Chapter 99
Amelioration of Anonymity Modus Operandi for Privacy Preserving Data Publishing

J. Indumathi
Anna University, India

ABSTRACT

The scientific tumultuous intonation has swept our feet's, of its balance and at the same time wheedled us to reach the take-off arena from where we can march equipped and outfitted into the subsequent century with confidence & self-assurance; by unearthing solutions for all information security related issues (with special emphasis on privacy issues). Examining various outstanding research problems that encompass to be embarked upon for effectively managing and controlling the balance between privacy and utility, the research community is pressurized to propose suitable elucidations. The solution is to engender several Privacy-Preserving Data Publishing (PPDP) techniques like Perturbation, swapping, randomization, cryptographic techniques etc., Amongst the various available techniques k-anonymity is unique in facet of its association with protection techniques that preserve the truthfulness of the data. The principal chip in of this sketch out comprises: 1) Motivation for this exploration for Amelioration Of Anonymity Modus Operandi For Privacy Preserving Data Mining; 2) investigation of well-known research approaches to PPDM; 3) argue solutions to tackle the problems of security threats and attacks in the PPDM in systems; 4) related survey of the various anonymity techniques; 5) exploration of metrics for the diverse anonymity techniques; 6) performance measures for the various anonymity techniques; and 7) contradistinguish the diverse anonymity techniques and algorithms.

1. INTRODUCTION

Everyone designs who devises courses of action aimed at changing existing situations into preferred ones — Herbert Simon (1996)

Knowledge-based decision making is growing leaps and bounds owing to the fabulous escalation in technology development for collection of digital information by governments, corporations, and individuals. There is an incessant stipulation

DOI: 10.4018/978-1-4666-9562-7.ch099

for data mining and data publishing. For example, licensed hospitals in California are mandatory to put forward detailed demographic data on every patient discharged from their facility (Carlisle et al. 2007). In June 2004, the Information Technology Advisory Committee released a report entitled *Revolutionizing Health Care through Information Technology* (President Information Technology Advisory Committee 2004). This pointed out the sharing of a nation-wide medical knowledge database through computer-assisted clinical decision support. Data publishing is equally omnipresent in other domains. For example, Netflix, a popular online movie rental service, recently published a data set containing movie ratings of 500,000 subscribers, in a drive to improve the accuracy of movie recommendations based on personal preferences (New York Times, Oct. 2, 2006); AOL published a release of query logs but quickly removed it due to the re-identification of a searcher (Barbaro and Zeller 2006).

2. IMPETUS FOR PPDP

Currently, the so-called privacy fortification modus operandi is deficient in privacy protection. They simply remove the explicit identifier of the record holders before releasing the data. L.Sweeney (2002) performed to show a privacy attack on William Weld, who is a former governor of the state of Massachusetts. By linking a voter list with some publicly available medical data on some shared quasi-identifier namely zip code, date of birth, and sex together with his medical information like diagnosis and medication he easily identified Weld's name. Weld's case is not an unexpected occurrence because L.Sweeney (2002) additionally pointed out that 87% of the U.S. population had reported characteristics that likely made them unique based on only such quasi-identifier.

This type of attack is called the *linking attack*, where the attacker needs two pieces of a priori knowledge: (1) the record holder is (likely to be) involved in the released data, and (2) the quasi-identifier of the record holder. By simple observation and common sense we can obtain this priori knowledge. For example, knowing his boss was absent for staying in a hospital, the attacker knew that his boss' medical information would appear in the released medical data from that hospital. For example, Maclean's (2005) was able to purchase months of phone logs of Jennifer Stoddart, who is the privacy commissioner of the federal government of Canada, from a U.S. data broker for US$200. The requested information returned within several hours and it even included an updated monthly statement which Stoddart herself had not received yet.

The current practice of privacy protection first and foremost focuses on policies and guidelines to confine the types of publishable data and on concords on the use and storage of sensitive data. This has *serious limitations* like either it deforms data excessively or requires a trust level that is impractically high in many data-sharing scenarios.

The need of the hour is to press the research communities into service to devise methods of privacy protection, which, simultaneously publishes the data in hostile environments and preserves it. We already have two approaches PPDM and PPDP. We will start to discuss the assumptions and desirable properties for PPDP, clarify the differences and requirements that distinguish PPDP from other related problems, 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 and is focused on preventing this kind of linking attack.