Information Sciences 324 (2015) 32–43

Contents lists available at ScienceDirect

Information Sciences

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

Conditional anonymity with non-probabilistic adversary

Weien Chen[a,b], Yongzhi Cao[a,b,*], Hanpin Wang[a,b]

[a] Institute of Software, School of Electronics Engineering and Computer Science, Peking University, China
[b] Key Laboratory of High Confidence Software Technologies, Ministry of Education, China

ARTICLE INFO

Article history:
Received 1 June 2014
Revised 27 March 2015
Accepted 12 June 2015
Available online 26 June 2015

Keywords:
Anonymity
Conditional anonymity
Strong anonymity
Non-probabilistic adversary
Anonymity metric

ABSTRACT

Conditional anonymity, in comparison to classical strong anonymity, provides a novel perspective on anonymity and has been applied to analyzing anonymizing protocols. While the existing research on conditional anonymity is limited to the probabilistic setting, in this paper we introduce the notion of *set-theoretic conditional anonymity* by considering the threat from non-probabilistic adversary. Then we refine the understanding of the relationship between strong anonymity and conditional anonymity. Moreover, in order to quantitatively evaluate system's degree of anonymity, we propose a metric for set-theoretic conditional anonymity and a variant of an existing metric for probabilistic conditional anonymity. We formally show that a system will lose more (at best preserve equal) anonymity when adversary obtains more observable outputs from the system, which confirms the intuition that observations reveal sensitive information.

1. Introduction

The past two decades have seen a tremendous growing of the electronic world. The advent of phenomena like electronic commerce [1], online social networks [2,3] and RFID [4,5] has blurred the boundaries between our daily life and the online society. As a side effect, a large amount of sensitive information including user profiles and business data can be obtained online by malicious third parties. The threat to privacy is getting increasingly public concern and research attention [6–10].

In this study, we focus on the request of *anonymity*, which is a key notion in the field of information hiding. The aim of anonymity is to protect the privacy of user identities. Though, on the practical side, a number of anonymity protocols and systems have been implemented and deployed, e.g., [11–14], a firm formal basis for clearly and precisely understanding anonymity is still under construction. In the literature, two noticeable perspectives of anonymity have emerged: *strong anonymity* and *conditional anonymity*.

To the best of our knowledge, the term strong anonymity[1] was first used by Schneider and Sidiropoulos [15]. Taking Pfitzmann and Köhntopp's *anonymity set* [16] as the underlying adversary's knowledge, they defined strong anonymity by requiring the size of each anonymity set to reach its maximum. The intuition of their definition is that by enlarging the anonymity set the adversary's uncertainty about the real subject is increased. *Anonymity set size* [17] and *k-anonymity* [18] can be seen as quantitative extensions of Schneider and Sidiropoulos's strong anonymity.

[*] Corresponding author. Tel.: +86 1062 765 818.
E-mail addresses: cwe@pku.edu.cn (W. Chen), caoyz@pku.edu.cn, caoyongzhi@gmail.com (Y. Cao), whpxhy@pku.edu.cn (H. Wang).

[1] In the following sections, without explicit explanation, the term "strong anonymity" will refer to the high-level perspective of anonymity, which encompasses definitions including STSA and PSA (cf. Section 4).

http://dx.doi.org/10.1016/j.ins.2015.06.018
0020-0255/© 2015 Elsevier Inc. All rights reserved.

By considering the probabilistic knowledge of adversary, Halpern and O'Neill proposed a probabilistic extension of Schneider and Sidiropoulos's strong anonymity, called *strong probabilistic anonymity* [19]. The key observation therein is that, a system's anonymity degree depends not only on the size of the corresponding anonymity set, but also on how the adversary distributes her suspicion among the elements in the anonymity set. The ideal situation is when no chief suspect exists, i.e., when the adversary's suspicion is uniformly distributed. Entropy-based anonymity metrics [20–22] and Reiter and Rubin's anonymity spectrum [23] are all due to this probabilistic concern.

We point out that strong anonymity takes a *static* viewpoint in the sense that it does not take into account the change of adversary's knowledge. In contrast, by assuming an adversary who infers system's sensitive information based on her observation of system's public outputs, researchers have proposed a *dynamic* viewpoint of anonymity [19,24,25], called *conditional anonymity*.[2] Before an adversary obtains the outcome of a targeted system, she has a priori knowledge about the system's sensitive information. After she has obtained the actual outcome, she then forms a posteriori knowledge based on the observation. Conditional anonymity requires that there should be no change between adversary's a priori knowledge and her a posteriori one. As we will point out in Remark 2, conditional anonymity is a requirement for a system itself. In contrast, strong anonymity puts security constraints not only on the concerned system but also on the system's input.

So far, the idea of conditional anonymity has only been studied in the probabilistic setting [19,24–27]. To be more precise, the existing definitions of conditional anonymity are all defined with respect to the adversary who has probabilistic distribution of system's sensitive information as her knowledge. However, it is not always reasonable to assume that the adversary is able to perform probabilistic analysis. On one hand, it is sometimes not easy for adversary to obtain probabilistic information about a targeted system. On the other hand, as was also pointed by other researchers [26], user behavior may be intrinsically nondeterministic. It means that sometimes there may not exist a distribution of system's sensitive information. In light of this, we consider the adversary who has only anonymity set as her knowledge in this study.

For adversary with set-theoretic knowledge, we propose a set-theoretic version of conditional anonymity. With the help of this new notion, we then refine the understanding of the relationship between strong anonymity and conditional anonymity. Furthermore, instructed by the original concern of conditional anonymity, we propose a quantitative metric for our set-theoretic conditional anonymity and extend it to systems which generate multiple outputs one by one. We also confirm the intuition that a targeted system will lose more (at best preserve equal) anonymity when the adversary obtains more observable outputs from the system.

Related work. The concept of anonymity has been formalized in various frameworks. Representative examples are Pfitzmann et al.'s informal terminology [16], the epistemic logic based approach [19,28–30], the process algebra based approach [15,22,25,31], the "function view" based approach [32,33]. Combinational frameworks, which benefit from process algebra and epistemic logic, have been developed [34,35].

Since first introduced by Chaum [24], the concept of conditional anonymity has gained an extensive study [19,25–27,36]. However, only probabilistic adversary has been considered. That is, the adversary's knowledge is assumed to be based on probabilistic distributions of system's sensitive information. By requiring a priori distribution to be identical to a posteriori one, qualitative definitions of probabilistic conditional anonymity have been proposed [19,25].

There also exist several attempts to quantifying probabilistic conditional anonymity. In order to measure Bhargava and Palamidessi's *probabilistic anonymity*, Deng et al. proposed *α-anonymity* [26]. Informally speaking, α-anonymity characterizes the maximal improvement of adversary's knowledge. We compare α-anonymity and our metric for probabilistic conditional anonymity in Section 5.2. In [27], Chatzikokolakis et al. viewed systems as noisy channels with secret input and quantified maximal loss of anonymity by channel capacity. Chatzikokolakis and Palamidessi has also taken the Bayesian approach [37]. In that work, they regarded the inference of channel's input as a hypothesis-testing problem. The corresponding probability of error, called Bayes risk, is used to measure the degree of protection. A comparison between our metric and these different approaches is subject to future work. It is worth noting that, though our set-theoretic conditional anonymity is a weaker condition than probabilistic conditional anonymity, existing metrics of the later do not address the concern of the former. This is another motivation of this study.

As was pointed out in [27], from an abstract point of view, the fields of *information flow analysis* [38–47] and of *information hiding* share the same concern: high classified information should not be deduced by adversary via observing low classified information. Therefore, we may apply those metrics considered in the context of quantitative information flow such as Rényi min-entropy [48], guesswork [49] and marginal guesswork [50] to measuring conditional anonymity. Nonetheless, none of these approaches is based on the assumption of non-probabilistic adversary. We point out that our set-theoretic conditional anonymity, similar to its probabilistic counterpart, can also be seen as a generalization of *non-interference* [38], which is a key concept in information flow analysis.

The rest of the paper is organized as follows. In Section 2, notations, system model and adversarial model are introduced. In Section 3, we give the formal definition of set-theoretic conditional anonymity. In Section 4, we first recall strong anonymity and probabilistic conditional anonymity, and then reexamine the relationship between strong anonymity and conditional anonymity. After generalizing system model with multiple outputs, we propose a metric for our set-theoretic conditional anonymity and an averaged version of Deng et al.'s α-anonymity in Section 5. The adversary's side information is then incorporated into our

[2] By "conditional anonymity", we follow the terminology of Halpern and O'Neill.