Controlled Query Evaluation over Lightweight Ontologies*

B. Cuenca Grau, E. Kharlamov, E. V. Kostylev, and D. Zheleznyakov

Department of Computer Science, University of Oxford

Abstract. We study confidentiality enforcement in ontologies under the Controlled Query Evaluation (CQE) framework. In a CQE system, a *policy* specifies the sensitive information, and a *censor* ensures that answers to user's queries that could violate the policy are not returned. Our goal is the design of *optimal* CQE algorithms, which ensure confidentiality while maximising access to information. We study two natural classes of censors that can be realised using existing infrastructure for query answering and propose optimal CQE algorithms for the standardised profiles of the ontology language OWL 2.

1 Introduction

As ontology-based information systems are becoming increasingly mature, there is a pressing need to devise mechanisms for ensuring that data is only made accessible to authorised users [1, 7–12, 20, 21].

Controlled Query Evaluation (CQE) is a prominent formal framework for confidentiality enforcement. In CQE, sensitive information is declaratively specified by means of a *policy* and confidentiality is enforced by a *censor*: when given a user query, a censor checks whether returning the answer might lead to a policy violation, in which case it returns a distorted answer. CQE was introduced in [18], and studied in [3, 4, 6] for propositional databases with complete information. It was extended to (propositional) incomplete databases in [5]. Beyond propositional logic, CQE remains largely unexplored [2].

In this paper we study CQE in the context of ontologies. Our basic framework is described in Section 3. We assume data to be hidden and that users interact with the system by means of a query interface. An ontology, which we assume to be known to all users, provides the vocabulary and background knowledge needed for users to formulate accurate queries, as well as to enrich query answers with implicit information. Policies are given as a set of ground atoms that follow from the ontology and data. When given a (conjunctive) query, the system returns the subset of certain answers determined by the censor; in this way, the role of the censor is to preserve confidentiality by filtering out those answers that could lead to a violation of the policy. Formally, we model the information that users could gather by querying the system as an (infinite) first-order theory;

[*] This paper recapitulates and extends our previous work [11]. It comes with a technical report available at http://tinyurl.com/DL14paper55.

confidentiality preservation then amounts to ensuring that such theory together with the ontology does not entail any atom in the policy. In this setting, there is a danger that confidentiality enforcement may over-restrict users' access. Thus, we focus on *optimal* censors, which maximise answers to queries while ensuring confidentiality of the policy. Furthermore, we are interested in censors that can be implemented by reusing off-the-shelf query answering infrastructure. To fulfil this requirement, we introduce in Section 4 two classes of censors, which we call *view* and *obstruction* censors, respectively.

View censors return only those answers that follow from the ontology and a materialised dataset (a *view*). In this way, a view encodes the information that users are authorised to access: the censor answers faithfully all queries against the view, and any information not captured by the view is inaccessible by default. View censors require the ability to materialise implicit data, and hence are especially well-suited for RDF-based applications in which reasoning is performed by a triple store. Obstruction censors are dual to view censors in the sense that they explicitly specify information which users are denied access to (with all other information being accessible by default). Obstruction censors are specified by a finite set of "forbidden query patterns" (*obstructions*), and all query answers that instantiate such patterns are filtered out. In contrast to view censors, obstruction censors do not require modification of the data and hence are well-suited for OBDA applications, where data is typically managed by an RDBMS. We finally characterise the duality of views and obstructions and argue that it is not always possible to "simulate" one with another.

In Section 5 we focus on view censors. First, we investigate their intrinsic limitations, and then show how these limitations can be circumvented. We propose algorithms for computing optimal view censors for knowledge bases with OWL 2 RL, EL and QL ontologies under relatively minor restrictions. Our algorithms, however, rely on views that can be of exponential size in the worst case. So, we identify natural conditions on ontologies that guarantee polynomial size of optimal views. In particular, all OWL 2 QL ontologies satisfy these conditions.

In Section 6 we turn our attention to obstruction censors and provide sufficient and necessary conditions for an optimal such censor to exist. Then, we propose algorithms for computing optimal obstruction censors for knowledge bases with OWL 2 QL as well as restricted OWL 2 RL ontologies, which are based on obstructions of polynomial size.

2 Preliminaries

We adopt standard notions in first-order logic over finite function-free signatures. We treat equality $\approx$ as an ordinary predicate, but assume that any set of formulae Σ contains all the axioms of equality for Σ. A *fact* is a ground, equality-free atom, and a *dataset* is a finite set of facts. A *structure* $\mathcal{I}$ is a pair $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ of a domain and interpretation function for the symbols in the signature. We define *homomorphisms* between structures $\mathcal{I}$ and $\mathcal{J}$ in the standard way and $\mathcal{I} \hookrightarrow \mathcal{J}$ denotes the fact that such a homomorphism from $\mathcal{I}$ to $\mathcal{J}$ exists.

(*1*) $A(x) \wedge R(x, y_1) \wedge B(y_1) \wedge R(x, y_2) \wedge B(y_2) \rightarrow y_1 \approx y_2$,	
(*2*) $R(x, y) \rightarrow S(x, y)$,	(*3*) $A(x) \rightarrow \exists y.[R(x, y) \wedge B(y)]$,
(*4*) $A(x) \rightarrow x \approx a$,	(*5*) $R(x, y) \wedge S(y, z) \rightarrow T(x, z)$,
(*6*) $A(x) \rightarrow R(x, x)$,	(*7*) $A(x) \wedge B(x) \rightarrow C(x)$,
(*8*) $R(x, x) \rightarrow A(x)$,	(*9*) $A(x) \wedge R(x, y) \rightarrow B(y)$,
(*10*) $R(x, y) \rightarrow S(y, x)$,	(*11*) $R(x, a) \rightarrow B(x)$,
(*12*) $R(x, y) \rightarrow A(y)$,	(*13*) $A(x) \rightarrow R(x, a)$,
(*14*) $A(x) \rightarrow B(x)$,	(*15*) $R(x, y) \wedge B(y) \rightarrow A(x)$.

Table 1. Horn-$\mathcal{SROIF}$ rules; unary predicates can be $\top$.

A *rule* is a sentence of the form $\forall \boldsymbol{x}.\forall \boldsymbol{z}.[\varphi(\boldsymbol{x}, \boldsymbol{z}) \rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y})]$, where $\boldsymbol{x}$, $\boldsymbol{z}$, and $\boldsymbol{y}$ are pairwise disjoint vectors of variables, the *body* $\varphi(\boldsymbol{x}, \boldsymbol{z})$ is an equality-free conjunction of atoms with variables $\boldsymbol{x} \cup \boldsymbol{z}$, and the *head* $\psi(\boldsymbol{x}, \boldsymbol{y})$ is a conjunction of atoms with variables $\boldsymbol{x} \cup \boldsymbol{y}$. For simplicity universal quantifiers in rules are usually omitted. A rule is *(i) Datalog* if its head consists of a single atom and $\boldsymbol{y}$ is empty; *(ii) guarded* if it has a body atom (a *guard*) mentioning all universally quantified variables; *(iii) linear* if it has a single body atom; and *(iv) multi-linear* if the body contains only guards. An *ontology* is a finite set of rules. We assume that both rule heads and bodies are non-empty and they do not contain the nullary atoms $\top$ and $\bot$. Thus, $\mathcal{O} \cup \mathcal{D}$ is satisfiable for each ontology $\mathcal{O}$ and dataset $\mathcal{D}$, and $\mathcal{O} \not\models \alpha$ for each fact α, which ensures a separation between schema and data.

To capture all OWL 2 profiles, we focus on *Horn-$\mathcal{SROIF}$*. Table 1 provides the normalised axioms of this DL in the form of rules. To capture the semantics of $\top$, usually allowed in DLs, we treat it as a unary predicate and assume that each ontology $\mathcal{O}$ contains the rule $P(x_1, \ldots, x_n) \rightarrow \top(x_i)$ for each predicate P and $1 \leq i \leq n$. A Horn-$\mathcal{SROIF}$ ontology is in *(i) RL* if it has *no* rules of Type (3); *(ii) QL* if it contains *only* rules of Types (2), (3), (10), (12), and (14); *(iii) EL* if it *does not* contain rules of Types (1), (9), and (10).

A *conjunctive query* (*CQ*) is a formula $Q(\boldsymbol{x})$ of the form $\exists \boldsymbol{y}.\varphi(\boldsymbol{x}, \boldsymbol{y})$, with $\varphi(\boldsymbol{x}, \boldsymbol{y})$ a conjunction of atoms. A *union of CQs* (*UCQ*) is a formula $\bigvee_i Q_i(\boldsymbol{x})$, with each $Q_i(\boldsymbol{x})$ a CQ. A CQ is *Boolean* (*BCQ*) if $\boldsymbol{x}$ is empty. A tuple of constants $\boldsymbol{t}$ is a (*certain*) *answer* to $Q(\boldsymbol{x})$ over ontology $\mathcal{O}$ and dataset $\mathcal{D}$ if $\mathcal{O} \cup \mathcal{D} \models Q(\boldsymbol{t})$. Then, $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ is the set of answers to $Q(\boldsymbol{x})$ over $\mathcal{O}$ and $\mathcal{D}$. Given a BCQ Q, $\mathbf{A}[Q]$ denotes the structure interpreting each relation R with $(f(u_1), \ldots f(u_n))$ for every atom $R(u_1, \ldots, u_n)$ in Q, where f maps each constant in Q to itself and each variable y to a fresh element d_y in the structure.

3 Basic Framework

Given an ontology $\mathcal{O}$ and dataset $\mathcal{D}$, we assume that $\mathcal{D}$ is hidden while $\mathcal{O}$ is known to users, who can formulate arbitrary CQs via a query interface. A policy, which is unknown to users, is given as a set of facts entailed by $\mathcal{O} \cup \mathcal{D}$. It is assumed that system administrators are in charge of specifying policies, and that each policy is assigned to a specific (group of) users by means of standard mechanisms such as role-based access control techniques [17].

Definition 1. *A* policy $\mathcal{P}$ *for* $\mathcal{O}$ *and* $\mathcal{D}$ *is a dataset such that* $\mathcal{O} \cup \mathcal{D} \models \mathcal{P}$. *A* CQE-instance $\mathbf{I}$ *is a triple* $(\mathcal{O}, \mathcal{D}, \mathcal{P})$*, with* $\mathcal{P}$ *a policy for* $\mathcal{O}$ *and* $\mathcal{D}$*.*

Example 1. Consider the following ontology and dataset that describe an excerpt of a social network including information about movies:

$$\begin{aligned}\mathcal{O}_{\mathsf{ex}} = \{&\mathit{FOf}(x,y) \rightarrow \mathit{FOf}(y,x), \mathit{Susp}(x) \wedge \mathit{Cr}(x) \rightarrow \mathit{Thr}(x),\\ &\mathit{Likes}(x,y) \wedge \mathit{Thr}(y) \rightarrow \mathit{ThrFan}(x)\},\\ \mathcal{D}_{\mathsf{ex}} = \{&\mathit{FOf}(\mathit{John}, \mathit{Bob}),\ \mathit{FOf}(\mathit{Bob}, \mathit{Mary}), \mathit{Likes}(\mathit{John}, \mathit{Seven}),\\ &\mathit{Likes}(\mathit{Bob}, \mathit{Seven}),\ \mathit{Susp}(\mathit{Seven}),\ \mathit{Cr}(\mathit{Seven})\}.\end{aligned}$$

Here, $\mathcal{O}_{\mathsf{ex}}$ states that friendship is symmetric; movies that are both suspense and crime are thrillers; and everyone who likes a thriller is a thriller fan. Assume that John wants to hide his friend list. Then, $\mathcal{P}_{\mathsf{ex}} = \{\alpha_{\mathsf{ex}}\}$ with $\alpha_{\mathsf{ex}} = \mathit{FOf}(\mathit{John}, \mathit{Bob})$, and $\mathbf{I}_{\mathsf{ex}} = (\mathcal{O}_{\mathsf{ex}}, \mathcal{D}_{\mathsf{ex}}, \mathcal{P}_{\mathsf{ex}})$.

A key component of a CQE system is the *censor*, whose goal is to decide according to the policy which query answers can be safely returned to users.

Definition 2. *A* censor *for a CQE-instance* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *is a function* cens *that maps each CQ* Q *to a subset of* $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$*. The* characteristic theory $\mathsf{Th}_{\mathsf{cens}}$ *of* cens *is the (possibly infinite) set of sentences*

$$\{Q(\boldsymbol{t}) \mid \boldsymbol{t} \in \mathsf{cens}(Q) \text{ and } Q(\boldsymbol{x}) \text{ is a CQ}\}.$$

Censor cens *is* confidentiality preserving *if* $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \not\models \alpha$ *for each* $\alpha \in \mathcal{P}$*. It is* optimal *if* (i) *it is confidentiality preserving and* (ii) *no confidentiality preserving censor* $\mathsf{cens}' \neq \mathsf{cens}$ *exists such that* $\mathsf{cens}(Q) \subseteq \mathsf{cens}'(Q)$ *for every* Q*.*

Intuitively, $\mathsf{Th}_{\mathsf{cens}}$ represents the information that a user can potentially gather by asking an unbounded number of queries to the system. If the censor is confidentiality preserving, then no information can be obtained about $\mathcal{P}$, regardless of the CQs asked. Finally, optimal censors maximise information accessibility without compromisig the policy.

4 View and Obstruction Censors

The idea behind view censors is to associate to a CQE instance $\mathbf{I}$ a new dataset, called a *view*. Intuitively, a view encodes the information that a user is allowed to see. The user gets only those query answers that follow from $\mathcal{O}$ and this view. In this way, the main workload of the censor boils down to the computation of certain answers, which can be fully delegated to the query answering engine.[1]

[1] We assume that all the definitions in this section are parameterised by a (fixed) instance $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$.

Definition 3. *The* view censor $\mathsf{vcens}^{\mathcal{V}}_{\mathbf{I}}$ *for* $\mathbf{I}$ *based on a dataset* $\mathcal{V}$ *(a* view*), is the function mapping each CQ* $Q(\boldsymbol{x})$ *to the set* $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) \cap \mathsf{cert}(Q, \mathcal{O}, \mathcal{V})$.

Obviously, if we want the censor to enjoy the properties we are after, the view $\mathcal{V}$ must be constructed with care. For the censor to be confidentiality preserving, $\mathcal{O} \cup \mathcal{V}$ must not entail any atom from the policy $\mathcal{P}$, and to be optimal $\mathcal{V}$ must encode as much information from the hidden dataset as possible.

Example 2. Consider a view $\mathcal{V}_{\mathsf{ex}}$ obtained from the dataset $\mathcal{D}_{\mathsf{ex}}$ by replacing *Bob* with a fresh constant an_b. Intuitively, $\mathcal{V}_{\mathsf{ex}}$ is the result of "anonymising" the constant *Bob*, while keeping the structure of the data intact. Since $\mathcal{V}_{\mathsf{ex}}$ contains no information about *Bob*, we have $\mathcal{O}_{\mathsf{ex}} \cup \mathcal{V}_{\mathsf{ex}} \not\models \alpha_{\mathsf{ex}}$, that is the censor based on $\mathcal{V}_{\mathsf{ex}}$ is confidentiality preserving. View $\mathcal{V}_{\mathsf{ex}}$, however, is not optimal: $\mathcal{O}_{\mathsf{ex}} \cup \mathcal{V}_{\mathsf{ex}}$ does not entail the fact $\mathit{Likes}(\mathit{Bob}, \mathit{Seven})$, which is harmless for confidentiality. Indeed, $\mathcal{O}_{\mathsf{ex}} \cup \mathcal{V}'_{\mathsf{ex}} \not\models \alpha_{\mathsf{ex}}$ holds for the extension $\mathcal{V}'_{\mathsf{ex}}$ of $\mathcal{V}_{\mathsf{ex}}$ with $\mathit{Likes}(\mathit{Bob}, \mathit{Seven})$.

View censors require the ability to materialise implicit data, and hence are especially well-suited for RDF-based applications, in which reasoning is performed by a triple store. In OBDA scenarios, however, data is typically managed by an RDBMS and materialisation is not possible. To fulfill the requirement of OBDA applications, we need a different kind of censors.

The idea behind obstruction censors is to associate to $\mathbf{I}$ an *obstruction* in the form of a Boolean UCQ U, such that given a query $Q(\boldsymbol{x})$ and an answer $\boldsymbol{t}$ over $\mathcal{O}$ and $\mathcal{D}$, the censor returns $\boldsymbol{t}$ only if no CQ in U follows from $Q(\boldsymbol{t})$. Thus, the obstruction can be seen as a collection of "forbidden query patterns", which should not be disclosed.

Definition 4. *The* obstruction censor $\mathsf{ocens}^{U}_{\mathbf{I}}$ *for* $\mathbf{I}$ *based on a Boolean UCQ* U *(called an* obstruction*) is the function mapping each CQ* $Q(\boldsymbol{x})$ *to the set*

$$\{\boldsymbol{t} \mid \boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) \textit{ and } \mathbf{A}[Q(\boldsymbol{t})] \not\models U\}.$$

Similarly to view censors, obstruction censors do not require dedicated algorithms: checking $\mathbf{A}[Q(\boldsymbol{t})] \models U$ can be delegated to an RDBMS. Also, obstructions can be maintained virtually without the need of data materialisation.

Example 3. The censor based on $\mathcal{V}_{\mathsf{ex}}$ from Example 2 can also be realised with following obstruction U_{ex}:

$$\exists x.\mathit{FOf}(x, \mathit{Bob}) \lor \exists x.\mathit{FOf}(\mathit{Bob}, x) \lor \exists x.\mathit{Likes}(\mathit{Bob}, x) \lor \mathit{ThrFan}(\mathit{Bob}).$$

Intuitively, U_{ex} "blocks" query answers involving *Bob*; and all other answers are the same as over $\mathcal{O}_{\mathsf{ex}} \cup \mathcal{D}_{\mathsf{ex}}$.

As seen in Examples 2 and 3, the same censor may be based on both a view and an obstruction. View and obstruction censors, however, behave rather differently: a view explicitly encodes the information accessible to users, whereas obstructions specify information which users are denied access to. Thus, obstructions are *dual* to views. Unsurprisingly, even in simple cases it is not obvious whether (and how) a view can be realised by an obstruction, or vice-versa.

We next focus on Datalog ontologies and characterise when a given view $\mathcal{V}$ and obstruction U yield the same censor. Each Datalog ontology $\mathcal{O}$ and dataset $\mathcal{D}$ have a unique *least Herbrand model* $\mathcal{H}_{\mathcal{O},\mathcal{D}}$, that is a finite structure that satisfies $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ iff $\mathbf{A}[Q(\boldsymbol{t})] \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$ and hence captures the information relevant to query answering. We can then formalise the duality between views and obstructions in a natural way: U and $\mathcal{V}$ implement the same censor iff U captures the structures *not* homomorphically embeddable into $\mathcal{H}_{\mathcal{O},\mathcal{V}}$. To formalise this statement, we recall the notion of (non-uniform) constraint satisfaction [13].

Definition 5. *Let $\mathcal{J}$ be a finite structure and $\mathcal{C}$ a class of finite structures. The* CSP *of $\mathcal{J}$ relative to $\mathcal{C}$ (denoted $\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{J})$) is the set $\{\mathcal{I} \in \mathcal{C} \mid \mathcal{I} \hookrightarrow \mathcal{J}\}$.*

A central problem is to determine whether a class of finite structures can be captured by a single formula.

Definition 6. *Let $\mathcal{C}$ be a class of finite structures and let $\mathcal{C}' \subseteq \mathcal{C}$. First-order sentence ψ* defines *$\mathcal{C}'$ if $\mathcal{I} \in \mathcal{C}'$ is equivalent to $\mathcal{I} \models \psi$ for every structure $\mathcal{I} \in \mathcal{C}$.*

The correspondence between view and obstruction censors is then as follows.

Theorem 1. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ Datalog ontology, and $\mathcal{C}$ the class of finite structures $\mathcal{I}$ with $\mathcal{I} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$. Then, $\mathsf{vcens}^{\mathcal{V}}_{\mathbf{I}} = \mathsf{ocens}^{U}_{\mathbf{I}}$ iff U defines $\neg\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$, for any view $\mathcal{V}$ and obstruction U.*

Using Theorem 1 together with definability results in Finite Model Theory, we can show that views and obstructions cannot simulate one another in general.

Theorem 2. *There is a CQE-instance for which there exists a view censor, but no obstruction censor. There is a CQE-instance for which there exists an obstruction censor, but no view censor.*

5 Optimal View Censors

Our discussion in Section 4 shows that view and obstruction censors should be studied independently. In this section, we focus on view censors.

Before investigating the design of view-based CQE algorithms, we first establish the theoretical limitations of our approach. We show that an optimal view-based censor for an instance $\mathbf{I}$ is not guaranteed to exist since the optimality requirement may lead to infinite "views", even for EL and RL ontologies.

Theorem 3. *There are CQE-instances $\mathbf{I}_1$ and $\mathbf{I}_2$ such that*

- *the ontology of $\mathbf{I}_1$ uses rules of Types* (1) *and* (9)*, and*
- *the ontology of $\mathbf{I}_2$ uses rules of Types* (5), (8) *and* (15)*,*

for which no optimal view censors exist.

The construction of $\mathbf{I}_1$ shows that equality rules lead to non-existence of optimal views; in turn, the construction of $\mathbf{I}_2$ shows that equality is not needed to preclude optimality in the presence of recursion, transitivity axioms, and Self restrictions.

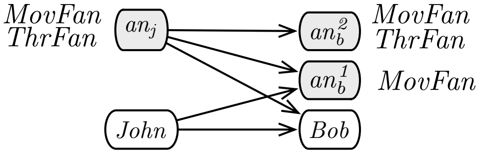

Fig. 1. Essential part of an exhaustive view (solid arrows represent *FOf* relation).

5.1 View Censors for Guarded RL

We next describe how to compute an optimal view for guarded RL ontologies. The idea is to create anonymised copies of constants in the data to encode the information required for optimality. Such a view may use exponentially many anonymised copies of constants.

Definition 7. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a guarded RL ontology. A view $\mathcal{V}$ consisting of unary atoms $\mathcal{V}_c^1$ on constants from $\mathbf{I}$, unary atoms $\mathcal{V}_\exists^1$ on new constants (anonymised copies), and binary atoms $\mathcal{V}^2$, is* exhaustive *on $\mathbf{I}$ if it satisfies all of the following.*

1. *The part $\mathcal{V}_c^1$ is a maximal set of unary atoms from $\mathcal{H}_{\mathcal{O},\mathcal{D}}$ such that $\mathcal{O} \cup \mathcal{V}_c^1 \not\models \alpha$ for each $\alpha \in \mathcal{P}$.*
2. *For each constant a from $\mathbf{I}$, and each set $\mathcal{A}$ of unary predicates, such that*
   - *$A(a) \in \mathcal{H}_{\mathcal{O},\mathcal{D}}$ for every $A \in \mathcal{A}$,*
   - *if $A, B \in \mathcal{A}$, and $\mathcal{O} \models A(x) \wedge B(x) \to C(x)$, then $C \in \mathcal{A}$,*
   - *if $A \in \mathcal{A}$, then $\mathcal{O} \not\models A(x) \to x \approx a$ for each a,*
   - *$\mathcal{O} \cup \mathcal{V}_c^1 \cup \{A(a') | A \in \mathcal{A}\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$ and a fresh constant a',*

   the view $\mathcal{V}$ uses a fresh constant $a_\mathcal{A}$, and the part $\mathcal{V}_\exists^1$ contains all unary atoms $A(a_\mathcal{A})$ such that $A \in \mathcal{A}$.
3. *For each constant a from $\mathbf{I}$, let σ_a be the set of constants consisting of a itself and all the constants $a_\mathcal{A}$. The binary part $\mathcal{V}^2$ of the view $\mathcal{V}$ contains the atom $R(a_1^*, a_2^*)$ for constants a_1^*, a_2^* iff*
   - *$R(a_1, a_2) \in \mathcal{H}_{\mathcal{O},\mathcal{D}}$, where $a_1^* \in \sigma_{a_1}$ and $a_2^* \in \sigma_{a_2}$,*
   - *$\mathcal{O} \cup \{R(a_1^*, a_2^*)\} \not\models \alpha$ for any $\alpha \in \mathcal{P}$, and*
   - *$\mathcal{O} \cup \mathcal{V}_c^1 \cup \mathcal{V}_\exists^1 \cup \{R(a_1^*, a_2^*)\} \models A(a^*)$ implies that $A(a^*) \in \mathcal{V}_c^1 \cup \mathcal{V}_\exists^1$.*

This definition is constructive and it is routine to devise an algorithm, which for any instance (non-deterministically) constructs an exhaustive view.

Example 4. Consider the following CQE-instance $(\mathcal{O}, \mathcal{D}, \mathcal{P})$:

$$\begin{aligned}
\mathcal{O} &= \{\mathit{ThrFan}(x) \to \mathit{MovieFan}(x), \mathit{ThrFan}(y) \wedge \mathit{FOf}(x, y) \to \mathit{MovieFan}(x)\},\\
\mathcal{D} &= \{\mathit{FOf}(\mathit{John}, \mathit{Bob}), \mathit{ThrFan}(\mathit{John}), \mathit{ThrFan}(\mathit{Bob})\},\\
\mathcal{P} &= \{\mathit{MovieFan}(\mathit{Bob}), \mathit{MovieFan}(\mathit{John})\}.
\end{aligned}$$

The essential part of the exhaustive view on this CQE-instance is given in Figure 1, where $\mathcal{V}_c^1 = \emptyset$, $\mathcal{V}_\exists^1$ contains unary atoms over the anonymised copies an_b^1, an_b^2 of Bob, and an_j of John, and $\mathcal{V}^2$ contains the depicted binary atoms. Two anonymised copies of *Bob* are necessary in any optimal view for $\mathbf{I}$ to answer correctly "harmless" queries like

$$\exists x, y, z. \mathit{ThrFan}(z) \wedge \mathit{MovieFan}(z) \wedge \mathit{FOf}(y, z) \wedge \mathit{ThrFan}(y) \wedge \\ \mathit{MovieFan}(y) \wedge \mathit{FOf}(y, x) \wedge \mathit{MovieFan}(x) \wedge \mathit{FOf}(\mathit{John}, x).$$

The following theorem formulates the desired properties of exhaustive views.

Theorem 4. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a guarded RL ontology, and $\mathcal{V}$ an exhaustive view on $\mathbf{I}$. Then $\mathcal{V}$ is optimal. Furthermore, if $\mathcal{O}$ is linear, then $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ is the only optimal censor for $\mathbf{I}$.*

The proof relies on the following facts. First, the construction ensures that $\mathcal{H}_{\mathcal{O},\mathcal{V}} = \mathcal{V}$ for any exhaustive view $\mathcal{V}$, that is, no rules are applicable to $\mathcal{V}$. Also, properties of $\mathcal{V}_c^1$, $\mathcal{V}_\exists^1$, and $\mathcal{V}^2$ guarantee that $\mathcal{V}$ does not entail any policy atom. Optimality follows from the fact that for any a in $\mathbf{I}$, each combination of its unary atoms that satisfies the relevant axioms in $\mathcal{O}$ is "witnessed" by a new constant from σ_a, and all possible binary atoms which are compatible with those combinations are added to the view. Then, no essentially new atom can be "added" to the view $\mathcal{V}$ without disclosing a policy. The uniqueness of the optimal censor for linear ontologies follows from the lack of choices in the construction of $\mathcal{V}$. An exhaustive view may use exponentially many constants. However, for multi-linear ontologies, optimal views are of polynomial size.

Proposition 1. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a multi-linear RL ontology. There is an optimal censor for $\mathbf{I}$ based on a view of polynomial size.*

5.2 View Censors for EL and QL

In contrast to OWL 2 RL, the QL and EL profiles can capture existentially quantified knowledge. To bridge this gap, we show that, under some mild conditions, we can transform an ontology $\mathcal{O}$ into a Datalog ontology $\mathcal{O}'$ such that an optimal view for $(\mathcal{O}, \mathcal{D}, \mathcal{P})$ can be directly obtained from such a view for $(\mathcal{O}', \mathcal{D}, \mathcal{P})$. Thus, devising an optimal view censor for an instance is reduced to devising one for an instance with a Datalog ontology.

Definition 8. *Let σ be a set of constants. A Datalog ontology $\mathcal{O}'$ is a (*Datalog*)* σ-rewriting *of an ontology $\mathcal{O}$ if for each fact β and dataset $\mathcal{D}$ over constants from σ we have that $\mathcal{O} \cup \mathcal{D} \models \beta$ iff $\mathcal{O}' \cup \mathcal{D} \models \beta$.*

Proposition 2. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{D}$ using set of constants σ, $\mathcal{O}'$ a σ-rewriting of $\mathcal{O}$ such that $\mathcal{O}' \models \mathcal{O}$, and $\mathcal{V}'$ an optimal view for $(\mathcal{O}', \mathcal{D}, \mathcal{P})$. Then $\mathcal{H}_{\mathcal{O}',\mathcal{V}'}$ is an optimal view for $\mathbf{I}$.*

Now, we just need to transform a QL (or guarded EL) ontology into a stronger guarded RL ontology, which, however, entails the same facts for any dataset. We exploit techniques developed for the *combined approach* to query answering [14–16, 19]. The idea is to transform rules of Type (3) into Datalog by Skolemising existentially quantified variables into globally fresh constants. Such transformation strengthens the ontology; however, if applied to a QL or guarded EL ontology, it preserves entailment of facts for any dataset over σ [19].

Definition 9. *Let $\mathcal{O}$ be an ontology and σ a set of constants. The ontology $\Xi_\sigma(\mathcal{O})$ is obtained from $\mathcal{O}$ by replacing each rule of the form $A(x) \to \exists y.[R(x,y) \wedge B(y)]$ with $A(x) \to P(x,a), P(x,y) \to R(x,y), P(x,y) \to B(y)$, where P is a fresh predicate and a is a globally fresh constant not from σ, unique to A and R.*[2]

Proposition 3. *If $\mathcal{O}$ is a Horn-$\mathcal{SROIF}$ ontology, then $\Xi_\sigma(\mathcal{O}) \models \mathcal{O}$. If also $\mathcal{O}$ is either a QL or guarded EL ontology, then $\Xi_\sigma(\mathcal{O})$ is a σ-rewriting of $\mathcal{O}$.*

Propositions 2 and 3 ensure that $\mathcal{H}_{\Xi_\sigma(\mathcal{O}),\mathcal{V}}$ is optimal for $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ with $\mathcal{O}$ a QL or guarded EL ontology, whenever $\mathcal{V}$ is such a view for $(\Xi_\sigma(\mathcal{O}), \mathcal{D}, \mathcal{P})$. The transformation of $\mathcal{O}$ to $\Xi_\sigma(\mathcal{O})$ preserves linearity and guardedness, so $\Xi_\sigma(\mathcal{O})$ is a guarded RL ontology, and the results of Section 5.1 are applicable.

Theorem 5. *A CQE-instance with a QL or guarded EL ontology has an optimal view censor. For QL, it is unique and can be based on a polynomial size view.*

6 Obstruction Censors

We start our study of obstruction censors by focusing on Datalog ontologies and characterising optimality in terms of resolution proofs of the policy. To this end, we first recapitulate the standard notions on (clause) SLD resolution.

Definition 10. *A* goal *is a conjunction of atoms. The* SLD resolution step *takes a goal $\beta_1 \wedge \ldots \wedge \beta_m$ and a Datalog rule $\bigwedge_{i=0}^{k} \gamma_i \to \delta$ and produces a new goal $(\bigwedge_{i=0}^{k} \gamma_i\theta) \wedge \beta_2\theta \wedge \ldots \wedge \beta_m\theta$, where θ is a most general unifier (MGU) of β_1 and δ. A* proof *of a goal G_0 in a Datalog ontology $\mathcal{O}$ and dataset $\mathcal{D}$ is a sequence $G_0 \xrightarrow{r_1,\theta_1} G_1 \xrightarrow{r_2,\theta_2} \ldots \xrightarrow{r_n,\theta_n} G_n$, where $G_n = \top$ and the goal G_i is obtained from the goal G_{i-1} and sentence $r_i \in \mathcal{O} \cup \mathcal{D}$ by an SLD resolution step with MGU θ_i.*

SLD resolution is sound and complete: for each satisfiable $\mathcal{O} \cup \mathcal{D}$ and goal G, a proof of G exists in $\mathcal{O} \cup \mathcal{D}$ iff $\mathcal{O} \cup \mathcal{D} \models \exists^* G$, with $\exists^* G$ the existential closure of G. We next provide a characterisation of optimality based on proofs. Consider a policy atom $\alpha \in \mathcal{P}$ and some proof π of α in $\mathcal{O} \cup \mathcal{D}$. If a censor answers positively sufficiently many BCQs $\exists^* G$ for goals G in π, then a user could "reconstruct" (a part of) π and compromise the policy. Also, there can be many proofs of α, and a user can compromise the policy by reconstructing any of them. Thus, to ensure that a censor is confidentiality preserving, we must guarantee that the obstruction contains enough CQs to prevent reconstruction of any π. If we want the censor to be optimal, the obstruction should not "block" too many queries. As we will see later on, these requirements may be in conflict and lead to an infinite "obstruction". To formalise this intuition we need an auxiliary notion. A *core* of a set of Boolean CQs $\mathbb{Q}$ is a minimal subset $\mathbb{C}$ of $\mathbb{Q}$ such that for each $Q \in \mathbb{Q}$ there exists $Q' \in \mathbb{C}$ with $Q \models Q'$.

[2] To correctly deal with Self restrictions (rules (6) and (8)) a slightly more complex transformation is required. These changes are straightforward but require introducing further notation, so we present here only the basic transformation for simplicity.

Definition 11. *Given a CQE-instance* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *with* $\mathcal{O}$ *a Datalog ontology, let* $\mathbb{Q}(\mathbf{I})$ *be the set of all Boolean CQs* $\exists^* G$ *with* $G \neq \top$ *a goal in a proof of a fact* $\alpha \in \mathcal{P}$ *in* $\mathcal{O} \cup \mathcal{D}$*, and let* $\mathbb{S}$ *be a maximal subset of* $\mathbb{Q}(\mathbf{I})$ *such that* $\mathcal{O} \cup \mathbb{S} \not\models \alpha$ *for any* $\alpha \in \mathcal{P}$*. Then, a* pseudo-obstruction Υ *of* $\mathbf{I}$ *is a core of* $\mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$*.*

We now relate pseudo-obstructions and optimality.

Theorem 6. *Let* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *be a CQE-instance with* $\mathcal{O}$ *a Datalog ontology.*

1. *If* Υ *is a finite pseudo-obstruction for* $\mathbf{I}$*, then* $U = \bigvee_{Q \in \Upsilon} Q$ *is an optimal obstruction for* $\mathbf{I}$*.*
2. *If each pseudo-obstruction for* $\mathbf{I}$ *is infinite, then no optimal obstruction censor for* $\mathbf{I}$ *exists.*

This theorem has consequences on the expressive power of obstructions. Using the results from Section 5.1 we can see that optimal view and obstruction censors are incomparable. This complements Theorem 2, which talks about not necessarily optimal censors.

Theorem 7. *There is a CQE-instance with ontology in both RL and EL (respectively, RL) for which an optimal view (respectively, obstruction) censor exists, but no optimal obstruction (respectively, view) censor exists.*

Next, we show how to apply resolution-based techniques to compute optimal obstructions for instances with linear RL ontologies. These results are then adapted to the case of QL. The algorithm for linear RL is based on the computation of the set $\mathbb{Q}(\mathbf{I})$. To do this computation efficient, we need the following auxiliary structure.

Definition 12. *Let* $\mathcal{O}$ *be a linear RL ontology,* $\mathcal{D}$ *a dataset,* x *and* y *fresh variables, and* $\mathcal{A}$ *the set of all equality-free atoms over the signature of* $\mathcal{O} \cup \mathcal{D}$ *extended with* x *and* y*. The* proof graph *of* $\mathcal{O} \cup \mathcal{D}$ *is the directed graph with the set of nodes* $\mathcal{A} \cup \{\top\}$*, and edges* (β, γ) *such that* γ *can be derived from* β *by means of a single SLD resolution step with a rule from* $\mathcal{O} \cup \mathcal{D}$*.*

The following example illustrates proof graphs.

Example 5. Consider a CQE-instance $\mathbf{I}^1_{\mathsf{ex}}$ with ontology $\mathcal{O}^1_{\mathsf{ex}} = \{\mathit{Likes}(x, y) \rightarrow \mathit{Movie}(y),\ \mathit{Likes}(x, y) \rightarrow \mathit{MovieFan}(x)\}$, dataset $\mathcal{D}^1_{\mathsf{ex}} = \{\mathit{Likes}(\mathit{John}, \mathit{Seven})\}$, and the policy of single atom $\alpha^1_{\mathsf{ex}} = \mathit{MovieFan}(\mathit{John})$. A fragment of the proof graph is given in Figure 2.

Using proof graphs we can compute optimal censors.

Theorem 8. *Let* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *be a CQE-instance with* $\mathcal{O}$ *a linear RL ontology. For each* $\alpha \in \mathcal{P}$*, let* S_α *be the set of nodes in the proof graph of* $\mathcal{O} \cup \mathcal{D}$ *in a path from* α *to* $\top$*. Finally, let* U *be the Boolean UCQ*

$$\bigvee_{\alpha \in \mathcal{P}} \bigvee_{G \in S_\alpha \setminus \{\top\}} \exists^* G.$$

Then, $\mathsf{ocens}^U_{\mathbf{I}}$ *is the unique optimal censor for* $\mathbf{I}$*, and* U *can be computed in polynomial time in the size of* $\mathbf{I}$*.*

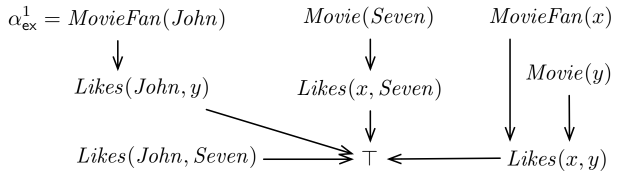
$\alpha^1_{\mathsf{ex}} = \mathit{MovieFan}(\mathit{John})$
$\mathit{Movie}(\mathit{Seven})$
$\mathit{MovieFan}(x)$
$\mathit{Likes}(\mathit{John}, y)$
$\mathit{Likes}(x, \mathit{Seven})$
$\mathit{Movie}(y)$
$\mathit{Likes}(\mathit{John}, \mathit{Seven})$
$\top$
$\mathit{Likes}(x, y)$

Fig. 2. Fragment of proof graph for $\mathcal{O}^1_{\mathsf{ex}} \cup \mathcal{D}^1_{\mathsf{ex}}$

Example 6. For $\mathbf{I}^1_{\mathsf{ex}}$ from Example 5, there is only one path in the proof graph from α^1_{ex} to $\top$ and $S_{\alpha^1_{\mathsf{ex}}} \setminus \{\top\} = \{\mathit{MovieFan}(\mathit{John}), \mathit{likes}(\mathit{John}, y)\}$. Thus, $U = \mathit{MovieFan}(\mathit{John}) \vee \exists y.\mathit{Likes}(\mathit{John}, y)$ is optimal.

Finally, note that the transformation of a QL ontology $\mathcal{O}$ to an RL ontology $\Xi_\sigma(\mathcal{O})$ given in Definition 9, preserves linearity of rules. Hence, Proposition 3 and Theorem 8 yield the following result.

Theorem 9. *For every CQE-instance with a QL ontology there exists a unique optimal obstruction censor.*

7 Discussion and Conclusions

We have studied CQE in the context of ontologies. Our results yield a flexible way for system designers to ensure selective access to data and provide insights on the fundamental tradeoff between accessibility and confidentiality of information. We have proposed algorithms applicable to the profiles of OWL 2, which can be implemented using off-the-shelf query answering infrastructure. Thus, our algorithms provide a starting point to the development of CQE systems.

The problems studied here remain rather unexplored and we see many open questions. From a theoretic point of view, we plan to consider policies beyond sets of facts (e.g., given as CQs). We also plan to study weaker notions of optimality that can ensure polynomiality of views and obstructions for more expressive languages. From a practical perspective, we will implement our algorithms and test their scalability using state-of-the art Datalog engines such as RDFox.[3]

The approach closest to ours is the view-based access authorisation framework in [9]. In this setting, policies are represented as *authorisation views*: CQs that define the only information accessible to the user; since queries are answered faithfully against the views, there is no explicit notion of policy violation. In contrast, in our setting policies express inaccessible information, and our goal is to maximally answer queries without violating the policy.

Acknowledgements. Work supported by the Royal Society, the EPSRC projects Score!, Exoda, and MaSI[3], and the FP7 project OPTIQUE.

[3] http://www.cs.ox.ac.uk/isg/tools/RDFox/

References

1. Bao, J., Slutzki, G., Honavar, V.: Privacy-Preserving Reasoning on the Semantic Web. In: WI. pp. 791–797. IEEE Computer Society (2007)
2. Biskup, J., Bonatti, P.: Controlled Query Evaluation with Open Queries for a Decidable Relational Submodel. Ann. Math. and Artif. Intell. 50(1-2), 39–77 (2007)
3. Biskup, J., Bonatti, P.A.: Lying Versus Refusal for Known Potential Secrets. Data Knowl. Eng. 38(2), 199–222 (2001)
4. Biskup, J., Bonatti, P.A.: Controlled Query Evaluation for Enforcing Confidentiality in Complete Information Systems. Int. J. Inf. Sec. 3(1), 14–27 (2004)
5. Biskup, J., Weibert, T.: Keeping Secrets in Incomplete Databases. Int. J. Inf. Sec. 7(3), 199–217 (2008)
6. Bonatti, P.A., Kraus, S., Subrahmanian, V.S.: Foundations of Secure Deductive Databases. IEEE Trans. Knowl. Data Eng. 7(3), 406–422 (1995)
7. Bonatti, P.A., Sauro, L.: A Confidentiality Model for Ontologies. In: ISWC. pp. 17–32 (2013)
8. Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: View-based Query Answering over Description Logic Ontologies. In: KR. AAAI Press (2008)
9. Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: View-based Query Answering in Description Logics: Semantics and Complexity. J. Comput. Syst. Sci. 78(1), 26–46 (2012)
10. Cuenca Grau, B.: Privacy in ontology-based information systems: A pending matter. Semantic Web 1(1-2), 137–141 (2010)
11. Cuenca Grau, B., Kharlamov, E., Kostylev, E.V., Zheleznyakov, D.: Controlled Query Evaluation over OWL 2 RL Ontologies. In: ISWC. pp. 49–65 (2013)
12. Cuenca Grau, B., Motik, B.: Reasoning over Ontologies with Hidden Content: The Import-by-Query Approach. J. Artif. Intell. Res. 45, 197–255 (2012)
13. Kolaitis, P.G., Vardi, M.Y.: A Logical Approach to Constraint Satisfaction. In: Complexity of Constraints. pp. 125–155 (2008)
14. Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyaschev, M.: The Combined Approach to Ontology-Based Data Access. In: IJCAI. pp. 2656–2661 (2011)
15. Lutz, C., Seylan, I., Toman, D., Wolter, F.: The Combined Approach to OBDA: Taming Role Hierarchies Using Filters. In: ISWC. pp. 314–330 (2013)
16. Lutz, C., Toman, D., Wolter, F.: Conjunctive Query Answering in the Description Logic EL Using a Relational Database System. In: IJCAI. pp. 2070–2075 (2009)
17. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-Based Access Control Models. IEEE Computer 29(2), 38–47 (1996)
18. Sicherman, G.L., de Jonge, W., van de Riet, R.P.: Answering Queries Without Revealing Secrets. ACM Trans. Database Syst. 8(1), 41–59 (1983)
19. Stefanoni, G., Motik, B., Horrocks, I.: Introducing Nominals to the Combined Query Answering Approaches for EL. In: AAAI (2013)
20. Stouppa, P., Studer, T.: A Formal Model of Data Privacy. In: PSI (2007)
21. Tao, J., Slutzki, G., Honavar, V.: Secrecy-Preserving Query Answering for Instance Checking in $\mathcal{EL}$. In: RR. pp. 195–203 (2010)

A Appendix

The following proposition will be used in several places throughout this appendix.

Proposition 4. *The censor $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ based on a view $\mathcal{V}$ is confidentiality preserving if and only if $\mathcal{O} \cup \mathcal{V} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. Additionally, it is optimal if and only if for each CQ $Q(\boldsymbol{x})$ and each $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$, the fact that $\mathcal{O} \cup \mathcal{V} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$ for any $\alpha \in \mathcal{P}$ implies that $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{V})$.*

Proof. Assume that $\mathcal{O} \cup \mathcal{V} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. Trivially, $\mathcal{O} \cup \mathcal{V} \models \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}}$ and hence we have $\mathcal{O} \cup \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \not\models \alpha$ for each $\alpha \in \mathcal{P}$, as required. Assume that cens is confidentiality preserving, in which case $\mathcal{O} \cup \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. Next, assume for the sake of contradiction that $\mathcal{O} \cup \mathcal{V} \models \alpha$ for some $\alpha \in \mathcal{P}$; since $\mathcal{O} \cup \mathcal{D} \models \alpha$ by the definition of policy we have that $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(\alpha) = \textbf{True}$ and thus $\alpha \in \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}}$; thus, $\mathcal{O} \cup \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \models \alpha$, which is a contradiction.

We next focus on the optimality statement. Assume that $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ is not optimal. Then, there is a confidentiality preserving censor cens that extends $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$; this means that for some CQ $Q(\boldsymbol{x})$ and, $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ we have $\boldsymbol{t} \in \mathsf{cens}(Q)$, but $\boldsymbol{t} \notin \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q)$. The fact that $\boldsymbol{t} \notin \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q)$ and $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ implies that $\boldsymbol{t} \notin \mathsf{cert}(Q, \mathcal{O}, \mathcal{V})$. Furthermore, the fact that cens is confidentiality-preserving implies that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$; but then, since cens extends $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$, we have that $\mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \subseteq \mathsf{Th}_{\mathsf{cens}}$ and hence $\mathcal{O} \cup \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$, as required.

Finally, assume that there exists some CQ $Q(\boldsymbol{x})$ and, $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ such that $\mathcal{O} \cup \mathcal{V} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$, but $\mathcal{O} \cup \mathcal{V} \not\models Q(\boldsymbol{t})$; then, we can define a censor cens that behaves exactly like $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$, with the exception of answering $Q(\boldsymbol{t})$ positively. Thus, $\mathsf{Th}_{\mathsf{cens}} = \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}} \cup \{Q(\boldsymbol{t})\}$. But then, since $\mathcal{O} \cup \mathcal{V} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$ and $\mathcal{O} \cup \mathcal{V} \models \mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}}$ we have that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \not\models \alpha$, which implies that cens is confidentiality preserving and $\mathsf{Th}_{\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}}$ is not optimal, as required.

A.1 Proofs for Section 4

Before proving Theorem 1 we present the following notation and a lemma. Let $\mathcal{I}$ be a finite structure and f a function associating a fresh variable to each domain element of $\mathcal{I}$. The query $Q^{\mathcal{I}}$ for $\mathcal{I}$ is the Boolean CQ defined as follows, with $R_1, \ldots R_n$ the predicates interpreted by $\mathcal{I}$:

$$Q^{\mathcal{I}} = \exists^* \bigwedge_{1 \leq i \leq n} \{R_i(f(u_1), \ldots, f(u_{m_i})) \mid (u_1, \ldots, u_{m_i}) \in R_i^{\mathcal{I}}\}.$$

Lemma 1. *Let $\mathcal{J}$ be a finite structure and let $\mathcal{C}$ be a class of all finite structures. Then, the following holds:*

$$\neg \mathsf{CSP}_{[\mathcal{C}]}(\mathcal{J}) = \{\mathcal{I} \in \mathcal{C} \mid \mathcal{I} \models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{J}} Q^{\mathcal{K}}\}.$$

Proof. We need to show the following:

$$\{\mathcal{I} \in \mathcal{C} \mid \mathcal{I} \not\to \mathcal{J}\} = \{\mathcal{I} \in \mathcal{C} \mid \mathcal{I} \models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{J}} Q^{\mathcal{K}}\}.$$

Let $\mathcal{I} \in \mathcal{C}$ be such that $\mathcal{I} \not\to \mathcal{J}$; clearly, $\mathcal{I} \models Q^{\mathcal{I}}$ and hence $\mathcal{I} \models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{J}} Q^{\mathcal{K}}$, as required. Conversely, assume that $\mathcal{I} \in \mathcal{C}$ is such that $\mathcal{I} \models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{J}} Q^{\mathcal{K}}$; then, there exists $\mathcal{K}$ such that $\mathcal{K} \in \mathcal{C}$, $\mathcal{K} \not\to \mathcal{J}$ and $\mathcal{I} \models Q^{\mathcal{K}}$. The latter implies that $\mathcal{K} \hookrightarrow \mathcal{I}$ and hence we can deduce $\mathcal{I} \not\to \mathcal{J}$, as required (otherwise, we would have by composition of homomorphisms that $\mathcal{K} \hookrightarrow \mathcal{J}$, which is a contradiction).

Theorem 1. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ Datalog ontology, and $\mathcal{C}$ the class of finite structures $\mathcal{I}$ with $\mathcal{I} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$. Then, $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}} = \mathsf{ocens}_{\mathbf{I}}^{U}$ iff U defines $\neg\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$, for any view $\mathcal{V}$ and obstruction U.*

Proof.
(⇐) Assume that U defines $\neg\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$. Then, for each $\mathcal{I} \in \mathcal{C}$ we have that $\mathcal{I} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}$ iff $\mathcal{I} \models U$. By Lemma 1, the following holds for each $\mathcal{I} \in \mathcal{C}$:

$$\mathcal{I} \models U \quad \text{iff} \quad \mathcal{I} \models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}} Q^{\mathcal{K}}. \tag{1}$$

Let $Q(\boldsymbol{x})$ be a CQ, and let $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$, which implies that $\mathbf{A}[Q(\boldsymbol{t})] \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$ and hence $\mathbf{A}[Q(\boldsymbol{t})] \in \mathcal{C}$. We show that $\boldsymbol{t} \in \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q)$ iff $\boldsymbol{t} \in \mathsf{ocens}_{\mathbf{I}}^{U}(Q)$.

For the forward direction, assume that $\boldsymbol{t} \in \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q)$; then, $\mathcal{O} \cup \mathcal{V} \models Q(\boldsymbol{t})$ and hence $\mathbf{A}[Q(\boldsymbol{t})] \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}$. We can then conclude $\mathbf{A}[Q(\boldsymbol{t})] \not\models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}} Q^{\mathcal{K}}$ (otherwise, $\mathcal{K} \hookrightarrow \mathbf{A}[Q(\boldsymbol{t})]$ for some $Q^{\mathcal{K}}$ in U and since we have established that $\mathbf{A}[Q(\boldsymbol{t})] \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}$ and homomorphism compose we would have $\mathcal{K} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}$ which is a contradiction). But then, (1) implies that $\mathbf{A}[Q(\boldsymbol{t})] \not\models U$ and by the definition of obstruction-censor that $\boldsymbol{t} \in \mathsf{ocens}_{\mathbf{I}}^{U}(Q)$, as required.

For the backward direction, assume now that $\boldsymbol{t} \in \mathsf{ocens}_{\mathbf{I}}^{U}(Q)$. Then, by the definition of obstruction censor we have $\mathbf{A}[Q(\boldsymbol{t})] \not\models U$. By (1) we then have $\mathbf{A}[Q(\boldsymbol{t})] \not\models \bigvee_{\mathcal{K} \in \mathcal{C}, \mathcal{K} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}} Q^{\mathcal{K}}$; Lemma 1 immediately implies that $\mathbf{A}[Q(\boldsymbol{t})] \notin \neg\mathsf{CSP}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$. From this, we must conclude that $\mathbf{A}[Q(\boldsymbol{t})] \in \mathsf{CSP}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$ and hence $\mathbf{A}[Q(\boldsymbol{t})] \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}$, which implies $\mathcal{O} \cup \mathcal{V} \models Q(\boldsymbol{t})$ and $\boldsymbol{t} \in \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q)$, as required.

(⇒) Assume that $\mathsf{ocens}_{\mathbf{I}}^{U} = \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$. To show that U defines $\neg\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$ we prove that $\mathcal{I} \models U$ iff $\mathcal{I} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}$ for every structure $\mathcal{I}$ such that $\mathcal{I} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$. If $\mathcal{I} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$ and $\mathcal{I} \models U$, then $\mathsf{ocens}_{\mathbf{I}}^{U}(Q^{\mathcal{I}}) = \mathbf{False}$. Since $\mathsf{ocens}_{\mathbf{I}}^{U} = \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ we also have that $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q^{\mathcal{I}}) = \mathbf{False}$ and hence $\mathcal{O} \cup \mathcal{V} \not\models Q^{\mathcal{I}}$. Consequently, $\mathcal{I} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}$, as required. If $\mathcal{I} \not\to \mathcal{H}_{\mathcal{O},\mathcal{V}}$, then $\mathcal{O} \cup \mathcal{V} \not\models Q^{\mathcal{I}}$; consequently, $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}(Q^{\mathcal{I}}) = \mathbf{False}$. Since $\mathsf{ocens}_{\mathbf{I}}^{U} = \mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ we have $\mathsf{ocens}_{\mathbf{I}}^{U}(Q^{\mathcal{I}}) = \mathbf{False}$ and hence since $\mathcal{I} \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{D}}$ we necessarily have $\mathcal{I} \models U$.

Theorem 2. *There is a CQE-instance for which there exists a view censor, but no obstruction censor. There is a CQE-instance for which there exists an obstruction censor, but no view censor.*

Proof. First we illustrate that obstruction censors cannot always simulate view censors. Consider CQE-instance $\mathbf{I} = (\emptyset, \mathcal{D}, \emptyset)$, where $\mathcal{D}$ represents an undirected graph with nodes "green" g and "blue" b, which are connected by *edge* in all possible ways:

$$\mathcal{D} = \{edge(g,b), edge(b,g), edge(b,b), edge(g,g)\}.$$

Clearly, $\mathcal{D}$ entails every Boolean CQ over the *edge* relation and thus every graph can be homomorphically embedded into $\mathcal{D}$. Consider $\mathcal{V} = \{edge(g,b), edge(b,g)\}$. Since the ontology is empty, $\mathcal{H}_{\emptyset,\mathcal{V}} = \mathcal{V}$ and $\neg\mathsf{CSP}(\mathcal{V})$ is the class of all graphs that are not 2-colourable. It is well-known that this class of graphs is not first-order definable and hence cannot be captured by a UCQ.

Next we construct an obstruction censor which cannot be simulated by a view censor. Consider the instance $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \emptyset)$, where $\mathcal{D} = \{edge(a,a)\}$ and $\mathcal{O}$ consists of the single transitivity rule

$$edge(x,y) \land edge(y,z) \to edge(x,z).$$

Clearly, $\mathcal{O} \cup \mathcal{D}$ entails each Boolean CQ over the *edge* relation. Consider obstruction $U = \exists y.edge(y,y)$, which defines the class of directed graphs with self loops. Suppose that some view $\mathcal{V}$ realises $\mathsf{ocens}_{\mathbf{I}}^{U}$. By Theorem 1, the obstruction U must define $\neg\mathsf{CSP}_{[\mathcal{C}]}(\mathcal{H}_{\mathcal{O},\mathcal{V}})$, where $\mathcal{C}$ is the class of all directed graphs. Thus, any graph G must satisfy the property

$$G \text{ has no self loops iff } G \hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}.$$

Due to the rule in $\mathcal{O}$, we conclude that $\mathcal{V}$ is a DAG, that is, it has no *edge*-loops. Take a DAG G extending (a graph isomorphic to) $\mathcal{H}_{\mathcal{O},\mathcal{V}}$ with a new node v and edges connecting all its sink nodes to v. Clearly G has no self loops, but $G \not\hookrightarrow \mathcal{H}_{\mathcal{O},\mathcal{V}}$, which is a contradiction.

A.2 Proofs for Section 5

Theorem 3. *There are CQE-instances* $\mathbf{I}_1$ *and* $\mathbf{I}_2$ *such that*

- *the ontology of* $\mathbf{I}_1$ *uses rules of Types* (1) *and* (9)*, and*
- *the ontology of* $\mathbf{I}_2$ *uses rules of Types* (5)*,* (8) *and* (15)*,*

for which no optimal view censors exist.

Proof. The first case of the theorem was proved in the work [11] for censors based on sound views, and the same proof extends to the general case. So, we next focus on showing the second case. To this end, let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be the CQE-instance defined as follows:

$$\begin{aligned}
\mathcal{O} = \{&E(x,y) \land E(y,z) \to E(x,z),\\
&E(x,x) \to F(x), E(x,y) \land F(y) \to F(x)\},\\
\mathcal{D} = \{&E(a,a)\},\\
\mathcal{P} = \{&F(a)\}.
\end{aligned}$$

Consider also the following (infinite) sequence of Boolean conjunctive queries (for $k \geq 1$):

$$Q_k(a) = \exists x_1, \ldots, x_k. E(x, x_1) \wedge \ldots \wedge E(x_{k-1}, x_k).$$

Clearly, $\mathcal{O} \cup \mathcal{D} \models Q_k(a)$ for each $k \geq 1$. Next, we show the following properties:

(P1) if an interpretation $\mathcal{I}$ satisfies $\mathcal{I} \models \mathcal{O} \cup \{Q_k(a) | k \geq 1\}$ and $\mathcal{I} \not\models F(a)$, then $\mathcal{I}$ has an infinite domain;
(P2) if cens is optimal for $\mathbf{I}$, then $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \models \{Q_k(a) | k \geq 1\}$.

To prove Property (P1), assume for the sake of contradiction that a finite model $\mathcal{I}$ of $\mathcal{O} \cup \{Q_k(a) | k \geq 1\}$ exists such that $\mathcal{I} \not\models F(a)$. Let m be the size of the interpretation domain of $\mathcal{I}$. Since $\mathcal{I} \models Q_m(a)$, there must exist elements $a_0, a_1, \ldots, a_{m-1}, a_m$ in the domain of $\mathcal{I}$ such that $a_0 = a$, and $(a_0, a_1), \ldots, (a_{m-1}, a_m) \in E^{\mathcal{I}}$. Since the size of the domain is just m, there must exist objects a_i and a_j for $0 \leq i < j \leq m$ such that $a_i = a_j$ and as a result $E^{\mathcal{I}}$ encodes a cycle. Since $\mathcal{O}$ axiomatises E to be a transitive relation, we have $(a_i, a_i) \in E^{\mathcal{I}}$ as well. This implies that $a_i \in F^{\mathcal{I}}$ and hence rule $E(x, y) \wedge F(y) \rightarrow F(x)$ will ensure that $a \in F^{\mathcal{I}}$, which contradicts our assumption that $\mathcal{I} \not\models F(a)$.

To prove Property (P2), consider an arbitrary confidentiality preserving censor cens for $\mathbf{I}$. It suffices to show that the censor cens' defined such that $\mathsf{Th}_{\mathsf{cens}'} = \mathsf{Th}_{\mathsf{cens}} \cup \{Q_k(a) | k \geq 1\}$ is also a confidentiality preserving censor for $\mathbf{I}$. Since cens is assumed to be confidentiality preserving, it holds that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \not\models F(a)$; thus, there exists an interpretation $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}}$, but $\mathcal{J} \not\models F(a)$. Let $\mathcal{J}'$ be the (infinite) interpretation over elements of $\mathcal{J}$ and fresh elements a_k, $k \geq 1$, defined as follows, where Trans of a relation instance is the fixpoint of the program axiomatising transitivity of the relation together with the instance:

$$E^{\mathcal{J}'} = \mathsf{Trans}(E^{\mathcal{J}} \cup \{(a, a_1)\} \cup \{(a_k, a_{k+1}) | k \geq 1\}),$$
$$F^{\mathcal{J}'} = F^{\mathcal{J}}.$$

We can check that $\mathcal{J}' \models \mathcal{O}$. Also, $\mathcal{J}' \models \mathsf{Th}_{\mathsf{cens}}$ because $\mathcal{J}$ does and $\mathcal{J}'$ simply extends $\mathcal{J}$ with new information (remember that $\mathsf{Th}_{\mathsf{cens}}$ consists of existentially quantified positive formulas). Also, $\mathcal{J}' \models \{Q_k | k \geq 1\}$ and $\mathcal{J}' \not\models F(a)$; thus, $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}'} \not\models F(a)$, as required.

Next, we use properties (P1) and (P2) to show the claim of the theorem. Assume for the sake of contradiction that for some view $\mathcal{V}$ we have that $\mathsf{vcens}^{\mathcal{V}}_{\mathbf{I}}$ is optimal. Then, Proposition 4 and Property (P2) imply $\mathcal{O} \cup \mathcal{V} \models \{Q_k(a) | k \geq 1\}$. Clearly, Since $\mathcal{O} \cup \mathcal{V}$ is satisfiable and $\mathcal{O}$ is Datalog, we have $\mathcal{H}_{\mathcal{O},\mathcal{V}} \models \mathcal{O} \cup \mathcal{V}$. Furthermore, since $\mathcal{O} \cup \mathcal{V} \models \{Q_k(a) | k \geq 1\}$ and $\mathcal{H}_{\mathcal{O},\mathcal{V}} \models \mathcal{O} \cup \mathcal{V}$ we must also have that $\mathcal{H}_{\mathcal{O},\mathcal{V}} \models \mathcal{O} \cup \{Q_k | k \geq 1\}$. However, $\mathcal{H}_{\mathcal{O},\mathcal{V}} \not\models F(a)$ since cens is confidentiality preserving, which together with the fact that $\mathcal{H}_{\mathcal{O},\mathcal{V}}$ is finite contradicts Property (P1).

Theorem 4. *Let* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *be a CQE-instance with* $\mathcal{O}$ *a guarded RL ontology, and* $\mathcal{V}$ *an exhaustive view on* $\mathbf{I}$*. Then* $\mathcal{V}$ *is optimal. Furthermore, if* $\mathcal{O}$ *is linear, then* $\mathsf{vcens}^{\mathcal{V}}_{\mathbf{I}}$ *is the only optimal censor for* $\mathbf{I}$.

Proof. First, note that by the construction of exhaustive view $\mathcal{H}_{\mathcal{O},\mathcal{V}} = \mathcal{V}$, that is, no rules from $\mathcal{O}$ can be applied to $\mathcal{V}$. Indeed, this follows from the facts that

- the set $\mathcal{V}_c^1$ is maximal, so it is closed, i.e., $\mathcal{H}_{\mathcal{O},\mathcal{V}_c^1} = \mathcal{V}_c^1$,
- the sets $\mathcal{V}_\exists^1$ and $\mathcal{V}^2$ are also closed,
- binary atoms in $\mathcal{V}^2$ cannot influence unary atoms in $\mathcal{V}_c^1 \cup \mathcal{V}_\exists^1$,
- no rules can be applied to identify any constants: constants from $\mathbf{I}$ are already identified in $\mathcal{H}_{\mathcal{O},\mathcal{D}}$, and new constants (anonymous copies) cannot be identified by their second requirement.

Since it is explicitly required that no policy atom is in the $\mathcal{V}$, we can conclude that the censor $\mathsf{vcens}^{\mathcal{V}}_{\mathcal{O},\mathcal{D}}$ is confidentiality preserving.

From the construction we immediately have that if $\mathcal{O} \cup \mathcal{V} \models Q(\boldsymbol{t})$ then $\mathcal{O} \cup \mathcal{D} \models Q(\boldsymbol{t})$ for any CQ Q and tuple $\boldsymbol{t}$, that is $\mathcal{V}$ is sound.

Next we show that $\mathsf{vcens}^{\mathcal{V}}_{\mathcal{O},\mathcal{D}}$ is optimal. Consider arbitrary CQ $Q(\boldsymbol{x})$ and tuple of constants $\boldsymbol{t}$ such that $\mathcal{O} \cup \mathcal{D} \models Q(\boldsymbol{t})$ and $\mathcal{V} \cup Q(\boldsymbol{t}) \not\models \alpha$ for any $\alpha \in \mathcal{P}$. We need to prove that $\mathcal{V} \models Q(\boldsymbol{t})$, i.e., there exists a homomorphism from $\mathbf{A}[Q(\boldsymbol{t})]$ to $\mathcal{V}$. Consider the structure $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$. Since $\mathcal{O} \cup \mathcal{D} \models Q(\boldsymbol{t})$, i.e., $\boldsymbol{t}$ is a certain answer to Q over $\mathcal{O} \cup \mathcal{D}$, there exists a homomorphism h from $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ to $\mathcal{H}_{\mathcal{O},\mathcal{D}}$. For every element d in $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ denote $\mathcal{A}(x)$ the set

$$\{A \mid A(d) \in \mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}\}.$$

Note, that for every element d in $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ which is not a constant from $\mathbf{I}$, there exists a variable x in $Q(\boldsymbol{x})$, such that $d = d_x$. Recall, that d_x is the individual in $\mathbf{A}[Q(\boldsymbol{t})]$ which corresponds to x. Moreover, the set $\sigma_{h(d)}$ in the view $\mathcal{V}$ contains the constant $(h(d))_{\mathcal{A}(x)}$—the anonymous copy of $h(d)$ with exactly the same set of unary predicates, as d. Hence, we can define a mapping h' from the domain of $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ to the domain of $\mathcal{V}$ as follows:

- $h'(a) = a$ for every constant a,
- $h'(d) = (h(d))_{\mathcal{A}(x)}$ for every d which is not a constant.

This mapping is a homomorphism from $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ to $\mathcal{V}$. Indeed,

- for every constant a from $\mathbf{I}$ if $A(a)$ is in $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$, then it is in $\mathcal{V}_c^1$ (and, hence, in $\mathcal{V}$), because $\mathcal{V}_c^1$ is maximal and $\mathcal{V} \cup Q(\boldsymbol{t})$ does not disclose the policy;
- for every element d which is not a constant we have $A(h'(d)) \in \mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ iff $A((h(d))_{\mathcal{A}(x)}) \in \mathcal{V}$ by the construction above;
- finally, for any pair of elements d_1, d_2, if $R(d_1, d_2) \in \mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ then it holds that $R(h'(d_1), h'(d_2)) \in \mathcal{V}^2$ (and, hence, in $\mathcal{V}$), because *(i)* $\mathcal{V}^2$ is constructed in a way that it is maximal among whose that does not change unary atoms, and do not disclose the policy; and *(2)* $\mathcal{V} \cup Q(\boldsymbol{t})$ also does not disclose the policy.

Since h' is a homomorphism from $\mathcal{H}_{\mathcal{O},\mathbf{A}[Q(\boldsymbol{t})]}$ to $\mathcal{V}$, there exists also a homomorphism from $\mathbf{A}[Q(\boldsymbol{t})]$ to $\mathcal{V}$, which means that $\mathcal{V} \models Q(\boldsymbol{t})$.

Finally, the fact that if $\mathcal{O}$ is linear then $\mathbf{I}$ has a unique optimal censor is shown in the proof of Theorem 8 below (of course, without using this theorem or any of its consequences).

Proposition 1. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a multi-linear RL ontology. There is an optimal censor for $\mathbf{I}$ based on a view of polynomial size.*

Proof. Consider an exhaustive view $\mathcal{V}$ on $\mathbf{I}$. For every constant a the set σ_a contains the constant $a_{\mathcal{A}^*}$ such that

$$\mathcal{A}^* = \{A \mid A(a) \in \mathcal{H}_{\mathcal{O},\mathcal{D}}, \mathcal{O} \not\models A(x) \to x \approx a \text{ for each } a\}.$$

The set $\mathcal{A}^*$ is subset maximal among all the constants in σ_a, i.e., if $a_{\mathcal{A}} \in \sigma_a$ then $\mathcal{A} \subseteq \mathcal{A}^*$.

First, note that if $R(a_{\mathcal{A}}, b_{\mathcal{B}})$ is in the binary part $\mathcal{V}^2$ for two new constants (anonymous copies) $a_{\mathcal{A}}$, and $b_{\mathcal{B}}$), then so is $R(a_{\mathcal{A}^*}, b_{\mathcal{B}^*})$. Similarly, from the fact that $\mathcal{O}$ is multi-linear, we conclude that if $S(a_{\mathcal{A}}, b)$ is in $\mathcal{V}^2$ for a new constant $a_{\mathcal{A}}$, constant b from $\mathbf{I}$ and binary predicate or inverse of a binary predicate S, then so is $S(a_{\mathcal{A}^*}, b)$.

Hence, for any CQ $Q(\boldsymbol{x})$ and any tuple $\boldsymbol{t}$ of constants, $\mathcal{V} \models Q(\boldsymbol{t})$ holds if and only if $\mathcal{V}^* \models Q(\boldsymbol{t})$, where $\mathcal{V}^*$ is the sub-view of $\mathcal{V}$ based on the constants a and $a_{\mathcal{A}^*}$. So, these two views realise the same censor, which is optimal by Theorem 4. The fact that $\mathcal{V}^*$ is of polynomial size completes the proof.

Proposition 2. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{D}$ using set of constants σ, $\mathcal{O}'$ a σ-rewriting of $\mathcal{O}$ such that $\mathcal{O}' \models \mathcal{O}$, and $\mathcal{V}'$ an optimal view for $(\mathcal{O}', \mathcal{D}, \mathcal{P})$. Then $\mathcal{H}_{\mathcal{O}',\mathcal{V}'}$ is an optimal view for $\mathbf{I}$.*

Proof. First we show the confidentiality preservation of the censor. Since $\mathsf{vcens}_{\mathbf{I}'}^{\mathcal{V}'}$ is confidentiality-preserving, we have that $\mathcal{O}' \cup \mathcal{V}' \not\models \alpha$ for each $\alpha \in \mathcal{P}$. Since $\mathcal{O}'$ is Datalog, it is clear that $\mathcal{H}_{\mathcal{O}',\mathcal{V}} = \mathcal{H}_{\mathcal{O}',\mathcal{V}'}$; thus, $\mathcal{O}' \cup \mathcal{V} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. But then, since $\mathcal{P}$ contains only facts and $\mathcal{O}'$ is a rewriting of $\mathcal{O}$ we have $\mathcal{O} \cup \mathcal{V} \not\models \alpha$ for each $\alpha \in \mathcal{P}$, as required.

Now we concentrate of the optimality of the view. Assume by contradiction that $\mathsf{vcens}_{\mathbf{I}}^{\mathcal{V}}$ is not optimal, then there exists a BCQ Q such that *(i)* $\mathcal{O} \cup \mathcal{D} \models Q$; *(ii)* $\mathcal{O} \cup \mathcal{V} \not\models Q$; and *(iii)* $\mathcal{O} \cup \mathcal{V} \cup \{Q\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. Since $\mathcal{O} \cup \mathcal{D} \models Q$ and $\mathcal{O}' \models \mathcal{O}$ we have *(iv)* $\mathcal{O}' \cup \mathcal{D} \models Q$. Furthermore, condition *(iii)* implies that $\mathcal{O} \cup \mathcal{V} \cup \mathbf{A}[Q] \not\models \alpha$ and since α is a fact and $\mathcal{O}'$ is a rewriting of $\mathcal{O}$ we have $\mathcal{O}' \cup \mathcal{V} \cup \mathbf{A}[Q] \not\models \alpha$, which by the fact that $\mathcal{V} \models \mathcal{V}'$ then also implies that *(v)* $\mathcal{O}' \cup \mathcal{V}' \cup \{Q\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. But then, *(iv)* and *(v)* and the fact that $\mathcal{V}'$ is optimal for $(\mathcal{O}', D, \mathcal{P})$ we must have $\mathcal{O}' \cup \mathcal{V}' \models Q$. Since $\mathcal{V} = \mathcal{H}_{\mathcal{O}',\mathcal{V}'}$ we have $\mathcal{V} \models Q$, which contradicts *(ii)*.

A.3 Proofs for Section 6

For the sake of ease in the proofs for theorems and propositions of this section we will consider only the class of BCQs with constants. Clearly, any results obtained for this class will also hold for the class of all CQs. Before proceeding to the main proofs, we introduce few definitions and lemmas.

Let $\mathcal{O}$ be a Datalog ontology and $\mathcal{D}$ a dataset; let $\mathbb{Q}$ be a possibly infinite set of queries such that $\mathcal{O} \cup \mathcal{D} \models Q$ for each $Q \in \mathbb{Q}$. Then a censor $\mathsf{cens}_{\mathbb{Q}}$ is defined as follows:

$$\mathsf{cens}_{\mathbb{Q}}(Q) = \mathbf{True} \quad \text{iff} \quad \mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) = \mathbf{True} \text{ and } \mathbf{A}[Q] \not\models Q' \text{ for each } Q' \in \mathbb{Q}.$$

Lemma 2. *Let* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *be a CQE-instance; let* Υ *be a pseudo-obstruction based on a subset* $\mathbb{S}$ *of* $\mathbb{Q}(\mathbf{I})$*. Then,* $\mathsf{cens}_{\Upsilon} = \mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$*.*

Proof. Let Q be a CQ such that $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) = \mathbf{True}$.

Assume that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}(Q) = \mathbf{False}$; this yields that $\mathbf{A}[Q] \models Q'$ for some $Q' \in \mathbb{Q}(\mathbf{I})\setminus\mathbb{S}$. Then there exists $Q'' \in \Upsilon$ such that $Q' \models Q''$ and thus $\mathbf{A}[Q] \models Q''$, i.e., $\mathsf{cens}_{\Upsilon}(Q) = \mathbf{False}$.

Assume that $\mathsf{cens}_{\Upsilon}(Q) = \mathbf{False}$; this yields that $\mathbf{A}[Q] \models Q''$ for some $Q'' \in \Upsilon$. Note that $Q'' \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$ since $\Upsilon \subseteq \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$ and thus $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}(Q) = \mathbf{False}$.

The proposition above allows us to speak of obstruction censors in terms of either Υ or $\mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$, whatever way is more convenient to show the required results. We are going to show now that a censor cens is optimal for a given CQE-instance $\mathbf{I}$ iff there exists a maximal subset $\mathbb{S}$ of $\mathbb{Q}(\mathbf{I})$ such that $\mathsf{cens} = \mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$. But first we need the following notion of a normalised proof.

Definition 13. *Let* $\mathcal{O}$ *be a Datalog ontology,* $\mathcal{D}$ *a dataset, and* G_0 *a goal. A proof* π *of length* n *of* G_0 *in* $\mathcal{O} \cup \mathcal{D}$ *is* normalised *if there is* $k \leq n$ *such that* $r_i \in \mathcal{O}$ *for each* $i < k$ *and* $r_j \in \mathcal{D}$ *for each* $j \geq k$*. Moreover, the number* k *is called the* frontier *of* π*, denoted* $\mathsf{fr}(\pi)$*.*

Intuitively, a normalised proof π works as follows: first we rewrite the initial query G_0 over the ontology $\mathcal{O}$ until we obtain the query $G_{\mathsf{fr}(\pi)-1}$ that can be mapped into $\mathcal{D}$, and then we perform such a mapping applying (r_i, θ_i) with $i \geq \mathsf{fr}(\pi)$. Observe that for every G_i with $i < \mathsf{fr}(\pi)$ it holds that $\mathcal{O} \cup G_i \models G_0$.

We exploit the following known result about SLD-resolution over Datalog ontologies.

Lemma 3. *Let* $\mathcal{O}$ *be a Datalog ontology, let* $\mathcal{D}$ *be a dataset, and let* G_0 *be a goal such that* $\mathcal{O} \cup \mathcal{D} \models G_0$*. Then there exists a normalised SLD-proof* π *of* G_0 *in* $\mathcal{O} \cup \mathcal{D}$*.*

Lemma 4. *Let* $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ *be a CQE-instance with* $\mathcal{O}$ *a Datalog ontology and* cens *a censor for* $\mathcal{O}$ *and* $\mathcal{D}$*. Then* cens *is optimal for* $\mathbf{I}$ *iff there exists a maximal subset* $\mathbb{S}$ *of* $\mathbb{Q}(\mathbf{I})$ *such that (i)* $\mathcal{O} \cup \mathbb{S} \not\models \alpha$ *for each* $\alpha \in \mathcal{P}$ *and (ii)* $\mathsf{cens} = \mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$*.*

Proof. We start with the "only if"-direction. Let us assume that such maximal subset $\mathbb{S}$ exists. We show that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$ is optimal.

First, we show that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$ is confidentiality preserving. Assume the contrary; then, there is a (finite) subset $\mathbb{F}$ of $\mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}}$ such that $\mathcal{O} \cup \mathbb{F} \models \alpha$ for some $\alpha \in \mathcal{P}$. This yields the existence of proof π of α in $\mathcal{O} \cup \mathbf{A}[\mathbb{F}]$, where $\mathbf{A}[\mathbb{F}] = \bigcup_{Q \in \mathbb{F}} \mathbf{A}[Q]$. Due to Lemma 3, we can assume that π is normalised with frontier $k+1$. Let G_k be the goal right before frontier in π. Since π is normalised, then G_k is proved by using only facts from $\mathbf{A}[\mathbb{F}]$. So, we can write G_k as $G_k = B_1 \wedge \ldots \wedge B_m$, where each B_j is the conjunction of all atoms that are proved using facts only from a particular $\mathbf{A}[Q_j]$. Obviously, the order in which these B_j are proved is irrelevant, so let us assume that all B_j have been proved except for B_i; since, the different B_j can share variables, the remaining goal to prove may not be just B_1, but rather $B_i\theta_i$, with θ_i some substitution. We make the following observations:

1. $B_i\theta_i$ does not mention any constants not in $\mathcal{O} \cup \mathcal{D}$. Indeed, for any distinct queries Q_k, Q_j in $\mathbb{F}$ we have that $\mathbf{A}[Q_k]$ and $\mathbf{A}[Q_j]$ only share constants from $\mathcal{O} \cup \mathcal{D}$ $\mathbf{A}[Q_k]$; thus, if $B_i\theta_i$ contains some constant coming from $\mathbf{A}[Q_j]$ with $j \neq i$, it wouldn't be possible to prove $B_i\theta_i$ using only facts from $\mathbf{A}[Q_i]$.
2. There exists a proof of α in $\mathcal{O} \cup \mathcal{D}$ such that $B_i\theta_i$ occurs as a subgoal. We construct such proof as follows. First, we can "reach" goal G_k because it only requires rules from $\mathcal{O}$. Note also that each B_j follows from $\mathcal{O} \cup \mathcal{D}$, so we can continue the proof by showing all B_j except for B_i. Then, we can do it in such a way we reach precisely $B_i\theta_i$ as a subgoal.
3. $Q_i \models \exists^* B_i\theta_i$ since $B_i\theta_i$ is provable from $\mathbf{A}[Q_i]$.

Observation 2 means that $B_i\theta_i \in \mathbb{Q}(\mathbf{I})$ for all $1 \leq i \leq m$. Furthermore, since the censor answers **True** for each Q_i we have that $B_i\theta_i \in \mathbb{S}$. But then, $\mathcal{O} \cup \mathbb{S} \models \alpha$, which is a contradiction.

Now we show the optimality of $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$. To do this we make use of the following result from [11]: a censor cens for $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ is optimal if and only if for each CQ $Q(\boldsymbol{x})$ and each $\boldsymbol{t} \in \mathsf{cert}(Q, \mathcal{O}, \mathcal{D})$ the fact that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \cup \{Q(\boldsymbol{t})\} \not\models \alpha$ holds for each $\alpha \in \mathcal{P}$ implies that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \models Q(\boldsymbol{t})$.

Due to this result, $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$ is optimal if and only if for each Q such that $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) = \mathbf{True}$ and $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}} \cup \{Q\} \not\models \alpha$, it holds that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}} \models Q$. Assume to the contrary that there exists a CQ Q such that $\mathsf{cert}(Q, \mathcal{O}, \mathcal{D}) = \mathbf{True}$ and $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}} \cup \{Q\} \not\models \alpha$, but $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}} \not\models Q$. The latter means that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}(Q) = \mathbf{False}$, that is, $\mathbf{A}[Q] \models Q'$, for some $Q' \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$.

Recall that for any $Q \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$ it holds that $\mathcal{O} \cup \mathbb{S} \cup \{Q\} \models \alpha$ by maximality of $\mathbb{S}$. Observe that $\mathbb{S} \subseteq \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}}$; this yields $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}} \cup \{Q\} \models \alpha$, which contradicts the initial assumption and concludes the "only if"-direction.

Now we consider the "if"-direction. Let us now assume that cens is optimal, and let $\mathbb{Q} = \{Q \mid \mathsf{cens}(Q) = \mathbf{False}\}$. Consider the following subset $\mathbb{S}$ of $\mathbb{Q}(\mathbf{I})$: $\mathbb{S} = \mathbb{Q}(\mathbf{I}) \setminus \mathbb{Q}$. To prove the theorem, it suffices to prove the following two conditions: *(i)* $\mathbb{S}$ is a maximal subset of $\mathbb{Q}(\mathbf{I})$ such that $\mathcal{O} \cup \mathbb{S} \not\models \alpha$ for each $\alpha \in \mathcal{P}$ and *(ii)* $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}} = \mathsf{cens}$.

To show *(i)*, assume that $\mathcal{O} \cup \mathbb{S} \cup \{Q\} \models \alpha$ for some $\alpha \in \mathcal{P}$. Clearly, since by construction $\mathbb{S} \subseteq \mathsf{Th}_{\mathsf{cens}}$, it holds that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \cup \{Q\} \models \alpha$, and therefore $\mathsf{cens}(Q) = \mathbf{False}$, i.e. $Q \in \mathbb{Q}$, which implies *(i)*.

To show *(ii)*, let us pick an arbitrary Q such that $\mathcal{O} \cup \mathcal{D} \models Q$ but $\mathsf{cens}(Q) = \mathbf{False}$ and hence $Q \in \mathbb{Q}$. Since cens is optimal, we have that $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \cup \{Q\} \models \alpha$

for some $\alpha \in \mathcal{P}$, so let $\mathbb{F}$ be any minimal subset of $\mathsf{Th}_{\mathsf{cens}}$ such that $\mathcal{O} \cup \mathbb{F} \cup \{Q\} \models \alpha$. Following the same arguments as we used in the "only if" direction we have that there exists $G \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$ such that $Q \models G$; since G is part of the obstruction, then $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}(Q) = \mathbf{False}$. Finally, assume that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}(Q) = \mathbf{False}$; then, $Q \models G$ for some $G \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$. Since $\mathbb{Q}(\mathbf{I}) \setminus \mathbb{S} \subseteq \mathbb{Q}$, we have that $\mathsf{cens}(Q) = \mathbf{False}$, as required.

Theorem 6. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a Datalog ontology.*

1. *If Υ is a finite pseudo-obstruction for $\mathbf{I}$, then $U = \bigvee_{Q \in \Upsilon} Q$ is an optimal obstruction for $\mathbf{I}$.*
2. *If each pseudo-obstruction for $\mathbf{I}$ is infinite, then no optimal obstruction censor for $\mathbf{I}$ exists.*

Proof. Let us prove Statement 1. Assume that Υ is a finite pseudo-obstruction. By Lemma 2, we have that $\mathsf{cens}_{\Upsilon} = \mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$. By the "only if" statement in Lemma 4, we have that $\mathsf{cens}_{\mathbb{Q}(\mathbf{I})\setminus\mathbb{S}}$ is optimal. But then, since Υ is finite, then U is an obstruction.

Next, we show Statement 2. Assume by contradiction that each pseudo-obstruction is infinite, but there is an optimal censor based on an obstruction U. Since $\mathsf{ocens}^{U}_{\mathbf{I}}$ is an optimal censor, then the "if" direction of Lemma 4 tells us that there exists a pseudo-obstruction Υ such that $\mathsf{ocens}^{U}_{\mathbf{I}} = \mathsf{cens}_{\Upsilon}$. We can show that this contradicts the fact that Υ is both a core and infinite. Pick any CQ Q from U; then, clearly, $\mathsf{ocens}^{U}_{\mathbf{I}}(Q) = \mathbf{False}$ and hence $\mathsf{cens}_{\Upsilon}(Q) = \mathbf{False}$. The latter implies that there exists $Q' \in \Upsilon$ such that $Q \models Q'$. Let us now construct $U' = \bigvee_{Q \in U} Q'$, which is finite and also a "subset" of Υ. To obtain a contradiction, it thus suffices to show now that $\mathsf{ocens}^{U'}_{\mathbf{I}} = \mathsf{cens}_{\Upsilon}$. Indeed, for each CQ Q such that $\mathsf{cert}(Q, \mathcal{D}, \mathcal{O}) = \mathbf{True}$ (recall that $\mathsf{ocens}^{U}_{\mathbf{I}} = \mathsf{cens}_{\Upsilon}$):

- Assume that $\mathsf{ocens}^{U}_{\mathbf{I}}(Q) = \mathbf{False}$; then there is Q' in U such that $\mathbf{A}[Q] \models Q'$, which yields $\mathbf{A}[Q] \models Q''$ with Q'' from U', and therefore $\mathsf{ocens}^{U'}_{\mathbf{I}}(Q) = \mathbf{False}$.
- Assume that $\mathsf{ocens}^{U'}_{\mathbf{I}}(Q) = \mathbf{False}$; then $\mathbf{A}[Q] \models Q''$ for some Q'' in U', and consequently, since $Q'' \in \mathbb{Q}(\mathbf{I}) \setminus \mathbb{S}$, we conclude that $\mathsf{cens}_{\Upsilon}(Q) = \mathbf{False}$.

The obtained contradiction concludes the proof.

Theorem 7. *There is a CQE-instance with ontology in both RL and EL (respectively, RL) for which an optimal view (respectively, obstruction) censor exists, but no optimal obstruction (respectively, view) censor exists.*

Proof. To show the first statement, consider $\mathbf{I}_1 = (\mathcal{O}_1, \mathcal{D}_1, \mathcal{P}_1)$, where $\mathcal{D}_1 = \{R(a,a), A(a)\}$, $\mathcal{P}_1 = \{A(a)\}$, and the guarded RL (and EL) ontology $\mathcal{O}_1 = \{R(x,y) \wedge A(y) \rightarrow A(x)\}$. Since the ontology $\mathcal{O}_1$ is guarded, by Theorem 4 we can devise an optimal view. In contrast, there are infinitely many "non-redundant" proofs of $A(a)$ in $\mathcal{O}_1 \cup \mathcal{D}_1$. To see this, note that we have all of the following (infinitely many) normalised proofs for $n \geq 1$, and where the first line marks the frontier

$$A(a) \to R(a, y_1) \wedge A(y_1) \to \ldots \to R(a, y_1) \wedge \ldots \wedge R(y_{n-1}, y_n) \wedge A(y_n) \to \\ R(a, y_1) \wedge \ldots \wedge R(y_{n-1}, a) \to \ldots \to R(a, a) \to \top.$$

This means, that for every n, $\mathbb{Q}(\mathbf{I}_1)$ contains all of the following CQs for $n \geq 1$:

$$Q_n = \exists \boldsymbol{y}.[R(a, y_1) \wedge \ldots \wedge R(y_{n-1}, y_n) \wedge A(y_n)].$$

It is clear that no maximal subset $\mathbb{S}$ of $\mathbb{Q}(\mathbf{I}_1)$ such that $\mathcal{O} \cup \mathbb{S} \not\models A(a)$ can contain any of such queries; therefore, for any such $\mathbb{S}$ we have $\mathbb{Q}(\mathbf{I}_1) \setminus \mathbb{S}$ contains all such Q_n. But then, the core of such set also contains all these Q_n every pseudo-obstruction not entailing the policy atom must contain all such queries. Therefore, each pseudo-obstruction is infinite and by Theorem 6, no optimal obstruction censor can exist.

To show the second statement, consider CQE-instance $\mathbf{I}_2 = (\mathcal{O}_2, \mathcal{D}_2, \mathcal{P}_2)$, whith $\mathcal{D}_2 = \{R(a, a)\}$, $\mathcal{P}_2 = \{A(a)\}$, and $\mathcal{O}_2 = \{R(x_1, y) \wedge R(x_2, y) \to x_1 \approx x_2, R(x, y) \to A(y)\}$. From [11] we know that no optimal view exists for this instance, and the proof extends also to the case where views are not required to be sound. However, $U = A(a) \vee \exists x. R(x, a)$ is an optimal obstruction, since there is only one proof of $A(a)$ with subgoal $R(x, a)$.

Theorem 8. *Let $\mathbf{I} = (\mathcal{O}, \mathcal{D}, \mathcal{P})$ be a CQE-instance with $\mathcal{O}$ a linear RL ontology. For each $\alpha \in \mathcal{P}$, let S_α be the set of nodes in the proof graph of $\mathcal{O} \cup \mathcal{D}$ in a path from α to $\top$. Finally, let U be the Boolean UCQ*

$$\bigvee_{\alpha \in \mathcal{P}} \bigvee_{G \in S_\alpha \setminus \{\top\}} \exists^* G.$$

Then, $\mathsf{ocens}_{\mathbf{I}}^{U}$ is the unique optimal censor for $\mathbf{I}$, and U can be computed in polynomial time in the size of $\mathbf{I}$.

Proof. Optimality and uniqueness follows from Theorem 6 and the facts that *(i)* the set $\bigvee_{\alpha \in \mathcal{P}} S_\alpha$ is exactly $\mathbb{Q}(\mathbf{I})$ *(ii)* the only maximal subset $\mathbb{S}$ of $\mathbb{Q}(\mathbf{I})$ such that $\mathcal{O} \cup \mathbb{S}$ does not entail any α is the empty set. To prove the former fact, first observe that any goal that can appear in any SLD-proof in $\mathcal{O} \cup \mathcal{D}$ is isomorphic to one of the nodes of the proof-graph of $\mathcal{O} \cup \mathcal{D}$; then Fact (i) follows directly from the construction of the proof-graph. Fact (ii) follows from the observation that each SLD-proof is normalised, and therefore for each $Q \in S_\alpha$ it holds that $\mathcal{O} \cup Q \models \alpha$.

Finally, polynomiality follows from the fact that in linear RL the size of the proof-graph of α is at most cubic in $|\mathcal{O} \cup \mathcal{D} \cup \{\alpha\}|$.

Theorem 9. *For every CQE-instance with a QL ontology there exists a unique optimal obstruction censor.*

Proof. Let cens' be the optimal censor for $\mathbf{I}' = (\Xi_\sigma(\mathcal{O}), \mathcal{D}, \mathcal{P})$, where σ is a set of constants of $\mathbf{I}$ and $\Xi_\sigma(\mathcal{O})$ is a linear RL ontology. By Theorem 8, $\mathsf{cens}' = \mathsf{ocens}_{\mathbf{I}'}^{U}$

for the UCQ U as defined in the theorem. Let $\mathsf{cens} = \mathsf{ocens}_{\mathbf{I}}^{U}$. We are going to show that cens is an optimal censor for $\mathbf{I}$.

Confidentiality preservation. Assume that cens is not confidentiality preserving for $\mathbf{I}$, that is, $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \models \alpha$ for some $\alpha \in \mathcal{P}$. This means that there exist $Q_1, \ldots, Q_n \in \mathsf{Th}_{\mathsf{cens}}$ such that $\mathcal{O} \cup \{Q_1, \ldots, Q_n\} \models \alpha$; clearly, $\mathcal{O} \cup \mathcal{D} \models Q_i$ for each $i \in \{1, \ldots, n\}$. By Proposition 3, $\Xi_\sigma(\mathcal{O}) \models \mathcal{O}$ and consequently $\Xi_\sigma(\mathcal{O}) \cup \mathcal{D} \models Q_i$ for each $i \in \{1, \ldots, n\}$. Since cens' is confidentiality preserving for $\mathbf{I}'$, we conclude that $\{Q_1, \ldots, Q_n\} \not\subseteq \mathsf{Th}_{\mathsf{cens}'}$, so there is $j \in \{1, \ldots, n\}$ such that $\mathsf{cens}'(Q_j) = \mathbf{False}$; i.e., $\mathbf{A}[Q_i] \models U$. The last entailment implies that $\mathsf{cens}(Q_j) = \mathbf{False}$, i.e., $Q_j \notin \mathsf{Th}_{\mathsf{cens}}$, which yields a contradiction and thus cens is confidentiality preserving for $\mathbf{I}$.

Optimality. Assume for the sake of contradiction, that cens is not optimal for $\mathbf{I}$, that is, there exists Q such that *(i)* $\mathcal{O} \cup D \models Q$, *(ii)* $Q \notin \mathsf{Th}_{\mathsf{cens}}$, and *(iii)* $\mathcal{O} \cup \mathsf{Th}_{\mathsf{cens}} \cup \{Q\} \not\models \alpha$ for each $\alpha \in \mathcal{P}$. This yields $\mathbf{A}[Q] \models u$ for some disjunct u in U and consequently $\mathsf{cens}'(Q) = \mathbf{False}$. Note that for each disjunct u in U, it holds that $\Xi_\sigma(\mathcal{O}) \cup \{u\} \models \alpha$ for some $\alpha \in \mathcal{P}$; thus $\Xi_\sigma(\mathcal{O}) \cup \{Q\} \models \alpha$. There are the following cases depending on the form of u.

- If u is of the form $A(a)$ or $R(a, b)$ with $a, b \in \sigma$, then $\mathcal{O} \cup \{u\} \models \alpha$ since, due to Proposition 3, $\Xi_\sigma(\mathcal{O})$ is a σ-rewriting of $\mathcal{O}$; thus, $\mathcal{O} \cup \{Q\} \models \alpha$ which yields a contradiction with *(iii)*.
- If u is of the form $\exists y.R(a, y)$ with $a \in \sigma$, then let $O_{\min}$ be a minimal subset of $\Xi_\sigma(\mathcal{O})$ such that $\mathcal{O}_{\min} \cup \{u\} \models \alpha$. Due to the assumption, it holds $\mathcal{O} \cup \{u\} \not\models \alpha$; thus, $\mathcal{O}_{\min} \not\subseteq \mathcal{O}$ and therefore $\mathcal{O}_{\min}$ includes one of the rules introduced by Ξ. That is, $\mathcal{O}_{\min}$ contains (some of) the following rules that come from the Skolemisation $\Xi_\sigma(r)$ of some rule $r = A(x) \to \exists y.[S(x, y) \wedge B(y)]$ of Type (3) in $\mathcal{O}$:

  $$A(x) \to P_S(x, c_{A,S}), \quad P_S(x, y) \to S(x, y), \quad \text{and } P_S(x, y) \to B(y). \quad (2)$$

  Consider a proof $\pi = G_0 \to \ldots \to G_n$ of α in $\Xi_\sigma(\mathcal{O}) \cup \mathbf{A}[\exists y.R(a, y)]$, where $G_0 = \alpha$. Clearly, G_i can be obtained from G_{i-1} by applying a rule from $\mathcal{O}_{\min}$ for each $i = 1, \ldots, n-1$, and $G_{n-1} = R(a, x')$ for some x' since the last step of the proof is applying the only rule from $\mathbf{A}[\exists y.R(a, y)]$. Let G_k be the first goal in π obtained from G_{k-1} by applying a rule from Equation (2); clearly, $\mathcal{O} \cup \{\exists^* G_{k-1}\} \models G_0$. We have the following cases.
  - Assume that we apply the third rule from Equality (2) to $G_{k-1} = B(b)$ for some constant b (note that a goal $B(x)$ with x a Skolem constant cannot appear by applying QL rules except for Type (3)). Then $G_k = P_S(x, b)$, and the only rule that has P_S in its head is the first one from Equality (2); however, this rule cannot be applied to G_k since we cannot unify b and $c_{A,S}$. Thus, this case is invalid.
  - Assume that we apply the second rule from Equality (2) to $G_{k-1} = S(b, d)$ for some constants b and d. This case is always invalid due to the same reason as the previous one.

- Assume that we apply the third rule from Equality (2) to $G_{k-1} = S(b, x)$ for some constant b and Skolem constant x. Then, $G_k = P_S(b, x)$ and G_{k+1} is obtained from G_k by applying the first rule from Equation (2); that is, $G_{k+1} = A(b)$. But then we have that $A(x) \to \exists y.[S(x, y) \wedge B(y)] \in \mathcal{O}$ and consequently $\mathcal{O} \cup \{A(b)\} \models \exists^* G_{k-1}$. W.l.o.g. we can assume that starting from G_{k+1} rules only from $\mathcal{O}$ are used, which means that $\mathcal{O} \cup \mathbf{A}[\exists y.R(s, y)] \models A(b)$.
- No other case is possible.

Thus $O \cup \{u\} \models \alpha$ which contradicts *(iii)*.

Thus, cens is optimal for $\mathbf{I}$, which concludes the proof.