International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 24th Jan 2020. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-01

Title: **CRYPTOGRAPHIC TECHNIQUES FOR PRIVACY-PRESERVING STATISTICAL PARAMETER**

Volume 09, Issue 01, Pages: 76-80.

Paper Authors

B NAGABHUSHANA BABU, S. ZAHEER AHAMMED

JNTUACEK, Kalikiri, Chittoor, AP, India

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CRYPTOGRAPHIC TECHNIQUES FOR PRIVACY-PRESERVING STATISTICAL PARAMETER

[1]B NAGABHUSHANA BABU, [2]S. ZAHEER AHAMMED
[1,2]Ad-hoc Assistant Professor,Dept of CSE, JNTUACEK, Kalikiri, Chittoor, AP, India.

Abstract— Statistical population parameters are sensitive in the todays' international. The dreams of privacy maintaining records mining comprise the strategies of distortion and noise addition. In order to preserve privateness, many measures have be propounded in the studies, identifying the quantity of sensitivity and along with the desired noise is a manner, despite the reality that through estimating the possibility distributions of severa trials of data, the adversaries have to look at some PII from the disseminated statistics. In the parameter pushed privacy, we recommend that the unique facts in the message is distorted and may be recognized to the recipient handiest, which can not be guessed with the resource of the intruders. In this paper a gesture of software of statistical parameter privacy has been mentioned that have been drawn from severa studies belongings.

Keywords—privacy, statistical parameter privacy, crypto systems

.

I. INTRODUCTION

The pinnacle of the world business is information, information rules world. Information digitized, stored at the reducing costs, prunes human lives into malice. The human lives encumber costs of usage of technology towards digitizing data. Information out of control is information breached and infested. Data is shared into public domains, keeping individual in anesthetic senses. There is a wide scope of benefits in sharing data on social media, for such is usually a benefit or utility to be obtained by sharing data. Personally identifiable information might be undesired for revealing, however if the benefits of sharing proves fidelity, a natural trade-off between utility and privacy shall arise. The research in privacy and security faces surge on the problem of utility versus privacy [20][21]. Differential privacy and k-anonymity are the famous approaches that are used for providing privacy.

Statistical population parameters are estimated from different users to select and implemenstrategies of anonymity towards preserving privacy. There parameters are true and some estimate parameters are mapped and tested whether information can be revealed. Often, these are checked with specific users or a user group. The objectives of privacy preserving are primarily by suppression. Similar values are clustered and a suppression technique is applied then the values are distorted with noise. Most common examples exists in healthcare, where a group of users of a medical instance survey might remove their names, ages and reserved weights are assigned by addition of a zero-mean random variable. The disseminated data shall always look like the original, except that the end users or observer shall not be able to learn patterns that lead to sensitive statistics of the raw data [1][2]. The linear philosophy comes that a user shares a stream in order to obtain service, under the belief the data is faithful to obtain better quality of services and easy analysis of hidden information. A general example states here, a copy of document is given

to the online service agencies and expect that no personality features, such as signature, identification marks are inferred[19][20].

Exploration into the privacy research contributes enormous ideas, out of which the idea in this article only pave some direction for the better opportunities of preserving privacy[1][2][3][6][7][8]. Nevertheless, the technologies run to provide strong security and implement privacy policies, but still tiny bits of information forms into vast number of data by vast users and compiled into databases that can reveal the personally identifiable information.

II. Cost of Privacy

In the world of hefty business options life style and programs have become a sacrifice. Only certain folds of sectors grow but still choking the effectiveness of the privacy measures and regulations. Hundreds of law books preach about privacy and privacy-related services particularly by the organizations that practice law and accounting. Breach of privacy runs the company into loss of billions of costly issues. Privacy compliance is sought be run with easy and just with a compliance of book rules, but a growing firm cannot afford to establish a "privacy official". Law enabled for keeping the consumers apprised and assured of privacy issue, even resulted in deluge of privacy purge. Even a simple signup for a medical consultation online leads to derailment of privacy, requiring so much paper work about lifestyle information, which an individual wants to be private or personal. Torrent of notifications and privacy reforms fall as desensitized and careless. Government and Federal institutions preclude private organizations to hold the personal information of an individual, still as an understanding between the organizations they practice collecting the individual private information by any means. Institutes dealing with Healthcare and Medicine shall have to implement norms of privacy at any cost where, there are mediums which they are stored and represented for the reserved enquiries in an emergency. Preserving on digital mediums is always prone to breach by hackers. Adversaries have talent to understand the medium online and offline timings and crack on the firewall to intercept the individual data.

III. Privacy Preserving

Privacy preserving on various kinds of data sets both dynamic and static have been studied in the research very extensively. An important policy of privacy attacks on datasets is to again identify individuals' information available in the published datasets, which may contain sensitive information with some external influence of modeling [9][10][11]. To combat with re-identification attacks of adversaries, the mechanism are introduced that use some clustering and iterative procedures such as *k*-anonymity. Quasi-identifiers are set up that act as attributes on the data sets which work to enable the join of attributes and prevent the re-identification[9][12][14]. In k-anonymity, each and every record is indistinguishable from at least k-1 records. The larger the *k* the better the privacy. For certain specific data sets in altered forms the *k*-anonymity is found not suited. Diversity of sensitive attributes is the inundated lack in the existing works. The degree of privacy protection lies on the size and form of data sets. Classification sometimes proves to reconsolidate the corpus of data sets in order to enable easy sampling and continue the process of *k*-anonymization. The size of equivalence classes on quasi-identifiers and attributes which contain tuples appear identical to the class attributes, which is not preferred to be a classical solving method of privacy preservation [12][13][15][16]. Further, even though k-anonymity is implemented, if all the records for a sensitive attribute within the group sized *k* is same, the personally identifiable values represented by the sensitive attribute for the group containing *k*

records can be revealed easily [11][13]. Therefore, *l*-diversity is proposed that maintains the minimum group size and focuses on maintaining the diversity of the attributes. A data block identified by a quasi-identifier group, as a set of values, such that they are non-sensitive values generalize to quasi-block, the quasi-block is *l*-diverse if it contains ***l*** well represented data sets for the sensitive attributes out of the group of attributes, thus a data block or a table is *l*-divers if a quasi-block in it is *l*-diverse.

Further, if the distribution of tuples in the l-diverse block is predictable, then the real values of the data are also predictable, which may not occur on the sparse data. Generally huge datasets, containing repeating values in the datasets, which does not guarantee the sparsity of data, fail to secure privacy. For this reason, any adversary with background knowledge can guess the information in the anonymized data, the distance between the distributions of the quasi-block is maintained and the closeness of the data with respect to the domain is ordered, a *t*-closeness, is proposed which uses the distance property of the quasi-blocks with respect to the sensitive attributes within an anonymized group. The individual distributions are not different from the global distribution of the group of datasets that are anonymized [16][17][18].

IV. STATISTICAL PARAMETER PRIVACY

Statistical parameter privacy [7]is a filter already proposed by eminent researchers in the area, which can inherently apply complexity, and the statistics of sequences of data shall be taken into account.

The relationship between utility and privacy has been connoted by many research contributions in the privacy preserving data mining. The rate of distortion function decides the trade-off, whether the priority is utility or privacy. Entropy or information in mutual share shall decide the privacy measure, whether it is privacy or utility[1][2][5]. However, the challenge in the parameter driven privacy is to provide either perfect privacy or perfect utility. In certain contributions the concept of maximal correlation is considered as privacy measure, also the other contributions purported that this measure is equivalent to the principle 'maximize MMSE' – (Mini Mental State Examination: a principle of memorizing the sensitive information) of private data when shared to unsecured public. The problem of privacy preserving is technically connoted as rate-distortion, where privacy is defined by mutual information and characterized as optimal asymptote of leakage for *independent identically distributed* (or *personally identifiable information*) and generic privacy preserving mechanism.

State theorem of defining privacy of privacy preserving average consensus algorithms, indicatively measure privacy with the begin state and the target state of achieving the privacy.

V. NOTATIONS

The notations emphasize the importance of privacy and their methods of preserving. The lowercase letters such as ***x*** and ***y*** are mainly used to represent constants or realizations of random variables, capital letters such as ***X*** and ***Y*** stand for the random variables in itself, and calligraphic letters such as ***X*** and ***Y*** are reserved for sets.

The sequence of independent identically distributed random variables $\{X_k\}_{k=1}^{n}$ is denoted as X^n. Markov Chain methods are also used to represent state oriented privacy preserving forms such as $X \rightarrow Y \rightarrow Z$, where ***X***, ***Y*** and ***Z*** are random variables. Entropy is denoted by H(°) and mutual information.

VI. MEASURE THEORY

Measure Theory is a branch of mathematics, invented by Henri Lebesgue, a French Mathematician. Measure Theory is a standard way

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

of assigning a measure to subsets of n-dimensional Euclidean space. A finite series of integers always coincides with the standard measure of length. A set of integers often coincide with area, volume and necessary dimensions. It is also referred as an n-dimensional volume, n-volume, or simply volume. Sets of integers / values that are considered for a Lebesgue measure are also known as Lebesgue-measurable; the measure of the Lebesgue-measurable set A is here denoted by λ(A).

VII. A General System Model

In a classical three user problem, A, B and C, A has information to share with B by observing it as the random parametric source. The parameter contains private information that A do not want to disclose in public, might even be unknown to A. The parameter may also represent personality traits of A, where while messaging to B, they are not revealed except the message. By not considering any constraint A chooses to directly send the observed sequence of values (in message), which is overheard by C, who is interesting in portraying A, by characterizing the statistical properties of the random parametric source, where the parameter need to be estimated. In order to protect privacy, A can send the distorted version of the source, but only that is useful to B.

A observes several samples (say $\boldsymbol{n}$) of the X (information) $\in \chi$ with all the samples that are *independent identically distributed* in a probability distribution of P with the probability measure $P_{\theta\theta}$. Where $P_{\theta\theta}$ is considered as parameterized family of distributions $P_{\Theta} = \{ P_{\theta} : \theta \in \Theta\}$ on a measurable space which is believed as continuous with Lebesgue measure. $\boldsymbol{\theta_0}$ is considered as a data point in the interior of Θ, the $p_{\theta\theta}$ is the probability density function of $P_{\theta\theta}$ with respect to a fixed finite measure with mean difference. $p_{\theta\theta}(x)$ is a non-zero almost everywhere on χ and the corresponding probabilistic measure $P_{\theta\theta}$ is observed as absolutely continuous with respect to Lebesgue measure.

As the probability distribution is known $P_{\theta\theta}$, the value of the parameter is chosen randomly priori to the distribution p(θ) with complies with Lebesgue measure assumed as above. This parameter is considered as random variable Θ. As believed earlier A sends the distorted information to B, on the free channel, only just to make C not to understand the personality traits of A and increase knowledge. Further, it may be considered that B has no advantages when compared to C. The information in the message is communicated to both users and the level of quality is same, the strategy employed is also known to everyone, if and only if the stochastic information used to increase the privacy is not learnt by any recipient.

VIII. Overview of Results

Without properly distorting the information in the message $\boldsymbol{X^n}$, the probability of information about Θ is gained by the eavesdropper increases $\boldsymbol{n}$ times. As it is known as a basic principle, the differential entropy of X, where X is absolutely random variable with probability density function $p_X(x)$, the player eavesdropper C in the communication shall be able to estimate the parameter with arbitrarily higher precision.

Alternatively, let us assume that $\chi = \Re^{\mathbf{d}}$ and the parameter set Θ has a finite cardinality. In each instance of communication the player A generates random variable with distorted information, as in $Y_k = \Phi(X_k, \theta, \tilde{\theta})$, where $\tilde{\theta}$ is part of parameter set Θ and $\Phi(.,\theta,\tilde{\theta})$ is a map from $\Re^{\mathbf{d}}$ to $\Re^{\mathbf{d}}$. The map $\Phi(.,\theta,\tilde{\theta})$ is designed such that the common probability density function of $\{\boldsymbol{Y_1}, \boldsymbol{Y_2}, \ldots \boldsymbol{Y_n}\}$, is equal to $p_{\tilde{\theta}}(x)$. And it is assumed that θ is known to the player A and $\tilde{\theta}$ is

selected, by the specified privacy filter, to parallelly ensure the privacy of θ and accuracy of the communicated information to player B.

CONCLUSIONS

The schemes discussed in the overview of results propose for the statistical parameter privacy problem, where the mutual information was used as the privacy measure. In the first scheme, stochastic privacy filter and deterministic privacy filter are used under the belief that the parameter belongs to a continuous alphabet and the mutual information was exploited as proxy. The parameter was assumed a it belongs to a finite set of all the possible values in the second scheme. The mutual information was directly used as a privacy measure. In this paper we urge to propose the variant of statistical parameter privacy theoretically.

REFERENCES

[1] Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." ACM Sigmod Record. Vol. 29. No. 2. ACM, 2000.

[2] Aggarwal, Charu C., and S. Yu Philip, eds. "Privacy-preserving data mining: models and algorithms". Springer Science & Business Media, 2008.

[3] Aldeen, Yousra Abdul Alsahib S., and Mazleena Salleh. "Privacy Preserving Data Utility Mining Architecture." Smart Cities Cybersecurity and Privacy. Elsevier, 2019. 253-268.

[4] Dennis, Simon, et al. "Privacy versus open science." Behavior research methods (2019): 1-10.

[5] Pinkas, Benny. "Cryptographic techniques for privacy-preserving data mining." ACM SIGKDD Explorations Newsletter 4.2 (2002): 12-19.

[6] Ruan, Minghao, Huan Gao, and Yongqiang Wang. "Secure and privacy-preserving consensus." IEEE Transactions on Automatic Control (2019).

[7] Showkatbakhsh, Mehrdad. "Security and Privacy in Dynamical Systems". Diss. UCLA, 2019.

[8] Sharma, Shivani, and Sachin Ahuja. "Privacy Preserving Data Mining: A Review of the State of the Art." Harmony Search and Nature Inspired Optimization Algorithms. Springer, Singapore, 2019. 1-15.

[9] Zhou, Bin, Jian Pei, and WoShun Luk. "A brief survey on anonymization techniques for privacy preserving publishing of social network data." ACM Sigkdd Explorations Newsletter 10.2 (2008): 12-22.

[10] Aggarwal, Charu C., and S. Yu Philip. "A general survey of privacy-preserving data mining models and algorithms." Privacy-preserving data mining. Springer, Boston, MA, 2008. 11-52.

[11] Kantarcioglu, Murat. "A survey of privacy-preserving methods across horizontally partitioned data." Privacy-preserving data mining. Springer, Boston, MA, 2008. 313-335.

[12] Wang, Jian, et al. "A survey on privacy preserving data mining." 2009 First International Workshop on Database Technology and Applications. IEEE, 2009.

[13] Lindell, Yehuda, and Benny Pinkas. "Privacy preserving data mining." Annual International Cryptology Conference. Springer, Berlin, Heidelberg, 2000.

[14] Pinkas, Benny. "Cryptographic techniques for privacy-preserving data mining." ACM Sigkdd Explorations Newsletter 4.2 (2002): 12-19.

[15] Aggarwal, Charu C., and S. Yu Philip, eds. Privacy-preserving data mining: models and algorithms. Springer Science & Business Media, 2008.