Information Systems 36 (2011) 859–880

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

ASAP: Eliminating algorithm-based disclosure in privacy-preserving data publishing

Xin Jin [a], Nan Zhang [a,1,*], Gautam Das [b,2]

[a] Department of Computer Science, George Washington University, 20052, United States
[b] Department of Computer Science and Engineering, University of Texas at Arlington, 76019, United States

ARTICLE INFO

Article history:
Received 25 August 2010
Received in revised form
10 January 2011
Accepted 8 March 2011
Recommended by L. Wong
Available online 15 March 2011

Keywords:
Privacy preservation
Data publishing
Algorithm-based disclosure
Algorithm-SAfe Publishing

ABSTRACT

Numerous privacy-preserving data publishing algorithms were proposed to achieve privacy guarantees such as ℓ-diversity. Many of them, however, were recently found to be vulnerable to algorithm-based disclosure—i.e., privacy leakage incurred by an adversary who is aware of the privacy-preserving algorithm being used. This paper describes generic techniques for correcting the design of existing privacy-preserving data publishing algorithms to eliminate algorithm-based disclosure. We first show that algorithm-based disclosure is more prevalent and serious than previously studied. Then, we strictly define *A*lgorithm-*SA*fe *P*ublishing (ASAP) to capture and eliminate threats from algorithm-based disclosure. To correct the problems of existing data publishing algorithms, we propose two generic tools to be integrated in their design: *global look-ahead* and *local look-ahead*. To enhance data utility, we propose another generic tool called *stratified pick-up*. We demonstrate the effectiveness of our tools by applying them to several popular ℓ-diversity algorithms: Mondrian, Hilb, and MASK. We conduct extensive experiments to demonstrate the effectiveness of our tools in terms of data utility and efficiency.

1. Introduction

1.1. Privacy-preserving data publishing

Many organizations, such as hospitals, require publishing microdata with personal information, such as medical records, for facilitating research and serving public interests. Nonetheless, such publication may incur privacy concerns for the individual owners of tuples being published (e.g., patients). To address this challenge, privacy-preserving data publishing (i.e., PPDP) was proposed to generate the published table in a way that enables analytical tasks (e.g., aggregate query answering, data mining) over the published data, while protecting the privacy of individual data owners.

In general, a *microdata* table (denoted by T) can contain three types of attributes: (1) *personal identifiable* attributes (e.g., *SSN*), each of which is an explicitly unique identifier of an individual, (2) *quasi-identifier* (QI) attributes (e.g., *Age*, *Sex*, *Country*), which are not explicit identifiers but, when combined together, can be empirically unique for each individual, and (3) *sensitive attributes* (SA) (e.g., *Disease*), each of which contains a sensitive value (set) that must be protected. In privacy-preserving data publishing, personal identifiable attributes are usually removed prior to publishing. QI and/or SA attributes are perturbed to achieve a pre-defined privacy model while maximizing the utility of published data.

Samarati and Sweeney [1] first defined a privacy model, k-anonymity, for PPDP. It requires each tuple in the published table (denoted by T^*) to have at least $k-1$ other QI-indistinguishable tuples—i.e., tuples with the same QI attribute values. To protect individual SA information, Machanavajjhala et al. [2] introduced another

* Corresponding author. Tel.: +1 2029945919; fax: +1 2029944875.
E-mail addresses: xjin@gwu.edu (X. Jin), nzhang10@gwu.edu (N. Zhang), gdas@uta.edu (G. Das).

[1] Partially supported by NSF grants 0852673, 0852674, 0845644 and 0915834 and a GWU Research Enhancement Fund.
[2] Partially supported by NSF grants 0845644, 0812601 and 0915834 and grants from Microsoft Research and Nokia Research.

0306-4379/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.is.2011.03.001

privacy model, ℓ-diversity, which further requires each group of QI-indistinguishable tuples to have diverse SA values. Variations of the ℓ-diversity include (α,k)-anonymity [3], t-closeness [4], (k, e)-anonymity [5], m-invariance [6], etc. To satisfy these privacy models, numerous PPDP algorithms have been proposed [5,7–11].

1.2. Algorithm-based disclosure

It was traditionally believed that, to determine whether a privacy model is properly satisfied, one only needs to look at the published table, i.e., the output of a data publishing algorithm, without investigating the algorithm itself. The recently discovered algorithm-based disclosure [12] contradicts this traditional belief as it demonstrates that privacy disclosure can be incurred by the *design* of a data publishing algorithm. In particular, if a privacy-preserving algorithm is vulnerable to algorithm-based disclosure, then once an adversary learns the design of the algorithm, s/he may utilize this knowledge to reverse-engineer the published table to compromise additional private information. We shall discuss the details in Section 2.

Algorithm-based disclosure poses a significant threat to the privacy of published data, because the data publishing algorithm is usually considered public and may be learned by an adversary. One might argue that, given the large number of public algorithms that are available for PPDP, it is difficult for an adversary to precisely identify which algorithm has been used and thereby to launch the algorithm-based attack. This is a typical "security through obscurity" argument which counts on the secrecy of an algorithm to ensure the security of its output. However, such arguments have been repeatedly argued against and aborted in the literature of security and cryptography. As Kerckhoff's principle [13] in cryptography states, "The cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience." Similarly, we argue that, to design an effective algorithm for privacy-preserving data publishing, one must eliminate algorithm-based disclosure.

1.3. Existing work and limitations

Wong et al. [12] demonstrated the first known case of algorithm-based disclosure by showing that the minimality principle used by many existing algorithms, i.e., to perturb QI with the minimum degree possible for satisfying the privacy model, may lead to the disclosure of private SA information when the adversaries have the original QI as external knowledge. An example of this disclosure will be described in Section 2. To counteract this attack, Wong et al. proposed a new privacy model called m-confidentiality [12], which guarantees that even an adversary with knowledge of QI cannot have confidence of more than $1/m$ on the SA value of an individual tuple. This attack was also studied in [14], with a new privacy model p-safety proposed as a countermeasure.

The new privacy models studied in the existing work, i.e., m-confidentiality [12] and p-safety [14], are by definition safe against (at least certain types of) algorithm disclosure. In addition, some recently proposed privacy models such as differential privacy [15] are also by definition immune from algorithm-based disclosure. While defining these new privacy models and developing their corresponding new algorithms provides a clean-slate solution for eliminating algorithm-based disclosure, limiting the investigation of algorithm-based disclosure to this realm has a number of problems.

First, the state-of-the-art PPDP calls for a proper understanding of the scope of algorithm-based disclosure for the existing data publishing algorithms. Currently, unless a data publishing algorithm is designed for an inherently algorithm-disclosure-safe privacy model such as differential privacy, it is unclear how to determine whether the algorithm is vulnerable to algorithm-based disclosure. Meanwhile, there are considerable ongoing efforts [16,17] on developing data publishing algorithms for popular privacy models such as ℓ-diversity which do not provide such definition-inherent guarantee against algorithm-based disclosure. To enable the safe deployment of these algorithms in practice, it is important to understand whether and how algorithm-based disclosure may occur for a given data publishing algorithm.

Furthermore, the wide prevalence of data publishing algorithms calls for a generic method to revise the design of an existing algorithm for eliminating algorithm-based disclosure. In the literature, for popular privacy models such as ℓ-diversity, there have been not only a myriad of algorithms for publishing tabular data, but also numerous others that publish application-specific data such as location [18], social network [19], and transaction information [20]. Instead of re-inventing algorithms for all these applications, we argue that a more cost-effective way is to develop a generic method that eliminates algorithm-based disclosure from the existing algorithms.

1.4. Outline of technical results

In this paper, we attack the problem of algorithm-based disclosure from a novel algorithmic angle. In particular, we first illustrate the challenge of identifying algorithm-based disclosure by demonstrating that the space of such disclosure is substantially larger than previously recognized. Then, we provide a testing tool to determine whether a given data publishing algorithm is subject to algorithm-based disclosure. Finally, we develop two tools, *global look-ahead* and *local look-ahead*, to revise the design of existing data publishing algorithms for eliminating algorithm-based disclosure. To recover the utility loss incurred by applying these tools, we develop *stratified pick-up*, another tool to retain a high level of utility for the published table.

Our detailed results can be stated as follows:

First, we find that the space of algorithm-based disclosure is much broader than previously discovered. While the previous work identifies algorithm-based disclosure when an adversary holds external knowledge about the QI attributes, we find that other forms of external knowledge, such as the distribution of SA values and/or certain negative association rules [21] can also give rise to algorithm-based disclosure. Our further investigation even eliminates the dependency of algorithm-based disclosure on external