academicJournals

Vol. 10(7), pp. 239-247, 16 April, 2015
DOI: 10.5897/IJPS11.1795
ISSN 1992 - 1950
Article Number: BF6C9B552057
Copyright © 2015
Author(s) retain the copyright of this article
http://www.academicjournals.org/IJPS

International Journal of Physical Sciences

Review

Privacy preserving data publishing: Review

Asmaa Hatem Rashid* and Norizan Binti Mohd Yasin

Department of Information Science, Faculty of Computer Science and IT, University of Malaya, Kuala Lumpur, 50603 KL, Malaysia.

Received 30 December, 2011; Accepted 24 November, 2014

Privacy preserving data publishing (PPDP) methods a new class of privacy preserving data mining (PPDM) technology, has been developed by the research community working on security and knowledge discovery. It is common to share data between two organizations in many application areas. When data are to be shared between parties, there could be some sensitive patterns which should not be disclosed to the other parties. These methods aims to keep the underlying data useful based on privacy preservation "utility based method based on privacy preservation, and created tremendous opportunities for knowledge- and information-based decision making. Recently, PPDP has received considerable attention in research communities, and many approaches have been proposed for different data publishing scenarios. In this survey, we will systematically summarize and evaluate different approaches to PPDP, study the challenges in practical data publishing, clarify the differences and requirements that distinguish PPDP from other related problems, and propose future research directions.

Key words: Privacy preserving, privacy preserving data publishing, privacy preserving data mining, republishing, security, privacy, decision making, knowledge.

INTRODUCTION

The development of IT and the collection of electronic information by data owners, such as governments, corporations, and individuals, have resulted in higher instances of data sharing. Many organizations are often willing to collaborate with other entities to perform a common action for mutual benefit (Gkoulalas-Divanis and Verykiosc, 2009; Qi and Zong, 2012). Driven by mutual benefits. Recent developments have helped improve decision making especially in the fields of medical information, research, and public health organization, among others. Many approaches have been proposed for different data publishing needs in different fields. Collaboration is an important factor in HISs (Ahmed and Yasin, 2012). According to Ohno-Machado (2013), privacy is an important requirement for collaboration in data sharing (Ohno-Machado, 2013). However, privacy concerns tend to become obstacles. According to Gkoulalas-Divanis and Loukides (2011) stated that 62% of patients were concerned about the disclosure of their EMRs. The sharing of data needs control and management to ensure system integration. Integration is required especially in the management of patient data to secure sensitive information such as patient identification.

*Corresponding author. E-mail: asmaarashid@siswa.um.edu.my

Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data privacy. Recently, PPDP has received considerable attention in research communities, and many approaches have been proposed for different data publishing scenarios (Gkoulalas-Divanis and Loukides, 2011). Several studies have focused on the management of data such as in medical applications to ensure system integration. However, the management and sharing of data in different fields but the challenge in misuse of information, and data owner's identification and others related problem. Privacy protection and data-keeping utility remain problems that must be solved (Gkoulalas-Divanis and Loukides, 2011). Information privacy in the healthcare sector is an issue of increasing importance. The adaption of healthcare HISs and the increasing need for information among patients, providers, and payers, all point toward the need for better information protection (Appari and Johnson, 2010). The frequency of identity theft continues to increase. Consequently, concerns about the ability of organizations to protect the personally identifiable data with which they are entrusted has also increased (Appari and Johnson, 2010; Fung et al., 2010).

In June 2004, the President's Information Technology Advisory Committee (PITAC) published a report entitled "Revolutionizing Health Care Through Information Technology" (Committee, 2004). One of the key points of this report was the establishment of a nationwide system of EHRs that encourages the sharing of medical knowledge through computer-assisted clinical decisions. Data publishing is equally ubiquitous in other domains. EHRs are a type of health IT that assist in storing health data and improve collaboration to provide better care. EHRs also reduce the necessity for paperwork by eliminating the need for paper-based records and by improving administrative efficiency, thereby decreasing healthcare costs. EHRs improve healthcare by decreasing medical errors with an assurance that all healthcare providers will have accurate and timely information (Bowman, 2012; Wu et al., 2006). For example, contracts and agreements cannot guarantee that sensitive data will not be carelessly misplaced and end up in the wrong hands.

A task of the utmost importance is to develop methods and tools for publishing data in a more hostile environment, so that the published data remains practically useful while individual privacy is preserved (Fung et al., 2010). This undertaking is called PPDP. In the past few years, research communities have responded to this challenge and proposed many approaches. While the research field is still rapidly developing, it is a good time to discuss the assumptions and desirable properties for PPDP, clarify the differences and requirements that distinguish PPDP from other related problems, and the current gaps and systematically summarize and evaluate different approaches to PPDP.

PRIVACY PRESERVING DATA PUBLISHING

PPDP provides methods and tools for publishing useful information while preserving data privacy (Chen et al., 2012; Fung et al., 2010). Recently, PPDP has received considerable attention in research communities, and many approaches have been proposed for different data publishing scenarios. According to Fung et al. (2010) a typical scenario for data collection and publishing (Fung et al., 2010), as show in Figure 1. In the data collection phase, the data publisher collects data from record owners (e.g., X1 and X2 to Xn). In the data publishing phase, the data publisher releases the collected data to a data miner or to the public, called the data recipient, who will then conduct data mining on the published data.

In this context, data mining has a broad sense, not necessarily restricted to pattern mining or model building. For example, a hospital collects data from patients and publishes the patient records to an external medical centre. In this example, the hospital is the data publisher, patients are record owners, and the medical centre is the data recipient. The data mining conducted at the medical centre could be anything from a simple count of the number of men with diabetes to a sophisticated cluster analysis. According to Gehrke (2006) proposed two models for privacy preserving data analysis and publishing (Gehrke, 2006).

(i) The untrusted model. The data publisher is not trusted and may attempt to identify sensitive information from record owners. Various cryptographic solutions (Yang et al., 2005), anonymous communications (Chaum, 1981; Jakobsson et al., 2002), and statistical methods (Warner, 1965) have been proposed to collect records anonymously from their owners without revealing the identities of the owners.

(ii) The trusted model. The data publisher is trustworthy, and record owners are willing to provide personal information to the data publisher. However, the trust is not transitive to the data recipient. Models of data publisher are described in Figure 2.

This study assume the trusted model of data publishers and consider privacy issues in the data publishing phase. According to Fung (2010) mentioned that in practice, every data publishing scenario has its own assumptions and requirements of the data publisher, the data recipients, and the data publishing purpose. The following are several desirable assumptions and properties in practical data publishing, according to Fung et al. (2010).

(1) The non expert data publisher.
(2) The data recipient could be an attacker.
(3) Publish data, not the data mining result.
(4) Truthfulness at the record level.

The initial idea of PPDM was to extend traditional data

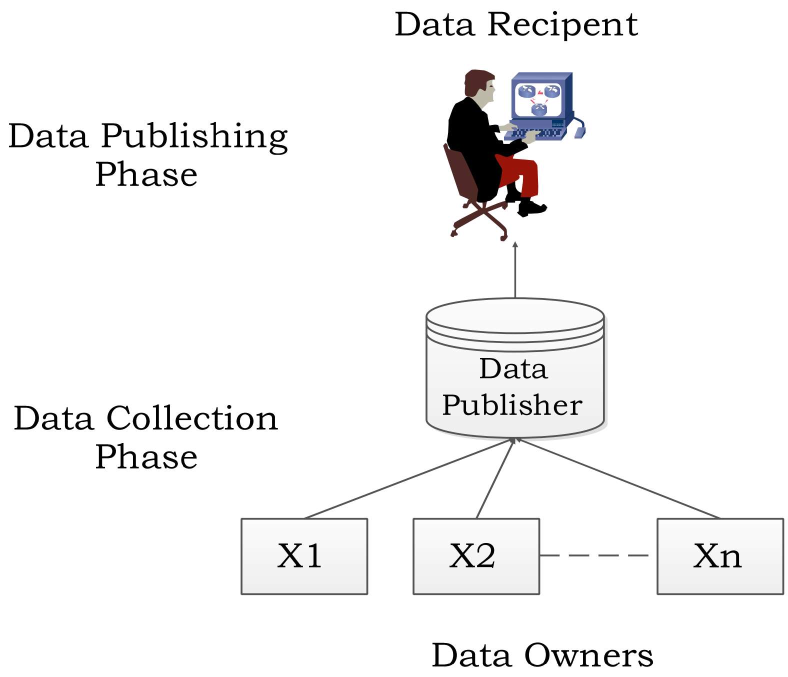

Figure 1. Scenario collection and publishing of data (Fung et al., 2010).

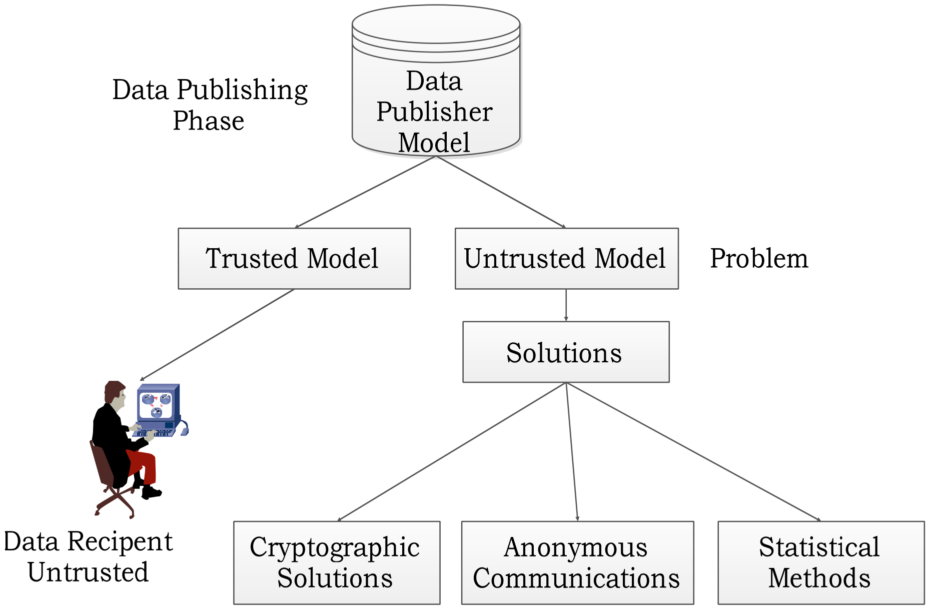

Figure 2. Models classification for data publishing.

mining techniques to work with the data modified to mask sensitive information. The key issues were how to modify the data and how to recover the data mining result from the modified data. The solutions were often tightly coupled with the data mining algorithms under consideration. In contrast, PPDP may not necessarily be tied to a specific data mining task, and the data mining task may be unknown at the time of data publishing. Furthermore, some PPDP solutions emphasize preserving the data truthfulness at the record level as discussed earlier, but often PPDM solutions do not preserve such a property in recent years, the term

Table 1. Comparison between PPDM and PDPP.

Variables	PPDM	PPDP
General Idea	PPDM is to allow data mining from a modified version of the data that contains no sensitive information	A new class of PPDM methods. PPDP allows the publication of useful information, while preserving data privacy (Benjamin et al., 2010; Gehrke, 2006). PPDP allow to anonymize the data by hiding identify of individuals, not hiding sensitive data.
Definition	Algorithms a new class of data mining methods, has been developed by the research community working on security and knowledge discovery (Bertino et al., 2005a; Fung et al., 2010).	Methods and tools for publishing useful information while preserving data privacy (Chen et al., 2012; Fung et al., 2010).
Aim	Extraction of relevant knowledge from large amounts of data, while protecting at the same time sensitive information (Bertino et al., 2005a).	Keep the underlying data useful based on privacy preservation "utility based method" (Fung et al., 2010).
Example	**Example to describe the scenario between them** A hospital may publish the patient data to a cancer research institute; although willing to contribute its data to cancer research, the hospital is not interested in and has expertise in data mining algorithms because cancer research is normal work.	
Demonstration	PPDM focuses on the data without sensitive information (Bertino et al., 2005b; Fung et al., 2007).	PPDP focuses on the data. Therefore, published records should be meaningful when examined individually (Chen et al., 2012).
Techniques	PPDM is to allow data mining techniques such as Association Rule Mining, Classification, Clustering (Fung et al., 2010)	PPDP seeks to anonymize the data by hiding identify of individuals, not hiding sensitive data. Hiding techniques such as k-anonymity, l-diversity, m-Invariance, T-Closeness (Fung et al., 2010).

"PPDM" has evolved to cover many other privacy research problems, even though some of them may not directly relate to data mining (Fung et al., 2010). Another related area is the study of the non-interactive query model in statistical disclosure control (Adam and Worthmann, 1989; Brand, 2002), in which the data recipients can submit one query to the system. This type of non-interactive query model may not fully address the information needs of data recipients because, in some cases, it is very difficult for a data recipient to accurately construct a query for a data mining task in one shot.

Consequently, there are a series of studies on the interactive query model (Blum et al., 2005; Dinur and Nissim, 2003; Dwork, 2008), in which the data recipients, unfortunately including attackers, can submit a sequence of queries based on previously received query results. One limitation of any privacy-preserving query system is that it can only answer a sub-linear number of queries in total; otherwise, an attacker (or a group of corrupted data recipients) will be able to reconstruct all but fraction of the original data (Blum et al., 2008), which is a very strong violation of privacy. When the maximum number of queries is reached, the system must be closed to avoid privacy leak. In the case of a non-interactive query model, the attacker can issue an unlimited number of queries and, therefore, a non-interactive query model cannot achieve the same degree of privacy defined by the interactive model. This study focuses mainly on the non-interactive query model (Fung et al., 2010), in this study; cover the review of recent studies on anonymization approaches to PPDP and provide our own insights into this topic. There are several fundamental differences between the recent work on PPDP and the previous work proposed by the official statistics community.

COMPARING BETWEEN (PPDM) AND (PPDP)

The general principle of this study is to release all data to facilitate the use of data sent or published in scientific fields, but the identities of people who are owners of such data and other sensitive properties found in the data must be protected. Therefore, the aim of this study falls outside the traditional work on access and authentication control (Sweeney, 2002). The latter area, PPDM and PPDP, explains the differences between two subjects. The results of the comparison are shown in Table 1.

CLASSIFYING THE PRIVACY PRESERVATION TECHNIQUES AND APPROACHES

The primary goal in privacy preserving is to protect the

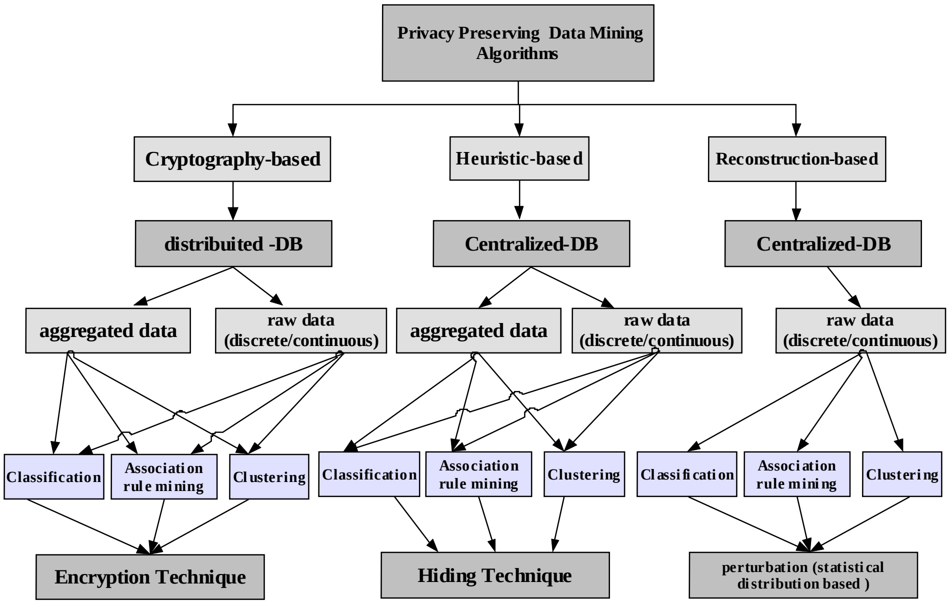

Figure 3. A taxonomy of the developed PPDM algorithms (Bertino et al., 2005b).

sensitive data before it is released for analysis or re-publishing. However, the data may reside at centralized or distributed data storage. In such a scenario appropriate algorithms or techniques should be used which preserves any sensitive information in the knowledge discovery process. To address this issue there are many approaches adopted for privacy preserving data mining (Parmar et al., 2011). Classifying the proposed privacy preservation techniques according to five different dimensions:

(1) Data distribution (centralized or distributed);
(2) The modification applied to the data (encryption, perturbation, generalization, and so on) in order to sanitize them;
(3) The data mining algorithm which the privacy preservation technique is designed for;
(4) The data type (single data items or a complex data correlation) that needs to be protected from disclosure;
(5) The approach adopted for preserving privacy (heuristic, reconstruction or cryptography-based approaches).

Figure 3 shows taxonomy of the existing PPDM algorithms according to those dimensions. Obviously, it represents a first organization in this new area and does not cover all the possible PPDM algorithms. However, it gives one overview of the algorithms that have been proposed so far, focusing on their main features. While heuristic and reconstruction-based techniques are mainly conceived for centralized datasets, cryptography based algorithms are designed for protecting privacy in a distributed scenario by using encryption techniques. Reconstruction-based algorithms recently proposed aim at hiding sensitive raw data by applying perturbation techniques based on probability distributions. Moreover, several heuristic-based approaches for hiding both raw and aggregated data through a hiding technique (perturbation, blocking, data swapping, aggregation, generalization and sampling) have been developed, first, in the context of association rule mining and classification and, more recently, for clustering techniques. Now, we briefly describe some of the algorithms proposed in the PPDM area (Bertino and Sandhu, 2005). Figure 4 show the approaches of privacy preserving data mining based on the above dimensions.

THE PRIVACY MODELS IN PRIVACY PRESERVING DATA PUBLISHING

The privacy protection is important issues when related with personal data we need to provide stringent definition about protection of privacy. The clear definition: access to the published data should not enable the attacker to learn anything extra about any target victim compared to no access to the database, even with the presence of any attacker’s background knowledge obtained from other sources (Dalenius, 1977).

Most literature on PPDP considers a more relaxed, more practical notion of privacy protection by assuming the attacker has limited background knowledge. A privacy threat occurs either when an identify is linked to a record or when an identify is linked to a value on some sensitive, these threats are called record linkage, attribute linkage, table linkage. Below, we can broadly classify privacy

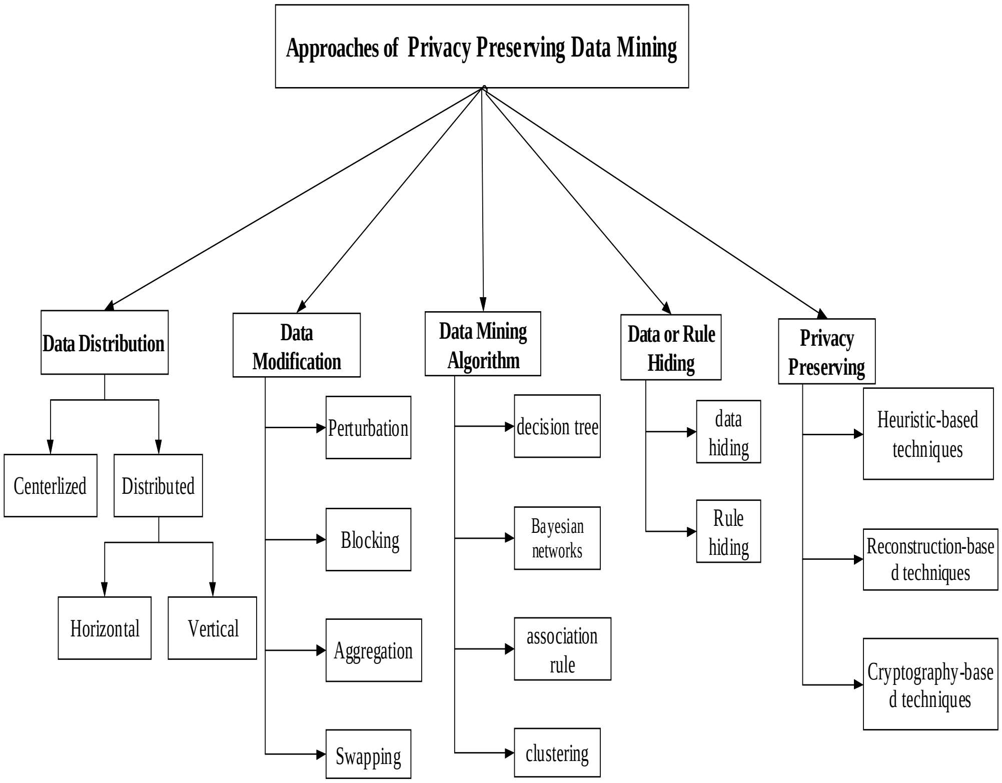

Figure 4. The approaches of privacy preserving data mining.

models into two categories based on their attack principles. Attack; refer to unauthorized access to this data. The victim refers to data owner targeted by the attacker. We can broadly classify privacy models into two categories based on their attack principles (Fung et al., 2010)

The first category considers that a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call these record linkage, attribute linkage, and table linkage, respectively. In all three types of linkages, we assume that the attacker knows the QID of the victim. In record and attribute linkages, we further assume that the attacker knows that the victim's record is in the released table, and seeks to identify the victim's record and/or sensitive information from the table. In table linkage, the attack seeks to determine the presence or absence of the victim's record in the released table. A data table is considered to be privacy preserving if it can effectively prevent the attacker from successfully performing these linkages (Fung et al., 2010). The second category aims at achieving the uninformative (not providing knowledge) principle: The published table should provide the attacker with little additional information beyond the background knowledge. If the attacker has a large variation between the prior and posterior beliefs, we call it the probabilistic attack (Machanavajjhala et al., 2007). Many privacy models in this family do not explicitly classify attributes in a data table into QID and sensitive attributes, but some of them could also thwart the sensitive linkages in the first category, so the two categories overlap (Machanavajjhala et al., 2007). The following Table 2 summarizes the attack models addressed by the privacy models.

TYPES OF LINKAGES

Record linkage

In the attack of record linkage, some value qid on QID identifies a small number of records in the released table T, called a group. If the victim's QID matches the value qid, the victim is vulnerable to being linked to the small number of records in the group. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group (Fung et al., 2010).

Attribute linkage

According to Fung (2010) in the attack of attribute

Table 2. Privacy models in privacy preserving data publishing (Fung et al., 2010).

Privacy models	Attack models			
	Record linkage	**Attribute linkage**	**Table linkage**	**Probabilistic attack**
k-Anonymity	/			
Multi R k-Anonymity	/			
ℓ Diversity	/	/		
Confidence Bounding		/		
(a; k)-Anonymity	/	/		
(X; Y)-Privacy	/	/		
(k; e)-Anonymity		/		
(€;m)-Anonymity		/		
Personalized Privacy		/		
t-Closeness		/		/
£, Presence			/	
(c; t)-Isolation	/			/
E-Differential Privacy			/	/
(d; y)-Privacy			/	/
Distributional Privacy			/	/

linkage, the attacker may not precisely identify the record of the target victim, but could infer his/her sensitive values from the published data T, based on the set of sensitive values associated to the group that the victim belongs to. In case some sensitive values predominate in a group, a successful inference becomes relatively easy even if k-anonymity is satisfied. According to Clifton et al. (2002) suggested eliminating attribute linkages by limiting the released data size. Limiting data size may not be desirable if data records such as HIV patient data are valuable and are difficult to obtain (Clifton et al., 2002). Several other approaches have been proposed to address this type of threat. The general idea is to diminish the correlation between QID attributes and sensitive attributes.

Table linkage

Both record linkage and attribute linkage assume that the attacker already knows the victim's record is in the released table T. However, in some cases, the presence or the absence of the victim's record in T already reveals the victim's sensitive information. Suppose a hospital releases a data table with a particular type of disease. Identifying the presence of the victim's record in the table is already damaging. A table linkage occurs if an attacker can confidently infer the presence or the absence of the victim's record in the released table (Fung et al., 2010).

Probabilistic linkage

There is another family of privacy models that does not focus on exactly what records, attributes, and tables the attacker can link to a target victim, but focuses on how the attacker would change his/her probabilistic belief on the sensitive information of a victim after accessing the published data. In general, this group of privacy models aims at achieving the uninformative principle, whose goal is to ensure that the difference between the prior and posterior beliefs is small (Fung et al., 2010; Machanavajjhala et al., 2007). In sum, the privacy models in privacy preserving data publishing based on the linkage types.

RESULTS AND DISCUSSION

Privacy rights include the collection, storage and usage of personal data have only been partially protected under a variety of context-specific privacy laws, the protection of data privacy is an important problem that organizations must solve (LeFevre et al., 2006). As the frequency of identity theft continues to increase, there are increasing concerns about the competency of the organization's ability to protect the personally identifiable data the organization is entrusted with.

The problem has three main areas that are combined to create a Personally Identifiable Information Program (PIIP). The areas that must be considered are privacy, data security programs, and authentication of the requester. It is not so far behind as the privacy preserving topic is still a hot information technology and hinders the development of health information technology issue. From the findings of the literatures above, there were several gaps in the privacy preserving subject, in the form of: (a) when data are to be shared between parties, there could be some sensitive patterns which should not be disclosed to the other parties'. Many methods have been proposed for privacy preserving in various fields. They have monitored and analyze .the accuracy of the performance is still poor, because the methods not provide high level from privacy, efficiency, data quality, and negatively affects the accuracy of such methods performance. Moreover; sharing data will bring the problem of misuse. This is the main drawback of the privacy preserving of data. This research will address this main drawback through analyzing and evaluating the following sub-gaps.

(1) There is no model that can identify the number of quasi identifier attributes in such a way that protect the privacy of original data and keep the new version of data usable.
(2) There is a lack of connectivity between providers the health care.
(3) Beside the lack of performing centralized database to keep the confidentiality and privacy of data or to collect data, the problem of case indexing still not solved.
(4) The lack of high quality of data and the possibility of errors that adversely affect the results of researches and studies, which depend on the new version data.

CONCLUSION

The information sharing has become part of the routine activity of many individuals, companies, organizations, and government agencies. Privacy-preserving data publishing is a promising approach to information sharing, while preserving individual privacy and protecting sensitive information. Recent developments have helped improve decision making especially in the fields of medical information, research, and public health organization. Privacy protection is a complex social issue, which involves policy-making, technology, psychology, and politics. Finally, we emphasize that privacy-preserving technology solves only one side of the problem. It is equally important to identify and overcome the nontechnical difficulties faced by decision makers when they deploy a privacy-preserving technology. Their typical concerns include the degradation of data/service quality, loss of valuable information, increased costs, and increased complexity. The findings and reviews outlined in this chapter have indeed contributed to the researcher's study and in depth understanding of the subject matter. This has served as the impetus needed to further the research and ultimately meet the research objectives stated in the beginning.

Conflict of Interest

The authors have not declared any conflict of interest.

ACKNOWLEDGMENTS

The researchers appreciate the role of the Cancer Treatment Institute, Cairo University - Egypt. We are also grateful to all the participants in this study and to the University of Malaya for their interest in and support for this research.

REFERENCES

Adam NR, Worthmann JC (1989). Security-control methods for statistical databases: A comparative study. ACM Computing Surveys (CSUR). 21(4):515-556.

Ahmed NS, Yasin NM (2012). Improvement the cooperation feature in distributed healthcare information systems based on the fractal approach: An empirical study. Adv. Mater. Res. 463:861-867.

Appari A, Johnson ME (2010). Information security and privacy in healthcare: current state of research. Int. J. Internet enterprise manage. 6(4):279-314.

Bertino E, Fovino IN, Provenza LP (2005a). A framework for evaluating privacy preserving data mining algorithms. Data Mining Knowledge Discovery. 11(2):121-154. doi: 10.1007/s10618-005-0006-6.

Bertino E, Fovino IN, Provenza LP (2005b). A framework for evaluating privacy preserving data mining algorithms*. Data Mining Knowledge Discovery. 11(2):121-154.

Bertino E, Sandhu R (2005). Database security - Concepts, approaches, and challenges. IEEE Trans. Dependable Secure Computing. 2(1):2-19.

Blum A, Dwork C, McSherry F, Nissim K (2005). Practical privacy: the SuLQ framework. In: Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. pp. 128-138.

Blum A, Ligett K, Roth A (2008). A learning theory approach to non-interactive database privacy. In: Proceedings of the 40th ACM SIGACT Symposium on Thoery of Computing. pp. 609-618.

Bowman S (2012). Impact of Electronic Health Record Systems on Information Integrity: quality and safety implications. Perspectives in Health Information Management. P. 10.

Brand R (2002). Microdata protection through noise addition. Infer. Contr. Statistical Databases pp. 61-74.

Chaum DL (1981). Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM. 24(2):84-90.

Chen L, Yang JJ, Wang Q, Niu Y (2012). A framework for privacy-preserving healthcare data sharing. Paper presented at the e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on; 01/2012.

Clifton C, Kantarcioglu M, Vaidya J, Lin X, Zhu MY (2002). Tools for privacy preserving distributed data mining. ACM Sigkdd Explorations Newsletter 4(2):28-34.

Committee PITAC (2004). Revolutionizing health care through information technology. Report to the President of the United States.

Dalenius T (1977). Towards a methodology for statistical disclosure control. Statistik Tidskrift, 15(429-444):2-1.

Dinur I, Nissim K (2003). Revealing information while preserving privacy. In: Proceedings of ACM PODS, pp.202–210.

Dwork C (2008). Differential privacy: A survey of results. Theory Appli. Models Comput. pp.1-19.

Fung BCM, Wang K, Chen R, Yu PS (2010). Privacy-preserving data publishing: A survey on recent developments. Computing. 5(4):1-53.

Fung BCM, Wang K, Yu PS (2007). Anonymizing classification data for privacy preservation. IEEE Trans. Knowledge Data Eng. 19(5):711-725. doi: 10.1109/tkde.2007.1015.

Gehrke J (2006). Models and methods for privacy-preserving data analysis and publishing. In ICDE 2006: Proceedings Of The Twenty second International Conference on Data Engineering, P. 105. IEEE Computer Society, Washington, Dc, USA.

Gkoulalas-Divanis A, Loukides G (2011). Medical Data Sharing: Privacy Challenges and Solutions. Dublin, Ireland. https://books.google.com.ng/books?id=HFLXlOdkpD8C&pg=PR6&dq

=2011.+Medical+Data+Sharing:+Privacy+Challenges+and+Solutions.&hl=en&sa=X&ei=QqMKVYfHEIfcatSdgcgF&ved=0CCQQ6AEwAA#v=onepage&q=2011.%20Medical%20Data%20Sharing%3A%20Privacy%20Challenges%20and%20Solutions.&f=false

Gkoulalas-Divanis A, Verykiosc VS (2009). An overview of privacy preserving data mining. Crossroads 15(4):6.

Jakobsson M, Juels A, Rivest RL (2002). Making mix nets robust for electronic voting by randomized partial checking. http://people.csail.mit.edu/rivest/JakobssonJuelsRivest-MakingMixNetsRobustForElectronicVotingByRandomizedPartialChecking.pdf

LeFevre K, DeWitt DJ, Ramakrishnan R (2006). Mondrian multidimensional k-anonymity. In: Proc. of the 22nd International conference on Data Engineering (ICDE, 2006). IEEE. pp. 25-35.

Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M (2007). l-diversity: Privacy beyond k-anonymity. ACM Trans. Knowledge Discovery Data (TKDD). 1(1):3.

Ohno-Machado L (2013). Sharing data for the public good and protecting individual privacy: informatics solutions to combine different goals. J. Am. Medical Info. Association. 20(1):1-1.

Parmar AA, Rao UP, Patel DR (2011). Blocking based approach for classification Rule hiding to preserve the Privacy in Database. Paper presented at the Computer Science and Society (ISCCS), 2011 International Symposium on; 08/2011

Qi X, Zong M (2012). An Overview of Privacy Preserving Data Mining. Procedia. Environ. Sci. 12:1341-1347.

Sweeney L (2002). k-anonymity: A model for protecting privacy. Int. J. Uncertainty Fuzziness Knowledge Based Syst. 10(5):557-570.

Warner SL (1965). Randomized response: A survey technique for eliminating evasive answer bias. J. Am. Statistical Association. 63-69.

Wu S, Chaudhry B, Wang J, Maglione M, Mojica W, Roth E, Shekelle PG (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. Annals internal medicine. 144(10):742-752.

Yang Z, Zhong S, Wright RN (2005). Anonymity-preserving data collection. In Proc. eleventh ACM SIGKDD int. Confer. Knowledge discovery data mining ACM. pp. 334-343.