Journal of Information Security and Applications 000 (2017) 1–16

Contents lists available at ScienceDirect

Journal of Information Security and Applications

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

A comprehensive view on quantity based aggregation for cadastral databases

Firas Al Khalil[a,*], Alban Gabillon[b], Patrick Capolsini[b]

[a] University College Cork, 13 South Mall, Cork, Ireland
[b] Université de la Polynésie française, BP 6570 Faa'a Aéroport, French Polynesia

ARTICLE INFO

Keywords:
Security
Database
Access control
Inference
Aggregation

ABSTRACT

Quantity Based Aggregation (QBA) control is a subject that is closely related to inference control in databases. The goal is to enforce k out of n disclosure control. In this paper we work on QBA problems in the context of cadastral databases: how to prevent a user from knowing 1) the owners of all parcels in a region, and 2) all parcels belonging to the same owner. This work combines and extends our previous work on the subject [1, 2, 3]. We overview the legislative context surrounding cadastral databases. We give important definitions related to the QBA concept. We present a complete model for QBA control in cadastral databases. We show how to implement the security policy efficiently, and we present our prototype of secure cadastral databases with some performance evaluations.

1. Introduction

The inference problem [4–7] in databases [8–13], privacy preserving data publishing [14–16] and mining [17,18] (and other domains [19–22]) has been heavily studied in the last couple of decades. The inference problem arises whenever (partial or complete) knowledge about some classified information can be derived (inferred, deduced) using unclassified information. In this work we address a problem that is very close to the inference problem and usually discussed with it: the aggregation problem. Actually, we are interested in a very special type of aggregation problems, called quantity-based aggregation problems (QBA henceforth).

QBA problems were, most probably, distinguished from inference and other aggregation problems for the first time in the work of Hinke [23], under the name "cardinality aggregation". Lunt [24] analyzed inference and aggregation problems and showed the difference between them. She coined the term quantity-based aggregation, and she gave the following example to illustrate QBA: suppose that there is a phonebook of n phone entries, where a user has the right to know k entries—at most—out of n; the goal of QBA control is to enforce this "k out of n" disclosure control.

The abovementioned example is known in the literature as the NSA (National Security Agency) or the SGA (Secretive Government Agency) phonebook problem. After reviewing the literature, it seems like the problem has not been addressed fully. The only work we are aware of, treating QBA directly, is that of Motro, Marks and Jajodia [25,26] two decades ago. In a previous paper [1] we addressed two QBA problems in the context of cadastral databases. In fact, we proposed a solution to enforce the following security policy: in the cadastral database, any user has the right to know the owner's name of any given parcel. However, this permission is constrained with the following prohibitions that represent two QBA problems: the user is forbidden to know:

Pr_1: the list of all parcels in a region,
Pr_2: the list of all parcels belonging to the same family.

In [1] we presented our model and an implementation based on graphs. However, this implementation was inefficient (for reasons we present in Section 9.1). In [2] we proposed an alternative implementation (using the relational model) and we showed empirically that the execution time of our QBA control algorithm grows linearly with respect to the number of users of the database; we tackled Pr_1 only, and presented our prototype demonstrating this prohibition on the cadastral database of the island of Maupiti, a test database provided to us by the real-estate service of French Polynesia. This work aims to synthesize and extend all of our previous work on QBA in cadastral databases. Moreover, we include results of experiments comparing the basic model [1,2] to the extension [3] that was developed to improve the availability of the data while preserving their confidentiality.

This paper is organized as follows: Section 2 presents the legislative context, discussing current policies regarding online

* Corresponding author.
E-mail addresses: firas.alkhalil@ucc.ie (F. Al Khalil), alban.gabillon@upf.pf (A. Gabillon), patrick.capolsini@upf.pf (P. Capolsini).

http://dx.doi.org/10.1016/j.jisa.2016.11.007
2214-2126/

publication of cadastral data in different countries including French Polynesia. Section 3 gives some preliminary definitions that help identifying different types of aggregation problems, and distinguishing them from inference problems. Section 4 gives details about QBA problems in the cadastral application: the security policy and some aggregation properties that should be enforced to implement both Pr_1 and Pr_2. Section 5 talks about additional aspects that should be taken into account while enforcing QBA: database updates, resetting access, and inference channels that could arise from QBA enforcement itself. Section 6 talks about some recommendations that should be considered for the choice of different parameters of our model. Section 7 talks about the application of QBA control to the French Polynesian cadaster, discussing the desired workflow and authentication. Section 8 presents the prototype developed to implement Pr_1. Section 9 presents performance benchmarks for the enforcement of Pr_1 comparing the base model [1,2] with the new one [3]. Section 10 gives a review of related work in both inference control and QBA control. Finally, Section 11 concludes this paper.

2. Legislative context

After investigating the state of some online cadastral applications, we will give a couple of examples from different countries reflecting the legal point of view on the publication of parcel ownership information. We will also explain the French point of view on the subject and the case of French Polynesia motivating this work.

Access to the Spanish [27] cadaster is provided through a mapping interface built with Google Maps. Parcel ownership information is considered sensitive and it is not available to the public.[1] Land owners form a different level of users (more privileged than the public) and they are granted access to all information related to their own properties if they provide a valid X509 certificate associated with their national electronic ID.

Similarly, the Belgian cadaster is available online for the public,[2] and ownership information is considered sensitive, thus prohibited. Using their national electronic ID, authenticated users can access through another website[3] to information related to their own parcels only.

In Croatia, parcel ownership information is public. Users can access the online website[4] where they can submit a query on any parcel and get a list of information related to the parcel, including land ownership. Queries are submitted by selecting the desired department, office and parcel ID or deed ID (using simple rudimentary lists). The query interface is protected against repeated automatic querying/scrapping/crawling.

Similarly, the state of Montana, US, considers land ownership as public information and they provide the cadaster for online browsing through a mapping interface.[5] Access to cadastral data in the US depends on state-level legislation.

Canada publishes its cadaster freely.[6] No ownership information is present, but all parcels can be downloaded as vector data (shapefiles) from an FTP site, after agreeing on a user-license agreement.

In France, the cadaster is available through a mapping interface,[7] however, only land boundaries are available to the public. This is due to the CNIL[8] recommendation [28] where it is stated that[9] *"the diffusion of any identifying information (directly or indirectly) on interactive terminals or public websites entails the risk of using this information for other purposes, including commercial, without the concerned people's consent."*

However, the Cada[10] indicates that *"punctual demands"* of cadastral excerpts are allowed [29]. Furthermore, cadastral excerpts may contain the name of land owners, but no other identifying information such as their national ID or their address. The frequency of demands and the number of parcels requested should be analyzed to ensure that these demands do not infringe the principle of free communication of cadastral documents. There is no clear definition of *"punctual demands"* and it is subject to various interpretations, therefore the Cada recommends a restrictive interpretation of the term.

French Polynesia is an overseas territory of France, where the recommendations of the CNIL and Cada are applicable. Currently, the punctuality of demands issued by citizens is ensured by employees of the real-estate service of French Polynesia when they are physically present at their desks (which is also the current situation in France). The work presented here is a requirement of the IT service of French Polynesia expressing their interpretation of the recommendations of both CNIL and Cada in order to provide the same facilities offered by the real estate service through the internet: a user should have access to the ownership information of any parcel, at random, but s/he is not allowed to exploit the service for commercial ends (or social, etc.) This interpretation is the foundation of prohibitions Pr_1 and Pr_2 presented in detail in Section 4.1.

3. Definitions

In this section we give a set of definitions (relying on earlier work described in Section 10) that helps identifying inference, aggregation and QBA problems:

Definition 1 (Inference problem)**.** The inference problem arises whenever a collection of information can be used to derive (infer, deduce) partial or complete knowledge about information stored in the database and classified higher than the classification of each subset of the collection. This collection forms an inference channel. Inference control is a mechanism used to eliminate inference channels and prevent users from performing inferences.

To illustrate an inference problem, let us consider the phonebook example. A phonebook is represented by the relation `PHONEBOOK (NAME, TEL, DEPT)` where the classifications of `NAME` and `TEL` (say, `UNCLASSIFIED`) are lower than that of `DEPT` (say, `CONFIDENTIAL`). A user with an `UNCLASSIFIED` clearance can access both `NAME` and `TEL`, and naturally `DEPT` is prohibited. However, if we consider that `TEL` depends on `DEPT` (e.g. one telephone per department, or numbers of the same department have the same suffix, etc.), then a user can infer, using `NAME + TEL`, to which department a given employee is affiliated, or even the list of employees who work in the same department.

Notice that the definition of the inference problem does not specify the source(s) of information in the collection. They could be partially derived from the database as in the inference from external knowledge, where a user combines his *a priori* knowledge with partial knowledge acquired from objects (that s/he has the

[1] http://www.maps.data-spain.com/cadastral/
[2] http://ccff02.minfin.fgov.be/cadgisweb/
[3] https://eservices.minfin.fgov.be/portal/fr/public/citizen/welcome
[4] http://www.katastar.hr/
[5] http://svc.mt.gov/msl/mtcadastral/
[6] http://clss.nrcan.gc.ca/cadastraldata-donneescadastrales-eng.php
[7] http://www.geoportail.gouv.fr/

[8] *Commission Nationale de l'Informatique et des Libertés*. An independent administrative authority whose mission is to ensure that information technology is at the service of citizens and does not undermine human identity, rights, private life, or individual and public liberties.
[9] Translated from its original language, French, by the authors.
[10] *Commission d'accès aux documents administratifs*. An independent administrative authority responsible for ensuring freedom of access to administrative documents.

Please cite this article as: F. Al Khalil et al., A comprehensive view on quantity based aggregation for cadastral databases, Journal of Information Security and Applications (2017), http://dx.doi.org/10.1016/j.jisa.2016.11.007