Research Journal of Applied Sciences, Engineering and Technology 12(7): 782-789, 2016
DOI:10.19026/rjaset.12.2753
ISSN: 2040-7459; e-ISSN: 2040-7467
© 2016 Maxwell Scientific Publication Corp.
Submitted: November 3, 2015 Accepted: January 8, 2016 Published: April 05, 2016

Research Article

State-of-the-art in Privacy Preserved K-anonymity Revisited

[1, 2]Yousra Abdul Alsahib S. Aldeen and [1]Mazleena Salleh
[1]Faculty of Computing, Universiti Teknologi Malaysia, UTM, 81310 UTM Skudai, Johor, Malaysia
[2]Department of Computer Science, College of Education_Ibn Rushd, Baghdad University, Baghdad, Iraq

Abstract: The prevalent conditions in data sharing and mining have necessitated the release and revelation of certain vulnerable private information. Thus the preservation of privacy has become an eminent field of study in data security. In addressing this issue, *K*-anonymity is amongst the most reliable and valid algorithms used for privacy preservation in data mining. It is ubiquitously used in myriads of fields in recent years for its characteristic effective prevention ability towards the loss of vulnerable information under linking attacks. This study presents the basic notions and deep-insight of the existing privacy preserved *K*-anonymity model and its possible enhancement. Furthermore, the present challenges, excitements and future progression of privacy preservation in *K*-anonymity are emphasized. Moreover, this study is grounded on the fundamental ideas and concepts of the existing *K*-anonymity privacy preservation, *K*-anonymity model and enhanced the *K*-anonymity model. Finally, it extracted the developmental direction of privacy preservation in *K*-anonymity.

Keywords: Generalization, privacy preservation, suppression and Quasi Identifiers (QI)

INTRODUCTION

The massive information technology development with the ensuing extensive formation of application network has resulted in a huge repository and publication of information (Xu *et al*., 2014). The situation has necessitated the dire need for effective data mining applications used for knowledge discovery and information retrieval. Concurrently, the data mining initiatives has brought about problems in terms of the protection of privacy and the revelation of sensitive information; which in turn has made the preservation of privacy a crucial research area of great significance (Dev *et al*., 2012). This phenomenon occurs at the current era of electronic availability of vast public information, with historical depth. The huge amount of data could be traced and linked together to form explicit electronic image profiling of an individual, even though in circumstances where no overt identifiers or obvious markers such as name or phone number are present (Yang *et al*., 2014). According to Sweeney (2002a, 2002b) other unique combination of identity markers such as birth date and postal code, could be used to triangulate or be used in cross referencing through publically available data to re-identify individuals.

Thus in the current data rich matrix which are technologically enabled, there exist questions on the protection and security issues with regards to privacy protection from the disclosure of identity of individuals; in the setups such as a medical institution, financial organizations or public health organizations (Zhu and Peng, 2007). How do custodians of data ensure that person-specific shared-records and released information do not disclose or lead to the disclosure of the identity of the subject of the data? How can the determination of the subjects of the data be obstructed or kept anonymous?

Hence strict compliance to *K*-anonymity is needed in order to maintain and protect the anonymity of the individual through the use and release of information (Pan *et al*., 2012). It is imperative that released data adheres to *K*-anonymity ruling and is considered to be as such on the condition that each of the record released possesses at least (*k*-1) other data which is noticeable in the released records with indistinguishable values extending across a particular set of fields known as the quasi-identifier. The quasi-identifier attributes encompass fields with a high probability of emerging in other identified data sets (Zhang *et al*., 2013a). Hence, the provision of privacy protection is assured by *K*-anonymity by ensuring that each released records that are directly associated with external information are related to at least *k* individuals of each record released. The fundamental theory of this study is grounded on *K*-anonymity approach, which is the pervasive current privacy preservation approach. It analyzed the central

Corresponding Author: Yousra Abdul Alsahib S. Aldeen, Faculty of Computing, Universiti Teknologi Malaysia, UTM, 81310 UTM Skudai, Johor, Malaysia

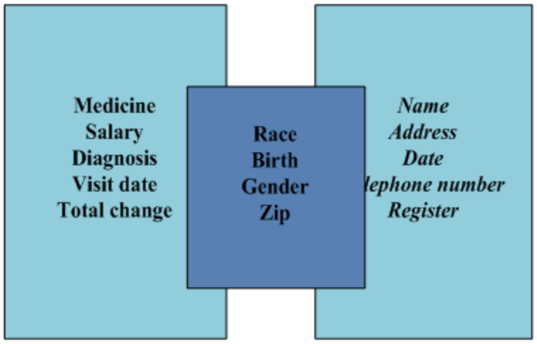

Fig. 1: A typical linking attack

premises, ideas and models of existing *K*-anonymity algorithm and finally extracted the future developmental direction of *K*-anonymity on privacy protection. The main goal of this study is to provide readers with the fundamentals insight of *K*-anonymity that relies on the persistent privacy preservation approach. It analyzes the central premises, ideas, models of existing *K*-anonymity algorithm and highlights the future trends of privacy protection facilitated *K*-anonymity development.

***K*-anonymity model:** There is a global demand to certain agencies for publicizing their data involving research and statistics on medical, health and population aspects. However, utter risks are foreseen towards the disclosure of such identity through linking attack or inference from other channels by the phishers. This attacks are unavoidable even if the discharge is based on quasi-identifiers with two or more data table, where identification markers such as name, ID number and other attributes are suppressed or hidden (Pan *et al*., 2012). In this viewpoint, *K*-anonymity is considered to be the most prospective approach against the linking attack mediated disclosure and maintaining the privacy protection.

Figure 1 illustrates a typical linking attack. Samarati and Sweeney (1998) first brought *K*-anonymity to attention, where they showed a published data must contain a specific number (at least for the *K*) with indistinguishable records so that an attacker is unable to discern the particular privacy information of unique individuals. Thus, personal privacy is maintained through leakage prevention. The parameter *K* can be specified by a user for the maximum risk of information leakage in *K*-anonymous. Furthermore, it is capable to protecting an individual's privacy to a certain degree by reducing the availability of the data. This discovery propelled gamut research activities in developing the technology with absolutely secured data mining and sharing.

The objectives of data mining is centered on the efficient development of the original data via anonymization, with the goal of attaining the ultimate anonymity, the minimal time spent and space utilized and the optimum data availability. It is dissimilar as compared to data perturbation technique, where interference methods are deployed such as randomization, distortion and disturbance. It should be noted that the authenticity of data can be preserved by *K*-anonymity.

Definition 1:
***QI* (Quasi-identifier):** Given a table *U*, *a* $T\,(A^{1} \ldots A^{n})$, $f^{C}\!:U \rightarrow T$, $f^{g}\!:T \rightarrow \acute{U}$, where $U \subseteq V$, a quasi-identifier of *T* (Q^{T}) defines a set of attributes $\{A^{i} \ldots\ A^{J}\} \subseteq \{A^{1} \ldots.\ A^{n}\}$, where $\exists\ p^{i} \in U$ such that $f^{g}\ (f^{c}\ (p^{i})[Q^{T}]) = p^{i}$ (Sweeney, 2002a).

Definition 2:
***K*-anonymity:** A table *T* satisfies *K*-anonymity if for every tuple $t \in T$ there exist (k-1) other tuples $t^{i1} t^{i2} \ldots\ t^{k-1} \in T$ such that $t^{i1}\ [C] = t^{i2}\ [C] = \ldots\ ikt^{ik-1}[C]$ for all $C \in Q^{T}$ (Machanavajjhala *et al*., 2007).

Definitions 1 and 2 signify the property of maintenance and the truthfulness of the data. Using the proposed generalization and suppression techniques we implement the *K*-anonymity on a private table. By referring to the values in the domain, generalization attributes the values through the operations of replacing or recoding them. It is based on a domain generalization hierarchy and a corresponding value of generalization hierarchy. Characteristically, it is composed of domain generalization hierarchy and the tree is made up of the corresponding value generalization hierarchy. For instance, the association of the parent/child relationship represents the direct generalization/specialization relationship. Conversely, the suppression involves the removal or exclusion of sensitive confidential data/information. Thus, through enforced generalization it attains the *K*-anonymity and reduces the information loss via the application at the single cell level, entire tuple, or entire column. Intuitively, the removal of outliers from a table would result less generalization and eventually leads to the achievement of *K*-anonymity. It performs without requiring the compulsion of implementing an enormous amount of generalization to satisfy *K*-anonymity (Ciriani *et al*., 2008). Table 1 enlists the typical information in a medical form. It demonstrates that, even after hindering and excluding the specific sensitive confidential data such as identification markers comprising of names, home address, Medicare and ID numbers, there are still probabilities of linking or inferences risks. Actually, some identification markers including quasi-identifier data (age, gender, zip code and others) still exists that can be used by attackers to disclose the identity of the patient and their medical records. Thus, to protect the patient's privacy and to prevent the disclosure of confidential information the data generalization is must

Table 1: An original medical information document

	Quasi-identifier			
ID	Gender	Zip	Age	Sensitive information
1	Female	344625	35	Gastric ulcer
2	Female	350500	32	AIDS
3	Female	371060	37	Flu
4	Female	375067	39	Neurasthenia
5	Male	371060	44	Flu
6	Male	371060	41	Hepatitis
7	Male	350500	46	Neurasthenia
8	Male	350500	47	Flu

Table 2: Generalization of the data in Table 1

	Quasi-identifier			
ID	Gender	Zip	Age	Sensitive information
1	Person	344***	30-40	Gastric ulcer
2	Person	350***	30-40	AIDS
3	Person	371***	30-40	Flu
4	Person	375***	30-40	Neurasthenia
5	Person	371***	40-50	Flu
6	Person	371***	40-50	Hepatitis
7	Person	350***	40-50	Neurasthenia
8	Person	350***	40-50	Flu

as summarized in Table 2. This places the data in a larger cohort range, where the form of 'Generalization' operations symbolizes the gender as person, age range instead of actual age and indicates the zip code via "*" signifying any number with a form of 'suppression' (Table 2).

Weakness of K-anonymity: Weaknesses and vulnerabilities in the original *K*-anonymity through attacks via linking and inferences through various channels still persist. This often leads to information disclosure. These limitations of *K*-anonymity is resolved (Machanavajjhala *et al*., 2007), where a *K*-anonymity model with two attacking methods such as homogeneity and background knowledge are proposed. In the former one, the attacker obtains *K*-anonymous table and sensitive information of an individual. Conversely, in the latter case the attacker acquires the information prior to the attack. Disclosure of sensitive confidential data in *K*-anonymity originates from these two types of attacks. Thus, an enhancement to the original *K*-anonymity through enhanced model of *K*-Anonymity, *L*-Diversity and (α, k) -Anonymity are mandatory (Wong *et al*., 2006). Here, we provide a brief description of each of them.

***L*-Diversity model:** The *L*-diversity model is introduced (Machanavajjhala *et al*., 2007) to enhance the diversity of an anonymous cohort. It reduces the likelihood of inference or deduction of high confidentiality of sensitive values, which in turn result in disclosure or leaks. In a released equivalence cluster of published table of a *K*-anonymous group, the *L*-diversity model must consist of at least *L* well-presented records. This involves the possession of at least one different value for every sensitive attributes per equivalence. Conversely, this particular diversity model does not accommodate for the probable background knowledge of the attacker or the vulnerability of any posted data at risk if the attacker possesses background knowledge on the patient. This considers the setting of parameters in *L*-diversity model as unsuitable.

(*a*, *k*)-Anonymous model: The anonymity property of (α, k)-anonymous model restricts the frequency contained in a sensitive group to be less than α. The occurrence of a high percentage of sensitive information in the matrix of data must be circumvented. Therefore, by amplifying the multiplicity of the sensitive values one can prevent the consistency of attacks. This eliminates the situational risks with enhanced protection as acknowledged by Pan *et al*. (2012). This model enables the selective protection of sensitive property values rated to confidential sensitive information such as AIDS (Pan *et al*., 2012), where the sensitivity of attackers inferences is always below α. Thus, it protects the selected sensitive data from homogeneity attacks. The (α, k)-Anonymous model processing of the sensitive property with a higher degree of value indeed obstructs the attacker from viewing the specific protected susceptible data. This model is rather limited in terms of parameters choice, where only the highest level of sensitive property value is considered. The absence of other level in the model discounts several facets of privacy disclosure issues. Certainly, being a NP-hard problem the optimum K-anonymity is difficult to achieve (Pan *et al*., 2012).

Overview of *K*-anonymity: The data anonymization approach of Loukides and Gkoulalas-Divanis (2012) having low information loss fulfils the data publishers' utility demand and afford (Loukides *et al*., 2012). It presents an accurate information loss measure and efficient anonymization algorithm by discovering a huge part of the problem space. Friedman *et al*. (2008) extended the meanings of K-anonymity and used them to demonstrate that α assumed data mining model does not disclose the K-anonymity of the individuals represented in the learning examples (Friedman *et al*., 2008). This expansion offers a tool to determine the volume of anonymity preserved data mining. This model can be applied to address different data mining problems including classification, association and clustering. *K*-anonymity is often applied in data mining exercise for preserving the individuality of the respondent's data to be extracted. Ciriani *et al*. (2008) defined the possible risks to *K*-anonymity aroused from implementing mining on a collection of data. Two main approaches are proposed to combine *K*-anonymity in data mining (Ciriani *et al*., 2008). Various methods are used to detect *K*-anonymity violations and their removal in association rule mining and classification mining. Based on clustering, He *et al*. (2012) proposed an algorithm to produce a utility-friendly anonymized version of micro data. The performance evaluation revealed that this method outperformed the non-homogeneous technique, where the size of *QI*-attribute is greater than 3. The proposed clustering-based *K*-anonymity algorithm is found to enhance the utility when implemented on actual datasets. *K*-anonymous privacy preservation being a diversified approach indicates that its future expansion may face numerous challenges unless overcome. Patil and Patankar (2013) in their analyses on current *K*-anonymity model and its applications (Patil and Patankar, 2013) summarized some of the multidimensional *K*-anonymous researches used to achieve the basic knowledge. Currently, the majority of *K*-anonymity algorithms depend on multidimensional datasets. The main aim is to improve the quality of anonymity and to reduce the information loss by adding the nearest neighborhood strategy.

K-anonymity is topically attractive due to the feasibility of enhanced privacy preservation in data mining. It effectively avoids the link attacks mediated privacy leakages and extensively used in several areas (Zhu and Chen, 2012). Continual developments and improvements on the current *K*-anonymity privacy protection are in progression. This development gave the birth of an improved *K*-anonymity model. Soodejani *et al*. (2012) employed a version of the chase called standard chase by imposing some restrictions (such as positive and conjunctive) on the dependencies and constrains (Soodejani *et al*., 2012). Further studies on the applicability of other versions of the chase are an open research avenue. This anonymity model principle is similar to that of *L*-diversity privacy model. The demand for achieving other stronger privacy models such as the *t*-closeness is ever-increasing. A numerical method to mine maximal frequent patterns with privacy preserving capability is proposed (Karim *et al*., 2012). It revealed an efficient data transformation with a novel encoded and compressed lattice structure using MFPM algorithm. The proposed lattice structure and MFPM algorithm reduced both search space as well time. The experimental results demonstrated that MFPM algorithm outperformed PC_Miner and existing maximal frequent pattern mining algorithms. Besides the lattice structure, it outperformed FP-like tree and PC_tree algorithm as well.

Recently, a rule-based privacy model is introduced by developing two anonymization algorithms (Loukides *et al*., 2012). It allowed data publishers to express fine-grained protection requirements for both identity and sensitive information disclosure. The first algorithm worked in a top-down fashion, which employed an efficient strategy to recursively generalize the data with low information loss. Conversely, the second algorithm used a sampling and a mixture of bottom-up and top-down generalized heuristics, which greatly improved the scalability and maintained low information loss. Extensive experiments showed that these algorithms significantly outperformed the state-of-the-art in terms of data utility retention, superior protection and scalability. Lately, *K*-anonymity is considered as an interesting approach to protect micro-data associated with public or semi-public datasets from linking attacks (Vijayarani *et al*., 2010). They defined the possible risks to *K*-anonymity that could arise from execution mining on a group of data and the approaches to achieve *K*-anonymity in data mining. Clifton *et al*. (2008) extended the meaning of *K*-anonymity to multiple relations definitions of *K*-anonymity relationships. It is shown that earlier methods either failed to protect privacy, or reduced the data utility in the preservation in various relations setting (Nergiz *et al*., 2009). New clustering algorithms are proposed to achieve multi-relational anonymity. The enhanced efficiency of the approach in terms of utility and effectiveness is clearly demonstrated. The problem of secured outsourcing of frequent item set mining on the multi-cloud environments is examined (Tai *et al*., 2013).

Concerning the challenges in big data analysis, the data partitioning into several parts and independent outsourcing of each part to different cloud based on pseudo-taxonomy anonymization technique called KAT is suggested. A DKNT is proposed to ensure the privacy security for each partial data outsourced to different clouds. Experimental results demonstrated that these approaches achieved good protection and better computation efficiency than the one obtained on a single machine. A *K*-support anonymity mediated protection against a knowledgeable attacker with exact support information is presented (Tai *et al*., 2010).This *K*-support anonymity is achieved by developing a pseudo taxonomy tree with the third party having generalized frequent item sets. The creation of the pseudo taxonomy tree allowed to concealing the original items and limiting the fake items introduced in the encrypted database. The experimental results revealed that this *K*-support anonymity achieved very good privacy protection with moderate overhead storage. Pan *et al*. (2012) analyzed and compared some of the current K-anonymity model and its applications to determine their enhanced efficiency (Pan *et al*., 2012). This enhancement allowed protecting the data privacy against two attacks including the homogeneity and the background knowledge. Models such as *L*-diversity, (α, k) -anonymity and (α, L) -diversification *K*-anonymity are enhanced versions of the *K*-anonymity.

The (α, L)-diversification *K*-anonymity model is considered to be the best one. This approach is considered to be more reliable because it overcomes all the drawbacks present in other two techniques. A suppression dependent novel method for realizing *K*-

anonymity called 'kactus' is introduced (Deivanai *et al*., 2011). This technique fulfils the *K*-anonymity and identifies the characteristics possessing fewer effects on the data records classification and suppresses them if required. It assessed numerous datasets to evaluate its accuracy, where it compared and matched with other *K*-anonymity based methods. Anonymization can be combined with perturbation for privacy preservation in a multiparty environment. Monreale *et al*. (2014) introduced a new definition of *K*-anonymity for personal sequential data, which provided an effective privacy protection model. This method transformed the sequential datasets into a *K*-anonymous form and preserved the data utility with varieties of analytical attributed (Monreale *et al*., 2014). Intensive experiments are carried out with various real-life sequential datasets to demonstrate that the developed method substantially protect the consecutive mining pattern in terms of their number and support. Results are discerned to be extremely interesting for dense datasets. Yet, utility protection is considered to be a key issue particularly for generalization-based methods. The main reasons for over-generalization are the presence of outliers in private datasets. The occurrence of less population in neighborhood outliers in the higher dimensions makes it challenging for an anonymization algorithm to create an equality group of sufficient size. The negative influences of outliers and over-generalization is addressed (Nergiz and Gök, 2014; Nergiz *et al*., 2013), where a hybrid generalizations with only a certain number of repositioned data elements are made (Nergiz *et al*., 2013). This type of generalization is realized by limiting the number of relocations, which the algorithm could apply by controlling the trade-off between utility and truthfulness. Zhang *et al*. (2013b) studies the scalability issue of sub-tree anonymization over big data on cloud using a hybrid method by combining the Top-Down Specialization (TDS) and Bottom-Up Generalization (BUG). The hybrid approach mechanically selected one of the two components by comparing the user defined *K*-anonymity parameter with workload complementary point. Both TDS and BUG showed extreme scalability via a sequence of deliberately designed Map Reduce jobs (Zhang *et al*., 2013b). A highly scalable two-phase TDS approach is introduced using Map Reduce on cloud (Zhang *et al*., 2014a, 2014b). In the first phase, datasets are divided and anonymized in parallel to creating intermediate results. Second phase combined these results and more anonymized to produce consistent *K*-anonymous datasets. Map Reduction is creatively applied on cloud for data anonymization and to deliberately design a group of advanced jobs to achieve highly scalable specialized computation. Results on real-world datasets confirmed significant enhancement of the scalability and efficiency of TDS over existing approaches. Privacy preservation for the growing data being a critical challenge requires in-depth research. Current methods suffer from inefficiency and poor scalability because of centralization and access difficulty to complete data when updated. An effective technique with limited overhead performance is presented (Zhang *et al*., 2013). In cloud system, integration of privacy-aware efficient preparation of anonymized datasets is performed by placing privacy preservation as a metric along with others such as storage and cloud computation. A distributed anonymization protocol for privacy-preserving data publishing from multiple data providers in a cloud system is introduced (Ding *et al*., 2013). This method enabled the performance of a personalized anonymization, which satisfied every data provider's requirements and the union formed a global anonymization for publishing. Furthermore, a new anonymization algorithm is developed using R-tree index structure. Rapid advancements in the social networks raised alarming privacy concerns in terms of data releasing, disclosure and the publishing to third parties. The execution of identifying attributes removal before publishing the data obtained from the social network does not dispel the revelation of the user's identity. The process of privacy protection from disclosure is not easy because it affects the user. The established techniques of social network data publications is widely explored (Liu *et al*., 2014), where a model with confidence p on the preservation of privacy on the social network is defined. Consequently, a privacy preserving hybrid algorithm is developed for social network data publication. By integrating the randomization with *K*-anonymity features this algorithm concealed the sensitive data into the natural groups of social network data. The residual data is processed through the utilization of the random approach. Gkountouna, *et al*. (2014) introduced an anonymized data protection method from identity disclosure attacks and concentrated on a set of values which is anonymized in the constant domains including numerical data (Gkountouna *et al*., 2014). Interestingly, through the decision-making ability of the anonymization algorithm different values are generalized in contrast to the utilization of a fixed generalized hierarchy. This approach possesses two notable advantages:

- The generalization of datasets ensures the non-requirement hierarchy definition
- The limited scope of generalization ensures restricted information loss.

Gkountouna, *et al*. (2014) acknowledged the benefits of the algorithm in relations to the quality of information as compared to the state-of-the-art in a succession of experiments. *K*-anonymity possesses several advantages than other methods (Gkountouna *et al*., 2014). The characteristic approach towards

Table 3: Comparison of various models on privacy preservation

References	Privacy	Utility	Accuracy	Efficiency	Scalability
Loukides *et al*. (2012)	√	√			
Friedman *et al*. (2008)	√				
Ciriani *et al*. (2008)	√				
He *et al*. (2012)	√	√			
Patil and Patankar (2013)	√				
Zhu and Chen (2012)	√				
Soodejani *et al*. (2012)	√				
Karim *et al*. (2012)	√				
Loukides *et al*. (2012)	√	√			√
Vijayarani (2010)	√				
Nergiz *et al*. (2009)	√			√	
Tai *et al*. (2013)	√			√	
Tai *et al*. (2010)	√				
Pan *et al*. (2012)	√				
Deivanai *et al*. (2011)	√		√		
Monreale *et al*. (2014)	√				
Nergiz and Gök (2014)	√	√			
Nergiz *et al*. (2013)	√	√			
Zhang *et al*. (2013a)	√				√
Zhang *et al*. (2014a)	√				√
Zhang *et al*. (2014b)	√				√
Zhang *et al*. (2013b)	√			√	
Ding *et al*. (2013)	√				
Liu *et al*. (2014)	√				
Gkountouna *et al*. (2014)	√				
Wong *et al*. (2006)	√				

generalization involves the division of the tuples in a table into numerous *QI* (quasi-identifier)-groupings. In this order the value of *K* must be smaller than the size of individual *QI*-groupings. Commonly, the enhancement of the anonymized data utility can occur in circumstances where a reduction of each *QI*-group size occurs. He *et al*. (2014) proposed a linking-based anonymity model, where K-anonymity is attained with QI-groupings with the restriction that K must be greater than these groupings (He *et al*., 2014). An efficient heuristic solution is obtained by utilizing the simple yet effective local recoding methodology. Broad range of experiments are performed on real datasets Results on utility demonstrated considerable enhancement than other state-of-the-art methods. Table 3 illustrates the comparison of various models on privacy preservation.

CONCLUSION AND FURTHER OUTLOOK

This study provides an up to date overview on privacy preserved *K*-anonymity. Most of the current findings, theories, practices, applications, challenges, basic insight and future trends of privacy preservation as well as protection against disclosure and leakages of sensitive individual information in terms of data mining and data sharing in existing *K*-anonymity model are underscored. Major significant algorithms and methodologies are emphasized. An analysis with a check list of various methods is tabulated, where five main elements are focused including:

- Privacy
- Utility
- Accuracy
- Efficiency
- Scalability

Upon analyzing the issues of occurrences and potential risks of user's private information leakages into the environment, it is asserted that *K*-anonymity is the most reliable and valid algorithm for privacy preservation in data mining. *K*-anonymity is used for widespread applications due its effective prevention ability towards the loss of vulnerable information under linking attacks. In addition, the merits and demerits of *K*-anonymity are identified and discussed at length. Further strengthening and enhancements of the models are proposed (Machanavajjhala *et al*., 2007; Wong *et al*., 2006; Pan *et al*., 2012). Various existing *K*-anonymity model are analyzed and the feasible solutions in overcoming these limitations are addressed. The main challenges faced by the *K*-anonymous are emphasized. *K*-anonymity is found to be efficient in the protection of data privacy entrenched in a data distribution matrix. It is capable of ensuring the data authenticity and gaining increasing attention due to heuristic solutions capacities. The urgent necessity of privacy preservation in diversified areas of data mining and data sharing in cyberspace, cyber world, cloud computation and knowledge sharing is reaffirmed. All-encompassing phenomenon of the constant data alteration, changes in data forms and transformation, modifications in attributes, the addition of new data and deletion of the old data are the main challenges in *K*-anonymity algorithms because they are based on static datasets. Thus, due to the correlations and interrelationships between data sets, the data mining capabilities without disclosure of sensitive individual information certainly gets affected by these constant changes. Undoubtedly, the significant issues and challenges in privacy preservation implanted in the complex setting are real and need focused attention.

ACKNOWLEDGMENT

This research is supported by University Technology Malaysia, Johor, Malaysia /UTM library and Baghdad University.

REFERENCES

Ciriani, V., S. De Capitani di Vimercati, S. Foresti and P. Samarati, 2008. *k*-Anonymous Data Mining: A Survey. In: Aggarwal, C.C. and P.S. Yu (Eds.), Privacy-preserving Data Mining: Models and Algorithms. Advances in Database Systems, Springer Science+Business Media, LLC, 34: 105-136.

Deivanai, P., J.J.V. Nayahi and V. Kavitha, 2011. A hybrid data anonymization integrated with suppression for preserving privacy in mining multi party data. Proceeding of the International Conference on Recent Trends in Information Technology (ICRTIT, 2011). Chennai, Tamil Nadu, pp: 732-736.

Dev, H., T. Sen, M. Basak and M.E. Ali, 2012. An approach to protect the privacy of cloud data from data mining based attacks. Proceeding of 2012 SC Companion: High Performance Computing, Networking Storage and Analysis. Salt Lake City, UT, pp: 1106-1115.

Ding, X., Q. Yu, J. Li, J. Liu and H. Jin, 2013. Distributed Anonymization for Multiple Data Providers in a Cloud System. In: Meng, W. *et al.* (Eds.), Database Systems for Advanced Applications. Part I, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 7825: 346-360.

Friedman, A., R. Wolff and A. Schuster, 2008. Providing k-anonymity in data mining. VLDB J., 17(4): 789-804.

Gkountouna, O., S. Angeli, A. Zigomitros, M. Terrovitis and Y. Vassiliou, 2014. K^m -Anonymity for Continuous Data using Dynamic Hierarchies. In: Domingo-Ferrer, J. (Ed.), Privacy in Statistical Databases. Lecture Notes in Computer Science, Springer International Publishing, Switzerland, 8744: 156-169.

He, X., H.H. Chen, Y.F. Chen, Y.H. Dong, P. Wang and Z.H. Huang, 2012. Clustering-based k-Anonymity. In: Tan, P.N., S. Chawla, C.K. Ho and J. Bailey (Eds.), Advances in Knowledge Discovery and Data Mining. Part I, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 7301: 405-417.

He, X., D. Li, Y. Hao and H. Chen, 2014. Utility-friendly Heterogenous Generalization in Privacy Preserving Data Publishing. In: Yu, E. *et al*. (Eds.), Conceptual Modeling. Lecture Notes in Computer Science, Springer International Publishing, Switzerland, 8824: 186-194.

Karim, M.R., M.M. Rashid, B.S. Jeong and H.J. Choi, 2012. Privacy Preserving Mining Maximal Frequent Patterns in Transactional Databases. In: Lee, S.G. *et al.* (Eds.), Database Systems for Advanced Applications. Part I, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 7238: 303-319.

Liu, P., L. Cui and X. Li, 2014. A Hybrid Algorithm for Privacy Preserving Social Network Publication. In: Luo, K., J.X. Yu and Z. Li (Eds.), Advanced Data Mining and Applications. Lecture Notes in Computer Science, Springer International Publishing, Switzerland, 8933: 267-278.

Loukides, G. and A. Gkoulalas-Divanis, 2012. Utility-preserving transaction data anonymization with low information loss. Expert Syst. Appl., 39(10): 9764-9777.

Loukides, G., A. Gkoulalas-Divanis and J. Shao, 2012. Efficient and flexible anonymization of transaction data. Knowl. Inf. Syst., 36(1): 153-210.

Machanavajjhala, A., D. Kifer, J. Gehrke and M. Venkitasubramaniam, 2007. l-diversity: Privacy beyond k-anonymity. ACM T. Knowl. Discov. Data, 1(1): 1-12.

Monreale, A., D. Pedreschi, R.G. Pensa and F. Pinelli, 2014. Anonymity preserving sequential pattern mining. Artif. Intell. Law, 22(2): 141-173.

Nergiz, M.E. and M.Z. Gök, 2014. Hybrid k-anonymity. Comput. Secur., 44: 51-63.

Nergiz, M.E., C. Clifton and A.E. Nergiz, 2009. Multirelational k-anonymity. IEEE T. Knowl. Data En., 21(8): 1104-1117.

Nergiz, M.E., M.Z. Gök and U. Özkanli, 2013. Preservation of Utility through Hybrid k-anonymization. In: Furnell, S., C. Lambrinoudakis and J. Lopez (Eds.), Trust, Privacy and Security in Digital Business. Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 8058: 97-111.

Pan, Y., X.L. Zhu and T.G. Chen, 2012. Research on privacy preserving on K-anonymity. J. Softw., 7(7): 1649-1656.

Patil, B.B. and A.J. Patankar, 2013. Multidimensional k-anonymity for protecting privacy using nearest neighborhood strategy. Proceeding of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC, 2013). Enathi, pp: 1-4.

Samarati, P. and L. Sweeney, 1998. Protecting privacy when disclosing information: k-Anonymity and its enforcement through generalization and suppression. Proceeding of the IEEE Symposium on Research in Security and Privacy, 384-393.

Soodejani, A.T., M.A. Hadavi and R. Jalili, 2012. k-Anonymity-based Horizontal Fragmentation to Preserve Privacy in Data Outsourcing. In: Cuppens-Boulahia, N., F. Cuppens and J. Garcia-Alfaro (Eds.), Data and Applications Security and Privacy XXVI. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp: 263-273.

Sweeney, L., 2002a. Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertain. Fuzz., 10(5): 571-588.

Sweeney, L., 2002b. k-anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzz., 10(5): 557-570.

Tai, C.H., P.S. Yu and M. Chen, 2010. k-support anonymity based on pseudo taxonomy for outsourcing of frequent itemset mining. Proceeding of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp: 473-482.

Tai, C.H., J.W. Huang and M.H. Chung, 2013. Privacy preserving frequent pattern mining on multi-cloud environment. Proceeding of the International Symposium on Biometrics and Security Technologies (ISBAST). Chengdu, pp: 235-240.

Vijayarani, S., A. Tamilarasi and M. Sampoorna, 2010. Analysis of privacy preserving K-anonymity methods and techniques. Proceeding of the International Conference on Communication and Computational Intelligence (INCOCCI). Erode, pp: 540-545.

Wong, R.C.W., J. Li, A.W.C. Fu and K. Wang, 2006. (α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing. Proceeding of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp: 754-759.

Xu, Y., T. Ma, M. Tang and W. Tian, 2014. A survey of privacy preserving data publishing using generalization and suppression. Appl. Math. Inf. Sci., 8(3): 1103-1116.

Yang, J.J., J.Q. Li and Y. Niu, 2014. A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Gener. Comp. Sy., 43-44: 74-86.

Zhang, X., C. Liu, S. Nepal and J. Chen, 2013a. An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. J. Comput. Syst. Sci., 79(5): 542-555.

Zhang, X., C. Liu, S. Nepal, C. Yang, W. Dou and J. Chen, 2013b. Combining top-down and bottom-up: Scalable sub-tree anonymization over big data using mapreduce on cloud. Proceeding of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Melbourne, VIC, pp: 501-508.

Zhang, X., C. Liu, S. Nepal, C. Yang, W. Dou and J. Chen, 2014a. A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud. J. Comput. Syst. Sci., 80(5): 1008-1020.

Zhang, X., L.T. Yang, C. Liu and J. Chen, 2014b. A scalable two-phase top-down specialization approach for data anonymization using mapreduce on cloud. IEEE T. Parallel Distr., 25(2): 363-373.

Zhu, X.L. and T.G. Chen, 2012. Research on Privacy Preserving Based on k-Anonymity. In: He, X. *et al.* (Eds.), Computer, Informatics, Cybernetics and Applications. Lecture Notes in Electrical Engineering, Springer, Netherlands, 107: 915-923.

Zhu, Y. and L. Peng, 2007. Study on K-anonymity Models of Sharing Medical Information. Proceeding of the International Conference on Service Systems and Service Management. Chengdu, pp: 1-8.