy preserving data distributing is group in the form of anonymization, through which reports in the obsessed relative are segmented into groups. Each collection guarantees few possessions such as assortment so as to gratify the isolation obligation while preserving adequate information function. Privacy models structure, MASK, injector and t-closeness endure from severe isolation breaches. The major problem is the usefulness that is kept in the anonymized table benefits the opponent to breach individual isolation. Group based anonymization is forming groups of QI with similar tuples avoiding privacy breaches.

The quasi-identifiers are combined with information acquired from varied exterior source to re-identify the persons in the unconfined data which is called verification association attack. Generalization is an accepted approach to attain both k-anonymity and ℓ-diversity. In specific, the microdata are separated into anonymization group. For the tuples in the identical collection, their quasi-identifier values are established to be same. Quasi-identifier (QI) is interchangeable to each other with respect to their QI principles.

A microdata with QI, each tuple is contained in a cluster containing at least three dissimilar sensitive values. The aggressor possibly obtains the efficient dependencies from each source. It is essential to establish vigorous privacy principle and anonymization algorithm for the micro data that holds full practical dependencies. The (d,ℓ) suggestion model is designed to protect against the FFD in the form of attack with adequate anonymization algorithms. Finally re-identifying an individual sensitive value by the attacker is impossible.

II. LITERATURE REVIEW

A toolset developed for eliminating algorithm in the form of disclosure from existing privacy preserving data publishing algorithms. A generic tool for revising the design holds worst-case eligibility test and stratified pick-up [1]. The derived patterns form foreground data knowledge and the group based anonymization approach by bucketization basically hides each individual record behind a group to preserve data privacy[2]. The formalized FFD-based privacy attack and define the privacy model, (d,ℓ) -inference, to combat the FD-based attack[3]. The designed Versatile publishing scheme devise two algorithms, Guardian Decomposition (GD) and Utility-aware Decomposition (UAD), for decomposing a micro data table into GNF[4]. A method that combines micro aggregation and any synthetic data generator is formalized[5].

The derived data mining privacy by decomposition (DMPD) algorithm uses a genetic algorithm to investigate for optimal feature set partitioning[6]. The supports for complex classification methods are support vector machines, respecting mobile computing and communication constraints, and enabling user-determined privacy levels. Privacy preserving collaborative learning is a robust method that attempt to infer the original data or poison the model, and imposes minimal costs[7]. The presented several types of well-known data mining models deliver a comparable level of model quality over the geometrically perturbed dataset as over the original dataset[8]. The designed PPSCs that partition, encode, and compare strings yield highly accurate record linkage results[9]. A secret sharing allows the data to be divided into multiple shares and processed independently at different servers[10]. A framework with algorithms and mechanisms for privacy and security enhanced dynamic data collection, aggregation, and analysis with feedback loops [11].

III. METHODOLOGIES

The different works involved in "Performance Analysis of Group Based Anonymization by bucketization with Versatile Safe Publishing" are

IV. ALGORITHM-SAFE PUBLISHING (ASP)

The problem of algorithm in the form of disclosure from an algorithmic viewpoint is faced in Algorithm-safe publishing. Specifically, confront of identify algorithm in the form disclosure is illustrated by representative the space of disclosure is considerably better. Supporting a preliminary tool for difficult either an agreed data distributing algorithm might lead to algorithm based disclosure. Worst-case eligibility test and stratified pick-up method are developed to modify the plan of data publishing algorithms such that algorithm in the form of disclosure is excluded.

A. Discovering space of algorithm-based disclosure

The space of algorithm in the form of disclosure is much broader than already exposed. Earlier performance recognizes algorithm in the form of disclosure when an opponent controls exterior knowledge about the QI attributes. Algorithm-Safe Publishing finds the additional forms of external knowledge, such as the allocation of SA values. ASAP ignores the dependency of algorithm in the form of disclosure on exterior knowledge.

Algorithm-based disclosure happens at the occasion when the opponent holds no exterior knowledge about the published data. Minimality attack in privacy-preserving data distributing [MASK] eliminates the earlier discovered algorithm in the form of disclosure, but suffers from an additional type of algorithm in the form of disclosure which is solved by ASAP.

B. ASP is susceptible to Algorithm in the form of Disclosure

A preliminary tool is designed for verifying the known data publishing algorithm is susceptible to algorithm in the form disclosure or not. Algorithm- Safe data Publishing (ASP) model properly describes algorithm in the form of disclosure as the changes among two arbitrary worlds. The native one is each probable mapping between an original table and the published table is equally mapping which violates an adversary's external knowledge. The other smart one is mapping that follows the data publishing algorithm. An algorithm satisfies ASP as it continually conserves correspondence between these two worlds.

The susceptibility of several earlier data publishing algorithms is identified by two essential circumstances of ASP. An adequate criterion of ASP is to analyze the immunity of various other algorithms for algorithm in the form of disclosure. The simulatable publishing design paradigm refers agreeable condition. Apparently, simulatable publishing needs the published QI. The published QI is provisionally liberated of the unique SA given the unique QI and the published SA as prior information. In other words, no unpublished QI-SA correlation information is passed down for producing the available table. The union of these essential and adequate conditions forms an investigative tool. The preliminary tool investigates the agreed data publishing algorithm is susceptible to algorithm in the form of disclosure.

C. Worst-Case Eligibility Test and Stratified Pick-Up

Worst-case eligibility test and Stratified pick-up tools are outlined for revising the plan of algorithms to pursue the simulatable publishing example. The worst-case eligibility test tool is outlined to improve the mainly ordinary violation of ASP raise in earlier algorithms. Worst-case eligibility test make it stick to the simulatable publishing model. Worst-case eligibility test is applied to two well-known data publishing algorithms, Mondrian and Hilb, ℓ-diversity algorithms proving ASP efficiency.

The Stratified pick-up device is outlined to amend the usefulness of available data beyond violate the simulatable publishing example. Specifically, stratified pick-up benefits an anatomy-like method to reduce the figure of tuple in each available QI-group. Stratified pick-up is applied on top of the first device output of Mondrian and Hilb to create SP-Mondrian and SP-Hilb, respectively proving its efficiency. Stratified pick-up supports nearly equivalent more excellent usefulness than the original Mondrian and Hilb algorithms, correspondingly. SP-Mondrian and SP-Hilb are presenting in their corresponding versions of bucketization publishing scheme. The fairness of comparison, the SP-Mondrian and SP-Hilb are adapted into bucketization scheme compared to simulatable publishing algorithm. Bucketized data are published using algorithm safe publishing scheme.

V. VERSATILE PUBLISHING (VP)

Versatile publishing is an original structure which particularizes the privacy condition of announcing a micro data table as a random set of isolation rules. Versatile publishing captures the actual world obligation of enforcing numerous isolation rules over the distributing micro data table. Versatile publishing derives total set of conclusion axioms for isolation rules. Versatile publishing defines guardian normal form (GNF) which promises a set of several privacy rules over the group of several published tables. For decomposing a table into GNF, the value optimization is confirmed to be NP-hard, and prosper two heuristic algorithms GD and UAD. An inclusive set of test is attended over two real world datasets, a extensively used benchmark adult dataset and

the abovementioned Texas dataset. The effective result shows the advantage of GD and UAD over the multi-SA and solitary attribute publishing techniques.

A. Guardian Normal Form

Versatile publishing enables the production of various sub-tables with possible coinciding attributes. Each sub-table in multiple sub-tables only addresses one isolation rule and therefore processed by algorithms designed for the QI-SA structure. GNF is listening carefully on intend of available tables as each solitude rule rest in event over the compilation of all available tables.

It is valuable to comprehend the suggestion of distributing multiple tables on the enforcement of isolation rules. Due to the crossroads attack a privacy rule fulfilled by two anonymized tables independently as it's out of order by the mixture of both tables. Still an isolation rule denoted by none of the available tables certainly convinced due to the detachment of attributes, diagonally available tables.

GNF is a normal form for the design of published tables. GNF provides criterion for influential whether an isolation rule is fulfilled over a group of available tables. The basic concept of GNF in terms of two ways for an isolation rule to be noted is a remarkable case of non-reachability and a general case of the continuation of a guardian table.

B. Decomposition into GNF

Decomposition of a microdata table into GNF is done with definition of GNF. Since GNF promise the approval of all isolation rules, the attention here is on optimizing the value of available tables. In particular, the utilization of several available tables in applications such as data mining is discussed. The inflexibility of value optimization is established and developed a GD and UAD algorithms for determining on heuristic decomposition.

Algorithm GD follows a privacy rule which violates GNF. GD algorithm dissolves the sub-tables to deal with the isolation rule, and advance awaiting no more criminal isolation rule exists. UAD algorithm is advanced for leveraging the connection among utility optimization and the Min-Vertex Coloring difficulty. Min-Vertex Coloring problem indicates the problem of each vertex corresponds to an attribute in the microdata table. Bucketized data are published using versatile publishing scheme.

VI. BUCKETIZATION A GROUP BASED ANONYMIZATION TECHNIQUE (GBA)

A leading approach used in isolation preserving data publishing is bucketization a group in the form of anonymization. Bucketization records the agreed relation and devises data into collections where each cluster must guarantee several properties. The property such as variety convinces the privacy needs while conserving adequate data efficacy. There are numerous isolation models connected with group in the form of anonymization such as k-anonymity, ℓ-diversity, t-closeness, (k, e)-anonymity, Injector and m- confidentiality.

a. Foreground Knowledge Attack

The bucketized data is an uncertain data, and an opponent uncovers attractive patterns because the available data must preserve high data efficacy. The uncovered patterns are the foreground knowledge. In contrast the background knowledge needs much opponent attempt to get hold of data from anywhere exterior from the table. It is not difficult to get hold of the foreground information from the anonymized dataset; most approaches experience from isolation breaches

b. Generalization-Based Anonymization

After the anonymization, the anonymized data set are published. Data set consists of a deposit of QI-groups, where each QI-group is a group of tuples associated with a multi set of susceptible values. Counting on the anonymization mechanism, each QI-group agrees to either set a QI values. An attribute is included for the ID of the QI-group. QI-group is referred by its ID. Such bucketization is usually adopted in the literature of information publishing including k-anonymity, l-diversity, t-closeness and a huge amount of other isolation models. The connection among separate records and the sensitive attribute in every QI-group is broken. Best method to crack the connection is bucketization by producing two tables, QI table for the QI attributes and the sensitive table for the sensitive attribute. The two QI table and Sensitive table form the anonymized dataset.

The tables are anonymized by bucketization. The intention is each individual tuple are uncombined to a dataset with a more probability. The foreground information attack appears in a table produce by bucketization. The similar problem arises with a bucketization and generalization based method anonymization. Reason for issues is the adversary has disposal at the external table with which the details of individuals mapping to a QI-group are looked up. The two kinds of bucketization techniques are global recoding and local recoding. The foreground information attack happens in the table produced by either global recoding or local recoding. Under

global recoding, all amount of a particular attribute value are recoded to the equivalent value. Under local recoding, happening of the identical value of an attribute is recoded to diverse values.

VII. FULL FUNCTIONAL DEPENDENCIES (FFD) FOR BUCKETIZATION

FFDs allow de-generalization of the anonymized data facing isolation breaches. Based on the collision of FFDs to privacy distinguished safe FFDs block any FFD-based attack from the dangerous ones. The (d,ℓ)-inference model is planned to protect alongside the FFD-based attack. The (d,ℓ)-inference model needs each anonymization group includes sensitive values that are of same frequency, where the likeness is handled by d. In addition, it needs any two anonymization groups, either zero or at worst ℓ coinciding distinct sensitive values, and at least ℓ non-coinciding distinct sensitive values. The input to attain (d,ℓ)-inference is the suitable grouping of sensitive values. Analyze the quantity of information loss by tuple repression for each grouping strategy. The bucketization algorithm exists of two steps, phase-1 partition and phase-2 QI-group construction. The optimal separating scheme with minimal information loss of NP-hard is proved. A two heuristics approach namely top-down and bottom-up approaches to construct partitions with low information loss is designed. An efficient bucketization algorithm for numerous unsafe FFDs is calculated on both time performances and information loss of the anonymization algorithm empirically designed.

A. Bucket construction

The apparent sensitive values are sorted by their frequencies in ascending order. Based on the arranged result, the contiguous sensitive values satisfying d-closeness are clustered into the similar bucket. Only the buckets contain at least ℓ distinct sensitive values which are kept after bucketing process completion. For any residue sensitive value there are three options either it is detached, nor it is added to an obtainable bucket or it is added to a fresh bucket with other excess sensitive values. To add sensitive value, the bucket is picked. The maximum frequency is the largest value than the frequency of s. The other ℓ−1 remains sensitive values and are picked for the closest frequency. Among these three options, the one that rebound the minimal number of tuples are removed to make the buckets d-closeness satisfaction.

After all sensitive values are bucketed. The larger buckets are split into smaller ones. In particular, each bucket is split into smaller disjoint buckets. The splitting is order-preserving, i.e., the frequency of all impressionable values in the bucket is smaller than that of all sensitive values in the split bucket. After splitting, all the buckets are ordered by maximum frequency of their impressionable values in ascending order. At the end of the process, it returns a set of disjoint buckets and least ℓ distinct impressionable values.

B. Bucketization algorithm

The bucketization algorithm constructs the QI - groups for unsafe FFD attack. An unsafe FFDs necessity includes at least one sensitive attribute in its consideration. A naive bucketization method is applied on grouping strategies exactly for all definite values of the sensitive attributes. A tremendous information loss occurred while tuple suppression, and especially for the dataset whose sensitive values are of skewed frequency.

Phase one partition is to decrease the information loss into lower disjoint segments by applying tuple suppression splitting. Partition applies groupings on these sectors returning the smallest number of removed tuples. The phase two is of bucketization on QI-group construction, the information loss is reduced by data generalization while construct QI-groups. The multiple unsafe FFDs are present, because these FFDs share determinant dependent attributes. The bucketization algorithm applied for each FFD individually and solves inconsistent QI-grouping decisions on the tuples by Pick representative FFDs and FFD-chains construction methods.

Bucketization a group based anonymization technique is considered and the dataset are grouped with two phase partitions and construction. Bucketization for full functional dependencies protect against the record linkage attack. On privacy preserving data publishing is efficient if data publishing is safe. So Algorithm safe publishing and versatile publishing is carried on.

VIII. PERFORMANCE RESULT

Performance result section demonstrates the experimental analysis of bucketization a group based anonymization technique with versatile safe publishing. It is measured in terms of

A. Bucketization
B. Problematic Tuples
C. Adversary Knowledge
D. Time performance
E. Privacy Breaches

a. Bucketization

Bucketization is the process of grouping several records and mixing their sensitive values. In ASP, SP-Mondrian and SP-Hilb are bucketized by integrating worst-case eligibility test and stratified pick-up into the original Mondrian and Hilb, respectively. In Versatile Publishing GD and UAD are implemented in the bucketization based anatomy algorithm as the single-table anonymization subroutine. Group based anonymization approach by bucketization basically hides each individual record behind a group to preserve data privacy. In FFD approach The QI are grouped.

TABLE I. PPDP VS. BUCKETIZATION

PPDP	Bucketization (%)			
	ASP	***VP***	***GBA***	***FFD***
10	52	53	55	57
20	54	54	56	55
30	59	62	65	64
40	61	63	65	66
50	68	72	74	75
60	72	74	77	79
70	74	77	81	85

The above table I describes the percentage of bucketization grouping data set while publishing. Based on the above table I a graph is derived follows

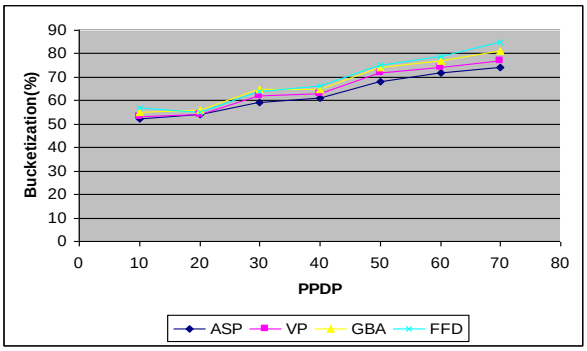


Fig. 1. PPDP vs. Bucketization

The above Fig. 1. for privacy preserving data publishing versus bucketization is elaborated. Around 2-8% differs in each methodology ASP, VP, GBA and FFD for bucketization. As grouping of data set are bucketed based on various algorithms and publication methods.

b. Problematic Tuples

Considering the quasi identifier (QI) initial size as 3 the table is formulated. The QI size is set because the algorithm needs to process more attribute sets increasing execution time. The proportion of problematic tuples among sensitive tuples increases with QI size. With a larger QI size, there is a higher chance that individual privacy breaches due to more attributes increases which can be used to construct the global distributions.

TABLE II. QI SIZE VS. PROBLEMATIC TUPLES

QI size	Problematic Tuples (%)			
	ASP	***VP***	***GBA***	***FFD***
3	32	37	40	35
4	34	38	44	36
5	39	47	50	48
6	41	53	55	51
7	48	62	64	58
8	52	64	67	63

GBA has fewer privacy breaches compared with ASP and FFD because the side-effect of the minimization of QI values in each QI-group adopted in GBA makes the difference in the global distribution

among all tuples in each QI group smaller. Thus, the number of individual with privacy breaches is smaller. Based on the above table II. QI size vs. problematic tuples a graph is derived as follows

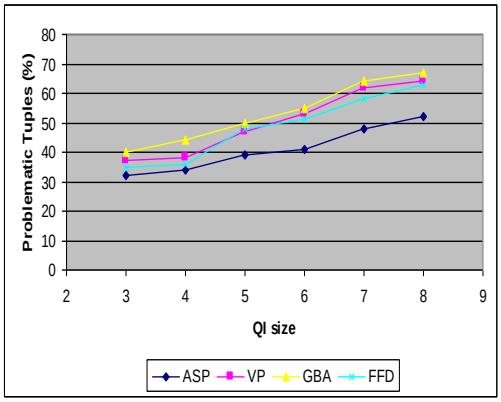


Fig. 2. QI size vs. Problematic Tuples

The above figure Fig. 2. QI size vs. problematic tuples shows GBA efficiency compared to ASP, VP and FFD. The reason is no privacy breaches is not allowed due to the large QI-groups formed by global recoding. GBA is 10-15% high compared to other methodology.

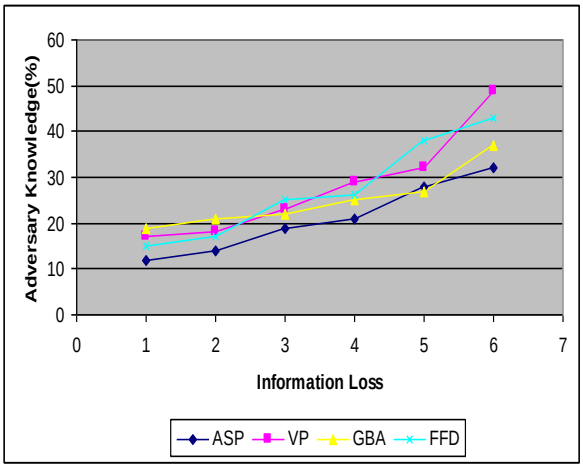
c. Adversary Knowledge

The major goal of all these methodology is to publish the data without the access of adversary knowledge.

TABLE III. INFORMATION LOSS VS. ADVERSARY KNOWLEDGE

Information Loss	**Adversary Knowledge (%)**			
	ASP	***VP***	***GBA***	***FFD***
1	12	17	19	15
2	14	18	21	17
3	19	23	22	25
4	21	29	25	26
5	28	32	27	38
6	32	49	37	43

The information loss is due to the interference of adversary to find sensitive attributes (SA) noticing quasi identifier (QI). The information loss is low in all these above methodology due to bucketization. Based on the above table III. a graph is derived as follows



f

Fig. 3. Information Loss vs. Adversary Knowledge

The above Fig. 3. elaborates the information loss versus adversary acts. ASP, VP, GBA and FFD works on low information loss strategy. The overall average percentage is 5-10% resulting in low information loss comparing all four methodologies.

d. Time performance

Time taken for publishing data privacy preservation is methodized. Performance of individual methodology is measured based on their time efficiency in completing the data publishing task effectively with versatile safe publishing.

TABLE IV. DATA PUBLISHING VS. TIME (SECS)

Data Publishing	Time (Secs)			
	ASP	***VP***	***GBA***	***FFD***
1	2	3	9	5
2	4	6	11	7
3	9	8	12	15
4	11	10	15	16
5	18	15	17	28

A less time is required to publish data for these methodologies. Based on the above Table IV a graph is derived as follows.

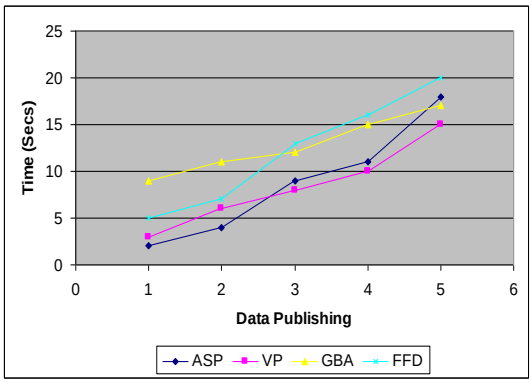


Fig. 4. Data Publishing vs. Time (Secs)

The above graph (figure 4.4) data publishing versus time (secs) is elaborated. Group based anonymization takes more time in publishing around 20-40% compared to ASP and VP and 55-60% compared to FFD. The reason is ASP uses worst-Case eligibility Test and stratified pick-up algorithm in publishing while versatile publishing uses Guardian Normal Form for publishing and FFD uses bucketization algorithm.

e. Privacy Breaches

Methods ASP, VP, GBA and FFD works against privacy breaches thus protecting the data from adversary in transmitting data.

TABLE V. DATASET VS. PRIVACY BREACHES

Dataset	Privacy Breaches (%)			
	ASP	**VP**	**GBA**	**FFD**
10	82	83	86	89
20	84	86	90	91
30	91	88	91	92
40	95	90	93	95
50	98	91	96	97
60	99	92	97	98

The above Table V for dataset versus privacy breaches is described. The value in each methodology, average one another as they aim to defend against same privacy breaches. Based on the above table 4.5 a graph is derived as follows

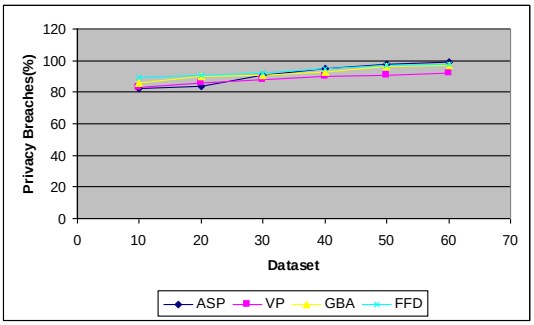


Fig. 5. Dataset vs. Privacy Breaches

The above Fig. 5. exactly shows the methods value point lying on the same graph as all the ASP, VP, GBA and FFD goals to defend against the privacy breaches. Full functional dependencies is little high in result producing a better attack as it uses bucketization algorithm which partitions tuples and constructs QI- group.

IX. CONCLUSION

Bucketization a group based anonymization technique is considered and the dataset are grouped with two phase partitions and construction. The two kinds of bucketization techniques global recoding and local recoding protect against adversary knowledge. Bucketization for full functional dependencies protect against the record linkage attack. The bucketization algorithm for each FFD individually solves inconsistent QI-grouping decisions on the tuples by Pick representative FFDs and FFD-chains construction methods.

On privacy preserving data publishing is efficient if data publishing is safe. So Algorithm safe publishing and versatile publishing is carried on. Algorithm- Safe data Publishing (ASP) model properly describes algorithm based disclosure as the changes among two random worlds. The native one is every possible mapping between an original table and the published table is equally mapping which violates an adversary's external knowledge. The other smart one is mapping that follows the data publishing algorithm. Versatile publishing enables the generation of multiple sub-tables with possible coinciding attributes. Each sub-table in multiple sub-tables only addresses one privacy rule and therefore can be processed by algorithms designed for the QI-SA framework. GNF is focused on design of published tables as each privacy rule rest in event over the collection of all published tables

REFERENCES

[1] Xin Jin, Nan Zhang and Gautam Das,Algorithm-Safe Privacy-Preserving Data Publishing, ACM transaction 2010
[2] Raymond Chi-Wing Wong, Ada Wai-Chee Fu, Ke Wang, Philip S. Yu, Jian Pei, Can the Utility of Anonymized Data be used for Privacy Breaches?, ACM Transactions on Knowledge Discovery from Data, Publication date: March 2010
[3] Hui Wang and Ruilin Liu, Privacy-preserving publishing microdata with full functional dependencies, ELSEIVER 2011
[4] Xin Jin, Mingyang Zhang, Nan Zhang and Gautam Das, Versatile Publishing For Privacy Preservation, ACM transaction 2010
[5] Josep Domingo-Ferrer and Ursula Gonzalez-Nicolas,Hybrid microdata using micro aggregation, ELSEIVER 2010
[6] Nissim Matatov, Lior Rokach and Oded Maimon, Privacy-preserving data mining: A feature set partitioning approach, ELSEIVER 2010
[7] Bin Liu, Yurong Jiang, Fei Sha, Ramesh Govindan, Cloud-Enabled Privacy-Preserving Collaborative Learning for Mobile Sensing., ACM transaction 2012
[8] Keke Chen1and Ling Liu2, Geometric Data Perturbation for Privacy Preserving Outsourced Data Mining, IEEE Transactions Knowledge and Data Engineering Year 2012
[9] Elizabeth Durham, Yuan Xue , Murat Kantarcioglu and Bradley Malin, Quantifying the correctness, computational complexity, and security of privacy-preserving string comparators for record linkage., ELSEIVER 2012
[10] Maneesh Upmanyu, Anoop M. Namboodiri, Kannan Srinathan, and C.V. Jawahar,Efficient Privacy Preserving K-Means Clustering, Springer- 2010
[11] Li Xiong, Vaidy Sunderam, Liyue Fan, Slawomir Goryczka and Layla Pournajaf,PREDICT: Privacy and Security Enhancing Dynamic Information Collection and Monitoring, ELSEIVER 2013

ABOUT AUTHORS

Dr. S. Sukumaran graduated in 1985 with a degree in Science from Bharathiar University, Coimbatore. He obtained his Master Degree in Science and M.Phil in Computer Science also from the Bharathiar University. He received the Ph.D degree in Computer Science from the Bharathiar University. He has 25 years of teaching experience starting from Lecturer to Associate Professor. At present he is working as Associate Professor of Computer Science in Erode Arts and Science College, Erode,

Tamilnadu. He has guided for more than 25 M.Phil research Scholars in various fields and guided one Ph.D Scholar. Currently he is Guiding 5 M.Phil Scholars and 8 Ph.D Scholars. He is member of Board studies of various Autonomous Colleges and Universities. He published around 12 research papers in national and international journals and conferences. His current research interests include Fractal Graphics, Multimedia and Data Mining.

G.Kesavaraj received B.Sc Computer science, M.Sc Computer Science and M.C.A Degree from Bharathiar University, Coimbatore. Now he is pursuing Ph.D degree in Computer Science from the Manonmaniam Sundaranar University. He has 10 years of teaching experience.He is working as an Assistant Professor of Computer Science in Vivekanandha College of Arts and Science for Women, Tiruchengode Namakkal DT, Tamilnadu. His research interests include Data Mining, Software Engineering and Mobile Computing.