Controlled Query Evaluation in Description Logics Through Instance Indistinguishability (Extended Abstract)

Gianluca Cima[1], Domenico Lembo[1], Riccardo Rosati[1], Domenico Fabio Savo[2]

[1]Sapienza University of Rome
{cima,lembo,rosati}@diag.uniroma1.it
[2]University of Bergamo
domenicofabio.savo@unibg.it

Abstract. This extended abstract summarizes our recent work [11] about Controlled Query Evaluation over Description Logic ontologies.

Controlled query evaluation (CQE) is a declarative framework for privacy-preserving query answering investigated in the literature on knowledge representation and database theory [16,7,3]. The basic idea of CQE is defining a *data protection policy* through logical statements. Specifically, we consider the case where the policy is a set of denial assertions, i.e., FO sentences of the form $\forall \boldsymbol{x}.\phi(\boldsymbol{x}) \rightarrow \bot$, such that $\exists \boldsymbol{x}.\phi(\boldsymbol{x})$ is a Boolean conjunctive query. Consider for instance an organization that wants to keep confidential the fact that it has suppliers involved in both Project A and Project B. This can be expressed over the information schema of the organization through a denial assertion of the form:

$$\forall x.\, \mathsf{Supplier}(x) \wedge \mathsf{ProjA}(x) \wedge \mathsf{ProjB}(x) \rightarrow \bot$$

In CQE, two different main approaches can be identified. The first one [5,4,6,2,1,17] models privacy preservation through the notion of *indistinguishable data instances*. In this approach, a system for CQE enforces data privacy if, for every data instance I, there exists a data instance I' that does not violate the data protection policy and is indistinguishable from I for the user, i.e., for every user query q, the system provides the same answers to q over I and over I'. We call this approach *(instance) indistinguishability-based* (IB). In continuation of the previous example, in the presence of an instance $\{\mathsf{Supplier}(c), \mathsf{ProjA}(c), \mathsf{ProjB}(c)\}$, an IB system should answer user queries as if the instance were, e.g., $\{\mathsf{Supplier}(c), \mathsf{ProjA}(c)\}$ (note that other instances not violating the policy can be considered as indistinguishable, e.g., $\{\mathsf{Supplier}(c), \mathsf{ProjB}(c)\}$).

The second approach [8,13,14] models privacy preservation by considering the whole (possibly infinite) set of answers to queries that the system provides to the user. In this approach, a CQE system protects the data if, for every data

instance I, the logical theory corresponding to the set of answers provided by the system to all queries over I does not entail any violation of the data protection policy. According to [14], we call this approach *confidentiality-preserving (CP)*. In our ongoing example, a CP system would entail, e.g., the queries $\mathsf{Supplier}(c) \wedge \mathsf{ProjA}(c)$ and $\exists x.\mathsf{Supplier}(x) \wedge \mathsf{ProjB}(x)$, but not also the query $\mathsf{Supplier}(c) \wedge \mathsf{ProjB}(c)$ (notice that the choice is non-deterministic, and in our example the system could have decided to disclose that c participates in Project B and hide its participation in Project A).

In both approaches, the ultimate goal is to realize *optimal* CQE systems, i.e., systems maximizing the answers returned to user queries, still respecting the data protection policy. Traditionally, this aim has been pursued through the usage of a *single optimal censor*, i.e., a specific implementation of the adopted notion of privacy-preservation, either IB or CP. Since in both approaches several optimal censors typically exist, this way of proceeding requires to select an optimal censor (thus discarding all the others). However, we argue that this should be motivated by a reasonable semantic preference criterion. Indeed, in the lack of further meta-information about the data domain, picking up just one optimal censor may end up in arbitrary behaviors. To avoid this, query answering over all optimal censors has been recently studied (limited to the CP approach) [13,15].

Despite their similarities, the precise relationship between the IB and CP approaches is still not clear and has not been fully investigated yet. Also, query answering over all optimal IB censors has not been previously studied. Moreover, among the complexity results obtained and the techniques defined so far for CQE, we still miss the identification of cases that are promising towards its practical usage.

In our work, we aim at filling some of the above mentioned gaps in the context of Description Logic (DL) ontologies.[1] We focus on the approach to CQE based on instance indistinguishability, and study its relationship with the CP approach. Specifically, we prove that the IB approach to CQE in DLs corresponds to a particular instance of the CP approach to CQE [15]. Based on such a correspondence, for ontologies specified in the well-known DL *DL-Lite*$_{\mathcal{R}}$ [9], we are able to transfer some complexity results for query answering over all optimal censors shown in [15] to the case of CQE under IB censors. Such results show that, even in the lightweight DL *DL-Lite*$_{\mathcal{R}}$, query answering in the IB approach is coNP-complete in data complexity, unless one relies on a single optimal censor chosen non-deterministically in the lack of further meta-information about the domain of the dataset.

To overcome the above problems and provide a practical, semantically well-founded solution, we define a *quasi-optimal* notion of IB censor, which corresponds to the best sound approximation of all the optimal IB censors.We then prove that, in the case of *DL-Lite*$_{\mathcal{R}}$ ontologies, query answering based on the quasi-optimal IB censor is tractable with respect to data complexity and is reducible to the evaluation of a first-order query over the data instance, i.e., it

[1] Privacy-preserving query answering in DLs has been investigated also in settings different from CQE: see, e.g., [12,10,18].

is *first-order rewritable*. We believe that this result has an important practical impact. Indeed, we have identified a setting in which privacy-preserving query answering formalized in a declarative logic-based framework as CQE, for a DL (i.e., *DL-Lite*$_{\mathcal{R}}$) specifically designed for data management, has the same data complexity upper bound as evaluating queries over a database (i.e., AC^0). This opens the possibility of defining algorithms for CQE of practical usage, amenable to implementation on top of traditional (relational) data management systems, as in Ontology-based Data Access [19]. We are currently working to achieve this goal.

Another important future direction is a deeper study of the user model. Indeed, our framework inherits from its predecessors a relatively simple model in which the user knows (at most) the TBox and all the query answers returned by the system, and considers only the *deductive* abilities of the user over such knowledge. This user model might need to be enriched to capture more realistic data protection scenarios.

References

1. M. Benedikt, P. Bourhis, L. Jachiet, and M. Thomazo. Reasoning about disclosure in data integration in the presence of source constraints. In *Proc. of the 28th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1551–1557, 2019.
2. M. Benedikt, B. Cuenca Grau, and E. V. Kostylev. Logical foundations of information disclosure in ontology-based data integration. *Artificial Intelligence*, 262:52–95, 2018.
3. J. Biskup. For unknown secrecies refusal is better than lying. *Data and Knowledge Engineering*, 33(1):1–23, 2000.
4. J. Biskup and P. A. Bonatti. Controlled query evaluation for enforcing confidentiality in complete information systems. *Int. J. of Information Security*, 3(1):14–27, 2004.
5. J. Biskup and P. A. Bonatti. Controlled query evaluation for known policies by combining lying and refusal. *Ann. of Mathematics and Artificial Intelligence*, 40(1-2):37–62, 2004.
6. J. Biskup and T. Weibert. Keeping secrets in incomplete databases. *Int. J. of Information Security*, 7(3):199–217, 2008.
7. P. A. Bonatti, S. Kraus, and V. S. Subrahmanian. Foundations of secure deductive databases. *IEEE Trans. Knowl. Data Eng.*, 7(3):406–422, 1995.
8. P. A. Bonatti and L. Sauro. A confidentiality model for ontologies. In *Proc. of the 12th Int. Semantic Web Conf. (ISWC)*, volume 8218 of *Lecture Notes in Computer Science*, pages 17–32, 2013.
9. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
10. D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. View-based query answering in description logics: Semantics and complexity. *J. of Computer and System Sciences*, 78(1):26–46, 2012.
11. G. Cima, D. Lembo, R. Rosati, and D. F. Savo. Controlled Query Evaluation in Description Logics Through Instance Indistinguishability. In *Proc. of the 29th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1791–1797, 2020.

12. B. Cuenca Grau and I. Horrocks. Privacy-preserving query answering in logic-based information systems. In *Proc. of the 18th Eur. Conf. on Artificial Intelligence (ECAI)*, pages 40–44, 2008.
13. B. Cuenca Grau, E. Kharlamov, E. V. Kostylev, and D. Zheleznyakov. Controlled query evaluation over OWL 2 RL ontologies. In *Proc. of the 12th Int. Semantic Web Conf. (ISWC)*, volume 8218 of *Lecture Notes in Computer Science*, pages 49–65, 2013.
14. B. Cuenca Grau, E. Kharlamov, E. V. Kostylev, and D. Zheleznyakov. Controlled query evaluation for datalog and OWL 2 profile ontologies. In *Proc. of the 24th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 2883–2889, 2015.
15. D. Lembo, R. Rosati, and D. F. Savo. Revisiting controlled query evaluation in description logics. In *Proc. of the 28th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1786–1792, 2019.
16. G. L. Sicherman, W. de Jonge, and R. P. van de Riet. Answering queries without revealing secrets. *ACM Trans. Database Syst.*, 8(1):41–59, 1983.
17. T. Studer and J. Werner. Censors for boolean description logic. *Trans. Data Privacy*, 7(3):223–252, 2014.
18. J. Tao, G. Slutzki, and V. G. Honavar. A conceptual framework for secrecy-preserving reasoning in knowledge bases. *ACM Trans. on Computational Logic*, 16(1):3:1–3:32, 2014.
19. G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, and M. Zakharyaschev. Ontology-based data access: A survey. In *Proc. of the 27th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 5511–5519, 2018.