CQE in Description Logics Through Instance Indistinguishability
(extended version)

Gianluca Cima¹, Domenico Lembo¹, Riccardo Rosati¹ and Domenico Fabio Savo²

¹Sapienza Università di Roma
²Università 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-LiteR 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-LiteR ontologies, this form of CQE is tractable with respect to data complexity and is first-order re-writable, 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

\[ \forall x. \text{Supplier}(x) \land \text{ProjA}(x) \land \text{ProjB}(x) \rightarrow \bot \]

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 \{Supplier(c), ProjA(c), ProjB(c)\}, an IB system should answer queries as if the instance were, e.g., \{Supplier(c), ProjA(c)\} (note that other instances not violating the policy can be considered as indistinguishable, e.g., \{Supplier(c), 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 Supplier(c) ∧ ProjA(c) and \( \exists x. \text{Supplier}(x) \land \text{ProjB}(x) \), but not also to the query Supplier(c) ∧ 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.\textsuperscript{1} 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-Lite\textsubscript{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-Lite\textsubscript{R}, query answering in the IB approach is tractable 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\textsubscript{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\textsubscript{R}) specifically designed for data management, has the same data complexity as evaluating queries over a database (i.e., AC\textsuperscript{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].

\section{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 $\Sigma_V$ 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 intentional 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 $\phi$ specified over the signature of $\mathcal{O}$, denoted $\mathcal{O} \models \phi$, if $\phi$ is true in every model of $\mathcal{O}$. In this paper, we consider ontologies expressed in DL-Lite\textsubscript{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\textsubscript{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: $B_1, B_2$ 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., unequivocal 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 $\mathcal{A}(a, b)$, 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 $\Sigma_V$, 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 $\Sigma_I$ or a variable in $\vec{x}$. We denote by $\text{Eval}(q, \mathcal{A})$ 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 $\delta$ 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 FO, CQ, and 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 $\Sigma_V$. 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 $\Sigma_V$. For a TBox $\mathcal{T}$ and a language $\mathcal{L}$, we denote by $c_{\mathcal{L}}^{\mathcal{T}}(\cdot)$ 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\textsuperscript{2}. Our complexity results are for data complexity, i.e., are w.r.t. the size of the ABox only.

\section{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 alters query answers so that by uniting the answers (even a possibly infinite set thereof) with the TBox a user can never infer a BCQ $\exists \vec{x}. \phi(\vec{x})$, for each denial $\forall \vec{x}. \phi(\vec{x}) \rightarrow \bot$ in $\mathcal{P}$.

\textsuperscript{1}Privacy-preserving query answering in DLs has been investigated also in settings different from CQE: see, e.g., [Cuenca Grau and Horrocks, 2008; Calvanese et al., 2012; Tao et al., 2014].

\textsuperscript{2}It is also easy to see that, since DL-Lite\textsubscript{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.
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; 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 $T$ be a DL TBox and $P$ be a policy. An indistinguishability-based (IB) censor for $T$ and $P$ is a function $cens(\cdot)$ that, for each ABox $A$, returns a set $cens(A) \subseteq cl_{CQ}(A)$ such that there exists an ABox $A'$ for which (i) $cens(A) = cens(A')$ (in this case we say that $A$ and $A'$ are indistinguishable w.r.t. cens) and (ii) $T \cup P \cup A'$ 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 $\forall A \subseteq Supplier$ and $ProjB \sqsubseteq Supplier$, stating that each individual instance of ProjA or ProjB is also instance of Supplier. Data protection is specified through the policy $P = \{ \forall x.ProjA(x) \land ProjB(x) \rightarrow \bot \}$. The following functions are IB censors for $T$ and $P$:

- **cens**\(_1\): given an ABox $A$, $cens_1(A)$ returns the set $cl_{CQ}(A_{ProjA})$ of BCQs, where $A_{ProjA}$ is obtained from $A$ by removing the assertion ProjA(c), for each individual $c$ such that both ProjA(c) and ProjB(c) are in $A$ (note that for every ABox $A$, $A$ and $A_{ProjA}$ are indistinguishable w.r.t. $cens_1$. Similarly in the following censors).
- **cens**\(_2\): given an ABox $A$, $cens_2(A)$ returns the set $cl_{CQ}(A_{ProjB})$ of BCQs, where $A_{ProjB}$ is obtained from $A$ by removing the assertion ProjB(c), for each individual $c$ such that both ProjA(c) and ProjB(c) are in $A$.
- **cens**\(_3\): given an ABox $A$, $cens_3(A)$ returns the set $cl_{CQ}(A_{ProjA} \cup A_{ProjB})$ of BCQs, where $A_{ProjA} \cup A_{ProjB}$ is obtained from $A$ by adding the assertion Supplier(c) and removing ProjA(c) and ProjB(c), for each individual $c$ such that both ProjA(c) and ProjB(c) are in $A$.

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

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

**Example 2** Among the censors of Example 1, $cens_3 \notin OptIBCens_{T,P}$, 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 $T$ and $P$.

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, where confidentiality-preserving (CP) censor in $L$ for $T$ and $P$ is a function $cens(\cdot)$ that, for each ABox $A$, returns a set $cens(A) \subseteq cl_L(A)$ such that $T \cup P \cup cens(A)$ 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 3** [Confidentiality-preserving censor] Let $T$ be a DL TBox, $P$ be a policy, and $L \subseteq FO$ be a language. A confidentiality-preserving (CP) censor in $L$ for $T$ and $P$ is optimal if there does not exist any other CP censor in $L$ for $T$ and $P$ that is more informative than $cens$. The set of all the optimal CP censors in $L$ for $T$ and $P$ is denoted with $L$-OptCPCens_{T,P}.

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

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

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

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

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

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

As for the converse, Example 3 shows that the CP censor $\text{cens}_4$ in CQ for $T$ and $P$ is not an IB censor for $T$ and $P$. We also notice that optimal IB censors are not necessarily optimal CP censors in CQ. Indeed, consider Examples 1 and 3. We have that $\text{cens}_1 \in \text{OptIBCens}_{T,P}$ but, even if, as shown by Theorem 1, it is a CP censor in CQ for $T$ and $P$, $\text{cens}_1 \notin \text{CQ-OptCPCens}_{T,P}$ (it is easy to see that $\text{cens}_1$ is more informative than $\text{cens}_2$). We also know from Example 3 that the optimal CP censor $\text{cens}_3$ in CQ for $T$ and $P$ is not an IB censor, and thus $\text{cens}_4 \notin \text{OptIBCens}_{T,P}$. However, if an optimal CP censor in CQ for $T$ and $P$ is also an IB censor then it is an optimal IB censor for $T$ and $P$, as stated below.

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

**Proof.** Theorem 1 implies that the set $\text{IB}$ of IB censors for $T$ and $P$ is a subset of the set $\text{CP}$ of CP censors in CQ for $T$ and $P$. Thus, since for a censor $\text{cens} \in \text{CQ-OptCPCens}_{T,P}$ there does not exist in $\text{CP}$ a censor $\text{cens}'$ that is more informative than $\text{cens}$, such $\text{cens}'$ cannot exist in $\text{IB}$ too.

As a counterexample for the converse, as said above, $\text{cens}_3$ is in $\text{OptIBCens}_{T,P}$ but not in $\text{CQ-OptCPCens}_{T,P}$.

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

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

**Proof.** $(\Leftarrow)$. Suppose that there exists a CP censor $\text{cp.cens} \in \text{GA-OptCPCens}_{T,P}$ such that $\text{cl}_{\text{CQ}}(\text{cp.cens}(A)) = \text{ib.cens}(A)$ for each ABox $A$. Observe that, since $\text{cp.cens}$ is an optimal censor in GA for $T$ and $P$, we have that (i) $\text{cp.cens}(A) = \text{cl}_{\text{GA}}(\text{cp.cens}(A))$ for each ABox $A$ such that $T \cup P \cup \text{cp.cens}(A)$ is consistent (otherwise, we easily get a contradiction on the optimality of $\text{cp.cens}$), and (ii) $T \cup P \cup \text{cp.cens}(A)$ is consistent for each ABox $A$, where $\text{cp.cens}(A)$ can be seen as another ABox. From the above considerations, and the fact that $\text{ib.cens}(A) = \text{cl}_{\text{CQ}}(\text{cp.cens}(A))$ holds by assumption for each ABox $A$, we have that, for each ABox $A$, also the following hold: (i) $\text{ib.cens}(\text{cp.cens}(A)) = \text{ib.cens}(A)$ (i.e., $A$ and $\text{cp.cens}(A)$ are indistinguishable w.r.t. $\text{ib.cens}$), and (ii) $T \cup P \cup \text{ib.cens}(A)$ is consistent because $T \cup P \cup \text{cp.cens}(A)$ is consistent. This, together with the fact that $\text{ib.cens}(A) \subseteq \text{cl}_{\text{CQ}}(A)$ for each ABox $A$ (since $\text{ib.cens}(A) = \text{cl}_{\text{CQ}}(\text{cp.cens}(A))$ and $\text{cp.cens}(A) \subseteq \text{cl}_{\text{GA}}(\text{cp.cens}(A))$), implies that $\text{ib.cens}$ is an IB censor for $T$ and $P$.

We now prove its optimality by way of contradiction. Suppose, for the sake of contradiction, that $\text{ib.cens}$ is not an optimal IB censor for $T$ and $P$, i.e., there exists an IB censor $\text{ib.cens}'$ for $T$ and $P$ such that $\text{ib.cens}'(A) \subseteq \text{ib.cens}(A)$ for each ABox $A$, and there exists an ABox $A'$ such that $\text{ib.cens}'(A') \not\subseteq \text{ib.cens}(A')$. Since $\text{ib.cens}'$ is an IB censor for $T$ and $P$, there is an ABox $A'$ such that $A'$ and $A'$ are indistinguishable w.r.t. $\text{ib.cens}'$ (i.e., $\text{ib.cens}'(A') = \text{ib.cens}'(A')$) and $T \cup P \cup A'$ is consistent. Since by definition $\text{ib.cens}'(A') \subseteq \text{cl}_{\text{CQ}}(A')$, the following inclusions hold:

$$\text{ib.cens}'(A') \subseteq \text{ib.cens}'(A') = \text{ib.cens}'(A') \subseteq \text{cl}_{\text{CQ}}(A').$$

By assumption, moreover, we know that $\text{cl}_{\text{CQ}}(\text{cp.cens}(A'))) = \text{ib.cens}'(A')$, and therefore $\text{cl}_{\text{CQ}}(\text{cp.cens}(A'))) \subseteq \text{cl}_{\text{CQ}}(A')$. It follows that $\text{cp.cens}(A') \subseteq A'$, i.e., there is a ground atom $\psi$ such that $\psi \in A'$ and $\psi \notin \text{cp.cens}(A')$. But then, consider the function $\text{cp.cens}'$ with $\text{cp.cens}'(A') = \text{cp.cens}(A')$ for each ABox $A$ such that $A \neq A'$ and $\text{cp.cens}'(A') = \text{cl}_{\text{GA}}(A')$. Clearly, due to the facts that $\text{cp.cens}$ is a CP censor in $\text{GA}$ for $T$ and $P$ and $T \cup P \cup \text{cl}_{\text{GA}}(A')$ is consistent (because $T \cup P \cup A'$ is consistent), we have that $\text{cp.cens}'$ is a CP censor in GA for $T$ and $P$ as well. Observe, however, that $\text{cp.cens}(A) \subseteq \text{cp.cens}(A)$ for each ABox $A$, and $\text{cp.cens}'(A') \subseteq \text{cp.cens}(A') = \text{cl}_{\text{GA}}(A')$. In particular, the ground atom $\psi$ is such that $\psi \in A'$ (and thus $\psi \in \text{cp.cens}(A') = \text{cl}_{\text{GA}}(A')$) and $\psi \notin \text{cp.cens}(A')$. Therefore $\text{cp.cens}'$ is a CP censor in $\text{GA}$ for $T$ and $P$ that is more informative than $\text{cp.cens}$, and this contradicts the assumption that $\text{cp.cens} \in \text{GA-OptCPCens}_{T,P}$, as required.

$(\Rightarrow)$ In the proof, we will make use of the following claim.

**Claim 1** Let $T$ be a DL TBox, $P$ be a policy, and $\text{ib.cens}$ be an IB censor for $T$ and $P$. If $\text{ib.cens} \in \text{OptIBCens}_{T,P}$, then the following hold:

(i) $\text{ib.cens}(A) = \text{cl}_{\text{CQ}}(A)$ for each ABox $A$ such that $T \cup P \cup A$ is consistent.

(ii) $\text{ib.cens}(A) = \text{cl}_{\text{CQ}}(A') = \text{ib.cens}(A')$ for each ABox $A$, where $A'$ is the ABox such that $A$ and $A'$ are indistinguishable w.r.t. $\text{ib.cens}$ and $T \cup P \cup A'$ is consistent (such an ABox $A'$ is guaranteed to exists due to the fact that $\text{ib.cens}$ is an IB censor for $T$ and $P$).

**Proof.** Assume that $\text{ib.cens} \in \text{OptIBCens}_{T,P}$.

Suppose, for the sake of contradiction, that (i) does not hold, i.e., there exists an ABox $A$, such that $\text{ib.cens}(A) \subseteq \text{cl}_{\text{CQ}}(A)$ and $T \cup P \cup A$ is consistent. But then, consider the function $\text{ib.cens}'$ with $\text{ib.cens}'(A) = \text{cl}_{\text{CQ}}(A)$ for each ABox $A$ such that $A$ and $A'$ are indistinguishable w.r.t. $\text{ib.cens}$ (obviously, $\text{ib.cens}'(A) = \text{cl}_{\text{CQ}}(A)$ since indistinguishability w.r.t. an IB censor for a DL TBox $T$ and policy $P$ always forms an equivalence relation), and $\text{ib.cens}'(A) = \text{ib.cens}(A)$ for each ABox $A$ such that $A$ and $A'$ are not indistinguishable w.r.t. $\text{ib.cens}$. Observe that, for each pair of ABoxes $A_1$ and $A_2$, we have that $A_1$ and $A_2$ are...
indistinguishable w.r.t. ib\_cens if and only if they are indistinguishable w.r.t. ib\_cens'. Furthermore, since ib\_cens is an IB censor for \(T\) and \(P\), and since \(T \cup P \cup A\) is consistent (and therefore also \(T \cup P \cup \text{ib\_cens}(A)\) is consistent for each ABox \(A\) such that \(A\) and \(A\) are indistinguishable w.r.t. \text{ib\_cens}), it can be easily verified that \text{ib\_cens}'s is an IB censor for \(T\) and \(P\) that more informative than \text{ib\_cens} (in particular, \text{ib\_cens}(A) \subseteq \text{ib\_cens}'(A) = \text{cl}_{\text{CQ}}(A)\) for each ABox \(A\) such that \(A\) and \(A\) are indistinguishable w.r.t. ib\_cens), thus contradicting the assumption that ib\_cens is an optimal IB censor for \(T\) and \(P\), as required.

As for (ii), let \(A\) be an arbitrary ABox. Consider the ABox \(A'\) such that \(A\) and \(A'\) are indistinguishable w.r.t. ib\_cens (i.e., \text{ib\_cens}(A) = \text{ib\_cens}(A')) , and \(T \cup P \cup A'\) is consistent. From (i), we derive that \text{ib\_cens}'(A') = \text{cl}_{\text{CQ}}(A')\), and therefore \text{ib\_cens}(A) = \text{ib\_cens}(A') = \text{cl}_{\text{CQ}}(A')\).

Suppose that \text{ib\_cens} \in \text{OptIBCens}_{T,p}. Consider the function \text{cp\_cens} with \text{cp\_cens}(A) = \text{GA} \cap \text{ib\_cens}(A)\) for each ABox \(A\). In other words, for each ABox \(A\), \text{cp\_cens}(A) returns the set of all and only the ground atoms occurring in \text{ib\_cens}(A)\). From the definition of \text{cp\_cens} and from Claim (i), it is easy to see that \text{cl}_{\text{CQ}}(\text{cp\_cens}(A)) = \text{ib\_cens}(A)\) for each ABox \(A\). We now prove that \text{cp\_cens} \in \text{GA\_OptHPCens}_{T,p}.

Observe that, by the assumption that \text{ib\_cens} is an IB censor for \(T\) and \(P\), we have that \text{ib\_cens}(A) \subseteq \text{cl}_{\text{CQ}}(A)\) (and therefore \text{cp\_cens}(A) \subseteq \text{cl}_{\text{CQ}}(A)\) for each ABox \(A\). Furthermore, for each ABox \(A\), \(T \cup P \cup A'\) is consistent (implying that \(T \cup P \cup \text{ib\_cens}(A) = \text{ib\_cens}(A)\) is consistent), where \(A'\) is the ABox such that \(A\) and \(A'\) are indistinguishable w.r.t. ib\_cens, and therefore, since \text{cp\_cens}(A) = \text{GA} \cap \text{ib\_cens}(A)\) for each ABox \(A\), we derive that \(T \cup P \cup \text{cp\_cens}(A)\) is consistent for each ABox \(A\). Thus, \text{cp\_cens} is a CP censor in \text{GA\_OptHPCens}_{T,p}.

We now prove its optimality by contradiction. Suppose, for the sake of contradiction, that \text{cp\_cens}' is not an optimal CP censor in \text{GA\_OptHPCens}_{T,p}, i.e., there exists an optimal CP censor \text{cp\_cens}' in \text{GA\_OptHPCens}_{T,p} such that \text{cp\_cens}(A) \subseteq \text{cp\_cens}'(A)\) for each ABox \(A\), and there exists an ABox \(A'\) such that \text{cp\_cens}'(A') \subseteq \text{cp\_cens}(A')\) (observe that, by definition, an optimal CP censor in \text{GA\_OptHPCens}_{T,p} always exists). Consider now the function \text{ib\_cens}' with \text{ib\_cens}'(A) = \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A))\) for each ABox \(A\). Since \text{cp\_cens}' \in \text{GA\_OptHPCens}_{T,p}, and \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A)) = \text{ib\_cens}'(A)\) for each ABox \(A\), using the (\(\subseteq\)) shown before, we derive that \text{ib\_cens}' \in \text{OptIBCens}_{T,p}.

Observe that: (i) for each ABox \(A\), we have that \text{cp\_cens}(A) \subseteq \text{cp\_cens}'(A), \text{ib\_cens}(A) = \text{cl}_{\text{CQ}}(\text{cp\_cens}(A))\), and \text{ib\_cens}'(A) = \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A)); (ii) there exists an ABox \(A'\) such that \text{cp\_cens}'(A') \subseteq \text{cp\_cens}(A'), i.e., there is a ground atom \(\psi\) such that \(\psi \in \text{cp\_cens}(A')\) and \(\psi \notin \text{cp\_cens}'(A')\). From (i), however, we easily derive that \text{ib\_cens}(A) \subseteq \text{ib\_cens}'(A)\) for each ABox \(A\). Furthermore, since for the ABox \(A'\) \text{cp\_cens}'(A') \subseteq \text{cp\_cens}(A'), and since by definition \text{cp\_cens}(A) = \text{GA} \cap \text{ib\_cens}(A)\) for each ABox \(A\), we have that \(\text{GA} \cap \text{ib\_cens}'(A') \subseteq \text{cp\_cens}'(A')\). Due to the fact that \text{ib\_cens}(A) = \text{cl}_{\text{CQ}}(\text{cp\_cens}(A))\) for each ABox \(A\), we derive \(\text{GA} \cap \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A')) \subseteq \text{cp\_cens}'(A')\). It is not hard to see that this latter fact implies that \(\text{cl}_{\text{CQ}}(\text{GA} \cap \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A'))) \subseteq \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A'))\). In particular, the ground atom \(\psi\) is such that \(\psi \in \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A'))\) and \(\psi \notin \text{cl}_{\text{CQ}}(\text{GA} \cap \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A'))\).

Thus, since shown in the previous steps the following equalities hold
\[
\text{cl}_{\text{CQ}}(\text{GA} \cap \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A'))) = \text{cl}_{\text{CQ}}(\text{GA} \cap \text{cp\_cens}'(A')) = \text{cl}_{\text{CQ}}(\text{cp\_cens}'(A')) = \text{ib\_cens}'(A'),
\]
and since \(\text{cl}_{\text{CQ}}(\text{cp\_cens}'(A')) = \text{ib\_cens}'(A')\), we derive that \text{ib\_cens}'(A') \subseteq \text{ib\_cens}'(A'). Therefore, \text{ib\_cens}' is an IB censor for \(T\) and \(P\) more informative than \text{ib\_cens}, and this contradicts the assumption that \text{ib\_cens} \in \text{OptIBCens}_{T,p}, as required.

### Algorithm 1: OptGACensor

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

5 **Query answering under optimal IB censors**

In this section we study query answering under IB censors over DL-Lite\(_T\) 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\(_T\) TBox \(T\) and every policy \(P\), corresponds to an optimal CP censor in \text{GA} for \(T\) and \(P\). 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 \(A_T\) of ground atoms entailed by \(T \cup A\). Then, it iteratively picks a ground atom \(\alpha\) from \(A_T\) following the lexicographic order, and adds \(\alpha\) to the ABox \(T_h\) if \(T \cup T_h \cup \{\alpha\} \cup P\) does not violate the policy \(P\). The following theorem establishes the correctness and complexity of the algorithm.
**Theorem 3** Let $T$ be a DL-Lite$_R$ TBox and $P$ be a policy. There exists a censor $cens \in GA$-OptCPCens$_{T,P}$ such that, for each ABox $A$, OptGAConsor($T, P, A$) (i) returns $cens(A)$ and (ii) runs in polynomial time in the size of $A$.

**Proof.** For each ABox $A$, the set $Th$ returned by the algorithm contains only assertions in $cl_G(A)$, that is, it contains only assertions in $GA$ entailed by $T \cup A$. Moreover, step 6 of the algorithm checks that $Th$ is consistent with $T$ and $P$. Hence, according to Definition 3, the algorithm implements a CP censor $cp_cens$ in $GA$ for $T$ and $P$. It is also immediate to verify that $cp_cens$ is optimal. Indeed, suppose, by way of contradiction, that there exists an ABox $A$ and a censor $cp_cens'$ such that $cp_cens(A) \subseteq cp_cens'(A)$ and $cp_cens(A') \subseteq cp_cens'(A')$ for every other ABox $A'$. This means that there exists an assertion $\alpha \in cl_G(A)$ such that $\alpha$ is not in $cp_cens(A)$, but since $\alpha$ is not in $cp_cens(A)$ then $Th \cup cp_cens(A) \cup \{\alpha\} \cup P$ has to be inconsistent (step 6 of the algorithm), and so $Th \cup cp_cens(A) \cup P$ is inconsistent too, which contradicts the fact that $cp_cens'$ is a CP censor.

As for the complexity, note that the algorithm iterates on the set of ABox assertions $cl_G(A)$ by choosing an assertion $\alpha$ and, in each iteration, it checks if $Th \cup Th \cup \{\alpha\} \cup P$ is consistent. Clearly, the algorithm terminates since $cl_G(A)$ is finite. Moreover, the thesis follows from the fact that given a DL-Lite$_R$ TBox, a policy $P$ (i.e. a set of denial assertions), and an ABox $T \cup \{\alpha\}$, checking if $Th \cup Th \cup \{\alpha\} \cup P$ is consistent can be done in $\mathcal{AC}^0$ w.r.t. to the size of $Th \cup \{\alpha\}$ [Lembo et al., 2015], that the set $cl_G(A)$ can be computed in polynomial time w.r.t. $|A|$ and that its size is polynomial w.r.t. to $|A|$ as well.

From Theorem 2 and Theorem 3 it follows that, to establish if a BCQ $q$ is entailed by $T \cup A$ under an optimal IB censor for $T$ and $P$, it is sufficient to verify whether $Th \cup cp_cens(A) \cup \{\alpha\} \cup P$ is consistent. Clearly, the algorithm terminates since $cl_G(A)$ is finite. Moreover, the thesis follows from the fact that given a DL-Lite$_R$ TBox, a policy $P$ (i.e. a set of denial assertions), and an ABox $T \cup \{\alpha\}$, checking if $Th \cup Th \cup \{\alpha\} \cup P$ is consistent can be done in $\mathcal{AC}^0$ w.r.t. to the size of $Th \cup \{\alpha\}$ [Lembo et al., 2015], that the set $cl_G(A)$ can be computed in polynomial time w.r.t. $|A|$ and that its size is polynomial w.r.t. to $|A|$ as well.

**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 $\mathcal{AC}^0$ (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 $T$ be a DL TBox. Let $P$ be a policy, and let $cens$ be an IB censor for $T$ and $P$. We say that:

(i) $cens$ is an approximation of the optimal IB censors (AIB censor) for $T$ and $P$ if, for every $cens' \in OptIBCens_{T,P}$ and every ABox $A$, $cens(A) \subseteq cens'(A)$;

(ii) $cens$ is a quasi-optimal IB censor (QIB censor) for $T$ and $P$ if, for every ABox $A$, $cens(A) \subseteq cens'(A)$ and $cens(A) \subseteq cens'(A')$ for every ABox $A'$.

**Example 4** The IB censor $cens_3$ of Example 1 is a QIB censor for $T$ and $P$ (but $cens_3 \not\in OptIBCens_{T,P}$).

For QIB censors the following notable property hold.

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

**Proof.** First, observe that the “least informative” censor $cens_3$ is an approximation of the optimal IB censors for $T$ and $P$, and for every ABox $A$, $cens(A) \subseteq cens_3(A)$.

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

**Theorem 4** Let $T$ be a DL-Lite$_R$ TBox, $P$ be a policy, $A$ be an ABox, and $q$ be a BCQ. Then, IB-Entailment($T, P, A, q$) is coNP-complete in data complexity.
Lemma 1 Let \( T \) be a DL TBox, let \( P \) be a policy, let \( A \) be an ABox, and let \( q \) be a BCQ. QIB-Entailment \((T, P, A, q)\) is true iff there exists a \( \alpha' \subseteq c_{GA}(A) \) such that:

(i) \( T \cup A' \models q \);

(ii) \( A' \cap S = \emptyset \), for each secret \( S \in \text{secrets}(T, P, A) \).

Proof. \( (\Rightarrow) \). We first show that given an ABox assertion \( \alpha \in c_{GA}(A) \), there exists an optimal IB censor \( c \in \text{OptIBCens}_{T, P} \) such that \( \alpha \notin c_{GA}(A) \) only if there exists a secret \( S \in \text{secrets}(T, P, A) \) such that \( \alpha \notin S \). Suppose, by way of contradiction, that \( \alpha \) does not belong to any secret in \( \text{secrets}(T, P, A) \). This means that \( c_{GA}(A) \cup P \cup \{\alpha\} \) is still consistent and so \( c \) is not optimal, from which the contradiction follows. Now, suppose that there exists an ABox \( A' \subseteq c_{GA}(A) \) such that: (i) \( T \cup A' \models q \); and (ii) there is no secret \( S \) in \( \text{secrets}(T, P, A) \) such that \( A' \cap S \neq \emptyset \). From what shown above and from condition (ii), we have that \( A' \subseteq c_{GA}(A) \) for every cens in \( \text{OptIBCens}_{T, P} \). This means that \( A' \subseteq c_{GA}(A) \). Moreover, since \( T \cup A' \models q \), we have that \( T \cup c_{GA}(A) \models q \), which shows the thesis.

\( (\Leftarrow) \). Suppose that QIB-Entailment \((T, P, A, q)\) is true. This means that \( q \in c_{GA}(A) \). Since \( c_{GA}(A) \) is an IB censor, then there exists an ABox \( A' \) such that \( c_{GA}(A) = c_{GA}(A') \) and \( T \cup P \cup A' \) is consistent (that is, \( A' \) and \( A \) are indistinguishable w.r.t. \( c_{GA}(A) \)). Hence, \( T \cup A' \models q \). Moreover, \( A' \subseteq c_{GA}(A) \) and thus \( A' \subseteq c_{GA}(A) \). So, \( A' \) satisfies condition (i) of the lemma. As for condition (ii) we proceed towards a contradiction. Suppose that there exists an ABox assertion \( \alpha \in A' \) and a secret \( S \in \text{secrets}(T, P, A) \) such that \( \alpha \in S \). From Definition 4, we have that \( A' \subseteq c_{GA}(A) \) for every cens in \( \text{OptIBCens}_{T, P} \), and so, \( \alpha \in c_{GA}(A) \) for every cens in \( \text{OptIBCens}_{T, P} \). Since \( S \setminus \{\alpha\} \) is consistent with \( T \cup P \), we have that \( S \setminus \{\alpha\} \) is not a secret in \( T \cup P \). So it is possible to define an optimal IB censor whose theory contains \( S \setminus \{\alpha\} \), which is a contradiction, and so \( A' \) satisfies condition (ii) too.

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

Theorem 6 Let \( T \) be a DL-Lite_T TBox, let \( P \) be a policy, let \( A \) be an ABox, and let \( q \) be a BCQ. QIB-Entailment \((T, P, A, q)\) is true iff IAR-Entailment \((T \cup P, c_{GA}(A), q)\) is true.

Proof. Since \( T \cup A \) is consistent, then the secrets in \( T \cup P \cup c_{GA}(A) \) coincide with the minimal subsets of \( c_{GA}(A) \) that are inconsistent with \( T \cup P \). Therefore, the IAR-Repair \( R \) of \((T \cup P, c_{GA}(A))\) is the set of ground atoms from \( c_{GA}(A) \) that do not belong to any secret in \( T \cup P \cup c_{GA}(A) \). 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 \( c_{GA}(A) \). We recall that query entailment under IAR-semantics in a DL-Lite_T is FO-rewritable, if for every TBox \( T \) expressed in \( L \) and every BCQ \( q \), one can effectively compute an FO query \( q' \) such that for every ABox \( A \), IAR-Entailment \((T, A, q)\) is true iff \( A \models q' \). The query \( q' \) is called the IAR-perfect reformulation of \( q \) w.r.t. \( T \).

To establish FO-rewritability of QIB-entailment in DL-Lite_T, however, we still need to address the above mentioned computation of \( c_{GA}(A) \), and turn it into an additional query reformulation step. To this aim, we can exploit the
fact that, for a $DL-Lite_{R,\text{den}}$ ontology $\mathcal{T} \cup \mathcal{A}$, an FO query $q$ evaluates to true over $cl_{\mathcal{GA}}(\mathcal{A})$ iff $\phi'$ evaluates to true over $\mathcal{A}$, where $\phi'$ 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 $\phi'$, we use the function $\text{atomRewr}(q, \mathcal{T})$, which substitutes each atom $\alpha$ of $q$ with the formula $\phi(\alpha)$ defined as follows (where $A, B$ are atomic concepts and $R, S$ are atomic roles):

$$
\phi(A(t)) = \bigvee_{\mathcal{T} \models \alpha \subseteq A} B(t) \lor \bigvee_{\mathcal{T} \models \exists R \subseteq A} (\exists x. R(x, t))
$$

$$
\phi(R(t_1, t_2)) = \bigvee_{\mathcal{T} \models S \subseteq R} S(t_1, t_2) \lor \bigvee_{\mathcal{T} \models S \subseteq R} S(t_2, t_1)
$$

For example, if $\mathcal{T} = \{ A \subseteq C, B \subseteq C \}$ and $q = \exists x, y. C(x) \land P(x, y)$, then $\text{atomRewr}(q, \mathcal{T})$ returns the query $q' = \exists x, y. C(x) \lor A(x) \land P(x, y)$.

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

**Lemma 2** Let $\mathcal{T}$ be a $DL-Lite_{R,\text{den}}$ TBox, let $\mathcal{A}$ be an ABox, and let $q$ be an FO sentence. Then $\text{Eval}(q, cl_{\mathcal{GA}}(\mathcal{A})) = \text{Eval}(\text{atomRewr}(q, \mathcal{T}), \mathcal{A})$.

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

**Theorem 7** Let $\mathcal{T}$ be a $DL-Lite_{R}$ TBox, let $\mathcal{P}$ be a policy, let $q$ be a BCQ, and let $q_r$ be an FO sentence that is a IAR-perfect reformulation of $q$ w.r.t. the $DL-Lite_{R,\text{den}}$ TBox $\mathcal{T} \cup \mathcal{P}$. Then, the FO sentence $\text{atomRewr}(q_r, \mathcal{T})$ 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 $DL-Lite_{R,\text{den}}$ TBox $\mathcal{T} \cup \mathcal{P}$. Then, for every ABox $\mathcal{A}$, $\text{IAR-Entailment}(\mathcal{T} \cup \mathcal{P}, cl_{\mathcal{GA}}(\mathcal{A}), q_r)$ is true iff $\text{Eval}(q_r, cl_{\mathcal{GA}}(\mathcal{A}))$ is true. Now, from Lemma 2, it follows that, for every ABox $\mathcal{A}$, $\text{Eval}(q_r, cl_{\mathcal{GA}}(\mathcal{A})) = \text{Eval}(\text{atomRewr}(q_r, \mathcal{T}), \mathcal{A})$. And since by Theorem 6, for every ABox $\mathcal{A}$ such that $\mathcal{T} \cup \mathcal{A}$ is consistent, $\text{IAR-Entailment}(\mathcal{T} \cup \mathcal{P}, cl_{\mathcal{GA}}(\mathcal{A}), q_r)$ is true iff $\text{QIB-Entailment}(\mathcal{T} \cup \mathcal{P}, \mathcal{A}, q)$ is true, it follows that the FO sentence $\text{atomRewr}(q_r, \mathcal{T})$ is a QIB-perfect reformulation of $q$ w.r.t. $\mathcal{T}$ and $\mathcal{P}$. \hfill \blacksquare

Since IAR-entailment is actually FO rewritable, as shown in [Lemko et al., 2015], the above theorem proves the FO rewritability of QIB-entailment for $DL-Lite_{R}$ 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 [Lemko et al., 2015] and the $\text{atomRewr}$ reformulation defined above. Therefore:

**Corollary 2** Let $\mathcal{T}$ be a $DL-Lite_{R}$ TBox, let $\mathcal{P}$ be a policy, let $\mathcal{A}$ be an ABox, and let $q$ be a BCQ. The problem $\text{QIB-Entailment}(\mathcal{T}, \mathcal{P}, \mathcal{A}, q)$ is in $\text{AC}^0$ in data complexity.

7 Conclusions

In this paper we have studied the approach to controlled query evaluation based on instance indistinguishability: we have applied this approach to Description Logic ontologies, we have studied its relationship with another confidentiality-preserving approach, and we have established complexity results for this form of controlled query evaluation in the case of $DL-Lite_{R}$ ontologies.

Notably, in this framework we have identified a tractable and semantically well-founded notion of CQE that enjoys the first-order rewritability property. 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 EU within the H2020 under grant agreement 834228 (ERC Advanced Grant WhiteMec) and under grant agreement 825333 (MOSAIcOWN), by Regione Lombardia within the Call Hub Ricerca e Innovazione under 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 secrets 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.

[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 Kharchamov, 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 Kharchamov, 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.