Chapter 25
A New View of Privacy in Social Networks:
Strengthening Privacy During Propagation

Wei Chang
Saint Joseph's University, USA

Jie Wu
Temple University, USA

ABSTRACT

Many smartphone-based applications need microdata, but publishing a microdata table may leak respondents' privacy. Conventional researches on privacy-preserving data publishing focus on providing identical privacy protection to all data requesters. Considering that, instead of trapping in a small coterie, information usually propagates from friend to friend. The authors study the privacy-preserving data publishing problem on a mobile social network. Along a propagation path, a series of tables will be locally created at each participant, and the tables' privacy-levels should be gradually enhanced. However, the tradeoff between these tables' overall utility and their individual privacy requirements are not trivial: any inappropriate sanitization operation under a lower privacy requirement may cause dramatic utility loss on the subsequent tables. For solving the problem, the authors propose an approximation algorithm by previewing the future privacy requirements. Extensive results show that this approach successfully increases the overall data utility, and meet the strengthening privacy requirements.

INTRODUCTION

Learning others' social features can significantly improve the performance of many mobile social network-related tasks, such as data routing (Wu & Wang, 2012), personalized recommendation (Feng & Wang, 2012) and social relationship prediction (Aiello et.al. 2012). In these scenarios, a participant needs access to a large volume of personal information in order to spot the pattern (Meyerson & Williams, 2004). A dataset, which consists of the information at the level of individual respondents, is known as microdata dataset. In order to protect the privacy of each individual respondent, data holders must carefully sanitize

DOI: 10.4018/978-1-5225-8897-9.ch025

(also known as anonymize) the dataset before publishing. In the past decade, many privacy standards have been proposed, such as *k*-anonymity (Sweeney, 2002), *l*-diversity (Machanavajjhala et. al., 2007), and *t*-closeness (Li et.al, 2007).

Unlike the conventional centralized database system, where data requesters *directly* interact with data owners, information on a mobile social network is disseminated from user to user via *multi-hop relays*. Considering the well-known limitations with centralized systems, such as system bottlenecks or a single point of attacks problem, in this paper, we study the problem of multi-hop relay-based privacy-preserving data publishing, where a microdata table is gradually propagated from its original owner to distant people. However, under this scheme, the recipients will present different trust-levels regarding to the original data owner. Intuitively, after each time of relay, one should further provide more privacy protections on the data. For example, in Figure 1(a), along a social path with length *K*, each user eventually will get one copy of v_0's table, and we need the tables' privacy to be gradually reinforced, as shown by Figure 1(b). Data privacy and data utility are naturally at odds with each other (Meyerson & Williams, 2004): The more privacy a dataset preserves, the less utility the dataset has. This propagation scheme creates a unique problem: `for a group of friends, how can they create a series of tables with maximal overall data utility, and assure that the tables' privacy is increasingly protected at the same time?' To our best knowledge, this unique problem has never been proposed or solved.

Take Table 1 as an example. Suppose that *l*-diversity is the privacy requirement, and the total target propagation distance l is equal to 2. Assume that the corresponding participants are v_0, v_1 and v_2. With the growing of the propagation distance, the parameter l becomes larger and larger (i.e. after the first hop, the table should satisfy 2-diversity, and after the second hop, it should satisfy 3-diversity). The original dataset is given by Table 1 (a). Figures 2(b) and (c) give the results by directly using anonymizing operations on the original table T_0. We can see that the sanitized values are different in these two tables. However, during multi-hop relays, a participant can only observe the table passed from the previous one, and therefore, if v_0 gives T_1 to v_1, v_2 can only obtain Table 1(c), instead of T_2. Consider that the tables, which satisfy $(l + 1)$-diversity, must satisfy l-diversity. For v_0, he has two options for sending the dataset to v_1: he either sends T_3 or T_2. For the first case, the user v_1 only needs to forward T_3 to v_2 without any changes, while for the other case, v_1 should further sanitize T_2 and send the result T_3 to v_2.

Figure 1. An example of the target problem. (a), the black nodes consists of a social chain with length 4. The source v_0 possesses a data table and wants to propagate it to the other four nodes (from v_1 to v_4). Since the source node has a different trust level to these destination nodes, the privacy protection of the table's content should be further enhanced after each time of relays. (b), the dash lines represent the enhanced privacy requirements and the bars stand for the real privacy value of the table