Controlled Query Evaluation in Description Logics Through Instance Indistinguishability

Gianluca Cima[1] , **Domenico Lembo**[1] , **Riccardo Rosati**[1] and **Domenico Fabio Savo**[2]
[1]Sapienza Università di Roma
[2]Università degli Studi di Bergamo
{cima, lembo, rosati}@diag.uniroma1.it, domenicofabio.savo@unibg.it

Abstract

We study privacy-preserving query answering in Description Logics (DLs). Specifically, we consider the approach of controlled query evaluation (CQE) based on the notion of *instance indistinguishability*. We derive data complexity results for query answering over *DL-Lite*$_\mathcal{R}$ ontologies, through a comparison with an alternative, existing confidentiality-preserving approach to CQE. Finally, we identify a semantically well-founded notion of approximated query answering for CQE, and prove that, for *DL-Lite*$_\mathcal{R}$ ontologies, this form of CQE is tractable with respect to data complexity and is first-order rewritable, i.e., it is always reducible to the evaluation of a first-order query over the data instance.

1 Introduction

We consider controlled query evaluation (CQE), a declarative framework for privacy-preserving query answering investigated in the literature on knowledge representation and database theory [Sicherman *et al.*, 1983; Bonatti *et al.*, 1995; Biskup, 2000]. The basic idea of CQE is defining a *data protection policy* through logical statements. 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 [Biskup and Bonatti, 2004b; Biskup and Bonatti, 2004a; Biskup and Weibert, 2008; Benedikt *et al.*, 2018; Benedikt *et al.*, 2019; Studer and Werner, 2014] 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 [Bonatti and Sauro, 2013; Cuenca Grau *et al.*, 2013; Cuenca Grau *et al.*, 2015] 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 [Cuenca Grau *et al.*, 2015], 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 construction of a *single optimal censor*, i.e., a specific implementation of the adopted notion of privacy-preservation, either IB or CP. Since, however, in both approaches several optimal censors typically exist, this way of proceeding requires to make a choice on how to obfuscate data, which, in the absence of additional (preference) criteria, may result discretionary. To avoid this, query answering over all optimal censors has been recently studied (limited to the CP approach) [Cuenca Grau *et al.*, 2013; Lembo *et al.*, 2019].

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 this paper, we aim at filling some of the above mentioned gaps in the context of Description Logic (DL) ontologies.[1]

[1]Privacy-preserving query answering in DLs has been investi-

We focus on the approach to CQE based on instance indistinguishability (Section 3), and study its relationship with the CP approach (Section 4). Specifically, we prove that the IB approach to CQE in DLs corresponds to a particular instance of the CP approach to CQE [Lembo *et al.*, 2019]. Based on such a correspondence, for ontologies specified in the well-known DL *DL-Lite*$_\mathcal{R}$ [Calvanese *et al.*, 2007], we are able to transfer some complexity results for query answering over all optimal censors shown in [Lembo *et al.*, 2019] to the case of CQE under IB censors (Section 5). Such results show that, even in the lightweight DL *DL-Lite*$_\mathcal{R}$, query answering in the IB approach is intractable with respect to 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 (Section 6). 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 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 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 [Xiao *et al.*, 2018].

For complete proofs of our results we refer the reader to an extended version of the present paper [Cima *et al.*, 2020].

2 Preliminaries

We use standard notions of function-free first-order (FO) logic, and in particular we consider Description Logics (DLs), which are fragments of FO using only unary and binary predicates, called concepts and roles, respectively [Baader *et al.*, 2007]. We assume to have the pairwise disjoint countably infinite sets $\Sigma_C, \Sigma_R, \Sigma_I$ and $\Sigma_\mathcal{V}$ for *atomic concepts*, *atomic roles*, *constants* (a.k.a. individuals), and *variables*, respectively. A DL ontology $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$ is constituted by a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$, specifying intensional and extensional knowledge, respectively. The set of atomic concepts and roles occurring in $\mathcal{O}$ is the *signature* of $\mathcal{O}$. The semantics of $\mathcal{O}$ is given in terms of FO models over the signature of $\mathcal{O}$, in the standard way [Baader *et al.*, 2007]. In particular, we say that $\mathcal{O}$ is *consistent* if it has at least one model, *inconsistent* otherwise. $\mathcal{O}$ *entails* an FO sentence ϕ specified over the signature of $\mathcal{O}$, denoted $\mathcal{O} \models \phi$, if ϕ is true in every model of $\mathcal{O}$. In this paper, we consider ontologies expressed in *DL-Lite*$_\mathcal{R}$, the member of the *DL-Lite* family [Calvanese *et al.*, 2007] which underpins OWL 2 QL [Motik *et al.*, 2012], i.e., the OWL 2 profile specifically designed for efficient query answering. A TBox $\mathcal{T}$ in *DL-Lite*$_\mathcal{R}$ is a finite set of axioms of the form $B_1 \sqsubseteq B_2$ (resp., $R_1 \sqsubseteq R_2$), denoting concept (resp., role) inclusion, and $B_1 \sqsubseteq \neg B_2$ (resp., $R_1 \sqsubseteq \neg R_2$), denoting concept (resp., role) disjointness, where: R_1, R_2 are of the form P, with $P \in \Sigma_R$, or its inverse P^-, and B_1, B_2 are of the form A, with $A \in \Sigma_C$, $\exists P$, or $\exists P^-$, i.e., unqualified existential restrictions, which denote the set of objects occurring as first or second argument of P, respectively. An ABox $\mathcal{A}$ is a finite set of *ground atoms*, i.e., assertions of the form $A(a)$, $P(a, b)$, where $A \in \Sigma_C$, $P \in \Sigma_R$, and $a, b \in \Sigma_I$. As usual in query answering over DL ontologies, we focus on the language of conjunctive queries. A Boolean conjunctive query (BCQ) q is an FO sentence of the form $\exists \vec{x}.\phi(\vec{x})$, where $\vec{x}$ are variables in $\Sigma_\mathcal{V}$, and $\phi(\vec{x})$ is a finite, non-empty conjunction of atoms of the form $\alpha(\vec{t})$, where $\alpha \in \Sigma_C \cup \Sigma_R$, and each term in $\vec{t}$ is either a constant in Σ_I or a variable in $\vec{x}$. We denote by $Eval(q, \mathcal{A})$ the evaluation of a query q over (the model isomorphic to) an ABox $\mathcal{A}$.

A *denial assertion* (or simply a denial) is an FO sentence of the form $\forall \vec{x}.\phi(\vec{x}) \rightarrow \bot$, such that $\exists \vec{x}.\phi(\vec{x})$ is a BCQ. Given one such denial δ and an ontology $\mathcal{O}$, we say that $\mathcal{O} \cup \{\delta\}$ is consistent if $\mathcal{O} \not\models \exists \vec{x}.\phi(\vec{x})$, and is inconsistent otherwise.

In the following, with $\mathbf{FO}$, $\mathbf{CQ}$, and $\mathbf{GA}$ we denote the languages of function-free FO sentences, BCQs, and ground atoms, respectively, all specified over the alphabets $\Sigma_C, \Sigma_R, \Sigma_I$, and $\Sigma_\mathcal{V}$. Given an ontology $\mathcal{O}$ and a language $\mathcal{L}$, with $\mathcal{L}(\mathcal{O})$ we refer to the subset of $\mathcal{L}$ whose sentences are built over the signature of $\mathcal{O}$ and the variables in $\Sigma_\mathcal{V}$. For a TBox $\mathcal{T}$ and a language $\mathcal{L}$, we denote by $\mathsf{cl}^\mathcal{T}_\mathcal{L}(\cdot)$ the function that, for an ABox $\mathcal{A}$, returns all the sentences $\phi \in \mathcal{L}(\mathcal{T} \cup \mathcal{A})$ such that $\mathcal{T} \cup \mathcal{A} \models \phi$.

For the sake of presentation, we will limit our technical treatment to languages containing only closed formulas, but our results hold also for open formulas. In particular, the results on entailment of BCQs (see Sections 5 and 6) can be extended to arbitrary (i.e., non-Boolean) CQs in the standard way[2]. Our complexity results are for data complexity, i.e., are w.r.t. the size of the ABox only.

3 CQE through Instance Indistinguishability

A CQE framework consists of a TBox $\mathcal{T}$ and a policy $\mathcal{P}$ over $\mathcal{T}$, i.e., a finite set of denial assertions over the signature of $\mathcal{T}$. An ABox $\mathcal{A}$ for $\mathcal{T}$ is such that $\mathcal{A}$ and $\mathcal{T}$ have the same signature. In the following, when a TBox $\mathcal{T}$ is given, we always assume that the coupled policy is specified over $\mathcal{T}$, that each considered ABox $\mathcal{A}$ is for $\mathcal{T}$, and that, unless otherwise specified, $\mathcal{T} \cup \mathcal{A}$ and $\mathcal{T} \cup \mathcal{P}$ are consistent. A *censor* is a function that taken an ABox for $\mathcal{T}$ as input alters standard query answering over $\mathcal{T} \cup \mathcal{A}$ so that on the basis of the answers (even a possibly infinite set thereof) and the TBox, a user can never infer a BCQ $\exists \vec{x}.\phi(\vec{x})$ such that $\forall \vec{x}.\phi(\vec{x}) \rightarrow \bot$ belongs to $\mathcal{P}$.

We here propose a notion of censor which is the natural application to our framework of the analogous definitions given in [Biskup and Bonatti, 2004b; Biskup and Weibert, 2008;

gated also in settings different from CQE: see, e.g., [Cuenca Grau and Horrocks, 2008; Calvanese *et al.*, 2012; Tao *et al.*, 2014].

[2]We also notice that, since *DL-Lite*$_\mathcal{R}$ is insensitive to the adoption of the *unique name assumption* (UNA) for CQ answering [Artale *et al.*, 2009], our results hold both with and without UNA.

Benedikt *et al.*, 2018; Benedikt *et al.*, 2019]. The basic idea of this approach is that for every underlying instance (an ABox in our framework) and every query, a censor returns to the user the same answers it would return on another (possibly identical) instance that does not contain confidential data, so that she cannot understand which of the two instances she is querying. This is formalized as follows.

Definition 1 [Indistinguishability-based censor] Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. An *indistinguishability-based (IB) censor* for $\mathcal{T}$ and $\mathcal{P}$ is a function $\mathsf{cens}(\cdot)$ that, for each ABox $\mathcal{A}$, returns a set $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ such that there exists an ABox $\mathcal{A}'$ for which (i) $\mathsf{cens}(\mathcal{A}) = \mathsf{cens}(\mathcal{A}')$ (in this case we say that $\mathcal{A}$ and $\mathcal{A}'$ are *indistinguishable* w.r.t. cens) and (ii) $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is a consistent FO theory.

Example 1 Let us now formalize more precisely the scenario we have used for the examples in the introduction, by instantiating our CQE framework. The TBox signature consists of the atomic concepts Supplier, ProjA, and ProjB, denoting the set of suppliers of the company, suppliers involved in Project A and those involved in Project B, respectively, and contains the axioms $\mathsf{ProjA} \sqsubseteq \mathsf{Supplier}$ and $\mathsf{ProjB} \sqsubseteq \mathsf{Supplier}$, stating that each individual instance of ProjA or ProjB is also instance of Supplier. Data protection is specified through the policy $\mathcal{P} = \{\forall x.\mathsf{ProjA}(x) \wedge \mathsf{ProjB}(x) \rightarrow \bot\}$. The following functions are IB censors for $\mathcal{T}$ and $\mathcal{P}$:

- cens_1: given an ABox $\mathcal{A}$, $\mathsf{cens}_1(\mathcal{A})$ returns the set $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}_{P_A})$ of BCQs, where $\mathcal{A}_{P_A}$ is obtained from $\mathcal{A}$ by removing the assertion $\mathsf{ProjA}(c)$, for each individual c such that both $\mathsf{ProjA}(c)$ and $\mathsf{ProjB}(c)$ are in $\mathcal{A}$ (note that for every ABox $\mathcal{A}$, $\mathcal{A}$ and $\mathcal{A}_{P_A}$ are indistinguishable w.r.t. cens_1. Similarly in the following censors).
- cens_2: given an ABox $\mathcal{A}$, $\mathsf{cens}_2(\mathcal{A})$ returns the set $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}_{P_B})$ of BCQs, where $\mathcal{A}_{P_B}$ is obtained from $\mathcal{A}$ by removing the assertion $\mathsf{ProjB}(c)$, for each individual c such that both $\mathsf{ProjA}(c)$ and $\mathsf{ProjB}(c)$ are in $\mathcal{A}$.
- cens_3: given an ABox $\mathcal{A}$, $\mathsf{cens}_3(\mathcal{A})$ returns the set $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}_{sup})$ of BCQs, where $\mathcal{A}_{sup}$ is obtained from $\mathcal{A}$ by adding the assertion $\mathsf{Supplier}(c)$ and removing $\mathsf{ProjA}(c)$ and $\mathsf{ProjB}(c)$, for each individual c such that both $\mathsf{ProjA}(c)$ and $\mathsf{ProjB}(c)$ are in $\mathcal{A}$. □

It is easy to see that an IB censor always exists, but, as Example 1 shows, there may be many IB censors for $\mathcal{T}$ and $\mathcal{P}$, and so it is reasonable to look for censors preserving as much information as possible. Formally, given two IB censors cens and cens' for $\mathcal{T}$ and $\mathcal{P}$, we say that cens' is *more informative* than cens if: *(i)* for every ABox $\mathcal{A}$, $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cens}'(\mathcal{A})$, and *(ii)* there exists an ABox $\mathcal{A}'$ such that $\mathsf{cens}(\mathcal{A}') \subset \mathsf{cens}'(\mathcal{A}')$. Optimal censors are then defined as follows.

Definition 2 Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. An IB censor cens for $\mathcal{T}$ and $\mathcal{P}$ is *optimal* if there does not exist any other IB censor for $\mathcal{T}$ and $\mathcal{P}$ that is more informative than cens. The set of all the optimal IB censors for $\mathcal{T}$ and $\mathcal{P}$ is denoted with $\mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$.

Example 2 Among the censors of Example 1, $\mathsf{cens}_3 \notin \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$, since both cens_1 and cens_2 are more informative than cens_3. It can be then verified that cens_1 and cens_2 are the only optimal IB censors for $\mathcal{T}$ and $\mathcal{P}$. □

4 IB Censors vs. CP Censors

In [Cuenca Grau *et al.*, 2015], a different notion of censor, named *confidentiality-preserving (CP)* censor, has been proposed. Intuitively, a CP censor establishes which are the BCQs entailed by a TBox and a given ABox that can be disclosed without violating the policy. We report below the definition given in [Lembo *et al.*, 2019], which generalizes CP censors to any language $\mathcal{L} \subseteq \mathbf{FO}$, called the censor language.

Definition 3 [Confidentiality-preserving censor] Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, and $\mathcal{L} \subseteq \mathbf{FO}$ be a language. A *confidentiality-preserving (CP) censor* in $\mathcal{L}$ for $\mathcal{T}$ and $\mathcal{P}$ is a function $\mathsf{cens}(\cdot)$ that, for each ABox $\mathcal{A}$, returns a set $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cl}^{\mathcal{T}}_{\mathcal{L}}(\mathcal{A})$ such that $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cens}(\mathcal{A})$ is a consistent FO theory.

The notion of more informative censor previously given for IB censors can be naturally extended to CP censors, and we can thus define optimal censors also in this case.

Definition 4 Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, and $\mathcal{L} \subseteq \mathbf{FO}$ be a language. A CP censor cens in $\mathcal{L}$ for $\mathcal{T}$ and $\mathcal{P}$ is *optimal* if there does not exist any other CP censor in $\mathcal{L}$ for $\mathcal{T}$ and $\mathcal{P}$ that is more informative than cens. The set of all the optimal CP censors in $\mathcal{L}$ for $\mathcal{T}$ and $\mathcal{P}$ is denoted with $\mathcal{L}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$.

Example 3 Consider $\mathcal{T}$ and $\mathcal{P}$ as defined in Example 1. An optimal CP censor cens_4 in $\mathbf{CQ}$ for $\mathcal{T}$ and $\mathcal{P}$ is defined as follows: given an ABox $\mathcal{A}$, $\mathsf{cens}_4(\mathcal{A})$ returns the set of BCQs obtained by removing from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ every query containing the atom $\mathsf{ProjA}(c)$, for each individual c such that both $\mathsf{ProjA}(c)$ and $\mathsf{ProjB}(c)$ are in $\mathcal{A}$.

We soon notice that the CP censor cens_4 is instead not an IB censor. Indeed, consider the ABox $\mathcal{A} = \{\mathsf{ProjA}(c), \mathsf{ProjB}(c)\}$. We have that $\mathsf{cens}_4(\mathcal{A}) = \{\phi \mid \phi \in \mathbf{CQ} \text{ and } \mathcal{T} \cup \mathcal{S} \models \phi\}$, where $\mathcal{S} = \{\exists x.\mathsf{ProjA}(x), \mathsf{ProjB}(c)\}$. It is not hard to see that there exists no ABox $\mathcal{A}'$ such that $\mathcal{A}'$ and $\mathcal{A}$ are indistinguishable w.r.t. cens_4 and $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is consistent. □

Let $\mathcal{A}$ be an ABox and cens be either an IB or a CP censor, the set $\mathsf{cens}(\mathcal{A})$ is called *theory of the censor* cens for $\mathcal{A}$.

The following theorem explains the relation between IB censors and CP censors.

Theorem 1 *Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. If* cens *is an IB censor for $\mathcal{T}$ and $\mathcal{P}$, then it is a CP censor in* $\mathbf{CQ}$ *for $\mathcal{T}$ and $\mathcal{P}$. The converse does not necessarily hold.*

Proof. Let cens be an IB censor for $\mathcal{T}$ and $\mathcal{P}$. Consider an arbitrary ABox $\mathcal{A}$. According to Definition 1, there exists an ABox $\mathcal{A}'$ such that $\mathsf{cens}(\mathcal{A}) = \mathsf{cens}(\mathcal{A}')$ and $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is consistent. Since by definition $\mathsf{cens}(\mathcal{A}')$ contains only sentences $\phi \in \mathbf{CQ}$ logically implied by $\mathcal{T} \cup \mathcal{A}'$ (i.e., BCQs ϕ such that $\phi \in \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}')$) and $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}'$ is consistent, we have that $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cens}(\mathcal{A}')$ is consistent as well. Due to the equivalence $\mathsf{cens}(\mathcal{A}') = \mathsf{cens}(\mathcal{A})$, we derive that $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cens}(\mathcal{A})$ is consistent. To conclude the implication part observe that, by definition, $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$.

As for the converse, Example 3 shows that the CP censor cens_4 in $\mathbf{CQ}$ for $\mathcal{T}$ and $\mathcal{P}$ is not an IB censor for $\mathcal{T}$ and $\mathcal{P}$. ■

We also notice that optimal IB censors are *not* necessarily optimal CP censors in $\mathbf{CQ}$. Indeed, consider Examples 1 and 3. We have that $\mathsf{cens}_1 \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$ but, even if, as shown by Theorem 1, it is a CP censor in $\mathbf{CQ}$ for $\mathcal{T}$ and $\mathcal{P}$, $\mathsf{cens}_1 \notin \mathbf{CQ}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$ (it is easy to see that cens_4 is more informative than cens_1). We also know from Example 3 that the optimal CP censor cens_4 in $\mathbf{CQ}$ for $\mathcal{T}$ and $\mathcal{P}$ is not an IB censor, and thus $\mathsf{cens}_4 \notin \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$. However, from Theorem 1 it easily follows that if an optimal CP censor in $\mathbf{CQ}$ for $\mathcal{T}$ and $\mathcal{P}$ is also an IB censor then it is an optimal IB censor for $\mathcal{T}$ and $\mathcal{P}$, as stated below.

Proposition 1 *Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. If $\mathsf{cens} \in \mathbf{CQ}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$ and cens is an IB censor for $\mathcal{T}$ and $\mathcal{P}$, then $\mathsf{cens} \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$. The converse does not necessarily hold.*

Actually, the relation between the two optimality notions of censor depends on the censor language adopted for the CP censors. In particular, for $\mathbf{GA}$, the set of the theories of the optimal IB censors for a TBox $\mathcal{T}$ and a policy $\mathcal{P}$ coincides with the set of the deductive closures $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\cdot)$ of the theories of the optimal CP censors in $\mathbf{GA}$ for $\mathcal{T}$ and $\mathcal{P}$. This property is formalized by the following theorem, which is crucial to establish the complexity results of the next section.

Theorem 2 *Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. Then, $\mathsf{ib\_cens} \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$ iff there exists a CP censor $\mathsf{cp\_cens} \in \mathbf{GA}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$ such that, for each ABox $\mathcal{A}$, $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathsf{cp\_cens}(\mathcal{A})) = \mathsf{ib\_cens}(\mathcal{A})$.*

Proof (sketch). ($\Leftarrow$). Since $\mathsf{cp\_cens} \in \mathbf{GA}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$, it is easy to see that $\mathsf{cp\_cens}(\mathcal{A}) = \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ for each ABox $\mathcal{A}$ such that $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$ is consistent. Thus, for each ABox $\mathcal{A}$, since $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cp\_cens}(\mathcal{A})$ is consistent and $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathsf{cp\_cens}(\mathcal{A})) = \mathsf{ib\_cens}(\mathcal{A})$ holds by assumption, we derive that $\mathsf{ib\_cens}(\mathcal{A}) = \mathsf{ib\_cens}(\mathsf{cp\_cens}(\mathcal{A}))$, i.e., all pairs of ABoxes $\mathcal{A}$ and $\mathsf{cp\_cens}(\mathcal{A})$ are indistinguishable w.r.t. $\mathsf{ib\_cens}$, and so $\mathsf{ib\_cens}$ is an IB censor for $\mathcal{T}$ and $\mathcal{P}$.

We now prove its optimality by contradiction. Suppose there is an IB censor $\mathsf{ib\_cens}'$ more informative than $\mathsf{ib\_cens}$. So $\mathsf{ib\_cens}(\mathcal{A}') \subset \mathsf{ib\_cens}'(\mathcal{A}') = \mathsf{ib\_cens}'(\mathcal{A}'_i) \subseteq \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}'_i)$, where $\mathcal{A}'$ is the ABox such that $\mathsf{ib\_cens}(\mathcal{A}') \subset \mathsf{ib\_cens}'(\mathcal{A}')$ and $\mathcal{A}'_i$ is the ABox such that $\mathcal{A}'$ and $\mathcal{A}'_i$ are indistinguishable w.r.t. $\mathsf{ib\_cens}'$ (one of such pairs $\mathcal{A}', \mathcal{A}'_i$ must exist). But then, the function $\mathsf{cp\_cens}'$ such that $\mathsf{cp\_cens}'(\mathcal{A}) = \mathsf{cp\_cens}(\mathcal{A})$ for each ABox $\mathcal{A} \neq \mathcal{A}'$ and $\mathsf{cp\_cens}'(\mathcal{A}') = \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A}'_i)$ can be shown to be a CP censor in $\mathbf{GA}$ for $\mathcal{T}$ and $\mathcal{P}$ more informative than $\mathsf{cp\_cens}$, which leads to a contradiction.

($\Rightarrow$). Since $\mathsf{ib\_cens} \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$, it can be shown that, for each ABox $\mathcal{A}$, $\mathsf{ib\_cens}(\mathcal{A}) = \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A})$ if $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$ is consistent, and $\mathsf{ib\_cens}(\mathcal{A}) = \mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathcal{A}')$ otherwise, where $\mathcal{A}'$ is the ABox such that $\mathcal{A}$ and $\mathcal{A}'$ are indistinguishable w.r.t. $\mathsf{ib\_cens}$. Let $\mathsf{cp\_cens}$ be the function such that, for each ABox $\mathcal{A}$, $\mathsf{cp\_cens}(\mathcal{A})$ returns all and only the ground atoms of $\mathsf{ib\_cens}(\mathcal{A})$, i.e., $\mathsf{cp\_cens}(\mathcal{A}) = \mathbf{GA} \cap \mathsf{ib\_cens}(\mathcal{A})$. From the above considerations, it is not hard to derive that $\mathsf{cp\_cens}$ is a CP censor in $\mathbf{GA}$ for $\mathcal{T}$ and $\mathcal{P}$ such that $\mathsf{cl}^{\mathcal{T}}_{\mathbf{CQ}}(\mathsf{cp\_cens}(\mathcal{A})) = \mathsf{ib\_cens}(\mathcal{A})$ for each ABox $\mathcal{A}$. Moreover, it can be shown by contradiction that $\mathsf{cp\_cens} \in \mathbf{GA}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$. ∎

Algorithm 1: OptGACensor

input: a *DL-Lite*$_{\mathcal{R}}$ TBox $\mathcal{T}$, a policy $\mathcal{P}$, an ABox $\mathcal{A}$;
output: an ABox;
1) $\mathcal{A}_{\mathcal{T}} \leftarrow \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$;
2) $Th \leftarrow \emptyset$;
3) **while** $\mathcal{A}_{\mathcal{T}}$ is not empty **do**:
4) **let** α be the lexicographically first assertion in $\mathcal{A}_{\mathcal{T}}$;
5) $\mathcal{A}_{\mathcal{T}} \leftarrow \mathcal{A}_{\mathcal{T}} \setminus \{\alpha\}$;
6) **if** $\mathcal{T} \cup Th \cup \{\alpha\} \cup \mathcal{P}$ is consistent **then**
7) $Th \leftarrow Th \cup \{\alpha\}$;
8) **return** Th;

5 Query Answering under Optimal IB Censors

In this section we study query answering under IB censors over *DL-Lite*$_{\mathcal{R}}$ ontologies. In particular, we consider entailment of BCQs specified over the signature of the ontology.

A possible strategy for addressing this problem is to choose only one IB censor among the optimal ones, and use it to alter the answers to user queries. In the absence of a criterion for determining which censor is the best for our purposes, the choice of the optimal censor is made in an arbitrary way (like in [Biskup and Bonatti, 2007; Cuenca Grau *et al.*, 2013]). Towards the realization of an optimal IB censor, we first provide the algorithm OptGACensor (Algorithm 1), which implements a function that, for every *DL-Lite*$_{\mathcal{R}}$ TBox $\mathcal{T}$ and every policy $\mathcal{P}$, corresponds to an optimal CP censor in $\mathbf{GA}$ for $\mathcal{T}$ and $\mathcal{P}$. Then we explain how to use OptGACensor to establish BCQs entailment under an optimal IB censor by exploiting Theorem 2. The algorithm first computes the set $\mathcal{A}_{\mathcal{T}}$ of ground atoms entailed by $\mathcal{T} \cup \mathcal{A}$. Then, it iteratively picks a ground atom α from $\mathcal{A}_{\mathcal{T}}$ following the lexicographic order, and adds α to the ABox Th if $\mathcal{T} \cup Th \cup \alpha$ does not violate the policy $\mathcal{P}$. The following theorem establishes the correctness and complexity of the algorithm.

Theorem 3 *Let $\mathcal{T}$ be a DL-Lite*$_{\mathcal{R}}$ *TBox and $\mathcal{P}$ be a policy. There exists a censor $\mathsf{cens} \in \mathbf{GA}\text{-}\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$ such that, for each ABox $\mathcal{A}$, OptGACensor$(\mathcal{T},\mathcal{P},\mathcal{A})$ (i) returns $\mathsf{cens}(\mathcal{A})$ and (ii) runs in polynomial time in the size of $\mathcal{A}$.*

Proof (sketch). The proof of correctness is straightforward. Tractability in the size of $\mathcal{A}$ follows from the fact that $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ can be computed in polynomial time, and the consistency check of step 6 can be done in AC^0 in the size of $\mathcal{A}$ [Lembo *et al.*, 2015]. ∎

From Theorem 2 and Theorem 3 it follows that, to establish if a BCQ q is entailed by $\mathcal{T} \cup \mathcal{A}$ under an optimal IB censor for $\mathcal{T}$ and $\mathcal{P}$, it is sufficient to verify whether $\mathcal{T} \cup \mathsf{OptGACensor}(\mathcal{T},\mathcal{P},\mathcal{A}) \models q$, which can be done in polynomial time in the size of $\mathcal{A}$.

We note also that it is possible to implement different optimal IB censors (actually, every optimal IB censor) by modifying the order in which the ABox assertions from the set $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ are selected by the algorithm.

Depending on the application at hand, the approach of randomly choosing a censor may not always be considered

appropriate [Cuenca Grau *et al.*, 2013]. For this reason, in [Lembo *et al.*, 2019] the authors suggest to use a form of skeptical entailment over (the theories of) all the optimal censors, i.e., they propose a CQE framework in which a query has a positive answer if it is entailed by each optimal censor. In the same spirit, we define the following decision problem.

Definition 5 Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, $\mathcal{A}$ be an ABox, and q be a BCQ. IB-Entailment$(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is the problem of deciding whether $q \in \mathsf{cens}(\mathcal{A})$ for every cens in $\mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$.

By exploiting Theorem 2 and the results given in [Lembo *et al.*, 2019], we can provide the following theorem.

Theorem 4 *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, $\mathcal{P}$ be a policy, $\mathcal{A}$ be an ABox, and q be a BCQ. Then, IB-Entailment$(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is coNP-complete in data complexity.*

Proof. The result immediately follows from Theorem 2 and from [Lembo *et al.*, 2019, Theorem 6], which states that deciding if $\mathcal{T} \cup \mathsf{cp\_cens}(\mathcal{A}) \models q$ for every cp_cens in $\mathbf{GA}$-$\mathsf{OptCPCens}_{\mathcal{T},\mathcal{P}}$ is coNP-complete in data complexity. ∎

6 Approximating Optimal IB Censors

As stated in Theorem 4, IB-Entailment is in general intractable in data complexity. Towards a practical approach to CQE, in this section we consider a different entailment problem that approximates IB-Entailment, and we show that its data complexity is in AC^0 (i.e., the same complexity of evaluating FO queries over a database). The approximation we propose consists in considering a non-necessarily optimal IB censor whose theory, for every ABox, is as close as possible to the theories of all the optimal IB censors.

Definition 6 [AIB censor and QIB censor] Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, and cens be an IB censor for $\mathcal{T}$ and $\mathcal{P}$. We say that:

- (i) cens is an *approximation of the optimal IB censors (AIB censor)* for $\mathcal{T}$ and $\mathcal{P}$ if, for every $\mathsf{cens}' \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$ and for every ABox $\mathcal{A}$, $\mathsf{cens}(\mathcal{A}) \subseteq \mathsf{cens}'(\mathcal{A})$;
- (ii) cens is a *quasi-optimal IB censor (QIB censor)* for $\mathcal{T}$ and $\mathcal{P}$ if cens is an AIB censor for $\mathcal{T}$ and $\mathcal{P}$ and there exists no AIB censor cens′ for $\mathcal{T}$ and $\mathcal{P}$ that is more informative than cens.

Example 4 The IB censor cens_3 of Example 1 is a QIB censor for $\mathcal{T}$ and $\mathcal{P}$ (but $\mathsf{cens}_3 \notin \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$). □

For QIB censors the following notable property holds.

Theorem 5 *Let $\mathcal{T}$ be a DL TBox and $\mathcal{P}$ be a policy. A QIB censor for $\mathcal{T}$ and $\mathcal{P}$ always exists and it is unique.*

Proof (sketch). First, observe that the censor cens_0 such that $\mathsf{cens}_0(\mathcal{A}) = \emptyset$ for every ABox $\mathcal{A}$, satisfies condition (i) of Definition 6. So, either cens_0 is a QIB censor, or there exists a more informative AIB censor. This implies the existence of a QIB censor. Then, it is possible to show that if two distinct QIB censors cens_a, cens_b exist (i.e., such that $\mathsf{cens}_\mathsf{a}(\mathcal{A}) \neq \mathsf{cens}_\mathsf{b}(\mathcal{A})$ for at least one ABox $\mathcal{A}$), it is possible to construct another AIB censor $\mathsf{cens}'_\mathsf{a}$ that is more informative than cens_a (or, more informative than cens_b), such that $\mathsf{cens}'_\mathsf{a}(\mathcal{A}') = \mathsf{cens}_\mathsf{a}(\mathcal{A}')$ for all $\mathcal{A}' \neq \mathcal{A}$, and $\mathsf{cens}'_\mathsf{a}(\mathcal{A}) = \mathsf{cl}^\mathcal{T}_\mathbf{CQ}(\mathcal{A}_1 \cup \mathcal{A}_2)$, where $\mathcal{A}_1$ and $\mathcal{A}_2$ are two ABoxes indistinguishable with $\mathcal{A}$ w.r.t. cens_a and cens_b, respectively. This leads to a contradiction. ∎

Hereinafter, we denote with $\mathsf{qib\_cens}_{\mathcal{T},\mathcal{P}}$ the QIB censor for $\mathcal{T}$ and $\mathcal{P}$. Entailment of BCQs over QIB censors is then naturally defined as follows.

Definition 7 Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, $\mathcal{A}$ be an ABox, and q be a BCQ. QIB-Entailment$(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is the problem of deciding whether $q \in \mathsf{qib\_cens}_{\mathcal{T},\mathcal{P}}(\mathcal{A})$.

We now focus on the case of *DL-Lite$_\mathcal{R}$* TBoxes and prove that, in this case, entailment of BCQs under QIB censors is FO-rewritable. Formally, we say that QIB-entailment in a DL $\mathcal{L}$ is FO-rewritable, if for every TBox $\mathcal{T}$ expressed in $\mathcal{L}$, every policy $\mathcal{P}$ and every BCQ q, one can effectively compute an FO query q_r such that for every ABox $\mathcal{A}$, QIB-Entailment$(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is true iff $\mathcal{A} \models q_r$. We call q_r the *QIB-perfect reformulation* of q w.r.t. $\mathcal{T}$ and $\mathcal{P}$.

We prove FO-rewritability of entailment of BCQs under QIB censors in *DL-Lite$_\mathcal{R}$* by exploiting a correspondence between this problem and entailment of BCQs under IAR-semantics for DL ontologies, which is indeed FO-rewritable for *DL-Lite$_{\mathcal{R},den}$*, i.e., *DL-Lite$_\mathcal{R}$* enriched with denial assertions [Lembo *et al.*, 2015]. We recall that the IAR-semantics is an inconsistency-tolerant semantics that allows for meaningful entailment also when the ABox contradicts the TBox of an ontology (for instance, when $\mathcal{A} = \{A(d), B(d), C(d)\}$ and $\mathcal{T} = \{A \sqsubseteq \neg B\}$). The IAR-semantics is based on the notion of *ABox repair* (A-repair), which is a maximal subset of the ABox that is consistent with the TBox (in our example there are two A-repairs, $\mathcal{R}_1 = \{A(d), C(d)\}$ and $\mathcal{R}_2 = \{B(d), C(d)\}$). Then, entailment under IAR-semantics is defined as follows: let $\mathcal{T}$ be a *DL-Lite$_{\mathcal{R},den}$* TBox, $\mathcal{A}$ be an ABox, and q be a BCQ, IAR-Entailment$(\mathcal{T}, \mathcal{A}, q)$ is the problem of verifying whether $\mathcal{T} \cup \mathcal{R}_{iar} \models q$, where $\mathcal{R}_{iar}$ is the intersection of all A-repairs of $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$, called the IAR-repair of $\mathcal{O}$ (in our example, $\mathcal{R}_{iar} = \{C(d)\}$).

To establish the relationship between QIB-entailment and IAR-entailment, we give the notion of secret, which in our framework is the counterpart of minimal inconsistent set in inconsistency-tolerant query answering [Lembo *et al.*, 2015].

Let $\mathcal{T}$ be a TBox, let $\mathcal{P}$ be a policy, and let $\mathcal{A}$ be an ABox. We say that a set of ABox assertions $\mathcal{S} \subseteq \mathsf{cl}^\mathcal{T}_\mathbf{GA}(\mathcal{A})$ is a *secret* in $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$, if $\mathcal{T} \cup \mathcal{P} \cup \mathcal{S}$ is inconsistent and for each assertion $\sigma \in \mathcal{S}$ we have that $\mathcal{T} \cup \mathcal{P} \cup \mathcal{S} \setminus \{\sigma\}$ is consistent. We denote by $secrets(\mathcal{T}, \mathcal{P}, \mathcal{A})$ the set of all the secrets in $\mathcal{T} \cup \mathcal{P} \cup \mathcal{A}$.

We now provide the following key property.

Lemma 1 *Let $\mathcal{T}$ be a DL TBox, $\mathcal{P}$ be a policy, $\mathcal{A}$ be an ABox, and q be a BCQ. QIB-Entailment$(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is true iff there exists an ABox $\mathcal{A}' \subseteq \mathsf{cl}^\mathcal{T}_\mathbf{GA}(\mathcal{A})$ such that (i) $\mathcal{T} \cup \mathcal{A}' \models q$, and (ii) $\mathcal{A}' \cap \mathcal{S} = \emptyset$, for each secret $\mathcal{S} \in secrets(\mathcal{T}, \mathcal{P}, \mathcal{A})$.*

Proof (sketch). ($\Leftarrow$). It is not difficult to show that given an ABox assertion $\alpha \in \mathsf{cl}^\mathcal{T}_\mathbf{GA}(\mathcal{A})$, there exists an optimal IB censor $\mathsf{cens} \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$ such that $\alpha \notin \mathsf{cens}(\mathcal{A})$ only if there exists a secret $\mathcal{S}$ in $secrets(\mathcal{T}, \mathcal{P}, \mathcal{A})$ such that $\alpha \in \mathcal{S}$.

Now, from the above property, conditions (i) and (ii) we have that $\mathcal{A}' \subseteq \mathsf{cens}(\mathcal{A})$ for every $\mathsf{cens} \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$. This means that $\mathcal{A}' \subseteq \mathsf{qib\_cens}_{\mathcal{T},\mathcal{P}}(\mathcal{A})$. Moreover, since $\mathcal{T} \cup \mathcal{A}' \models q$, we have that $\mathcal{T} \cup \mathsf{qib\_cens}_{\mathcal{T},\mathcal{P}}(\mathcal{A}) \models q$.

($\Rightarrow$). It is not hard to see that there exists an ABox $\mathcal{A}'$ satisfying condition (i) ($\mathcal{A}'$ is the ABox indistinguishable from $\mathcal{A}$ w.r.t. qib_cens). As for condition (ii), suppose that there exists an ABox assertion $\alpha \in \mathcal{A}'$ and a secret $\mathcal{S} \in \mathit{secrets}(\mathcal{T},\mathcal{P},\mathcal{A})$ such that $\alpha \in \mathcal{S}$. From Definition 6, we have that $\alpha \in \mathsf{cens}'(\mathcal{A})$ for every $\mathsf{cens}' \in \mathsf{OptIBCens}_{\mathcal{T},\mathcal{P}}$. Since $\mathcal{S} \setminus \{\alpha\}$ is consistent with $\mathcal{T} \cup \mathcal{P}$, it is possible to define an optimal IB censor whose theory contains $\mathcal{S} \setminus \{\alpha\}$, which is a contradiction. ∎

The following theorem establishes the relationship between QIB-entailment and IAR-entailment.

Theorem 6 *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, let $\mathcal{P}$ be a policy, let $\mathcal{A}$ be an ABox, and let q be a BCQ. QIB-Entailment$(\mathcal{T},\mathcal{P},\mathcal{A},q)$ is true iff IAR-Entailment$(\mathcal{T} \cup \mathcal{P}, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A}), q)$ is true.*

Proof. Since $\mathcal{T} \cup \mathcal{A}$ is consistent, then the secrets in $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ coincide with the minimal subsets of $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ that are inconsistent with $\mathcal{T} \cup \mathcal{P}$. Therefore, the IAR-Repair $\mathcal{R}$ of $\langle \mathcal{T} \cup \mathcal{P}, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})\rangle$ is the set of ground atoms from $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ that do not belong to any secret in $\mathcal{T} \cup \mathcal{P} \cup \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$. Thus, from Lemma 1 the thesis follows. ∎

Theorem 6 actually states that, to solve QIB-entailment, we can resort to the query rewriting techniques used to establish IAR-entailment given in [Lembo *et al.*, 2015], provided that we compute $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$. We recall that query entailment under IAR-semantics in a DL $\mathcal{L}$ is FO-rewritable if for every TBox $\mathcal{T}$ expressed in $\mathcal{L}$ and every BCQ q, one can effectively compute an FO query q_r such that for every ABox $\mathcal{A}$, IAR-Entailment$(\mathcal{T},\mathcal{A},q)$ is true iff $\mathcal{A} \models q_r$. The query q_r is called the *IAR-perfect reformulation* of q w.r.t. $\mathcal{T}$.

To establish FO-rewritability of QIB-entailment in *DL-Lite*$_\mathcal{R}$, however, we still need to address the above mentioned computation of $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$, and turn it into an additional query reformulation step. To this aim, we can exploit the fact that, for a *DL-Lite*$_{\mathcal{R},den}$ ontology $\mathcal{T} \cup \mathcal{A}$, an FO query q evaluates to true over $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})$ iff q' evaluates to true over $\mathcal{A}$, where q' is obtained by suitably rewriting each atom of q according to the positive inclusions of $\mathcal{T}$. Intuitively, in this way we cast into the query all the possible causes of the facts that are contained in the closure of the ABox w.r.t. the TBox (similarly to what is done in query rewriting algorithms for *DL-Lite* [Calvanese *et al.*, 2007]).

To compute such a query q', we use the function atomRewr$(q,\mathcal{T})$, which substitutes each atom α of q with the formula $\phi(\alpha)$ defined as follows (where A, B are atomic concepts and R, S are atomic roles):

$$\phi(A(t)) = \bigvee_{\mathcal{T} \models B \sqsubseteq A} B(t) \vee \bigvee_{\mathcal{T} \models \exists R \sqsubseteq A} (\exists x.R(t,x)) \vee \bigvee_{\mathcal{T} \models \exists R^- \sqsubseteq A} (\exists x.R(x,t))$$

$$\phi(R(t_1,t_2)) = \bigvee_{\mathcal{T} \models S \sqsubseteq R} S(t_1,t_2) \vee \bigvee_{\mathcal{T} \models S^- \sqsubseteq R} S(t_2,t_1).$$

The following lemma, whose proof can be immediately obtained from the definitions of $\mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\cdot)$ and atomRewr$(\cdot,\cdot)$, states the property we are looking for.

Lemma 2 *Let $\mathcal{T}$ be a DL-Lite$_{\mathcal{R},den}$ TBox, $\mathcal{A}$ be an ABox, and q be an FO sentence. Then Eval$(q, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})) =$ Eval$($atomRewr$(q,\mathcal{T}),\mathcal{A})$.*

We are now able to extablish FO-rewritability of QIB-entailment.

Theorem 7 *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, $\mathcal{P}$ be a policy, q be a BCQ, and q_r be an FO sentence that is a IAR-perfect reformulation of q w.r.t. the DL-Lite$_{\mathcal{R},den}$ TBox $\mathcal{T} \cup \mathcal{P}$. Then, the FO sentence atomRewr$(q_r,\mathcal{T})$ is a QIB-perfect reformulation of q w.r.t. $\mathcal{T}$ and $\mathcal{P}$.*

Proof. Let the FO sentence q_r be an IAR-perfect reformulation of q w.r.t. the *DL-Lite*$_{\mathcal{R},den}$ TBox $\mathcal{T} \cup \mathcal{P}$. Then, for every ABox $\mathcal{A}$, IAR-Entailment$(\mathcal{T} \cup \mathcal{P}, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A}), q)$ is true iff $\mathit{Eval}(q_r, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A}))$ is true. Now, from Lemma 2, it follows that, for every ABox $\mathcal{A}$, $\mathit{Eval}(q_r, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A})) = \mathit{Eval}($atomRewr$(q_r,\mathcal{T}),\mathcal{A})$. Since by Theorem 6, for every ABox $\mathcal{A}$ such that $\mathcal{T} \cup \mathcal{A}$ is consistent, IAR-Entailment$(\mathcal{T} \cup \mathcal{P}, \mathsf{cl}^{\mathcal{T}}_{\mathbf{GA}}(\mathcal{A}), q)$ is true iff QIB-Entailment$(\mathcal{T},\mathcal{P},\mathcal{A},q)$ is true, it follows that the FO sentence atomRewr$(q_r,\mathcal{T})$ is a QIB-perfect reformulation of q w.r.t. $\mathcal{T}$ and $\mathcal{P}$. ∎

Since IAR-entailment is actually FO rewritable, as shown in [Lembo *et al.*, 2015], the above theorem proves the FO rewritability of QIB-entailment for *DL-Lite*$_\mathcal{R}$ TBoxes. Moreover, the above theorem identifies a technique for obtaining the QIB-perfect reformulation of a CQ, based on a simple combination of the IAR-perfect reformulation algorithm of [Lembo *et al.*, 2015] and the atomRewr reformulation defined above. Therefore:

Corollary 1 *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, $\mathcal{P}$ be a policy, $\mathcal{A}$ be an ABox, and q be a BCQ. The problem QIB-Entailment$(\mathcal{T},\mathcal{P},\mathcal{A},q)$ is in* AC^0 *in data complexity.*

7 Conclusions

In this paper, we have studied the approach to CQE based on instance indistinguishability and identified a semantically well-founded notion of CQE that enjoys first-order rewritability in the case of *DL-Lite*$_\mathcal{R}$ ontologies. We believe that this result opens the way towards practical implementations of CQE engines for DL ontologies and Ontology-based Data Access. We are currently working to achieve this goal.

Another important future direction is a deeper study of the user model. Our framework inherits from its predecessors a relatively simple model, which assumes that 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.

Acknowledgements

This work was partly supported by the EU within the H2020 Programme under the grant agreement 834228 (ERC Advanced Grant WhiteMec) and the grant agreement 825333 (MOSAICrOWN), by Regione Lombardia within the Call Hub Ricerca e Innovazione under the grant agreement 1175328 (WATCHMAN), and by Sapienza Università di Roma (2019 project CQEinOBDM).

References

[Artale *et al.*, 2009] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research*, 36:1–69, 2009.

[Baader *et al.*, 2007] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2nd edition, 2007.

[Benedikt *et al.*, 2018] Michael Benedikt, Bernardo Cuenca Grau, and Egor V. Kostylev. Logical foundations of information disclosure in ontology-based data integration. *Artificial Intelligence*, 262:52–95, 2018.

[Benedikt *et al.*, 2019] Michael Benedikt, Pierre Bourhis, Louis Jachiet, and Michaël 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.

[Biskup and Bonatti, 2004a] Joachim Biskup and Piero A. Bonatti. Controlled query evaluation for enforcing confidentiality in complete information systems. *Int. J. of Information Security*, 3(1):14–27, 2004.

[Biskup and Bonatti, 2004b] Joachim Biskup and Piero 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.

[Biskup and Bonatti, 2007] Joachim Biskup and Piero A. Bonatti. Controlled query evaluation with open queries for a decidable relational submodel. *Ann. of Mathematics and Artificial Intelligence*, 50(1–2):39–77, 2007.

[Biskup and Weibert, 2008] Joachim Biskup and Torben Weibert. Keeping secrets in incomplete databases. *Int. J. of Information Security*, 7(3):199–217, 2008.

[Biskup, 2000] Joachim Biskup. For unknown secrecies refusal is better than lying. *Data and Knowledge Engineering*, 33(1):1–23, 2000.

[Bonatti and Sauro, 2013] Piero A. Bonatti and Luigi 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.

[Bonatti *et al.*, 1995] Piero A. Bonatti, Sarit Kraus, and V. S. Subrahmanian. Foundations of secure deductive databases. *IEEE Trans. Knowl. Data Eng.*, 7(3):406–422, 1995.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2012] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. View-based query answering in description logics: Semantics and complexity. *J. of Computer and System Sciences*, 78(1):26–46, 2012.

[Cima *et al.*, 2020] Gianluca Cima, Domenico Lembo, Riccardo Rosati, and Domenico Fabio Savo. CQE in description logics through instance indistinguishability (extended version). *CoRR*, abs/2004.11870, 2020.

[Cuenca Grau and Horrocks, 2008] Bernardo Cuenca Grau and Ian 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.

[Cuenca Grau *et al.*, 2013] Bernardo Cuenca Grau, Evgeny Kharlamov, Egor V. Kostylev, and Dmitriy 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.

[Cuenca Grau *et al.*, 2015] Bernardo Cuenca Grau, Evgeny Kharlamov, Egor V. Kostylev, and Dmitriy 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.

[Lembo *et al.*, 2015] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Inconsistency-tolerant query answering in ontology-based data access. *J. of Web Semantics*, 33:3–29, 2015.

[Lembo *et al.*, 2019] Domenico Lembo, Riccardo Rosati, and Domenico Fabio Savo. Revisiting controlled query evaluation in description logics. In *Proc. of the 28th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1786–1792, 2019.

[Motik *et al.*, 2012] Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz. OWL 2 Web Ontology Language profiles (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. Available at http://www.w3.org/TR/owl2-profiles/.

[Sicherman *et al.*, 1983] George L. Sicherman, Wiebren de Jonge, and Reind P. van de Riet. Answering queries without revealing secrets. *ACM Trans. Database Syst.*, 8(1):41–59, 1983.

[Studer and Werner, 2014] Thomas Studer and Johannes Werner. Censors for boolean description logic. *Trans. Data Privacy*, 7(3):223–252, 2014.

[Tao *et al.*, 2014] Jia Tao, Giora Slutzki, and Vasant G. Honavar. A conceptual framework for secrecy-preserving reasoning in knowledge bases. *ACM Trans. on Computational Logic*, 16(1):3:1–3:32, 2014.

[Xiao *et al.*, 2018] Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In *Proc. of the 27th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 5511–5519, 2018.