Preserving Confidentiality in Ontologies: Can we develop secure ontologies?

Erika Guetti Suca, Supervisor: Flávio Soares Corrêa da Silva

[1] Institute of Mathematics and Statistics – University of São Paulo – Brazil

{eguetti, fcs}@ime.usp.br

Abstract. *Many semantic web applications require selective sharing of ontologies due to copyright, confidentiality, business, security concerns and others. Our motivation is to protect the possibility to infer sensitive information and improperly extract it from the connected knowledge bases represented in ontologies. Supporting the design knowledge bases in order to overcome possible types of attacks. We give a brief description of a software tool we are currently building. We propose heuristics based on properties of ontologies and their planned use to identify and apply confidentiality preservation techniques that minimize the risk of breaches. Finally, we discuss several open problems.*

1. Introduction

Controlling access to information is important, but is not sufficient for protecting confidentiality. Many cybercrimes do not require advanced technology and are caused by human errors in the preservation of sensitive data. Technology itself cannot solve all issues related to confidentiality, yet it is important for information systems designers to take responsibility for the privacy of managed data. To do so, access to data must be designed and implemented to prevent confidentiality breaches. We aim to assist the knowledge engineer in improving the design of the knowledge base overcoming possible types of attacks. Therefore, we start with motivating examples, then we describe our proposal, expected results, and finally our conclusions.

2. Motivating Examples

We classify two levels of information ownership: (1)Full access to data: data will be processed before releasing and (2) Partial access to data: distributed data access. Our examples are related to Healthcare, but we can easily imagine similar needs in many other scenarios. The following example refers to the first category.

Example 1: Anonymizing Healthcare Data

Consider the raw patient data in Table 1 where each record represents a surgery case with the patient's specific information. *Job*, *Sex*, and *Age* are quasi-identifying (QID) attributes, which uniquely identify an individual. Quasi-identifiers can thus, when combined, become personally identifying facts. This process is called re-identification[Fung et al. 2010].

The hospital wants to release Table 1 for the purpose of classification analysis on the class attribute, *Transfuse*, which has two values, *YES* and *NO*, indicating whether or not the patient has received a blood transfusion. Without a loss of generality, we assume

Table 1. Raw patient data

	Quasi-identifier (QID)			*Class*	*Sensitive*
ID	**Job**	**Sex**	**Age**	**Transfuse**	**Surgery**
1	Janitor	M	34	Yes	Transgender
2	Doctor	M	58	No	Plastic
3	Mover	M	34	Yes	Transgender
4	Lawyer	M	24	No	Vascular
5	Mover	M	58	No	Urology
6	Doctor	M	24	No	Urology
7	Lawyer	F	58	No	Plastic
8	Carpenter	F	63	Yes	Vascular
9	Technician	F	63	Yes	Plastic

Table 2. Anonymous data(L=2, K=2, C = 0.5)

	Quasi-identifier (QID)			*Class*	*Sensitive*
ID	**Job**	**Sex**	**Age**	**Transfuse**	**Surgery**
1	Non-Technical	M	[30,60)	Yes	Transgender
2	Professional	M	[30,60)	No	Plastic
3	Non-Technical	M	[30,60)	Yes	Transgender
4	Professional	M	[1,30)	No	Vascular
5	Non-Technical	M	[30,60)	No	Urology
6	Professional	M	[1,30)	No	Urology
7	Professional	F	[30,60)	No	Plastic
8	Technical	F	[60,99)	Yes	Vascular
9	Technical	F	[60,99)	Yes	Plastic

that the only sensitive value in *Surgery* is *Transgender*. Table 2 shows the data after the anonymization using the LKC-privacy model[Fung et al. 2010]. The general intuition of LKC-privacy is to ensure that every combination of values in QID with maximum length L is shared by at least K records, and that the confidence of inferring any sensitive values in S is not greater than C, where L , K , C are thresholds and S is a set of sensitive values specified by the data holder(the hospital). We mentioned that Table 2 output is a secure view, after its processing in order to generalize the records into equivalence groups, so that each group contains at least k records with respect to some QID attributes, because it preserves the confidentiality of sensitive data (in this case the identity of patients). Hence, the sensitive values in each qid group are diversified enough to disorient confident inferences [Mohammed et al. 2009].

We need to establish a balance between retaining context and protecting participants aiming to preserve data that is of interest to the end user. These techniques are based on Statistical Disclosure Control(SDC) and they are called Privacy Preserving Data Publishing (PPDP), which eliminate privacy threats while, preserving useful information for data analysis. There are several anonymization models, depending on the requirements of the publication of data[Fung et al. 2010].

The following examples are related to partial access to data. Suppose that a system has generated an answer to a user that preserves confidential information. However, some works [Bonatti et al. 2014, Bonatti et al. 2015] showed that it is possible to break confidentiality once a smart user considers the following possibilities (examples 2 and 3).

Example 2: Attacks using User's Background Knowledge

Generally, only one part of the domain is modeled. The user may exploit various sources of background knowledge and meta-knowledge to reconstruct the hidden part of the knowledge base. The additional knowledge, from a public ontology could be used to infer secrets and confidentiality breach, for instance. We define $C_n(KB)$ as the set of logical consequences of a knowledge base. In effect, the condition $Cn(KB) \cap S = \emptyset$ is not sufficient to protect confidentiality. We suppose that there is one secret

$S = \{OncologyPatient(Bob)\}$ and

$KB = \{SocialSecurityNumber(Bob, 12345),\ OncologyPatient(user123), SocialSecurityNumber(user123, 12345)\}$.

KB does not entail $OncologyPatient(Bob)$, KB is a secure view. However, it is common knowledge that a $SocialSecurityNumber$ uniquely identifies one person, then the attacker can infer that $Bob = user123$, and he may consequently discloses the secret. From a probabilistic perspective, the attacker should not change the probability distribution distribution over the possible answers to a sensitive query that represents a set of secrets, even though revealing the secret is also important[Cuenca Grau and Horrocks 2008].

Example 3: Attacks to Complete Knowledge

Suppose the attacker has complete knowledge about a certain set of axioms. Then the attacker may be able to reconstruct some secrets from the "I don't know" answers of a secure view KB_u . In particular, for all instances, the system constantly answers "I do not know" to any query over one secret. Consider a hospital's knowledge base that defines a concept $Patient$ and a role $Patient\_of$ that describes which patient belongs to which of the Y hospital department. The KB consists of assertions like: $Patient(X)$ and $Patient\_of(X, Y)$ where we assume that each patient p belongs to exactly one department i. $1 \leq i \leq n \leq k$ A user u is authorized to see all assertions except the instances of $Y = n$, because n is a special department, dedicated to highly dangerous diseases. So S (set of secrecies) are all assertions where $Patient\_of(X, n)$, i.e., the members of n are all the patients that are treated for a special disease.

Based on the knowledge that for each patient e, KB contains exactly one assertion $Patient\_of(e, i)$ that is assumed to be known by the attackers, a smart attack can easily identify all the members of n .

3. Related Work

The work [Grau and Motik 2014] propose the partition of the knowledge base into a visible part KB_v and a hidden part KB_h of the secrets to be protect. The KB_h is accessible through a limited query interface. Their objective is to publish sensitive information in K_h, ensuring it is not exposed in any way, but preserving all the consequences the users could reason over $K_v \cup K_h$ This methodology works under the assumption that the signatures of KB_v and KB_h are disjointed, i.e., it does not consider protecting the axioms that are implied by a combination of KB_v and KB_h. Their work establishes lower bounds on the size and the number of queries that an import-by-query algorithm may need to ask an oracle in order to solve a reasoning task. About the possibility of reuse, their model is flexible and useful for KB's designers to ensure selective access to parts of KB's. Nevertheless, it is not discussed how to select the hidden part given a set of target secrets ; moreover the user's background knowledge that should be maintained is not analyzed.

A probabilistic perspective is introduced by the work [Cuenca Grau and Horrocks 2008]. Enlarging the public view should not change the probability distribution over the possible answers to a sensitive query that can represent a set of secrets. Their confidentiality condition allows P to be replaced with a different P' after enlarging the public view. But, taking a closer look, P does not really consider the user's a *priori* knowledge about the KB.

The work [Grau and Kostylev 2016] focus on data publishing and anonymization, but not on access control. They lay theoretical foundations for Privacy-Preserving Data Publishing(PPDP) in the context of Linked Data, formalizing anonymization in terms of suppressor functions. Working the computational complexity of the decision problems underlying the policy compliance, safety, and optimality requirements. Their policy compliance ensures that sensitive information remains protected when the anonymized data is considered in isolation. They can ensure safety by replacing all occurrences of sensitive data with blank nodes. As a consequence, they ensure the published linked datasets are protected against disclosure of sensitive information while remaining practically useful. This framework does not receive OWL 2 ontologies yet, and the authors expect that the introduction of OWL2 will lead to significant technical challenges, especially in combination with closed-world semantics.

Lastly, the works [Bonatti et al. 2014, Bonatti et al. 2015] introduce a stronger confidentiality model that take both object level and meta-level background knowledge into consideration and define a method for computing secure knowledge views. They illustrate attacks using an object-level background, complete knowledge, and signatures including in the the formalization of user's prior knowledge. Their methods are inspired in Controlled Query Evaluation (CQE) based on lies and/or refusals. Technically, they use lies because rejected queries are not explicitly marked. Once the user requests access to information by means of a query, the policy requirements are enforced. Their model proposes a safe approximation of background meta-knowledge and it checks its answers to users queries that do not entail any secret. They applied the data filtering without considering a context analysis that helps to distinguish the relevance or interest of the end user from the data.

4. The Proposal

We summarize the confidentiality problem in how to create secure views for answers to user queries that preserve the confidentiality given an insecure ontology. We consider that the secure view of the ontology does not entail a secret. So, we assume that every scenario of published data has its own assumptions and requirements on the data holder, the data recipients, and the purpose of published data. Our framework implements a conceptual structure, a Simple Confidentiality Model (SCM), based on the works [Cuenca Grau and Horrocks 2008, Bonatti et al. 2015]. The SCM considers the user's background knowledge to create a secure view, then, it is not vulnerable to the attacks mentioned in Examples 2 and 3. We show in Figure 1 the general components of our framework. We are currently building it up.

The framework considers two ways of interaction with a user: (1) Configure features of KB's:- the user sets up inherent characteristics of a knowledge base (i.e. the user's background knowledge, a finite set of secrecies that should not be disclosed for each user). Furthermore, there are characteristics that concerns the objective of information utility as policy representations, hidden features (it is important to preserve the semantics of the data or not) are also essentials. (2) User Queries: if the queries are compliant with the preservation of confidentiality, the system will show a secure view. Otherwise, it will show a justification because it is not compliant with the policies.

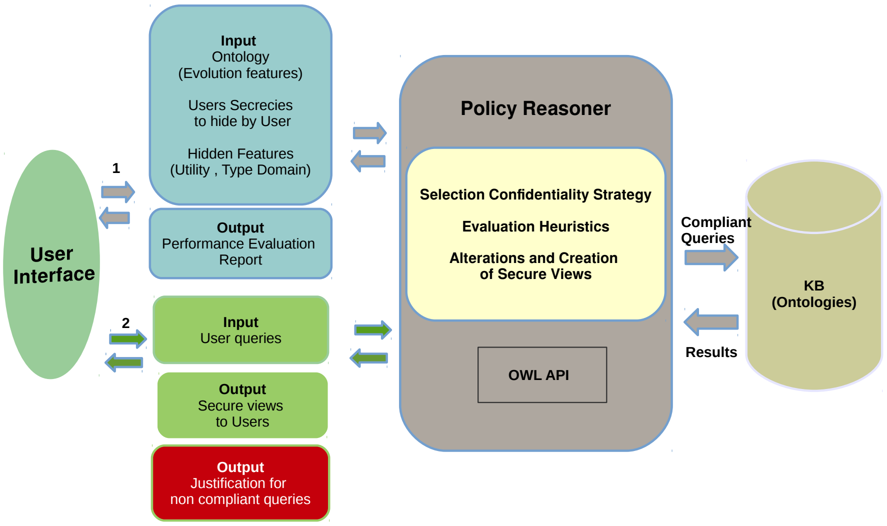

Figure 1. A Framework to Preserve Confidentiality in Ontologies

5. Evaluation plan

We propose heuristics based on properties of knowledge bases and their planned use to identify and apply confidentiality preservation techniques that minimize the risk of breaches. Our intention is to systematize, as far as possible, the heuristics performance evaluation connected to confidentiality preservation techniques. We propose the following:

1. OWL 2 Profiles: The W3C distinguishes three OWL 2 profiles (*OWL 2 QL, OWL 2 RL, OWL 2 EL*). Reasoning over ontologies in each of these profiles is performed using a set of rules that vary with each profile, there is some overlap among the different rule sets. The essential differences between the OWL 2 profiles are in their restrictions on inverse properties, and universal and existential quantifiers. We want to understand whether the differences between OWL 2 profiles influence creating a secure view.
2. Reasoning Strategies: Forward chaining, backward chaining or a hybrid strategy. We want to understand how the strategies of ontology reasoners influence to creating secure view.
3. Ontology Development Strategies: There are two main ways of comprehending ontology development, top-down and bottom-up. Ontology construction can not be considered in isolation, without relation to its use. We want to understand whether the strategies of ontology development impact on creating create a secure view. Our intention is to work in the Healthcare domain.

6. Work Plan

We present the properties and general requirements for a software tool for preventing access to sensitive information. Our immediate future work is to develop our proposal in case studies related to healthcare. The doctoral defense is expected to be in February 2018. Some specific anticipated outcomes of our work include:

- Systematization and comparative analysis of proposed heuristics.
- Basing ourselves heuristics, our goal is to identify the characteristics of ontologies that allow preserving the confidentiality, or the properties of non-viability.
- Creating scenarios in which ontology developers have restricted access to the parts of the ontology by others with the objective to improve the computational complexity.
- Identify others kinds of attacks and propose privacy models that ensure the confidentiality offering support for conflict resolution: semantic inconsistencies and ambiguities.
- A future direction of our work is to consider other forms of distortion of the knowledge base to ensure confidentiality. For example, we can explore not only remove elements of the knowledge base, but we may also add new elements.

References

Bonatti, P. A., Petrova, I. M., and Sauro, L. (2015). Optimized construction of secure knowledge-base views. In Calvanese, D. and Konev, B., editors, *Description Logics*, volume 1350 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Bonatti, P. A., Sauro, L., and Petrova, I. M. (2014). A mechanism for ontology confidentiality. In *Proceedings of the 29th Italian Conference on Computational Logic, Torino, Italy, June 16-18, 2014.*, pages 147–161.

Cuenca Grau, B. and Horrocks, I. (2008). Privacy-preserving query answering in logic-based information systems. In *Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence*, pages 40–44, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Fung, B. C., Wang, K., Fu, A. W.-C., and Yu, P. S. (2010). *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. Chapman & Hall/CRC, 1st edition.

Grau, B. C. and Kostylev, E. V. (2016). Logical foundations of privacy-preserving publishing of linked data. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 943–949.

Grau, B. C. and Motik, B. (2014). Reasoning over ontologies with hidden content: The import-by-query approach. *CoRR*, abs/1401.5853.

Mohammed, N., Fung, B. C., Hung, P. C., and Lee, C.-k. (2009). Anonymizing healthcare data: A case study on the blood transfusion service. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1285–1294. ACM.