Anonymity and k-Choice Identities*

Jacek Cichoń and Mirosław Kutyłowski

Institute of Mathematics and Computer Science,
Wrocław University of Technology, Poland
jacek.cichon@pwr.wroc.pl, miroslaw.kutylowski@pwr.wroc.pl

Abstract. We consider pervasive systems and identifiers for objects in these systems. Using unique global identifiers for these objects increases the size of the ID's and requires some global coordination. However, severe privacy threats are the key issues here.

On the other hand, for performing the goals of a pervasive system the identifiers are normally used in *small local environments*, and we need uniqueness limited to these environments only. This yields an opportunity to re-use the ID's and in this way anonymize the objects. The problem is that we cannot predict assignment of the objects to local environment or set it in advance, while on the other hand in many application scenarios we cannot change an ID already assigned to an object. Random predistribution of ID's is a technique that partially solves this problem, but has drawbacks due to the birthday paradox.

We propose a solution in which each object holds k preinstalled ID's (where k is a small parameter like $k = 2, 3, ...$). While entering a local environment, one of its ID's not used so far in this local environment is chosen for the object. We analyze probability of a conflict, i.e. of the event that no identity can be chosen for this object. We show that the size of ID's may be significantly reduced compared to random predistribution without increasing conflict probability. Apart from implementation advantages it contributes to privacy protection: since globally a large number of objects holds the same ID, privacy threats are reduced.

Keywords: anonymity set, identifier, two-choice paradigm, birthday paradox.

1 Introduction

Recently, it has been widely recognized that processing electronic data may yield severe privacy problems (see for instance a report [13]). The problem is that single data transaction might be regarded as fully safe from a privacy point of view, but collections of these data may leak sensitive private information. Among others, this is caused by the fact that electronic identification is expanding and used not only to recognize physical persons or machines, but also simple electronic

* The paper is partially supported by EU within the 6th Framework Programme under contract 001907 (DELIS).

Dingyi Pei et al. (Eds.): Inscrypt 2007, LNCS 4990, pp. 283–297, 2008.

devices used for diverse purposes. Since the number of such devices (like for instance RFID tags) is growing, the scope of the problem is expanding as well.

Technology for privacy protection has been recently developed by many research groups and companies. For instance, policy based approach has been proposed within an EU Project PRIME [12]. Most of the work is devoted to identity management and limiting privacy threats by appropriate organization, privacy policies and similar measures. This approach has limitations – it is not much useful when the devices concerned have very limited resources or the mentioned measures are poorly implemented. However, some papers propose solutions that are not based on proper behavior of system actors (see e.g. [9,5]). The concept is to guarantee some level of security by technical means only.

Privacy Problems for Pervasive Systems and RFID's. The objects in pervasive systems must be identified for the purposes of running these systems. Once the objects are given unique identifiers, the system can be used for tracing objects and, in this way, for tracing people holding these objects. This turns out to be one of the major problems in usage of RFID systems is retail stores. Alone the possibility of illegitimate tracing the clients and their preferences brings severe legal problems for the enterprises deploying such systems. Since more and more simple devices can include RFID's, this becomes one of the most acute security problems for emerging technologies of pervasive systems. Unfortunately, the devices need to be extremely simple, so the solutions designed for traditional networks composed of powerful computational units are of no use in majority of cases. In the simplest case we have to do with a memory unit with just a few bytes, which can be read by an external reader with no authorization performed before giving access to data.

Identification Problem in Local Environments. Our approach is inspired by the fact that even if a global system may consist of a huge number of possible ID's, the system is composed of a number of small environments with a limited number of objects and each function of the system is performed in some small environment. We have the following conflicting demands:

- all objects in a small environment should have different ID's,
- each ID may be linked to a large number of different objects in the global system, so that tracing a single object becomes complex.

Let us remark that the social mechanisms are exactly of this kind: ID's used in most cases of everyday life are not unique. For instance, even the first name and family name of a person does not identify a physical person uniquely, nevertheless the goals of identification are achieved. The general framework looks as follows:

- each person holds a couple of ID's (such as the first names) that are not unique in the whole society,
- in each system concerned there is a limited number of participants, identification within the system is performed with the ID's mentioned.

Nevertheless, there is a tendency to build information systems where each unit is assigned a unique digital ID. This makes design of databases much easier, but