Chapter 79
A Survey on Privacy Preserving Dynamic Data Publishing

Salheddine Kabou

LabRI Laboratory, Ecole Superieure en Informatique, Sidi Bel-Abbes, Algeria

Sidi mohamed Benslimane

https://orcid.org/0000-0002-7008-7434

LabRI Laboratory, Ecole Superieure en Informatique, Sidi Bel-Abbes, Algeria

Mhammed Mosteghanemi

Ecole Nationale Supérieure d'Informatique, Bab Ezzouar, Algeria

ABSTRACT

Many organizations, especially small and medium business (SMB) enterprises require the collection and sharing of data containing personal information. The privacy of this data must be preserved before outsourcing to the commercial public. Privacy preserving data publishing PPDP refers to the process of publishing useful information while preserving data privacy. A variety of approaches have been proposed to ensure privacy by applying traditional anonymization models which focused only on the single publication of datasets. In practical applications, data publishing is more complicated where the organizations publish multiple times for different recipients or after modifications to provide up-to-date data. Privacy preserving dynamic data publication PPDDP is a new process in privacy preservation which addresses the anonymization of the data for different purposes. In this survey, the author will systematically evaluate and summarize different studies to PPDDP, clarify the differences and requirements between the scenarios that can exist, and propose future research directions.

1. INTRODUCTION

Nowadays, Government regulations and many organizations, especially Small and Medium Business (SMB) enterprises require the collection, exchange and sharing of enormous repositories of digital information. In the case where information contains personally identifiable information, such data sharing is subject to constraints imposed by security and privacy of data owners (Chang et al., 2016a). In the study of (Fung et al., 2010), authors announce that 87% of the population of the United States can be uniquely identified by a given dataset published for the public, which extremely reflects privacy violation in the publishing scenario. In August 2006, America Online (AOL) published 20 Million anonymous logs of search queries collected from 658,000 users to facilitate information retrieval research for academic purposes, after mapping each user to a randomly generated identifier (Adeel, 2013). The privacy of this data which is the number one factor for security based on 400 IT professionals' opinions (Chang et al., 2016b) must be preserved, i.e. any sensitive information should not be disclosed to guarantee that individuals privacy cannot be inferred from dataset directly. Keeping and improving security and privacy is also essential for all users and services, such for the Internet Of Things and the Big Data paradigms (Yang et al., 2018; Kuo et al., 2018).

The most important task is to develop methods and tools for publishing data as a remedy of this awkward situation for finding the right balance between data utility and information privacy when publishing dataset. This area of research is called privacy-preserving data publishing (PPDP), which can be considered as a technical answer to complement the privacy approaches. Data anonymization is one of the privacy preserving techniques that translate the information making the original data worthless to anybody except the owners (Kabou and Benslimane, 2015).

It has been widely discussed in the literature such as k-anonymity (Samarati and Sweeney, 1998) (Sweeney, 2002), l-diversity (Machanavajjhala et al., 2006), k-concealment (Tassa et al., 2012).Since the appearance of k-anonymization, several privacy preserving models have been proposed, generally known as Privacy-Preserving Static Data Publishing PPSDP which ensure privacy protection up to a certain level i.e., they are focused on single publication of datasets.

In practical applications, data publishing is more complicated. For example, the organizations can publish a dataset multiple times for different recipients statically or after modifications (insertions, deletions or update) for providing up-to-date data. Each time, the data is anonymized differently for different purposes, or the data is published incrementally as new data is collected. In dynamic data publication problem, the above-mentioned paradigms could provide protection pertaining to a single release. This need opens a new era in privacy preservation called privacy preserving dynamic data publication. (Adeel, 2013).

1.1. Privacy-Preserving Dynamic Data Publishing

Like privacy-preserving static data publishing (Figure 2), the process of privacy preserving dynamic data publication has two phases, data collection and data publish phase (Figure 1). In the data collection phase, the data holder (who could be corporate organizations and institutions, private companies, etc.) collects and stores records of their clients (record owners) in an electronic storehouse due to some business or social interactions.

Related Content

Internet Voting: Beyond Technology

Trisha Woolleyand Craig Fisher (2004). *Information Security and Ethics: Social and Organizational Issues (pp. 117-138).*

www.irma-international.org/chapter/internet-voting-beyond-technology/23347

The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty

Jonathan Ford (2012). *International Journal of Risk and Contingency Management (pp. 75-77).*

www.irma-international.org/article/flaw-averages-underestimate-risk-face/70234

A Visualization Dashboard for COVID-19 Tweets Sentiment Analysis

Devang Pathak, Ishita Kumar, Maheswari Rajaand Carol Anne Hargreaves (2022). *Handbook of Research on Technical, Privacy, and Security Challenges in a Modern World (pp. 223-242).*

www.irma-international.org/chapter/a-visualization-dashboard-for-covid-19-tweets-sentiment-analysis/312424

Role of Artificial Intelligence on Cybersecurity and Its Control

Ramesh Chandra Rath, Sukanta Kumar Baraland Richa Goel (2022). *Cross-Industry Applications of Cyber Security Frameworks (pp. 15-35).*

www.irma-international.org/chapter/role-of-artificial-intelligence-on-cybersecurity-and-its-control/306790

Personalized Key Drivers for Individual Responses in Regression Modeling

Stan Lipovetsky (2020). *International Journal of Risk and Contingency Management (pp. 15-30).*

www.irma-international.org/article/personalized-key-drivers-for-individual-responses-in-regression-modeling/252179